Virginia Mannori
Università di Bologna, Laurea Magistrale in Scienze criminologiche per l'investigazione e la sicurezza.
Socio ASIS Chapter Italy Student Member (Member ID: 19018041)

La Gestione del Rischio Umano nell'Era dell'Intelligenza Artificiale Generativa

Abstract

L'utilizzo di strumenti di intelligenza artificiale generativa, come i chatbot avanzati, sul luogo di lavoro introduce una nuova categoria di minaccia ibrida. Questi rischi si concretizzano quando un dipendente utilizza queste piattaforme, in modo accidentale o intenzionale, per compromettere asset aziendali critici o alterare processi operativi, spesso in modo poco tracciabile. Tale minaccia, ad oggi, non è solamente basata sulla tecnologia ma si sviluppa anche su fattori umani e organizzativi. Attraverso un approccio criminologico possiamo analizzare come il pericolo venga prodotto attraverso l'unione di tre elementi, ovvero la predisposizione individuale del dipendente definita dal Critical Pathway to Insider Risk (CPIR), l'opportunità offerta dagli strumenti di IA e l'inadeguatezza dei controlli e della cultura aziendale. L'articolo propone che unendo cultura, tecnologia e processi, sia possibile trasformare una minaccia complessa in un'opportunità per rendere le aziende più resilienti, attraverso tre fasi, agendo sia sulla dimensione umana che su quella tecnologica. La prima fase si sviluppa sulla prevenzione del rischio, concentrandosi sulla costruzione di una consapevolezza attraverso politiche chiare e formazione specifica, promuovendo il dipendente come primo attore della sicurezza. La seconda fase è il monitoraggio ibrido, basato sull'osservazione di anomalie comportamentali nell'uso dell'IA e di deviazioni nel comportamento del dipendente stesso. La terza fase definisce una risposta investigativa strutturata, con protocolli per la raccolta delle evidenze digitali e la conduzione di accertamenti finalizzati a comprendere l'evento, contenere i danni e ad apprendere per il futuro. In conclusione, la sfida chiave non risiede nel controllare la tecnologia in sé, ma nell'applicare metodologie di analisi del rischio umano e di investigazione interna a questo nuovo contesto. La prospettiva criminologica si rivela dunque una risorsa fondamentale per sviluppare una difesa aziendale resiliente e adatta alla complessità delle minacce ibride moderne.

Introduzione

Nel contesto attuale della sicurezza informatica, la minaccia che proviene dall'interno si configura come una delle difficoltà maggiori per le aziende. La crescente diffusione di strumenti basati sull'intelligenza artificiale generativa, sempre più potenti e accessibili, allarga di fatto le possibilità per chi già dispone di credenziali e permessi di accedere, modificare o diffondere dati sensibili. A differenza di quanto accade con gli attacchi esterni, in questo caso chi agisce si trova già dentro il perimetro organizzativo e questo rende

oggettivamente complesso capire quando un'attività è regolare e quando invece nasconde un'intenzione dannosa.

Sebbene gli strumenti avanzati di monitoraggio (“advanced monitoring tools”) e le soluzioni basate sull'intelligenza artificiale contribuiscono sempre di più a individuare comportamenti anomali, il fattore umano resta uno degli elementi più critici e spesso trascurati (Bamashmoos, 2025). Le minacce che nascono all'interno possono avere origini diverse, da un lato ci sono quelle intenzionali, mosse magari da un desiderio di guadagno personale o da qualche forma di risentimento nei confronti dell'azienda. Dall’altro lato, ci sono quelle che derivano da errori non voluti, magari perché qualcuno non era abbastanza informato ed ha agito con superficialità, o semplicemente non aveva ricevuto una preparazione adeguata (Zangana et al., 2025).

Proprio perché il fenomeno è così articolato, non basta guardare solo la parte tecnologica ma serve un approccio più ampio, che consideri anche il comportamento delle persone e il modo in cui l'organizzazione è strutturata al suo interno. Le organizzazioni che promuovono una solida cultura della sicurezza e investono nella formazione continua dei dipendenti registrano infatti meno incidenti legati a minacce interne (Zangana et al., 2025). Rimane però una sfida importante ovvero quella di trovare un equilibrio tra monitoraggio efficace, tutela della privacy e capacità di adattarsi ai cambiamenti nei comportamenti degli utenti (Bamashmoos, 2025).

Questo articolo offre una prospettiva criminologica sul fenomeno, mostrando come la combinazione tra caratteristiche individuali, opportunità create dagli strumenti di IA generativa e controlli aziendali insufficienti possa generare nuove forme di rischio ibrido. Il percorso operativo proposto si articola in tre momenti fondamentali, il primo riguarda la prevenzione e punta a ridurre i comportamenti a rischio lavorando sulla consapevolezza delle persone. Il secondo è il monitoraggio, che cerca di cogliere eventuali anomalie nel modo in cui ciascuno utilizza gli strumenti a disposizione. Il terzo, infine, è la fase investigativa ovvero quella in cui si interviene quando qualcosa è già successo.

È in questo contesto che i modelli tradizionali per gestire l'insider threat mostrano un limite che non si può ignorare, sono pensati per cogliere comportamenti fuori dalla norma, mentre con l'intelligenza artificiale generativa anche le azioni quotidiane, quelle che sembrano normali, possono diventare rischiose. A questo si aggiunge un problema di fondo che viene dalla teoria economica, noto come problema principale-agente. Il rapporto tra chi dà le direttive (l’azienda) e chi le esegue (il dipendente) è per natura asimmetrico, perché chi esegue ha sempre maggiori informazioni di chi controlla. In contesti come quello dell’IA generativa, questa asimmetria si allarga ulteriormente e può tradursi in comportamenti non allineati con gli interessi aziendali (Acharyya & Houston, 2025).

Il problema non è più solo riconoscere l'anomalia, ma comprendere come si formano certi comportamenti e intervenire prima che diventino critici. È qui che gli strumenti della criminologia diventano necessari e solo integrando la dimensione umana, quella tecnologica e quella organizzativa si può avere un quadro completo sulle minacce ibride moderne.

Capitolo 1 - Il fenomeno

Quando l'intelligenza artificiale generativa entra negli ambienti di lavoro, il rischio interno assume forme nuove e più difficili da riconoscere. Oggi un dipendente che interagisce con chatbot avanzati può innescare conseguenze rilevanti per la sicurezza, in modo accidentale o deliberato, spesso senza lasciare tracce facilmente individuabili. Diventa quindi necessario integrare strategie di gestione del rischio che tengano conto del fattore umano insieme ai progressi tecnologici (Peca & Turcanu, 2025).

I dati aiutano a comprendere le dimensioni del fenomeno, infatti, il rapporto Ponemon Institute del 2024 segnala un aumento del 26% degli incidenti rispetto al 2022, con costi che superano i 15 milioni di dollari a evento (Bamashmoos, 2025). A questo si aggiunge che quasi un terzo dei dipendenti ha vissuto o è venuto a conoscenza di un episodio di minaccia interna nel proprio contesto lavorativo. La maggior parte di questi episodi è dovuta ad errori non intenzionali, più che a comportamenti dolosi (Zangana et al., 2025). Questi numeri non sono solo statistiche, ma descrivono un fenomeno in crescita che coinvolge trasversalmente organizzazioni di ogni settore, con conseguenze che vanno dalla perdita di dati alla compromissione della reputazione.

Accanto ai numeri, c'è un aspetto meno visibile, ma altrettanto importante. Fattori come lo stress o il burnout aumentano la probabilità che i comportamenti rischiosi possano manifestarsi e questi sono elementi centrali nel percorso che può portare una persona a violare le regole. Non si tratta solo di fragilità individuali, ma anche di un problema di incentivi che non funzionano, quello che la letteratura economica chiama problema principale-agente. L’azienda fa fatica a capire se il dipendente sta usando l’IA per migliorare la sua produttività o per impegnarsi meno quindi, il dipendente, in assenza di controlli efficaci, può avere convenienza a sfruttare questa situazione (Acharyya & Houston, 2025).

Sul versante tecnico, emerge un limite strutturale, ovvero chi agisce dall'interno dispone già di credenziali valide e accessi legittimi rendendo i sistemi di sicurezza tradizionali, nati per difendere il perimetro, inefficaci nel distinguere un'attività regolare da una malevola (Lishchynsky, 2025; Bamashmoos, 2025). A ciò si aggiunge che solo una minoranza dei dipendenti riceve una formazione adeguata, il che rende il quadro ancora più complesso (Zangana et al., 2025).

Casi noti di violazione di dati sensibili da parte di personale interno ci fanno capire che non esiste mai un’unica causa; infatti spesso si è trattato dell'incontro tra fragilità individuali, contesti organizzativi permissivi e misure tecniche inadeguate. In alcuni casi,

l'uso dell'ingegneria sociale ha permesso di ottenere credenziali altrui, mentre in altri, l'assenza di controlli ha consentito il download massiccio di documenti senza che alcun allarme scattasse (Lishchynsky, 2025). Le vulnerabilità si intrecciano e si amplificano a vicenda e la criminologia offre uno sguardo d'insieme necessario per capire cosa accade.

Capitolo 2 - La lente criminologica

Per capire come si presenta oggi la minaccia interna, soprattutto quando entra in gioco l'intelligenza artificiale generativa, non basta guardare alla tecnologia, ma serve uno sguardo diverso, che aiuti a leggere ciò che sta dietro ai comportamenti. La criminologia e in particolare i modelli pensati per studiare le condotte devianti dentro le organizzazioni, possono offrire strumenti utili per interpretare il fenomeno e individuare dove e come intervenire.

Uno dei modelli più noti in questo ambito è il Critical Pathway to Insider Risk (CPIR), il quale aiuta a comprendere come una persona, inizialmente considerata affidabile, possa arrivare a compiere un atto dannoso. Non è un percorso obbligatorio, ma piuttosto un accumularsi di fattori che, se non intercettati, possono portare a conseguenze gravi (Lishchynsky, 2025). Il modello individua quattro elementi che insieme, aumentano il rischio. Il primo riguarda le caratteristiche personali di ciascun individuo, come ad esempio certi tratti del carattere, condizioni psicologiche, una storia di violazione delle regole, difficoltà economiche o scarse capacità relazionali. Il secondo è rappresentato dagli eventi scatenanti, come una brutta valutazione sul lavoro, un conflitto con i colleghi, problemi familiari o finanziari. Il terzo riguarda comportamenti che iniziano a destare preoccupazioni, ovvero le assenze ingiustificate, scarso rendimento, lamentele continue o le provocazioni. L’ultimo elemento, forse il più delicato, è il modo in cui l'organizzazione reagisce a questi segnali. Una risposta sbagliata, troppo dura o distratta, può spingere la persona ancora più giù, mentre un intervento tempestivo e attento può fermare tutto (Lishchynsky, 2025).

Applicare questo modello al contesto dell’IA generativa significa riconoscere che il rischio non nasce dalla tecnologia in sé, ma dall'incontro tra chi la usa, come la usa e il contesto in cui la usa. Da un lato ci sono le persone, con le loro motivazioni e fragilità, dove stress, esaurimento e disattenzione giocano un ruolo importante. Le ragioni che spingono qualcuno a comportarsi in modo rischioso possono essere le più varie; ad esempio desiderio di guadagno, rancore, pressioni esterne, ma anche semplice disinformazione o superficialità (Zangana et al., 2025). Dall'altro lato c'è lo strumento, in questo caso il modello LLM, progettato per essere facile e intuitivo, ma che spesso viene usato male senza che nessuno se ne accorga. Un dipendente che chiede al chatbot di riassumere i documenti riservati, ad esempio, può pensare di star facendo un’azione lecita, ma in realtà potrebbe non essere autorizzato e causare una fuga di informazioni. Il problema è che quando l'uso dell’IA diventa routine, diventa difficile capire cosa sia normale e cosa no (Bamashmoos, 2025).

Un ulteriore argomento è il contesto, cioè le regole, i controlli e la cultura che si respira in un'azienda. Un ambiente tossico, dove ci si sente trattati ingiustamente o lasciati soli, può amplificare i rischi. Al contrario, un'organizzazione che investe sul benessere delle persone e sulla fiducia crea le condizioni in cui i dipendenti stessi diventino parte della soluzione (Lishchynsky, 2025; Zangana et al., 2025). Il vero valore della criminologia, in tutto questo, è che aiuta a tenere insieme i pezzi e non guarda solo alla tecnologia, non si ferma alla persona, non si limita all'organizzazione, ma cerca di capire come tutto si intreccia. Come è stato osservato, i diversi livelli di analisi (individuale, tecnologico e organizzativo), non vivono separati ma le vulnerabilità si alimentano a vicenda e solo guardando all'insieme si può sperare di intervenire in modo efficace (Lishchynsky, 2025). In quest'ottica, partire dal perché le cose accadono può aiutare a costruire strategie più solide e condivise (Peca & Turcanu, 2025).

In sintesi, la prospettiva criminologica offre una chiave di lettura che non si accontenta di descrivere il problema, ma aiuta a progettare interventi capaci di agire sulla persona, sullo strumento e sul contesto. È proprio su questa base che nei prossimi capitoli verranno sviluppate le tre fasi del percorso operativo: prevenzione, monitoraggio e risposta investigativa.

Capitolo 3 - Prevenzione

La prima fase del percorso operativo riguarda la prevenzione, cioè tutto ciò che si può fare per ridurre la possibilità che una minaccia interna si manifesti. Quando si parla di intelligenza artificiale generativa, la prevenzione non può limitarsi a mettere barriere tecniche, ma deve lavorare soprattutto sulle persone e sull’organizzazione, puntando su consapevolezza, responsabilità e regole chiare. Un primo passo fondamentale è definire politiche precise sull'uso dell’IA generativa in azienda. L’arrivo di strumenti come Chat GPT ha cambiato le regole del gioco, oggi anche utenti privi di competenze tecniche possono creare contenuti o automatizzare processi. Questa facilità d'uso è un'opportunità, ma richiede anche confini chiari su cosa si può fare, cosa no e chi è responsabile di cosa (Schmidt et al., 2025).

Accanto alle regole, serve formazione, non quella generica ma una formazione specifica che faccia capire i rischi legati all'uso degli LLM. Si può osservare un fenomeno simile in ambito educativo, dove gli studi mostrano che gli studenti tendono a delegare agli LLM compiti complessi, finendo per saltare i passaggi importanti del loro apprendimento. Lo stesso rischio esiste in azienda, dove un dipendente che si abitua a far fare all’IA i ragionamenti al posto suo, nel tempo perde la capacità di giudizio e diventa lui stesso un punto debole (Shepherd, 2025).

I dati mostrano che solo una piccola parte dei dipendenti riceve una preparazione regolare in materia di cybersecurity e meno della metà rispetta in modo costante le politiche

di sicurezza (Zangana et al., 2025). Per questo è importante che le persone capiscano non solo cosa devono fare, ma perché le stanno facendo. Infatti, partire dal senso delle cose aiuta a creare coinvolgimento e responsabilità (Peca & Turcanu, 2025). Questo è esattamente il cuore dell'idea del dipendente come primo attore della sicurezza, non si tratta di imporre regole dall'alto, ma di costruire un ambiente in cui ciascuno si senta parte della soluzione. Una cultura positiva, che valorizza le persone e le sostiene, favorisce la lealtà e la collaborazione, chi si sente trattato bene è più propenso a seguire le regole e a segnalare eventuali problemi (Lishchynsky, 2025). I dati confermano che le organizzazioni con una buona cultura della sicurezza registrano meno incidenti (Zangana et al., 2025). In questo senso, la prevenzione agisce su due livelli del modello CPIR; da un lato riduce le fragilità individuali (primo elemento) attraverso formazione e supporto; dall'altro migliora la qualità della risposta organizzativa (quarto elemento) affinché i segnali deboli vengano colti e gestiti prima che si trasformino in qualcosa di più grave.

Infine, il tema della comunicazione, dove gli strumenti di IA sono pensati per essere semplici e accessibili e questo è senza dubbio un vantaggio. Tuttavia, se non si illustrano con chiarezza quali sono i rischi e le modalità d'uso, quella stessa semplicità può trasformarsi in una trappola. Le regole devono essere scritte in modo comprensibile, facilmente reperibili e richiamate con regolarità (Schmidt et al., 2025). Per quanto efficace, la prevenzione da sola non basta ma occorre anche intercettare tempestivamente i comportamenti che sfuggono alle regole o emergono in modo inatteso. È qui che entra in gioco la seconda fase del percorso, ovvero il monitoraggio ibrido.

Capitolo 4 - Monitoraggio ibrido

La seconda fase del percorso operativo è il monitoraggio ibrido, cioè l'insieme degli strumenti e delle attività che servono a intercettare per tempo i comportamenti anomali, prima che si trasformino in incidenti. In un contesto dove l'IA generativa è ormai diffusa, non basta osservare solo le variabili tecniche, ma serve uno sguardo più ampio che tenga conto anche del comportamento delle persone. I sistemi tradizionali faticano a distinguere tra le attività reali e quelle sintetiche. L'IA generativa, infatti, consente di creare profili utente così convincenti da risultare autentici, rendendo sempre più difficile l'individuazione delle potenziali minacce (Kotb et al., 2025).

Per rispondere a questa sfida, alcuni studi propongono modelli di deep learning in grado di riconoscere le differenze tra profili reali e sintetici, con risultati che superano il 99% di accuratezza in condizioni ottimali (Kotb et al., 2025). Altri approcci si basano sui sistemi multi-agente, in cui agenti specializzati analizzano diverse categorie di log e producono report dettagliati, che vengono successivamente valutati da un agente supervisore per arrivare a una classificazione finale. I risultati mostrano che questi modelli gestiscono efficacemente situazioni complesse e si adattano a scenari diversi (Ferraro et al., 2025).

Tale architettura distribuisce l'analisi su più livelli, riducendo il margine di errore e garantendo la tracciabilità dell'intero processo.

Un elemento centrale che emerge da entrambi gli studi è l'importanza della "baseline comportamentale" (Bamashmoos, 2025). Per capire se un comportamento è anomalo, è necessario conoscere prima cosa sia normale per quella specifica persona. I sistemi più efficaci sono quelli che imparano nel tempo le abitudini di ciascun utente e segnalano solo le deviazioni significative (Bamashmoos, 2025). In chiave CPIR, è in questa fase che diventa operativo il terzo elemento del modello dove accessi insoliti, richieste anomale agli strumenti di IA o deviazioni delle abitudini consolidate possono essere intercettati proprio perché il sistema riconosce ciò che si discosta dalla norma.

Un ulteriore contributo viene dall'analisi dei big data. La sfida nel monitorare l'IA generativa è soprattutto quantitativa. L'interazione continua tra dipendenti e LLM genera una mole di dati talmente vasta da rendere obsoleti i sistemi di analisi tradizionali. Analizzare in tempo reale grandi quantità di dati provenienti da fonti diverse, come log di sistema, tracce di accesso e comunicazioni, permette di individuare pattern sospetti che altrimenti resterebbero nascosti. Confrontare i nuovi dati con quelli storici aiuta a ridurre i falsi positivi, perché il sistema impara a riconoscere meglio ciò che è davvero anomalo (Yedalla, 2025). L'analisi dei big è il presupposto per rendere operative le baseline comportamentali su larga scala.

Da queste prospettive emergono tre pilastri su cui costruire un monitoraggio efficace. Il primo è la personalizzazione, ogni persona ha le sue abitudini e il sistema deve imparare a riconoscerle. Il secondo è l'integrazione, cioè più fonti vengono incrociate, più il quadro risulta affidabile. Il terzo è l'architettura multi-agente, che consente di distribuire l'analisi su più livelli e sintetizzare le informazioni in una valutazione complessiva. Il monitoraggio non ha lo scopo di sostituire i professionisti della sicurezza, ma di fornire loro strumenti più efficaci. Approcci come il ragionamento "Chain-of-Thought" permettono di ottenere report che non si limitano a segnalare un'anomalia, ma ne spiegano le ragioni, gli elementi che l'hanno generata e gli indicatori emessi nel processo. Questa trasparenza è fondamentale per consentire a chi deve prendere decisioni di farlo in modo consapevole e informato (Ferraro et al., 2025). La tecnologia supporta il giudizio umano, non lo sostituisce.

In sintesi, il monitoraggio ibrido è il cuore operativo del percorso proposto dove la sua efficacia dipende dalla capacità di personalizzare le baseline, integrare fonti diverse, usare architetture distribuite e preservare uno spazio per il giudizio umano. Non si tratta di sorveglianza, ma della capacità di ascoltare i segnali che il sistema già produce, prima che diventino troppo forti per essere ignorati. Quando uno di questi segnali si rivela una minaccia concreta, si passa alla terza fase, cioè la risposta investigativa, che ha il compito di gestire l'incidente in modo ordinato e di trarne insegnamento.

Capitolo 5 - Risposta investigativa

La terza fase del percorso operativo è la risposta investigativa, ovvero l'insieme delle azioni che si attivano nel momento in cui una potenziale minaccia interna viene individuata. Se la prevenzione mira ad evitare che l'evento si verifichi e il monitoraggio cerca di intercettarlo tempestivamente, la risposta investigativa ha il compito di gestire l'incidente in modo ordinato andando a limitare i danni, raccogliendo prove utilizzabili e, soprattutto, trarre insegnamento da quanto successo per migliorare in futuro.

Oggi la risposta agli incidenti non può basarsi solo su procedure manuali e reattive, ma servono strumenti intelligenti, capaci di adattarsi in tempo reale. Un esempio è il framework ARCS (Adaptive Reinforcement Learning Framework for Automated Cybersecurity Incident Response Strategy Optimization), che propone un'architettura a due livelli, uno dedicato alla gestione immediata della minaccia e l'altro alla pianificazione strategica. I risultati mostrano che questo approccio riduce i tempi di risoluzione e aumenta l'efficacia difensiva, soprattutto negli attacchi complessi che si sviluppano in più fasi (Ren et al., 2025).

Un aspetto centrale della risposta è la capacità di raccogliere e conservare le evidenze digitali correttamente. Per fare ciò servono architetture dati ben progettate, in grado di integrare fonti diverse (log, sensori, comportamento degli utenti) e garantire al tempo stesso sicurezza e conformità alle normative. La protezione dei dati in transito e in archivio, l'uso di firewall e sistemi di rilevamento e la solidità delle comunicazioni sono elementi essenziali per mantenere integra la catena di custodia. Un altro aspetto cruciale riguarda la gestione degli accessi durante la fase investigativa, infatti non tutti possono vedere tutto, ed è quindi necessario definire ruoli chiari e ben definiti e meccanismi di autenticazione robusti, così da garantire che solo il personale autorizzato possa accedere alle informazioni sensibili (Shaffi & Sidhick, 2025).

Una prospettiva interessante arriva dall'uso di sistemi multi-agente basati su Large Language models. In contesti simulati, questi agenti si sono dimostrati capaci di collaborare in modo efficace, adattandosi a strutture di team diverse, centralizzate, decentralizzate e ibride e mostrando che un gruppo equilibrato, con differenti livelli di esperienza, può risultare più performante di uno composto esclusivamente da esperti (Liu, 2025). Anche la fase di contenimento e recupero trae vantaggio dall'integrazione tra automazione e giudizio umano. Sistemi come ARCS, grazie a meccanismi di apprendimento personalizzati, mettono in campo contromisure rapide che riducono in modo significativo i falsi positivi. In alcuni scenari reali, come nel caso di attacchi ransomware, questi sistemi hanno dimostrato di poter proteggere la grande parte di dispositivi coinvolti, contenendo la diffusione del malware e ripristinando quelli compromessi (Ren et al., 2025).

Infine, un aspetto fondamentale ma spesso trascurato è quello di imparare dall'incidente. Il reporting automatizzato permette di produrre analisi dettagliate su cosa è accaduto, perché è accaduto e come evitare che si possa ripetere. Nel tempo, questi report possono individuare problemi ricorrenti o vulnerabilità nascoste, trasformando ogni episodio in un'occasione per migliorare (Shaffi & Sidhick, 2025). In questa prospettiva, la fase investigativa non si esaurisce nella gestione tecnica dell'incidente, ma comprende anche la comprensione degli eventi scatenanti che, secondo il CPIR, hanno preceduto il comportamento dannoso. Capire se alla base c'era un conflitto irrisolto, una pressione esterna o una fragilità ignorata permette di intervenire in modo più mirato, correggendo non solo le vulnerabilità tecniche, ma anche quelle organizzative e umane. Si tratta, in altri termini, di un'analisi propriamente criminologica, cioè risalire alle cause profonde del comportamento per agire alla radice e non limitarsi a gestire gli effetti. È proprio questa capacità di adattarsi e apprendere in modo continuo a rendere un sistema davvero resiliente (Ren et al., 2025).

Una risposta investigativa efficace richiede automazione intelligente, architetture dati solide, collaborazione tra persone e strumenti e una chiara definizione dei ruoli. L'obiettivo non è solo gestire l'emergenza, ma imparare da essa, passando da una logica reattiva a una proattiva. La risposta investigativa non è il punto finale del percorso, bensì il momento in cui l'organizzazione decide se reagire o imparare da quello che è successo. Su questa capacità di apprendere si fonda la resilienza.

Conclusioni

Quanto emerso nei capitoli precedenti mostra che le minacce interne, sono oggi tutt'altro che semplici da inquadrare, specialmente con l'arrivo dell'intelligenza artificiale generativa il fenomeno è diventato ancora più complesso. Non si tratta mai di una sola causa isolata, tecnica o personale che sia, ma di un intreccio di elementi diversi, come il modo in cui le persone reagiscono a determinate situazioni, gli strumenti che hanno a disposizione e l'ambiente in cui lavorano.

Adottare la prospettiva criminologica ha permesso di superare una visione troppo riduttiva del problema, anziché di concentrarsi esclusivamente sulla tecnologia o sul comportamento individuale, si è cercato di capire come questi elementi si combinano tra loro. La connessione tra individuo, strumenti e contesto, aiuta a mettere in ordine e a pensare a interventi più mirati. A questo si aggiunge una prospettiva complementare, quella economico-organizzativa, che aiuta a vedere il rischio interno anche come un problema principale-agente, ovvero come il risultato di incentivi disallineati tra chi dirige e chi esegue. La criminologia non si limita a descrivere il rischio, ma offre strumenti concreti per intervenire su ciascuna delle sue fasi: dalla prevenzione alla risposta investigativa, fino all'apprendimento organizzativo.

Il percorso proposto, articolato in tre fasi, traduce tutto questo in qualcosa di concreto. La prevenzione richiede di agire sulle persone e sul clima che si respira in azienda, non solamente di definire regole. Il monitoraggio dimostra che oggi è possibile rendere visibile ciò che un tempo era difficile da tracciare, a patto di usare strumenti capaci di incrociare fonti eterogenee e di adattarsi a ciascun contesto specifico. La risposta investigativa, infine, è il momento in cui si vede se quello che è stato preparato funziona davvero, ovvero, raccogliere le prove, gestire l'emergenza e imparare dall'accaduto. Ciò che tiene insieme queste tre fasi è la logica sottostante al modello CPIR.

Ciò che resta sullo sfondo, ma rappresenta forse l'elemento più importante, è un cambiamento di prospettiva; nell'era degli LLM il rischio interno non è più un evento eccezionale da individuare, ma un comportamento quotidiano da interpretare. I modelli tradizionali di controllo, progettati per intercettare anomalie, rischiano di rivelarsi insufficienti in un contesto in cui anche l'uso apparentemente legittimo degli strumenti può generare vulnerabilità. È in questo scenario che la criminologia smette di essere un supporto accessorio per diventare uno strumento necessario, non solo per capire cosa è successo, ma per comprendere perché le persone agiscono in un certo modo e costruire così ambienti più sicuri e consapevoli.

La vera sfida non è esclusivamente tecnologica né esclusivamente culturale ma la capacità di tenere insieme entrambe le dimensioni. Costruire organizzazioni capaci di riconoscere i segnali deboli, rispondere in modo proporzionato e apprendere da ogni episodio richiede un cambiamento di mentalità profondo. Le tecnologie continueranno ad evolversi, portando con sé nuove opportunità e nuovi rischi, quindi mettere insieme cultura, tecnologia e processi è l'unico modo per far sì che la complessità non si trasformi in un ostacolo, ma diventi una risorsa. Organizzazioni che non si limitano a resistere, ma apprendono da ciò che accade e si rafforzano un passo alla volta.

Bibliografia

Acharyya, M., & Houston J. (2025). *Strategic Interaction in Cyber Risk Governance: A Four-Agent Game-Theoretic and Institutional Framework.* Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5481107.

Bamashmoos, F. (2025). *Adaptive Privacy-Preserving Insider Threat Detection Using Generative Sequence Models*. Available at https://www.mdpi.com/1999-5903/18/1/11.

Ferraro, A., Orlando, G., &Russo, D. (2025). *Generative Agent-Based Modeling with Large Language Models for insider threat detection.* Available at https://www.sciencedirect.com/science/article/abs/pii/S0952197625013454?casa_token=o8nk6Mnd4uoAAAAA:uZQwGkEIl4G1FdFHoHco5SkNQp4fO6lVslbKVVNvTNjcOIGGmbG3YHrFwRn6jTasW5g7T4yjMg.

Kobt, H., Gaber, T., AlJanah, S., Zawbaa H. M., & Alkhathami, M. (2025). *A novel deep synthesis-based insider intrusion detection (DS-IID) model for malicious insiders and AI-generated threats.* Available at https://www.nature.com/articles/s41598-024-84673-w.

Liu, Z., (2025). *AutoBnB: Multi-Agent Incident Response with Large Language Models.* Available at https://ieeexplore.ieee.org/abstract/document/11012055?casa_token=GMb_8khEb1gAAAAA:GMdTTq8lmmuFDtPdjKWCa3-3-xsnuteywdqKa7PHNlFNlZ18Ad11lX4EcWXVWEelcm-_JXJw.

Lishchynsky, M. (2025). *The insider threat: A socio-technical analysis of preventing data breaches and espionage within governmental agencies.* Available at https://politics-security.net/index.php/ojsdata/article/view/284/270.

Peca, L., & Turcanu, D. (2025). *Reducing cyber risk through a human–centred approach.* Available at https://ibn.idsi.md/vizualizare_articol/229306.

Ren, S., Jin, J., Niu, G., & Liu, Y. (2025). *ARCS: Adaptive Reinforcement Learning Framework for Automated Cybersecurity* Incident Response Strategy Optimization. Available at https://www.mdpi.com/2076-3417/15/2/951.

Schmidt, R., Alt, R., & Zimmermann, A. (2025). *Characteristics of Platform Shifts - A Single Case Study of ChatGPT.* Available at https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/cd9a099f-25d4-4acf-a32d-748b19f464a6/content.

Shaffi, S. M., & Sidhick, J. N. (2025). *Real-time incident reporting and intelligence framework: Data architecture strategies for secure and compliant decision support.* 110–118. Available at https://www.researchgate.net/profile/Shamnad-Mohamed-Shaffi-2/publication/392406974_Real-time_incident_reporting_and_intelligence_framework_Data_architecture_strategies_for_secure_and_compliant_decision_support/links/684116008a76251f22eb9dc1/Real-time-incident-reporting-and-intelligence-framework-Data-architecture-strategies-for-secure-and-compliant-decision-support.pdf.

Shepherd, C. (2025). *Generative AI Misuse Potential in Cyber Security Education: A Case Study of a UK Degree Program*. Available at https://arxiv.org/pdf/2501.12883.

Yedalla, J. (2025). *Unmasking insider threats: How big data analytics is revolutionizing cybersecurity defense.* International Research Journal of Modernization in Engineering, Technology and Science. Available at https://www.researchgate.net/profile/Jayasudha-Yedalla/publication/392726801_UNMASKING_INSIDER_THREATS_HOW_BIG_DATA_ANALYTICS_IS_REVOLUTIONIZING_CYBERSECURITY_DEFENSE/links/6850272c7869fe75c5596d93/UNMASKING-INSIDER-THREATS-HOW-BIG-DATA-ANALYTICS-IS-REVOLUTIONIZING-CYBERSECURITY-DEFENSE.pdf.

Zangana, H. M., Sallow, Z. B., & Omar, M. (2025). *The Human Factor in Cybersecurity: Addressing the Risks of Insider Threats.* Jurnal Ilmiah computer science (JICS). Available at https://www.ejurnal.snn-media.com/index.php/jics/article/view/37/37.