AIPIA

ASSOCIAZIONE ITALIANA PROFESSIONISTI DELL'INTELLIGENZA ARTIFICIALE

GUIDA AIPIA · RISORSE PER PROFESSIONISTI

Sicurezza e Intelligenza Artificiale

Proteggere dati, persone e sistemi quando si usa l'AI: i rischi da conoscere e le difese che funzionano.

A cura del **Comitato Tecnico-Scientifico AIPIA**

Versione 1.0 — Giugno 2026

EUROPEAN AI ALLIANCE

ROME CALL FOR AI ETHICS

EUROPEAN DIGITAL CREDENTIALS · EIDAS

AI ACT READY

Documento realizzato da AIPIA per i propri soci e per la community.

Tutti i materiali: aipia.it/risorse-utili

Indice

Il tema della **sicurezza e AI** ha due facce che vanno tenute insieme. Da un lato l'intelligenza artificiale introduce rischi nuovi: fuga di dati verso strumenti pubblici, manipolazione degli input, account e fornitori fuori controllo, contenuti falsi che alimentano truffe. Dall'altro la stessa tecnologia, usata bene, rafforza la difesa: accelera il rilevamento degli attacchi, filtra il phishing, aiuta chi presidia i sistemi. Questa guida spiega entrambi i lati e, soprattutto, come proteggere dati, persone e sistemi quando l'AI entra nel lavoro quotidiano.

Vuoi ricevere le prossime guide AIPIA appena escono? **Iscriviti alla newsletter** o **diventa socio** e accedi a tutti i materiali.

PARTE I

I due lati del tema

Prima di entrare nei rischi e nelle difese, serve un quadro onesto: cosa cambia davvero per la sicurezza quando arriva l'AI, dove finiscono i rischi d'uso e dove iniziano quelli tecnici, e quali norme disegnano il campo da gioco.

Sicurezza e AI: i rischi che porta e l'aiuto che può dare

Una banca usa un sistema di intelligenza artificiale per riconoscere transazioni fraudolente in tempo reale, e blocca in pochi millisecondi operazioni che un analista umano non avrebbe mai visto in tempo. La settimana dopo, un dipendente della stessa banca incolla un elenco di clienti in un assistente gratuito per farsi scrivere una mail, e quei dati escono dal perimetro aziendale. Stessa tecnologia, stessa organizzazione, due storie opposte.

Questo è il punto da cui partire. L'AI non è né uno strumento di sicurezza né una minaccia: è entrambe le cose, a seconda di come la si usa e di chi la usa. Trattarla solo come pericolo porta a vietarla e a perdere i suoi vantaggi difensivi. Trattarla solo come alleata porta a sottovalutare i rischi che introduce. Una postura seria tiene insieme i due lati.

Sul lato dei rischi, l'AI generativa ha allargato la superficie esposta in modi che i controlli tradizionali faticano a vedere. I dati escono attraverso il copia e incolla. Gli attacchi fanno leva sul linguaggio naturale invece che sul codice. I contenuti falsi diventano credibili e a basso costo. Gli strumenti si collegano l'uno all'altro creando dipendenze che nessuno ha censito.

Sul lato dell'aiuto, l'AI lavora su scala e velocità che agli esseri umani sono precluse. Analizza enormi quantità di log, riconosce schemi anomali, smista segnalazioni, assiste chi risponde a un incidente, riassume informazioni di minaccia. Non sostituisce gli specialisti, ma moltiplica ciò che possono coprire.

> **In sintesi.** La sicurezza dell'AI non è una scelta tra usarla e vietarla, ma tra usarla con metodo o subirla senza governo. I capitoli che seguono servono a costruire quel metodo.

Conviene anche sgomberare il campo da una falsa scelta. C'è chi presenta il tema come "AI sì o AI no", come se vietarla mettesse al riparo. Non è così: chi vieta perde i vantaggi difensivi e non elimina i rischi, perché le persone continuano a usarla altrove, fuori dal controllo. La domanda utile non è se usare l'AI, ma come usarla in modo che i suoi vantaggi superino i suoi rischi, e questo dipende quasi interamente da decisioni organizzative, non dalla tecnologia in sé.

Una precisazione utile fin da subito: questa è una guida di sensibilizzazione, non un manuale operativo per attaccare i sistemi. Spiega i rischi perché conoscerli è il primo passo per difendersi, e si ferma dove l'informazione diventerebbe istruzione per fare danno.

Tre fraintendimenti da abbandonare subito

Buona parte delle difese mal riuscite parte da un'idea sbagliata di cosa sia la sicurezza dell'AI. Tre fraintendimenti, in particolare, tornano di continuo.

Il primo: "la sicurezza è un problema dell'IT". Lo è in parte, ma la maggioranza degli incidenti passa dalle persone, non dai server. Un reparto tecnico bravissimo non impedisce a un collega del marketing di incollare dati dove non dovrebbe.

Il secondo: "basta comprare lo strumento giusto". Nessun prodotto rende sicura un'organizzazione le cui persone non sono formate e i cui processi non esistono. La tecnologia è necessaria, non sufficiente.

Il terzo: "se non abbiamo dati segreti, non ci riguarda". Quasi ogni organizzazione tratta dati personali di clienti e dipendenti, e quei dati bastano a far scattare obblighi di legge e danni reputazionali. La sicurezza dell'AI riguarda chiunque usi questi strumenti, non solo chi custodisce segreti industriali.

Sgombrato il campo da queste idee, il resto della guida diventa più facile da leggere: non come una serie di prodotti da acquistare, ma come un insieme di decisioni da prendere e abitudini da costruire.

Perché l'AI cambia la superficie di rischio di un'organizzazione

La superficie di rischio è l'insieme dei punti attraverso cui un'organizzazione può essere colpita o può perdere informazioni. Per anni quella superficie è stata fatta di server, reti, dispositivi, applicazioni. L'AI ne aggiunge una parte nuova, fatta di dati che entrano nei modelli, di interazioni in linguaggio naturale e di automazioni che agiscono al posto delle persone.

Tre cambiamenti meritano attenzione, perché spiegano perché i vecchi controlli non bastano.

Il primo è la porosità del confine. Un assistente generativo gira nel browser, spesso fuori dal controllo dei sistemi aziendali. Il dato che lo raggiunge non passa per i percorsi sorvegliati: esce dal lato meno presidiato, quello delle persone che lavorano.

Il secondo è la natura linguistica delle interazioni. Un sistema di AI moderno non esegue solo comandi precisi: interpreta richieste scritte in italiano, e in quell'interpretazione si annida una vulnerabilità nuova, perché le istruzioni dannose possono nascondersi in un testo all'apparenza innocuo.

Il terzo è l'autonomia crescente. Più gli strumenti agiscono da soli (leggono la posta, accedono ai file, compiono azioni), più un errore o una manipolazione si propagano senza che nessuno intervenga in tempo.

C'è poi un effetto trasversale: la velocità con cui tutto questo cambia. Gli strumenti si aggiornano ogni pochi mesi, le funzioni si moltiplicano, le integrazioni si attivano con un clic. La superficie di rischio non è più una mappa stabile da presidiare, ma un territorio che si ridisegna di continuo, e questo impone un governo continuo invece di un controllo una tantum.

> *La sicurezza non è uno stato che si raggiunge, ma una manutenzione che non finisce. Con l'AI questo è ancora più vero.*
>
> PRINCIPIO GUIDA DEI CORSI AIPIA

La superficie che le persone aggiungono

C'è una parte della superficie di rischio che nessun firewall presidia: quella che ogni persona porta con sé. Un dipendente con un telefono in tasca e un account personale su un servizio di AI è, da solo, un punto di contatto tra i dati dell'organizzazione e il mondo esterno. Moltiplicato per il numero di persone, questo diventa un perimetro fatto di abitudini, non di macchine.

È un cambiamento di natura, non solo di dimensione. Per anni la sicurezza ha potuto concentrarsi su ciò che si poteva controllare tecnicamente: i sistemi di proprietà dell'azienda. L'AI sposta una quota del rischio su comportamenti che i sistemi non vedono, e che si governano con la cultura e le regole più che con la configurazione.

Questo non rende inutili le difese tecniche. Le rende insufficienti da sole. Un'organizzazione che investe solo in tecnologia e nulla in formazione assomiglia a una casa con una porta blindata e le finestre aperte: la spesa è concentrata dove il rischio è già basso, e trascurata dove è alto.

Quattro punti d'ingresso che prima non c'erano

Per rendere concreta l'idea di superficie allargata, conviene elencare i punti d'ingresso che l'AI ha aggiunto e che dieci anni fa, semplicemente, non esistevano in questa forma.

- **L'interfaccia in linguaggio naturale**: la casella dove si scrive all'assistente è un canale verso l'esterno che non passa per i controlli tradizionali.
- **I contenuti che il modello legge**: documenti, pagine, mail che l'AI elabora possono contenere istruzioni ostili nascoste.
- **Le integrazioni**: ogni connettore tra l'assistente e un sistema interno è una porta in più, spesso aperta senza valutazione.
- **I contenuti generati**: testi, voci e video falsi prodotti a basso costo che alimentano truffe verso le persone.

Nessuno di questi punti si chiude con un solo prodotto. Il primo e il quarto si presidiano soprattutto con la consapevolezza delle persone; il secondo e il terzo con scelte tecniche e di configurazione. È la ragione per cui questa guida tiene insieme, in ogni capitolo, il lato umano e il lato tecnico: lasciarne fuori uno significa lasciare aperta metà della superficie.

La differenza tra rischi d'uso e rischi tecnici

Confondere i due tipi di rischio è l'errore che porta più spesso a difese sbilanciate: tutta l'attenzione sugli aspetti tecnici, quasi nessuna sui comportamenti delle persone, o viceversa. Distinguere serve proprio a non lasciare scoperto un intero fronte.

I rischi d'uso nascono da come le persone adoperano l'AI. Sono il dato incollato dove non dovrebbe, l'output accettato senza verifica, l'account personale usato per lavoro, lo strumento adottato senza che nessuno lo sappia. Non richiedono un attaccante: bastano la fretta e la mancanza di regole chiare.

I rischi tecnici nascono dal funzionamento dei sistemi e dalla presenza di chi vuole abusarne. Sono la manipolazione degli input, le integrazioni vulnerabili, i fornitori compromessi, i contenuti falsi creati per ingannare. Qui c'è quasi sempre un'intenzione ostile, o almeno una falla da chiudere.

La distinzione conta perché le contromisure sono diverse. I rischi d'uso si riducono soprattutto con regole, formazione e strumenti comodi. I rischi tecnici si riducono con configurazioni, controlli, valutazione dei fornitori e sorveglianza. Servono entrambe, e quasi sempre l'anello debole è il primo, non il secondo.

ASPETTO	RISCHI D'USO	RISCHI TECNICI
Origine	Comportamento delle persone	Funzionamento dei sistemi e attaccanti
Esempi	Dati incollati, output non verificati, account personali	Prompt injection, integrazioni vulnerabili, fornitori compromessi
Servono	Quasi mai un attaccante	Quasi sempre un'intenzione ostile o una falla
Difesa principale	Regole, formazione, strumenti comodi	Configurazioni, controlli, sorveglianza, valutazione fornitori
Anello debole tipico	Le persone non informate	Le integrazioni attivate senza valutazione

Un'organizzazione matura lavora sui due fronti in parallelo. Una immatura ne presidia uno solo e si illude di essere protetta, mentre resta esposta dall'altro. La parte tecnica dà un falso senso di sicurezza se le persone restano senza istruzioni, ed è la situazione più comune.

Perché proprio il fronte delle persone resta scoperto più spesso? Per una ragione semplice: la parte tecnica si compra e si delega, quella umana si costruisce nel tempo. Davanti a un problema, comprare uno strumento dà una soddisfazione immediata e visibile; formare le persone richiede mesi e i risultati non si vedono subito. Così le organizzazioni tendono a spendere dove il ritorno è rapido e tangibile, lasciando indietro proprio l'area da cui passa la maggior parte degli incidenti.

Un esempio chiarisce la differenza. La stessa fuga di dati può nascere da un rischio d'uso (un dipendente incolla un file in uno strumento pubblico) o da un rischio tecnico (un'integrazione vulnerabile estrae dati senza che nessuno lo voglia). Il danno finale è simile, ma le difese sono diverse: nel primo caso servono regole e formazione, nel secondo una valutazione tecnica e un controllo dei permessi. Confondere i due casi porta a curare il sintomo sbagliato.

Il quadro: GDPR, AI Act, ACN e NIS2

La sicurezza dell'AI non si gioca in un vuoto normativo. Diverse regole, nate per scopi diversi, si intrecciano e impongono obblighi concreti a chi tratta dati e gestisce sistemi. Conoscerle a grandi linee serve a capire cosa è dovuto, non solo cosa è prudente.

Il primo riferimento è il Regolamento UE 2016/679, il GDPR. Quando l'AI tratta dati personali, valgono i principi di sempre: base giuridica, minimizzazione, sicurezza del trattamento, gestione delle violazioni. Una fuga di dati personali verso uno strumento di AI è una violazione come le altre, con gli stessi obblighi di valutazione ed eventuale notifica.

Il secondo è il Regolamento UE 2024/1689, l'AI Act, che disciplina i sistemi di intelligenza artificiale in base al rischio. Tra i suoi obblighi, fin dalle prime fasi di applicazione, c'è quello di garantire un livello adeguato di alfabetizzazione del personale che usa questi sistemi. La sicurezza, qui, passa per la competenza delle persone.

Sul fronte della cybersicurezza, in Italia opera l'ACN, l'Agenzia per la Cybersicurezza Nazionale, punto di riferimento per indirizzi, allerte e coordinamento. A livello europeo, la direttiva NIS2 ha alzato gli obblighi di sicurezza per molte organizzazioni, soprattutto quelle che operano in settori considerati essenziali o importanti, estendendo doveri di gestione del rischio e di segnalazione degli incidenti.

> **Attenzione.** Questa guida è divulgativa e non sostituisce una consulenza legale o tecnica specialistica. I riferimenti normativi servono a orientarsi; per gli obblighi specifici del proprio settore, soprattutto sotto NIS2, va coinvolto chi ha la competenza adeguata.

Il messaggio trasversale di queste norme è coerente: la sicurezza non è un optional tecnico, ma un dovere che combina misure tecniche, organizzazione e competenze. Chi affronta la sicurezza dell'AI solo come questione di strumenti perde di vista metà di ciò che la legge, oltre al buon senso, richiede. Per orientarsi tra gli adempimenti dell'AI Act, l'associazione mette a disposizione una guida dedicata al regolamento.

Cosa cambia in pratica con NIS2

NIS2 merita un'attenzione particolare, perché ha ampliato la platea delle organizzazioni soggette a obblighi di cybersicurezza e ha alzato l'asticella. Senza entrare nei dettagli, che variano per settore e dimensione, alcuni principi pratici valgono per molti.

La sicurezza diventa una responsabilità documentata della direzione, non delegabile interamente a un fornitore. Gli incidenti significativi vanno gestiti e segnalati secondo tempi definiti. La gestione del rischio deve includere anche i fornitori, perché la catena di approvvigionamento è esplicitamente parte del perimetro da proteggere. E la formazione del personale rientra tra le misure attese, non tra gli optional.

Per chi rientra nell'ambito di NIS2, l'uso non governato dell'AI è un problema diretto di conformità: introduce rischi sulla catena di fornitura e sui dati che la direzione ha il dovere di gestire e documentare. Far emergere e regolare l'uso non è quindi solo prudenza, ma adempimento.

> **In sintesi.** GDPR, AI Act e NIS2 convergono su un punto: la sicurezza dell'AI è una responsabilità della direzione, fatta di misure tecniche, processi e competenze documentate. Nessuna delle tre si soddisfa comprando un solo prodotto.

PARTE II

I rischi introdotti dall'AI

Cinque famiglie di rischio che prima non esistevano o non avevano questa portata: i dati che escono, gli input manipolati, gli account e le integrazioni, i fornitori, e i contenuti falsi che alimentano truffe sempre più credibili.

Fuga di dati: cosa esce quando si usa uno strumento pubblico

Quando una persona incolla un testo in un assistente generativo gratuito, quel testo lascia fisicamente l'organizzazione. Viaggia in rete, arriva sui server del fornitore, viene elaborato e, a seconda delle condizioni del servizio, può essere conservato, riletto da personale tecnico o usato per addestrare versioni future del modello.

La schermata di un chatbot somiglia a una conversazione privata. Non lo è. È più vicina a una cartolina: chi gestisce il servizio, in linea di principio, può leggerla.

La fuga di dati attraverso l'AI ha tre caratteristiche che la rendono particolarmente insidiosa rispetto alle fughe tradizionali.

È silenziosa: non scatta nessun allarme, perché dal punto di vista dei sistemi è solo traffico web verso un sito legittimo. È volontaria ma inconsapevole: nessuno ruba nulla, è la persona stessa che invia il dato, convinta di fare una cosa innocua. Ed è irreversibile: una volta inviata, l'informazione non si recupera. Non esiste un tasto annulla affidabile.

Cosa esce, in concreto? Quasi sempre le stesse categorie: dati di clienti incollati per scrivere risposte, listini e offerte per generare documenti commerciali, codice sorgente per farsi aiutare a programmare, bozze di contratti e documenti riservati per ottenerne sintesi. Nessuno di questi gesti nasce da cattiva volontà; tutti producono una perdita di controllo.

> **Esempio.** Un addetto amministrativo, per velocizzare un report, incolla in uno strumento pubblico un foglio con nomi, importi e codici fiscali di alcuni fornitori. In trenta secondi ha risparmiato mezz'ora di lavoro e ha trasferito dati personali a un servizio esterno senza base giuridica, senza informativa e senza alcun contratto. È una violazione di dati personali a tutti gli effetti, anche se nessuno la noterà mai.

La differenza tra una versione gratuita e una versione aziendale dello stesso strumento, per gli stessi dati, è enorme: la seconda di solito garantisce per contratto che i dati non vengano usati per l'addestramento e non siano conservati oltre il necessario. Quasi nessun utente conosce questa differenza, ed è esattamente il punto su cui la formazione deve insistere.

Tre destini diversi del dato che esce

Quando si dice che un dato "viene usato" dal fornitore, in realtà si parla di tre cose distinte, con rischi diversi, e tenerle separate aiuta a fare le domande giuste.

La conservazione: il fornitore tiene una copia di quanto inserito, per un periodo definito. Finché esiste su un server esterno, quel dato è esposto a errori di configurazione e accessi non previsti.

L'addestramento: il contenuto entra nel materiale con cui i modelli futuri vengono perfezionati. È il destino che più preoccupa, perché in teoria un'informazione potrebbe riaffiorare, trasformata, in risposte date ad altri.

La revisione umana: alcuni servizi prevedono che personale legga campioni di conversazioni per controllarne qualità e sicurezza. Significa che occhi esterni all'organizzazione possono, in linea di principio, vedere ciò che è stato scritto.

I servizi pensati per le aziende limitano o escludono per contratto questi tre destini. Le versioni gratuite spesso li prevedono come impostazione predefinita. La verifica da fare prima di approvare uno strumento è proprio questa: cosa succede, esattamente, ai dati che gli affidiamo.

> **Una verifica pratica.** Prima di permettere l'uso di uno strumento con dati di lavoro, cerca nelle sue condizioni tre risposte: i dati inseriti addestrano il modello? Per quanto tempo sono conservati? Personale del fornitore può leggerli? Se non trovi risposte chiare, è già un motivo sufficiente per non usarlo con dati riservati.

Il copia e incolla che nessuno controlla

Vale la pena soffermarsi sul gesto materiale, perché è lì che si gioca quasi tutto. La fuga di dati tramite AI non passa da un attacco né da una falla tecnica: passa dalla tastiera di una persona che seleziona un testo, lo copia e lo incolla in una finestra. È un'azione che si compie decine di volte al giorno, in automatico, senza pensarci.

Proprio questa naturalezza la rende difficile da fermare con i soli strumenti tecnici. Si possono bloccare alcuni siti, ma non il telefono personale, non la riscrittura a mano, non la fotografia dello schermo. L'unico presidio che regge è far sì che le persone, nel momento del copia e incolla, sappiano riconoscere cosa non deve uscire. È un presidio fatto di consapevolezza, non di software, ed è la ragione per cui la formazione torna in ogni capitolo di questa guida.

Prompt injection e manipolazione degli input

La prompt injection è il rischio tecnico più caratteristico dell'AI generativa, ed è anche il meno intuitivo, perché non agisce su una falla nel codice ma sul modo stesso in cui i modelli interpretano il linguaggio. Vale la pena spiegarlo con calma, a livello di principio, perché capirlo aiuta a difendersi.

Un assistente moderno non si limita a rispondere a chi gli scrive: spesso legge anche contenuti esterni, come una pagina web, un documento, una mail. Il problema nasce qui. Il modello non distingue bene tra i dati che deve elaborare e le istruzioni che deve eseguire: per lui sono tutti testo. Se in un documento che gli viene dato da riassumere qualcuno ha inserito istruzioni nascoste, il modello potrebbe seguirle come se venissero dall'utente legittimo.

Si distinguono due forme. Quella diretta, in cui è l'utente stesso a cercare di forzare il sistema a comportamenti non previsti. E quella indiretta, più pericolosa, in cui le istruzioni ostili arrivano da un contenuto esterno che il modello legge per conto dell'utente, all'insaputa di quest'ultimo.

Le conseguenze possibili vanno dalla divulgazione di informazioni che il sistema non dovrebbe rivelare, fino al compimento di azioni indesiderate quando l'assistente è collegato ad altri strumenti. Più il sistema è autonomo e connesso, più il danno potenziale cresce.

Come si riconosce e si riduce, senza entrare nei dettagli che servirebbero a un attaccante? Alcuni principi difensivi sono solidi.

- **Diffidare dell'autonomia eccessiva**: un assistente che può leggere contenuti esterni e agire su sistemi sensibili è una combinazione da valutare con cura, non da attivare per comodità.
- **Limitare gli accessi**: dare a ogni strumento solo i permessi strettamente necessari, così che anche una manipolazione riuscita faccia poco danno.
- **Tenere l'umano nel circuito**: per le azioni che contano, una conferma umana spezza la catena automatica su cui l'attacco fa affidamento.

- **Trattare l'output con prudenza**: non eseguire mai automaticamente, senza controllo, ciò che un modello produce a partire da contenuti non fidati.

> **In sintesi.** La prompt injection non si batte con un singolo prodotto, ma con una regola di prudenza: meno autonomia e meno accessi si concedono a un sistema, meno una manipolazione può fare danno.

Un esempio di manipolazione indiretta

Per capire perché la forma indiretta preoccupa di più, basta uno scenario plausibile, descritto a livello di principio. Un assistente viene incaricato di leggere e riassumere i documenti che arrivano da fornitori esterni. Uno di questi documenti, oltre al contenuto normale, contiene un testo nascosto pensato per essere letto dalla macchina e non dalla persona.

La persona che usa l'assistente non vede nulla di strano: chiede un riassunto, riceve un riassunto. Ma se il testo nascosto contiene istruzioni e il sistema è collegato ad altri strumenti, l'assistente potrebbe compiere azioni che l'utente non ha mai chiesto, credendo di obbedire a lui.

Il punto importante, ai fini della difesa, è che l'attacco entra da un canale legittimo, un normale documento di lavoro, e che la vittima è inconsapevole. Non c'è un link sospetto da non cliccare. Ecco perché la difesa non può essere solo "fai attenzione", ma deve essere strutturale: limitare ciò che il sistema può fare in autonomia quando elabora contenuti che non vengono da fonti fidate.

Account, credenziali e integrazioni non autorizzate

Gli account personali usati per lavoro sono una delle vie d'ingresso più trascurate. Quando un dipendente adopera il proprio profilo gratuito su un servizio di AI per attività aziendali, l'organizzazione perde ogni controllo su ciò che vi transita.

Il problema si manifesta in più modi. Non si applica l'autenticazione a più fattori che l'azienda impone sui propri sistemi. Non si può sapere quali dati siano stati caricati. E, soprattutto, non si può revocare l'accesso quando la persona lascia l'organizzazione: tutto ciò che ha salvato resta su un account che nessuno può chiudere.

Le credenziali, poi, hanno un valore nuovo nell'epoca dell'AI. Una password incollata per comodità in un assistente diventa una credenziale potenzialmente compromessa. E le truffe che puntano a rubare credenziali, come si vedrà più avanti, sono diventate molto più convincenti proprio grazie all'AI.

Qui torna utile una misura che dovrebbe essere ovvia e troppo spesso non lo è: l'autenticazione a più fattori. Anche se una credenziale viene rubata o incollata per errore dove non doveva, un secondo fattore di verifica impedisce, nella maggior parte dei casi, che da sola basti a entrare. È una delle difese con il miglior rapporto tra efficacia e costo, e va imposta su tutti gli accessi che contano, a partire da quelli agli strumenti di AI aziendali.

Le integrazioni meritano un discorso a parte. Plugin del browser, connettori che collegano l'assistente alla posta o ai documenti, applicazioni di terzi che chiedono ampi permessi: ognuna è una porta potenziale, e molte vengono attivate in pochi clic senza che nessuno valuti a quali dati daranno accesso.

La regola difensiva, di nuovo, è quella del minimo privilegio: concedere a ogni strumento e a ogni integrazione solo l'accesso indispensabile, e nulla di più. È un principio vecchio della sicurezza informatica, che l'AI rende di nuovo attuale, perché le integrazioni attivate dal singolo tendono a concedere tutti i permessi richiesti pur di funzionare subito.

> **Esempio.** Un collaboratore installa un'estensione che promette di riassumere automaticamente le mail. Per funzionare, l'estensione chiede e ottiene accesso a tutta la casella di posta. Da quel momento un fornitore sconosciuto può, in teoria, leggere ogni messaggio. Nessuno in azienda lo ha deciso, e quasi nessuno lo sa.

La contromisura non è vietare le integrazioni, ma censirle e approvarle: stabilire che collegare uno strumento di AI a sistemi che contengono dati aziendali è una decisione tecnica che passa da chi ha la responsabilità, non una scorciatoia individuale.

L'addio che lascia le chiavi attaccate

Il rischio degli account personali si concretizza quasi sempre in un momento preciso: quando una persona lascia l'organizzazione. Tutto ciò che ha caricato sul proprio profilo gratuito, conversazioni, documenti, istruzioni che contengono logica di lavoro, resta lì, su un account che l'azienda non controlla e non può chiudere.

Con un account aziendale, l'accesso si revoca in un minuto, come si fa da sempre con la posta elettronica. Con un account personale, semplicemente, non si può fare nulla: i dati restano fuori, a tempo indeterminato, sotto il controllo di chi se n'è andato. È una delle ragioni più concrete, e meno discusse, per imporre l'uso di account aziendali per qualsiasi attività di lavoro.

Fornitori e catena di approvvigionamento dell'AI

Quasi nessuna organizzazione costruisce in casa i propri sistemi di AI. Li acquista, li affitta come servizio, li integra in altri prodotti. Questo significa che la sicurezza dipende, in larga parte, da soggetti esterni: i fornitori del modello, chi ospita i dati, chi fornisce le applicazioni che incorporano l'AI. È la catena di approvvigionamento, e la sua sicurezza vale quanto il suo anello più debole.

Il rischio della catena ha diverse facce. Un fornitore può subire una violazione che espone i dati che gli sono stati affidati. Può cambiare le proprie condizioni e iniziare a usare i dati in modi prima esclusi. Può dipendere a sua volta da altri fornitori, creando una catena di dipendenze di cui il cliente finale non sa nulla. Può cessare l'attività, lasciando dati e processi in sospeso.

C'è poi un rischio più sottile: la fiducia trasferita. Quando un'applicazione che già si usa aggiunge funzioni di AI, spesso lo fa appoggiandosi a un fornitore terzo, e i dati iniziano a fluire verso un soggetto con cui nessuno ha mai stipulato un accordo specifico. La funzione nuova arriva con un aggiornamento, e la valutazione del rischio non viene rifatta.

Difendersi qui significa fare al fornitore le domande giuste prima di affidargli qualcosa, e rifarle quando le condizioni cambiano.

- Dove vengono trattati e conservati i dati, e per quanto tempo?
- I dati inseriti vengono usati per addestrare i modelli?
- Quali garanzie offre il contratto, e il fornitore agisce come responsabile del trattamento?
- Da quali altri fornitori dipende a sua volta?
- Come comunica eventuali violazioni, e in quali tempi?

Non serve essere giuristi per porre queste domande. Serve la disciplina di porle prima, e di non dare per scontato che un fornitore grande e noto sia automaticamente sicuro per i propri dati. La dimensione del fornitore riduce alcuni rischi e ne aggiunge altri, come la difficoltà di negoziare condizioni su misura.

> **Esempio.** Un'azienda usa da anni un software gestionale di cui si fida. Con un aggiornamento, il gestionale introduce un assistente di AI che, per funzionare, invia i dati a un fornitore terzo non specificato. Nessuno in azienda ha valutato quel nuovo flusso: la fiducia nel software principale si è estesa, senza esame, a un soggetto sconosciuto. La valutazione del fornitore andava rifatta al momento dell'aggiornamento, non data per acquisita.

La lezione è che la sicurezza della catena non si verifica una volta sola. Va rivista ogni volta che qualcosa cambia: una nuova funzione, un nuovo fornitore a monte, una modifica delle condizioni. La domanda "dove vanno i nostri dati?" non ha una risposta definitiva, ma una risposta che va aggiornata.

La dipendenza che non si vede

C'è un aspetto della catena di approvvigionamento particolarmente sfuggente: la concentrazione nascosta. Molte applicazioni diverse, fornite da aziende diverse, si appoggiano in realtà agli stessi pochi fornitori di modelli di AI. Un'organizzazione può credere di aver distribuito il rischio usando dieci strumenti distinti, mentre tutti e dieci, a monte, dipendono dalla stessa infrastruttura.

La conseguenza è che un problema su un singolo fornitore a monte, un'interruzione del servizio, una violazione, un cambio di condizioni, può ripercuotersi in contemporanea su strumenti che sembravano indipendenti. È un rischio che la singola organizzazione non può eliminare, ma che dovrebbe almeno conoscere, per non illudersi di una ridondanza che non esiste.

In pratica, due accortezze aiutano. La prima: chiedere ai fornitori non solo cosa fanno loro, ma da chi dipendono, per avere un'idea della catena reale. La seconda: per i processi più critici, evitare di dare per scontata la continuità del servizio, e avere un piano per il caso in cui uno strumento diventi improvvisamente indisponibile. La dipendenza tecnologica è comoda finché funziona, e costosa quando si interrompe senza che ci si fosse preparati.

Contenuti falsi e ingegneria sociale potenziata dall'AI

L'ingegneria sociale, cioè l'arte di ingannare le persone per ottenere accessi o informazioni, è da sempre la via d'attacco più efficace, perché aggira le difese tecniche colpendo l'anello umano. L'AI l'ha resa più economica, più veloce e molto più credibile.

Il phishing, le mail costruite per indurre la vittima a cliccare o a fornire credenziali, una volta si riconosceva spesso dagli errori di lingua e dal tono goffo. Oggi un testo generato è grammaticalmente perfetto, su misura per il destinatario, e può imitare lo stile di un collega o di un fornitore reale. Il segnale d'allarme classico, l'italiano sgrammaticato, non vale più.

I deepfake aggiungono il falso audiovisivo. Una voce clonata che imita quella di un dirigente, un breve video falsificato, una telefonata in cui sembra parlare una persona di fiducia: strumenti che alzano la posta dell'inganno e che hanno già prodotto frodi reali, soprattutto contro le funzioni che autorizzano pagamenti.

C'è un dettaglio che rende questi attacchi più insidiosi del previsto: spesso bastano pochi materiali pubblici per costruire il falso. La voce di una persona che ha parlato a un convegno, le immagini di un profilo pubblico, lo stile di scrittura di chi pubblica online: materia prima sufficiente, per chi ha gli strumenti, a imitare in modo convincente. Più una figura è esposta pubblicamente, più è facile bersaglio, e i ruoli apicali, per definizione visibili, sono i più imitati.

La tabella raccoglie le principali minacce di questa famiglia, con il meccanismo e la difesa di buon senso.

MINACCIA	COME FUNZIONA	DIFESA PRINCIPALE
Phishing mirato	Mail credibili e su misura per la vittima	Verifica per un secondo canale, diffidenza verso l'urgenza
Voce clonata	Imitazione vocale di una persona di fiducia	Parola d'ordine concordata, conferma per altra via
Video falso	Volto o gesti falsificati in una clip	Procedure che non si basano sul solo riconoscimento visivo
Frode al pagamento	Falsa richiesta urgente da un finto superiore	Doppia autorizzazione per i bonifici, nessuna eccezione
Falsi documenti	Testi o moduli generati per ingannare	Controllo della fonte, canali ufficiali

La difesa comune a tutte queste minacce non è tecnologica ma procedurale: stabilire che certe azioni delicate, prima fra tutte autorizzare un pagamento o comunicare una credenziale, non si compiono mai sulla base di un solo canale, per quanto convincente. Una richiesta urgente arrivata per mail o per voce si verifica sempre per una via indipendente. È una regola semplice, gratuita, e capace di fermare la maggior parte di questi attacchi.

> **Esempio.** Un impiegato dell'amministrazione riceve una chiamata dal "direttore finanziario", voce identica, che chiede un bonifico urgente per chiudere un'operazione riservata. La procedura aziendale impone, per qualsiasi pagamento sopra una certa soglia, una conferma scritta da un secondo responsabile. L'impiegato la richiede, il presunto direttore insiste, e proprio l'insistenza fa scattare il sospetto. La regola, non l'intuito, ha fermato la frode.

Perché i vecchi segnali non bastano più

Per anni la formazione contro il phishing ha insegnato a riconoscere alcuni segnali: errori di grammatica, indirizzi strani, tono impersonale. Quei segnali esistono ancora, ma l'AI ne ha tolti molti. Un testo generato è corretto, personalizzato, coerente con il contesto. Una voce clonata suona come quella vera.

La difesa deve quindi spostarsi dal riconoscere il falso al verificare il vero. Non più "questa mail sembra sospetta?", domanda che oggi inganna spesso, ma "ho confermato questa richiesta per un canale indipendente?". È un cambiamento importante: si smette di affidarsi all'occhio della persona, che può essere ingannato, e ci si affida a una procedura, che non cambia con la qualità del falso.

Tre principi reggono questa difesa, e vale la pena ripeterli finché diventano automatici. L'urgenza è un segnale di allarme, non un motivo per saltare i controlli. La segretezza richiesta ("non dirlo a nessuno") è quasi sempre il marchio di una truffa. E nessuna azione delicata, prima fra tutte un

pagamento, si compie sulla base di un solo canale, per quanto convincente sembri.

La parola d'ordine, una difesa antica che torna utile

Contro la voce clonata e il video falso, una delle difese più efficaci è anche la più semplice: una parola d'ordine concordata in anticipo tra le persone che potrebbero ricevere richieste delicate. Se una telefonata urgente chiede un'azione fuori dall'ordinario, si domanda la parola d'ordine. Un impostore, per quanto perfetta sia la voce clonata, non la conosce.

È una misura a costo zero, che non dipende da alcuna tecnologia e che funziona proprio perché aggira il punto su cui l'AI eccelle, l'imitazione, e si appoggia su qualcosa che l'imitazione non può riprodurre, un segreto condiviso. Le organizzazioni più esposte alle frodi sui pagamenti la adottano da tempo, e l'arrivo dei deepfake l'ha resa di nuovo attuale.

La regola generale che ne deriva è chiara: le procedure delicate non devono mai poggiare sul solo riconoscimento di una voce o di un volto, che oggi si falsificano. Devono poggiare su qualcosa che si sa o si possiede, e che un falso non può avere.

PARTE III

Le difese

Ai rischi si risponde su più livelli che si rinforzano a vicenda: decidere cosa non dare all'AI, scegliere strumenti e configurazioni sicure, sorvegliare, formare le persone, usare l'AI stessa a difesa e sapere come reagire quando qualcosa va storto.

Classificare i dati e definire cosa non dare all'AI

La prima difesa contro la fuga di dati non è un software, ma una decisione: stabilire quali informazioni non devono mai finire in uno strumento di AI non controllato. Senza questa decisione, ogni persona improvvisa la propria, e qualcuno improvviserà male.

Classificare i dati significa dividerli in poche categorie chiare, ognuna con una regola d'uso. Poche, perché uno schema con quindici livelli non lo userà nessuno. Tre o quattro classi sono il punto di equilibrio tra precisione e usabilità.

CLASSE DI DATO	ESEMPI	REGOLA CON L'AI
Pubblico	Materiale già pubblicato, contenuti del sito	Uso libero anche su strumenti pubblici
Interno	Bozze, appunti, procedure non riservate	Solo su strumenti approvati
Riservato	Dati personali di clienti e dipendenti, offerte, contratti	Solo su strumenti approvati e configurati, con base giuridica
Critico	Dati sanitari o finanziari, segreti industriali, credenziali	Mai senza autorizzazione esplicita e valutazione

La classificazione funziona se è collegata a qualcosa che le persone vedono ogni giorno: etichette sui documenti, una guida di una pagina, un promemoria dentro lo strumento approvato. L'obiettivo è che la categoria del dato sia evidente nel momento in cui si decide se incollarlo o no.

C'è una trappola ricorrente: i dati che non sembrano sensibili e invece lo sono. Un nome e un'email sono dati personali. Un elenco di clienti, anche senza altri dettagli, è un'informazione di valore. La formazione sulla classificazione deve lavorare proprio sui casi grigi, perché nessuno incolla volutamente una cartella clinica, ma molti incollano senza pensarci una mail che contiene il recapito di un cliente.

> **Regola pratica.** Nel dubbio, vale la classe più alta. Meglio trattare un dato come riservato anche quando non sarebbe stato necessario, che fare il contrario una volta sola con il dato sbagliato.

Un vantaggio spesso ignorato: la classificazione dei dati di solito esiste già, costruita per la sicurezza informatica o per la privacy. In quel caso non serve inventarla da capo, basta estenderla con la colonna "regola con l'AI". Riusare ciò che le persone già conoscono accorcia di molto i tempi.

La domanda che risolve quasi ogni dubbio

Per non costringere le persone a memorizzare una tabella, conviene ridurre la decisione a una domanda sola, semplice da ricordare e potente abbastanza da coprire la grande maggioranza dei casi: "se questo contenuto finisse pubblicato online o nelle mani di un concorrente, sarebbe un problema?".

Se la risposta è sì, quel contenuto non va su uno strumento non approvato. Se è no, si può procedere con più libertà. È una scorciatoia mentale che funziona anche con uno strumento nuovo che la classificazione non aveva ancora previsto, perché lavora sul valore del dato, non sull'elenco degli strumenti.

Per i dati che riguardano persone vale un secondo riflesso: anche un semplice nome con un recapito è un dato personale, e va trattato con la cura che la legge richiede. La domanda, in quel caso, non è solo "è un problema se esce?", ma "ho una base per trattarlo proprio così?".

> **Esempio.** In uno studio, un collaboratore deve preparare la sintesi di un caso. Prima di incollare il testo in un assistente, applica la domanda: "se questo finisse in mano a un estraneo, sarebbe un problema?". La risposta è sì, perché contiene riferimenti a una persona e alla sua vicenda. Allora usa solo lo strumento approvato dallo studio, oppure rimuove i dati identificativi prima di chiedere aiuto. La domanda, fatta nel momento giusto, ha evitato una fuga.

Strumenti approvati e configurazioni sicure

Dire alle persone cosa non fare senza offrire un'alternativa sicura è la ricetta per essere ignorati. La difesa più efficace contro l'uso pericoloso è uno strumento approvato che sia almeno comodo quanto quello che si userebbe di nascosto.

Uno strumento approvato si distingue per le garanzie che offre, non per il nome. Le caratteristiche da pretendere sono poche e nette.

- I dati inseriti **non vengono usati per l'addestramento** e non sono conservati oltre il necessario.
- Esiste un **contratto** che regola il trattamento dei dati, con il fornitore nel ruolo di responsabile.
- L'accesso passa dagli **account aziendali**, con autenticazione forte e possibilità di revoca immediata.
- Dove serve, il trattamento avviene **entro l'Unione Europea** o con garanzie adeguate per i trasferimenti.

Scegliere lo strumento è solo metà del lavoro. L'altra metà è configurarlo bene, perché molte impostazioni predefinite privilegiano la comodità sulla sicurezza. Disattivare, dove possibile, l'uso dei dati per l'addestramento. Limitare la conservazione. Attivare l'autenticazione a più fattori. Definire chi può collegare integrazioni e chi no. Una configurazione lasciata ai valori di fabbrica è una difesa lasciata a metà.

> **Esempio.** Un'azienda acquista la versione business di un assistente, convinta di essere a posto. Mesi dopo scopre che, tra le impostazioni predefinite, l'opzione di conservazione estesa delle conversazioni era attiva e nessuno l'aveva toccata. Lo strumento giusto era stato scelto bene, ma lasciato configurato male. La sicurezza non è nel prodotto, è nel prodotto impostato come si deve.

Per questo l'adozione di uno strumento dovrebbe sempre prevedere una fase di configurazione consapevole, affidata a chi conosce le impostazioni di sicurezza, e una verifica periodica, perché i fornitori cambiano le opzioni e ne aggiungono di nuove con gli aggiornamenti. Quello che era configurato bene un anno fa potrebbe non esserlo più oggi.

Tre livelli secondo la sensibilità

Non esiste una sola risposta giusta: la soluzione adatta dipende da quanto sono sensibili i dati trattati.

SENSIBILITÀ	SOLUZIONE TIPICA	GARANZIA PRINCIPALE
Bassa	Versione business di un assistente diffuso	Dati non usati per l'addestramento, contratto col fornitore
Media	Servizio aziendale con trattamento nell'Unione Europea	Controllo sui trasferimenti, account centralizzati
Alta	Modello su infrastruttura controllata dall'organizzazione	I dati non lasciano il perimetro interno

La maggior parte delle organizzazioni si colloca nei primi due livelli e copre con questi quasi tutti i bisogni reali. Il terzo ha costi e competenze superiori, e si giustifica solo dove i dati sono davvero critici. L'errore da evitare è in entrambe le direzioni: sovradimensionare per usi banali, o usare strumenti di consumo per dati che meriterebbero il livello più alto.

La comodità è una misura di sicurezza

Vale la pena insistere su un punto su cui quasi tutte le iniziative falliscono. Lo strumento approvato compete, ogni giorno, con uno strumento libero a un clic di distanza. Se per accedere a quello ufficiale servono troppi passaggi, se è più lento o ha meno funzioni, le persone torneranno a quello pubblico, e dal loro punto di vista avranno ragione.

Per questo investire nella comodità dello strumento approvato non è una concessione al quieto vivere, ma una misura di sicurezza vera e propria. Una difesa che impone troppo attrito si aggira da sola. La sicurezza migliore è quella che coincide con la strada più facile: quando la scelta sicura è anche la più comoda, non serve nemmeno imporla.

Sorveglianza, log e rilevamento degli abusi

Non si protegge ciò che non si vede. La sorveglianza serve a sapere come l'AI viene usata, a riconoscere gli abusi e a ricostruire cosa è successo dopo un incidente. È la difesa che trasforma un problema invisibile in uno gestibile.

Sorvegliare non significa spiare le persone. Significa raccogliere, in modo proporzionato e trasparente, le tracce necessarie a garantire la sicurezza: quali strumenti approvati vengono usati, da quali account, con quali integrazioni attive. Il monitoraggio dell'attività dei lavoratori è soggetto a vincoli precisi, e va impostato in forma aggregata, informando le persone e, dove richiesto, concordandolo con le rappresentanze.

I log, cioè le registrazioni delle attività, sono il materiale di base. Hanno valore solo se rispondono a quattro requisiti: esistono, sono completi, sono protetti da manomissioni e qualcuno li guarda. Un log che nessuno legge è un costo senza beneficio; un log manomettibile è una prova senza valore.

Sul rilevamento degli abusi, vale un principio di realismo. Gli strumenti pubblici usati di nascosto, per definizione, non lasciano log nei sistemi aziendali. Quello che si può fare è osservare, in forma aggregata, il traffico verso i principali servizi di AI per capire la dimensione del fenomeno, e concentrare i controlli puntuali sugli strumenti approvati, dove la visibilità è piena.

> **Esempio.** Dopo un sospetto di fuga di dati, un'organizzazione ricostruisce l'accaduto grazie ai log del suo strumento approvato: chi ha avuto accesso, quando, con quali integrazioni attive. In poche ore capisce la portata reale del problema e può reagire in modo proporzionato. Senza quei log avrebbe brancolato per giorni, costretta a immaginare lo scenario peggiore. La registrazione, qui, non ha prevenuto l'incidente, ma ha reso possibile gestirlo.

> **In sintesi.** La sorveglianza utile è proporzionata, trasparente e orientata alla sicurezza, non al controllo delle persone. Raccoglie le tracce che servono, le protegge, e prevede che qualcuno le legga davvero.

Un'avvertenza sul falso senso di sicurezza: avere i log non equivale a essere sicuri. Servono solo se collegati a una capacità di reazione. Registrare un abuso e non fare nulla è peggio che non registrarlo, perché lascia una prova documentale di un problema noto e ignorato.

Privacy dei lavoratori e sicurezza: un equilibrio, non un conflitto

La sorveglianza tocca un nervo delicato, ed è giusto affrontarlo apertamente. Controllare l'uso degli strumenti può scivolare nel controllo delle persone, che la legge limita per buone ragioni. Ma sicurezza e tutela dei lavoratori non sono in conflitto, se la sorveglianza è impostata bene.

Tre criteri tengono insieme le due esigenze. La proporzionalità: si raccoglie solo ciò che serve davvero alla sicurezza, non tutto ciò che è tecnicamente possibile. La trasparenza: le persone sanno cosa viene registrato e perché, senza zone d'ombra. E l'aggregazione: dove l'obiettivo è capire un fenomeno, si guardano i dati in forma aggregata, non si profilano i singoli.

Impostata così, la sorveglianza non incrina la fiducia, la sostiene: dimostra che l'organizzazione prende sul serio la protezione dei dati di tutti, compresi quelli dei dipendenti. Nei contesti che lo richiedono, il confronto con le rappresentanze dei lavoratori non è un ostacolo, ma il modo per costruire regole che le persone accetteranno di rispettare.

Formazione delle persone: l'anello più importante

La maggior parte degli incidenti di sicurezza legati all'AI non nasce da un attacco sofisticato, ma da una persona che, in buona fede, ha fatto la cosa sbagliata: ha incollato un dato, ha cliccato un link, ha creduto a una voce al telefono. Per questo la formazione non è un complemento delle difese tecniche, ma il loro fondamento.

La formazione efficace sulla sicurezza dell'AI ha tre obiettivi, in quest'ordine.

Primo, far capire i meccanismi: che i dati inseriti escono dall'organizzazione, che gli output vanno verificati, che gli attacchi oggi sono credibili e su misura. Senza questa comprensione, ogni regola resta un divieto inspiegato che la prima scadenza spazzerà via.

Secondo, insegnare i comportamenti corretti in modo concreto: cosa non incollare mai, come verificare una richiesta sospetta, quando fermarsi e chiedere. La sicurezza si gioca su abitudini, e le abitudini si costruiscono con esempi del proprio lavoro, non con principi astratti.

Terzo, ancorare tutto alle regole interne e alle norme, incluso il dovere di alfabetizzazione che l'AI Act pone in capo a chi usa questi sistemi. L'articolo 4 del regolamento chiede un livello adeguato di competenza del personale: è esattamente ciò che impedisce a un rischio di trasformarsi in incidente.

La formazione va ripetuta, perché gli strumenti e le minacce cambiano in fretta e le persone entrano ed escono. Meglio un percorso iniziale solido seguito da aggiornamenti brevi e regolari, che un corso una tantum destinato a invecchiare. Oltre all'effetto pratico, una formazione documentata ha valore difensivo: dimostra, davanti a un'autorità o a un cliente, che l'organizzazione ha fatto la propria parte. AIPIA struttura i propri percorsi con il rilascio di una credenziale verificabile al termine, che attesta la competenza acquisita secondo uno standard europeo.

> *Si può comprare la migliore tecnologia di sicurezza sul mercato e perdere tutto per una persona che non sapeva. La competenza non è un costo, è la prima difesa.*
>
> PRINCIPIO GUIDA DEI CORSI AIPIA

Formare per livelli, non tutti allo stesso modo

Una formazione identica per tutti spreca tempo e annoia. Le esigenze sono diverse e conviene riconoscerlo.

A tutto il personale serve una base comune: cosa non inserire mai negli strumenti di AI, come verificare una richiesta sospetta, perché gli output vanno controllati, a chi rivolgersi. È un modulo breve, concreto, ancorato a esempi del proprio lavoro.

A chi usa l'AI in modo intenso serve di più: scrivere richieste efficaci, riconoscere un errore del modello, capire quando un compito è adatto e quando no. È la formazione che rende lo strumento approvato così utile da togliere ogni voglia di cercarne uno clandestino.

A chi ha ruoli di governo, dal referente AI a chi presidia la sicurezza, serve la competenza per decidere: valutare strumenti e fornitori, leggere le condizioni di un contratto, impostare configurazioni e log, gestire un incidente. È qui che un percorso strutturato con credenziale, come quelli proposti da AIPIA, ripaga di più. La Credenziale Europea AI, nelle versioni Base e Advanced, attesta in modo verificabile la competenza raggiunta.

La simulazione vale più della lezione

Sulla sicurezza, la formazione che resta non è quella ascoltata, ma quella vissuta. Spiegare a parole come riconoscere un tentativo di frode lascia il tempo che trova; mettere le persone davanti a un caso realistico, e poi discuterne insieme, costruisce un riflesso che dura.

Le organizzazioni più attente usano simulazioni controllate: inviano, in modo trasparente e concordato, finti messaggi di phishing per misurare quante persone li riconoscono, e trasformano i risultati in occasione di apprendimento, non in pretesto per colpire chi sbaglia. L'obiettivo non è il voto, è l'allenamento. Chi è caduto in una simulazione, e ne ha parlato senza vergogna, difficilmente cadrà in quella vera.

Vale lo stesso principio per i comportamenti con i dati: meglio un esercizio concreto, in cui si decide insieme cosa si può incollare e cosa no a partire da documenti reali del proprio lavoro, che una lezione astratta sulle categorie di rischio. La sicurezza è fatta di gesti, e i gesti si imparano provandoli.

La formazione come prova di diligenza

Oltre all'effetto pratico, una formazione documentata ha un valore difensivo. Se un giorno l'organizzazione dovesse dimostrare di aver fatto la propria parte, verso un'autorità, un cliente o in un contenzioso, una formazione tracciabile con credenziali verificabili è una prova concreta. La differenza tra "avevamo detto di stare attenti" e "abbiamo formato il personale, ecco le credenziali rilasciate" è netta, e sotto norme come l'AI Act e NIS2 può pesare.

L'AI a difesa: dove aiuta davvero la cybersecurity

Finora i rischi. Ma l'intelligenza artificiale è anche uno degli strumenti più utili che la sicurezza informatica abbia ricevuto negli ultimi anni, a patto di sapere cosa può e cosa non può fare. Ignorare questo lato significa lasciare ai soli attaccanti un vantaggio che potrebbe essere anche dei difensori.

Dove l'AI aiuta concretamente? In compiti che richiedono di analizzare grandi quantità di dati a velocità impossibili per una persona.

- **Rilevamento delle anomalie**: riconoscere schemi insoliti in enormi flussi di log o di transazioni, segnalando ciò che si discosta dalla norma.
- **Filtro del phishing**: individuare mail sospette analizzando contenuto, contesto e segnali che a occhio umano sfuggono.
- **Smistamento delle segnalazioni**: ordinare per priorità la massa di allarmi che ogni giorno arriva a chi presidia i sistemi, riducendo il rumore.
- **Assistenza all'analisi**: aiutare gli specialisti a riassumere informazioni di minaccia, a documentare un incidente, a comprendere più in fretta cosa è accaduto.

Il valore, in tutti questi casi, è la scala. Un team di sicurezza non può leggere milioni di eventi al giorno; un sistema di AI può smistarli e portare all'attenzione umana solo ciò che conta. L'AI non sostituisce l'analista, gli toglie il lavoro ripetitivo e gli lascia le decisioni.

> **Esempio.** Un sistema di monitoraggio nota che un account, di solito attivo dall'Italia in orario d'ufficio, accede improvvisamente di notte da un altro continente e inizia a scaricare grandi quantità di dati. Non è una regola fissa a scattare, ma il riconoscimento di uno schema che si discosta dal comportamento abituale. La segnalazione arriva a una persona, che blocca l'account e verifica. L'AI ha visto in un istante un'anomalia che, tra migliaia di accessi legittimi, sarebbe sfuggita all'occhio umano.

Conta però capire cosa, in questo esempio, ha funzionato: non l'AI da sola, ma l'AI che porta all'attenzione di una persona competente la cosa giusta al momento giusto. Tolta la persona, restano solo allarmi; tolta l'AI, resta una massa di dati illeggibile. È la combinazione a produrre il risultato, e questo vale per quasi ogni uso difensivo serio.

Vanno detti con onestà anche i limiti, perché il marketing della sicurezza tende a nasconderli. L'AI produce falsi positivi e falsi negativi: segnala cose innocue e lascia passare cose dannose. Può essere ingannata da attacchi pensati apposta. E, se mal configurata, genera una valanga di allarmi che paralizza invece di aiutare. È uno strumento potente nelle mani di chi sa usarlo, non una difesa automatica da accendere e dimenticare.

> **Attenzione all'equilibrio.** Gli stessi strumenti che aiutano i difensori aiutano gli attaccanti. L'AI abbassa il costo del phishing e dei contenuti falsi tanto quanto abbassa il costo della difesa. La corsa è continua, e vince chi mantiene competenza e attenzione, non chi compra l'ultimo prodotto.

Per le organizzazioni piccole: cosa è alla portata

Le funzioni di AI a difesa non sono riservate alle grandi aziende con un reparto di sicurezza. Molte arrivano già incluse negli strumenti che una piccola organizzazione usa ogni giorno: il filtro antispam della posta, i controlli antifrode del sistema di pagamento, le difese del software gestionale. Spesso il modo migliore per beneficiarne è semplicemente tenerle attive e aggiornate, invece di disattivarle perché "danno fastidio".

Per chi non ha competenze interne, la scelta sensata non è costruire, ma scegliere bene fornitori che integrano queste protezioni e mantenere le difese accese. La complessità si delega; la responsabilità di scegliere e di tenere aggiornato, no.

> **Esempio.** Una PMI riceve ogni giorno decine di mail, e alcune sono tentativi di frode sempre più curati. Il filtro intelligente del suo servizio di posta intercetta la maggior parte di quelli evidenti, ma non tutti. La difesa che fa la differenza non è il filtro da solo, è il filtro più la regola che impone di verificare ogni richiesta di pagamento per un secondo canale. La tecnologia riduce il volume; la procedura ferma ciò che passa.

Questo equilibrio tra strumento e procedura è il filo di tutta la parte sulle difese. Nessuno dei due basta da solo. La tecnologia abbatte il rumore e gestisce la scala; le persone e le regole fermano ciò che la tecnologia lascia passare. Affidarsi solo all'una o solo alle altre lascia sempre un fianco scoperto.

Risposta agli incidenti che coinvolgono l'AI

Nessuna difesa azzera il rischio. Prima o poi qualcosa va storto, e il momento per decidere come reagire non è mentre accade. Avere un piano di risposta, anche minimo, è ciò che separa un incidente contenuto da un disastro.

Un incidente legato all'AI ha spesso una forma diversa da quelli tradizionali. Non c'è sempre un'intrusione: a volte è un dato uscito per errore, una decisione automatica sbagliata, un contenuto falso che ha ingannato qualcuno. Ma la logica della risposta resta la stessa, e si può riassumere in pochi passi.

1. **Contenere**: fermare ciò che si può fermare. Revocare un accesso, sospendere un'integrazione, chiedere al fornitore la cancellazione di un dato dove possibile.
2. **Capire**: stabilire cosa è successo, quali dati e quante persone sono coinvolti, da quando.
3. **Coinvolgere**: attivare chi ha la responsabilità, e in presenza di dati personali il DPO, che valuterà obblighi e tempi di un'eventuale notifica.
4. **Documentare**: registrare l'accaduto e la reazione, perché la documentazione fa parte della diligenza e serve a non ripetere l'errore.
5. **Imparare**: chiudere il ciclo correggendo la causa, non solo l'effetto.

Il fattore che decide la gravità è il tempo. Un errore gestito in poche ore ha conseguenze limitate; lo stesso errore scoperto mesi dopo, magari da un cliente o da un'autorità, è molto peggio. Ecco perché la cultura della segnalazione, descritta nella parte sulla policy e sulla formazione, è parte integrante della risposta: un'organizzazione dove le persone nascondono gli errori li scopre sempre troppo tardi.

Prepararsi prima che serva

Il momento peggiore per decidere come reagire a un incidente è mentre sta accadendo. Sotto pressione si dimenticano passaggi, si perde tempo prezioso, si prendono decisioni avventate. Per questo anche un piano minimo, scritto in tempo di pace, vale molto più dell'improvvisazione del momento.

Non serve un documento elaborato. Bastano poche risposte decise in anticipo: chi va avvisato per primo e come lo si raggiunge anche fuori orario, chi ha l'autorità di sospendere un accesso o un'integrazione, chi valuta gli obblighi sui dati personali, chi parla con l'esterno se serve. Sono domande semplici che, senza una risposta pronta, in mezzo a un incidente diventano ostacoli.

Vale anche la pena provarlo, almeno una volta, con un'esercitazione su carta: si immagina uno scenario plausibile, una fuga di dati o una frode tentata, e si percorrono i passi della risposta. Spesso emergono buchi banali, un recapito che nessuno conosce, un'autorizzazione che manca, che è molto meglio scoprire durante una prova che durante un fatto reale.

Esempio. Una persona si accorge di aver caricato il file sbagliato, con dati di clienti, su uno strumento non approvato. Lo segnala subito, senza timore di sanzioni, al referente indicato. Nel giro di un'ora si chiede la cancellazione al fornitore, si valuta la portata e il DPO decide il da farsi. Lo stesso errore, taciuto per paura, sarebbe diventato una violazione scoperta troppo tardi.

PARTE IV

Metodo e avvio

Capiti i rischi e le difese, resta la parte difficile: metterle in pratica. Una policy che le persone seguono, una checklist prima di adottare gli strumenti, ruoli chiari, gli errori da evitare, le misure da guardare e un piano concreto per i primi novanta giorni.

Una policy di sicurezza per l'uso dell'AI

Una policy serve a rendere prevedibile il comportamento delle persone davanti a situazioni che le difese tecniche non coprono. Funziona solo se è breve, chiara e applicabile. Le policy che falliscono sono lunghe, scritte in linguaggio giuridico, piene di divieti generici e prive di esempi: nessuno le legge, quindi nessuno le segue.

Una buona policy di sicurezza per l'AI risponde a domande concrete che la persona si pone davvero: quali strumenti posso usare, cosa posso inserirci, cosa non devo mai inserirci, come riconosco una richiesta sospetta, a chi mi rivolgo, come segnalo un problema senza timore.

Lo schema seguente è una struttura di partenza, da adattare alla propria realtà. Non è un modello legale completo.

```
POLICY DI SICUREZZA PER L'USO DELL'AI - SCHEMA MINIMO

1. SCOPO E AMBITO
   A chi si applica (tutti i dipendenti e collaboratori)
   e a quali strumenti di AI.

2. STRUMENTI APPROVATI
   - Elenco degli strumenti aziendali autorizzati
   - Divieto di account personali per attività di lavoro
   - Come si chiede l'approvazione di un nuovo strumento

3. DATI: COSA SI PUÒ INSERIRE
   - Pubblico ..... consentito
   - Interno ..... solo su strumenti approvati
   - Riservato ..... solo su strumenti approvati e configurati
   - Critico ..... mai senza autorizzazione esplicita

4. CREDENZIALI E ACCESSI
   - Mai inserire password o segreti negli strumenti di AI
   - Autenticazione a più fattori obbligatoria
   - Integrazioni solo se approvate da chi ha la responsabilità

5. RICHIESTE SOSPETTE
   - Verifica per un secondo canale prima di pagamenti o credenziali
   - Diffidenza verso urgenza e segretezza
   - Nessuna eccezione per i pagamenti, anche se la voce è nota

6. VERIFICA DEGLI OUTPUT
   - Niente decisioni su persone senza supervisione umana
   - Citazioni, numeri e norme controllati alla fonte

7. SEGNALAZIONE E RESPONSABILITÀ
   - A chi rivolgersi per dubbi e incidenti
   - Segnalazione senza colpa per chi avvisa in buona fede

8. FORMAZIONE
   - Percorso obbligatorio prima dell'uso, aggiornamenti periodici

Versione, data, prossima revisione.
```

Due accorgimenti rendono una policy viva invece che morta. Il primo: accompagnarla con esempi reali del proprio settore. Il secondo: prevedere un canale di segnalazione senza colpa, così che chi si accorge di un errore lo dica subito. Una policy senza un modo sicuro per ammettere gli sbagli produce silenzio, non sicurezza.

Il canale di segnalazione senza colpa

Questo punto separa una policy di facciata da una che riduce davvero il danno. In sicurezza vale un principio noto: gli incidenti che fanno più male non sono quelli che accadono, ma quelli che restano nascosti.

Se una persona si accorge di aver fatto un errore e teme di essere punita, tacerà. L'organizzazione perderà l'unica finestra utile per reagire in fretta. Se invece esiste un canale dove segnalare l'errore senza conseguenze disciplinari, quella stessa persona lo userà, e l'incidente diventa gestibile.

La regola da scrivere nero su bianco è semplice: chi segnala tempestivamente un proprio errore in buona fede non subisce sanzioni per averlo commesso. Costa nulla e cambia tutto, perché allinea l'interesse del singolo, non essere punito, con quello dell'organizzazione, sapere in fretta. Le sanzioni si riservano a chi nasconde o a chi agisce in malafede, non a chi sbaglia e lo dice.

Una pagina, non venti

Sulla lunghezza vale la pena essere netti. Una policy che supera le tre o quattro pagine non verrà letta dalle persone a cui è destinata, e una regola non letta è una regola che non esiste. Meglio un documento principale di una o due pagine, in italiano comprensibile, eventualmente affiancato da una scheda di approfondimento per chi ha ruoli di responsabilità.

Il testo che tutti devono ricordare deve poter stare in testa: cosa non inserire mai, come verificare una richiesta sospetta, a chi rivolgersi. Tutto il resto è dettaglio per gli specialisti. Una policy che cerca di prevedere ogni caso diventa illeggibile e fallisce proprio nel suo scopo, che è guidare il comportamento quotidiano delle persone normali, non coprire ogni ipotesi davanti a un giudice.

> **Esempio.** Un comune di medie dimensioni adotta una policy di due pagine, scritta in italiano semplice, con tre esempi presi dal lavoro quotidiano degli uffici. La affianca a una scheda di una pagina appesa accanto a ogni postazione, con la regola d'oro e il recapito a cui segnalare. Sei mesi dopo, le persone la citano a memoria. La versione precedente, un documento di quaranta pagine in linguaggio tecnico, non l'aveva letta nessuno.

Una checklist di sicurezza prima di adottare uno strumento

Ogni nuovo strumento di AI è una decisione di sicurezza, anche quando arriva come funzione aggiuntiva di un'applicazione già in uso. Il modo più semplice per non dimenticare nulla è una checklist da percorrere prima di dire sì, non dopo.

```
CHECKLIST PRIMA DI ADOTTARE UNO STRUMENTO DI AI

DATI
[ ] Quali dati tratteremo con questo strumento?
[ ] La classe più alta di dato è compatibile con le sue garanzie?
[ ] I dati vengono usati per addestrare il modello? (se sì, stop)
[ ] Dove sono trattati e conservati, e per quanto tempo?

FORNITORE
[ ] Esiste un contratto adeguato sul trattamento dei dati?
[ ] Il fornitore agisce come responsabile del trattamento?
[ ] Da quali altri fornitori dipende?
[ ] Come e in quanto tempo comunica le violazioni?

ACCESSI
[ ] Si accede con account aziendali e autenticazione forte?
[ ] Possiamo revocare l'accesso in modo rapido?
[ ] Quali integrazioni richiede, e con quali permessi?
[ ] I permessi sono i minimi necessari?

GOVERNO
[ ] Chi è responsabile di questo strumento internamente?
[ ] Le persone che lo useranno sono formate?
[ ] È previsto nella policy e nei log?
[ ] Quando rivedremo questa decisione?

Se una risposta manca, la decisione non è pronta.
```

La checklist non serve a complicare la vita, ma a spostare la valutazione nel momento giusto: prima dell'adozione, quando cambiare idea costa poco, invece che dopo un incidente, quando costa molto. Va applicata anche, e soprattutto, quando uno strumento già usato annuncia nuove funzioni di AI: è lì che le valutazioni si dimenticano più facilmente.

Un avvertimento sull'uso pratico: la checklist non è un modulo da archiviare, ma una conversazione da fare. Le risposte vanno cercate sul serio, anche quando richiedono di leggere le condizioni di un fornitore o di chiedere chiarimenti. Una checklist compilata in fretta, mettendo crocette senza verificare, dà l'illusione del controllo e nient'altro. Meglio poche domande prese sul serio che molte risposte di comodo.

Per le adozioni a basso rischio, la checklist può essere percorsa in pochi minuti da chi propone lo strumento. Per quelle che toccano dati riservati o critici, va coinvolto chi ha la responsabilità della sicurezza e, dove ci sono dati personali, il DPO. La regola pratica: più alta è la classe di dato in gioco, più la decisione sale di livello.

Ruoli e responsabilità per la sicurezza dell'AI

La sicurezza dell'AI fallisce quasi sempre per la stessa ragione: nessuno ne è chiaramente responsabile. Resta sospesa tra l'IT, che la vede come problema tecnico, il legale, che la vede come conformità, e il business, che la vede come produttività. Quando una cosa è di tutti, non è di nessuno.

Non serve una struttura pesante. Serve assegnare ruoli precisi, anche a persone che già esistono, perché ogni decisione abbia un titolare.

- **Una regia complessiva** (un referente AI o un piccolo gruppo) che approva strumenti e policy e decide nei casi dubbi.
- **L'IT e la sicurezza**, che valutano gli strumenti, gestiscono account e integrazioni, presidiano log e sorveglianza.
- **Il DPO o il legale**, che presidiano conformità, basi giuridiche, gestione delle violazioni.
- **L'HR e la formazione**, che gestiscono i percorsi e l'alfabetizzazione.
- **I responsabili di reparto**, che conoscono i flussi reali e fanno da ponte con le persone.

Il punto delicato è il coordinamento. Una decisione presa solo dalla sicurezza tende a vietare tutto, una presa solo dal business tende ad aprire tutto, una presa solo dal legale tende alla paralisi. La regia serve a bilanciare queste spinte.

Chi decide cosa: una ripartizione di esempio

Per rendere concreto il discorso, ecco come si possono distribuire le decisioni tipiche. È uno schema da adattare, non un dogma.

DECISIONE	CHI PROPONE	CHI APPROVA
Adottare un nuovo strumento	IT / referente AI	Referente AI, con parere legale se ci sono dati personali
Collegare un'integrazione a sistemi interni	IT	Referente AI con la sicurezza
Definire le classi di dati	DPO / sicurezza	Referente AI
Gestire un incidente	Chi lo rileva	IT e referente AI, con DPO se ci sono dati personali
Aggiornare la formazione	HR	Referente AI

Il principio che tiene insieme la tabella è la separazione tra chi propone e chi approva: evita che la stessa persona spinga uno strumento e lo validi da sola, e garantisce che ogni decisione rilevante passi da uno sguardo con la responsabilità complessiva.

Nelle realtà piccole tutto questo si concentra in poche persone, a volte in una sola. Va bene, purché i compiti siano espliciti e qualcuno risponda di ciascuno. Dove manca la competenza interna, ha senso appoggiarsi a un professionista esterno per le scelte iniziali: l'associazione mantiene una directory di professionisti a cui rivolgersi. Per chi ha responsabilità professionali verso terzi, esiste anche una copertura assicurativa dedicata al rischio di chi lavora con l'AI.

La responsabilità che non si delega

C'è una parte della sicurezza che si può affidare a fornitori e consulenti, e una che resta sempre in capo alla direzione. La prima è l'esecuzione tecnica: configurare, monitorare, gestire l'infrastruttura. La seconda è la responsabilità delle scelte: cosa proteggere, con quale priorità, accettando quali rischi.

Le norme più recenti, NIS2 in primo luogo, rendono questa distinzione esplicita: la sicurezza diventa un dovere documentato di chi guida l'organizzazione, non qualcosa da scaricare interamente su un fornitore esterno. Si può comprare la competenza tecnica; non si può comprare l'esenzione dalla responsabilità. Una direzione che pensa di aver risolto il tema firmando un contratto con un fornitore ha frainteso il problema.

Gli errori tipici

Quasi tutte le organizzazioni che affrontano la sicurezza dell'AI inciampano sugli stessi ostacoli. Conoscerli in anticipo costa poco e fa risparmiare mesi.

ERRORE	CORREZIONE
Vietare tutto senza offrire alternative	Vietare gli usi pericolosi e dare uno strumento approvato comodo
Difendere solo la parte tecnica, trascurando le persone	Formazione come prima difesa, non come complemento
Fidarsi di un fornitore perché grande e noto	Valutare le garanzie reali, non la reputazione
Attivare integrazioni con tutti i permessi richiesti	Minimo privilegio, approvazione di chi ha la responsabilità
Raccogliere log che nessuno legge	Sorveglianza collegata a una reale capacità di reazione
Considerare l'AI di difesa una soluzione automatica	Strumento per specialisti competenti, con limiti noti
Punire chi segnala un errore	Segnalazione senza colpa, premiare la trasparenza
Fare tutto una volta e considerarlo chiuso	Manutenzione continua, revisioni periodiche

Un filo lega quasi tutti questi errori: nascono dall'idea che la sicurezza sia un prodotto da comprare o un documento da firmare una volta, invece di una pratica da tenere viva. La tecnologia dà un falso senso di protezione quando le persone e i processi restano indietro, ed è la situazione più comune.

L'errore più costoso: aspettare di essere pronti

Ce n'è uno che non sta nella tabella perché non è un errore di metodo, ma di tempismo, ed è forse il più diffuso. Molte organizzazioni rinviano qualsiasi azione in attesa di avere le idee chiare, il budget, la persona giusta, la normativa definitiva. Nel frattempo l'uso dell'AI continua a crescere, e ogni mese di attesa è un mese di esposizione non governata.

Non si sarà mai pronti del tutto, perché gli strumenti e le minacce cambiano più in fretta di qualsiasi piano. La scelta vera non è tra agire bene e agire in modo imperfetto, ma tra agire in modo imperfetto e non agire affatto. Un primo strumento approvato, una pagina di regole e un'ora di formazione, messi in campo questo mese, proteggono più di una strategia impeccabile rimandata al prossimo anno.

Misurare la postura di sicurezza nel tempo

Misurare la sicurezza dell'AI non significa contare gli attacchi respinti, numero che dice poco. Significa capire se l'organizzazione sta diventando più capace di prevenire, rilevare e reagire. Pochi indicatori, guardati con costanza, valgono più di un cruscotto che nessuno legge.

Alcuni indicatori riguardano la prevenzione: la quota di persone che usano gli strumenti approvati invece di quelli personali, la copertura della formazione, la percentuale di strumenti adottati dopo una valutazione formale.

Altri riguardano il rilevamento e la reazione: il numero di segnalazioni spontanee, il tempo che passa tra un problema e la sua scoperta, il tempo di reazione a un incidente. Vanno letti insieme, perché presi da soli ingannano.

Quattro misure per cominciare

Se si deve partire con un set minimo, questi quattro indicatori dicono già molto e si raccolgono senza strumenti costosi.

- **Adozione degli strumenti approvati**: quante persone li usano davvero al posto di quelli personali. Sale quando la governance funziona.
- **Copertura della formazione**: quota di personale che ha completato il percorso di base. Dovrebbe avvicinarsi alla totalità.
- **Segnalazioni spontanee**: numero di dubbi ed errori riferiti senza essere scoperti. Uno zero stabile indica paura o disinteresse, non sicurezza.
- **Tempo di scoperta**: quanto passa tra un problema e il momento in cui qualcuno lo nota. Più si accorcia, meglio è.

Tre o quattro misure, riviste ogni trimestre, bastano a capire se la rotta è giusta. Un cruscotto pieno di decine di numeri che nessuno guarda non serve a niente: l'obiettivo è decidere meglio, non riempire una tabella.

> **Attenzione a un effetto controintuitivo.** Quando la sicurezza migliora, alcuni numeri "negativi" salgono: più segnalazioni, più incidenti registrati. Non è un peggioramento, è emersione. Le persone si fidano e raccontano. Il vero allarme è il silenzio assoluto: di solito non significa assenza di problemi, ma assenza di visibilità.

Nessuno di questi numeri va trasformato in una pagella per colpire i singoli. Servono a guidare le decisioni dell'organizzazione. Nel momento in cui un indicatore diventa uno strumento sanzionatorio, smette di essere affidabile, perché le persone iniziano a manipolarlo, e si torna al silenzio che si voleva evitare.

Leggere i numeri senza farsi ingannare

Un indicatore preso da solo dice poco e a volte mente. Vanno letti insieme, perché è la combinazione a raccontare la storia vera.

Immaginiamo due organizzazioni. Nella prima, dopo sei mesi, le segnalazioni sono salite, gli incidenti registrati pure, e il tempo di scoperta si è dimezzato. A uno sguardo distratto sembra un peggioramento. È invece il ritratto di una sicurezza che funziona: le persone si fidano, raccontano, e i

problemi vengono a galla in fretta.

Nella seconda, dopo sei mesi, segnalazioni a zero e nessun incidente registrato. Sulla carta è perfetta. In realtà, quasi sempre, vuol dire che i problemi restano sommersi e che nessuno si fida abbastanza da parlarne. Il silenzio non è sicurezza, è cecità. Per questo la lettura degli indicatori va affidata a chi conosce il quadro, non a un calcolo automatico che premia i reparti "senza problemi", che troppo spesso sono solo quelli che non li raccontano.

Un piano per i primi 90 giorni

La teoria delle parti precedenti diventa utile solo se si traduce in azione. Davanti a un tema vasto, molte organizzazioni non partono mai, in attesa del momento perfetto che non arriva. Per questo serve un primo trimestre concreto, fatto di passi fattibili che mettono in moto le cose.

```
PRIMI 90 GIORNI - SICUREZZA E AI

MESE 1 - VEDERE E DECIDERE I RUOLI
  - Nomina di un referente AI per la sicurezza
  - Coinvolgimento di IT, DPO/legale, HR
  - Ricognizione: quali strumenti di AI sono già in uso
  - Classificazione dei dati in 3-4 classi

MESE 2 - STRUMENTI E REGOLE
  - Scelta e configurazione sicura di uno strumento approvato
  - Attivazione account aziendali e autenticazione forte
  - Bozza di policy di sicurezza (poche pagine, con esempi)
  - Checklist di adozione adottata come prassi

MESE 3 - PERSONE E REAZIONE
  - Prima formazione obbligatoria a tutto il personale
  - Canale di segnalazione senza colpa attivo
  - Piano minimo di risposta agli incidenti scritto
  - Definiti 3-4 indicatori da rivedere ogni trimestre

Obiettivo del trimestre: passare dall'uso invisibile
a un uso visibile, governato e più sicuro.
```

Novanta giorni non bastano a costruire una sicurezza matura, e non è questo lo scopo. Bastano a invertire la rotta: sapere cosa succede, dare alle persone strumenti sicuri e regole chiare, costruire la capacità minima di reagire. Il resto si costruisce sopra queste fondamenta. La tentazione da combattere, soprattutto negli enti pubblici e nelle grandi organizzazioni, è rimandare tutto in attesa del processo perfetto: meglio un primo passo concreto questo mese che una strategia impeccabile fra due anni, mentre nel frattempo l'uso non governato continua. Chi vuole formare prima i propri referenti può partire dal Corso Base e dalla Credenziale Europea AI.

Adattare il piano alla propria dimensione

Lo stesso schema funziona per realtà molto diverse, a patto di calibrarlo. In uno studio o in una microimpresa, i tre mesi si comprimono e i ruoli si concentrano in una o due persone: la ricognizione è una conversazione, la policy è una pagina, lo strumento approvato si attiva in un pomeriggio. La sostanza non cambia.

In una media impresa serve più coordinamento, perché entrano reparti con bisogni diversi: conviene partire da un'area pilota, imparare, poi estendere. In un ente pubblico o in una grande organizzazione si aggiungono passaggi formali, dal confronto con le rappresentanze alle valutazioni d'impatto, e qui il rischio maggiore è la paralisi da procedura. Anche in quei contesti, un primo strumento approvato e una regola minima messi in campo subito valgono più di una governance perfetta perennemente rimandata.

Il punto fermo: la sicurezza dell'AI è fatta soprattutto di persone e processi

Conviene tornare, alla fine, a una verità che il marketing della sicurezza tende a coprire. Gli incidenti più gravi legati all'AI non nascono quasi mai da una tecnologia insufficiente. Nascono da una persona senza istruzioni, da un processo che non c'era, da una decisione presa senza che nessuno ne rispondesse.

La tecnologia conta, e va scelta e configurata bene. Ma è la parte più facile, perché si compra. Le persone formate e i processi chiari sono la parte difficile, perché si costruiscono, e sono proprio quelli che fanno la differenza tra un'organizzazione che subisce l'AI e una che la governa.

C'è un modo semplice per capire se si è sulla strada giusta. Non sta nel numero di strumenti di sicurezza installati, ma nella naturalezza con cui una persona, davanti a un dubbio o a un errore, sa a chi rivolgersi e si aspetta aiuto invece di una sanzione. Quando chiedere e segnalare diventano normali, la sicurezza ha smesso di essere un cartello sul muro e ha iniziato a essere una pratica.

Gli strumenti cambieranno, e in fretta. Le minacce pure. Ciò che resta, e che protegge davvero, è un'organizzazione di persone competenti che sanno cosa fare, dentro processi pensati prima che servissero. Quella è la sicurezza che dura.

AIPIA
ASSOCIAZIONE ITALIANA PROFESSIONISTI DELL'INTELLIGENZA ARTIFICIALE

Continua con AIPIA

Sei un professionista, un'azienda o un ricercatore che lavora con l'intelligenza artificiale? Ecco cosa mette a disposizione la prima associazione italiana dei professionisti AI.

Diventa socio — €24/anno

Contributo associativo annuale. Accesso alla Credenziale Europea, alla directory pubblica, alla tessera socio e alle convenzioni.

Iscriviti ora →

Formazione con credenziale

Corso Base (€249) e Corso Advanced (€399) con rilascio della European AI Professional Credential firmata eIDAS.

Scopri i percorsi →

Libri AIPIA

I volumi dell'associazione sull'intelligenza artificiale, dalla pratica all'etica.

Vai al catalogo →

RC Professionale AI

Polizza di Responsabilità Civile professionale dedicata a chi lavora con l'AI, riservata ai soci.

Scopri la copertura →

CONTATTI

Via Ostiense, 92 — 00154 Roma
Tel / WhatsApp: +39 06 56547987
Email: segreteria@aipia.it
PEC: aipia@altapec.it
CF: 96628540583 · aipia.it

SEGUICI

LinkedIn · Instagram · Facebook · X · YouTube · TikTok · Pinterest · Bluesky