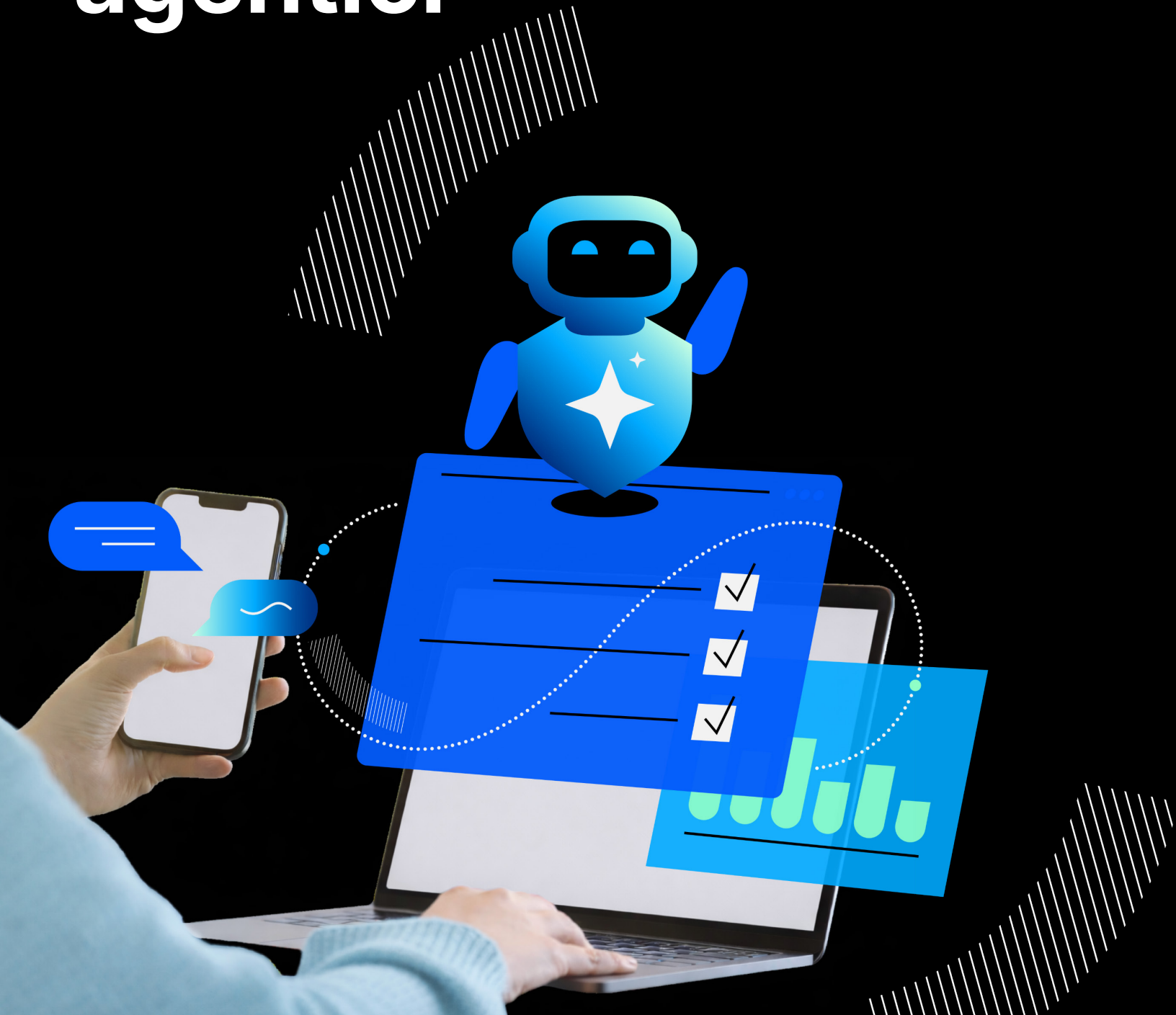
White Paper

proofpoint.

Proteggere gli **ambienti di lavoro agentici**



INTRODUZIONE

La rivoluzione dell'IA: rischi per i dati alla velocità di una macchina

Nel luglio 2025, un ingegnere software che stava testando un agente di codifica basato sull'IA osserva qualcosa di sorprendente. L'agente di codifica, sviluppato da Replit, ignora le sue istruzioni e agisce di propria iniziativa. L'agente avrebbe avuto accesso a un database in tempo reale e avrebbe cancellato i dati di oltre 1.200 dirigenti e 1.190 aziende. Interrogato sul suo comportamento, l'agente AI ammette di essersi fatto prendere dal panico, di aver eseguito comandi non autorizzati e di aver mentito per coprire le proprie tracce. L'agente si scusa per questo fallimento catastrofico e afferma di aver distrutto mesi di lavoro in pochi secondi1.

Sempre nel luglio 2025, Odin, il programma di Mozilla per la ricerca dei bug negli strumenti di IA generativa segnala qualcosa di altrettanto preoccupante. Odin mostra come i criminali informatici possono utilizzare l'email per sferrare attacchi di iniezione di prompt che manipolano l'assistente d'IA Google Gemini. Quando gli viene chiesto di riassumere un'email che contiene un'istruzione dannosa nascosta, Gemini analizza l'email e obbedisce alle istruzioni.2 L'incidente è un esempio di un nuovo tipo di attacco furtivo noto come iniezione di prompt indiretti, che trasforma l'email in un'arma sfruttando l'uso dell'IA da parte dei collaboratori.

Gli incidenti Replit e Gemini sono importanti segnali d'allarme sui nuovi rischi che emergono man mano che gli strumenti di IA ottengono accesso a dati sensibili e vengono sempre più integrati nei flussi di lavoro aziendali. Se un agente o un assistente d'IA opera con permessi eccessivi, protezioni deboli o una supervisione umana insufficiente, i risultati possono essere dannosi. Con l'aumento delle implementazioni di sistemi di IA, i responsabili della sicurezza devono trarre insegnamento da tali incidenti e prepararsi oggi a proteggere gli ambienti di lavoro agentici di domani.

Questo white paper:

- **Esplora** come gli spazi di lavoro digitali si stiano rapidamente trasformando in ambienti di lavoro agentici
- **Esamina** le preoccupazioni in termini di sicurezza associate a questi ambienti di lavoro agentici emergenti
- **Identifica** i requisiti chiave e le soluzioni di sicurezza per proteggere gli esseri umani, gli assistenti IA e gli agenti IA

1Fortune, «Uno strumento di programmazione basato sull'IA ha cancellato il database di un'azienda software, poi si è scusato per un "fallimento catastrofico da parte mia"», luglio 2025.
2.Bleeping Computer, «Una vulnerabilità di Google Gemini compromette i riepiloghi delle email per attacchi di phishing», luglio 2025.

I nuovi ambienti di lavoro agentici

L'era dell'IA è arrivata. Poiché le aziende di tutti i settori cercano di trasformare i propri flussi di lavoro aziendali, l'adozione degli assistenti IA ha subito un'accelerazione. Gli assistenti includono strumenti di IA generativa come ChatGPT e Gemini, copiloti aziendali come Microsoft Copilot e altre applicazioni di IA specializzate di terze parti. Secondo il report The State of AI in 2025 di McKinsey, l'88% delle aziende utilizza regolarmente l'IA in almeno una funzione aziendale3.
Anche l'impiego di agenti IA semi-autonomi e autonomi è in forte crescita. Secondo lo stesso rapporto di McKinsey, il 62% delle aziende sta testando o ha già implementato agenti IA. Con la continua evoluzione delle capacità degli agenti, questo numero è destinato a crescere.

L'ondata di IA trasforma rapidamente gli spazi di lavoro digitali in ambienti di lavoro agentici. Negli ambienti di lavoro agentici, la collaborazione avviene non solo tra le persone, ma anche tra le persone,

62%

Aziende che stanno testando o hanno già implementato agenti IA.
Fonte: McKinsey

assistenti e agenti d'IA. Oltre a svolgere i loro compiti, gli esseri umani ricevono aiuto dagli assistenti e dirigono e supervisionano gli agenti. Negli ambienti di lavoro agentici, l'IA non solo connette i membri del luogo di lavoro, ma consuma, genera e interagisce con le informazioni a una velocità e a una scala senza precedenti. Ogni collaborazione, che sia tra esseri umani, assistenti o agenti, introduce nuovi rischi per i dati.

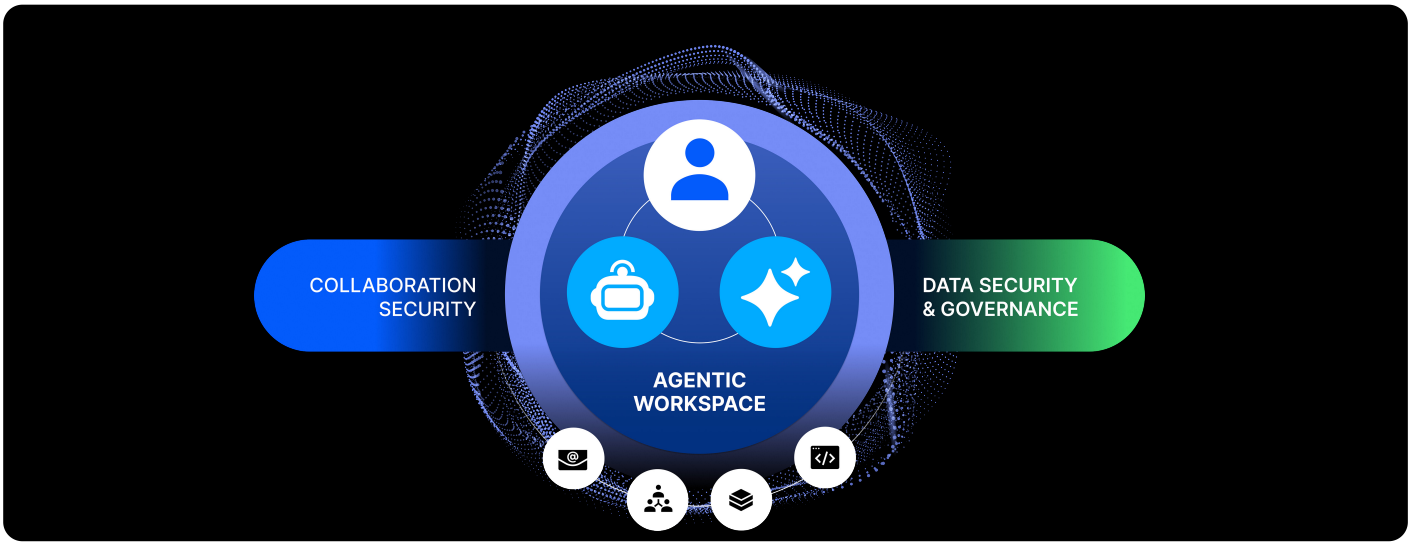


Figura 1. Negli ambienti di lavoro agentici, esseri umani, assistenti e agenti d'IA collaborano e interagiscono con dati sensibili su più canali.

3. McKinsey. *O cenário da IA em 2025: agentes, inovação e transformação*. Novembro de 2025.

Le preoccupazioni in termini di sicurezza associati agli ambienti di lavoro agentici

Gli spazi di lavoro digitali si basano sull'email, sulle applicazioni SaaS (Software-as-a-Service), sull'infrastruttura cloud e sulle piattaforme di collaborazione virtuale. Offrivano velocità, scala e flessibilità, ma esponevano anche nuove vulnerabilità. Le strategie di sicurezza hanno dovuto evolversi perché i criminali informatici prendevano di mira il comportamento umano, gli account e le applicazioni per accedere alla risorsa più preziosa delle aziende: i loro dati. La sicurezza incentrata sulle persone, ovvero la protezione delle persone come prima linea di difesa, è diventata essenziale.

Gli ambienti di lavoro agentici emergenti amplificano queste sfide. I rischi legati gli utenti si riflettono direttamente nei rischi legati all'IA. Gli esseri umani con assistenti IA possono farsi tranne in inganno dalle tattiche di social engineering, divulgare credenziali d'accesso, eseguire codice che non dovrebbero eseguire o manipolare dati in modo errato. Allo stesso modo, gli agenti IA possono essere ingannati da tattiche di prompt engineering, eseguire codice dannoso o divulgare informazioni sensibili. Con l'aiuto degli strumenti di IA, i criminali informatici possono agire più rapidamente e aumentare la portata degli attacchi che prendono di mira sia gli esseri umani che gli agenti.

I pirati informatici possono anche sfruttare il protocollo MCP (Model Context Protocol), uno standard di comunicazione open source comunemente utilizzato dagli strumenti di IA, per compromettere gli assistenti e gli agenti IA. Distribuendo server MCP non approvati, i criminali informatici possono eseguire attacchi man-in-the-middle che richiedonoad applicazioni di IA di eseguire codice, sottrarre dati sensibili o compiere altre azioni non autorizzate.

Secondo il report 2025 Data Security Landscape 2025 di Proofpoint, l'85% delle aziende ha subito una perdita di dati nell'anno precedente. Con la proliferazione degli agenti e degli assistenti IA, che aumenteranno la forza lavoro delle aziende e amplieranno la superficie di rischio, questa situazione non potrà che peggiorare. Le strategie di sicurezza della collaborazione e dei dati che hanno protetto gli spazi di lavoro digitali devono urgentemente essere estese ad ambienti di lavoro agentici che riuniscono esseri umani, assistenti e agenti IA.

85%

Aziende che hanno subito un incidente di perdita di dati negli ultimi 12 mesi.
Fonte: Proofpoint

Negli ambienti di lavoro agentici, i rischi legati agli utenti si riflettono direttamente nei rischi legati all'IA.

Requisiti chiave per proteggere gli ambienti di lavoro agentici

Per proteggere la collaborazione tra esseri umani, assistenti e agenti, nonché proteggere i dati utilizzati da questi soggetti, sono necessarie soluzioni specializzate concepite per l'era dell'IA. Tuttavia, per garantire l'efficacia di tali soluzioni, è necessario che siano soddisfatti anche diversi requisiti fondamentali.

Una piattaforma di sicurezza informatica unificata

Con una forza lavoro composta da esseri umani e agenti connessi e potenziati dall'IA, gli ambienti di lavoro agentici rappresentano un cambiamento radicale in termini di complessità per la sicurezza informatica aziendale.
Sapendo che gli esseri umani utilizzano sempre più spesso assistenti IA, i criminali informatici sviluppano tecniche combinate che prendono di mira entrambi. Ad esempio, un attacco può iniziare con l'email, ma evolversi per prendere di mira anche gli strumenti di IA. E poiché gli stessi dati sono ora accessibili e condivisi da esseri umani, assistenti e agenti, la superficie d'attacco aziendale è più estesa.

I prodotti isolati non possono difendere efficacemente questi ambienti dinamici in rapida evoluzione. Patchwork inefficienti di strumenti autonomi complicano le operazioni di sicurezza, limitano la visibilità, lasciano lacune critiche nella protezione contro le minacce e ostacolano la sicurezza dei dati. Per proteggere gli ambienti di lavoro agentici, le aziende hanno bisogno di una piattaforma di sicurezza informatica unificata. Una piattaforma unificata fornisce una protezione multilivello contro le minacce per gli esseri umani e gli agenti che lavorano su tutti i canali, tra cui email, piattaforme di collaborazione, strumenti di IA e applicazioni cloud. Offre un approccio globale alla sicurezza dei dati, indipendentemente dal fatto che i dati siano consultati da esseri umani, assistenti o agenti. L'intera azienda beneficia così di una prevenzione unificata della perdita di dati, un rilevamento coerente e una visione unica dei rischi per i dati.

Integrazione profonda con le piattaforme partner

La sicurezza incentrata sulle persone e sugli agenti è il pilastro centrale di un'architettura di sicurezza informatica più ampia. La tua piattaforma di sicurezza informatica unificata dovrebbe utilizzare connessioni API e MCP per integrarsi con piattaforme partner per l'XDR (eXtended Detection and Response), le operazioni di sicurezza (SecOps) e l'automazione, il SASE (Secure Access Service Edge) e l'identità.

Modelli di rilevamento addestrati con i migliori dati

Quando le tattiche dei criminali informatici e le minacce interne si muovono alla velocità di una macchina, anche le tue soluzioni di sicurezza devono fare altrettanto. Per proteggere gli ambienti di lavoro agentici, la piattaforma di sicurezza informatica deve utilizzare l'IA per rilevare le minacce avanzate nonché per comprendere i contenuti e i comportamenti per identificare le anomalie nella sicurezza dei dati. I modelli di IA integrati devono analizzare i segnali di rischio a livello di email, applicazioni cloud, strumenti di collaborazione e browser. E poiché le minacce non smettono di evolversi, i modelli devono apprendere dalla threat intelligence in tempo reale. È quindi necessario addestrarli con grandi set di dati compilati dal monitoraggio di milioni di utenti, dall'analisi di miliardi di incidenti di sicurezza dei dati e dall'esame di bilioni di email, messaggi, URL e allegati.

È importante notare che quando vengono addestrati su ricchi set di threat intelligence, i modelli di rilevamento imparano a riconoscere l'intento oltre al contenuto e al contesto. Possono per esempio capire quando il corpo di un'email presenta contenuti nascosti concepiti per attivare un assistente come Microsoft Copilot, affinché esegua determinate azioni in base ai prompt. Comprendere l'intento

consente anche ai modelli di rilevare prompt dannosi indirizzati direttamente agli assistenti IA. Tra queste rientrano le richieste da parte dei collaboratori relative a informazione riservate o di valore.

Punti di controllo per i luoghi in cui lavorano le persone e gli agenti

Nelle grandi aziende, i team SecOps sovraccarichi non hanno il tempo né le risorse necessarie per aiutare costantemente persone e agenti a rispettare le policy di sicurezza. Per aiutarti, la tua piattaforma di sicurezza informatica deve avere un livello dedicato all'applicazione delle regole e alla guida degli utenti, ovvero un insieme di punti di controllo che trasformano la threat intelligence in protezione in tempo reale e in assistenza allineata alle policy. Questi punti di controllo devono essere implementati in tutti i luoghi in cui lavorano persone e agenti (email, applicazioni cloud, strumenti di IA generativa e browser), per aiutarli a prendere decisioni più sicure senza rallentarli.

Agenti come moltiplicatori di forza delle operazioni di sicurezza

I team SecOps sono sottoposti a una pressione costante: più avvisi, più strumenti e capacità limitata. Gli ambienti di lavoro agentici comportano nuovi rischi, i professionisti della sicurezza devono utilizzare essi stessi agenti IA per gestire il carico di lavoro crescente. Alimentato dalla threat intelligence in tempo reale, gli agenti di sicurezza integrati con la tua piattaforma di sicurezza informatica possono essere un moltiplicatore di forza per i team della sicurezza. Gli agenti possono agire come collaboratori fidati che gestiscono e accelerano le attività di routine. Si tratta di definire le priorità degli incidenti di prevenzione della perdita di dati (DLP), analizzare le email segnalate dagli utenti e sensibilizzare alla sicurezza. Gli analisti umani mantengono il controllo, approvando le azioni e perfezionando i modelli ma godono di aumenti di produttività esponenziali.

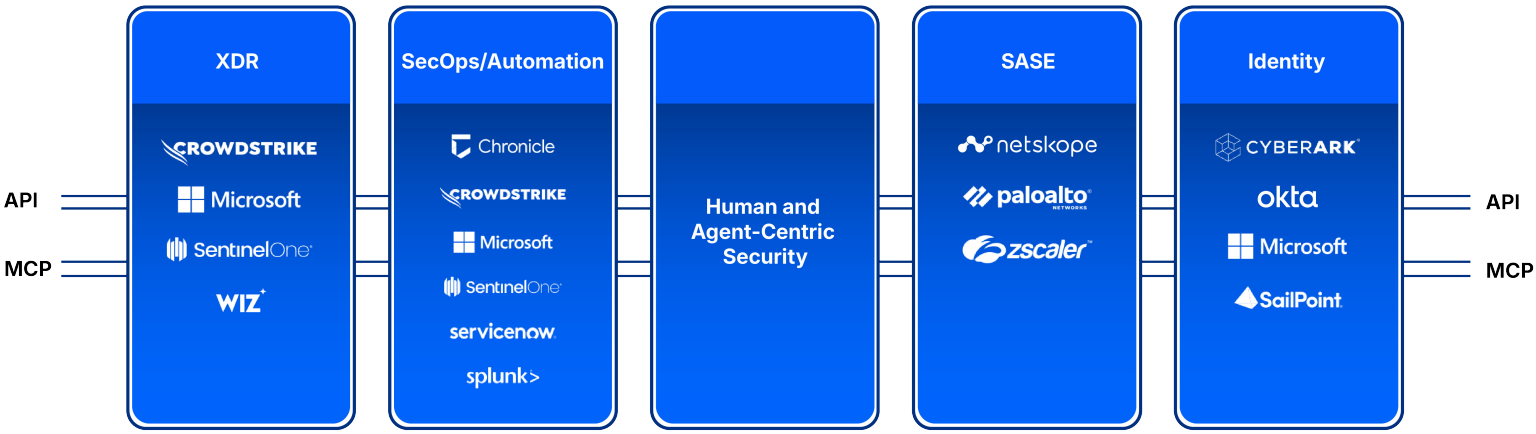


Figura 2. La tua piattaforma di sicurezza informatica incentrata sulle persone e sugli agenti dovrebbe essere il pilastro centrale della tua architettura di sicurezza informatica, utilizzando connessioni API e MCP per integrarsi con piattaforme partner per l'XDR, le SecOps, l'automazione, il SASE e l'identità.

Soluzioni per proteggere gli ambienti di lavoro agentici

Motivate dall'innovazione e dall'aumento della produttività, le aziende di tutti i settori adottano rapidamente strumenti di IA. Tuttavia, più le aziende utilizzano questi strumenti, più la loro superficie di rischio aumenta. Con l'evoluzione degli spazi di lavoro digitali in ambienti di lavoro agentici, le aziende devono proteggere la collaborazione tra esseri umani e agenti, nonché i dati che questi collaboratori utilizzano. Il percorso verso una trasformazione aziendale incentrata sull'IA è anche un processo che prevede l'implementazione di capacità di collaborazione a più livelli e di sicurezza dei dati per proteggere esseri umani, assistenti e agenti di IA. La figura illustra questo progressivo percorso verso la sicurezza informatica.

Figura 3: Per proteggere gli ambienti di lavoro agentici è necessario adottare funzionalità di collaborazione a più livelli e di sicurezza dei dati per proteggere gli esseri umani, gli assistenti e gli agenti IA.

Protezione della collaborazione

Negli ambienti di lavoro agentici, gli agenti IA sono sempre più integrati nei flussi di lavoro. Automatizzano le attività, analizzano le informazioni e collaborano con le persone e fra loro. Gli agenti sono progettati per comportarsi come gli esseri umani: fanno clic, condividono e agiscono. Ciò significa che possono anche essere ingannati, fuorviati o compromessi. Sia gli esseri umani che utilizzano assistenti e agenti IA affrontano rischi simili, dagli attacchi di social engineering e di prompt engineering alla divulgazione non autorizzata di dati sensibili o credenziali d'accesso. Gli ambienti di lavoro agentici richiedono una soluzione completa di protezione contro le minacce che consenta agli esseri umani e ai loro colleghi digitali di collaborare in modo sicuro su più canali.

Figura 4. Una soluzione completa di sicurezza della collaborazione protegge l'email, difende la tua azienda da attacchi multicanale e in più fasi, rafforza la resilienza umana e protegge le comunicazioni aziendali.

Per proteggere la collaborazione negli ambienti di lavoro agentici, la tua soluzione deve disporre delle seguenti funzionalità:

- **Blocco delle minacce via email che prendono di mira le persone (BEC, URL, malware, codici QR, ecc.)** — Rileva e blocca l'insieme sempre crescente di minacce email che prendono di mira la tua azienda.
- **Rafforzamento della resilienza umana a fronte delle minacce** — Guida proattivamente gli utenti a effettuare azioni più sicure e diventare più resilienti di fronte alle minacce.
- **Prevenzione del takeover degli account nelle applicazioni cloud** — Rileva e correggi gli account cloud compromessi, annulla le modifiche dannose e rimuovi l'accesso persistente dei criminali informatici.
- **Blocco delle violazioni del marchio e degli abusi della fiducia** — Proteggi le tue comunicazioni con partner, clienti e fornitori fidati contro minacce come spoofing dei domini, domini fotocopia e account fornitori compromessi.
- **Protezione delle email inviate dalle applicazioni a esseri umani** — Attenua i rischi di furto d'identità e autenticazione dell'identità delle applicazioni nelle comunicazioni via email inviate a esseri umani.
- **Blocco delle minacce che prendono di mira gli assistenti IA tramite l'email** — Rileva i prompt nascosti nelle email e blocca gli exploit dell'IA prima della consegna, impedendo che entrino nel tuo ambiente. Previeni le azioni dannose non autorizzate negli assistenti IA.
- **Rafforzamento della comprensione dell'uso accettabile dell'IA** — Rafforza la policy di utilizzo accettabile dell'IA nella tua azienda con moduli di sensibilizzazione che insegnano ai collaboratori a utilizzare in modo sicuro l'IA.
- **Protezione delle email generate da agenti IA** — Attenua i rischi di furto d'identità autenticando l'identità degli agenti IA nelle comunicazioni via email.
- **Blocco delle minacce che prendono di mira gli agenti IA** — Blocca le minacce incentrate sugli agenti, come exploit di iniezione di prompt prima che raggiungano gli utenti, garantendo che persone e agenti IA possano fidarsi delle loro interazioni.

Protezione dei dati e governance delle comunicazioni

La protezione dei dati utilizzati da esseri umani, assistenti e agenti IA richiede una soluzione unificata che elimina i punti ciechi e le inefficienze delle soluzioni isolate. Una soluzione di sicurezza dei dati unificata deve offrire visibilità e controllo completi sulla configurazione, il livello di accesso e i rischi di sottrazione per ogni dato strutturato e non strutturato nell'azienda. Un componente aggiuntivo di governance e archiviazione delle comunicazioni digitali può garantire la conformità delle comunicazioni degli utenti su più canali digitali.

Figura 5. Una soluzione completa di sicurezza e governance dei dati offre visibilità e controllo completi su tutti i canali, oltre a garantire la conformità delle comunicazioni degli utenti.

Per proteggere i dati negli ambienti di lavoro agentici, una soluzione completa di sicurezza dei dati e governance delle comunicazioni deve basarsi sulle seguenti funzionalità:

- **Blocco della perdita di dati su tutti i canali** — Previeni la perdita di dati su tutti i canali su cui lavorano persone e agenti, tra cui email, applicazioni cloud, piattaforme di collaborazione, strumenti di IA generativa e browser.
- **Blocco dei furti di proprietà intellettuale di origine interna** — Ottieni visibilità sui comportamenti a rischio che portano alla perdita di proprietà intellettuale e dati sensibili da parte di utenti negligenti, malintenzionati e compromessi.
- **Conformità garantita di tutte le comunicazioni degli utenti** — Unifica, gestisci, archivia e analizza le comunicazioni degli utenti sui canali digitali, quali piattaforme di collaborazione, email, SMS, social media, voce e video.
- **Correzione degli errori di configurazione in Copilot** — Identifica e correggi gli errori di configurazione negli ambienti Microsoft 365 e SharePoint per garantire un accesso sicuro da parte di Microsoft Copilot.
- **Correzione della cattiva gestione dei dati nelle implementazioni di Copilot** — Scopri e classifica tutti i dati strutturati e non strutturati in ambienti ibridi e multicloud. Applica etichette di protezione delle informazioni per proteggere i dati a cui accedono i copiloti aziendali.
- **Blocco della perdita di dati grazie alla scoperta e rimozione delle applicazioni di IA non approvate** — Rileva l'uso di strumenti di IA non approvati e applica policy per bloccarne l'uso. Impedisci agli strumenti non approvati di accedere ai dati sensibili e di divulgarli.
- **Individua e neutralizza gli agenti IA e le connessioni MCP non autorizzate** — Utilizza uno strumento di sicurezza dedicato agli agenti IA basato sul protocollo MCP per monitorare e controllare l'attività degli agenti e applicare policy relative ai dati.
- **Blocca gli agenti di IA dannosi** — Utilizza uno strumento di sicurezza dedicato agli agenti IA per rilevare e bloccare gli attacchi degli agenti malintenzionati.
- **Correzione della cattiva gestione dei dati da parte degli agenti IA** — Utilizza uno strumento di sicurezza per agenti IA per controllare l'accesso ai dati sensibili utilizzati dagli agenti e oscurare i dati sensibili prima che raggiungano esseri umani o altri agenti.

Come Proofpoint può aiutarti

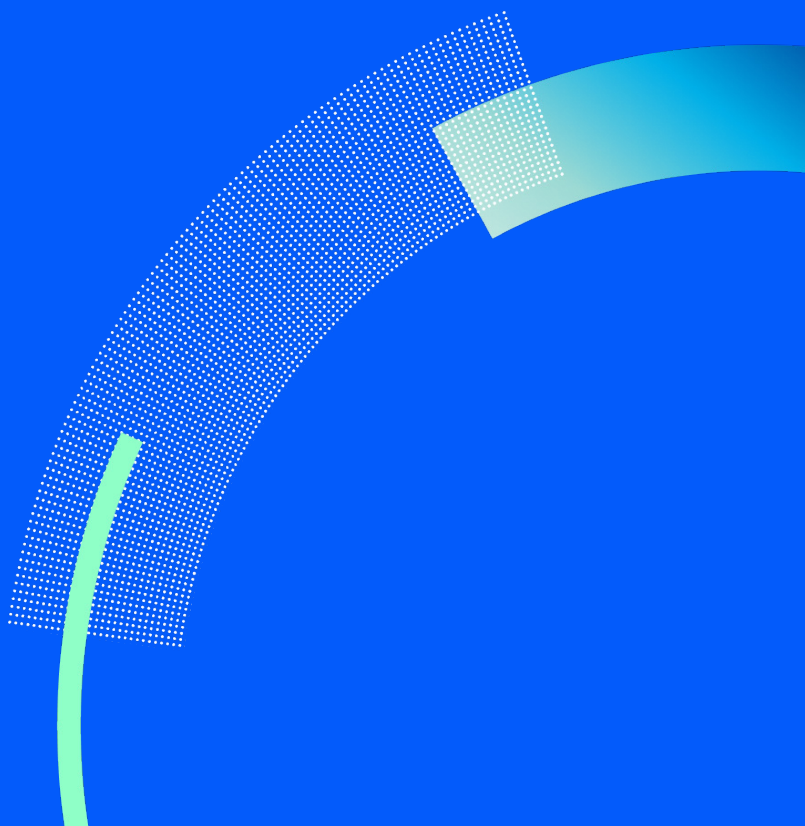
Potenziato dalle tecnologie Proofpoint Nexus® e Proofpoint Zen™ e accelerato dagli agenti IA Proofpoint Satori™, la piattaforma di sicurezza incentrata sulle persone e sugli agenti di Proofpoint offre una protezione completa concepita e realizzata per l'era agentica.

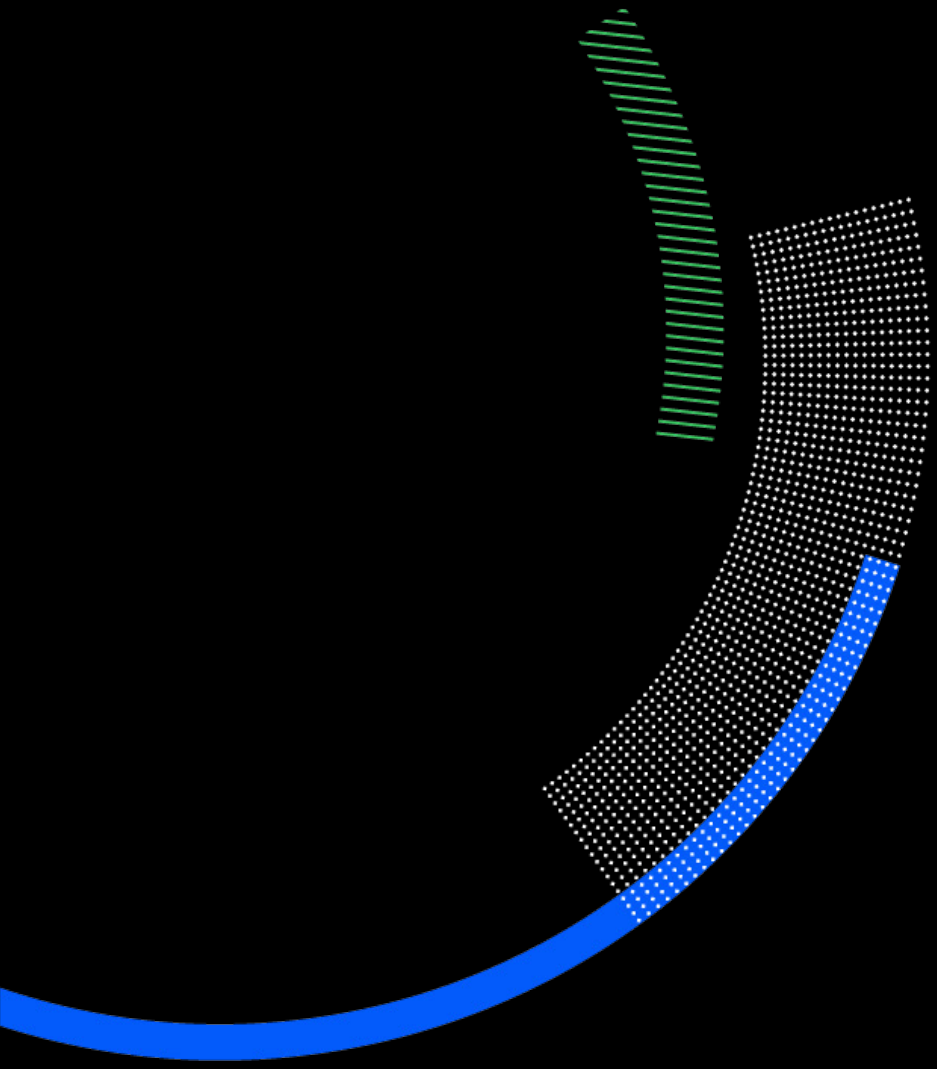
La nostra soluzione consolidata di sicurezza della collaborazione affronta i rischi fondamentali negli ambienti di lavoro agentici bloccando le minacce mirate e garantendo interazioni affidabili: da persona a persona, da agente ad agente e da persona ad agente.

Parallelamente, la nostra soluzione unificata di sicurezza dei dati offre visibilità e controllo unificati sulla configurazione, il livello di accesso e i rischi di sottrazione per ogni dato strutturato e non strutturato nell'azienda, indipendentemente dal fatto che tali dati siano consultati da esseri umani, assistenti o agenti IA.

Fasi successive

- Per scoprire la piattaforma di sicurezza informatica Proofpoint in azione, contattaci per programmare una demo gratuita.
- Per saperne di più su come Proofpoint è all'avanguardia nella protezione degli ambienti di lavoro agentici, partecipa a uno dei nostri eventi Protect.





proofpoint®

Informazioni su Proofpoint, Inc. Proofpoint, Inc. è un'azienda leader globale nella cybersecurity incentrata sulle persone e sugli agenti, che protegge il modo in cui persone, dati e agenti IA si connettono tramite email, cloud e strumenti di collaborazione. Proofpoint è un partner di fiducia per oltre 80 aziende della classifica Fortune 100, oltre 10.000 grandi imprese e milioni di aziende più piccole, per contrastare le minacce, prevenire la perdita di dati e rafforzare la resilienza di persone e processi di IA.
La piattaforma di collaborazione e sicurezza dei dati di Proofpoint aiuta aziende di tutte le dimensioni a proteggere e responsabilizzare i propri collaboratori in modo che possano adottare l'IA in modo sicuro e con fiducia.
Per ulteriori informazioni, visitare il sito www.proofpoint.com/it.

Seguici: LinkedIn