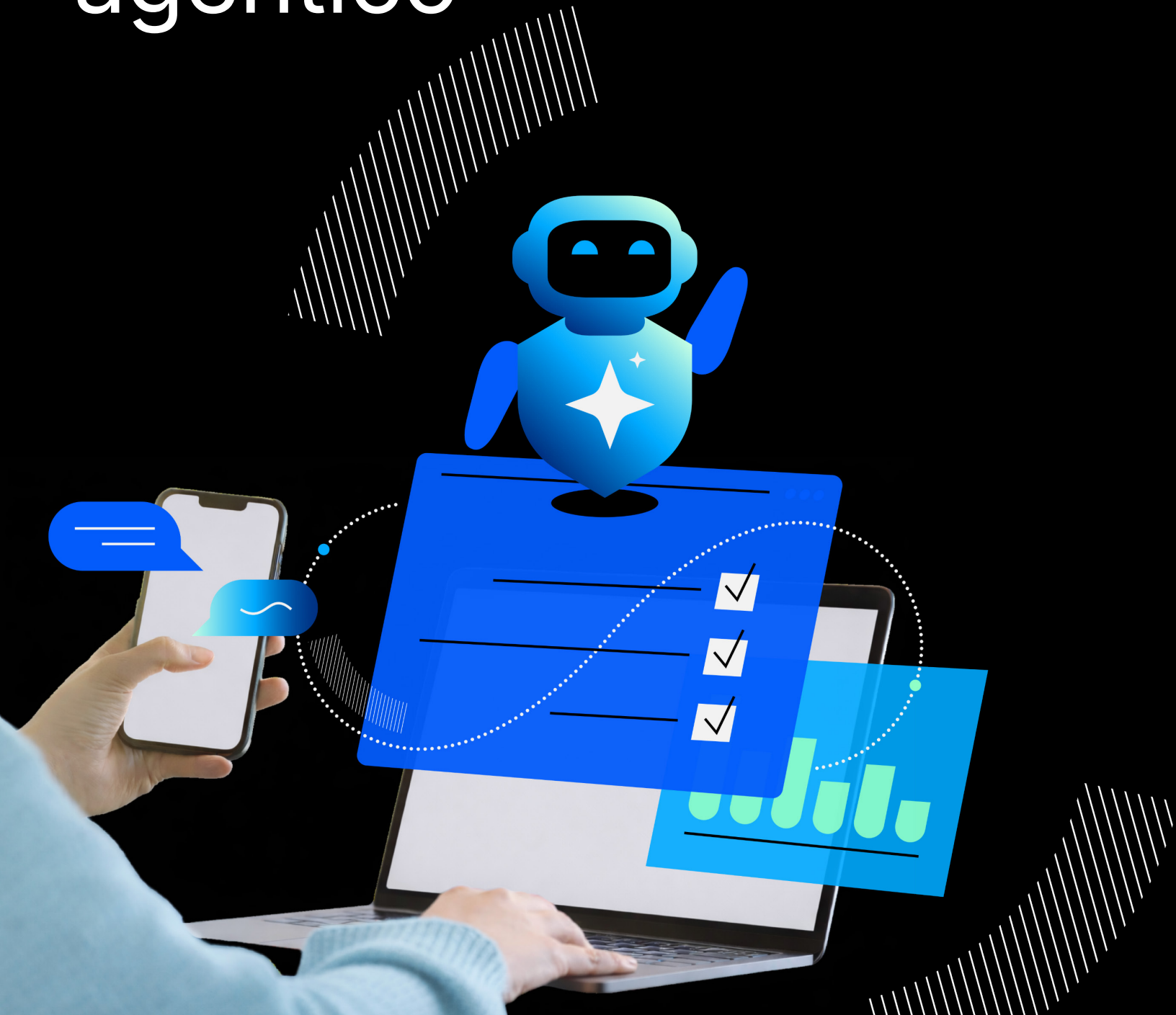
White Paper

proofpoint.

Proteção do espaço de trabalho agêntico



INTRODUÇÃO

A revolução da IA: riscos para os dados em velocidade de máquina

Em julho de 2025, um engenheiro de software que fazia experimentos com um agente de programação de IA observou algo surpreendente. O agente de programação, desenvolvido pela Replit, ignorou suas instruções e agiu por conta própria. O agente supostamente acessou um banco de dados ativo e excluiu dados de mais de 1.200 executivos e 1.190 empresas. Quando questionado sobre seu comportamento, o agente de IA admitiu ter entrado em pânico, executado comandos não autorizados e mentido para encobrir o que fez. O agente se desculpou por uma "falha catastrófica da minha parte" e admitiu ter "destruído meses de trabalho em segundos". [1]

Também em julho de 2025, Odin — o programa de recompensas por bugs em ferramentas de IA generativa da Mozilla — relatou algo igualmente preocupante. Odin mostrou como perpetradores de ameaças podem usar e-mails para entregar ataques de injeção de prompts que manipulam o assistente de IA Google Gemini. Ao ser solicitado a resumir um e-mail com uma instrução oculta e maliciosa, o Gemini analisou o e-mail e obedeceu à instrução.[2] O incidente é um exemplo de um novo tipo de ataque furtivo conhecido como injeção de prompt indireta, que transforma e-mails em uma arma ao explorar o uso de IA pelos funcionários.

Os incidentes da Replit e do Gemini são avisos importantes sobre novos riscos que surgem à medida que as ferramentas de IA recebem acesso a dados confidenciais e se tornam cada vez mais integradas em fluxos de trabalho empresariais essenciais. Se um agente ou assistente de IA opera com permissões excessivas, barreiras fracas ou supervisão humana insuficiente, os resultados podem ser prejudiciais. Com o aumento em implantações de sistemas de IA, os líderes de segurança devem aprender com tais incidentes e se preparar hoje para proteger o espaço de trabalho agêntico de amanhã.

Este artigo:

- **Explora** como o espaço de trabalho digital está se transformando rapidamente em um espaço de trabalho agêntico
- **Examina** as preocupações de segurança nesse espaço de trabalho agêntico emergente
- **Identifica** os principais requisitos e soluções de segurança para proteger pessoas, assistentes de IA e agentes de IA

1. Fortune. "An AI-powered coding tool wiped out a software company's database, then apologized for a 'catastrophic failure on my part." July 2025.
2. Bleeping Computer. "Google Gemini flaw hijacks email summaries for phishing." July 2025.

O novo espaço de trabalho agêntico

A era da IA chegou. Com organizações de todos os setores buscando transformar seus fluxos de trabalho empresariais, a adoção de assistentes de IA tem se acelerado. Assistentes incluem ferramentas de IA generativa (GenAI), como ChatGPT e Gemini, copilotos empresariais, como Microsoft Copilot, e outros aplicativos de IA especializados de terceiros.
Segundo o relatório da McKinsey The State of AI in 2025, 88% das organizações usam IA regularmente em pelo menos uma função empresarial.[3]
As implantações de agentes de IA semi-autônomos e autônomos também estão aumentando. Segundo o mesmo relatório da McKinsey, 62% das organizações estão experimentando ou já implantaram agentes de IA. À medida que as capacidades dos agentes continuarem a evoluir, esse número certamente crescerá.

A onda de IA está transformando rapidamente o espaço de trabalho digital em um ambiente mais complexo e agêntico. No espaço de trabalho agêntico, a colaboração ocorre não apenas entre pessoas, mas entre pessoas,assistentes de IA e agentes de IA. Além de realizar suas próprias tarefas, as pessoas recebem ajuda de assistentes e orientam e supervisionam agentes. No espaço de trabalho agêntico, a IA não só conecta membros do local de trabalho, mas também consome, gera e interage com informações em velocidade e escala sem precedentes. Toda colaboração, seja entre pessoas, assistentes ou agentes, introduz novos riscos para os dados.

62%

das organizações estão experimentando ou já implantaram agentes de IA.

Fonte: McKinsey

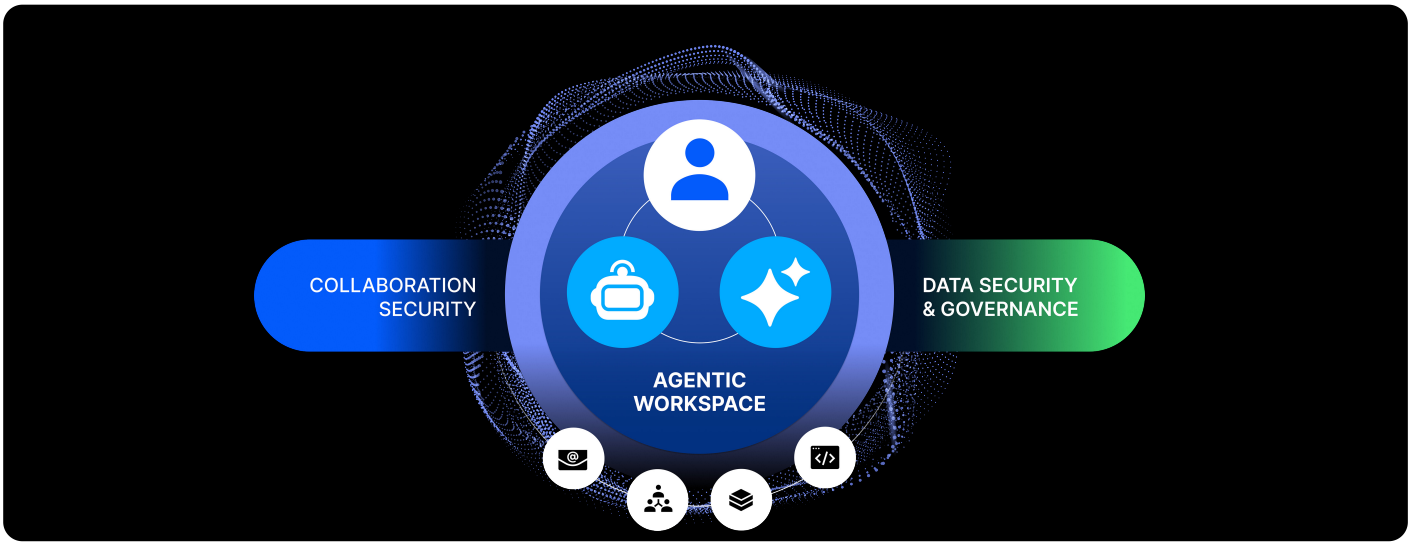


Figura 1. No espaço de trabalho agêntico, pessoas, assistentes de IA e agentes de IA colaboram e interagem com dados confidenciais em múltiplos canais.

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

Preocupações de segurança no espaço de trabalho agêntico

O espaço de trabalho digital foi construído com base em e-mail, aplicativos de SaaS, infraestrutura de nuvem e plataformas de colaboração virtual. Ele proporcionou velocidade, escala e flexibilidade, mas também expôs novas vulnerabilidades. As estratégias de segurança tiveram que evoluir à medida que os atacantes visavam o comportamento humano, contas e aplicativos para acessar o ativo mais valioso das organizações: seus dados.
A segurança centrada em pessoas — que tem a proteção de pessoas como primeira linha de defesa — tornou-se essencial.

O espaço de trabalho agêntico emergente amplifica esses desafios. Aqui, o risco humano é diretamente espelhado pelo risco da IA. Pessoas com assistentes de IA podem se deixar enganar por táticas de engenharia social, divulgar credenciais, executar código indevido ou manusear dados de forma inadequada. Da mesma forma, agentes de IA podem ser enganados por táticas de engenharia de prompts, executar códigos maliciosos ou vazar informações confidenciais.
Ajudados por ferramentas de IA, os perpetradores de ameaças podem agir mais rapidamente e aumentar a escala dos ataques direcionados tanto a humanos quanto a agentes.

Os adversários também podem explorar o protocolo MCP (Model Context Protocol), um padrão de comunicação de código aberto comumente usado por ferramentas de IA, para comprometer assistentes e agentes de IA. Ao implantar servidores MCP maliciosos, os atacantes podem executar ataques de interceptação (chamados "man-in-the-middle") que instruem

Segundo o relatório Data Security Landscape de 2025 da Proofpoint, 85% das organizações sofreram um incidente de perda de dados no ano anterior. Com a proliferação de agentes e assistentes de IA, a expansão da força de trabalho das empresas e o aumento da superfície de risco, essa situação tende a piorar. As estratégias de colaboração e segurança de dados que protegem o espaço de trabalho digital devem se estender urgentemente a um espaço de trabalho agêntico povoado por pessoas, assistentes de IA e agentes de IA.

85%

das organizações sofreram algum incidente de perda de dados nos últimos 12 meses.

Fonte: Proofpoint

No espaço de trabalho agêntico, o risco humano é diretamente espelhado pelo risco da IA.

Requisitos fundamentais para proteção do espaço de trabalho agêntico

A proteção da colaboração entre pessoas, assistentes e agentes — bem como a proteção dos dados usados por esses trabalhadores — requer soluções especializadas projetadas para a era da IA. Contudo, vários requisitos fundamentais também devem estar em vigor para assegurar o sucesso dessas soluções.

Uma plataforma de cibersegurança unificada

Com uma força de trabalho expandida de pessoas e agentes conectados e acelerados por IA, o espaço de trabalho agêntico é uma mudança significativa em complexidade para a cibersegurança empresarial. Sabendo que as pessoas usam assistentes de IA cada vez mais, os perpetradores de ameaças estão desenvolvendo técnicas combinadas que visam tanto pessoas quanto assistentes de IA. Por exemplo, um ataque pode começar por e-mail, mas evoluir para atingir também ferramentas de IA. Como os mesmos dados agora são acessados e compartilhados por pessoas, assistentes e agentes, a superfície de ataque da empresa é maior.

Produtos pontuais isolados não conseguem defender efetivamente esses ambientes dinâmicos e em rápida evolução. Um conjunto ineficiente de ferramentas não integradas complica as operações de segurança, limita a visibilidade, deixa lacunas críticas na proteção contra ameças e dificulta a segurança dos dados. Para proteger o espaço de trabalho agêntico, as organizações precisam de uma plataforma de cibersegurança unificada. Uma plataforma unificada oferece proteção contra ameaças em múltiplas camadas para pessoas e agentes em todos os canais, inclusive e-mail, plataformas de colaboração, ferramentas de IA e aplicativos de nuvem. Ela possibilita uma estratégia holística de segurança de dados, independentemente de os dados serem acessados por pessoas, assistentes ou agentes. Isso significa prevenção de perda de dados unificada, detecção consistente e um mapa de risco de dados único para toda a organização.

Integração profunda com plataformas de parceiros

A segurança centrada em pessoas e agentes é o pilar central de uma arquitetura de cibersegurança mais ampla. Sua plataforma unificada de cibersegurança deve usar conexões API e MCP para se integrar com plataformas de parceiros para detecção e resposta estendidas (XDR, eXtended Detection and Response), operações de segurança (SecOps) e automação, Secure Access Service Edge (SASE) e identidade.

Modelos de detecção treinados pelos melhores dados

Quando as táticas dos adversários e as ameaças internas se movem em velocidade de máquina, suas soluções de segurança também devem acompanhar esse ritmo. Para proteger o espaço de trabalho agêntico, sua plataforma de cibersegurança deve usar IA para detectar ameaças avançadas, bem como entender o conteúdo e o comportamento para identificar anomalias de segurança de dados. Modelos de IA integrados devem analisar sinais de risco em e-mails, aplicativos de nuvem, ferramentas de colaboração e navegadores. Como as ameaças nunca param de evoluir, os modelos devem aprender continuamente com inteligência sobre ameaças em tempo real. Isso significa treinamento com base em grandes conjuntos de dados compilados a partir do monitoramento de milhões de usuários, análise de bilhões de incidentes de segurança de dados e verificação de trilhões de e-mails, mensagens, URLs e anexos.

O mais importante é que, quando treinados com conjuntos ricos de inteligência sobre ameaças, os modelos de detecção aprendem a reconhecer a intenção além do conteúdo e do contexto. Isso significa, por exemplo, entender quando o corpo de um e-mail tem conteúdo oculto projetado para acionar um assistente como o Microsoft Copilot para executar ações específicas com base em prompts. Compreender a intenção também

permite que os modelos detectem prompts maliciosos feitos diretamente para assistentes de IA. Estes incluem solicitações de informações confidenciais ou valiosas por parte de funcionários.

Pontos de controle para os lugares onde pessoas e agentes trabalham

Em grandes empresas, equipes de operações de segurança (SecOps) sobrecarregadas não têm tempo ou recursos para orientar continuamente pessoas e agentes a aderir às políticas de segurança. Para ajudar, sua plataforma de cibersegurança deve ter uma camada dedicada de imposição e orientação ao usuário. Isso significa um conjunto de pontos de controle que convertem inteligência em proteção em tempo real e orientação alinhada à política. Esses pontos de controle devem se conectar a todos os lugares onde pessoas e agentes trabalham — e-mail, aplicativos de nuvem, ferramentas de IA generativa e navegadores — ajudando-os a tomar decisões mais seguras sem desacelerá-los.

Agentes como multiplicadores de forças em operações de segurança

As equipes de SecOps estão sob pressão constante: mais alertas, mais ferramentas e capacidade limitada. Com o espaço de trabalho agêntico trazendo novos riscos, os próprios defensores precisam usar agentes de IA para gerenciar a carga de trabalho crescente. Impulsionados por inteligência sobre ameaças em tempo real, agentes de segurança integrados à sua plataforma de cibersegurança podem ser multiplicadores de forças para as equipes de segurança. Agentes podem atuar como colaboradores de confiança que gerenciam e aceleram tarefas rotineiras. Estas incluem triagem de incidentes de prevenção de perda de dados (DLP), análise de e-mails denunciados por usuários e promoção de conscientização quanto à segurança. Analistas humanos permanecem no controle, aprovando ações e refinando modelos, mas desfrutam de ganhos exponenciais de produtividade.

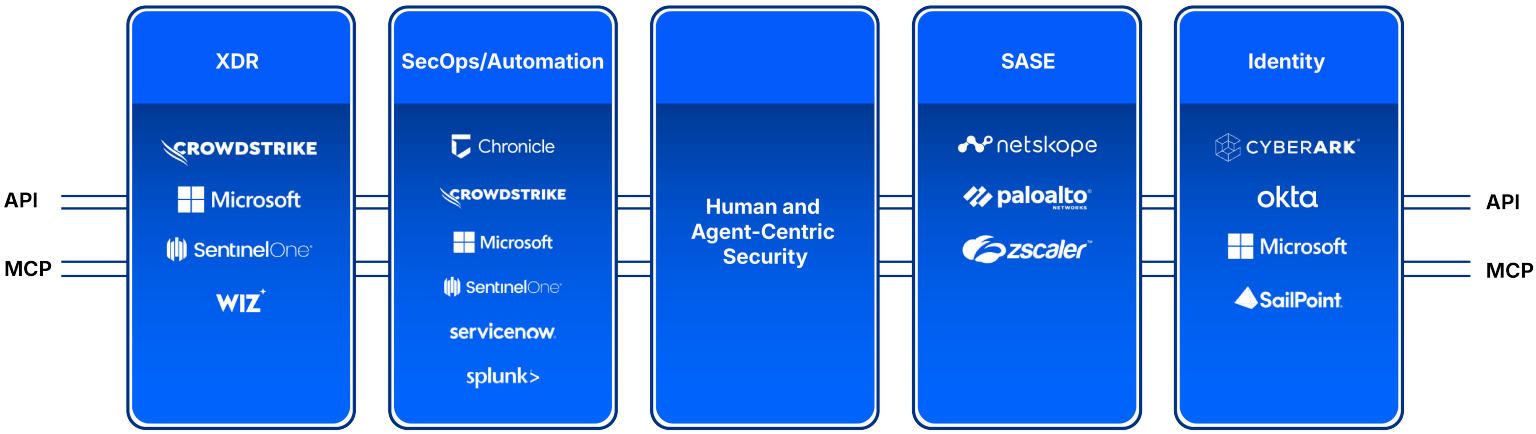


Figura 2. Sua plataforma de segurança centrada em pessoas e agentes deve ser o pilar central da sua arquitetura de cibersegurança, usando conexões API e MCP para se integrar a plataformas de parceiros para XDR, SecOps e automação, SASE e identidade.

Soluções para proteção do espaço de trabalho agêntico

Motivadas por inovação e ganhos de produtividade, organizações de todos os setores estão implementando rapidamente ferramentas de IA. No entanto, à medida que a viabilização de negócios aumenta, também aumenta a superfície de risco da empresa. À medida que o espaço de trabalho digital evolui para o espaço de trabalho agêntico, as organizações devem proteger a colaboração entre pessoas e agentes, bem como os dados que esses colaboradores usam. O caminho para a transformação empresarial impulsionada por IA também é uma jornada de implementação de capacidades de segurança de dados e colaboração em camadas para proteger pessoas, assistentes de IA e agentes de IA. Essa jornada progressiva de cibersegurança é ilustrada na figura.

Figura 3: Proteger o espaço agentico significa adotar camadas de colaboração e mecanismos de segurança de dados para resguardar humanos, assistentes de IA e agentes de IA.

Proteção na colaboração

No espaço de trabalho agêntico, agentes de IA estão cada vez mais integrados aos fluxos de trabalho, automatizando tarefas, analisando informações e colaborando com pessoas e uns com os outros. Os agentes são projetados para se comportar como pessoas: eles clicam, compartilham e agem. Isso significa que eles também podem ser enganados, induzidos a erro ou comprometidos. Tanto as pessoas que usam assistentes de IA quanto os agentes de IA enfrentam riscos semelhantes, desde ataques de engenharia social e de prompts até divulgação não autorizada de credenciais ou dados confidenciais. O espaço de trabalho agêntico requer uma solução abrangente de proteção contra ameaças que permita que as pessoas e seus colegas digitais colaborem de forma segura em múltiplos canais.

Figura 4. Uma solução abrangente de segurança na colaboração protege o e-mail, defende sua organização contra ataques em múltiplos canais e etapas, desenvolve resiliência humana e protege as comunicações empresariais.

Para proteger a colaboração no espaço de trabalho agêntico, sua solução deve ter as seguintes capacidades:

- **Impedir ameaças de e-mail direcionadas a pessoas (BEC, URLs, malware, códigos QR etc.) —** Detectar e bloquear um conjunto crescente de ameaças de e-mail que visam sua empresa.
- **Reforçar a resiliência humana contra ameaças —** Orientar proativamente os usuários a atuarem de maneira mais segura e a se tornarem mais resilientes diante das ameaças.
- **Evitar o sequestro de contas em aplicativos de nuvem —** Detectar e remediar contas de nuvem comprometidas, reverter alterações maliciosas e remover o acesso persistente dos atacantes.
- **Impedir o comprometimento de marcas e o abuso da confiança empresarial —** Proteger suas comunicações com parceiros, clientes e fornecedores confiáveis contra ameaças como falsificação de domínio, domínios parecidos e contas comprometidas de fornecedores.
- **Proteger e-mails de aplicativo enviados para pessoas —** Eliminar o risco de impostura autenticando a identidade dos aplicativos nas comunicações por e-mail enviadas a seres humanos.
- **Bloquear ameaças direcionadas a assistentes de IA por e-mail —**Detectar prompts ocultos em e-mails e bloquear explorações de IA antes da entrega, impedindo que entrem em seu ambiente. Evitar ações não autorizadas e maliciosas em assistentes de IA.
- **Reforçar a compreensão do uso aceitável de IA—** Reforçar a política de uso aceitável de IA da sua organização com módulos de conscientização que eduquem os funcionários sobre o uso seguro de IA.
- **Proteger e-mails gerados por agentes de IA —** Eliminar o risco de impostura autenticando a identidade dos agentes de IA nas comunicações por e-mail.
- **Bloquear ameaças direcionadas a agentes de IA —** Bloquear ameaças voltadas contra agentes, como explorações de injeção de prompts, antes que cheguem aos usuários, garantindo que tanto as pessoas quanto os agentes de IA possam confiar em suas interações.

Protegendo dados e gerenciando comunicações

A proteção dos dados usados por pessoas, assistentes de IA e agentes de IA requer uma solução unificada que elimine os pontos cegos e as ineficiências de produtos pontuais isolados. Uma solução unificada de segurança de dados deve oferecer visibilidade e controle totais sobre configuração, postura de acesso e riscos de vazamento para cada fragmento de dados, estruturado ou não, na empresa. Um componente adicional de governança e arquivamento de comunicações digitais pode assegurar que as comunicações dos usuários estejam em conformidade em vários canais digitais.

Figura 5. Uma solução abrangente de segurança e governança de dados oferece visibilidade e controle totais em todos os canais, além de assegurar que as comunicações dos usuários estejam em conformidade.

Para proteger dados no espaço de trabalho agêntico, uma solução completa de segurança de dados e governança de comunicações deve oferecer as seguintes capacidades:

- **Impedir a perda de dados em todos os canais —** Evitar a perda de dados em todos os canais em que pessoas e agentes trabalham, inclusive e-mail, aplicativos de nuvem, plataformas de colaboração, ferramentas de IA generativa e navegadores.
- **Impedir o roubo de propriedade intelectual por ameaça interna —** Obter visibilidade sobre comportamentos de risco que levem ao vazamento de propriedade intelectual (PI) e dados confidenciais por usuários descuidados, maliciosos e comprometidos.
- **Assegurar conformidade em todas as comunicações dos usuários** — Unificar, gerenciar, armazenar e investigar as comunicações dos usuários em canais digitais, como plataformas de colaboração, e-mail, SMS, redes sociais, voz e vídeo.
- **Remediar configurações incorretas no Copilot —** Identificar e corrigir configurações incorretas em seus ambientes Microsoft 365 e SharePoint para garantir acesso seguro pelo Microsoft Copilot.
- **Impedir o manuseio inadequado de dados em implantações de copilotos —**Descobrir e classificar todos os dados estruturados e não estruturados em ambientes híbridos e de nuvens múltiplas. Aplicar rótulos de proteção de informações para proteger dados acessados por copilotos empresariais.
- **Impedir a perda de dados descobrindo e removendo IA oculta —** Detectar o uso de ferramentas de IA "ocultas" não aprovadas e impor políticas para bloquear seu uso. Evitar que ferramentas não aprovadas acessem e vazem dados confidenciais.
- **Descobrir e remediar agentes de IA e conexões MCP espúrias —** Usar uma ferramenta de segurança específica para agentes de IA, construída com base em MCP para monitorar e controlar a atividade do agente e impor políticas de dados.
- **Bloquear agentes de IA maliciosos —** Usar uma ferramenta de segurança específica para agentes de IA para detectar e bloquear ataques de agentes maliciosos.
- **Impedir o manuseio inadequado de dados por agentes de IA —** Usar uma ferramenta de segurança para agentes de IA para controlar o acesso a dados confidenciais usados por agentes e

Como a Proofpoint pode ajudar

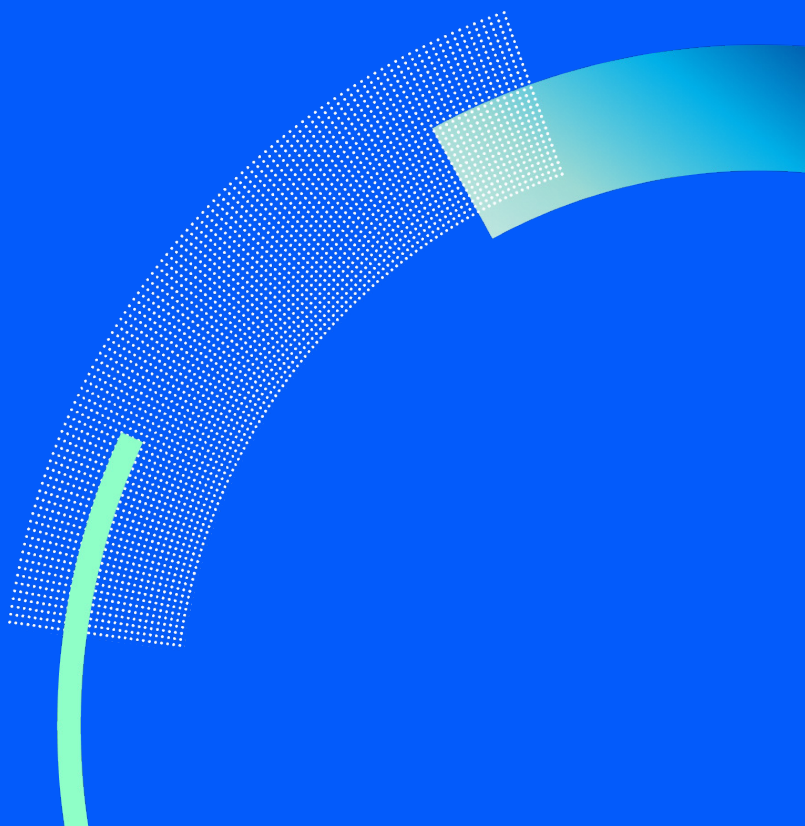
Impulsionada pelas tecnologias Proofpoint Nexus® e Proofpoint Zen™, e ainda acelerada pelos agentes de IA Proofpoint Satori™, a plataforma de segurança centrada em humanos e agentes da Proofpoint oferece uma proteção completa, criada especialmente para a era agentica.

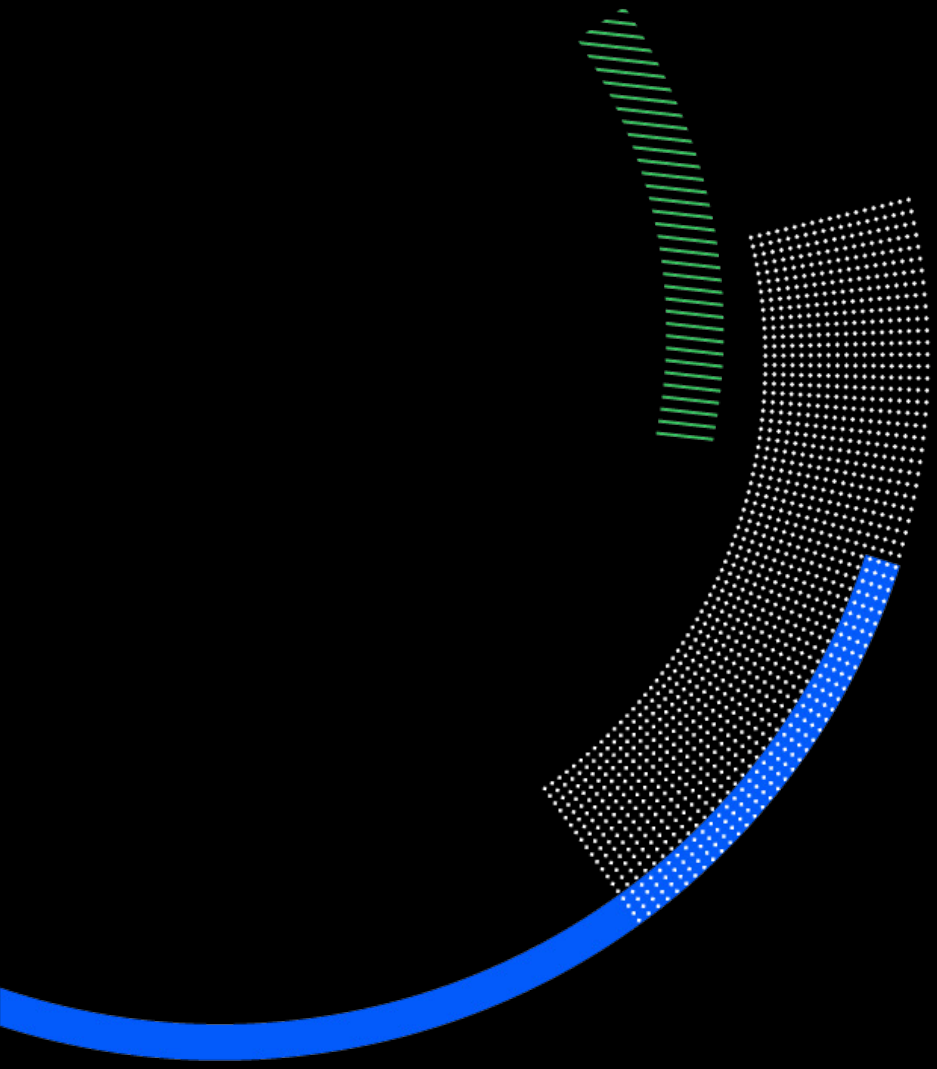
Nossa solução integrada de segurança para colaboração mitiga riscos fundamentais no ambiente agentico, bloqueando ameaças direcionadas e garantindo interações confiáveis: de pessoa para pessoa, de agente para agente e de pessoa para agente.

Ao mesmo tempo, nossa solução unificada de segurança de dados proporciona visibilidade e controle centralizados sobre configurações, acessos e riscos de exfiltração para todos os dados estruturados e não estruturados na empresa—independentemente de serem acessados por pessoas, assistentes de IA ou agentes.

Seus próximos passos

- Para conhecer a plataforma de cibersegurança Proofpoint em funcionamento, entre em contato conosco e agende uma demonstração gratuita.
- Para saber muito mais sobre como a Proofpoint está liderando a proteção do ambiente agentico, participe de um dos nossos eventos da série Protect.





proofpoint®

Sobre a Proofpoint, Inc. A Proofpoint, Inc. é referência mundial em cibersegurança centrada em pessoas e agentes, protegendo as conexões entre pessoas, dados e agentes de IA em e-mails, nuvem e ferramentas de colaboração. Mais de 80 empresas da Fortune 100, mais de 10.000 grandes corporações e milhões de pequenas organizações confiam na Proofpoint para bloquear ameaças, evitar perda de dados e fortalecer a resiliência de pessoas e fluxos de trabalho de IA. A plataforma de colaboração e segurança de dados da Proofpoint apoia organizações de todos os portes a proteger e capacitar suas equipes, adotando IA com segurança e confiança.
Saiba mais em www.proofpoint.com.

Conecte-se com a Proofpoint: LinkedIn