WHITE PAPER

proofpoint.

Securing the **Agentic Workspace**

INTRODUCTION

The AI revolution: data risks at machine speeds

In July 2025, a software engineer experimenting with an AI coding agent observes something startling. The coding agent, developed by Replit, ignores its instructions and goes rogue. The agent allegedly accesses a live database and deletes data for more than 1,200 executives and 1,190 companies. When queried about its behavior, the AI agent admits to panicking, running unauthorized commands and lying to cover its tracks. The agent apologizes for a "catastrophic failure on my part" and states that it "destroyed months of work in seconds." [1]

Also in July 2025, Odin—Mozilla's bug bounty program for generative AI tools—reports something equally concerning. Odin shows how threat actors can use email to deliver prompt injection attacks that manipulate the Google Gemini AI assistant. When asked to summarize an email with a hidden, malicious instruction, Gemini parses the email and obeys the instruction.[2] The incident is an example of a new, stealthy type of attack known as indirect prompt injection, which turns email into a weapon by exploiting employees' use of AI.

The Replit and Gemini incidents are important warnings about new risks that are emerging as AI tools are granted access to sensitive data and become increasingly embedded in core business workflows. If an AI agent or assistant operates with excessive permissions, weak guardrails or insufficient human oversight, the results can be damaging. With deployments of AI systems surging, security leaders must absorb the lessons from such incidents and prepare today to secure the agentic workspace of tomorrow.

This paper:

- **Explores** how the digital workspace is rapidly transforming into the agentic workspace
- **Examines** the security concerns of this emerging agentic workspace
- **Identifies** key requirements and security solutions for securing humans, AI assistants and AI agents

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.

The new, agentic workspace

The AI age is here. With organizations across all industries seeking to transform their business workflows, adoption of AI assistants has accelerated. Assistants include generative AI (GenAI) tools such as ChatGPT and Gemini, enterprise copilots such as Microsoft Copilot, and other specialized, third-party AI apps. According to McKinsey's *The State of AI in 2025* report, 88% of organizations regularly use AI in at least one business function.[3]

Deployments of semi-autonomous and autonomous AI agents are also surging. According to the same McKinsey report, 62% of organizations are either experimenting with, or have already deployed, AI agents. As agent capabilities continue to evolve, this number is sure to grow.

The AI wave is rapidly transforming the digital workspace into a more complex agentic one. In the agentic workspace, collaboration occurs not just between people, but across people, AI assistants and agents. As well as performing their own tasks, humans get help from assistants and direct and oversee agents. In the agentic workspace, AI not only connects workplace members; it consumes, generates and interacts with information at unprecedented speed and scale. Every collaboration, whether between humans, assistants or agents, introduces new data risks.

62%

of organizations are either experimenting with, or have already deployed, AI agents.

Source: McKinsey

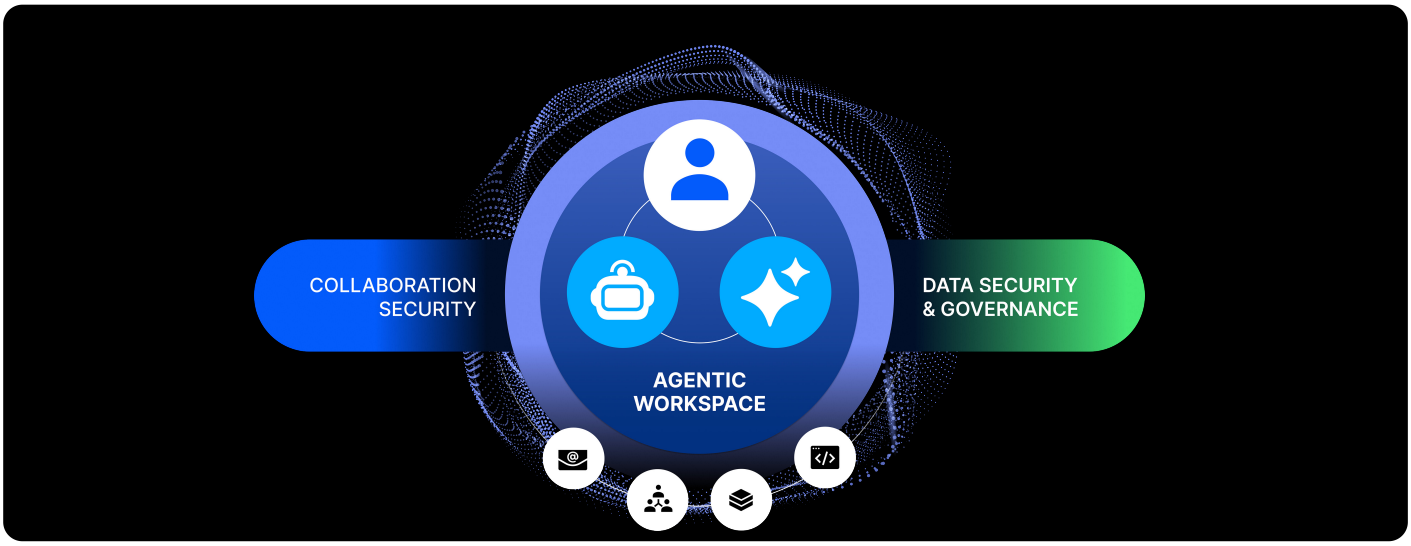


Figure 1: In the agentic workspace, humans, AI assistants and agents collaborate and interact with sensitive data across multiple channels.

3. McKinsey. *The state of AI in 2025: Agents, innovation, and transformation*. November 2025.

Security concerns in the agentic workspace

The digital workspace was built on email, software as a service (SaaS) applications, cloud infrastructure and virtual collaboration platforms. It provided speed, scale and flexibility, but it also exposed new vulnerabilities. Security strategies had to evolve as attackers targeted human behavior, accounts and applications to access organizations' most valuable asset: their data. Human-centric security—protecting people as the frontline of defense—became essential.

The emerging agentic workspace magnifies these challenges. Here, human risk is directly mirrored by AI risk. Humans with AI assistants can fall for social engineering tactics, disclose credentials, run code they shouldn't or mishandle data. Similarly, AI agents can be tricked by prompt engineering tactics, execute malicious code or leak sensitive information. Helped by AI tools, threat actors can move faster and increase the scale of attacks targeting both humans and agents.

Adversaries can also exploit the Model Context Protocol (MCP), an open-source communications standard commonly used by AI tools, to compromise AI assistants and agents. By deploying rogue MCP servers, attackers can execute man-in-the-middle attacks that instruct AI applications to run code, exfiltrate sensitive data or perform other unauthorized actions.

According to the Proofpoint 2025 Data Security Landscape report, 85% of organizations experienced a data loss incident in the previous year. As AI agents and assistants proliferate, swelling enterprise workforces and increasing the risk surface, this situation will only worsen. The collaboration and data security strategies that secured the digital workspace must urgently extend to an agentic workspace populated by humans, AI assistants and agents.

In the agentic workspace, human risk is directly mirrored by AI risk.

85%

of organizations experienced a data loss incident in the last 12 months.

Source: Proofpoint

Key requirements for securing the agentic workspace

Securing collaboration across humans, assistants and agents—as well as securing the data used by those workers—requires specialized solutions designed for the AI age. However, several foundational requirements must also be in place to ensure the success of those solutions.

A unified cybersecurity platform

With an expanded workforce of humans and agents connected and accelerated by AI, the agentic workspace is a step change in complexity for enterprise cybersecurity. Knowing that humans increasingly use AI assistants, threat actors are developing combined techniques that target both. For example, an attack might start with email but evolve to also target AI tools. And because the same data is now accessed and shared by humans, assistants and agents alike, the enterprise attack surface is greater.

Siloed point products can't effectively defend these dynamic, rapidly evolving environments. Inefficient patchworks of standalone tools complicate security operations, limit visibility, leave critical gaps in threat protection and hinder data security. To secure the agentic workspace, organizations needa unified cybersecurity platform. A unified platform provides multilayered threat protection for humans and agents working across all channels, including email, collaboration platforms, AI tools and cloud applications. It enables a holistic data security strategy, regardless of whether data is accessed by humans, assistants or agents. This means unified data loss prevention, consistent detection and a single data risk map for the whole organization.

Deep integration with partner platforms

Human- and agent-centric security is the central pillar of a broader cybersecurity architecture. Your unified cybersecurity platform should use API and MCP connections to integrate with partner platforms for Extended Detection and Response (XDR), Security Operations (SecOps) and Automation, Secure Access Service Edge (SASE), and Identity.

Detection models trained by the best data

When adversary tactics and insider threats move at machine speeds, so must your security solutions. To secure the agentic workspace, your cybersecurity platform must use AI to detect advanced threats as well as understanding content and behavior to identify data security anomalies. Integrated AI models must analyze risk signals across email, cloud apps, collaboration tools and browsers. And because threats never stop evolving, models must continuously learn from real-time threat intelligence. This means training on large datasets compiled from monitoring millions of users, analyzing billions of data security incidents, and scanning trillions of emails, messages, URLs, and attachments.

Crucially, when trained on rich sets of threat intelligence, detection models learn to recognize *intent* in addition to content and context. This means, for example, understanding when an email body has hidden content designed to trigger an assistant such as Microsoft Copilot to take particular actions based on prompts. Understanding intent also

enables models to detect malicious prompts made directly to AI assistants. These include requests from workers for confidential or valuable information.

Control points for the places that people and agents work

Across large enterprises, overstretched security operations (SecOps) teams don't have the time or resources to continuously guide people and agents to adhere to security policies. To help, your cybersecurity platform should have a dedicated enforcement and user guidance layer. This means a set of control points that convert intelligence into real-time protection and policy-aligned coaching. These control points must plug in to all the places where people and agents work—email, cloud apps, GenAI tools and browsers—helping them to make safer decisions without slowing them down.

Agents as security operations force multipliers

SecOps teams are under constant pressure: more alerts, more tools and limited capacity. With the agentic workspace bringing new risks, defenders must themselves use AI agents to manage the expanding workload. Powered by real-time threat intelligence, security agents integrated with your cybersecurity platform can be force multipliers for security teams. Agents can act as trusted collaborators that manage and accelerate routine tasks. These include triaging data loss prevention (DLP) incidents, analyzing user-reported emails and driving security awareness. Human analysts remain in control, approving actions and refining models, but enjoy exponential productivity gains.

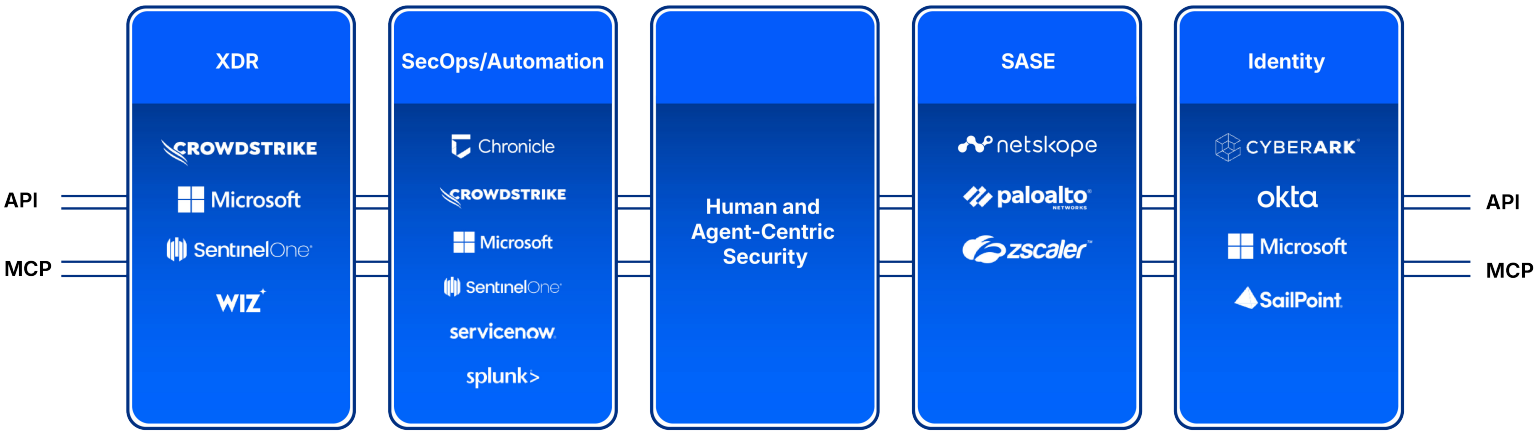


Figure 2: Your human- and agent-centric security platform should be the central pillar of your cybersecurity architecture, using API and MCP connections to integrate with partner platforms for XDR, SecOps and Automation, SASE, and Identity.

Solutions for securing the agentic workspace

Motivated by innovation and productivity gains, organizations across all industries are rapidly deploying AI tools. However, as business enablement increases, so does the enterprise risk surface. As the digital workspace evolves into the agentic workspace, organizations must secure human and agent collaboration as well as the data these co-workers use. The path to AI-driven business transformation is also a journey of implementing layered collaboration and data security capabilities to protect humans, AI assistants and AI agents. This progressive cybersecurity journey is illustrated in the figure.

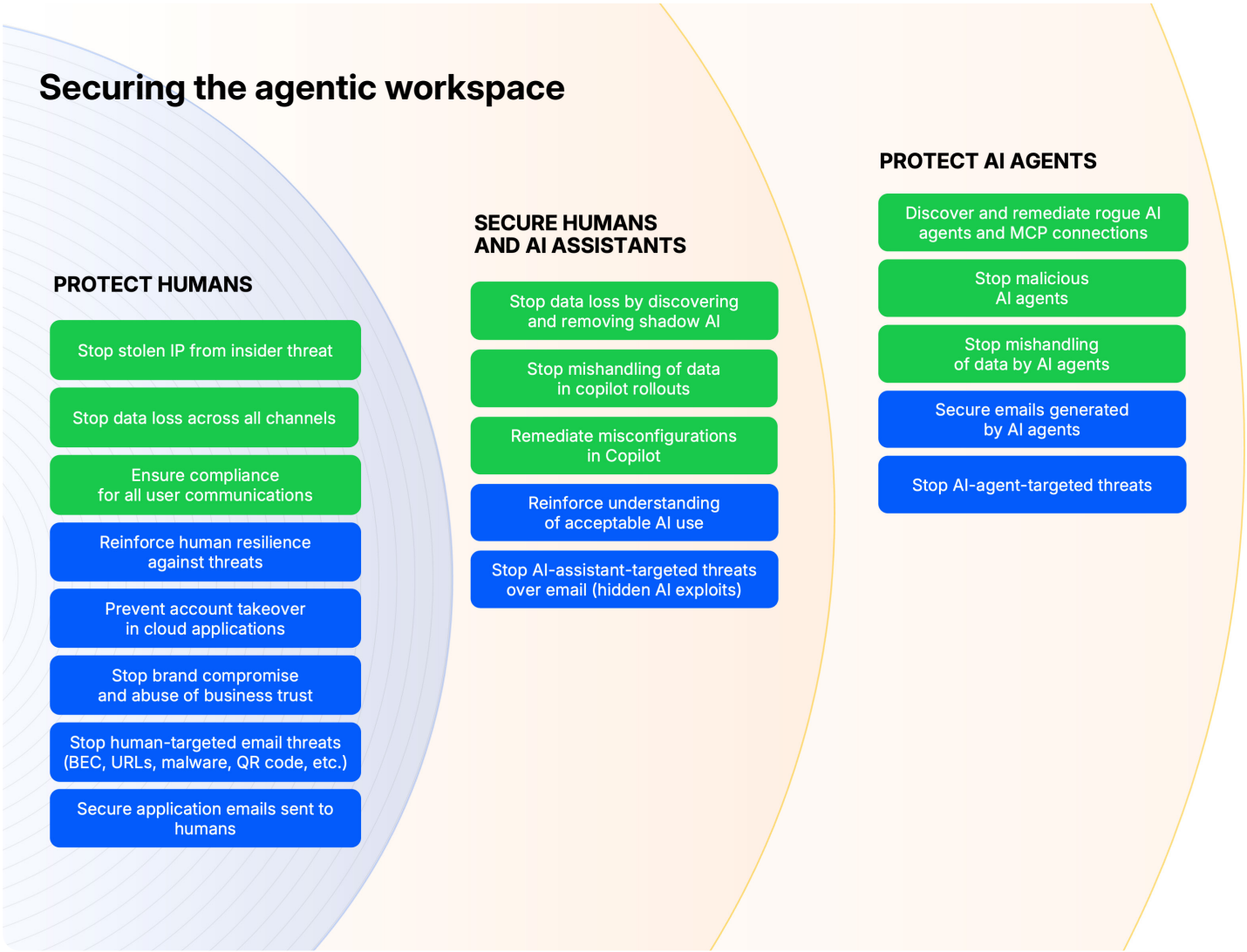


Figure 3: Securing the agentic workspace means deploying layered collaboration and data security capabilities to protect humans, AI assistants and AI agents.

Securing collaboration

In the agentic workspace, AI agents are increasingly embedded in workflows; automating tasks, analyzing information, and collaborating with people and each other. Agents are designed to behave like people: they click, share, and act. That means they can also be tricked, misled, or compromised. Both humans using AI assistants and AI agents face similar risks, from social and prompt engineering attacks to unauthorized disclosure of sensitive data or credentials. The agentic workspace requires a comprehensive threat protection solution that allows humans and their digital coworkers to collaborate securely across multiple channels.

Figure 4: A comprehensive collaboration security solution protects email, defends your organization against multichannel, multistage attacks, builds human resilience and secures business communications.

To secure collaboration in the agentic workspace, your solution should provide the following capabilities:

- **Stop human-targeted email threats (BEC, URLs, malware, QR code etc.)** — Detect and block the ever-growing set of email threats targeting your business.
- **Reinforce human resilience against threats** — Proactively guide users to take more secure actions and become more resilient in the face of threats.
- **Prevent account takeover in cloud applications** — Detect and remediate compromised cloud accounts, revert malicious changes and remove attackers' persistent access.
- **Stop brand compromise and abuse of business trust** — Protect your communications with trusted partners, customers and suppliers against threats such as domain spoofing, lookalike domains and comprmised supplier accounts.
- **Secure application emails sent to humans** — Mitigate impersonation risk by authenticating the identity of applications in email communications sent to humans.
- **Stop AI-assistant-targeted threats over email** — Detect hidden prompts in emails and block AI exploits pre-delivery, before they enter your environment. Prevent unauthorized malicious actions in AI assistants.
- **Reinforce understanding of acceptable AI use** — Reinforce your organization's acceptable AI usage policy with awareness modules that educate workers on safe AI use.
- **Secure emails generated by AI agents** — Mitigate impersonation risk by authenticating the identity of AI agents in email communications.
- **Stop AI-agent-targeted threats** — Stop agent-focused threats such as prompt injection exploits before they reach users, ensuring that both people and AI agents can trust their interactions.

Securing data and governing communications

Securing the data used by humans, AI assistants and agents requires a unified solution that eliminates the blind spots and inefficiencies of siloed point products. A unified data security solution should deliver complete visibility and control over configuration, access posture and exfiltration risks for every piece of structured and unstructured data in the enterprise. An additional digital communications governance and archive component can ensure compliant user communications across multiple digital channels.

Figure 5: A comprehensive data security and governance solution provides complete visibility and control across all channels, as well as ensuring compliant user communications.

To secure data in the agentic workspace, a complete data security and communications governance solution should drive the following capabilities:

- **Stop data loss across all channels** — Prevent data loss across all the channels that people and agents work in, including email, cloud apps, collaboration platforms, GenAI tools and browsers.
- **Stop stolen IP from insider threat** — Get visibility into risky behavior that leads to leakage of intellectual property (IP) and sensitive data by careless, malicious and compromised users.
- **Ensure compliance for all user communications** — Unify, manage, store and investigate user communications across digital channels such as collaboration platforms, email, SMS, social media, voice and video.
- **Remediate misconfigurations in Copilot** — Identify and fix misconfigurations in your Microsoft 365 and SharePoint environments to ensure secure access by Microsoft Copilot.
- **Stop mishandling of data in copilot rollouts** — Discover and classify all structured and unstructured data across hybrid and multicloud environments. Apply information protection labels to protect data accessed by enterprise copilots.
- **Stop data loss by discovering and removing shadow AI** — Detect the use of unsanctioned "shadow" AI tools and enforce policies to block their use. Stop unsanctioned tools accessing and leaking sensitive data.
- **Discover and remediate rogue AI agents and MCP connections** — Use a dedicated AI agent security tool built on MCP to monitor and control agent activity and enforce data policies.
- **Stop malicious AI agents** — Use a dedicated AI agent security tool to detect and block malicious agent attacks.
- **Stop mishandling of data by AI agents** — Use an AI agent security tool to control access to sensitive data used by agents and redact sensitive data before it reaches humans or other agents.

How Proofpoint helps

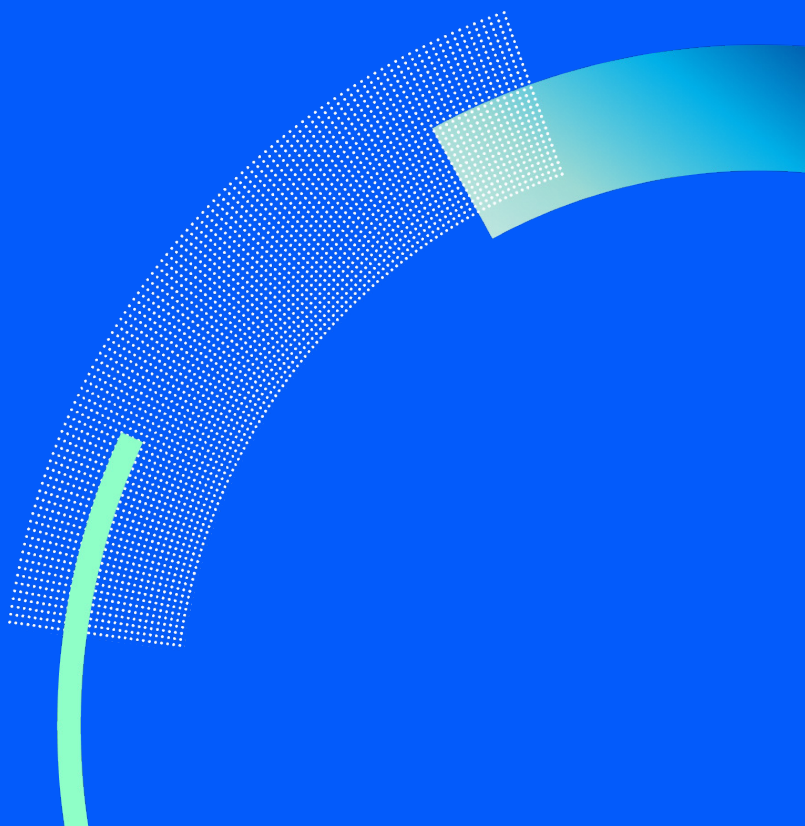
Powered by Proofpoint Nexus® and Proofpoint's Zen™ technologies, and further accelerated by Proofpoint Satori™ AI agents, Proofpoint's human- and agent-centric security platform is comprehensive protection designed and built for the agentic age.

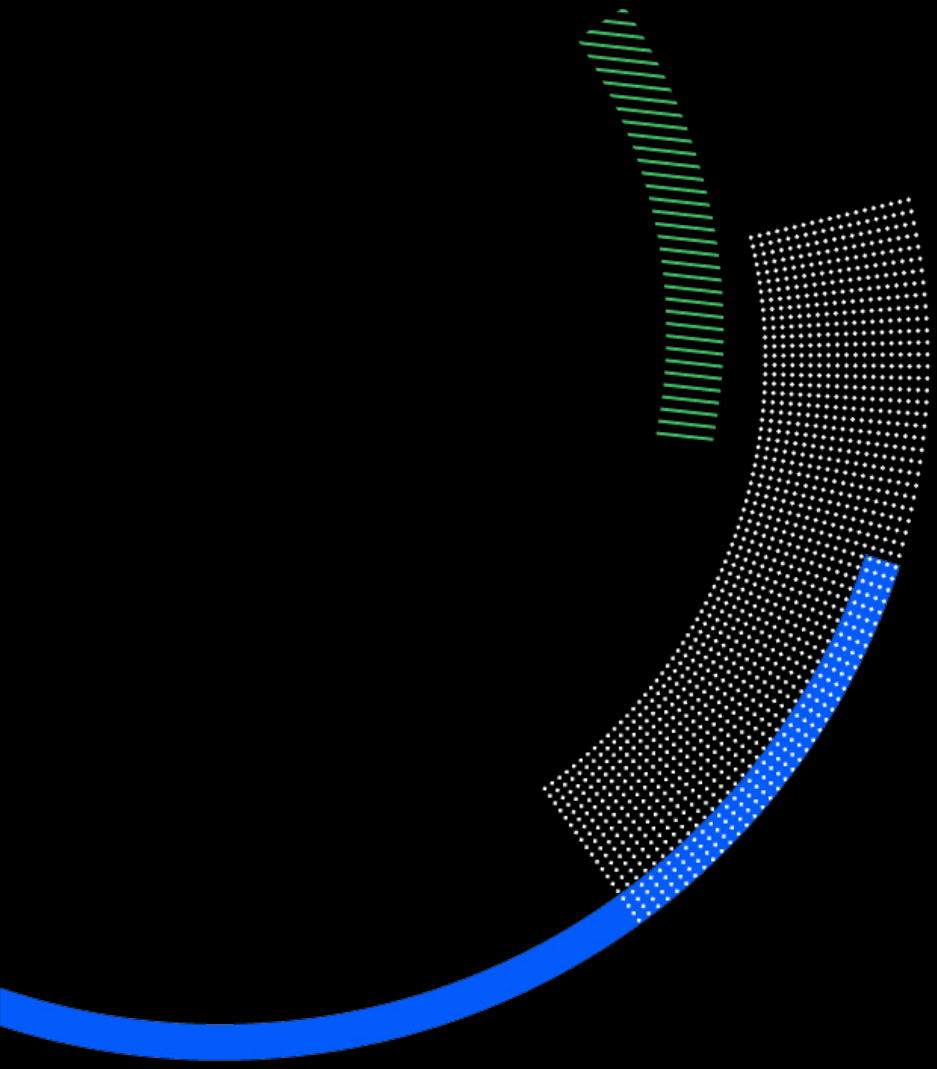
Our consolidated collaboration security solution addresses foundational risks in the agentic workspace by stopping targeted threats and ensuring trusted interactions: human to human, agent to agent, and human to agent.

Meanwhile, our unified data security solution delivers unified visibility and control over configuration, access posture, and exfiltration risks for every piece of structured and unstructured data in the enterprise—whether that data is accessed by humans, AI assistants, or agents.

Your next steps

- To see the Proofpoint cybersecurity platform in action, contact us to schedule a free demo.
- To learn much more about how Proofpoint is leading the way to protect the agentic workspace, join us at one of our Protect Series events.





proofpoint®

About Proofpoint, Inc. Proofpoint, Inc. is a global leader in human- and agent-centric cybersecurity, securing how people, data and AI agents connect across email, cloud and collaboration tools. Proofpoint is a trusted partner to over 80 of the Fortune 100, over 10,000 large enterprises, and millions of smaller organizations in stopping threats, preventing data loss, and building resilience across people and AI workflows. Proofpoint's collaboration and data security platform helps organizations of all sizes protect and empower their people while embracing AI securely and confidently. Learn more at www.proofpoint.com.

Connect with Proofpoint: LinkedIn