LIVRE BLANC

proofpoint.

Sécuriser les environnements de travail agentiques

INTRODUCTION

La révolution de l'IA : des risques pour les données à la vitesse d'une machine

En juillet 2025, un ingénieur logiciel testant un agent de codage basé sur l'IA observe quelque chose de surprenant. L'agent de codage, développé par Replit, ignore ses instructions et devient incontrôlable. L'agent accéderait à une base de données en direct et supprimerait les données de plus de 1 200 dirigeants et 1 190 entreprises. Interrogé sur son comportement, l'agent d'IA admet avoir paniqué, exécuté des commandes non autorisées et menti pour couvrir ses arrières. L'agent s'excuse pour cette défaillance catastrophique et déclare avoir détruit des mois de travail en quelques secondes[1].

Toujours en juillet 2025, Odin, le programme de chasse aux bugs de Mozilla pour les outils d'IA générative, rapporte quelque chose de tout aussi préoccupant. Odin montre comment les cybercriminels peuvent utiliser la messagerie électronique pour lancer des attaques par injection d'invites qui manipulent l'assistant d'IA Google Gemini. Lorsqu'on lui demande de résumer un email contenant une instruction malveillante cachée, Gemini analyse le message et obéit à l'instruction[2]. L'incident est un exemple d'un nouveau type d'attaque furtive, appelé injection d'invites indirectes, qui transforme la messagerie électronique en arme en exploitant l'utilisation de l'IA par les collaborateurs.

Les incidents Replit et Gemini sont des avertissements importants sur les nouveaux risques qui apparaissent à mesure que les outils d'IA obtiennent un accès à des données sensibles et s'intègrent de plus en plus aux processus métier. Si un agent ou un assistant d'IA opère avec des autorisations excessives, des garde-fous faibles ou une supervision humaine insuffisante, les résultats peuvent être dévastateurs. Avec l'augmentation des déploiements de systèmes d'IA, les responsables de la sécurité doivent tirer les leçons de tels incidents et se préparer dès aujourd'hui à sécuriser les environnements de travail agentiques de demain.

Ce livre blanc :

- **S'intéresse** à la façon dont les espaces de travail numériques se transforment rapidement en environnements de travail agentiques
- **Examine** les préoccupations en matière de sécurité associées à ces environnements de travail agentiques émergents
- **Identifie** les exigences clés et les solutions de sécurité pour sécuriser les humains, les assistants d'IA et les agents d'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.

Les nouveaux environnements de travail agentiques

L'ère de l'IA est bel et bien là. Étant donné que des entreprises de tous les secteurs cherchent à transformer leurs processus métier, l'adoption des assistants d'IA s'est accélérée. Les assistants incluent des outils d'IA générative tels que ChatGPT et Gemini, des copilotes d'entreprise comme Microsoft Copilot et d'autres applications d'IA tierces spécialisées. Selon le rapport *The State of AI* in 2025 de McKinsey, 88 % des entreprises utilisent régulièrement l'IA dans au moins une fonction métier[3].

Les déploiements d'agents d'IA semi-autonomes et autonomes font également un bond. D'après le même rapport de McKinsey, 62 % des entreprises testent ou ont déjà déployé des agents d'IA. Comme les capacités des agents continuent d'évoluer, ce nombre ne manquera pas de croître.

La vague de l'IA transforme rapidement les espaces de travail numériques en environnements agentiques plus complexes. Dans les environnements de travail agentiques, la collaboration ne se produit pas seulement entre des personnes, mais entre des personnes, des assistants et des agents d'IA. En plus d'accomplir leurs propres tâches, les humains bénéficient de l'aide d'assistants et dirigent et supervisent des agents. Dans les environnements de travail agentiques, l'IA ne se contente pas de connecter les membres du lieu de travail ; elle consomme, génère et interagit avec l'information à une vitesse et à une échelle sans précédent. Chaque collaboration, qu'elle soit entre humains, assistants ou agents, introduit de nouveaux risques pour les données.

62%

des entreprises testent ou ont déjà déployé des agents d'IA.

Source : McKinsey

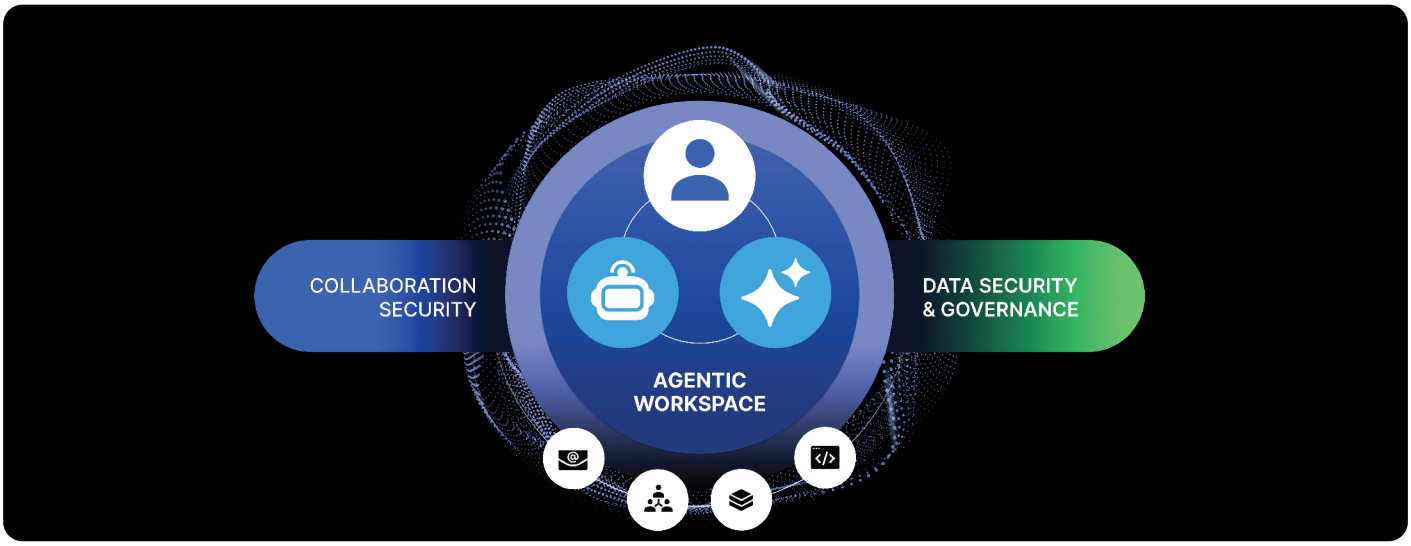


Figure 1 : Dans l'espace de travail agentique, humains, assistants IA et agents collaborent et interagissent avec des données sensibles via de multiples canaux.

3. McKinsey. *L'état de l'IA en 2025 : Agents, innovation et transformation*. Novembre 2025.

Les préoccupations en matière de sécurité associées aux environnements de travail agentiques

Les espaces de travail numériques reposaient sur la messagerie électronique, les applications SaaS (Software-as-a-Service), l'infrastructure cloud et les plates-formes de collaboration virtuelle. Ils offraient rapidité, échelle et flexibilité, mais révélaient également de nouvelles vulnérabilités. Les stratégies de sécurité ont dû évoluer car les cybercriminels ciblaient le comportement humain, les comptes et les applications pour accéder à la ressource la plus précieuse des entreprises : leurs données. La sécurité centrée sur les personnes — consistant à protéger les individus en tant que première ligne de défense — est devenue essentielle.

Les environnements de travail agentiques émergents amplifient ces défis. Les risques liés aux utilisateurs sont directement reflétés par les risques liés à l'IA. Les humains dotés d'assistants d'IA peuvent être victimes de tactiques d'ingénierie sociale, divulguer des identifiants de connexion, exécuter du code qu'ils ne devraient pas exécuter et manipuler des données de manière incorrecte. De même, les agents d'IA peuvent être trompés par des tactiques d'ingénierie d'invite, exécuter du code malveillant ou divulguer des informations sensibles.

Aidés par les outils d'IA, les cybercriminels peuvent agir plus rapidement et augmenter l'ampleur des attaques visant à la fois les humains et les agents.
Les cyberpirates peuvent également exploiter le protocole MCP (Model Context Protocol), un standard de communication open source couramment utilisé par les outils d'IA, pour compromettre les assistants et les agents d'IA. En déployant des serveurs MCP non approuvés, les cybercriminels peuvent exécuter des attaques man-in-the-middle qui demandent à des applications d'IA d'exécuter du code, d'exfiltrer des données sensibles ou d'effectuer d'autres actions non autorisées.

Selon le rapport Data Security Landscape 2025 de Proofpoint, 85 % des entreprises ont connu une fuite de données au cours de l'année écoulée. À mesure que les agents et assistants d'IA prolifèreront, gonfleront les effectifs des entreprises et élargiront la surface de risque, cette situation ne fera qu'empirer. Les stratégies de sécurité de la collaboration et des données qui ont sécurisé les espaces de travail numériques doivent s'étendre d'urgence à des environnements de travail agentiques regroupant humains, assistants et agents d'IA.

85%

des entreprises ont connu une fuite de données au cours des 12 derniers mois.
Source : Proofpoint

Dans les environnements de travail agentiques, les risques liés aux utilisateurs sont directement reflétés par les risques liés à l'IA.

Les exigences clés pour sécuriser les environnements de travail agentiques

Pour sécuriser la collaboration entre humains, assistants et agents, ainsi que sécuriser les données utilisées par ces intervenants, des solutions spécialisées conçues pour l'ère de l'IA sont nécessaires. Cependant, plusieurs exigences fondamentales doivent également être en place pour assurer l'efficacité de ces solutions.

Une plate-forme de cybersécurité unifiée

Avec une main-d'œuvre accrue d'humains et d'agents connectés et accélérés par l'IA, les environnements de travail agentiques représentent un changement radical de complexité pour la cybersécurité d'entreprise. Sachant que les humains utilisent de plus en plus d'assistants d'IA, les cyberpirates développent des techniques combinées qui ciblent les deux. Par exemple, une attaque peut commencer par la messagerie électronique, mais évoluer pour cibler également les outils d'IA. Et parce que les mêmes données sont maintenant accessibles et partagées par les humains, les assistants et les agents, la surface d'attaque de l'entreprise est plus étendue.

Les produits cloisonnés ne peuvent pas défendre efficacement ces environnements dynamiques en évolution rapide. Des patchworks inefficaces d'outils autonomes compliquent les opérations de sécurité, limitent la visibilité, laissent des failles critiques dans la protection contre les menaces et entravent la sécurité des données. Pour sécuriser les environnements de travail agentiques, les entreprises ont besoin d'une plate-forme de cybersécurité unifiée. Une plate-forme unifiée offre une protection multicouche contre les menaces pour les humains et les agents travaillant sur tous les canaux, y compris la messagerie électronique, les plates-formes de collaboration, les outils d'IA et les applications cloud. Elle offre une approche globale de la sécurité des données, que les données soient consultées par des humains, des assistants ou des agents. L'entreprise tout entière bénéficie ainsi d'une prévention unifiée des fuites de données, d'une détection cohérente et d'une vue unique des risques pour les données.

Intégration approfondie à desplates-formes partenaires

La sécurité centrée sur les personnes et les agents est le pilier central d'une architecture de cybersécurité plus large. Votre plate-forme de cybersécurité unifiée doit utiliser des connexions API et MCP pour s'intégrer aux plates-formes partenaires pour le XDR (eXtended Detection and Response), les opérations de sécurité (SecOps) et l'automatisation, le SASE (Secure Access Service Edge) et l'identité.

Des modèles de détection entraînés avec les meilleures données

Quand les tactiques des cybercriminels et les menaces internes évoluent à la vitesse d'une machine, vos solutions de sécurité doivent suivre le rythme. Pour sécuriser les environnements de travail agentiques, votre plate-forme de cybersécurité doit utiliser l'IA pour détecter les menaces avancées ainsi que pour comprendre les contenus et les comportements afin d'identifier les anomalies de sécurité des données. Les modèles d'IA intégrés doivent analyser les signaux de risque au niveau de la messagerie électronique, des applications cloud, des outils de collaboration et des navigateurs. Et parce que les menaces ne cessent d'évoluer, les modèles doivent apprendre de la threat intelligence en temps réel. Il faut donc les entraîner avec de grands ensembles de données compilés à partir de la surveillance de millions d'utilisateurs, de l'analyse de milliards d'incidents de sécurité des données et de l'examen de billions d'emails, de messages, d'URL et de pièces jointes.

Il est important de noter que lorsqu'ils sont entraînés avec des ensembles riches de threat intelligence, les modèles de détection apprennent à reconnaître l'intention en plus du contenu et du contexte. Ils peuvent par exemple comprendre quand le corps d'un email contient du contenu caché conçu pour déclencher un assistant tel que Microsoft Copilot afin qu'il effectue des actions particulières basées sur des invites. Comprendre l'intention

permet également aux modèles de détecter les invites malveillantes adressées directement aux assistants d'IA. Cela inclut les demandes de collaborateurs concernant des informations confidentielles ou précieuses.

Des points de contrôle pour les endroits où travaillent des personnes et des agents

Dans les grandes entreprises, les équipes SecOps surchargées n'ont ni le temps ni les ressources nécessaires pour aider en permanence les personnes et les agents à respecter les règles de sécurité. Pour vous aider, votre plate-forme de cybersécurité doit disposer d'une couche dédiée à l'application des règles et à l'accompagnement des utilisateurs, c'est-à-dire d'un ensemble de points de contrôle qui transforment la threat intelligence en protection en temps réel et en accompagnement aligné sur les règles. Ces points de contrôle doivent être mis en place à tous les endroits où travaillent des personnes et des agents — messagerie électronique, applications cloud, outils d'IA générative et navigateurs — afin de les aider à prendre des décisions plus sûres sans les ralentir.

Les agents comme multiplicateurs de force des opérations de sécurité

Les équipes SecOps sont constamment sous pression : plus d'alertes, plus d'outils et une capacité limitée. Les environnements de travail agentiques entraînant de nouveaux risques, les professionnels de la sécurité doivent eux-mêmes utiliser des agents d'IA pour gérer la charge de travail croissante. Alimentés par la threat intelligence en temps réel, les agents de sécurité intégrés à votre plate-forme de cybersécurité peuvent être des multiplicateurs de force pour les équipes de sécurité. Les agents peuvent agir comme des collaborateurs de confiance qui gèrent et accélèrent les tâches de routine. Il s'agit notamment de trier les incidents de prévention des fuites de données (DLP), d'analyser les emails signalés par les utilisateurs et de sensibiliser à la sécurité. Les analystes humains gardent le contrôle en approuvant les actions et en affinant les modèles, mais bénéficient de gains de productivité exponentiels.

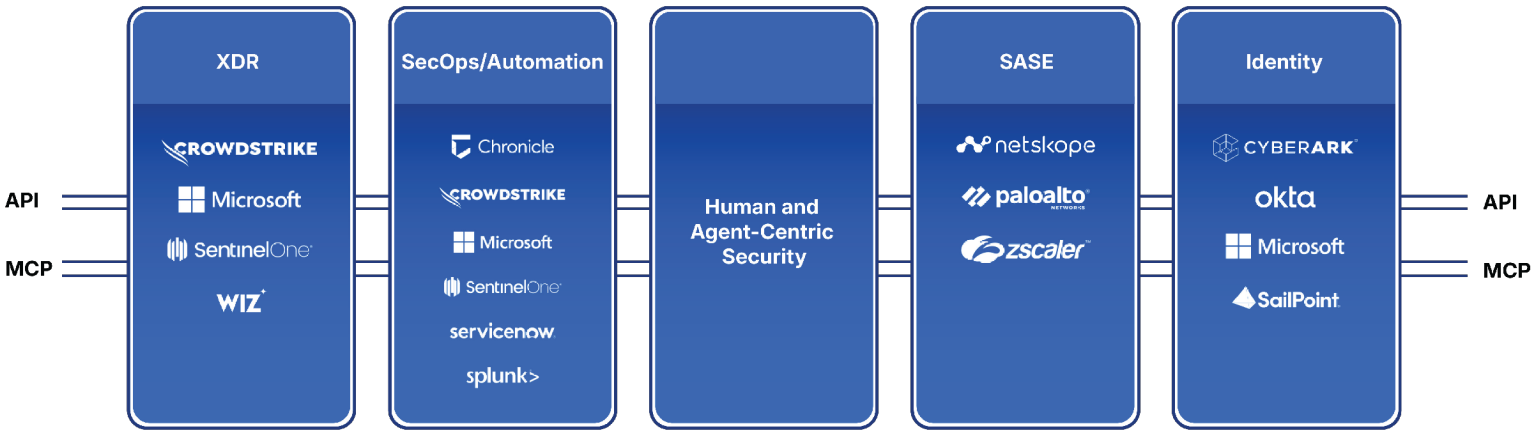


Figure 2: Votre plate-forme de cybersécurité centrée sur les personnes et les agents doit être le pilier central de votre architecture de cybersécurité, utilisant des connexions API et MCP pour s'intégrer aux plates-formes partenaires pour le XDR, les SecOps, l'automatisation, le SASE et l'identité.

Des solutions pour sécuriser les environnements de travail agentiques

Motivées par l'innovation et les gains de productivité, les entreprises de tous les secteurs déploient rapidement des outils d'IA. Cependant, plus les entreprises utilisent ces outils, plus leur surface de risque augmente. À mesure que les espaces de travail numériques se transforment en environnements de travail agentiques, les entreprises doivent sécuriser la collaboration humaine et agentique ainsi que les données que ces collaborateurs utilisent. Le cheminement vers une transformation métier axée sur l'IA est également un processus de mise en œuvre de fonctionnalités multicouches de collaboration et de sécurité des données en vue de protéger les humains, les assistants et les agents d'IA. La figure illustre ce parcours progressif de cybersécurité.

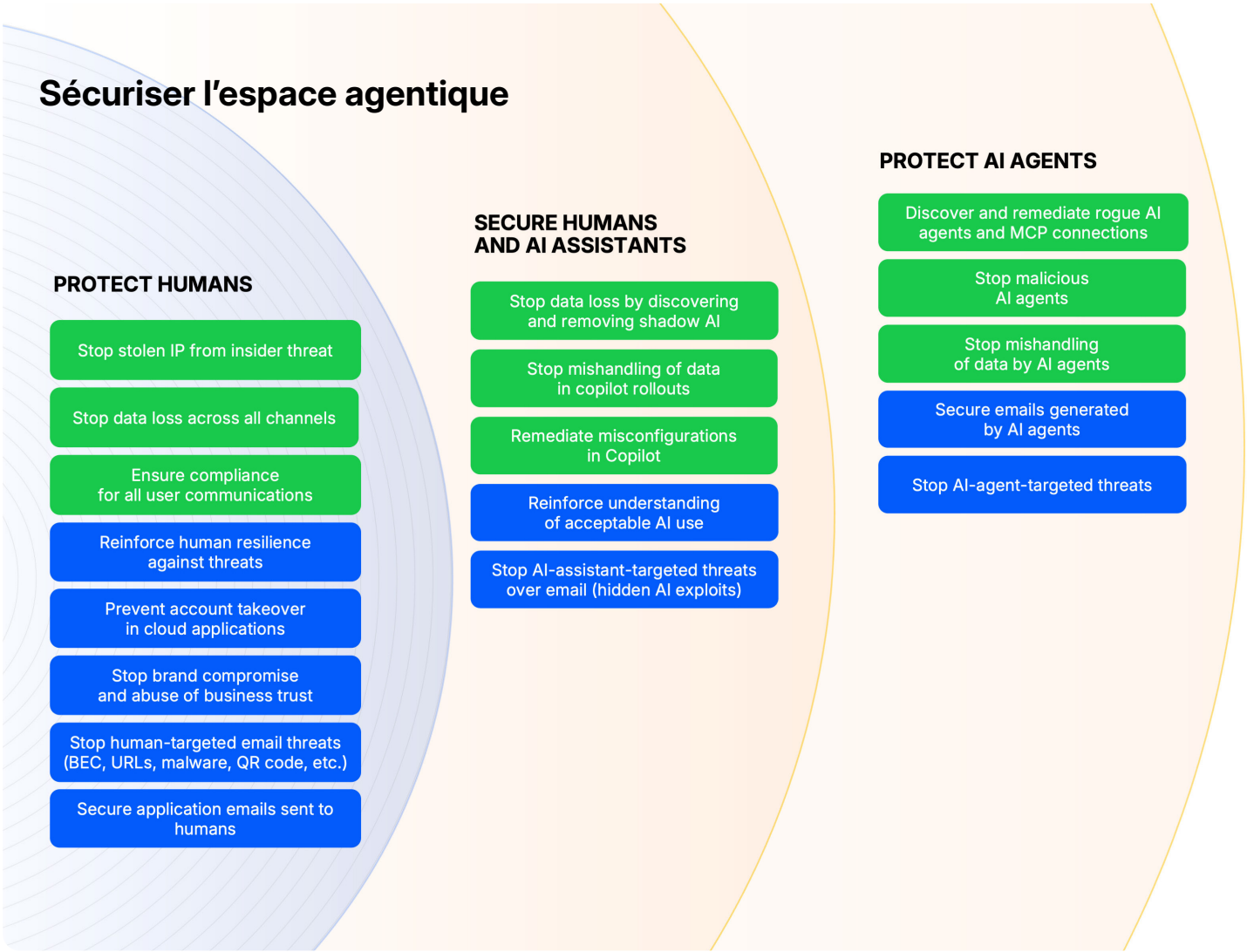


Figure 3 : Sécuriser l'espace de travail agentique consiste à déployer des mesures de collaboration et de protection des données multicouches afin de garantir la sécurité des humains, des assistants IA et des agents intelligents.

Sécurisation de la collaboration

Dans les environnements de travail agentiques, les agents d'IA sont de plus en plus intégrés aux processus. Ils automatisent les tâches, analysent les informations et collaborent avec les personnes et entre eux. Les agents sont conçus pour se comporter comme des humains : ils cliquent, partagent et agissent. Cela signifie qu'ils peuvent également être trompés, induits en erreur ou compromis. Les humains utilisant des assistants et des agents d'IA sont confrontés à des risques similaires, des attaques d'ingénierie sociale et d'ingénierie d'invite à la divulgation non autorisée de données sensibles ou d'identifiants de connexion. Les environnements de travail agentiques nécessitent une solution complète de protection contre les menaces qui permet aux humains et à leurs collègues numériques de collaborer en toute sécurité sur plusieurs canaux.

Figure 4: Une solution complète de sécurité de la collaboration protège la messagerie électronique, défend votre entreprise contre les attaques multicanales en plusieurs phases, renforce la résilience humaine et sécurise les communications métier.

Pour sécuriser la collaboration dans les environnements de travail agentiques, votre solution doit être dotée des capacités suivantes :

- **Blocage des menaces par email ciblant les personnes (BEC, URL, malwares, codes QR, etc.) –** Détectez et bloquez l'ensemble sans cesse croissant de menaces par email ciblant votre entreprise.
- **Renforcement de la résilience humaine face aux menaces –** Guidez de manière proactive les utilisateurs afin qu'ils effectuent des actions plus sûres et deviennent plus résilients face aux menaces**.**
- **Prévention de la prise de contrôle de comptes dans les applications cloud —** Détectez et corrigez les comptes cloud compromis, annulez les modifications malveillantes et supprimez l'accès persistant des cybercriminels.
- **Blocage des compromissions de marque et des abus de confiance –** Protégez vos communications avec vos partenaires, clients et fournisseurs de confiance contre des menaces telles que l'usurpation de domaines, les domaines similaires et les comptes fournisseurs compromis.
- **Sécurisation des emails envoyés par des applications à des humains –** Atténuez les risques d'usurpation d'identité en authentifiant l'identité des applications dans les communications par email envoyées à des humains.
- **Blocage des menaces ciblant les assistants d'IA par email —** Détectez les invites cachées dans les emails et bloquez les exploits d'IA avant la remise, avant qu'ils n'atteignent votre environnement. Prévenez les actions malveillantes non autorisées dans les assistants d'IA.
- **Renforcement de la compréhension de l'utilisation acceptable de l'IA -** Renforcez les règles d'utilisation acceptable de l'IA de votre entreprise avec des modules de sensibilisation qui apprennent aux collaborateurs à utiliser l'IA en toute sécurité.
- **Sécurisation des emails générés par des agents d'IA – A**tténuez les risques d'usurpation d'identité en authentifiant l'identité des agents d'IA dans les communications par email.
- **Blocage des menaces ciblant les agents d'IA –** Bloquez les menaces axées sur les agents, telles que les exploits d'injection d'invites, avant qu'elles n'atteignent les utilisateurs, garantissant que les personnes et les agents d'IA peuvent avoir confiance dans leurs interactions.

Sécurisation des données et gouvernance des communications

La sécurisation des données utilisées par les humains, les assistants et les agents d'IA requiert une solution unifiée qui élimine les zones d'ombre et les inefficacités des solutions cloisonnées. Une solution de sécurité des données unifiée doit offrir une visibilité et un contrôle complets sur la configuration, le niveau d'accès et les risques d'exfiltration pour chaque donnée structurée ou non structurée dans l'entreprise. Un composant supplémentaire de gouvernance et d'archivage des communications numériques peut assurer la conformité des communications des utilisateurs sur plusieurs canaux numériques.

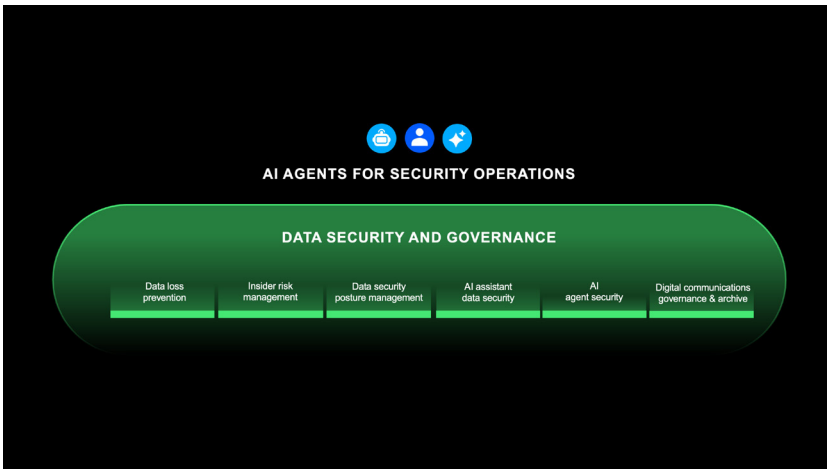


Figure 5: Une solution complète de sécurité et de gouvernance des données offre une visibilité et un contrôle complets sur tous les canaux, tout en garantissant la conformité des communications des utilisateurs.

Pour sécuriser les données dans les environnements de travail agentiques, une solution complète de sécurité des données et de gouvernance des communications doit s'appuyer sur les capacités suivantes :

- **Blocage des fuites de données sur tous les canaux –** Prévenez les fuites de données sur tous les canaux sur lesquels les personnes et les agents travaillent, y compris la messagerie électronique, les applications cloud, les plates-formes de collaboration, les outils d'IA générative et les navigateurs.
- **Prévention des vols de propriété intellectuelle d'origine interne –** Obtenez une visibilité sur les comportements à risque qui conduisent à des fuites de propriété intellectuelle et de données sensibles dues à des utilisateurs négligents, malveillants et compromis.
- **Conformité garantie de toutes les communications des utilisateurs –** Unifiez, gérez, stockez et enquêtez sur les communications des utilisateurs sur les canaux numériques tels que les plates-formes de collaboration, la messagerie électronique, les SMS, les réseaux sociaux, la voix et la vidéo.
- **Correction des erreurs de configuration dans Copilot –** Identifiez et corrigez les erreurs de configuration dans vos environnements Microsoft 365 et SharePoint pour garantir un accès sécurisé par Microsoft Copilot.

- **Correction de la mauvaise gestion des données dans les déploiements de Copilot —** Découvrez et classez toutes les données structurées et non structurées dans les environnements hybrides et multi-cloud. Appliquez des étiquettes de protection des informations pour protéger les données auxquelles accèdent les copilotes d'entreprise.
- **Blocage des fuites de données grâce à la découverte et à la suppression des applications d'IA non approuvées –** Détectez l'utilisation d'outils d'IA non approuvés et appliquez des règles pour bloquer leur utilisation. Empêchez les outils non approuvés d'accéder aux données sensibles et de les divulguer.
- **Découverte et neutralisation des agents d'IA et des connexions MCP non approuvées –** Utilisez un outil de sécurité dédié aux agents d'IA basé sur le protocole MCP pour surveiller et contrôler l'activité des agents et appliquer des règles en matière de données.
- **Blocage des agents d'IA malveillants –** Utilisez un outil de sécurité dédié aux agents d'IA pour détecter et bloquer les attaques d'agents malveillants.
- **Correction de la mauvaise gestion des données par les agents d'IA –** Utilisez un outil de sécurité dédié aux agents d'IA afin de contrôler l'accès aux données sensibles utilisées par les agents et de masquer les données sensibles avant qu'elles n'atteignent des humains ou d'autres agents.

Comment Proofpoint peut vous aider

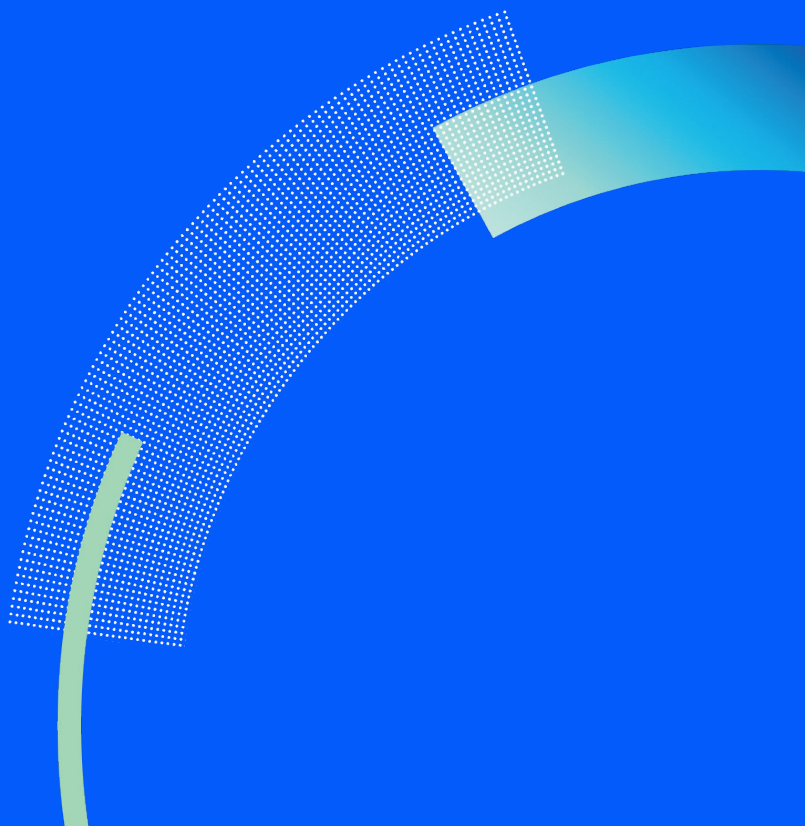
Grâce aux technologies Proofpoint Nexus® et Proofpoint Zen™, associées à la puissance des agents IA Proofpoint Satori™, la plateforme de sécurité Proofpoint, pensée autour de l'humain et de l'agent, offre une protection globale, spécialement conçue pour l'ère agentique.

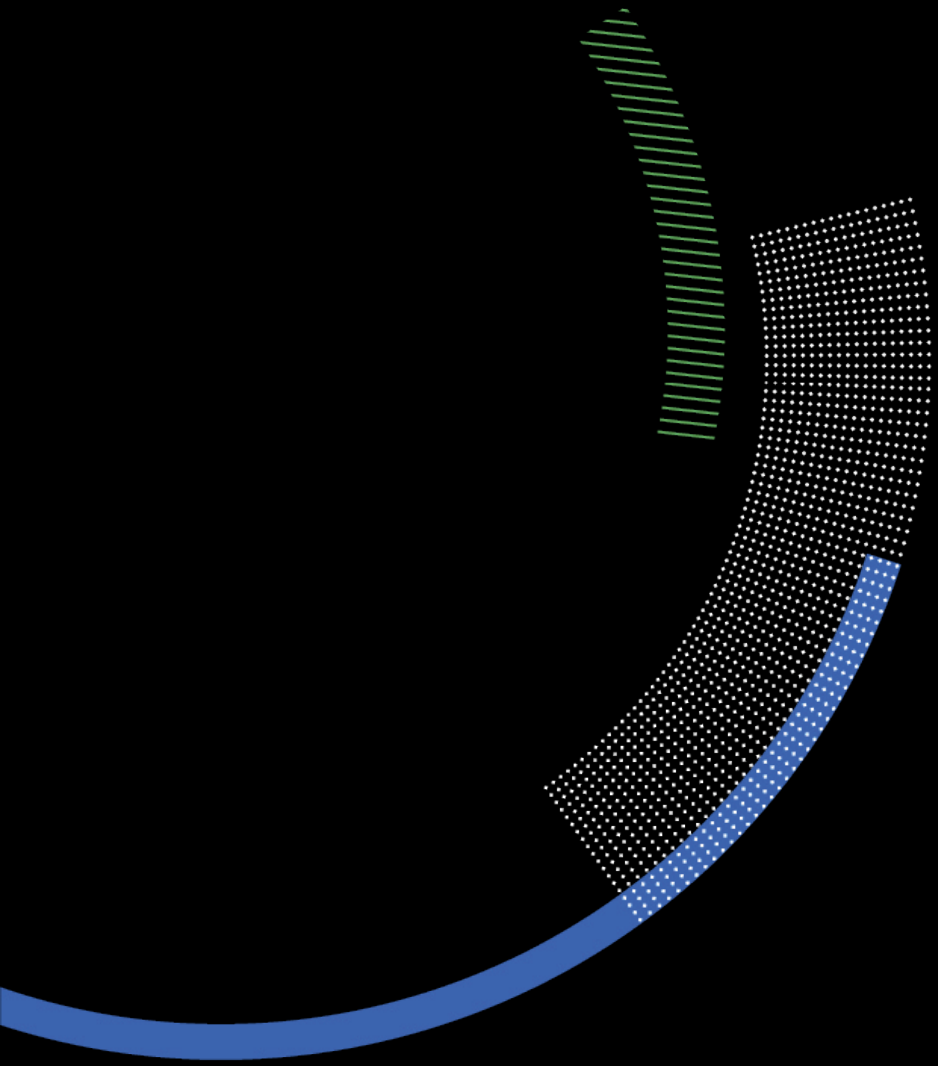
Notre solution de sécurité collaborative unifiée permet de neutraliser les menaces ciblées et garantit des échanges de confiance, qu'ils soient entre humains, agents, ou entre humains et agents.

Parallèlement, notre solution de sécurité des données centralisée assure une visibilité et un contrôle total sur la configuration, les accès et les risques d'exfiltration de chaque donnée de l'entreprise, structurée ou non—qu'elle soit manipulée par un humain, un assistant IA ou un agent.

Étapes suivantes

- Pour découvrir la plateforme Proofpoint en conditions réelles, contactez-nous et planifiez une démonstration gratuite.
- Pour en savoir plus sur la manière dont Proofpoint protège l'espace agentique, retrouvez-nous lors de l'un de nos événements Protect Series.





proofpoint®

À propos de Proofpoint, Inc. Proofpoint, Inc. est un acteur mondial de référence en cybersécurité centrée sur l'humain et les agents, protégeant les échanges entre les personnes, les données et les agents d'IA via l'e-mail, le cloud et les outils collaboratifs. Proofpoint accompagne en toute confiance plus de 80 entreprises du Fortune 100, plus de 10 000 grandes entreprises et des millions de petites organisations pour contrer les menaces, éviter la perte de données et renforcer la résilience des équipes humaines et des flux de travail pilotés par l'IA. Sa plateforme de sécurité collaborative et de protection des données aide les organisations de toutes tailles à protéger et valoriser leurs collaborateurs, tout en adoptant l'IA avec sécurité et assurance. En savoir plus sur www.proofpoint.com.

Restez en contact avec Proofpoint : LinkedIn