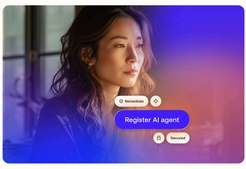
Plan d’action pour l’entreprise agentique sécurisée

Trois questions auxquelles tout responsable IT et sécurité doit pouvoir répondre avant que la multiplication des agents n’échappe à tout contrôle.



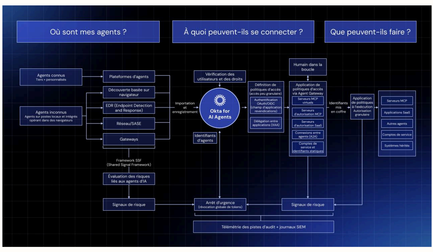
L’identité au cœur de la sécurité de l’IA

Par le passé, un logiciel exécutait docilement les commandes de l’utilisateur. À présent, il décide, agit et se connecte de sa propre initiative. Déjà, les agents d’IA aident les collaborateurs dans leurs tâches, servent les clients et remplissent diverses fonctions au sein des chaînes logistiques. Et ils opèrent à une vitesse bien plus rapide que le dispositif de sécurité qui leur est appliqué. Au cours de la dernière décennie, les entreprises ont renforcé la sécurité des identités humaines par la mise en coffre (vaulting), l’application du principe du moindre privilège et une authentification continue. Malgré toutes ces mesures, l’essor fulgurant des agents d’IA crée de nouvelles failles dans l’infrastructure d’identités. N’importe qui peut lancer un agent, un agent peut en créer d’autres, et chacun d’entre eux se connecte aux applications, aux API, aux outils SaaS et aux systèmes de données. Résultat : des milliers de nouvelles entités disposant d’un accès à privilèges et fonctionnant à la vitesse des machines — souvent en contournant les contrôles de sécurité existants.

Voilà pourquoi les agents doivent être considérés comme des identités à part entière. Pour que l’entreprise agentique soit sécurisée, il faut commencer par établir des responsabilités claires et bénéficier d’une visibilité complète avant que les agents ne deviennent incontrôlables. Les entreprises doivent être en mesure de répondre à trois questions :

1. Où sont mes agents ?
2. À quoi peuvent-ils se connecter ?
3. Que peuvent-ils faire ?

Ces questions définissent le plan d’action destiné à sécuriser les agents d’IA. Répondre à ces questions ne se limite pas à prendre acte de la situation : il faut disposer de systèmes, de contrôles d’identité et d’un modèle de gouvernance permettant de gérer l’entreprise agentique de façon sécurisée.



Question 1 : Où sont mes agents ?

La visibilité est la principale préoccupation qu’expriment les clients avec lesquels nous échangeons.

Demandez à votre équipe combien d’agents compte votre environnement : la plupart ne pourront pas vous fournir un chiffre précis. Les agents lancés par les collaborateurs à partir d’un navigateur ou exécutés silencieusement sur les ordinateurs de bureau sont souvent inconnus et échappent à tout contrôle.

Vous devez être en mesure de découvrir les agents, quel que soit l’emplacement où ils ont été créés ou déployés.

- **Intégration des plateformes agentiques :** enregistrez les agents des principales plateformes tierces ainsi que vos propres agents personnalisés auprès de votre fournisseur d’identité. Si vous ne pouvez pas les voir lors de leur création, vous ne pourrez pas en assurer la gouvernance.
- **Détection basée sur le navigateur :** découvrez les agents non validés exécutés via des navigateurs et des extensions qui ne sont pas hébergés par votre fournisseur d’identité. Il s’agit des agents que les collaborateurs initialisent sans demander d’autorisation.
- **Détection des terminaux :** identifiez les agents exécutés sur les terminaux gérés. Cette fonction doit être intégrée à la solution que vous utilisez pour gérer ou protéger les postes de travail et les terminaux mobiles.
- **Détection du réseau :** repérez le trafic non autorisé entre les agents, ou entre les agents et les ressources, au niveau de la couche réseau. Les agents communiquent entre eux et avec les services, et vous devez être en mesure de voir ces connexions.
- **Détection des gateways :** identifiez et gérez les agents d’IA et les clients OAuth non enregistrés qui interagissent avec vos gateways MCP, d’agents et d’API. Si des agents appellent vos API, ils doivent être authentifiés et consignés dans un journal.
- **Évaluation des risques posés par les agents d’IA :** surveillez et identifiez en permanence les erreurs de configuration laissant les agents d’IA vulnérables à l’exploitation. L’analyse de la posture de sécurité de chaque identité d’agent permet d’identifier les risques de manière proactive.

Quelle que soit votre pile, l’ingestion des signaux provenant de toutes ces sources est indispensable. Votre couche de découverte doit être compatible avec de multiples plateformes, outils et équipes. Une visibilité lacunaire n’est jamais acceptable.

Question 2 : À quoi peuvent-ils se connecter ?

Une fois un agent identifié, vous devez déterminer toutes les ressources auxquelles il peut accéder et lui appliquer des politiques d’accès. Sans contrôle centralisé des parcours de connexion, la compromission d’un seul agent peut avoir un effet domino et permettre l’accès à tout l’environnement à la vitesse des machines.

L’impact d’un agent compromis se mesure par l’étendue de ses connexions.

- **Serveurs et ressources MCP :** les serveurs MCP (Model Context Protocol) permettent aux agents d’accéder à des outils et à des sources de données à la fois internes (applications, API, bases de données, propriété intellectuelle) et
externes (clients MCP tiers tels que Slack, GitHub et Notion). Votre périmètre de sécurité s’étend désormais à toutes les ressources auxquelles votre agent peut accéder.
- **Applications SaaS :** les agents se connectent aux mêmes outils SaaS que ceux utilisés au quotidien par vos collaborateurs. La différence est que les agents exécutent leurs tâches beaucoup plus vite et accèdent à davantage de données. Un agent compromis peut exfiltrer des données ou apporter des modifications à toutes les applications SaaS connectées plus rapidement qu’un être humain ne pourrait le faire.
- **Connexions entre agents :** sécurisez l’établissement de liaisons et l’autorisation entre entités autonomes. L’appel d’un agent par un autre agent peut constituer le point de départ d’un déplacement latéral. Les deux parties doivent vérifier leur identité respective avant d’échanger des données ou de déléguer des tâches.
- **Comptes de service :** limitez la prolifération d’identifiants statiques de longue durée. Ils représentent des passe-partout que les agents héritent des anciens modèles machine-to-machine. Chaque identifiant statique est une porte dérobée persistante qui attend d’être exploitée.
- **Identifiants mis en coffre :** protégez et renouvelez automatiquement les secrets. Un token non renouvelé ouvre la porte aux attaquants. Les identifiants doivent être mis en coffre (vaulting), émis de manière dynamique et faire l’objet d’une rotation fréquente. Aucun agent ne devrait fonctionner avec des identifiants qui persistent au-delà de sa mission.

Toutes ces connexions doivent être consignées dans votre SIEM. Vous ne pouvez pas sécuriser ce que vous ne voyez pas. Chaque connexion d’agent, y compris les ressources auxquelles il accède, le moment de la connexion et les identifiants utilisés, doit être transmise à votre centre SOC à des fins de monitoring et d’examen.

Question 3 : Que peuvent-ils faire ?

Il ne suffit pas de savoir où résident les agents et à quoi ils peuvent se connecter si vous ne pouvez pas contrôler et interrompre leurs actions le cas échéant. Lorsqu’un agent commence à exfiltrer des données ou à exécuter des processus non autorisés, la réponse doit être immédiate.

- **Arrêt d’urgence :** si un agent s’écarte de la mission prévue, accède à des données sensibles de manière inattendue ou si une menace est détectée, vous devez être en mesure de révoquer instantanément l’accès dans tous les systèmes afin de limiter les risques.
- **Application de politiques à l’exécution :** attribuez des autorisations aux agents en temps réel en fonction des actions qu’ils tentent d’exécuter. Évaluez le contexte, la séquence et le volume. Une requête portant sur 10 enregistrements clients se distingue d’une requête en demandant 10 000. Détectez les attaques par injection d’invites et appliquez des politiques au niveau des outils avant que des actions ne soient exécutées.
- **Gestion du cycle de vie des agents :** les autorisations légitimement accordées à un agent à un instant T ont rarement du sens au bout de trois mois. Évaluez continuellement l’accès afin d’appliquer le principe du moindre privilège, d’automatiser les certifications et de révoquer immédiatement l’accès lorsque des agents sont mis hors service ou que des collaborateurs quittent l’entreprise.
- **Approbations « humain dans la boucle » :** exigez une approbation humaine pour les actions sensibles ou potentiellement dangereuses des agents. Bloquez les opérations destructrices, l’accès en masse aux données ou l’élévation des privilèges des agents.
- **Journaux d’audit et télémétrie :** chaque action d’agent doit être consignée et transmise à votre SIEM, à savoir chaque appel d’outil, chaque décision d’autorisation et chaque tentative d’accès. L’application de politiques à l’exécution et les mécanismes d’« arrêt d’urgence » ne sont efficaces que si vous disposez d’une visibilité totale.

Le plan d’action comme point de référence

Les trois questions (*où sont mes agents, à quoi peuvent-ils se connecter, que peuvent-ils faire*) ne sont pas un objectif à atteindre à plus ou moins long terme : il s’agit des critères à respecter a minima pour déployer des agents d’IA en production.

Les entreprises qui ne peuvent pas y répondre manquent cruellement de visibilité. Lors des échanges qui auront lieu avec le conseil d’administration, les auditeurs ou les autorités en cas de brèches, « je ne sais pas » ne constituera pas une réponse valable.

Les entreprises proactives l’ont déjà compris. Elles n’attendent pas le premier incident pour prendre des mesures. Elles traitent déjà les agents comme des identités de premier ordre. Elles intègrent la découverte, l’application de politiques et la gouvernance dans leurs déploiements d’agents dès le premier jour. Elles peuvent répondre aux trois questions avant de perdre le contrôle de leurs agents.

Le plan d’action n’est pas un audit ponctuel. Vous devez pouvoir répondre à ces trois questions à tout moment et en toutes circonstances pour faire face à la multiplication exponentielle des agents. Vous avez passé les dix dernières années à concevoir une infrastructure de sécurité des identités humaines. Ne laissez pas les agents réduire vos efforts à néant.

Okta a élaboré ce plan d’action en s’appuyant sur son expérience et ses interactions constantes avec des entreprises de premier plan qui sécurisent des agents d’IA à grande échelle. Découvrez comment Okta Platform l’implémente à la page okta.com/ai-agents.