Sécurité de l'IA : votre avantage concurrentiel dans la course à l'IA

Guide à l'intention des dirigeants pour accélérer l'adoption de l'IA, tout en maîtrisant les risques

Introduction

À l'heure où les entreprises s'empressent de maîtriser la puissance transformatrice de l'IA, l'adoption prend souvent la sécurité de vitesse. Pour chaque initiative autorisée, il existe d'innombrables occurrences d'IA fantôme imputables à des individus comme à des équipes. Concrètement, 85 % des décideurs informatiques affirment que leurs collaborateurs adoptent les outils IA avant qu'ils n'aient été évalués par les équipes responsables des technologies de l'information.[1] Par ailleurs, 93 % des collaborateurs admettent avoir saisi des informations dans des outils IA sans en avoir reçu l'autorisation.[1]

Pendant ce temps, les vecteurs d'attaque AI-native connaissent une forte augmentation, avec une hausse de 47 % au cours de l'année passée,[2] tandis que les outils de sécurité traditionnels peinent à suivre le rythme. Les nouvelles formes d'exposition des données et les lacunes en matière de conformité mettent à l'épreuve les pratiques de gouvernance existantes. Les dirigeants sont confrontés à une question essentielle : comment maîtriser les risques qu'entraîne l'IA sans paralyser l'innovation qu'elle permet ?

Les exigences sont claires. Les entreprises doivent créer un environnement dans lequel :

- Toute utilisation de l'IA est connue
- Toute communication avec l'IA peut être sécurisée
- Toute politique relative à l'IA peut être appliquée
- Tous les modèles IA peuvent être protégés contre les utilisations abusives

Cloudflare AI Security Suite fournit aux entreprises la confiance indispensable pour progresser plus rapidement avec l'IA en éliminant l'incertitude liée aux risques. Notre plateforme unifiée sécurise l'ensemble du cycle de vie de l'IA en détectant son utilisation, en appliquant une politique d'accès Zero Trust et en déployant, sur les points de terminaison web et d'API, une défense proactive contre les menaces. Avec une gouvernance intégrée des données, les équipes sont libres d'innover sans compromettre la sécurité.

85 %
des décideurs informatiques affirment que leurs collaborateurs adoptent les outils IA avant qu'ils n'aient été évalués par les équipes responsables des technologies de l'information.[1]

Le défi de la sécurité de l'IA

Les risques liés à l'adoption de l'IA alimentent un marché dynamique de solutions de sécurité de l'IA. Par exemple, les plateformes de protection des applications cloud-native (CNAPP, « Cloud-Native Application Protection Platform ») priorisent les flux de travail de développement de l'IA. Les premières offres de ce type constituent une ligne de défense essentielle, mais demeurent insuffisantes à elles seules. Les entreprises doivent sécuriser l'intégralité du cycle de vie de l'IA, notamment en protégeant les systèmes IA lorsqu'ils sont déployés en production.

L'ampleur du défi que constitue la sécurité de l'IA devient de plus en plus évidente. Aujourd'hui, les développeurs créent des fonctionnalités IA, le personnel utilise des outils IA externes et les clients interagissent avec des applications pilotées par IA. Sécuriser manuellement ces différents environnements est complexe, et accomplir cette tâche de manière cohérente l'est encore plus.

Prévenir l'utilisation abusive de l'IA par le personnel

À mesure que le personnel de l'entreprise adopte des outils IA approuvés et non approuvés, les données et les opérations sensibles sont exposées à des risques. Les équipes de sécurité doivent relever les défis suivants :

- **Visibilité insuffisante de l'utilisation de l'IA :** les dirigeants ne disposent souvent pas d'une visibilité exhaustive des outils IA utilisés par leurs collaborateurs, de l'endroit où sont traitées les données sensibles et de la manière dont ces systèmes IA se connectent aux applications de l'entreprise. Cet angle mort entraîne une exposition considérable à différents risques.
- **Risques liés à la sécurité des données et à la conformité :** l'IA transforme la manière dont les données circulent au sein de votre entreprise. Les informations personnelles, les données propriétaires et les dossiers de clients peuvent être introduits dans les systèmes IA, au risque d'entraîner des violations de la conformité ou d'exposer des informations décisionnelles stratégiques.
- **Contrôles des accès pour les flux de travail d'IA agentique :** l'accès des utilisateurs humains aux outils IA ne constitue pas l'unique paramètre devant être géré ; l'accès des IA agentiques aux serveurs MCP et aux autres systèmes critiques doit également être pris en compte. Or, cette condition nécessite de nouvelles approches de la gestion des identités et du contrôle des accès.

Protection des applications et modèles IA accessibles au public

Qu'ils soient développés en interne ou exploités par un tiers, les systèmes IA deviennent un pilier essentiel de l'expérience des clients et des utilisateurs et doivent être protégés en conséquence. Le classement OWASP Top 10 for LLM Applications met en évidence les menaces auxquelles les outils traditionnels ne permettent pas de répondre :

- **Attaques par surconsommation de ressources :** semblables aux attaques DoS traditionnelles, les attaques par surconsommation de ressources visent à submerger les LLM de requêtes consommant d'importantes quantités de ressources. Les modèles de tarification à l'utilisation de services cloud peuvent faire exploser l'impact financier de telles attaques, tandis que les utilisateurs légitimes se heurtent à une dégradation de la qualité de service.
- **Empoisonnement de modèles :** des acteurs malveillants implantent des portes dérobées, des biais ou des vulnérabilités dans les LLM en injectant des données corrompues dans des ensembles de données publics ou des référentiels utilisés par les développeurs pour l'apprentissage des modèles. Un modèle empoisonné de cette manière se comporte normalement jusqu'à ce que des déclencheurs spécifiques activent des comportements préjudiciables.
- **Attaques par injection de prompt :** tactique populaire d'exfiltration de données via les interfaces d'IA, l'injection de prompt permet de manipuler les entrées des LLM en intégrant des instructions malveillantes dans les prompts (ou invites) d'utilisateurs ou dans les contenus externes. Lors de ces attaques, les modèles ignorent les instructions originales et exécutent les commandes de l'acteur malveillant à la place.
- **Débridage :** des prompts soigneusement élaborés, tels que des scénarios de jeu de rôle, des commandes de substitution d'instructions et des stratégies conversationnelles à plusieurs tours peuvent outrepasser les systèmes de sécurité des LLM afin de générer des contenus interdits ou d'extraire des informations sensibles.

Afin d'encourager une innovation rapide en matière d'IA sans pour autant compromettre la sécurité de l'entreprise, nous recommandons aux dirigeants d'adopter une approche de la sécurité de l'IA répondant à l'ensemble des exigences suivantes :

- Protection de l'utilisation de l'IA générative par le personnel
- Protection des applications et des charges de travail pilotées par IA
- Développement d'applications pilotées par IA intrinsèquement sécurisées

Sécuriser les flux de travail de développement et d'apprentissage de l'IA

Les projets d'IA reposent sur d'immenses ensembles de données, des ressources coûteuses et des expérimentations itératives – autant d'éléments qui engendrent de nouvelles vulnérabilités et surfaces d'attaque. Les équipes responsables de la sécurité doivent gérer les problématiques suivantes :

- **Sécurité et intégrité des données d'apprentissage :** la compromission des données d'apprentissage peut créer des biais, des portes dérobées ou des vulnérabilités qui persistent dans les modèles de production. Les entreprises doivent sécuriser l'accès aux ensembles de données d'apprentissage, empêcher les modifications non autorisées et garantir la provenance des données sur l'ensemble du cycle de vie du modèle.
- **Gestion des informations d'identification et des secrets :** les flux de travail de développement de l'IA nécessitent un accès à différents systèmes, notamment un stockage dans le cloud pour les ensembles de données, des clusters de calcul pour l'apprentissage, des registres de modèles et des services tiers. Des secrets ou des clés d'API insuffisamment sécurisés peuvent révéler des informations d'identification sensibles, permettant ainsi l'accès non autorisé à des modèles propriétaires, des données d'apprentissage ou des systèmes de production.
- **Contrôle des accès à l'environnement de développement :** les ingénieurs spécialistes de l'IA ont souvent besoin de privilèges élevés pour tester des modèles et accéder à des données sensibles. En l'absence de contrôles d'accès appropriés, cette situation peut entraîner des menaces internes, l'exposition accidentelle de données ou l'extraction non autorisée de modèles.

Unifiez la sécurité avec Cloudflare

L'adoption réussie de l'IA a pour objectif d'améliorer la productivité, et non de la limiter. Lorsque les équipes peuvent utiliser l'IA avec confiance, avec la certitude que des protections adéquates ont été mises en place, elles peuvent innover plus rapidement et entreprendre des projets plus ambitieux.

Cloudflare AI Security Suite crée un environnement sécurisé pour l'innovation en matière d'IA. En unifiant des fonctionnalités SASE (Secure Access Service Edge) et de sécurité des applications web, les dirigeants peuvent connecter et protéger deux domaines fondamentaux :

- Les applications externes, publiques et pilotées par IA
- Les systèmes IA et charges de travail IA internes et privés

En priorisant l'ensemble du cycle de vie de l'IA, Cloudflare AI Security Suite répond aux besoins de sécurité, depuis la découverte et la gestion des risques jusqu'à la protection des données, afin de sécuriser l'accès des utilisateurs et de protéger les applications et les flux de travail de développement renforcés par IA. Pour offrir une couche essentielle de sécurité de la production en temps réel, le réseau mondial Cloudflare est déployé en ligne, afin d'inspecter et de filtrer chaque interaction avec l'IA. Il protège ainsi les données de l'ensemble des utilisateurs et applications.

Au lieu de simplement détecter les problèmes a posteriori, Cloudflare empêche les incidents de se produire et bloque les menaces avant qu'elles n'atteignent les modèles IA. Les équipes de sécurité acquièrent ainsi la visibilité indispensable pour garder une longueur d'avance sur les menaces émergentes.

Fonctionnalités essentielles de Cloudflare

Cloudflare AI Security Suite réunit une surveillance complète, une protection en temps réel et une gestion proactive des risques sur une plateforme unique, proposant une approche holistique de la sécurité de l'IA.

Découverte et visibilité exhaustives de l'IA

L'efficacité de la sécurité de l'IA dépend d'un inventaire complet, en temps réel, des ressources et de l'utilisation de l'IA, qu'elles soient autorisées ou non. La fondation que propose Cloudflare AI Security Suite est la surveillance continue et la découverte automatisée permettant d'identifier tous les modèles, assistants et agents IA, ainsi que les déploiements d'IA fantôme dans tous les types d'environnements : publics, privés ou internes.

Gestion proactive des risques liés à l'IA

Cloudflare AI Security Suite aide les entreprises à prévenir les attaques en détectant et en atténuant les vulnérabilités, les erreurs de configuration et les vecteurs d'attaque spécifiques à l'IA, notamment ceux énumérés dans le classement OWASP Top 10 for LLMs. Le score de confiance des applications permet de hiérarchiser les mesures correctives, afin que les équipes priorisent les risques les plus importants.

Sécurité des applications pour les applications pilotées par IA

Pour aider les équipes SecOps à se tenir informées des nouveaux vecteurs de menaces liés à l'IA, Cloudflare AI Security Suite inclut la détection et l'atténuation proactives des menaces pour les vulnérabilités spécifiques à l'IA, les erreurs de configuration et les chemins d'attaque dans les pipelines d'IA, notamment des protections contre l'injection de prompts, l'empoisonnement de données et l'utilisation abusive de modèles.

Les pare-feu IA spécialisés découvrent et identifient les points de terminaison d'IA agentique ou générative et les points de terminaison d'API, détectent les tentatives d'exfiltration d'informations d'identification personnelle et bloquent les prompts malveillants avant qu'ils n'altèrent les performances des modèles d'IA ou n'empoisonnent le modèle avec des contenus toxiques ou de la désinformation.

Accès Zero Trust pour les flux de travail d'IA générative et d'IA agentique

Les principes de la sécurité Zero Trust, tels que le moindre privilège, sont appliqués au personnel aussi bien qu'aux agents IA. Cloudflare AI Security Suite peut appliquer les politiques d'accès réseau Zero Trust (ZTNA, Zero Trust Network Access) pour les interactions humain-IA et IA-IA. La journalisation et les contrôles centralisés des serveurs MCP assurent que les IA agentiques accèdent uniquement aux ressources pour lesquelles elles sont autorisées, notamment les flux de travail à la demande.

Protection des données sensible à l'IA

La sécurisation efficace de l'IA nécessite une capacité de prévention des pertes de données qui exploite plusieurs modèles de langage pour comprendre le contenu d'un prompt, ainsi que l'intention sous-jacente. Cloudflare AI Security Suite intègre des fonctionnalités de prévention des pertes de données (DLP) couvrant l'ensemble du processus d'apprentissage, les prompts et les réponses, afin d'empêcher l'exposition des informations d'identification personnelle, les fuites de données et les accès non autorisés dans les modèles et les pipelines IA. La sécurité d'exécution déployée en ligne, priorisant les API, constitue une première couche de défense simple et rapide, qui complète l'approche « shift left » consistant à intégrer la sécurité le plus précocement possible dans le cycle de développement que permet une plateforme CNAPP.

Régionalisation des données

Les LLM et les applications IA sont soumis aux mêmes réglementations que tout autre type d'environnement de données. Cloudflare AI Security Suite aide les entreprises à assurer que les charges de travail IA respectent les frontières géographiques et juridictionnelles, grâce à des politiques permettant de conserver les données d'apprentissage et les requêtes d'inférence dans les régions approuvées.

Une sécurité intrinsèque pour le développement d'applications IA

À l'instar de tous les types de logiciels, la sécurité des applications IA doit être intégrée dès le début du cycle de développement, et non ajoutée a posteriori. Cloudflare AI Security Suite fournit aux développeurs les outils et infrastructures nécessaires pour créer des applications pilotées par IA dotées d'une sécurité intrinsèque.

L'impact opérationnel de Cloudflare AI Security Suite

L'essor de l'IA n'est pas seulement une évolution : c'est une révolution fondamentale dans l'ordre de l'industrialisation ou de l'informatisation. Par conséquent, l'adoption rapide et efficace devient un enjeu non seulement de dynamisation de la croissance, mais également de viabilité organisationnelle. L'adoption lente ou non sécurisée représente désormais une menace existentielle pour les entreprises de tous les secteurs et industries.

Cloudflare AI Security Suite peut encourager une transformation sécurisée, contrôlée et efficace, permettant de répondre aux attentes grandissantes des clients et aux exigences croissantes du marché dans des environnements hautement concurrentiels.

- **Accélération de l'innovation en matière d'IA :** lorsque la sécurité permet une utilisation sûre de l'IA, au lieu d'en grever l'adoption, les collaborateurs et les équipes peuvent explorer de nouvelles applications IA pour devenir plus productifs dans leur travail quotidien. Des cadres de sécurité appropriés donnent aux développeurs la confiance nécessaire pour créer des fonctionnalités IA ambitieuses, sans compromettre la sécurité des données ou des systèmes sensibles.
- **Réduction des risques liés à l'IA :** les entreprises peuvent gérer l'ensemble des risques résultant de l'adoption de l'IA, notamment les risques inhérents aux applications web pilotées par IA. Les applications IA en cours de développement peuvent être sécurisées afin d'éviter que les collaborateurs ne divulguent des données sensibles ou ne les exposent dans des ensembles de données d'apprentissage d'IA. Les équipes SecOps peuvent identifier et atténuer proactivement les menaces et les vulnérabilités spécifiques à l'IA, et ainsi, minimiser la surface d'attaque et protéger les données et modèles essentiels.
- **Opérations de sécurité rationalisées :** la visibilité et le contrôle centralisés de votre stratégie de sécurité de l'IA permettent de simplifier la gestion et de rationaliser la réponse aux incidents. Votre équipe SecOps peut prioriser les initiatives stratégiques au lieu de continuellement remédier aux incidents liés à l'IA.
- **Gouvernance et conformité robustes des données :** les contrôles de protection des données spécifiques à l'IA vous aident à sécuriser les informations sensibles et à satisfaire aux exigences réglementaires changeantes sur l'ensemble du cycle de vie de l'IA.
- **Réduction du coût total de possession :** il est plus économique d'utiliser une plateforme consolidée qui intègre les investissements de sécurité existants que de déployer des solutions dédiées distinctes pour chaque problématique liée à la sécurité de l'IA.

Scénarios d'utilisation de modèles pour Cloudflare AI Security Suite

Cloudflare AI Security Suite fournit une réponse idéale aux exigences essentielles concernant l'adoption de l'IA.

- **Sécurisation de l'utilisation des outils IA par le personnel :** appliquez des politiques de sécurité Zero Trust pour contrôler l'accès du personnel aux outils d'IA générative publics tels que ChatGPT, et aux applications pilotées par IA, développées en interne.
- **Protection des applications pilotées par IA accessibles au public :** protégez les applications web et les API intégrant des modèles IA (à l'image des chatbots et des moteurs de recommandation) contre les attaques susceptibles d'exposer des données sensibles ou d'entraîner une utilisation abusive du modèle.
- **Gestion de l'IA fantôme :** identifiez automatiquement les outils IA non autorisés dans l'ensemble de votre entreprise et appliquez les contrôles appropriés afin d'encourager une innovation continue, dans des limites de risque maîtrisées.
- **Prévention des pertes de données pilotée par IA pour les interactions avec l'IA :** empêchez l'exposition de données sensibles dans les prompts ou les réponses de l'IA, tout en assurant la protection des informations d'identification personnelle et des données confidentielles.
- **Sécurité du développement de l'IA :** fournissez aux équipes d'ingénierie des cadres de travail afin d'imposer la sécurisation du développement comme l'approche par défaut. Vos équipes pourront ainsi développer rapidement des fonctionnalités d'IA sans compromettre la sécurité.

Réflexions pour la mise en œuvre

Une stratégie de sécurité fondamentale consiste à aborder la sécurité de l'IA avec circonspection.

Commencez avec votre infrastructure actuelle	Les déploiements de solutions de sécurité de l'IA les plus performants reposent sur des outils SASE et de sécurité des applications existants, au lieu des remplacer. Cette approche permet de tirer parti des investissements actuels, tout en étendant la protection pour englober les risques spécifiques à l'IA.
Déployez des protections en ligne unifiées	La capacité de bloquer les activités malveillantes à la périphérie du réseau, à l'instant où elles se produisent, est essentielle pour la sécurité de l'IA. Déployez des contrôles en ligne en temps réel et complétez-les par une surveillance basée sur les API.
Veillez à déployer une couverture complète	Votre stratégie de sécurité de l'IA doit répondre à l'ensemble des besoins de l'entreprise : utilisation des outils d'IA générative par le personnel, protection des applications et des flux de travail pilotés par IA, sécurisation des flux de travail d'IA agentique et sécurité des flux de travail de développement d'IA.
Planifiez un déploiement à la mesure de votre entreprise	Optez pour des solutions capables d'évoluer à la mesure de l'adoption de l'IA par votre entreprise. Si une approche s'avère efficace dans le cadre d'un programme pilote, elle doit également évoluer à l'échelle de l'ensemble de l'entreprise, à mesure que l'utilisation de l'IA augmente.
Validez vos critères de réussite	Avant de vous engager complètement, validez la solution dans des scénarios d'utilisation réels. Optez pour une plateforme permettant l'activation gratuite, en libre-service des fonctionnalités pour entreprises. Cette approche vous permettra de tester la suite de sécurité intégrale à petite échelle (par exemple, pour une équipe ou une application unique), afin de démontrer rapidement sa valeur et de vous assurer qu'elle satisfait à vos critères de réussite.

Passer à l'étape suivante avec Cloudflare AI Security Suite

À mesure que l'adoption de l'IA s'accélère dans le monde entier, les entreprises capables de déployer l'IA à grande échelle en toute sécurité bénéficieront d'un avantage concurrentiel considérable. L'essentiel est de reconnaître que la sécurité de l'IA ne consiste pas à empêcher son utilisation, mais à l'encourager intelligemment.

Cloudflare AI Security Suite permet aux entreprises d'innover avec confiance. Bâtie sur nos célèbres plateformes SASE et de sécurité des applications, dont elle vient étendre les fonctionnalités, notre offre AI Security Suite réunit la découverte intégrée des applications IA, des contrôles d'accès Zero Trust, une défense proactive contre les menaces basée sur un corpus d'informations et une gouvernance robuste des données. En sécurisant l'utilisation de l'IA et les modèles IA sous tous les angles, les entreprises peuvent donner aux développeurs les moyens de créer plus rapidement et permettre aux collaborateurs d'être plus productifs, sans pour autant sacrifier l'expérience des utilisateurs finaux.

Demander un entretien

Découvrez comment la suite Cloudflare AI Security Suite peut transformer l'approche de votre entreprise concernant la sécurité de l'adoption de l'IA.

01 73 01 52 44

enterprise@cloudflare.com

www.cloudflare.com

1. ManageEngine. The Shadow AI Surge in Enterprises: Insights from the US & Canada
2. Check Point Research. Q1 2025 Global Cyber Attack Report from Check Point Software: An Almost 50% Surge in Cyber Threats Worldwide, with a Rise of 126% in Ransomware Attacks