CLOUDFLARE

E-BOOK

Garantir la sûreté des pratiques concernant l'IA

Guide de la création d'une stratégie évolutive
en matière d'IA à l'intention des RSSI

Sommaire

Synthèse

Nous vous souhaitons la bienvenue, RSSI !

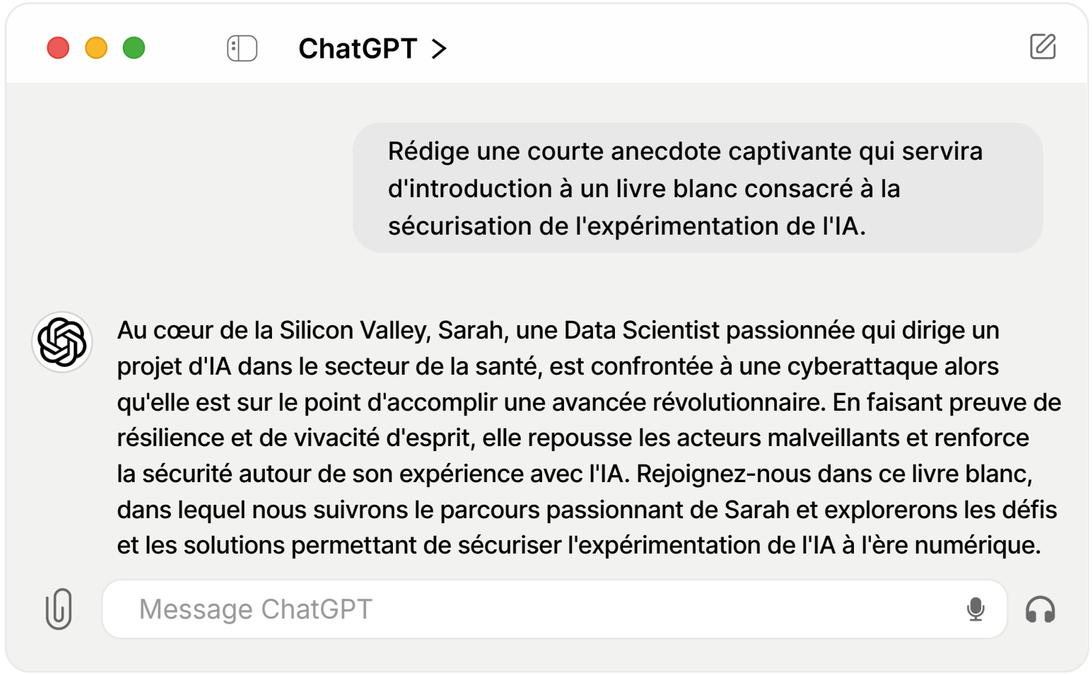
L'IA figure certainement parmi les termes les plus en vogue à l'heure actuelle, et elle incarne également l'une des problématiques les plus urgentes pour la communauté des professionnels de la sécurité. Son influence exige notre attention ; c'est pourquoi Cloudflare a rédigé ce guide, qui vous aidera à élaborer une réflexion approfondie sur l'expérimentation sécurisée de l'intelligence artificielle générative (GenAI) au sein de votre entreprise.

Les outils basés sur l'IA deviennent rapidement plus puissants et plus accessibles, présentant des opportunités d'innovation dans tous les secteurs. Cependant, à l'image d'autres changements de paradigme, la GenAI présente des défis uniques en matière de sécurité, de confidentialité et de conformité. L'adoption généralisée de la GenAI peut entraîner des pics d'utilisation imprévus, des scénarios d'utilisation abusive, des comportements malveillants et des pratiques dangereuses liées à l'informatique fantôme (Shadow IT) – autant de facteurs qui élèvent le risque de violations de données et de fuites d'informations sensibles.

À mesure que son adoption progresse sur votre lieu de travail, vous devez vous préparer en élaborant un schéma directeur dédié à la GenAI, qui précise comment utiliser, développer et sécuriser à grande échelle. Parlons des risques et examinons les conseils dont votre équipe peut bénéficier pour sécuriser la GenAI en fonction des niveaux de maturité et de l'utilisation. Avec ces stratégies, votre entreprise pourra élaborer une stratégie de déploiement de la GenAI qui réponde à ses besoins, tout en protégeant ses données et en garantissant la conformité.

– Dawn Parzych, Director of Product Marketing, Cloudflare

Sécuriser l'expérimentation de la GenAI



Nous sommes désolés de vous l'annoncer, mais l'histoire de Sarah s'arrête ici. Tandis que nous faisons nos adieux à notre personnage fictif, tandis que l'IA prédictive et la GenAI se développent, d'innombrables autres « Sarah » existent dans la vie réelle, incarnant chacune des héroïnes au sein d'équipes informatiques et de développement, des spécialistes des technologies pour entreprises ou des collaboratrices individuelles.

L'IA a enchanté les spécialistes de la technologie et les utilisateurs quotidiens, suscitant la curiosité et l'expérimentation. Si cette expérimentation est nécessaire pour révéler tout le potentiel de l'IA, en l'absence de précautions et de garde-fous, elle peut également entraîner des compromissions de la sécurité ou des défauts de conformité.

Pour parvenir à l'équilibre et pour comprendre et gérer plus efficacement les initiatives en matière d'IA, les entreprises doivent réfléchir à trois aspects essentiels :

1 L'utilisation de l'IA

L'utilisation de technologies IA (par exemple, ChatGPT, Bard et GitHub Copilot) proposées par des fournisseurs tiers, tout en veillant à la protection de leurs ressources (par exemple, les données sensibles, la propriété intellectuelle, le code source, etc.) et en atténuant les risques potentiels en fonction des scénarios d'utilisation

2 Le développement de l'IA

Le développement de solutions IA sur mesure, adaptées aux besoins spécifiques d'une entreprise (par exemple, des algorithmes propriétaires utilisés aux fins de l'analyse prédictive, des copilotes ou des chatbots accessibles aux clients ou un système de détection des menaces piloté par IA)

3 La sécurisation de l'IA

La protection des applications et systèmes basés sur l'IA contre leur manipulation par des acteurs malveillants, entraînant des comportements imprévisibles

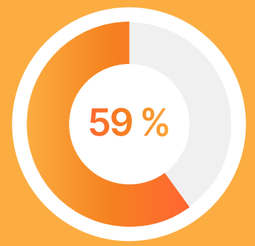
Sécuriser l'expérimentation de la GenAI

La transformation de la GenAI : aujourd'hui et à l'avenir

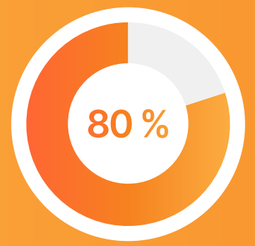
L'intérêt des consommateurs et des entreprises pour la GenAI permet à celle-ci de bénéficier d'une trajectoire d'adoption sans précédent. Un petit groupe d'utilisateurs chevronnés s'est rapidement développé, en partie grâce à une communauté open source active et à l'expérimentation, par le grand public, d'applications telles que ChatGPT et Stable Diffusion.

Ce que les utilisateurs ont découvert, à travers ces pérégrinations, c'est qu'en réalité, les robots ne vont pas « remplacer les humains ».

La GenAI offre aux humains la capacité d'affiner et d'augmenter, plutôt que de tout créer en partant de rien, et peut aider les entreprises à amplifier l'efficacité de leur personnel. L'IA prédictive offre des avantages similaires en facilitant l'exploitation des données dans le but d'améliorer la prise de décision, de développer des produits plus intelligents et de personnaliser les expériences des clients, entre autres initiatives.



Aujourd'hui, **59 % des développeurs** utilisent actuellement l'IA dans leurs flux de développement[1]



D'ici 2026, **80 % des entreprises** utiliseront des API, des modèles et/ou des applications orientés GenAI, déployés dans des environnements de production (contre 5 % aujourd'hui)[2]

D'ici 2030, la GenAI augmentera **50 % des tâches des travailleurs du savoir** (contre <1 % aujourd'hui), permettant d'accroître la productivité ou d'améliorer la qualité moyenne du travail[3]

1. SlashData, « How developers interact with AI technologies », mai 2024
2. Gartner, « A CTO's Guide to the Generative AI Technology Landscape », septembre 2023
3. Gartner, « Emerging Tech: The Key Technology Approaches That Define Generative AI », septembre 2023

Garantir l'utilisation sécurisée de la GenAI

L'expérimentation de l'IA englobe aussi bien l'utilisation d'outils et de services IA prêts à l'emploi que la création intégrale de solutions IA conçues sur mesure. Si certaines entreprises s'orientent vers la création de leurs propres modèles et applications IA, beaucoup se contenteront d'utiliser des outils IA tiers.

Dans ces cas, les outils IA tiers entraînent de nouveaux risques, car les contrôles directs dont les entreprises disposent sur leurs configurations de sécurité et de confidentialité sont limités.

56 % des collaborateurs ont utilisé des outils basés sur l'IA pour exécuter des tâches professionnelles. Toutefois, seulement 26 % des entreprises disposent d'une politique régissant le recours à l'IA[4]

Il est probable que le personnel utilise déjà des outils IA prêts à l'emploi dans le cadre de son travail, par l'intermédiaire de suites SaaS telles que Microsoft 365, de chatbots intégrés aux moteurs de recherche ou à d'applications publiques, et même d'API.

4. The Conference Board, septembre 2023

Si les entreprises aspirent à minimiser les risques, elles doivent prendre des dispositions préalables, notamment :

- **Évaluer** le risque lié à la sécurité des outils tiers
- **Répondre** aux préoccupations relatives à la confidentialité des données
- **Gérer** la dépendance (voire la dépendance excessive) à l'égard d'API externes
- **Surveiller** les vulnérabilités potentielles

C'est notamment le cas, par exemple, lorsque le personnel utilise des applications web publiques telles que ChatGPT. Toute information introduite dans une invite de commande devient un fragment de données qui échappe au contrôle de l'entreprise. Les utilisateurs sont susceptibles de divulguer par inadvertance des informations sensibles, confidentielles ou réglementées, telles que des informations d'identification personnelle (IPI), des données financières, des éléments de propriété intellectuelle ou du code source. Et même si les utilisateurs ne partagent pas d'informations sensibles explicites, il est possible de reconstituer le contexte des saisies afin d'en déduire des données sensibles.

Pour prévenir ce risque, le personnel peut activer un paramètre afin d'empêcher l'utilisation des informations saisies pour former davantage le modèle, mais il doit le faire manuellement. Pour assurer leur sécurité, les entreprises doivent donc trouver des moyens d'empêcher les utilisateurs de saisir des données privées.

Se préparer aux implications de l'IA en matière de sécurité

Exposition des données

Dans quelle mesure certains utilisateurs partagent-ils de manière inappropriée des données sensibles avec des services IA externes ? Les techniques d'anonymisation/de pseudonymisation sont-elles suffisantes ?

Risques liés aux API

Comment allez-vous remédier aux vulnérabilités d'API tierces pouvant potentiellement constituer des passerelles pour des acteurs malveillants ?

Systèmes opaques

Quels processus décisionnels des modèles IA externes sont susceptibles d'introduire des risques inattendus ?

Gestion du risque lié aux fournisseurs

Que savez-vous sur les pratiques de sécurité de vos fournisseurs d'IA tiers ? Et surtout, que ne savez-vous pas ?

Mesures pour défendre l'utilisation de l'IA

1 Maîtrisez la gouvernance et le risque

- Élaborez des politiques régissant l'utilisation de l'IA et les circonstances dans lesquelles son utilisation est autorisée, notamment les informations que l'entreprise autorise les utilisateurs à partager avec la GenAI, les directives régissant le contrôle des accès, les exigences en matière de conformité et la procédure de signalement des violations
- Réalisez une analyse de l'impact afin de collecter des informations, d'identifier et de quantifier les avantages et les risques liés à l'utilisation de l'IA

2 Améliorez la visibilité et les contrôles en matière de sécurité et de confidentialité

- Journalisez toutes les connexions, notamment aux applications IA, afin de surveiller continuellement les activités des utilisateurs, l'utilisation des outils IA et les modèles d'accès aux données, et ainsi, de détecter d'éventuelles anomalies
- Identifiez les éventuelles applications d'informatique fantôme (Shadow IT), notamment les outils IA, et prenez la décision d'approuver ou de bloquer les applications ou d'ajouter des contrôles supplémentaires
- Analysez les configurations d'applications SaaS afin d'identifier les risques potentiels pour la sécurité (par exemple, les autorisations OAuth accordées par des applications approuvées à des applications orientées IA non autorisées, entraînant un risque d'exposition de données)

3 Examinez les données transitant vers et depuis les outils IA, filtrez toutes les données susceptibles de compromettre la propriété intellectuelle, d'affecter la confidentialité ou d'enfreindre les restrictions en matière de droits d'auteur

- Déployez des contrôles de sécurité régissant les interactions des utilisateurs avec les outils IA (par exemple, interdire les transferts de données, empêcher le copier/coller, rechercher et bloquer les saisies de données sensibles ou propriétaires)
- Mettez en œuvre des mesures de protection afin de bloquer les bots IA et d'empêcher l'extraction des contenus de votre site web
- Bloquez purement et simplement les outils IA, si aucun autre contrôle n'est possible. Comme nous le savons, les utilisateurs trouveront des solutions de contournement pouvant compromettre votre emprise sur la sécurité

4 Contrôlez l'accès aux applications IA et à l'infrastructure

- Assurez-vous que chaque utilisateur et chaque appareil accédant aux outils IA fasse l'objet d'une vérification d'identité stricte, afin de déterminer qui est autorisé à utiliser les outils IA
- Déployez des contrôles des accès Zero Trust, basés sur l'identité. Appliquez le principe du moindre privilège, afin de limiter les éventuels dommages causés par la compromission de comptes ou des menaces internes

5 Rationalisez les coûts et l'efficacité opérationnelle

- Comprenez comment le personnel utilise les applications IA grâce à l'analyse et la journalisation des données afin de contrôler le volume de requêtes, la mise en cache et les renouvellements de requêtes, ainsi que le recours aux modèles de secours tandis que l'utilisation de ces applications progresse

Sécurisez ce que vous développez

Formez votre modèle IA

Les pipelines IA élargissent le spectre des vulnérabilités. Cependant, grâce à l'expérience acquise au début du processus de développement, et tout au long de celui-ci, nous connaissons désormais les paramètres qui en déterminent le succès. Pour la sécurité de l'IA, le point de départ naturel est votre modèle.

La fondation des applications IA est la suivante : toutes les données utilisées pour former votre modèle IA sont intégrées aux résultats générés par celui-ci. Réfléchissez à la manière dont vous allez sécuriser ces données, dans un premier temps, afin d'éviter des répercussions négatives ultérieures. En l'absence de protections, vous risquez d'étendre votre surface d'attaque et de voir apparaître, à l'avenir, des problèmes liés aux applications.

Une sécurité garantissant l'intégrité des données est essentielle pour atténuer la compromission délibérée ou accidentelle des données. Les risques liés à la sécurité inhérents au pipeline IA incluent notamment :

- **L'empoisonnement de données :** des ensembles de données malveillants influencent les résultats et créent des biais
- **L'utilisation abusive d'hallucinations :** les auteurs de menaces légitiment les hallucinations de l'IA (c'est-à-dire l'invention d'informations afin de générer des réponses) ; les résultats sont alors informés par des ensembles de données malveillants et illégitimes

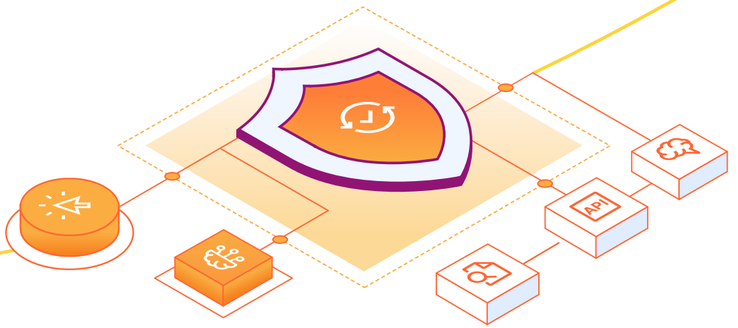
Par ailleurs, si vous ne formez pas de modèles, votre IA interne commencera par sélectionner un modèle pour exécuter des tâches. Dans ces situations, vous avez tout intérêt à explorer l'approche de la création et de la sécurisation du modèle par ses créateurs, car celle-ci joue un rôle dans l'inférence.

L'inférence est le processus qui suit l'apprentissage de l'IA. Plus un modèle est formé et affiné, meilleures seront les inférences – toutefois, sans aucune garantie qu'elles soient parfaites. Même des modèles très bien formés peuvent présenter des hallucinations.

Sécurité post-déploiement

Lorsque vous avez développé et déployé votre IA interne, vous devez protéger ses données privées et en sécuriser l'accès. Outre les recommandations que nous avons déjà formulées dans ce document, notamment la mise en œuvre de jetons pour chaque utilisateur et le contrôle du volume de requêtes, vous devez également réfléchir aux éléments suivants :

- **Gestion des quotas :** applique des limites afin d'éviter la compromission et la diffusion des clés d'API des utilisateurs
- **Blocage de certains numéros de systèmes autonomes (ASN) :** empêche des acteurs malveillants d'envoyer des volumes excessifs de trafic aux applications avec l'objectif de les saturer
- **Mise en œuvre de salles d'attente ou de vérifications des utilisateurs :** rend le traitement des requêtes plus complexe ou plus long, portant préjudice à l'économie des acteurs malveillants
- **Construction et validation d'un schéma d'API :** définit l'utilisation prévue en identifiant et en cataloguant tous les points de terminaison d'API, puis énumère l'ensemble des paramètres et limites de type spécifiques
- **Analyse de la profondeur et de la complexité des requêtes :** aide à prévenir les attaques par déni de service (DoS) et les erreurs des développeurs, tout en préservant l'intégrité de votre serveur d'origine et en servant les requêtes attendues à vos utilisateurs
- **Mise en place d'une discipline promouvant l'accès basé sur des jetons :** protection contre les accès compromis lors de la validation des jetons au niveau de la couche intergicielle (middleware) ou dans API Gateway



Une protection robuste contre les menaces pour l'ensemble de votre expérimentation de la GenAI

De l'adoption à la mise en œuvre, chaque étape du spectre d'expérimentation de la GenAI devrait progresser avec un risque minimal ou toléré. Grâce aux connaissances acquises dans ce document, vous avez le pouvoir de contrôler votre environnement numérique, que votre entreprise utilise, développe ou planifie l'utilisation de l'IA sous une forme ou une autre à l'avenir.

S'il est naturel de se montrer hésitant à adopter de nouvelles capacités, il existe des ressources qui vous donneront la confiance nécessaire pour expérimenter l'IA en toute sécurité. Parmi ces ressources, celle dont les entreprises ont le plus besoin aujourd'hui est un « tissu conjonctif » capable de connecter tous les éléments relatifs à l'informatique et à la sécurité. Il doit permettre de réduire la complexité en travaillant avec tous les éléments de l'environnement, offrir une accessibilité omniprésente et remplir les fonctions nécessaires en matière de sécurité, de connectivité réseau et de développement.

Ce tissu conjonctif vous apportera la confiance indispensable pour une multitude de scénarios d'utilisation, parmi lesquels :

- Assurer la conformité aux réglementations, grâce à la capacité de détecter et de contrôler les mouvements de données réglementées
- Regagner de la visibilité et reprendre le contrôle des données sensibles dans des scénarios tels que les applications SaaS, l'informatique fantôme (Shadow IT) et outils IA émergents
- Sécuriser le code des développeurs en détectant et en bloquant le code source lors des transferts et des téléchargements, mais également prévenir, identifier et corriger les erreurs de configuration dans les applications SaaS et les services cloud, notamment les référentiels de code

Tandis que l'IA continue d'évoluer, l'incertitude est une certitude ; c'est pourquoi une force stabilisatrice comme Cloudflare est si bénéfique.

Protéger l'entreprise contre les risques liés à l'IA pour trois types de LLM

Le niveau d'exposition au risque qu'engendre l'IA pour une entreprise varie en fonction de son utilisation. Il est essentiel de comprendre les différents risques associés à l'utilisation et au développement de grands modèles linguistiques (LLM, Large Language Model), puis de s'impliquer activement dans tous les déploiements de LLM.

Type de LLM	Risque essentiel
Interne	Accès aux données sensibles et à la propriété intellectuelle
Produit	Risque pour la réputation
Public	Fuites de données sensibles

Ampleur, facilité d'utilisation et intégration fluide

Le cloud de connectivité de Cloudflare vous donne le contrôle en améliorant la visibilité et la sécurité, rendant l'expérimentation de l'IA sûre et évolutive. Mieux encore, nos services renforcent tous ces aspects, vous assurant de ne devoir accepter aucun compromis entre l'expérience utilisateur et la sécurité.

Étant donné que la plupart des entreprises se contenteront d'utiliser l'IA ou utiliseront, puis développeront l'IA, le déploiement de Cloudflare vous assurera que vos projets de développement d'IA ne seront jamais bloqués.

- Notre **réseau mondial** vous permet d'étendre et d'appliquer des contrôles avec rapidité, partout où vous en avez besoin
- La **facilité d'utilisation** de notre solution simplifie le déploiement et la gestion des politiques relatives à l'utilisation de l'IA par votre personnel
- Une **architecture programmable** vous permet d'ajouter une couche de sécurité aux applications que vous développez, sans perturber l'utilisation de l'IA par vos collaborateurs

Le cloud de connectivité de Cloudflare protège chaque facette de votre expérimentation de l'IA, et notamment :

- Nos services **Zero Trust et SASE (Secure Access Service Edge)** contribuent à atténuer les risques liés à l'**utilisation** d'outils IA tiers par votre personnel
- Notre **plateforme pour développeurs** aide votre entreprise à **créer** ses propres outils et modèles IA de manière sûre et efficace
- Pour **sécuriser avec l'IA**, notre plateforme tire parti de l'IA et des techniques d'apprentissage automatique pour générer des informations sur les menaces, qui sont ensuite utilisées pour protéger les entreprises pendant leur expérimentation avec l'IA

	Comment vous utilisez l'IA	Comment vous développez l'IA
L'étendue de notre réseau mondial	Déployez et appliquez des contrôles partout, de manière cohérente	Accélérez l'inférence, le traitement des requêtes et la mise en cache
La simplicité de gestion de notre solution	Un plan de contrôle unique, avec un déploiement et des politiques simples	Des modèles pour une intégration rapide
L'architecture unifiée et programmable de notre réseau	Ajoutez une nouvelle couche de sécurité sans perturber votre utilisation de l'IA	Intégrez la confidentialité et la conformité

Prochaines étapes

De la protection de l'utilisation de l'IA par votre entreprise à la défense des applications IA que vous développez, Cloudflare for AI veille sur vous. Avec nos services, vous pouvez adopter de nouvelles fonctionnalités dans l'ordre de votre choix, avec une interopérabilité illimitée et des intégrations flexibles.

Discuter avec un expert

Pour plus d'informations, visitez cloudflare.com

01 73 01 52 44 | enterprise@cloudflare.com | www.cloudflare.com/fr-fr/

RÉV. : BDES-6307.2024AUG19