Microsoft Security

Quatre impératifs pour sécuriser et encadrer l'IA

Un manuel

Sommaire

Présentation

L'IA générative révolutionne la façon dont les entreprises innovent et se développent. Dès lors, il n'est pas étonnant que, selon Gartner, on s'attende à ce que 80 % des entreprises adoptent des applications d'IA d'ici 2026.[1] Mais l'adoption rapide s'accompagne d'une nouvelle vague de préoccupations en matière de sécurité. Plus de 80 % des dirigeants et des professionnels de la sécurité déclarent que leur principale préoccupation est la fuite de données sensibles à travers ces systèmes d'IA.[2]

Il ne s'agit pas seulement d'un problème hypothétique : les entreprises y sont déjà confrontées. Qu'il s'agisse d'outils d'IA de l'ombre introduits sur le lieu de travail sans surveillance ou du risque accru de fuites de données sensibles, les défis sont bien réels. Ajoutons à cela les craintes fondées sur la responsabilité réglementaire (plus de 62 % des dirigeants admettent qu'ils ne comprennent pas pleinement les réglementations sur l'IA[3]) et il apparaît clairement que les équipes de sécurité ont du pain sur la planche.

Il existe pourtant une solution. Pour aider votre équipe de sécurité à relever ces nouveaux défis, nous avons décrit quatre étapes essentielles.

1 Préparation de votre environnement

Tout d'abord, vous devez classer et étiqueter vos données pour adopter l'IA en toute confiance. Mais cela ne s'arrête pas là : la mise en place d'une gouvernance solide de l'identité et de l'accès est essentielle pour prévenir la surexposition des données.

2 Détection des risques

Ensuite, obtenez une visibilité sur la façon dont vos données circulent, sur les personnes qui peuvent y accéder et sur les emplacements des failles de sécurité. Vous éliminerez ainsi des problèmes tels que l'accès non autorisé par des utilisateurs surprovisionnés et le partage excessif de données sensibles.

3 Protection des applications d'IA et des données sensibles

Une fois vos systèmes d'IA activés, ils ont besoin d'une protection continue. Ils doivent ainsi être protégés contre les menaces telles que les fuites de données et les attaques spécifiques à l'IA, telles que les injections d'invites, qui peuvent manipuler le fonctionnement de votre IA.

4 Encadrement de l'utilisation

Enfin, à mesure que les réglementations en matière d'IA évoluent, votre approche doit évoluer. Encadrer vos systèmes d'IA pour détecter et résoudre des problèmes tels que la collusion, le harcèlement ou l'utilisation abusive de contenu sensible sera essentiel pour assurer votre conformité.

Ces impératifs ne se contentent pas de minimiser les risques, ils permettent aussi à votre équipe de tirer parti de la puissance de l'IA en toute sécurité. Approfondissons la façon dont vous pouvez transformer ces défis en un avantage stratégique.

Chapitre 1

Préparation de votre environnement

Avant d'adopter l'IA, il est essentiel de prendre en compte la manière dont les données de votre organisation seront collectées, stockées, traitées et utilisées. Un environnement d'IA est un écosystème complexe. Voici donc trois domaines clés sur lesquels vous concentrer dans le cadre de votre préparation.

Gestion des données

La première étape de la construction d'un système d'IA consiste à sécuriser les données dont il dépend. Classez et étiquetez vos données pour les gérer efficacement. Pour un RSSI, cela garantit le contrôle des flux de données, prévient les violations et maintient la conformité réglementaire.

Règlements de l'intelligence artificielle

> Un environnement réglementaire imprévisible attend tous ceux qui souhaitent intégrer l'IA générative. De nouvelles normes, telles que la loi sur l'intelligence artificielle de l'Union européenne (loi sur l'IA de l'UE) et le Artificial Intelligence Risk Management Framework du National Institute of Standards and Technology des États-Unis (NIST AI RMF), devraient avoir une incidence similaire à celle du Règlement général sur la protection des données (RGPD).

Identité et accès

Une fois que vous avez une visibilité sur vos données, l'étape suivante consiste à contrôler qui peut y accéder. Mettez en œuvre une solide gouvernance de l'identité et de l'accès pour atténuer les risques tels que l'accès non autorisé et les menaces internes.

Développement et déploiement rapides de l'IA

Après avoir sécurisé vos données, vous pouvez vous concentrer sur les applications d'IA dans lesquelles elles transiteront. Assurer un développement sécurisé et une surveillance régulière de ces applications permet de prévenir les vulnérabilités tout au long du cycle de vie de l'IA.

Chapitre 2

Détection des risques

Dans la phase opérationnelle de votre système d'IA, il est essentiel d'identifier les sources de risques. La visibilité dans trois domaines clés est vitale : les données qui circulent dans votre système, les applications qui traitent ces données et les personnes qui interagissent avec elles.

Risques liés aux données

La détection des risques liés aux données implique l'identification des vulnérabilités dans la façon dont les informations sensibles sont stockées, traitées et transmises. L'objectif est de prévenir l'exposition (par une fuite ou une violation) ou la corruption (par empoisonnement des données) de données sensibles.

Exemples de risques liés aux données dans les systèmes d'IA

> **Fuites de données :** se produisent lorsque des renseignements confidentiels sont exposés à des parties non autorisées. Dans un système d'IA, les données peuvent fuir pendant les étapes de formation, de traitement ou de génération.
>
> **Violations de données :** se produisent lorsque des acteurs malveillants contournent les mesures de sécurité pour obtenir un accès non autorisé à des données sensibles.
>
> **Empoisonnement des données :** implique de corrompre les données de formation d'un système d'IA, en compromettant ainsi son intégrité et en biaisant les sorties.

Risques liés aux applications

Les risques liés aux applications peuvent provenir à la fois de vos applications d'IA officielles et des applications d'IA SaaS non approuvées ou risquées que les employés peuvent utiliser seuls. Les applications officielles peuvent présenter des vulnérabilités telles que des erreurs de configuration ou des logiciels sans correctifs. Les applications non approuvées présentent des risques si elles n'ont pas été approuvées pour la sécurité. Il est donc primordial de détecter et de bloquer leur utilisation.

Exemple de risques liés aux applications dans les systèmes d'IA

> Les attaques par injection d'invites manipulent les applications d'IA en introduisant des entrées malveillantes qui provoquent un comportement involontaire du système. Par exemple, un utilisateur peut entrer une commande conçue pour contourner les protocoles de sécurité, ce qui amène l'IA à révéler des données sensibles ou à effectuer des actions non autorisées.

Risques liés aux utilisateurs

Les risques liés aux utilisateurs impliquent des vulnérabilités humaines de la part d'utilisateurs internes ainsi que d'attaquants externes. Les menaces internes peuvent provoquer des fuites avec une intention malveillante ou accidentelle. Les attaquants externes peuvent exploiter les systèmes d'IA par le biais d'actions malveillantes. La détection d'un comportement et d'une activité inhabituels est essentielle pour identifier ces risques rapidement.

Exemples de risques liés aux utilisateurs dans les systèmes d'IA

> **Menace interne :** un employé mécontent ayant accès à des systèmes d'IA sensibles modifie intentionnellement les données d'entraînement, manipulant les résultats pour saboter les opérations de l'organisation.
>
> **Attaque externe :** une attaque d'hameçonnage cible les administrateurs de l'IA, en les flouant pour qu'ils révèlent leurs informations d'identification, ce qui permet aux attaquants d'accéder aux données ou aux modèles sensibles et de les manipuler.

Chapitre 3

Protection des applications d'IA et des données sensibles

Une fois que vous avez établi la visibilité pour vous aider à détecter les risques, vous pouvez vous concentrer sur la protection continue à mesure que votre système d'IA s'exécute et interagit avec des données et des utilisateurs du monde réel. La protection implique la défense des données sensibles à chaque point de contact, l'adaptation des mesures de sécurité en fonction du risque, le contrôle de l'accès dans l'ensemble du système et la réponse rapide aux menaces émergentes.

Protéger les données sensibles

Protégez les données sensibles tout au long de leur cycle de vie. Cela implique le chiffrement pour sécuriser les données pendant la transmission et le stockage, les contrôles d'accès pour s'assurer que seuls les utilisateurs autorisés peuvent interagir avec elles, l'étiquetage pour classer les données en fonction de leur niveau de sensibilité et les mesures de prévention de la perte de données (DLP) pour détecter le partage et le mouvement non autorisés.

Adapter la sécurité en fonction du risque

Maintenez la protection sans restreindre inutilement l'accès ou nuire à l'efficacité en ajustant les contrôles pour s'adapter au niveau de risque posé par les différents utilisateurs, appareils et systèmes. Les contrôles adaptatifs permettent de mettre en place des politiques plus strictes pour les utilisateurs à haut risque tout en appliquant des contrôles plus légers pour les situations à faible risque.

Contrôler les accès

Assurez-vous que seuls les utilisateurs autorisés interagissent avec les applications d'IA et les données sensibles. Lorsque vous mettez en œuvre des stratégies flexibles et centralisées, vous pouvez les adapter aux rôles et aux comportements des utilisateurs. Les contrôles d'accès doivent être basés sur la sensibilité des données et le risque lié aux utilisateurs, en mettant en place l'authentification multifacteur, en limitant l'accès des utilisateurs à haut risque et en examinant régulièrement les autorisations pour assurer la conformité et la sécurité.

Réagir aux menaces

Lorsque vous détectez des risques, il convient d'agir sans tarder. Les outils de gestion des informations et des événements de sécurité (SIEM) peuvent aider à analyser les journaux pour identifier des modèles inhabituels, comme un comportement des utilisateurs ou une activité réseau suspects. Ces outils peuvent déclencher des réactions automatisées qui isolent les applications compromises, révoquent l'accès ou alertent les équipes de sécurité, ce qui garantit la protection de vos systèmes d'IA en temps réel.

Chapitre 4

Encadrement de l'utilisation

Une fois vos systèmes d'IA protégés, la dernière étape consiste à encadrer l'utilisation. À mesure que les réglementations en matière d'IA évoluent, il convient d'encadrer l'utilisation pour maintenir la conformité et traiter les violations des politiques, telles que la collusion, le harcèlement ou le partage de contenu dangereux. Cela implique la mise en œuvre de contrôles réglementaires et de conduite, l'établissement de politiques claires d'utilisation de l'IA, la définition de protocoles de rétention et de suppression des données et la préparation à l'évolution des réglementations.

Rester informé des obligations réglementaires

Assurez-vous que votre organisation se conforme aux réglementations externes, telles que la loi européenne sur l'IA et le RGPD, ainsi qu'aux politiques internes. Des examens et des audits réguliers aident à détecter rapidement les violations et à renforcer l'utilisation éthique de l'IA, pour rester sur la bonne voie.

Exemple d'exigence réglementaire

> L'article 5(1)(e) du RGPD[4] stipule que les données personnelles doivent être « conservées sous une forme qui permette l'identification des personnes concernées pour une durée qui n'est pas plus longue que nécessaire pour accomplir ce pour quoi les données sont traitées ». Cela oblige les organisations à définir des périodes de conservation et à supprimer les données personnelles une fois qu'elles ne sont plus nécessaires.

Se préparer au changement des réglementations

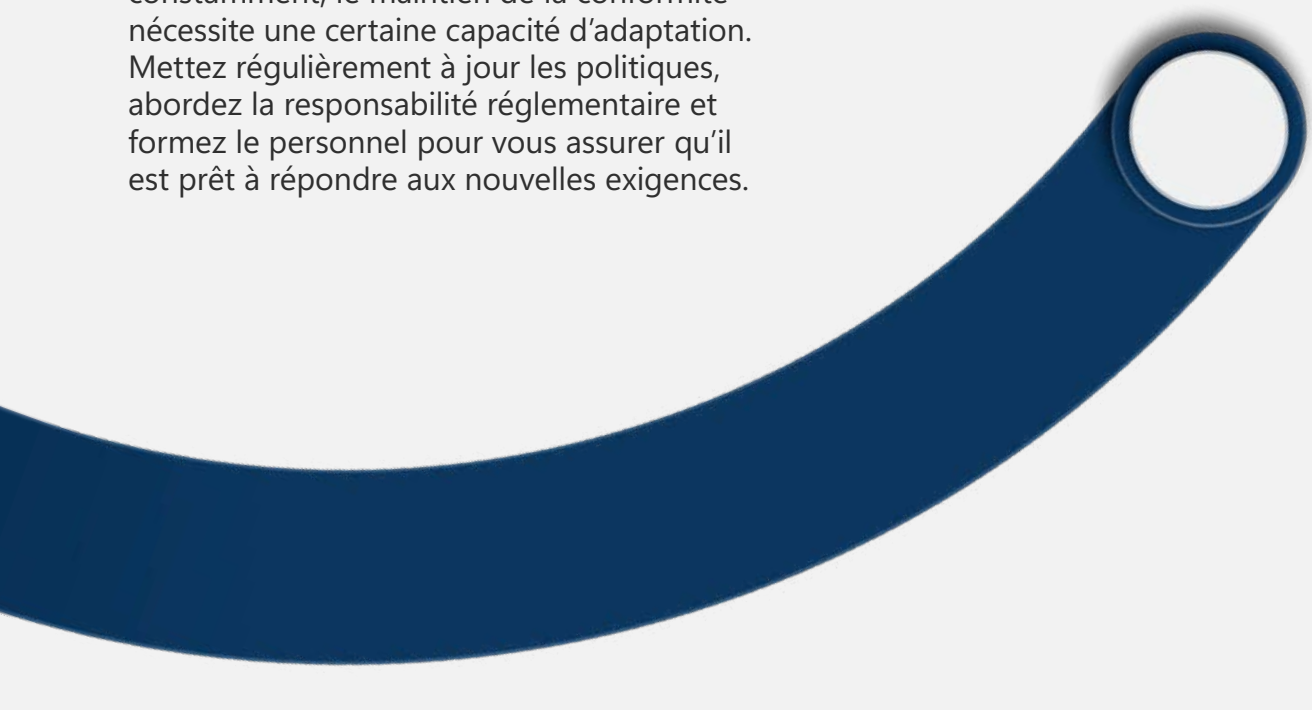
Alors que la réglementation de l'IA évolue constamment, le maintien de la conformité nécessite une certaine capacité d'adaptation. Mettez régulièrement à jour les politiques, abordez la responsabilité réglementaire et formez le personnel pour vous assurer qu'il est prêt à répondre aux nouvelles exigences.

Définir des directives d'utilisation claires de l'IA

Établissez des normes claires sur la façon dont l'IA devrait être appliquée, en garantissant des pratiques éthiques, un traitement sécurisé des données et le respect de la vie privée. Répondez aux préoccupations de gouvernance spécifiques telles que la prévention des hallucinations ou les violations du droit d'auteur, et passez régulièrement les politiques en revue pour s'assurer qu'elles demeurent pertinentes par rapport aux exigences de l'entreprise et des réglementations en constante évolution.

Définir des directives claires pour la conservation et la suppression des données

Définissez la durée de stockage des données et le moment où elles doivent être supprimées pour satisfaire aux exigences de sécurité et de réglementation. Il s'agit notamment d'établir des délais spécifiques pour les interactions avec l'IA, comme les invites et les achèvements, afin de garantir la conformité et de protéger les informations sensibles.



Assurer votre avenir de l'IA

L'IA générative offre d'incroyables possibilités d'innovation et de productivité, mais elle comporte également de nouveaux risques. Pour tirer le meilleur parti de l'IA tout en maintenant la sécurité de votre organisation, vous devez adopter une approche efficace de gestion des données, de protection de vos systèmes et de garantie de conformité aux réglementations fluctuantes.

En suivant les étapes décrites dans le présent manuel (préparer votre environnement, détecter les risques, protéger les applications d'IA et les données sensibles et encadrer l'utilisation), vous vous assurez que vos systèmes d'IA sont sécurisés et que votre organisation est prête à relever les défis à venir.

Sécurité Microsoft fournit les outils et le support pour vous aider à adopter l'IA en toute sécurité, à protéger les données sensibles et à gérer l'utilisation de l'IA de manière responsable.

En savoir plus sur l'approche complète de Microsoft en matière de sécurité de l'IA

[1] « Hype Cycle for Generative AI », Gartner, Inc., 11 septembre 2023

[2] « First Annual Generative AI Study : Business Rewards vs. Security Risks », page 8, Information Security Media Group (ISMG), 31 janvier 2024

[3] « First Annual Generative AI Study : Business Rewards vs. Security Risks », page 6, Information Security Media Group (ISMG), 31 janvier 2024

[4] « Article 5(1)(e) du Règlement général sur la protection des données (RGPD) », Règlement général sur la protection des données (RGPD), Parlement européen et Conseil de l'Union européenne, 25 mai 2018