Présentation de la solution

SkopeAI pour ChatGPT et l'IA générative

Introduction

L'émergence des applications SaaS basées sur l'intelligence artificielle (IA) a révolutionné les méthodes de travail des collaborateurs de l'entreprise au quotidien. Les applications d'IA générative telles que ChatGPT ont offert aux entreprises et à leurs employés d'innombrables opportunités d'accroître leur productivité, de simplifier les tâches, d'améliorer leurs services et d'optimiser leurs opérations. Que ce soit en équipe ou individuellement, ChatGPT permet notamment de générer du contenu, de traiter des textes, de traduire des données, d'élaborer des plans financiers et de déboguer et d'écrire du code. Cependant, les applications d'IA générative exposent également les données à des risques considérables et sans précédent.

L'enjeu de la sécurité des données

Bien que les applications d'IA puissent potentiellement améliore l'efficacité du travail, elles engendrent de nouveaux risques et exposent les données sensibles aux menaces externes. Les organisations doivent surmonter ces obstacles pour garantir la confidentialité, l'intégrité et la sécurité de leurs données. Voici quelques exemples de risques auxquels les données sensibles peuvent être exposées via ChatGPT et d'autres applications d'IA basées sur le cloud :

- Du texte contenant des informations personnelles identifiables (PII) peut être publié et donc exposé sur le chatbot pour demander des modèles d'e-mail, des réponses client, des lettres personnalisées ou une analyse des sentiments.
- Des informations de santé confidentielles, telles que des traitements individuels et des données d'imagerie médicale peuvent être entrées dans le chatbot et potentiellement compromettre les données personnelles des patients.
- Les développeurs de logiciels peuvent charger du code source propriétaire et inédit à des fins de débogage, de complétion de code et d'amélioration des performances.
- Les développeurs de logiciels peuvent même lier directement une application de l'entreprise contenant du code source ou une base de données à des applications d'IA générative via une API. Ce flux de données d'une application à l'autre permet la synchronisation automatique des informations dans le cloud et facilite les tâches répétitives telles que le perfectionnement de la structure du code et l'amélioration de la lisibilité. Toutefois, il est important de noter qu'un tel accès peut potentiellement exposer les données confidentielles à une application tierce dangereuse.
- Les documents confidentiels de l'entreprise tels que les ébauches de rapport d'activité, les documents de fusion-acquisition et les annonces non encore diffusées peuvent être téléchargés pour en vérifier la grammaire et l'orthographe, les exposant par négligence à des fuites de données potentielles.
- Les données financières, telles que les transactions de l'entreprise, les recettes non divulguées, les numéros de cartes de crédit et les notations de crédit des clients peuvent être traitées par ChatGPT à des fins de planification financière, de conformité, de détection des fraudes et d'ouverture de nouveaux comptes, sans aucune mesure de sécurité.
- Au sein du service marketing, un employé pourrait à l'avenir intégrer la totalité de la base de données clients à Salesforce.com par le biais de ChatGPT ou d'autres plugins basés sur l'IA générative, ainsi que de nombreuses autres applications non autorisées via une intégration à OAuth. Cette intégration inter-applications permettra à l'employé d'exploiter les fonctionnalités de GPT, notamment pour automatiser le processus de rédaction d'e-mails destinés aux contacts dont les contrats arrivent à expiration. C'est un autre exemple de flux de données inter-applications non détectable par des solutions réseau en ligne telles que les pare-feu et les passerelles Web sécurisées.

Sécuriser les données sensibles dans le cloud

Les organisations doivent adopter en priorité des contre-mesures efficaces pour protéger la confidentialité et la sécurité des données sensibles sur l'ensemble des applications SaaS managées et non managées, des instances et des comptes personnels. Ainsi, selon un récent rapport de Netskope sur les menaces cloud, 74 % des données usurpées sont transmises aux instances de stockage cloud personnelles d'applications populaires.

Les étapes suivantes sont indispensables pour protéger les informations sensibles et doivent être considérées comme des composantes essentielles d'une technologie de protection des données moderne :

1. **Surveillance et gestion des risques :** mise en œuvre de mécanismes de surveillance pour suivre l'utilisation (correcte et incorrecte) des applications et des instances SaaS présentant des risques. Réalisation d'évaluations des risques régulières pour identifier les failles de sécurité et les résoudre rapidement.
2. **Minimisation des données et contrôle d'accès :** limitation de l'exposition des informations sensibles via les applications SaaS grâce à l'adoption de des stratégies de minimisation des données. Mise en place d'un contrôle d'accès rigoureux permettant de s'assurer que seules les personnes autorisées peuvent accéder aux données sensibles et les manipuler.
3. **Chiffrement et prévention des pertes de données :** application de techniques de chiffrement fort pour protéger les données au repos ou en transit. Déploiement de solutions de prévention des pertes de données (DLP) pour surveiller et prévenir la perte accidentelle ou le vol de données.
4. **Sensibilisation et formation des utilisateurs :** sensibilisation des employés aux risques associés aux applications SaaS basées sur l'IA et formation de ces derniers aux meilleures pratiques de traitement sécurisé des données sensibles. Mise en place d'une culture de protection des données et insistance sur l'importance d'une utilisation responsable.

À mesure que les organisations adoptent des services cloud et des applications reposant sur l'IA telles que ChatGPT, il est primordial de garantir la sécurité des données sensibles. En mettant en œuvre des mesures de protection complètes, notamment la surveillance, le contrôle d'accès, le chiffrement et la formation des utilisateurs, les organisations peuvent atténuer les risques, protéger les informations confidentielles et rgarantir leur conformité dans l'univers en constante évolution du cloud.

Mesures et recommandations de sécurité générales à suivre avec les applications d'IA générative

L'utilisation de modèles d'IA tels que ChatGPT dans le contexte de l'entreprise offre des avantages importants en termes de productivité, de rendement et d'innovation. Toutefois, quand on emploie ces modèles d'IA, il est crucial de garantir la confidentialité et la sécurité des données. Voici quelques meilleures pratiques optimisées qui permettront aux équipes de sécurité et aux employés de protéger les données de l'entreprise :

1. **Déploiement local :** Autant que possible, déployez les modèles d'IA en local sur les machines de votre entreprise. Les données n'ont alors plus besoin de quitter le réseau de votre entreprise, ce qui réduit le risque de fuites de données.
2. **Anonymisation des données :** Demandez aux utilisateurs de l'entreprise de prendre le temps d'anonymiser ou de pseudonymiser les données sensibles avant de les utiliser dans les modèles d'IA. Cela consiste à remplacer les données identifiables par des identificateurs artificiels. Si une fuite venait à se produire, ces données seraient inutilisables sans les identificateurs d'origine.
3. **Chiffrement des données :** Dans la mesure du possible, mettez en œuvre un chiffrement des données les plus confidentielles de l'entreprise, qu'elles soient au repos et en transit. "Ainsi, même en cas d'exposition, ces données restent illisibles sans une clé de déchiffrement.
4. **Contrôle d'accès strict :** Utilisez des mécanismes robustes pour contrôler l'accès aux ressources de l'entreprise et aux référentiels de données afin de restreindre l'interaction avec les modèles d'IA et les données associées.
5. **Pistes d'audit :** Tenez à jour des journaux d'audit détaillés sur toutes les activités liées au traitement des données et à l'exploitation des modèles d'IA. Ces journaux aident à identifier les activités suspectes et servent de référence pour les enquêtes futures.
6. **Minimisation des données :** Formez tous les employés à afin qu'ils appliquent la règle qui consiste à n'utiliser que la quantité de données minimale nécessaire au fonctionnement efficace du modèle d'IA. En limitant l'exposition des données, il est possible de réduire l'impact potentiel d'une violation de données.
7. **Mises à jour régulières et correctifs :** Veillez à mettre à jour les logiciels locaux avec les derniers correctifs et mises à jour. Cette précaution permet de se prémunir contre les failles de sécurité connues.
8. **Audits et certifications tiers :** Sélectionnez les services d'IA de fournisseurs qui se sont soumis à des audits tiers rigoureux, possèdent des certifications telles que ISO 27001 ou SOC 2 et se conforment au RGPD RGPD.
9. **Règles d'utilisation des données :** Instaurez des règles claires en matière de gestion et d'utilisation des données au sein de votre organisation. Assurez-vous que les employés les connaissent bien et qu'ils comprennent l'importance de la sécurité des données.
10. **Sauvegarde des données :** Sauvegardez régulièrement les données pour pouvoir les restaurer en cas de perte ou de compromission.
11. **Examen constant :** Il est toujours judicieux d'examiner les toutes dernières règles et conditions d'utilisation d'un outil d'IA donné pour comprendre comment il utilise les données envoyées via l'API en vue d'améliorer ses modèles.

La sécurisation des données sensibles par Netskope dans le cadre de l'utilisation d'applications d'IA générative

Netskope est un leader du marché de la sécurité du cloud et de la protection des données. Forte de plus de dix ans d'expérience, la société offre la plus large visibilité et le contrôle le plus précis sur des milliers de nouvelles applications SaaS telles que ChatGPT. Netskope propose la solution de sécurité SkopeAI for GenAI, spécialement dédiée à l'utilisation d'applications d'IA générative telles que ChatGPT d'OpenAI, Bing AI, Google Bard et bien d'autres. Voici quelques fonctionnalités technologiques de base offertes par Netskope aux équipes de sécurité informatique pour protéger les données sensibles, ainsi que la marche à suivre pour les exploiter facilement afin de sécuriser ChatGPT et d'autres outils d'IA générative :

Contrôle de l'accès aux applications

1. Tout commence par la visibilité. Netskope fournit des outils automatisés qui permettent aux équipes de sécurité de surveiller constamment les applications (telles que ChatGPT) auxquelles les utilisateurs de l'entreprise tentent d'accéder, comment, quand, à partir d'où, à quelle fréquence, etc. Il est essentiel de comprendre les différents niveaux de risque que chaque application présente pour l'organisation et de pouvoir définir de façon granulaire des règles de contrôle d'accès en temps réel en fonction des catégories et des conditions de sécurité qui peuvent évoluer au fil du temps.
    - Par exemple, les équipes de sécurité gagneraient beaucoup à mieux connaître les nombreuses applications utilisées par les employés de l'entreprise. Face aux milliers de nouvelles applications disponibles, il est essentiel de pouvoir les filtrer et les classer par nom, utilisation ou catégorie (par exemple, ChatGPT, réseaux sociaux, collaboration, référentiels de fichiers, etc.). En outre, pour les équipes de sécurité, il est important comprendre le niveau de risque, les normes de conformité, les activités et les informations d'utilisation de chaque application.

2. Si les applications explicitement malveillantes doivent être bloquées, bien souvent, en matière de contrôle d'accès, la responsabilité de l'utilisation d'applications telles que ChatGPT devrait incomber aux utilisateurs, afin de tolérer, les activités qui ont du sens pour un sous-ensemble d'utilisateurs ou la majorité d'entre eux. Dans le même temps, les équipes de sécurité sont chargées de sensibiliser les utilisateurs aux applications et activités jugées risquées. Pour cela, il est notamment possible de recourir principalement à des alertes en temps réel et à des flux de travail d'encadrement automatisés qui impliquent l'utilisateur dans les décisions d'accès une fois qu'il a pris connaissance du risque. Netskope offre des options de sécurité souples permettant de contrôler l'accès aux applications SaaS d'IA générative telles que ChatGPT et de protéger automatiquement les données sensibles.
    - Les règles de contrôle d'accès comprennent par exemple des flux de travail d'encadrement en temps réel qui se déclenchent à chaque fois que les utilisateurs ouvrent ChatGPT, sous forme de fenêtres d'avertissement contextuelles et personnalisables contenant des recommandations sur l'utilisation responsable de l'application, le risque potentiel associé, ainsi qu'une demande d'accusé de réception ou de justification.

Détection avancée et protection des données sensibles

Les utilisateurs peuvent faire des erreurs et compromettre des données par négligence. S'il est possible d'accorder un accès à ChatGPT, il est crucial de limiter le téléchargement et la publication de données hautement sensibles dans le cloud via ChatGPT, que ce soit directement, indirectement ou par le biais de tout autre vecteur d'exposition des données potentiellement dangereux. Seul Netskope est en mesure d'offrir une telle protection avec ses techniques de prévention de la perte de données (DLP) modernes et ses contrôles de sécurité du cloud avancés. Grâce à la prévention de la perte de données (DLP) de Netskope reposant sur des modèles d'apprentissage automatique et d'IA, des milliers de types de fichiers, d'informations personnelles identifiables, de contenus relevant de la propriété intellectuelle, d'états financiers, et autres données sensibles sont identifiés de façon fiable et protégés automatiquement contre une exposition indésirable et non conforme. Netskope détecte et sécurise les données sensibles en mouvement, au repos et en cours d'utilisation, via toutes les connexions utilisateur possibles, au bureau, au sein du centre de données, à la maison et en déplacement.

1. Tout d'abord, la DLP avancée de Netskope peut identifier automatiquement les flux de données sensibles et classifier les publications sensibles avec un niveau de précision extrêmement fin.

   Grâce à cette précision, le système protège chaque élément d'information sensible de format structuré ou non" with "information, que son format soit structuré ou non, y compris des" with "notamment les images, captures d'écran, fichiers compressés, notes, messages chat, etc. Il est également fondamental de s'assurer que seules les données sensibles soient détectées et non les requêtes sans danger ou les tâches sûres traitées par le chatbot. Cela se fait automatiquement au moyen d'un ensemble complet de technologies de détection de données et d'algorithmes de classification avancés qui comprend à la fois des règles de reconnaissance des données définies manuellement et des moteurs de détection automatisés tels que des techniques d'apprentissage profond, le traitement automatique des langues l'analyse sémantique et des sentiments. L'apprentissage profond et le traitement automatique des langues tirent parti de l'apprentissage automatique de tâches complexes, avec et sans supervision, très semblables dans leur approche à celles qu'utilisent les modèles d'IA générative.
2. Netskope DLP permet une classification des images basée sur l'intelligence artificielle (IA) et l'apprentissage automatique (AA), ainsi qu'une reconnaissance optique des caractères. Elle est en outre capable de reconnaître automatiquement les fichiers et types de documents sensibles ssur la base de nombreuses caractéristiques d'identification. Ces modèles exploitent également les algorithmes de réseau neuronal convolutif (CNN) et la détection d'objets YOLOv5 basée sur l'intelligence artificielle afin d'analyser l'imagerie visuelle. Plus précisément, ces techniques permettent au système de détecter automatiquement les images électroniques de passeports, de permis de conduire, de photos d'identité, de formulaires fiscaux, de cartes médicales, de code source, de cartes Vitale, de cartes de crédit/débit, de CV, d'accords de confidentialité, de brevets, de fusions et acquisitions et de chèques, pour n'en citer que quelques-uns. Et ce, avec plus d'exactitude et d'efficacité, même quand ces images et documents sont partiellement endommagés, chiffonnés, flous, et pas très nets.
3. Netskope DLP permet en outre de créer des classificateurs personnalisés reposant sur l'apprentissage automatique. Grâce à la fonctionnalité Train Your Own Classifier (entraînez votre propre classifieur d'images) basée sur un apprentissage automatique supervisé, les organisations peuvent entraîner le système à identifier de nouveaux jeux de données uniques sous forme de fonctions d'apprentissage automatique irréversibles sans données personnelles identifiables.
4. Il est important de garantir lune protection optimale des documents propriétaires critiques en empêchant leur exfiltration ou duplication non autorisée. Des techniques de fingerprinting des fichiers et des documents peuvent servir à indexer des documents entiers et à identifier des copies précises ou partielles des informations qu'ils contiennent. En particulier,

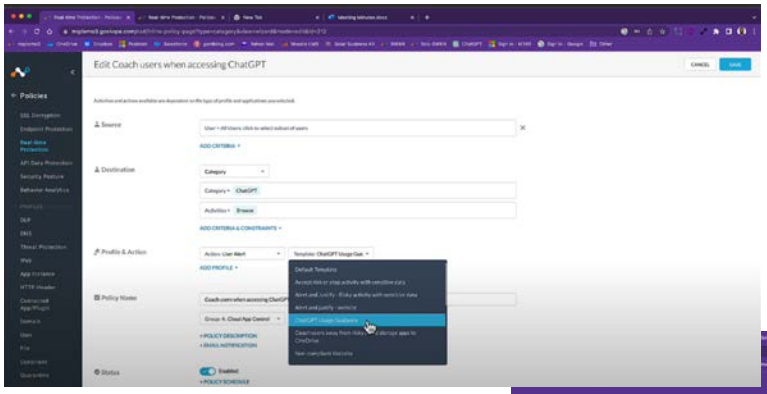
Netskope DLP peut procéder à l'analyse sémantique par apprentissage profonde d'intégrations de suites de mots dans les documents. Elle code ensuite ces intégrations sous forme de vecteurs numériques, puis calcule les similitudes de cosinus. En détectant les similitudes de contenu entre différents environnements et canaux de transmission, ces techniques permettent de mieux identifier et prévenir la dissémination non autorisée.

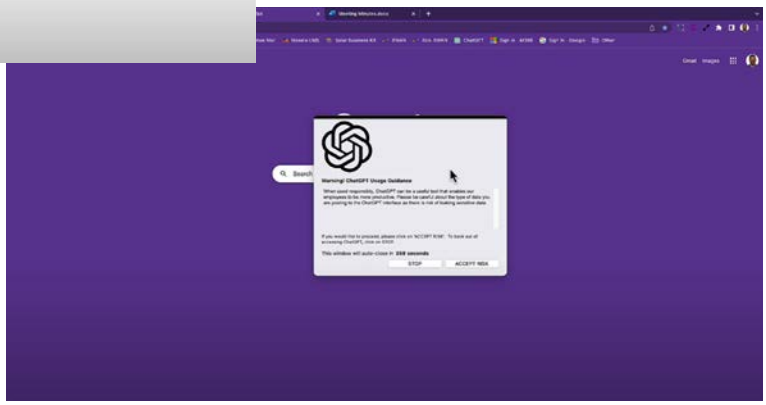
Protection des données en temps réel et encadrement automatique des utilisateurs

1. Netskope DLP offre plusieurs options coercitives imposant d'arrêter et de limiter le téléchargement et l'envoi de données hautement sensibles via ChatGPT. Ces fonctions en temps réel s'appliquent à chaque connexion de l'utilisateur, et garantissent" with "garantissant ainsi la protection des données dans l'environnement de travail hybride moderne au sein duquel les collaborateurs de l'entreprise se connectent depuis le bureau, leur domicile ou en déplacement. Par exemple, dans le cas de ChatGPT et d'autres applications d'IA générative, outre le blocage sélectif des téléchargements et des envois d'informations sensibles, des messages d'encadrement visuels peuvent être générés automatiquement en temps réel pour donnes des consignes en cas de violation d'envois de données, informer" with "afin de donner des consignes en cas d'infraction aux règles de publication de données, d'informer l'utilisateur sur les règles de sécurité de l'entreprise et de permettre de réduire les comportements à risque répétés au fil du temps, ceci afin d'alléger la charge de travail des équipes de sécurité.

   Netskope DLP s'intègre en natif à la solution complète Netskope Security Service Edge (SSE) et est constamment informée sur le comportement des utilisateurs, la géolocalisation, les postures de sécurité, les risques pesant sur les appareils, les risques liés aux applications et les réputations, les instances d'application personnelles, etc. La DLP peut ainsi s'adapter automatiquement au paysage des risques en constante évolution et personnaliser la réponse sécuritaire en fonction de la situation.

2. Avec applications d'IA générative, la protection des téléchargements de données peut s'avérer insuffisante. Les développeurs, par exemple, peuvent désormais intégrer ChatGPT et d'autres modèles à leurs applications et produits via l'API, ou des dérivés de ChatGPT (comme AutoGPT) à leurs flux de travail. Un développeur peut ainsi lier un code source propriétaire, une base de données entière dans le cloud, une feuille de calcul Excel 365 en ligne ou autoriser l'accès complet à une application. Le seul fait de surveiller les données sensibles en ligne au moyen des pare-feu traditionnels et de la DLP ne permet pas de détecter pas ces voies de sortie" with "ces chemins de fuite lorsque les données sont déjà dans le cloud et ne circulent plus en ligne. Netskope fournit une solution de protection des données complète destinée aux applications SaaS qui reconnaît et protège les données sensibles en ligne et dans le cloud. Cette solution sélectionne les données sensibles et les empêche d'êtres transférées" with" empêche leur transfert dans le cloud, offrant ainsi une protection contre l'accès non autorisé des applications aux données sensibles déjà présentes dans le cloud. En outre, Netskope offre une visibilité sur les intégrations de cloud à cloud à des fins d'évaluation et d'atténuation des risques.





À mesure que de nouvelles fonctionnalités et de nouveaux écosystèmes applicatifs voient le jour, cette approche offre la protection des données la plus performante et la plus complète pour préserver les informations confidentielles, le code source des applications des développeurs, les bases de données clients sur Salesforce.com et bien plus encore. Elle limite ou empêche l'exposition des données sensibles aux applications d'écosystèmes non dignes de confiance, y compris les applications d'IA générative.

Autres mécanismes de contrôle de Netskope et remarques finales

- Netskope propose également des mécanismes robustes reposant sur les principes du zero trust pour contrôler l'accès aux référentiels de données de l'entreprise afin de restreindre l'interaction avec les modèles d'IA et les données associées. Cela atténue considérablement le risque des menaces internes.
- Une autre mesure de sécurité importante consiste à identifier le comportement d'utilisateurs malveillants et les anomalies de comportement répétées. Le module d'analyse du comportement des utilisateurs et des entités (UEBA) intégré de Netskope est un élément constitutif de la plateforme de sécurité de Netskope axé sur l'analyse du comportement des utilisateurs et des entités qui détecte et atténue les menaces de sécurité potentielles. Les solutions UEBA recourent à une analyse avancée et à des algorithmes d'apprentissage automatique pour surveiller les activités des utilisateurs, le trafic réseau et les modèles d'accès aux données en vue d'identifier les comportements anormaux ou suspects. Plus spécifiquement, Netskope UEBA vise à explorer les comportements des utilisateurs, tels que leurs interactions avec les applications cloud, les transferts de données, les activités de connexion et les autorisations d'accès aux données. En analysant ces modèles de comportement, Netskope UEBA aide les organisations à identifier les menaces internes, les comptes compromis, les tentatives d'exfiltration de données et les autres risques de sécurité.
- Outre les cas d'utilisation décrits ci-dessus, Netskope offre un ensemble complet de fonctionnalités de sécurité reposant sur l'IA et l'apprentissage automatique au sein de sa plateforme de sécurité, notamment :
  - Des modèles d'apprentissage automatique avancés pour la détection des logiciels malveillants, lesquels viennent compléter les signature plus traditionnelles, les méthodes heuristiques et les techniques de « bac à sable »
  - Une protection contre le hameçonnage et le filtrage des URL, grâce à la génération de signature URLF automatisée, la détection des DGA, la détection des domaines de fast flux, le filtrage des contenus Web et la catégorisation
  - La sécurité de IdO assurant la classification et l'identification des appareils IdO, le regroupement dynamique des appareils et la détection des anomalies
  - La détection des accès anormaux au WAN
  - L'automatisation des flux de travail, la surveillance de la santé des applications, l'autoscaling du cloud et la priorisation adaptative des incidents.

Pour en savoir plus, consultez :

www.netskope.com/skopeai

www.netskope.com/solutions/netskope-for-chatgpt-and-generative-ai

www.netskope.com/products/security-service-edge