Comprendre le contrôle du contenu d'IA nouvelle génération et l'extension de la couverture des outils d'IA DLP

Table des matières

Introduction

Ce document décrit le nouveau contrôle de contenu d'IA génératif et l'extension de la couverture des outils d'IA DLP pour Umbrella.

Aperçu

Nous sommes heureux d'annoncer la disponibilité générale de Generative AI Content Control. Cette fonctionnalité vous permet de surveiller et, si nécessaire, de bloquer le contenu généré par ChatGPT.

Nous sommes également ravis de vous annoncer que nous avons étendu notre couverture DLP en temps réel pour les outils d'IA générative. Initialement limité à ChatGPT, nous prenons désormais en charge l'ensemble des 70 outils d'IA dans notre catégorie d'applications d'IA générative récemment publiée. Cette extension significative vous permet d'élargir l'exemple d'utilisation de l'IA en toute sécurité, offrant une solution plus complète et robuste pour la protection de l'utilisation de l'IA générative.

Comment DLP peut-il aider à contrôler le contenu généré par ChatGPT ?

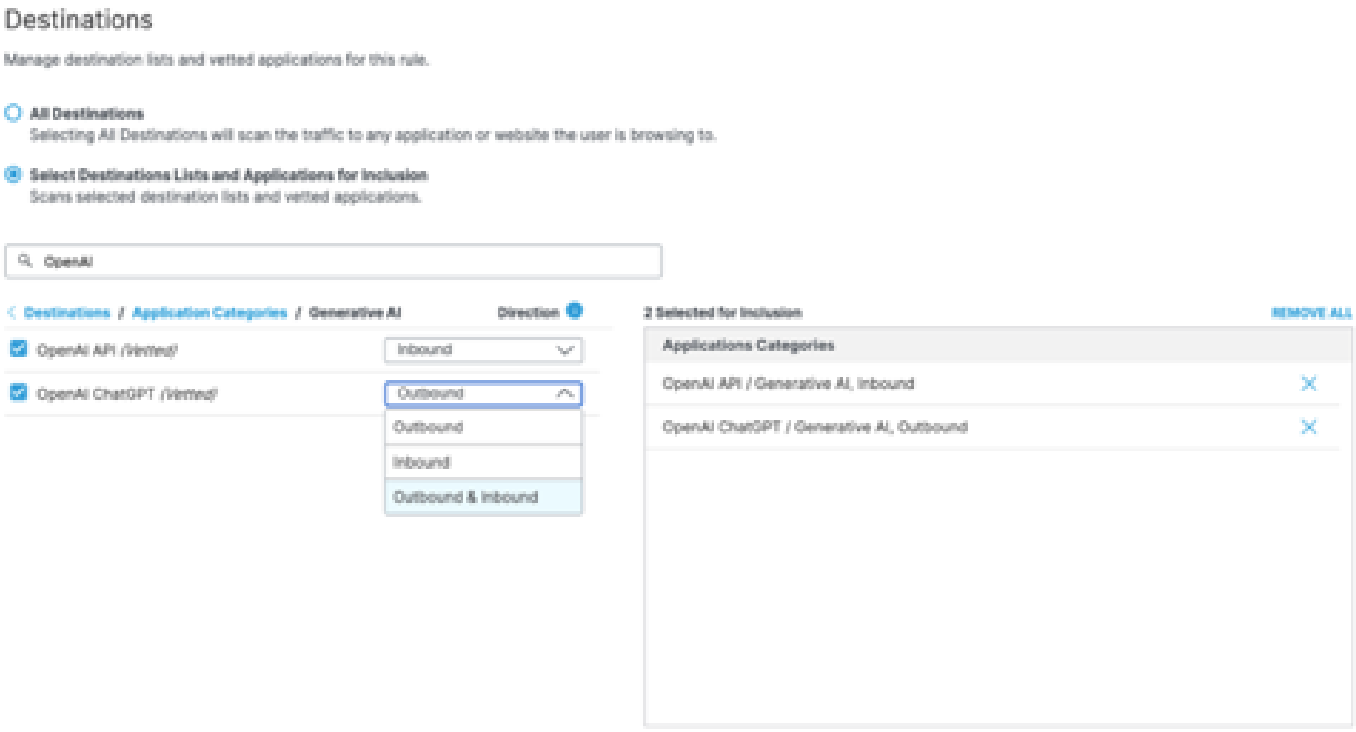
DLP peut aider les organisations à contrôler le contenu généré en analysant les réponses ChatGPT à l'aide d'une stratégie DLP en temps réel. Avec cette version, vous pouvez choisir d'analyser les réponses ChatGPT (c'est-à-dire le trafic entrant) pour tout type de contenu généré que vous souhaitez surveiller ou bloquer.

Pourquoi contrôler le contenu généré par l'IA ?

L'utilisation de contenu généré par l'IA présente des risques pour les entreprises pour diverses raisons, notamment la violation des droits d'auteur, des informations inexactes, un code défectueux, etc.
Par exemple, vous pouvez empêcher les utilisateurs d'utiliser du code source généré par IA pour empêcher l'utilisation de code protégé par des droits d'auteur ou dangereux, tandis que d'autres peuvent souhaiter empêcher l'utilisation de citations judiciaires générées par IA par crainte de fournir des informations inexactes.

Comment puis-je appliquer l'analyse DLP aux réponses ChatGPT ?

En général, la fonction DLP en temps réel analyse le trafic Web sortant, comme les invites ChatGPT, pour empêcher les fuites de données sensibles. Avec cette version, nous introduisons la possibilité d'analyser également le trafic entrant en choisissant la direction du trafic que la DLP en temps réel analyse, c'est-à-dire le trafic entrant, le trafic sortant, ou les deux. Cette fonctionnalité est actuellement disponible uniquement pour ChatGPT (chatbot et API). Le choix d'analyser le trafic entrant analyse les réponses ChatGPT.



23281122679316

Quelle est la catégorie d'application IA générative dans DLP ?

Avant cette version, les critères de destination des règles DLP en temps réel incluaient une liste sélectionnable finie d'environ 20 applications. ± Grâce à cette version, Real-Time DLP permet aux clients de choisir l'une de nos 38 catégories d'applications, y compris l'IA générative, ou l'une des 4 600 applications contrôlables disponibles qui y sont classées. La catégorie des applications d'IA générative, qui a été lancée il y a seulement quelques mois avec 20 applications, compte

maintenant 70 applications, et nous nous engageons à mettre à jour continuellement cette catégorie avec des outils d'IA haut de gamme.

Une règle DLP peut-elle être appliquée à l'ensemble de la catégorie d'applications d'intelligence artificielle générative ?

Oui, une règle DLP en temps réel peut être appliquée à une catégorie entière ou à un sous-ensemble d'applications qu'elle contient.

Où puis-je trouver la documentation associée ?

- Pour savoir comment contrôler la direction d'analyse pour surveiller ou bloquer les réponses ChatGPT, vérifiez :
  Ajouter une règle en temps réel à la stratégie de prévention contre la perte de données
- Pour savoir comment vérifier si une invite de chatGPT ou une réponse de chatGPT a été bloquée, vérifiez le rapport de direction d'analyse ici : Rapport Data Loss Prevention
- Pour consulter toutes les catégories d'applications désormais disponibles dans les règles de stratégie DLP en temps réel, cliquez ici : Catégories d'applications

Envisageons-nous de faire une annonce dans le cadre du prochain Cisco Live Amsterdam concernant ces cas d'utilisation passionnants de protection de l'IA générative ?

Oui, nous allons organiser une session séparée intitulée Protecting Your Sensitive Data from Generative AI Usage à Cisco Live Amsterdam, le mardi 6 février, de 15h00 à 16h30 CET.

Veuillez réserver votre place !

À propos de cette traduction
Cisco a traduit ce document en traduction automatisée vérifiée par une personne dans le cadre d'un service mondial permettant à nos utilisateurs d'obtenir le contenu d'assistance dans leur propre langue.

Il convient cependant de noter que même la meilleure traduction automatisée ne sera pas aussi précise que celle fournie par un traducteur professionnel.