Comprender el nuevo control de contenido de IA generativo y la expansión de la cobertura de herramientas de IA de DLP

Contenido

Introducción

Este documento describe el nuevo control generativo de contenido de IA y la expansión de la cobertura de herramientas de IA de DLP para Umbrella.

Overview

Nos complace anunciar la disponibilidad general del control de contenido de IA generativo. Esta función le permite supervisar y, si es necesario, bloquear el contenido generado por ChatGPT.

También estamos encantados de compartir que hemos ampliado el alcance de nuestra cobertura de DLP en tiempo real para herramientas de IA generativa. Inicialmente limitado a ChatGPT, ahora apoyamos las 70 herramientas de IA en nuestra recientemente lanzada categoría de aplicación de IA generativa. Esta importante expansión le permite ampliar el caso práctico de uso seguro de la IA, ofreciendo una solución más completa y robusta para la protección del uso generativo de la IA.

¿Cómo puede DLP ayudar a controlar el contenido generado por ChatGPT?

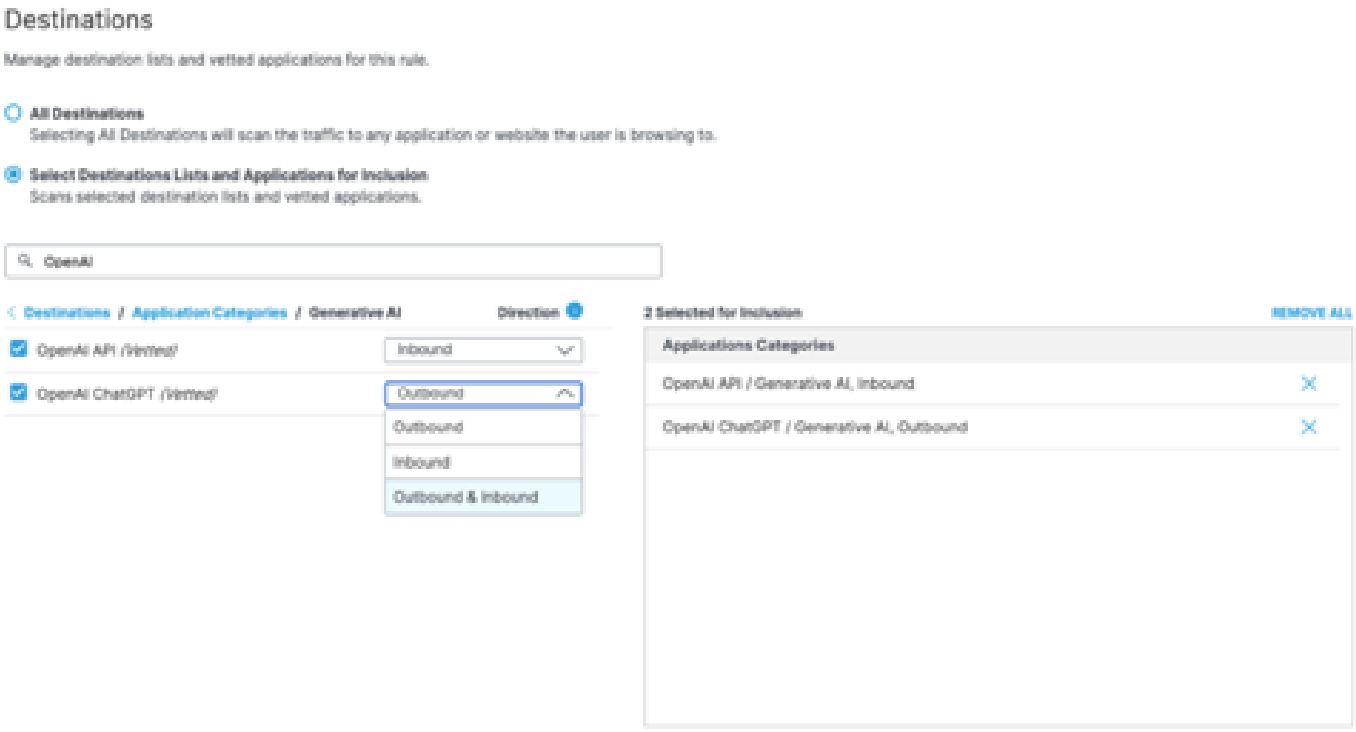
DLP puede ayudar a las organizaciones a controlar el contenido generado mediante el análisis de las respuestas de ChatGPT mediante la política de DLP en tiempo real. Con esta versión, puede optar por analizar las respuestas de ChatGPT (es decir, el tráfico entrante) en busca de cualquier tipo de contenido generado que desee supervisar o bloquear.

¿Por qué controlar el contenido generado por IA?

El uso de contenido generado por IA plantea riesgos para las organizaciones por diversas razones, incluyendo infracción de derechos de autor, información inexacta, código defectuoso, etc.
Por ejemplo, es posible que desee evitar que los usuarios utilicen código fuente generado por IA para evitar el uso de código con derechos de autor o no seguro, mientras que otros podrían querer evitar el uso de citas judiciales generadas por IA por miedo a presentar información inexacta.

¿Cómo puedo aplicar el escaneo de DLP a las respuestas de ChatGPT?

Generalmente, DLP en tiempo real analiza el tráfico web saliente, como los mensajes de ChatGPT, para evitar la filtración de datos confidenciales. Con esta versión, estamos introduciendo la capacidad de analizar también el tráfico entrante al elegir la dirección del tráfico que analiza DLP en tiempo real, es decir, el tráfico entrante, el tráfico saliente o ambos. Esta capacidad está actualmente disponible solo para ChatGPT (chatbot y API). Si elige analizar el tráfico entrante, se analizarán las respuestas de ChatGPT.



23281122679316

¿Cuál es la categoría de aplicaciones de IA generativa en DLP?

Antes de esta versión, los criterios de destino de las reglas de DLP en tiempo real incluían una lista seleccionable limitada de unas 20 aplicaciones. Con esta versión, Real-Time DLP permite a los clientes elegir cualquiera de nuestras 38 categorías de aplicaciones, incluida Generative AI, o cualquiera de las ± 4600 aplicaciones controlables disponibles categorizadas en ellas. La categoría de aplicaciones de IA generativa, que se lanzó hace solo unos meses con 20

aplicaciones, ahora tiene 70 aplicaciones, y estamos comprometidos a actualizar continuamente esta categoría con las herramientas de IA más importantes.

¿Se puede aplicar una regla de DLP a toda la categoría de aplicaciones de IA generativa?

Sí, una regla de DLP en tiempo real se puede aplicar a una categoría completa o a un subconjunto de aplicaciones de la misma.

¿Dónde puedo encontrar documentación relacionada?

- Para aprender a controlar la dirección de escaneo para monitorear o bloquear las respuestas de ChatGPT, marque:
  Agregar una regla en tiempo real a la directiva de prevención de pérdida de datos
- Para saber cómo comprobar si se bloqueó un mensaje de chatGPT o una respuesta de chatGPT, consulte aquí la información de la dirección de análisis: Informe de prevención de pérdida de datos
- Para revisar todas las categorías de aplicaciones que ahora están disponibles en las reglas de políticas de DLP en tiempo real, marque esta opción: Categorías de aplicaciones

¿Pretendemos hacer algún anuncio en el próximo Cisco Live Amsterdam en relación con estos emocionantes casos prácticos de protección de IA generativa?

Sí, vamos a celebrar una sesión introductoria titulada Protección de los datos confidenciales frente al uso generativo de IA en Cisco Live Amsterdam, el martes 6 de febrero de 15:00 a 16:30 CET.

¡Por favor, guarde su asiento!

Acerca de esta traducción
Cisco ha traducido este documento combinando la traducción automática y los recursos humanos a fin de ofrecer a nuestros usuarios en todo el mundo contenido en su propio idioma.

Tenga en cuenta que incluso la mejor traducción automática podría no ser tan precisa como la proporcionada por un traductor profesional.

Cisco Systems, Inc. no asume ninguna responsabilidad por la precisión de estas traducciones y recomienda remitirse siempre al documento original escrito en inglés (insertar vínculo URL).