Resumen de la solución

SkopeAI para ChatGPT e IA generativa

Introducción

La aparición de aplicaciones SaaS basadas en inteligencia artificial (IA) ha supuesto un cambio transformador en la forma en que los usuarios corporativos llevan a cabo su trabajo diario. Las aplicaciones de IA generativa, como ChatGPT, han abierto innumerables posibilidades para que las organizaciones y sus empleados aumenten la productividad empresarial, simplifiquen las tareas, mejoren los servicios y agilicen las operaciones. Con ChatGPT los trabajadores, tanto en equipo como de forma individual, podrán aprovechar cómodamente las capacidades que ofrece para generar contenidos, traducir textos, procesar datos, elaborar planes financieros, y depurar y escribir código, entre otros usos. Las aplicaciones de IA generativa también crean riesgos enormes y sin precedentes para la seguridad de los datos.

El reto de la seguridad de los datos

Aunque las aplicaciones de IA tienen el potencial de mejorar la eficiencia, también introducen nuevos riesgos y exponen los datos confidenciales a amenazas externas. Las organizaciones deben afrontar estos retos para garantizar la confidencialidad, integridad y seguridad de sus datos. A continuación se muestran algunos ejemplos de cómo los datos confidenciales pueden quedar expuestos a ChatGPT y otras aplicaciones de IA basadas en la nube:

- Textos que contienen información de identificación personal (IIP) pueden publicarse y, por tanto, exponerse en el chatbot para solicitar ideas de correo electrónico, respuestas a clientes, cartas personalizadas o análisis de sentimiento.
- Información sanitaria confidencial, incluidos planes de tratamiento individualizados y datos de imágenes médicas, puede introducirse en el chatbot, lo que podría comprometer la privacidad de los pacientes.
- Los desarrolladores de software pueden cargar código fuente con derechos de autor que no se haya publicado para depurarlo, completarlo y mejorar el rendimiento.
- Los desarrolladores de software podrían incluso conectar directamente una aplicación corporativa, que contenga el código fuente o una base de datos, a aplicaciones de IA generativa a través de API. Este movimiento de datos entre aplicaciones permite la sincronización automática de la información en la nube y facilita tareas rutinarias, como refinar la estructura del código y mejorar la legibilidad. Sin embargo, es importante tener en cuenta que dicho acceso podría exponer datos confidenciales a una aplicación de terceros poco segura.
- Archivos de documentos confidenciales de la empresa, como borradores de informes de resultados, documentos de fusiones y adquisiciones, y anuncios previos a la publicación podrían cargarse para someterlos a comprobaciones gramaticales y de redacción, lo que podría crear por negligencia un riesgo de una posible filtración de de datos.
- Datos financieros, incluidas las transacciones corporativas, los ingresos no divulgados, los números de tarjetas de crédito y las calificaciones crediticias de los clientes, pueden ser procesados por ChatGPT para la planificación financiera, el cumplimiento normativo, la detección de fraudes y la incorporación de clientes, sin ninguna medida de seguridad.
- Dentro del departamento de marketing, un empleado podría en el futuro integrar toda la base de datos de clientes en Salesforce.com con ChatGPT y otros complementos que utilicen la IA generativa y muchas otras aplicaciones no autorizadas a través de una integración OAuth. Esta integración entre aplicaciones permitirá a los empleados aprovechar las funciones de GPT, lo que facilitaría la automatización del proceso de redacción de mensajes de correo electrónico a los contactos cuyos contratos estén a punto de vencer. Este es otro ejemplo de movimiento de datos entre aplicaciones que no puede ser detectado por soluciones de red in-line como cortafuegos y puertas de enlace web seguras (SWG).

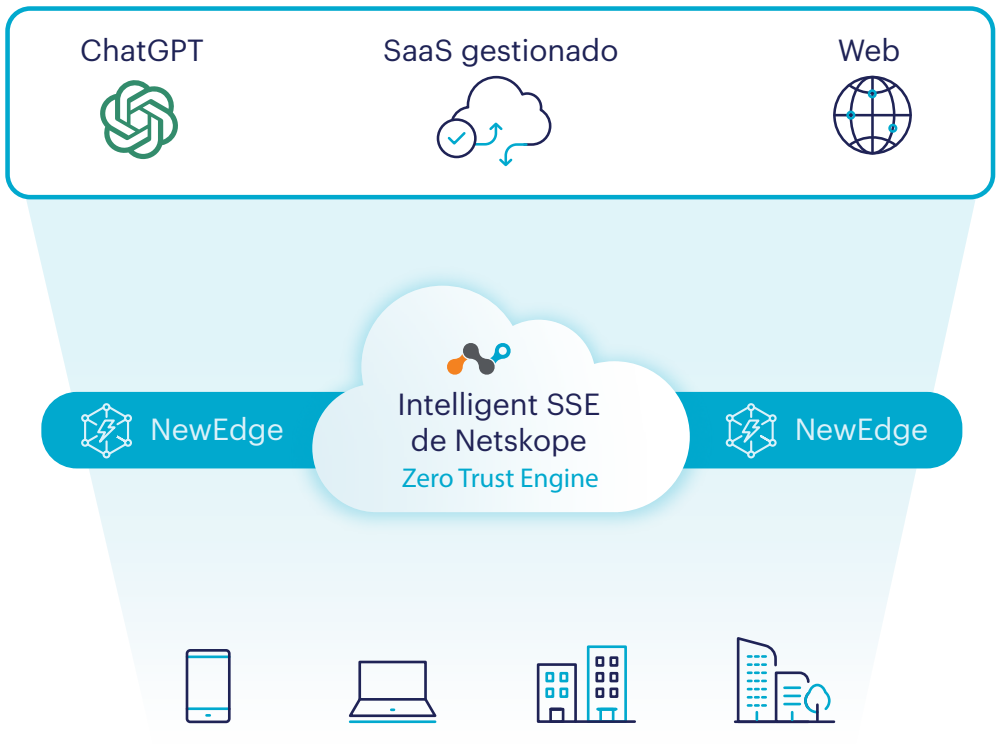
Protección de datos confidenciales en la nube

Las organizaciones deben dar prioridad a unas medidas defensivas sólidas que protejan la confidencialidad y seguridad de los datos confidenciales a través de aplicaciones SaaS gestionadas y no gestionadas, y a través de instancias y cuentas personales. Por ejemplo, según un reciente informe de Netskope sobre las amenazas en la nube, el 74 % de los robos de datos se dirige hacia instancias personales de almacenamiento en la nube de aplicaciones populares.

Los siguientes pasos clave son esenciales para proteger la información confidencial, y deben considerarse funciones básicas de la tecnología moderna de protección de datos:

1. **Supervisión y gestión de riesgos:** Implemente mecanismos de supervisión para hacer un seguimiento del uso y el posible uso indebido de las aplicaciones e instancias SaaS de riesgo. Lleve a cabo evaluaciones de riesgos con regularidad para identificar los puntos vulnerables y resolverlos con prontitud.
2. **Minimización de datos y controles de acceso:** Limite la exposición de la información confidencial a través de aplicaciones SaaS adoptando estrategias de minimización de datos. Aplique controles de acceso estrictos para garantizar que solo las personas autorizadas puedan acceder a los datos confidenciales y manipularlos.
3. **Cifrado y prevención de pérdida de datos:** Incorpore técnicas de cifrado seguras para proteger los datos tanto en reposo como en tránsito. Implemente soluciones de prevención de pérdida de datos (DLP) para controlar y evitar la pérdida o el robo accidental de los datos.
4. **Sensibilización y formación de los usuarios:** Informe a los empleados sobre los riesgos asociados a las aplicaciones SaaS basadas en IA y ofrézcales formación sobre las mejores prácticas para manipular los datos confidenciales de forma segura. Fomente una cultura de protección de datos y haga hincapié en la importancia de un uso responsable.

A medida que las organizaciones adoptan servicios basados en la nube y aplicaciones basadas en la IA, como ChatGPT, garantizar la seguridad de los datos confidenciales se convierte en algo primordial. Mediante la implementación de unas medidas integrales de protección de datos, como la supervisión, los controles de acceso, el cifrado y la formación de los usuarios, las organizaciones pueden mitigar los riesgos, salvaguardar la información confidencial y mantener el cumplimiento de las normativas en el entorno de la nube que está en constante evolución.



Precauciones generales y recomendaciones de seguridad con el uso de aplicaciones de IA generativa

El uso de modelos de IA como ChatGPT en un entorno empresarial ofrece ventajas importantes en términos de productividad, eficiencia e innovación. Sin embargo, garantizar la privacidad y la seguridad de los datos es crucial cuando se utilizan estos modelos de IA. He aquí algunas prácticas óptimas mejoradas para que los equipos de seguridad y los empleados protejan los datos de la empresa:

1. **Implementación local:** siempre que sea posible, implemente los modelos de IA a nivel local en las máquinas de la empresa. Esto elimina la necesidad de que los datos salgan de la red de la empresa, lo que reduce el riesgo de fuga de datos.
2. **Anonimización de los datos:** indique a los usuarios corporativos que dediquen algún tiempo a anonimizar o seudonimizar los datos confidenciales antes de utilizarlos en modelos de IA. Se trata de sustituir los datos de identificación por identificadores artificiales. Incluso en el caso de que se filtraran, los datos serían inútiles sin los identificadores originales.
3. **Cifrado de datos:** siempre que sea posible, aplique el cifrado tanto en reposo como en tránsito para los datos corporativos más confidenciales. Esto garantiza que, aunque los datos queden expuestos, sigan siendo ilegibles si se carece de la clave de descifrado.
4. **Control de acceso estricto:** utilice mecanismos sólidos para el control de acceso a los recursos corporativos y repositorios de datos a fin de restringir la interacción con los modelos de IA y los datos asociados.
5. **Pistas de auditoría:** mantenga registros de auditoría detallados de todas las actividades relacionadas con el tratamiento de datos y las operaciones con modelos de IA. Estos registros ayudan a identificar actividades sospechosas y sirven de referencia para futuras investigaciones.
6. **Minimización de los datos:** Forme a todos los empleados para que respeten el principio de utilizar la cantidad mínima de datos necesaria para el funcionamiento eficaz del modelo de IA. Al limitar la exposición de los datos, se puede reducir el posible impacto de una violación de seguridad de los datos.
7. **Actualizaciones y parches periódicos:** esté alerta para mantener el software local al día con los últimos parches y actualizaciones. Esto protegerá los datos contra vulnerabilidades conocidas.
8. **Auditorías y certificaciones de terceros:** elija servicios de IA de proveedores que se hayan sometido a rigurosas auditorías de terceros y posean certificaciones como ISO 27001, SOC 2 y cumplan el RGPD.
9. **Política de uso de los datos:** establezca políticas claras sobre la manipulación y uso de los datos en su organización. Asegúrese de que los empleados estén bien informados sobre estas políticas y comprendan la importancia de la seguridad de los datos.
10. **Copia de seguridad de los datos:** haga periódicamente copias de seguridad de los datos para garantizar su restauración en caso de que se pierdan o se vean comprometidos.
11. **Revisión constante:** siempre es aconsejable revisar las políticas de uso y las condiciones de servicio más actuales de cualquier herramienta de IA con el objetivo de entender cómo se utilizan los datos enviados a través de la API para mejorar sus modelos.

Cómo protege Netskope los datos confidenciales en el uso de las aplicaciones de IA generativa

Netskope es líder del mercado en seguridad y protección de datos en la nube, con más de una década de experiencia, y ofrece la más amplia visibilidad y el mejor control sobre miles de nuevas aplicaciones SaaS, como ChatGPT. Netskope ofrece SkopeAI para la IA generativa, una solución de seguridad que aborda específicamente el uso de aplicaciones de IA generativa como OpenAI ChatGPT, Bing AI, Google Bard, entre otras muchas. Estas son algunas de las principales funciones tecnológicas que Netskope ofrece a los equipos de seguridad de la información para proteger los datos confidenciales, haciendo hincapié en cómo se aprovechan estas funciones fácilmente para garantizar la seguridad de uso de ChatGPT y otras herramientas de IA generativa:

Control de acceso a las aplicaciones

1. Todo empieza por la visibilidad. Netskope proporciona herramientas automatizadas para que los equipos de seguridad supervisen continuamente a qué aplicaciones (como ChatGPT) intentan acceder los usuarios corporativos, cómo, cuándo, desde dónde, con qué frecuencia, etc. Es fundamental comprender los diferentes niveles de riesgo que cada aplicación supone para la organización y tener la capacidad de definir de forma pormenorizada políticas de control de acceso en tiempo real que se basen en categorizaciones y condiciones de seguridad que pueden cambiar con el tiempo.
    - Por ejemplo, los equipos de seguridad se beneficiarían enormemente si conocieran la amplia gama de aplicaciones que utilizan los empleados de la empresa. Con tantos miles de aplicaciones nuevas disponibles, la capacidad de filtrarlas y clasificarlas por nombre, uso o categoría (como ChatGPT, redes sociales, plataformas de colaboración, repositorios de archivos, etc.) es fundamental. Además, es importante que los equipos de seguridad conozcan el nivel de riesgo, las normas de cumplimiento, las actividades y los detalles de uso de cada aplicación.

2. Aunque deben bloquearse las aplicaciones que son más explícitamente maliciosas, cuando se trata del control de acceso, a menudo la responsabilidad del uso de aplicaciones como ChatGPT debe recaer en los usuarios, tolerando y no necesariamente deteniendo actividades que pueden tener sentido para un subconjunto de grupos de negocio o para la mayoría de ellos. Al mismo tiempo, los equipos de seguridad tienen la responsabilidad de concienciar a los empleados sobre las aplicaciones y actividades que se consideran de riesgo. Esto puede lograrse principalmente mediante alertas en tiempo real y flujos de trabajo automatizados de formación, de forma que el usuario participe en las decisiones de acceso tras reconocer el riesgo. Netskope ofrece opciones flexibles de seguridad para controlar el acceso a aplicaciones SaaS basadas en IA generativa, como ChatGPT, y para proteger los datos confidenciales de forma automática.
    - Entre los ejemplos de políticas de control de acceso se incluyen flujos de trabajo de asesoramiento en tiempo real que se activan cada vez que los usuarios abren ChatGPT, como ventanas emergentes de advertencia personalizables que ofrecen directrices sobre el uso responsable de la aplicación, el posible riesgo asociado y una solicitud de reconocimiento o justificación.

Detección avanzada y protección de los datos confidenciales

Los usuarios cometen errores y pueden poner en peligro los datos confidenciales por negligencia. Aunque se puede conceder acceso a ChatGPT, es primordial limitar la carga y publicación de datos sumamente confidenciales a través de ChatGPT directa e indirectamente y a través de otros vectores de exposición de datos en la nube que podrían ser peligrosos. Esto solo puede lograrse mediante las modernas técnicas de prevención de pérdida de datos (DLP) de Netskope y unos avanzados controles de seguridad en la nube. Con la prevención de pérdida de datos (DLP) de Netskope, basada en modelos de aprendizaje automático (ML) e IA, miles de tipos de archivos, información de identificación personal, propiedad intelectual (PI), registros financieros, y otros datos confidenciales, se identifican con seguridad y se protegen automáticamente de una exposición involuntaria y que no cumpla las normativas. Netskope detecta y protege los datos confidenciales en movimiento, en reposo y en uso y a través de todas las conexiones posibles por parte de los usuarios: en la oficina, en el centro de datos, en casa y durante los viajes.

1. En primer lugar, con la DLP avanzada de Netskope, se pueden identificar automáticamente los flujos de datos confidenciales y categorizar los mensajes sensibles con el máximo nivel de precisión.

   La precisión garantiza que el sistema proteja toda la información confidencial en cualquier formato estructurado y no estructurado, incluidas imágenes, capturas de pantalla, archivos comprimidos, portapapeles, mensajes de chat, etc. También es un aspecto fundamental para garantizar que solo se detecten datos confidenciales, y no consultas inofensivas y tareas seguras a través del chatbot. Esto se consigue automáticamente mediante un amplio conjunto de tecnologías de detección de datos y algoritmos de clasificación avanzados que incluyen tanto reglas de descubrimiento de datos definidas manualmente y motores de detección automatizados como técnicas de aprendizaje profundo, procesamiento del lenguaje natural (NLP), análisis de sentimiento y semántico. El aprendizaje profundo y el NPL aprovechan el aprendizaje automático supervisado y no supervisado para tareas complejas, un enfoque muy parecido al que utilizan los modelos de IA generativa
2. Netskope DLP ofrece una clasificación de imágenes basada en inteligencia artificial (IA) y aprendizaje automático (ML), junto con reconocimiento óptico de caracteres (OCR), lo que permite reconocer automáticamente archivos y tipos de documentos confidenciales en función de múltiples características identificables. Estos modelos también aprovechan los algoritmos de la red neuronal convolucional (CNN) y la IA de visión YOLOv5 para analizar las imágenes visuales. Precisamente, estas técnicas permiten al sistema detectar automáticamente imágenes electrónicas de pasaportes, permisos de conducir, documentos de identidad con fotografía, formularios fiscales, tarjetas médicas, código fuente, tarjetas de la seguridad social, tarjetas de crédito/débito, currículos, acuerdos de confidencialidad, patentes, documentación de fusiones y adquisiciones, y cheques, por mencionar algunos, con mayor precisión y rendimiento, incluso cuando dichas imágenes y documentos están parcialmente estropeados, arrugados, borrosos y, en general, no son nítidos.
3. Netskope DLP también ofrece la posibilidad de crear clasificadores personalizados basados en ML. Con la función «Train Your Own Classifier» (TYOC), basada principalmente en el ML supervisado, las organizaciones entrenan al sistema para que aprenda a identificar nuevos conjuntos de datos únicos en forma de características de ML irreversibles sin información de identificación personal (PII).
4. Es importante garantizar la máxima protección de los documentos con derechos de autor de gran importancia para la empresa, evitando cualquier filtración o duplicación no autorizada. Las técnicas basadas en la huella digital de archivos y documentos pueden emplearse para indexar documentos enteros e identificar copias exactas o parciales de la información que contienen. En particular, Netskope DLP puede examinar incrustaciones semánticas de aprendizaje profundo de secuencias de palabras en los documentos. A continuación, codifica las incrustaciones en vectores numéricos y calcula las similitudes del coseno. Al detectar similitudes en los contenidos de distintos entornos y canales de transmisión, estas técnicas mejoran la capacidad para identificar e impedir la difusión no autorizada.

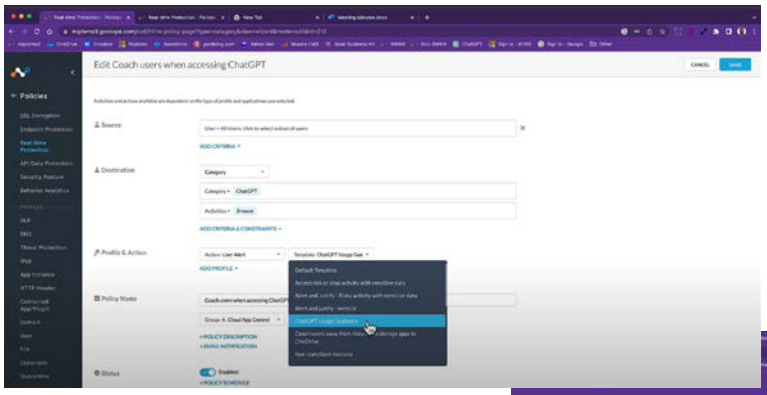
Protección de datos en tiempo real y asesoramiento automática para el usuario

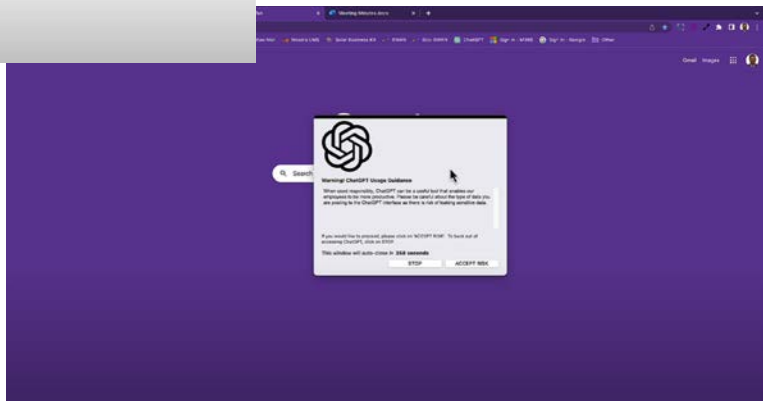
1. Netskope DLP ofrece varios mecanismos para detener y limitar la carga y publicación de datos sumamente confidenciales a través de ChatGPT. Este mecanismo en tiempo real se aplica a todas las conexiones de usuario, lo que garantiza la protección de los datos en el moderno entorno de trabajo híbrido, en el que los usuarios corporativos se conectan desde la oficina, desde casa y durante los viajes. Por ejemplo, en el caso de ChatGPT y otras aplicaciones de IA generativa, además de detener selectivamente las cargas y publicaciones de información confidencial, se pueden automatizar mensajes de asesoramiento visual en tiempo real para ofrecer directrices sobre las infracciones en la publicación de datos, informar al usuario sobre las políticas de seguridad corporativas y para reducir los comportamientos de riesgo repetidos a lo largo del tiempo, lo que reduce la carga de los equipos de respuesta de seguridad.

   Netskope DLP se integra de forma nativa en la exhaustiva solución de servicio de seguridad perimetral (SSE) de Netskope, y conoce continuamente el comportamiento de los usuarios, su geolocalización, las posiciones de seguridad, los riesgos de los dispositivos, los riesgos y reputaciones de las aplicaciones, las instancias de aplicaciones personales, etc. Esto permite a DLP adaptarse automáticamente a los cambios constantes en el contexto de riesgo y adaptar la respuesta de seguridad a diferentes situaciones.

2. Con las aplicaciones de IA generativa, puede que no sea suficiente proteger la carga de datos. Los desarrolladores, por ejemplo, ahora pueden integrar ChatGPT y otros modelos en sus aplicaciones y productos a través de las API, o derivados de ChatGPT (por ej., AutoGPT) en sus flujos de trabajo. Así, un desarrollador podría vincular un código fuente con derechos de propiedad, una base de datos completa en la nube, una hoja Excel 365 en Internet o proporcionar acceso completo a una aplicación. Si solo se supervisan los datos confidenciales en tiempo real a través de cortafuegos y DLP tradicionales, se pasarán por alto estas rutas de salida cuando los datos ya estén en la nube y no estén fluyendo en línea. Netskope ofrece una solución integral de protección de datos para aplicaciones SaaS que descubre y protege los datos confidenciales en línea y en la nube. La solución ayuda a impedir selectivamente que los datos confidenciales se transfieran a la nube, y protege contra el acceso no autorizado entre aplicaciones a los datos confidenciales que ya están en la nube. Además, Netskope proporciona visibilidad de las integraciones de nube a nube para la evaluación y disminución de riesgos.

   A medida que se desarrollan nuevas funciones y ecosistemas de aplicaciones, este enfoque proporciona la mejor y más completa protección de datos para salvaguardar la información confidencial, el código fuente de las aplicaciones de los desarrolladores, las bases de datos de clientes en Salesforce.com y mucho más, limitando o impidiendo la exposición de datos confidenciales a aplicaciones de ecosistemas poco fiables, incluidas las aplicaciones basadas en IA generativa.





Otros controles de Netskope y consideraciones finales

- Netskope también permite usar mecanismos eficaces para el control de acceso basados en los principios de Zero Trust a los repositorios de datos corporativos y limitar así la interacción con los modelos de IA y los datos asociados. Esto reduce significativamente el riesgo de amenazas internas.
- Otra medida de seguridad importante es la capacidad de identificar el comportamiento malicioso de los usuarios y las anomalías de comportamiento de los infractores reincidentes. El análisis de comportamiento de entidades y usuarios (UEBA) de Netskope es un componente de la plataforma de seguridad Netskope que se centra en analizar el comportamiento de usuarios y entidades para detectar y mitigar posibles amenazas a la seguridad. Las soluciones UEBA utilizan análisis avanzados y algoritmos de aprendizaje automático para supervisar las actividades de los usuarios, el tráfico de red y los patrones de acceso a los datos con el fin de identificar comportamientos anómalos o sospechosos. El objetivo específico de UEBA de Netskope es proporcionar información sobre los comportamientos de los usuarios, como su interacción con las aplicaciones en la nube, las transferencias de datos, las actividades de inicio de sesión y los permisos de acceso a los datos. Al analizar estos patrones de comportamiento, UEBA de Netskope ayuda a las organizaciones a identificar amenazas internas, cuentas comprometidas, intentos de exfiltración de datos y otros riesgos de seguridad.
- Además de los casos de uso descritos anteriormente, Netskope ofrece un amplio conjunto de funciones de seguridad basadas en IA y ML en la plataforma, entre las que se incluyen:
  - Modelos avanzados de ML para la detección de malware, que complementan las firmas más tradicionales, los métodos heurísticos y las técnicas de espacios seguros.
  - Filtrado de URL y contra el phishing, con generación automática de firmas URLF, detección DGA, detección rápida de dominios de flujo, filtrado de contenidos web y categorización.
  - Seguridad del IoT, que proporciona clasificación e identificación de los dispositivos IoT, agrupación dinámica de dispositivos y detección de anomalías.
  - Detección de anomalías en el acceso a la WAN.
  - Automatización de flujos de trabajo, supervisión del estado de las aplicaciones, escalado automático en la nube y priorización adaptativa de incidencias

Para más información, visite:

www.netskope.com/skopeai

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

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