Lösungsübersicht

SkopeAI für ChatGPT und generative KI

Einführung

Durch das Aufkommen von KI-gestützten SaaS-Anwendungen hat sich die tägliche Arbeit von Unternehmensbenutzern grundlegend geändert. Anwendungen mit generativer KI wie ChatGPT, haben Organisationen und ihren Mitarbeitern zahllose Möglichkeiten eröffnet, die geschäftliche Produktivität zu erhöhen, Aufgaben zu vereinfachen, Dienstleistungen zu verbessern und Betriebsabläufe zu rationalisieren. Mit ChatGPT können Teams und Einzelpersonen unter anderem Inhalte generieren, Texte übersetzen, Daten verarbeiten, Finanzpläne erstellen oder Code debuggen und schreiben. Anwendungen mit generativer KI bergen jedoch auch enorme, nie dagewesene, Sicherheitsrisiken.

Herausforderungen im Hinblick auf die Datensicherheit

KI-Anwendungen haben nicht nur das Potenzial, die Arbeitseffizienz zu verbessern, sondern bringen auch neue Risiken mit sich und setzen vertrauliche Daten externen Bedrohungen aus. Organisationen müssen diese Herausforderungen bewältigen, um die Vertraulichkeit, Integrität und Sicherheit ihrer Daten zu gewährleisten. Hier einige Beispiele, die zeigen, Beispiele, wie vertrauliche Daten für ChatGPT und andere cloudbasierte KI-Anwendungen preisgegeben werden:

- Text mit personenbezogenen Daten kann in Chatbots eingegeben, und somit auch Risiken ausgesetzt werden, um Ideen für E-Mails, Antworten an Kunden, persönliche Briefe oder Stimmungsanalysen anzufordern.
- Vertrauliche Gesundheitsdaten (wie zum Beispiel individuelle Behandlungspläne und radiologische Daten) werden in Chatbots eingegeben werden und gefährden potenziell die Privatsphäre von Patienten.
- Software-Entwickler laden unter Umständen unveröffentlichten, urheberrechtlich geschützten Quellcode hoch, um Fehler zu beheben, Code fertigzustellen oder die Leistung der Software zu verbessern.
- Software-Entwickler nutzen unter Umständen die Möglichkeit, Unternehmens-Apps, die Quellcode oder Datenbanken enthalten, per API sogar direkt mit Apps zu verbinden, die generative KI nutzen. Diese Datenübertragung von einer App zur nächsten ermöglicht die automatische Synchronisierung von Informationen in der Cloud und erleichtert Routineaufgaben, wie die Verbesserung der Struktur und die Lesbarkeit des Codes. Dabei ist aber zu beachten, dass ein derartiger Zugriff vertrauliche Daten für unsichere Anwendungen von Drittanbietern offenlegen kann.
- Dateien mit vertraulichen Unternehmensunterlagen, wie zum Beispiel Ergebnisberichtsentwürfe, Dokumente zu Fusionen und Übernahmen oder Ankündigungen von Produktveröffentlichungen, werden unter Umständen ohne, Rücksicht auf potenzielle Datenlecks, für Grammatik- und Rechtschreibprüfungen hochgeladen.
- Finanzdaten, wie zum Beispiel Unternehmenstransaktionen, nicht offengelegte, Einnahmen, Kreditkartennummern und Bonitätseinstufungen von Kunden, werden manchmal ohne Sicherheitsmaßnahmen für Finanzplanung, Compliance, Betrugserkennung und Kunden-Onboarding von ChatGPT verarbeitet.
- In der Marketing-Abteilung haben die Mitarbeiter in der nahen Zukunft die Möglichkeit, per OAuth-Integration die gesamte Kundendatenbank in Salesforce.com mit ChatGPT- und anderen Plugins mit generativer KI und vielen anderen unzulässigen Apps zu integrieren. Dank dieser Integration in mehrere Apps können Mitarbeiter die Funktionen von ChatGPT nutzen, um gegebenenfalls automatisierte E-Mails an Kontaktpersonen zu schreiben, deren Verträge bald auslaufen. Dies ist ein weiteres Beispiel für Datenverschiebungen von einer App in eine andere, die von Inline-Netzwerklösungen wie Firewalls und sicheren Web-Gateways (SWG) nicht erkannt werden können.

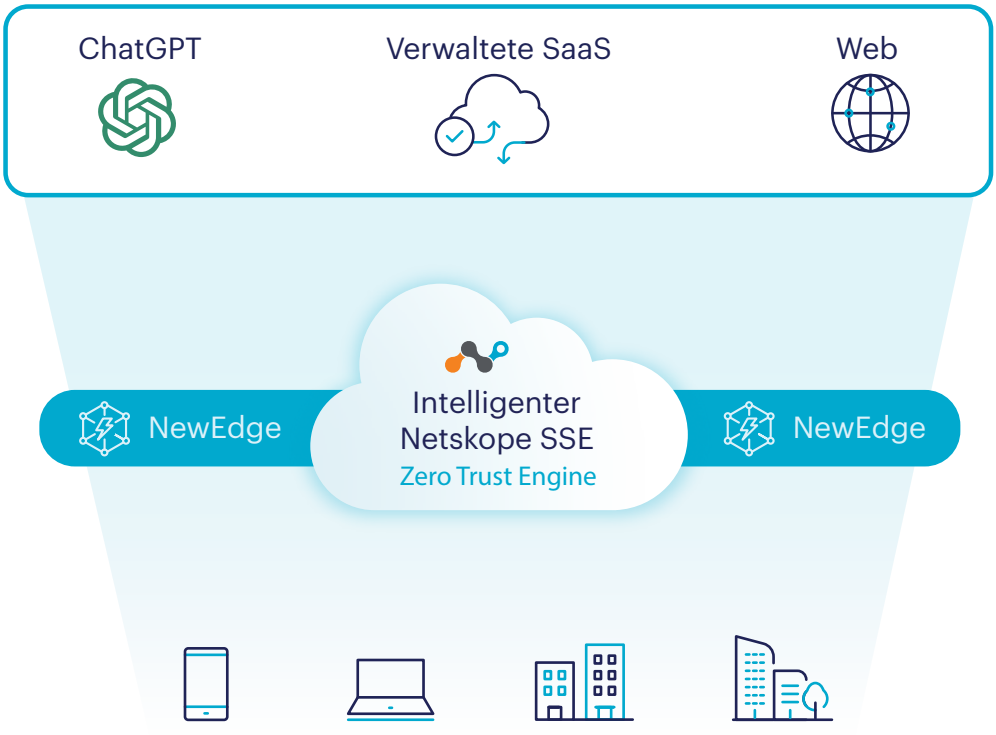
Sichern von vertraulichen Daten in der Cloud

Organisationen müssen wirkungsvolle Gegenmaßnahmen ergreifen, um den Schutz und die Sicherheit vertraulicher Daten in verwalteten und unverwalteten SaaS-Anwendungen, persönlichen Anwendungen und privaten Konten zu gewährleisten. Einem aktuellen Netskope-Bericht zu Cloud-Bedrohungen zufolge, werden 74 % der Datendiebstähle in den Cloudspeicher-Instanzen beliebter Apps begangen.

Die folgenden Maßnahmen sind äußerst wichtig für den Schutz vertraulicher Daten und zentral für moderne Datenschutztechnologien:

1. **Überwachung und Risikomanagement:** Führen Sie Überwachungsmechanismen ein, um die Nutzung und den potenziellen Missbrauch risikobehafteter SaaS-Anwendungen und ihrer Instanzen nachzuverfolgen. Führen Sie regelmäßige Risikobewertungen durch, um Schwachstellen zu erkennen und schnell zu beheben.
2. **Minimierung der Datenmenge und Zugriffskontrolle:** Begrenzen Sie die Offenlegung vertraulicher Daten mit SaaS-Anwendungen durch die Einführung von Strategien zur Minimierung der Datenmenge. Führen Sie strenge Zugriffskontrollen ein, damit nur autorisierte Personen auf vertrauliche Daten zugreifen und diese bearbeiten können.
3. **Verschlüsselung und Data Loss Prevention (DLP):** Schützen Sie Daten im Ruhezustand und bei der Übertragung mit starken Verschlüsselungstechnologien. Überwachen und verhindern Sie durch die Bereitstellung von DLP-Lösungen versehentliche Datenverluste und -diebstähle.
4. **Wachsamkeit und Schulung der Benutzer:** Schulen Sie Benutzer bezüglich der Risiken durch KI-gestützte SaaS-Anwendungen und bringen Sie ihnen die Best Practices für den sicheren Umgang mit vertraulichen Daten bei. Fördern Sie die Datenschutzkultur und betonen Sie, wie wichtig ein verantwortungsvoller Umgang mit den Daten ist.

Da immer mehr Organisationen mit cloudbasierten Services und KI-gestützten Anwendungen wie ChatGPT arbeiten, muss die Sicherheit personenbezogener Daten unbedingt gewährleistet werden. Durch die Einführung umfassender Datenschutzmaßnahmen wie Überwachung, Zugriffskontrolle, Verschlüsselung und Benutzerschulungen, können Organisationen Risiken eindämmen, vertrauliche Daten schützen und in dynamischen Cloud-Umgebungen die gesetzlichen Vorschriften jederzeit einhalten.



Allgemeine Sicherheitsvorkehrungen und -empfehlungen für Anwendungen mit generativer KI

KI-Modelle wie ChatGPT bringen im geschäftlichen Umfeld erhebliche Produktivitäts-, Effizienz- und Innovationsvorteile. Im Umgang mit solchen KI-Modellen haben Datenschutz und Datensicherheit jedoch höchste Priorität. Hier einige Best Practices, für mit dem Schutz der Unternehmensdaten beauftragte, Sicherheitsteams und -mitarbeiter:

1. **Lokale Bereitstellung:** Stellen Sie KI-Modelle, wann immer möglich, auf lokalen Rechnern des Unternehmens bereit. Dadurch verlassen die Daten nicht das Netzwerk des Unternehmens und die Gefahr eines Datenlecks ist geringer.
2. **Anonymisierung der Daten:** Weisen Sie die Benutzer im Unternehmen an, vertrauliche Daten vor der Eingabe in KI-Modelle zu anonymisieren oder mit Pseudonymen zu verfremden. Unter anderem sollten sie personenbezogene Daten durch fiktive Kennungen ersetzen. Sollte es dennoch zu einem Datenschutzverstoß kommen, sind die Datensätze ohne die Originalinformationen wertlos.
3. **Datenverschlüsselung:** Verschlüsseln Sie, wann immer möglich, vertrauliche Unternehmensdaten, sowohl im Ruhezustand als auch bei der Übertragung. Sollten die Daten offengelegt werden, sind sie ohne den Dechiffrierschlüssel unlesbar.
4. **Strenge Zugriffskontrolle:** Nutzen Sie zuverlässige Kontrollmechanismen für den Zugriff auf Ressourcen und Datenspeicher des Unternehmens, um die Interaktion mit KI-Modellen und die dazugehörigen Daten zu beschränken.
5. **Prüfpfade:** Pflegen Sie detaillierte Prüfprotokolle zu allen Aktivitäten rund um die Verarbeitung von Daten und den Betrieb von KI-Modellen. Diese Protokolle ermöglichen die Erkennung verdächtiger Aktivitäten und dienen als Referenz bei späteren Untersuchungen.
6. **Minimierung der Datenmenge:** Alle Mitarbeiter sollten geschult werden, aus Effizienzgründen so viele Daten wie nötig und so wenig wie möglich in das KI-Modell einzugeben. Die Limitierung der Datenmenge verringert die potenziellen Auswirkungen von Datenschutzverstößen.
7. **Regelmäßige Updates und Patches:** Achten Sie darauf, Ihre lokale Software regelmäßig mit den neuesten Patches und Updates zu aktualisieren. Dadurch schützen Sie die Daten des Unternehmens vor bekannten Schwachstellen.
8. **Externe Audits und Zertifizierungen:** Entscheiden Sie sich für KI-Dienstleistungen von Anbietern, die strenge externe Prüfungen durchlaufen haben und deren Leistungen nach von ISO 27001, SOC 2, der DSGVO und anderen Normen zertifiziert sind.
9. **Richtlinien zur Datennutzung:** Legen Sie klare Richtlinien für den Umgang mit Daten und deren Nutzung in Ihrer Organisation fest. Sorgen Sie dafür, dass alle Mitarbeiter mit diesen Richtlinien vertraut und sich der Bedeutung der Datensicherheit bewusst sind.
10. **Datensicherung:** Sichern Sie regelmäßig die Daten, um sie wiederherstellen zu können, falls sie verloren gehen oder gehackt wurden.
11. **Ständige Kontrolle:** Es ist immer ratsam, die aktuellen Nutzungsrichtlinien und Geschäftsbedingungen aller KI-Tools zu kennen, die beschreiben, wie diese mit den über die API gesendeten Daten ihre Modelle verbessern.

So sichert Netskope vertrauliche Daten beim Einsatz von Anwendungen mit generativer KI

Mit mehr als einem Jahrzehnt Erfahrung im Bereich Cloud-Sicherheit und Datenschutz, ist Netskope führender Anbieter von Lösungen, die eine umfassende Transparenz und eine detaillierte Kontrolle über Tausende neuer SaaS-Anwendungen wie ChatGPT bietet. Speziell für Apps mit generativer KI, wie zum Beispiel OpenAI ChatGPT, Bing AI oder Google Bard, bietet Netskope die Sicherheitslösung SkopeAI für GenAI an. Nachfolgend einige der zentralen Technologiemerkmale, die Netskope Informationssicherheitsteams für den Schutz vertraulicher Daten anbieten. Dabei wird der Schwerpunkt auf eine unkomplizierte Absicherung von ChatGPT und anderen Tools mit generativer KI gelegt:

Zugriffskontrolle für Anwendungen

1. Am Anfang steht die Transparenz. Netskope bietet automatisierte Tools, mit denen Sicherheitsteams kontinuierlich überwachen können, auf welche Anwendungen (z. B. ChatGPT) Unternehmensnutzer wie, wann, von wo, mit welcher Häufigkeit usw. zuzugreifen versuchen. Die Teams müssen vor allem nachvollziehen können, welches Risiko jede Anwendung für die Organisation darstellt. Und sie müssen in Echtzeit detaillierte Richtlinien für die Zugriffskontrolle erstellen können, die Kategorisierungen und Sicherheitsbedingungen berücksichtigen, die sich im Laufe der Zeit verändern können.
    - Sicherheitsteams würden zum Beispiel enorm davon profitieren, Einblicke in eine große Bandbreite an Anwendungen zu erhalten, die die Unternehmensnutzer nutzen. Da mittlerweile unzählige neue Apps verfügbar sind, müssen diese auch nach Namen, Nutzung und Kategorie (z. B. ChatGPT, soziale Netzwerke, Datei-Repositories usw.) gefiltert und kategorisiert werden können. Außerdem müssen Sicherheitsteams für jede App das Risiko, die Compliance-Standards, die Aktivitäten und die Nutzungsdaten kennen.

2. Wenngleich Anwendungen , die offensichtlich schädlich sind, gesperrt werden sollten, können die Benutzer bei Anwendungen wie ChatGPT die Verantwortung für die Zugriffskontrolle oft selbst übernehmen. In diesen Fällen sollten Sicherheitsteams die Aktivitäten tolerieren und nicht stoppen, wenn ihr Einsatz für manche oder die meisten Gruppen im Unternehmen sinnvoll ist. Gleichzeitig sind die Teams dafür verantwortlich, die Mitarbeiter über risikobehaftete Anwendungen und Aktivitäten zu informieren. Dies kann hauptsächlich durch Echtzeitwarnungen und automatisierte Coaching-Workflows erreicht werden, bei denen die Benutzer an der Zugriffsentscheidung beteiligt werden, sobald die Gefahr als solche erkannt worden ist. Netskope bietet flexible Sicherheitsoptionen an, um den Zugriff auf SaaS-Anwendungen mit generativer KI (wie zum Beispiel ChatGPT) zu kontrollieren und vertrauliche Daten automatisch zu schützen.
    - Bei diesen Richtlinien für die Zugriffskontrolle kann es sich unter anderem um Workflows für Echtzeit-Coaching handeln, die jedes Mal ausgelöst werden, wenn ein Benutzer ChatGPT öffnet. Das können beispielsweise Pop-up-Fenster mit Warnungen, Hinweise zum verantwortungsvollen Umgang mit der Anwendung und den damit verbundenen Gefahren oder eine Bitte um Bestätigung oder Legitimierung sein.

Erweiterte Erkennung und Absicherung vertraulicher Daten

Manchmal begehen Benutzer Fehler und gefährden aus Nachlässigkeit vertrauliche Daten. Auch wenn der Zugriff auf ChatGPT gestattet wird, müssen das Hochladen und Posten streng vertraulicher Daten über ChatGPT (auf direktem oder indirektem Wege) und andere potenziell risikobehaftete Datenvektoren in der Cloud zwingend beschränkt werden. Dies lässt sich nur durch moderne DLP-Techniken und erweiterte Cloud-Sicherheitskontrollen von Netskope erreichen. Mithilfe der DLP von Netskope, die durch ML- und KI-Modelle gestützt wird, werden Tausende von Dateitypen, personenbezogene Daten, geistiges Eigentum, Finanzdaten und andere vertrauliche Informationen in einer sicheren Umgebung eingelesen und automatisch vor unerwünschter und illegitimer Exponierung geschützt. Netskope erkennt und sichert vertrauliche Daten bei der Übertragung, im Ruhezustand, bei der Nutzung sowie über jede mögliche Benutzerverbindung: im Büro, im Rechenzentrum, zu Hause und unterwegs.

1. Zunächst einmal erkennt die moderne DLP von Netskope automatisch die Übertragung vertraulicher Daten und kategorisiert hochpräzise vertrauliche Daten und Beiträge.

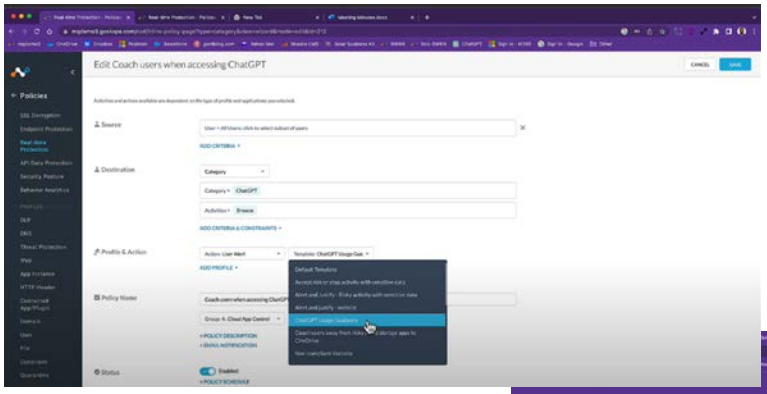
   Diese Genauigkeit garantiert, dass das System jedes vertrauliche Datenelement in strukturierten und unstrukturierten Formaten (z. B. Bilder, Screenshots, komprimierte Dateien, Zwischenablagen, Chat-Nachrichten usw.) schützt. Ein weiterer wichtiger Aspekt besteht darin, dass die Software ausschließlich vertrauliche Daten erkennt und keine harmlosen Anfragen und sicheren Aufgaben über den Chatbot als potenziell gefährdend einordne. Dies geschieht automatisch durch eine umfassende Kombination aus Datenerkennungstechnologien und erweiterten Klassifizierungsalgorithmen. Diese umfassen sowohl manuell definierte Datenerkennungsregeln als auch automatisierte Engines zur Datenerkennung wie Deep-Learning-Techniken, Natural Language Processing (NLP) und Stimmungs- und Sprachanalysen. Deep Learning und NLP nutzen überwachtes und nicht überwachtes maschinelles Lernen für komplexe Aufgaben in ganz ähnlicher Weise wie Modelle mit generativer KI.
2. Die DLP von Netskope klassifiziert Bilder mithilfe von künstlicher Intelligenz (KI) und maschinellem Lernen (ML), beherrscht optische Zeichenerkennung und erkennt vertrauliche Dateien und Dokumenttypen anhand verschiedener eindeutiger Merkmale. Diese Modelle analysieren visuelle Bilder außerdem mithilfe von CNN-Algorithmen (Convolutional Neural Network) und dem KI-Modell YOLOv5 Vision. Genau mit diesen Techniken erkennt das System automatisch elektronische Bilder, unter anderem auf Pässen, Führerscheinen, Lichtbildausweisen, Steuerformularen, Krankenkassenkarten, Quellcode, Sozialversicherungskarten, Kredit-/Debitkarten, Lebensläufen, Geheimhaltungsvereinbarungen, Patenten, Fusions- und Übernahmedokumenten und Schecks mit höchster Genauigkeit und Leistungsfähigkeit – sogar dann, wenn die Bilder zum Teil beschädigt, zerknittert, fleckig oder einfach unscharf sind.
3. Mithilfe der DLP von Netskope können Sie auch benutzerdefinierte ML-gestützte Klassifikatoren entwickeln. Mit der Funktion Train Your Own Classifier (TYOC), die auf überwachtem ML beruht, trainieren Organisationen das System darauf, neue einzigartige Datensätze in Form von PII-freien irreversiblen ML-Funktionen ohne personenbezogene Daten zu erkennen.
4. Dabei muss der Schutz urheberrechtlich geschützter, unternehmenskritischer Dokumente gewährleistet sein und eine nicht autorisierte Exfiltration oder Duplikation verhindert werden. Mit Datei- und Dokumentfingerabdrücken können ganze Dokumente indiziert und exakte oder teilweise identische Kopien von Informationen darin erkannt werden. Die DLP von Netskope kann insbesondere semantische Deep-Learning-Einbettungen von Wortfolgen in den Dokumenten untersuchen. Anschließend werden die Einbettungen als numerische Vektoren verschlüsselt und die Cosinus-Ähnlichkeiten berechnet. Durch die Erkennung von Ähnlichkeiten zwischen Inhalten in verschiedenen Umgebungen und Übertragungskanälen, verbessern diese Techniken die Erkennung und Verhinderung einer nicht autorisierten Verbreitung.

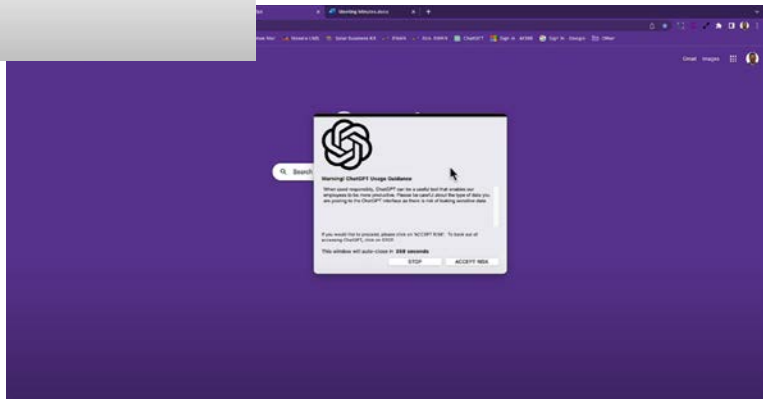
Datenschutz und automatisiertes Benutzer-Coaching in Echtzeit

1. Die DLP von Netskope bietet mehrere Durchsetzungsoptionen, um das Hochladen und Posten hochvertraulicher Daten in ChatGPT zu stoppen und einzuschränken. Die Durchsetzung geschieht in Echtzeit und betrifft jede Benutzerverbindung. Das ist moderner Datenschutz für hybride Arbeitsumgebungen, in denen Unternehmensnutzer im Büro, von zu Hause aus und unterwegs auf das System zugreifen. Im Fall von ChatGPT und anderen Apps mit generativer KI, werden zum Beispiel nicht nur Uploads und das Posten vertraulicher Daten verhindert. Automatisierte Coaching-Nachrichten weisen in Echtzeit auch visuell auf Unregelmäßigkeiten beim Posten von Daten hin, informieren den Benutzer über die Sicherheitsrichtlinien des Unternehmens und unterbinden wiederholt auftretendes risikobehaftetes Verhalten, was den Sicherheitsteams viel Arbeit erspart.

   Die DLP von Netskope ist nativ in die umfassende Lösung Security Service Edge (SSE) integriert und beobachtet kontinuierlich das Benutzerverhalten, den geografischen Standort, den Sicherheitsstatus, Risiken für Geräte und Anwendungen, die Reputation von Anwendungen, die Instanzen von persönlichen Apps usw. Auf diese Weise passt sich die DLP automatisch an dynamische Risikokontexte an und ergreift Sicherheitsmaßnahmen, die perfekt auf die jeweilige Situation abgestimmt sind.
2. Bei Apps mit generativer KI reicht es unter Umständen nicht, den Daten-Upload zu schützen. Entwickler haben jetzt zum Beispiel die Möglichkeit, ChatGPT und andere Modelle per API in ihre Apps und Produkte oder ChatGPT-Derivate (wie AutoGPT) in ihre Workflows zu integrieren. Entwickler könnten dann urheberrechtlich geschützte Quellcodes, ganze Datenbanken in der Cloud und Online-Excel-365-Dokumente verlinken oder Vollzugriff auf eine Anwendung gewähren. Eine reine Inline-Überwachung sensibler Daten durch die traditionellen Firewalls und DLP wird diese Eintrittswege übersehen, wenn sich die Daten bereits in der Cloud befinden und nicht Inline übertragen werden. Netskope bietet eine umfassende Datenschutzlösung für SaaS-Anwendungen an, die vertrauliche Daten sowohl inline als auch in der Cloud erkennen und schützen kann. Die Lösung verhindert selektiv die Übertragung vertraulicher Daten in die Cloud und schützt sie vor nicht autorisiertem Zugriff aus anderen Anwendungen, wenn sie schon in der Cloud sind. Außerdem schafft Netskope Transparenz in Cloud-to-Cloud-Integrationen, damit Risiken abgeschätzt und minimiert werden können.

   Es werden laufend neue Funktionen und App-Ökosysteme entwickelt Dieser Ansatz ermöglicht den besten und umfassendsten Schutz von vertraulichen Daten, Quellcode in Entwickler-Apps, Kundendatenbanken in Salesforce.com und vielem mehr. Er beschränkt oder verhindert die Offenlegung vertraulicher Daten in unsicheren Ökosystem-Apps (zum Beispiel Apps mit generativer KI).





Andere Kontrollmechanismen von Netskope und abschließende Überlegungen

- Netskope ermöglicht auch zuverlässige, auf Zero-Trust-Prinzipien basierende, Kontrollmechanismen für Daten-Repositories des Unternehmens, um Interaktionen mit KI-Modellen und den entsprechenden Datenfluss zu beschränken. Dadurch wird erheblich die Gefahr durch interne Bedrohungen verringert.
- Eine andere wichtige Sicherheitsmaßnahme ist die Erkennung böswilligen Benutzerverhaltens anhand sich wiederholender Verhaltensanomalien. Die integrierte Analyse des Benutzer- und Entitätsverhaltens (User and Entity Behavior Analytics, UEBA), ist eine weitere Komponente der Sicherheitsplattform von Netskope, die sich auf die Analyse des Verhaltens einzelner Benutzer und Entitäten konzentriert, um potenzielle Sicherheitsbedrohungen zu erkennen und abzuwehren. UEBA-Lösungen überwachen Benutzeraktivitäten, Netzwerkdatenverkehr und Datenzugriffsmuster mithilfe von erweiterten Algorithmen für Analysen und maschinelles Lernen, um anomale und verdächtige Verhaltensweisen zu identifizieren. Die UEBA von Netskope ist speziell auf die Gewinnung von Informationen über das Benutzerverhalten ausgelegt und analysiert Interaktionen mit Cloud-Anwendungen, Datenübertragungen, Anmeldeaktivitäten und Datenzugriffsgenehmigungen. Durch die Analyse solcher Verhaltensmuster hilft die Lösung Organisationen dabei, Insider-Bedrohungen, kompromittierte Konten, Exfiltrationsversuche und andere Sicherheitsrisiken zu erkennen.
- Neben den genannten Anwendungsfällen bietet Netskope auf der Plattform auch die folgenden umfassenden KI- und ML-gestützten Sicherheitsfunktionen an:
  - Erweiterte ML-Modelle zur Malware-Erkennung als Ergänzung zu herkömmlichen Signaturen, heuristischen Methoden und Sandboxing;
  - Anti-Phishing- und URL-Filter mit automatisierter Generierung von URL-Signaturen, DGA-Erkennung (Domain Generation Algorithm), Erkennung von Fast-Flux-Domains, Filterung von Webinhalten und Kategorisierung;
  - IoT-Sicherheit mit Klassifizierung und Identifizierung von IoT-Geräten, dynamischer Gerätegruppierung und Anomalie-Erkennung;
  - Erkennung von Anomalien beim WAN-Zugriff;
  - Automatisierung von Workflows, Überwachung der App-Integrität, automatische Cloud-Skalierung und adaptive Incident-Priorisierung.

Weitere Informationen finden Sie unter:

www.netskope.com/skopeai

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

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