netskope

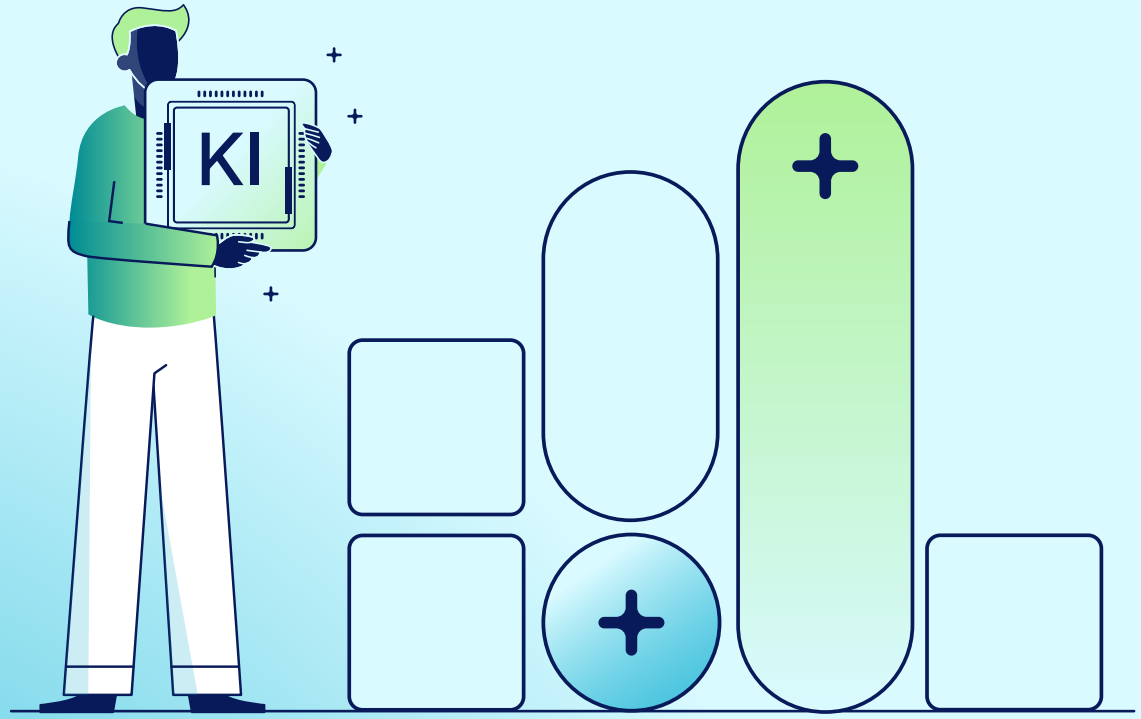
Das

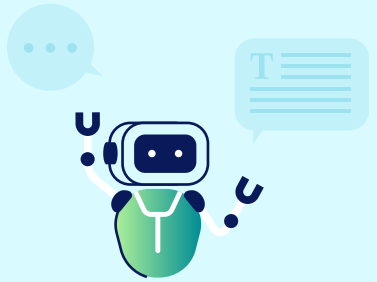
Playbook für KI-Sicherheit

Ein praktischer Leitfaden für KI mit ortsunabhängiger End-to-End-Sicherheit

Das

Playbook für KI-Sicherheit





Inhaltsverzeichnis

Einleitung

Technologien mit künstlicher Intelligenz (KI) haben sich für viele Unternehmen schnell als nützliche und wichtige Hilfsmittel erwiesen. Dank ständig neuer Fähigkeiten und Anwendungsfälle ist die KI inzwischen ein zentraler Bestandteil im IT-Stack der meisten Großunternehmen.

Der rasante Aufstieg der KI ging auch mit hohen Investitionen einher. IDC-Analysten zufolge werden die IT-Ausgaben für KI bis 2028 weltweit auf fast 750 Mrd. USD anwachsen. Davon werden etwas mehr als 300 Mrd. USD auf Investitionen in generative KI entfallen.[1]

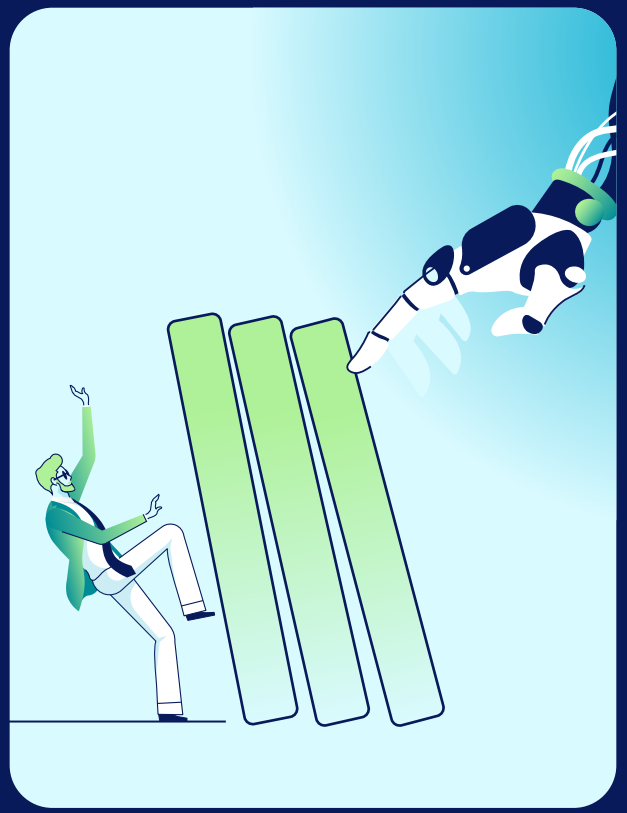
Die IT-Ausgaben für KI werden bis 2028 voraussichtlich weltweit auf fast 750 Mrd. USD anwachsen. Davon werden etwas mehr als 300 Mrd. USD auf Investitionen in generative KI entfallen.

Für Sicherheitstechniker sind die potenziellen Gefahren, die von KI-Anwendungen in ihren IT-Umgebungen ausgehen, offenkundig – und sie nehmen ständig zu. Selbst bei der einfachsten Form der KI-Nutzung werden Daten mit Drittanbieter-Anwendungen in der Cloud ausgetauscht. Hinsichtlich der Sicherheit wirft dies Fragen darüber auf, welche Daten von den Mitarbeitern in diese Systeme eingegeben werden und welche Kontrollmechanismen zu hrer Verwaltung bestehen. Dank der Weiterentwicklung von Standardprotokollen, die den Datenaustausch mit KI-Anwendungen noch einfacher machen, darunter das Model Context Protocol (MCP), lassen sich diese Risiken systematisch erfassen.[2]

Im Zuge der Weiterentwicklung von KI-Technologien werden die Herausforderungen für Großunternehmen künftig noch zunehmen. So können beispielsweise agentenbasierte KI-Systeme autonom auf bestimmte Ziele hinarbeiten oder festgelegte Aufgaben ohne ständige menschliche Eingriffe ausführen. Branchenanalysten von Gartner prognostizieren, dass bis zum Jahr 2028 der Missbrauch von KI-Agenten für 25 % aller Datenschutzverstöße in Großunternehmen verantwortlich sein wird.[3]

Angesichts dieser rasanten Entwicklung ist es kein Wunder, dass IT-Sicherheitstechniker nach Hilfestellungen im Umgang mit diesen neuen Herausforderungen suchen. In diesem E-Book beschreiben wir die wichtigsten Sicherheitsprobleme, vor denen Unternehmen heute stehen, und die Lösungen, die Netskope dafür anbietet.

Gartner prognostiziert, dass bis zum Jahr 2028 der Missbrauch von KI-Agenten für 25 % aller Datenschutzverstöße in Großunternehmen verantwortlich sein wird.



[1] IDC Market Forecast, Worldwide Artificial Intelligence IT Spending Forecast, 2024–2028, Rick Villars et al., Oktober 2024, Dok. US52635424.

[2] Netskope, Cloud and Threat Report 2025 https://www.netskope.com/netskope-threat-labs/cloud-threat-report/cloud-and-threat-report-2025

[3] Gartner's Top Predictions for 2025.

Sicherheitsprobleme bei KI

Die drei größten Probleme, vor denen Sicherheitsteams heutzutage stehen

1 Größere Angriffsfläche

Da die KI-Nutzung sich von Tools für den spielerischen Umgang mit KI (z. B. ChatGPT) zum integrierten Einsatz von KI in Enterprise-Apps und privat entwickelten KI-Anwendungen verlagert, vergrößert sich die Angriffsfläche immer weiter. Auf jeder Stufe drohen neue Gefahren:

- Bei der Verwendung öffentlicher GenAI-Tools besteht das Risiko einer unbeabsichtigten Offenlegung sensibler Daten.
- In bestehende SaaS-Anwendungen integrierte KI-Funktionen können Datenlecks und Datenmanipulation Tür und Tor öffnen.
- Privat gehostete LLMs und individuell angepasste KI-Anwendungen führen neue Bedrohungsvektoren ein, z. B. falsch konfigurierte Zugriffskontrollen oder Schwachstellen in Daten-Pipelines.
- Verbindungen zwischen KI-Anwendungen und Datenquellen über neue Protokolle wie MCP vergrößern die Angriffsfläche für Datenexfiltrationen.

2 Offenlegung und Exfiltration vertraulicher Daten

Die unmittelbarste Gefahr bei der Einführung von KI ist versehentlicher oder absichtlich herbeigeführter Datenverlust:

- Zu einer unbeabsichtigten Offenlegung kommt es, wenn Mitarbeiter vertrauliche Daten (z. B. personenbezogene Informationen, Geschäftsgeheimnisse, regulierte Daten) in öffentliche KI-Modelle eingeben, ohne sich der Folgen bewusst zu sein.
- Insider mit bösen Absichten oder externe Angreifer können mit KI-Tools Daten exfiltrieren oder die Ausgabekanäle des Modells missbrauchen.
- Auch das Training birgt Gefahren: Wenn Modelle mit unsachgemäß kuratierten Daten trainiert werden, legen sie unter Umständen vertrauliche Informationen offen.

3 Verantwortungsbewusste KI-Governance

Die Skalierung von KI-Systemen wirft kritische Compliance-Fragen und ethische Probleme mit Sicherheitsauswirkungen auf:

- KI-Modelle können unbeabsichtigt menschliche Vorurteile reproduzieren und verbreiten, was behördliche Kontrollen nach sich ziehen und den Ruf des Unternehmens schädigen kann.
- Der unsachgemäße Umgang mit den in KI-Workflows verwendeten Mitarbeiter- und Kundendaten kann gegen die DSGVO, den HIPAA und andere Datenschutzgesetze verstoßen.
- Das Ersetzen menschlicher Entscheidungsträger durch autonome KI kann (insbesondere auf wichtigen Gebieten wie Personal-, Sicherheits- und Finanzfragen) zu ethischen Dilemmasituationen und Zuständigkeitslücken führen.

Grundlagen der KI-Sicherheit

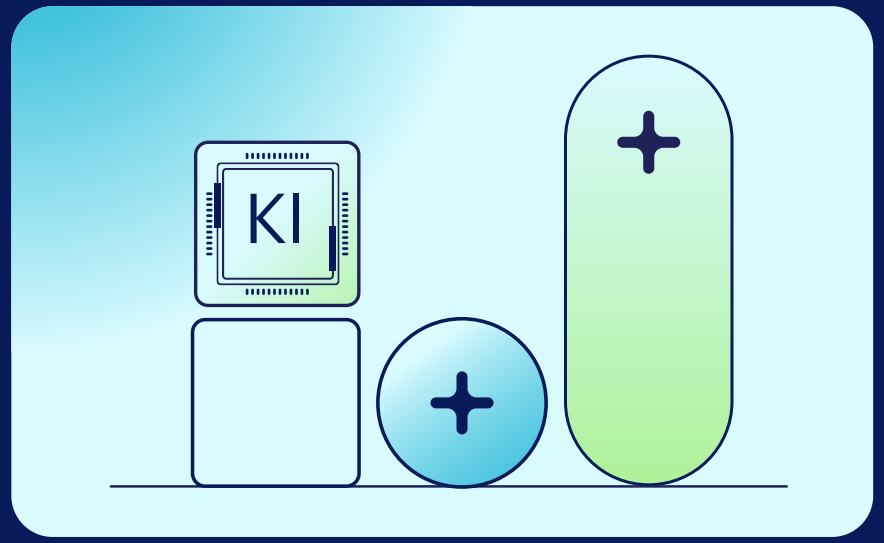
Zero Trust als Gebot der Stunde

Ähnlich wie SaaS-Sicherheit beruht auch KI-Sicherheit auf dem Zero-Trust-Prinzip. Durch die besondere Art, wie KI-Modelle Eingaben verarbeiten und Ausgaben generieren, bringt sie aber ganz besondere Herausforderungen mit sich.

Sowohl die KI- als auch die SaaS-Sicherheit erfordern zum Abwenden von Risiken strikte Zugriffskontrollen, eine kontinuierliche Überwachung und zuverlässigen Datenschutz. Doch während es bei der SaaS-Sicherheit vor allem um den Schutz der Anwendungen und Benutzerinteraktionen geht, muss KI-Sicherheit auch die Integrität der Trainingsdaten, die Zugriffsrechte für das Modell und potenzielle böswillige Manipulationen berücksichtigen. Dies erfordert kontextsensible Sicherheitsrichtlinien und Bedrohungserkennung in Echtzeit. Nur so lassen sich Datenlecks, unautorisierter Zugriff und der Missbrauch von KI-Modellen verhindern.

Ein starkes Zero-Trust-Framework für die KI-Sicherheit sorgt dafür, dass jede Anfrage verifiziert, jeder Daten-Flow überwacht und der Zugriff anhand dynamischer Risikobewertungen anstelle statischer Berechtigungen gewährt wird. Diese Methode erfordert eine granulare Sichtbarkeit der Datenbewegungen und adaptive Sicherheitskontrollen, die sich in Echtzeit an den Kontext anpassen.

Wenn die Zero-Trust-Grundsätze beachtet werden, können Unternehmen KI-gestützte Technologien gefahrlos einführen und skalieren, ohne die Sicherheit und Compliance zu gefährden.



Profi-Tipp

Ähnlich wie die SaaS-Sicherheit beruht auch die KI-Sicherheit auf dem Zero-Trust-Prinzip. Durch die besondere Art, wie KI-Modelle Eingaben verarbeiten und Ausgaben generieren, bringt sie aber ganz eigene Herausforderungen mit sich.

Der sichere Umgang mit KI

Die sechs größten Probleme und ihre Lösungen

Problem Nr. 1: Mangelnde Transparenz

Da KI-Tools in tägliche Arbeitsabläufe eingebettet werden, haben Unternehmen ein grundsätzliches Sicherheitsproblem: Sie können nicht schützen, was sie nicht sehen.

Die Beschäftigten greifen auf offizielle und inoffizielle Anwendungen sowohl mit firmeneigenen als auch mit persönlichen Anmeldedaten zu. So verschwimmt die Grenze zwischen genehmigter und nicht genehmigter Nutzung. Durch diese unkontrollierte Ausbreitung steigt das Risiko von Datenlecks, IP-Verlusten und Compliance-Verstößen, besonders dann, wenn vertrauliche Daten in nicht verwaltete KI-Services oder Schatten-KI eingegeben werden.

In den meisten Unternehmen fehlt die granulare Transparenz, um zwischen riskanter und legitimer KI-Nutzung zu unterscheiden. Herkömmliche Tools können oft weder spezifische Interaktionen von KI-Modellen erkennen, noch persönliche von firmeneigenen Accounts unterscheiden oder für Echtzeittransparenz auf Benutzer-, Anwendungs- oder Aktivitätenebene sorgen. Ohne tiefgreifende Einblicke, wie und wo die KI genutzt wird, sind Sicherheitsteams blind für die potenziellen Schwachstellen.

Welche Lösungen Netskope dafür anbietet

Bei der zunehmenden Einführung von KI-Tools in Unternehmen kommt es auf die Transparenz und die Kontrolle über ihre Nutzung an. Netskope bietet eine umfassende Lösung zur Nachverfolgung sowohl verwalteter als auch nicht verwalteter Anwendungen („Schatten-KI“), die Sicherheitsteams die nötigen Einblicke für die Ausübung ihrer Kontrollfunktion gewährt.

Die Hauptfunktionen umfassen:

- **Besondere Berücksichtigung der Instanz:** Unterscheidung zwischen persönlichen und firmeneigenen Instanzen von KI-Anwendungen wie ChatGPT, Gemini und Copilot.
- **KI-Dashboard**: Sie erhalten tiefe Einblicke in Trends bei der KI-Nutzung, die Beliebtheitsrangliste der Anwendungen, die Zugriffshäufigkeit und bestimmte Benutzeraktionen wie Anmelden, Posten, Hoch- und Herunterladen.
- **User and Entity Behavior Analytics (UEBA):** Erkennen Sie mit maschinellem Lernen Anomalien und risikobehaftete Verhaltensweisen, um Bedrohungen wie Datenexfiltration, Insiderrisiken und Richtlinienverstößen zu begegnen.
- **Grundlegende Transparenz:** Sie erlangen Transparenz über Ihr gesamtes KI-Ökosystem – vom Datenverkehr zwischen Benutzern und Apps bis zum API- und MCP-Traffic. Netskope bietet einen ganzheitlichen Überblick über Nutzungs- und Bestandsdaten sowie Datenflüsse.

Durch diese ganzheitliche Transparenz können Sicherheitsteams schnell reagieren und im ganzen Unternehmen Risiken durch KI-Nutzung eindämmen.

Problem Nr. 2:
Die Risiken von KI-Anwendungen

Nicht nur die Möglichkeiten der KI entwickeln sich rasant, sondern auch die Bedrohungslandschaft. Was einmal eine einfache SaaS-Anwendung war, kann jetzt ohne Weiteres integrierte KI-Funktionen wie Texterstellung, intelligente Antworten und KI-Copiloten einführen, ohne Benutzer und Sicherheitsteams zu benachrichtigen. Durch diesen Trend wird es immer schwieriger zu verstehen, welche Anwendungen KI nutzen, wie diese Nutzung erfolgt und welche Risiken damit für das Unternehmen verbunden sind.

Sicherheitsteams müssen Risiken dynamisch bewerten können und dabei berücksichtigen, wie KI-Funktionen integriert wurden, ob sie Unternehmensdaten aufbewahren, ob sie damit trainiert werden und inwiefern die Compliance-Anforderungen beachtet wurden. Ohne dieses Wissen drohen Unternehmen Risiken wie Datenlecks, die Entwendung geistigen Eigentums, Verstöße gegen behördliche Vorschriften oder sogar die Manipulation ihrer KI-Modelle. Da der KI-Anteil in SaaS immer mehr zunimmt, ist die Kenntnis der Anwendungsrisiken nicht nur eine Best Practice, sondern eine Notwendigkeit für jedes Unternehmen, das bei der KI-Einführung auf Sicherheit achtet.

Welche Lösungen Netskope dafür anbietet

Netskope begegnet der zunehmenden Komplexität der mit KI-Anwendungen verbundenen Risiken mit seinem Cloud Confidence Index (CCI), der kontinuierlich aktualisierte Echtzeit-Einblicke in mehr als 85.000 Cloud- und SaaS-Anwendungen bietet. Mit dynamischen, KI-sensiblen Risikobewertungen hilft CCI Sicherheitsteams, potenziellen Risiken vorzubeugen und für Compliance zu sorgen.

Die Hauptfunktionen umfassen:

- **KI-sensible Risikobewertung in Echtzeit:** Ermitteln Sie Anwendungen mit integrierten KI-Funktionen und finden Sie heraus, mit welchen Risiken sie verbunden sind.
- **Einblicke in den Umgang mit Unternehmensdaten:** Evaluieren Sie, wie Anwendungen Unternehmensdaten verwalten, mit Informationen zur Aufbewahrung der Daten, zum Trainieren der Modelle und zur Freigabe für externe Apps.
- **Compliance-Nachverfolgung:** Bleiben Sie auf dem aktuellen Stand behördlicher Anforderungen wie DSGVO, SOC 2 und ISO 27001.
- **Sichere LLMs und MCP:** Bewerten Sie über 85.000 SaaS-Anwendungen, darunter KI-Anwendungen und eingebettete KI-Funktionen sowie öffentliche MCP-Server, und identifizieren Sie dabei risikobehaftete Merkmale, Authentifizierungstypen und Protokollversionen.

Mit CCI können Sicherheitsteams zuverlässig die komplexen Risiken von KI-Anwendungen bewältigen und in ihren Unternehmen für Sicherheit und Compliance sorgen.

Problem Nr. 3: Integrität der KI-Modelle

Da Unternehmen immer mehr Tools mit generativer KI nutzen (sowohl individuelle Modelle als auch Unternehmensanwendungen wie Microsoft Copilot), kommt der Integrität der Trainingsdaten für diese Modelle höchste Bedeutung zu. Solche KI-Modelle werden oft mit gewaltigen Datensätzen trainiert, die vertrauliche Unternehmensunterlagen, E-Mails, Präsentationen, Tabellen und urheberrechtlich geschützte Informationen enthalten.

Wenn aus Versehen vertrauliche oder urheberrechtlich geschützte Daten in die Trainingsdatensätze aufgenommen werden, können sie nicht nur über die Ausgaben der Modelle, sondern auch durch böswillige Prompts, Datenlecks und potenzielle Compliance-Verstöße offengelegt werden. Da GenAI in immer mehr Unternehmensbereichen eingeführt wird, haben Sicherheitsteams zunehmend Schwierigkeiten damit, Beschaffung, Validierung und Schutz der Trainingsdaten zu kontrollieren.

+ Microsoft Copilot kann mit den Inhalten der Office-Suite eines Benutzers trainiert werden, von Word-Dokumenten bis hin zu Excel-Tabellen. Wenn dort vertrauliche Daten gespeichert sind und die Zugriffskontrollen nicht entsprechend konfiguriert wurden, legen die Antworten von Copilot möglicherweise vertrauliche Unternehmensstrategien, Finanz- oder Kundendaten offen.

Welche Lösungen Netskope dafür anbietet

Mit Netskope One DSPM (Data Security Posture Management) können Unternehmen vertrauliche Daten sowohl in Cloud-Umgebungen als auch in Daten-Repositorys überwachen und schützen. Durch die Überwachung und Klassifizierung kritischer Informationen wie Finanzdaten, personenbezogener Daten und geistigen Eigentums verhindert Netskope, dass sie ohne entsprechende Genehmigung beim KI-Training zum Einsatz kommen.

Die Hauptfunktionen umfassen:

- **Kontinuierliche Überwachung von Cloud-Umgebungen:** Erkennen und klassifizieren Sie vertrauliche Daten in Echtzeit, um ihre unbefugte Nutzung für das Training von KI-Modellen zu verhindern.
- **Transparenz von Datenzugriff und -freigabe:** Erhalten Sie Echtzeiteinblicke in den Zugriff auf Daten und ihre Verbreitung in der Cloud, um bei Bedarf sofort Gegenmaßnahmen zu ergreifen.
- **Compliance und Verhinderung von Datenlecks:** Schützen Sie sensible Daten, um Compliance-Anforderungen zu erfüllen, Datenlecks zu verhindern und die Kontrolle über geistiges Eigentum zu behalten.
- **Robustes Security Posture Management:** Gewährleisten Sie einen soliden Datensicherheitsstatus und identifizieren, kennzeichnen und klassifizieren Sie Ihre strukturierten und unstrukturierten Daten.

Mit Netskope One DSPM können Unternehmen proaktiv ihre vertraulichen Daten schützen und beim Training ihrer KI-Modelle für Sicherheit, Compliance und Kontrolle sorgen.

Problem Nr. 4: Gegen KI-Systeme gerichtete Bedrohungen

Angreifer entwickeln neue Taktiken, um KI-spezifische Schwachstellen auszunutzen. Mit Prompt Injection, Datenvergiftung und böswilligen Eingaben verfälschen sie die Ergebnisse oder exfiltrieren vertrauliche Daten. Außerdem werden KI-Anwendungen oft in allgemeine Unternehmenssysteme integriert und sind dann ein potenzielles Einfallstor für laterale Bewegungen, Rechteausweitung und Datendiebstahl.

Die Angriffsflächen wachsen ständig: Böswillige Akteure versuchen, die Ausgaben eines KI-Modells zu manipulieren, Trainingsdaten zu extrahieren und schwache Zugriffskontrollen von KI-APIs auszunutzen. Erschwerend kommt hinzu, dass es keine standardmäßigen Sicherheits-Frameworks für den Schutz von KI-Systemen gibt. Dadurch sind viele Unternehmen nicht auf den Kampf gegen neue Angriffsvektoren vorbereitet. Durch die fortschreitende Einführung von KI sind Sicherheitsteams mehr und mehr auf eine proaktive Erkennung und Bekämpfung von Bedrohungen angewiesen, die sich speziell gegen KI-Umgebungen richten, bevor diese vertrauliche Daten, Betriebsabläufe und Entscheidungsprozesse kompromittieren.

Welche Lösungen Netskope dafür anbietet

Netskope bekämpft die zunehmenden Bedrohungen, denen KI-Systeme ausgesetzt sind, mit einem mehrstufigen Sicherheitsansatz, der hochentwickelten Bedrohungsschutz, umfassende Transparenz und KI-spezifische Abwehrmaßnahmen vereint.

Die Hauptfunktionen umfassen:

- **Einheitlicher Schutz vor KI-Bedrohungen:** Netskope One AI Guardrails wehrt komplexe Angriffe, wie Prompt Injections und Jailbreak-Versuche, durch eine umfassende Echtzeitanalyse des gesamten Datenverkehrs ab.
- **Hochentwickelter Bedrohungsschutz:** Mit maschinellem Lernen, Sandboxing und heuristischen Analysen können Sie sowohl bekannte Bedrohungen als auch Zero-Day-Angriffe erkennen und blockieren. Hierzu gehört auch verborgene Malware in Dateien, die an KI-Tools übermittelt werden.
- **Angriffssimulationen und Schwachstellenevaluierung:** Automatisieren Sie Angriffssimulationen mit Netskope One Red Teaming, um Schwachstellen aufzudecken, damit Ihre privaten Modelle sicher, regelkonform und vor komplexen Bedrohungen geschützt sind.
- **Proaktive Überwachung von KI-Aktivitäten:** Durch die Echtzeitüberwachung aller KI-Interaktionen können neue Bedrohungen und Schwachstellen schnell erkannt werden, was die Entwicklung einer umfassenden Abwehrstrategie ermöglicht.

Netskope kombiniert all diese Technologien zu einer integrierten Lösung, mit der Unternehmen ihre KI-Systeme vor raffinierten Cyberangriffen und neuen Bedrohungsvektoren schützen können.

Problem Nr. 5: Offenlegung von Daten

Eines der dringendsten und potenziell kostspieligsten Probleme bei der KI-Sicherheit ist die Offenlegung von Daten. Wenn Beschäftigte in allen Abteilungen KI-Tools einführen, um ihre Produktivität zu steigern, können sie versehentlich vertrauliche Daten, wie zum Beispiel Quellcode, Kundendaten, Finanzdokumente oder urheberrechtlich geschütztes geistiges Eigentum, hochladen oder mit öffentlichen KI-Modellen teilen. Danach werden diese Daten möglicherweise gespeichert, zum Trainieren von Modellen verwendet oder sogar offengelegt, je nach Datenschutzrichtlinien und Datenverarbeitungsverfahren der Anwendung.

Anders als herkömmliche Kanäle zum Teilen von Daten können KI-Plattformen zu einer Blackbox werden, die kaum Einblicke in Datenspeicherung, -zugriff und -nutzung gewährt. Ohne wirksame Schutzmaßnahmen drohen Unternehmen ernsthafte Risiken – von Verstößen gegen behördliche Vorschriften und IP-Diebstählen bis hin zu Imageschäden und Wettbewerbsnachteilen.

+ Die Netskope Threat Labs haben bei fast 50 % aller Richtlinienverstöße mit KI-Bezug festgestellt, dass Quellcode offengelegt wurde. Dies unterstreicht, wie leicht kritische Wirtschaftsgüter durch wohlmeinende Handlungen wie das Kopieren eines Codeschnipsels in einen KI-Chatbot (um ihn zu debuggen oder zu optimieren) kompromittiert werden können.

Welche Lösungen Netskope dafür anbietet

Netskope bietet umfassenden, kontextbezogenen Schutz für Unternehmensdaten, sowohl im Ruhezustand als auch bei der Übertragung. Die einheitlichen Sicherheitsrichtlinien von Netskope kombinieren Risikobewertungen in Echtzeit, Inline- und API-Kontrollen und Prüfungen des Sicherheitsstatus. Sowohl Benutzer- als auch Dateninteraktionen können unternehmensweit präzise beaufsichtigt werden.

Die Hauptfunktionen umfassen:

- **Erweiterte Data Loss Prevention (DLP):** Schützen Sie vertrauliche Daten vor der Exfiltration mit KI-Tools – ganz gleich, ob die Benutzer im Büro, zu Hause oder unterwegs sind.
- **Granulare Kontrolle:** Blockieren oder beschränken Sie risikoreiche Aktionen wie das Hochladen von Quellcode oder vertraulichen Dokumenten.
- **Benutzer-Coaching in Echtzeit:** Informieren Sie Benutzer mithilfe visueller Hinweise über Richtlinienverstöße, um erneute Zuwiderhandlungen zu verhindern.
- **Überprüfen Sie jede Anfrage und Antwort:** Erkennen und verhindern Sie die Weitergabe patentierter oder urheberrechtlich geschützter Informationen in KI-Antworten, um sich proaktiv vor IP-Risiken im Zusammenhang mit den Ausgaben generativer Modelle zu schützen.
- **Sicherer API-Datenverkehr:** Authentifizieren und zentralisieren Sie die Verwaltung des Datenverkehrs und die Inhaltsprüfung zwischen privaten Anwendungen und LLMs.

Mit diesen Ressourcen sorgt Netskope für umfassenden, adaptiven Datenschutz, der sich in der gesamten KI- und Cloud-Umgebung eines Unternehmens skalieren lässt.

Problem Nr. 6: Governance, Compliance und ethische Nutzung

Da sich die KI-Einführung immer mehr beschleunigt, stehen Unternehmen unter zunehmendem Druck, Governance-Standards, behördliche Vorschriften und ethische Erwartungen einzuhalten – besonders in stark regulierten Branchen wie Finanzen, Gesundheit und Regierung. Länder in aller Welt sind dabei, rasch KI-spezifische Frameworks und Vorschriften einzuführen, wie sie im EU AI Act, dem NIST AI Risk Management Framework und den US-Durchführungsverordnungen zur KI-Sicherheit dargelegt werden. Diese Vorschriften sollen für eine verantwortungsvolle Entwicklung und Bereitstellung von KI-Systemen sorgen und sehen Transparenz, Datenschutz, Erklärbarkeit und Diskriminierungsfreiheit vor.

Die Einhaltung dieser Standards gestaltet sich allerdings alles andere als einfach. Sicherheits- und Compliance-Teams müssen die KI-Nutzung in ihrer Umgebung kennen, verhindern, dass vertrauliche Daten unberechtigterweise gespeichert oder als Trainingsmaterial verwendet werden, und die Einhaltung neuer rechtlicher und ethischer Richtlinien nachweisen.

Welche Lösungen Netskope dafür anbietet

Netskope sorgt durch tiefgreifende Einblicke, Richtlinienkontrollen und Echtzeittransparenz der KI-Nutzung im Unternehmen für KI-Governance und Compliance-Bereitschaft.

Die Hauptfunktionen umfassen:

- **Granulare Richtliniendurchsetzung:** Kontrollieren Sie die Weitergabe von Daten an KI-Tools, damit Modelle von Dritten nicht ohne Genehmigung mit vertraulichen oder regulierten Daten trainiert werden.
- **Compliance-Kontrolle in Echtzeit:** Blockieren Sie das Hochladen von gesetzlich geschützten Gesundheitsdaten in nicht konforme Apps, und stoppen Sie die Verarbeitung von Finanzdaten in Tools, die nicht entsprechend zertifiziert wurden.
- **Unterstützung bei regulatorischem Framework:** Frameworks wie das KI-Gesetz der EU und NIST AI RMF erleichtern die Einhaltung von Vorschriften.
- **Inhaltsmoderation in Echtzeit:** Schädliche oder diskriminierende Inhalte wie menschenverachtende Aussagen, Verbrechensdarstellungen sowie Waffen- und Gewaltverherrlichungen werden automatisch herausgefiltert und kontrolliert.
- **Effektive KI-Daten-Governance:** Schützen Sie Ihre Daten über ihren gesamten Lebenszyklus hinweg durch die automatisierte Erkennung, Klassifizierung und proaktive Härtung vor der Einführung, damit Ihr geistiges Eigentum stets geschützt ist und alle Vorschriften eingehalten werden.

Durch die Kombination von Transparenz, Compliance-Daten und adaptiver Richtliniendurchsetzung ermöglicht Netskope Unternehmen die Einführung von KI-Innovationen. Dabei werden nicht nur die ethischen Anforderungen und behördlichen Vorschriften von heute, sondern auch die von morgen erfüllt.

Die Zukunft der KI-Sicherheit

Neue Technologien und Bedrohungen

Die KI-Einführung nimmt zu, neue Anwendungsfälle (von Copiloten bis hin zu individuell entwickelten KI-Agenten) werden immer mehr zur Normalität – und leider hält die Bedrohungslandschaft mit dieser rasanten Entwicklung Schritt. Wenngleich sich die Sicherheitsmaßnahmen heutzutage vor allem auf Datenschutz und Modellintegrität konzentrieren, werden zwei neue Bereiche der technologischen Entwicklung in naher Zukunft voraussichtlich sogar noch größere Probleme mit sich bringen.

Erstens sind agentenbasierte KI-Systeme im Aufstieg begriffen, die mit minimaler menschlicher Aufsicht Entscheidungen treffen und Maßnahmen ergreifen. Laut Gartner werden bis zum Jahr 2028 mindestens 15 % der täglichen Geschäftsentscheidungen autonom von agentenbasierter KI getroffen – heute liegt dieser Anteil praktisch bei null.[4] Durch diese Veränderung wird sich die Angriffsfläche erheblich vergrößern, insbesondere wenn Agenten über MCP oder A2A (Agent-to-Agent-Protokoll) Zugriff auf Unternehmenssysteme und -daten erhalten.

Zweitens gewinnt physische KI, wie bei autonomen Fahrzeugen und Robotern, immer mehr Boden in Branchen wie Logistik, Transport und Fertigung. Diese Systeme bringen in der Praxis Sicherheitsrisiken mit sich, weil kompromittierte oder fehlerhafte KI nicht nur zu Datenverlusten führt, sondern auch die Beschäftigten und die Infrastruktur von Unternehmen schädigen kann.

Da sich KI-Technologien immer weiter entwickeln und tief in die alltäglichen Geschäftsabläufe integriert werden, müssen die Zuständigen für Sicherheit für strategische, zukunftsgerichtete Governance sorgen.

Die folgenden Überlegungen können dabei helfen, möglichen Problemen zuvorzukommen:

- **Transparenz der KI-Nutzung:** Bleiben Sie auf dem Laufenden, welche Teams KI-Modelle entwickeln oder nutzen (sowohl offene Modelle als auch Schatten-IT). Sorgen Sie für zentrale Transparenz und Aufsicht, ohne Innovationen zu hemmen.
- **Vertrauenswürdigkeit der Daten:** Die Modelle müssen mit sicheren, konformen und hochzuverlässigen Datensätzen trainiert werden. Minderwertige oder verunreinigte Daten führen zu fehlerhaften, unausgewogenen oder indiskreten Ausgaben.
- **Autonomie und Risikobegrenzung:** Die Fähigkeiten agentenbasierter KI nehmen immer weiter zu – legen Sie unbedingt klare Richtlinien für ihre Autonomie fest. Warten Sie nicht, bis die Agenten Entscheidungen mit weitreichenden Auswirkungen treffen, sondern sorgen Sie jetzt schon für Governance.
- **Lebenszyklusmanagement für Modelle:** Behandeln Sie Ihre KI-Modelle wie Code: mit Versionskontrollen, Schwachstellen-Scans, Zugriffssteuerung und Audit-Protokollen.
- **Sicherheit als Unternehmenskultur:** Sicherheit ist nicht nur eine Frage der Technik, sondern auch des Verhaltens. Informieren Sie Mitarbeiter und Führungskräfte über die Risiken und die sichere Nutzung von KI sowie über Veränderungen im regulatorischen Umfeld.

Für die Zukunft der KI-Sicherheit ist nicht nur entscheidend, wie gut sich Unternehmen vor den aktuellen Bedrohungen schützen, sondern auch, wie durchdacht sie sich auf die kommenden vorbereiten.

[4] Gartner 2024 https://www.gartner.com/en/newsroom/press-releases/2024-10-21-gartner-identifies-the-top-10-strategic-technology-trends-for-2025

Prognose

Das Marktforschungsunternehmen Gartner sagt voraus, dass agentenbasierte KI bis 2028 mindestens 15 % der täglichen Geschäftsentscheidungen autonom treffen wird. 2024 waren es noch 0 %.

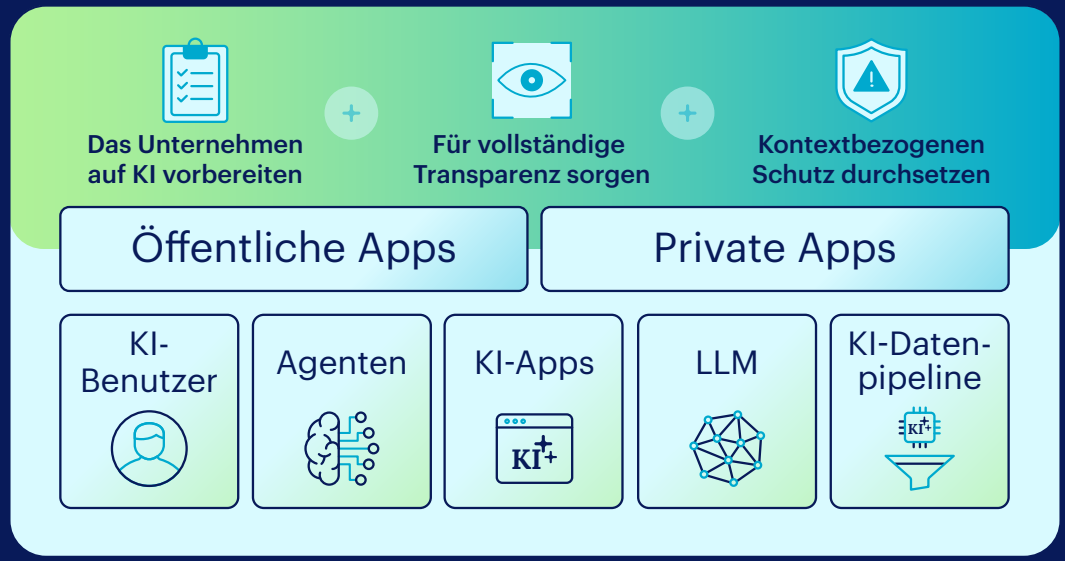
Fazit

Netskope One – ortsunabhängige End-to-End-Sicherheit für KI

Beim Wettrennen um die schnellstmögliche Einführung von KI stehen die Zuständigen für Sicherheitsfragen unter zunehmendem Druck, vertrauliche Daten zu schützen und neuen Risiken für ihr KI-Umfeld entgegenzutreten. Von mangelnder Transparenz bei der KI-Nutzung über die Offenlegung von Daten bis hin zu Compliance-Fragen haben wir sechs zentrale Herausforderungen beschrieben, vor denen Sicherheitsteams stehen, wenn KI im ganzen Unternehmen gefahrlos zum Einsatz kommen soll:

- Mangelnde Transparenz
- Die Risiken von KI-Anwendungen
- Integrität der KI-Modelle
- Gegen KI-Systeme gerichtete Bedrohungen
- Offenlegung von Daten
- Governance, Compliance und ethische Nutzung

Netskope One AI Securitystellt eine einzige Lösung zur Verwaltung Ihres gesamten KI-Ökosystems und zum Schutz Ihrer Daten bereit. Diese Lösung sichert Benutzer und automatisierte Agenten nicht nur in öffentlichen SaaS-Umgebungen ab, sondern schützt auch private KI-Tools und agentenbasierte Workflows. Durch die Kombination von Leistungsstärke mit kontextabhängigen Zero-Trust-Kontrollen hilft Netskope Unternehmen dabei, sich die Vorteile von KI auf sichere Weise zunutze zu machen.



Studien

Das Marktforschungsinstitut Forrester hat festgestellt, dass Netskope die Gefahr ernster Datenschutzverstöße durch Angriffe von außen um 80 % reduziert. Das entspricht pro Jahr 2 Mio. USD an Kosten, die durch Verhinderung schwerer Verstöße vermieden werden.[5]

[5] Forrester Report: The Total Economic Impact™ of Netskope SSE
https://www.netskope.com/resources/analyst-reports/forrester-the-total-economic-impact-of-netskope-sse

Über Netskope

Netskope (NASDAQ: NTSK), ein führender Anbieter von modernen Suchwachstellenevaluierung- und Netzwerklösungen für die Cloud- und KI-Ära, gibt Sicherheit- und Netzwerkteams genau das, was sie brauchen: optimierten Zugriff und kontextbasierte Sicherheit in Echtzeit für das gesamte KI-Ökosystem, einschließlich Agenten, Anwendungen, Tools, LLMs, Mitarbeitern, Geräten und Daten. Tausende Kunden, darunter mehr als 30 Unternehmen der Fortune 100, vertrauen der Netskope One-Plattform, ihrer Zero Trust Engine und ihrem leistungsstarken NewEdge-Netzwerk, wenn es um die Verringerung von Risiken geht. Sie erhalten vollständige Transparenz und Kontrolle über alle Cloud-, KI-, SaaS-, Web- und Privatanwendungen und können ihre Sicherheitslage und Leistung ohne Kompromisse verbessern.

Sie möchten mehr erfahren?

Demo anfordern

Ressourcen

- Sichere KI mit Netskope One
- Blog zu gelungener, ortsunabhängiger KI-Einführung mit End-to-End-Sicherheit
- Netskope Threat Labs: Bedrohungsbericht zu generativer KI in der Cloud
- Sichere generative KI für Dummies

netskope

©2026 Netskope, Inc. Alle Rechte vorbehalten. Netskope, NewEdge, SkopeAI und das stilisierte „N"-Logo sind eingetragene Marken von Netskope, Inc. Netskope Active, Netskope Cloud XD, Netskope Discovery, Cloud Confidence Index und SkopeSights sind Marken von Netskope, Inc. Alle anderen enthaltenen Marken sind Marken ihrer jeweiligen Inhaber. 04/26 EB-827-5-DE