Sicherheit – Ihr Wettbewerbsvorteil beim KI-Wettlauf

Leitfaden für Führungskräfte zur erfolgreichen und risikoarmen Beschleunigung der KI-Transformation

Einleitung

Unternehmen versuchen gerade, sich die transformative Kraft der KI möglichst schnell zunutze zu machen. Doch die Sicherheit hält mit dem Tempo der Einführung oft nicht Schritt. Jeder offiziellen Initiative stehen unzähligen Fälle gegenüber, in denen einzelne Mitarbeitende oder ganze Teams für die Entstehung von Schatten-KI sorgen. Tatsächlich geben 85 % der Entscheidungsträger im IT-Bereich an, dass Mitarbeitende KI-Tools einführen, bevor sie von der IT-Abteilung begutachtet wurden.[1] Darüber hinaus geben 93 % der Mitarbeitenden zu, dass sie ohne vorherige Genehmigung Informationen in KI-Tools eingeben.[1]

Unterdessen nehmen KI-basierte Angriffsvektoren rasant zu. Hier wird ein Anstieg um 47 % im Jahresvergleich verzeichnet.[2] Herkömmliche Sicherheitstools haben zunehmend Schwierigkeiten, mit der Entwicklung Schritt zu halten. Neue Formen ungewollter Datenoffenlegung und Compliance-Lücken stellen bestehende Governance-Verfahrensweise infrage. Führungskräfte müssen sich die entscheidende Frage stellen, wie sie die mit KI verbundenen Risiken in den Griff bekommen, ohne den dadurch ermöglichten Innovationen im Weg zu stehen.

Die Aufgabe ist klar. Unternehmen müssen eine Umgebung schaffen, die folgende Anforderungen erfüllt:

- Die gesamte KI-Nutzung ist bekannt
- Die vollständige KI-Kommunikation kann abgesichert werden
- Alle KI-Richtlinien können durchgesetzt werden
- Sämtliche KI-Modelle können vor Missbrauch geschützt werden

Die **Cloudflare AI Security Suite** verleiht Unternehmen die Gewissheit, die sie für einen schnelleren Einsatz von KI brauchen, indem sie die durch Risiken verursachte Unsicherheit beseitigt. Unsere übergreifende Plattform sichert den gesamten KI-Lebenszyklus ab, indem sie die KI-Nutzung aufdeckt, Zero Trust-Zugriff implementiert sowie das Web und API-Endpunkte mit proaktiver Bedrohungsabwehr schützt. Dank integrierter Data Governance können Innovationen frei und auf sichere Weise vorangetrieben werden.

85 %
der Entscheidungsträger im IT-Bereich geben an, dass Mitarbeitende KI-Tools einführen, bevor sie von der IT-Abteilung begutachtet werden können.[1]

Die Herausforderung der KI-Sicherheit

Die mit der KI-Einführung verbundenen Risiken geben einem schnell wachsenden Markt für KI-Sicherheitslösungen weiteren Auftrieb. Beispielsweise konzentrieren sich cloudnative Plattformen für Anwendungsschutz auf Arbeitsabläufe mit Bezug zur KI-Entwicklung. Erste Angebote dieser Art stellen eine wichtige Verteidigungslinie dar, reichen jedoch nicht aus. Unternehmen müssen den gesamten KI-Lebenszyklus – und somit auch KI-Systeme – schützen, sobald die entsprechenden Lösungen im Produktivbetrieb zum Einsatz kommen.

Das ganze Ausmaß der Herausforderung im Bereich der KI-Sicherheit wird immer deutlicher. Inzwischen erschaffen Entwickler KI-Funktionen, Mitarbeitende nutzen externe KI-Tools und Kunden interagieren mit KI-gestützten Anwendungen. Die manuelle Absicherung all dieser unterschiedlichen Umgebungen ist kompliziert. Und dies auf einheitliche Weise zu bewerkstelligen, ist sogar noch schwieriger.

KI-Missbrauch durch Mitarbeitende verhindern

Da Mitarbeitende sowohl genehmigte als auch nicht genehmigte KI-Tools verwenden, sind sensible Daten und Betriebsabläufe gefährdet. Sicherheitsabteilungen müssen sich deshalb mit Folgendem auseinandersetzen:

- **Mangelnder Überblick über die KI-Nutzung:** Führungskräfte können sich oft kein umfassendes Bild davon machen, welche KI-Tools ihre Mitarbeitenden verwenden, wo sensible Daten verarbeitet werden und wie diese KI-Systeme mit Geschäftsanwendungen verbunden sind. Diese blinden Flecken bergen ein erhebliches Risiko.
- **Risiken bezüglich Datensicherheit und Compliance:** Durch KI verändert sich die Art und Weise, in der Daten innerhalb des Unternehmens bewegt werden. Personen- und kundenbezogene Informationen sowie firmeneigene Daten können in KI-Systemen landen und so Compliance-Verstöße verursachen oder Betriebsgeheimnisse offenlegen.
- **Zugriffskontrollen für agentenbasierte KI-Workflows:** Es muss nicht nur der Zugriff des Menschen auf KI-Tools verwaltet werden, sondern auch der Zugang agentenbasierter KI zu MCP-Servern und anderen kritischen Systemen. Dies erfordert neue Konzepte für Identitätsmanagement und Zugriffskontrollen.

Schutz von öffentlich zugänglichen KI-Anwendungen und -Modellen

KI-Systeme werden zu einem wesentlichen Bestandteil der Kunden- und Nutzererfahrung und müssen als solche geschützt werden – unabhängig davon, ob sie intern entwickelt wurden oder über Dritte genutzt werden. Die OWASP Top 10 für LLM-Anwendungen hebt Bedrohungen hervor, denen herkömmliche Tools nicht gewachsen sind:

- **Angriffe durch unkontrollierten Verbrauch:** Ähnlich wie bei traditionellen DoS-Angriffen wird bei Attacken mittels unkontrolliertem Verbrauch versucht, mit ressourcenintensiven Anfragen eine Überlastung von LLM herbeizuführen. Cloud-Preismodelle nach dem Pay-per-Use-Prinzip können die finanziellen Auswirkungen solcher Angriffe in die Höhe treiben, während legitime Nutzer mit einer verminderten Servicequalität konfrontiert sind.
- **„Vergiftung" von Modellen:** Angreifer bauen Hintertüren, Verzerrungen oder Schwachstellen in LLM ein, indem sie schadhafte Daten in öffentliche Datensätze oder Repositorys einschleusen, mit denen Entwickler Modelle trainieren. Ein auf diese Weise „vergiftetes" Modell verhält sich normal, bis durch bestimmte Auslöser schädliche Verhaltensweisen aktiviert werden.
- **Prompt Injection-Angriffe:** Eine beliebte Taktik zur Ausschleusung von Daten über KI-Schnittstellen ist Prompt Injection. Dabei werden LLM-Eingaben manipuliert, indem man bösartige Anweisungen in Nutzer-Prompts oder externe Inhalte einbettet. Dies führt dazu, dass die Modelle die ursprünglichen Anweisungen ignorieren und stattdessen Befehle des Angreifers ausführen.
- **Jailbreaking:** Sorgfältig ausgearbeitete Prompts wie Rollenspiel-Szenarien, Befehle zum Überschreiben von Anweisungen und Multi-Turn-Strategien können die Sicherheitsvorkehrungen von LLM umgehen, um unzulässige Inhalte zu generieren oder sensible Informationen auszuschleusen.

Um KI-gestützte Innovationen schnell vorantreiben zu können, ohne dabei das Unternehmen Gefahren auszusetzen, empfehlen wir Führungskräften die Wahl eines KI-Sicherheitsansatzes, der das gesamte Anforderungsspektrum abdeckt:

- Absicherung der Nutzung generativer KI durch Mitarbeitende
- Schutz KI-gestützter Anwendungen und Workloads
- Entwicklung KI-gestützter, von Grund auf sicherer Anwendungen

Absicherung von Workflows bei Entwicklung und Training von KI

KI-Projekte umfassen riesige Datensätze, kostspielige Ressourcen und iterative Experimente, wodurch jeweils neue Angriffsflächen und Schwachstellen entstehen. Sicherheitsabteilungen müssen daher Folgendes berücksichtigen:

- **Sicherheit und Integrität von Trainingsdaten:** Kompromittierte Trainingsdaten können zu Verzerrungen, der Entstehung von Hintertüren oder Schwachstellen bei Modellen im Produktivbetrieb führen. Unternehmen müssen deshalb den Zugriff auf Trainingsdatensätze absichern, Änderungen durch Unbefugte verhindern und die Vertrauenswürdigkeit der Datenherkunft während des gesamten Modelllebenszyklus gewährleisten.
- **Verwaltung von Anmeldedaten und Secrets:** Für Workflows bei der KI-Entwicklung wird Zugriff auf Systeme wie Cloud-Speicher für Datensätze, Rechen-Cluster für das Training, Modellregistrierungen und externe Dienste benötigt. Durch unzureichend geschützte Secrets oder API-Schlüssel können sensible Anmeldedaten offengelegt und unbefugter Zugriff auf firmeneigene Modelle, Trainingsdaten oder Produktivsysteme ermöglicht werden.
- **Zugriffskontrollen für die Entwicklungsumgebung:** KI-Entwickler brauchen oft weiter gefasste Rechte, um mit Modellen experimentieren und auf sensible Daten zugreifen zu können. Ohne angemessene Zugriffskontrollen kann dies interne Bedrohungen schaffen, zu unbeabsichtigter Datenoffenlegung führen oder eine Modellausschleusung durch Unbefugte ermöglichen.

Einheitliche Sicherheit mit Cloudflare

Für eine erfolgreiche KI-Einführung sollte man sich darauf konzentrieren, die Produktivität nicht einzuschränken, sondern zu fördern. Wenn Mitarbeitende KI vertrauensvoll einsetzen, weil sie wissen, dass die richtigen Sicherheitsvorkehrungen getroffen wurden, können sie Innovationen schneller vorantreiben und ehrgeizigere Projekte in Angriff nehmen.

Die Cloudflare AI Security Suite schafft ein sicheres Umfeld für Innovation mit KI. Durch die Zusammenführung von Funktionen für SASE (Secure Access Service Edge) und Webanwendungen können Führungskräfte zwei grundlegende Bereiche verbinden und schützen:

- Externe, öffentlich zugängliche KI-fähige Anwendungen
- Interne, private KI-Systeme und -Workloads

Die Cloudflare AI Security Suite deckt den gesamten KI-Lebenszyklus ab. Dies umfasst Sicherheitsanforderungen von der Bedrohungserkennung und dem Risikomanagement bis hin zum Datenschutz, der Absicherung des Nutzerzugriffs und dem Schutz von KI-gestützten Anwendungen und Entwicklungsworkflows. Als wichtige Ebene einer Echtzeit-Produktivsicherheit ist das globale Cloudflare-Netzwerk intern zwischengeschaltet, um jede KI-Interaktion zu überprüfen und zu filtern und so die Daten aller Nutzer und Anwendungen zu schützen.

Anstatt Probleme erst nach ihrem Auftreten zu erkennen, geht Cloudflare diese schon bei ihrer Entstehung an. So werden Bedrohungen blockiert, bevor sie KI-Modelle überhaupt erreichen können. Sicherheitsabteilungen erhalten den nötigen Überblick, um neuen Gefahren immer einen Schritt voraus zu sein.

Kernfunktionen von Cloudflare

Die Cloudflare AI Security Suite bietet Überwachung, Echtzeitschutz und proaktives Risikomanagement umfassend und unter einem einzigen Dach. Dies ermöglicht einen ganzheitlichen Ansatz für KI-Sicherheit.

Umfassende KI-Erkennung und -Übersicht

Wirksame KI-Sicherheit erfordert eine vollständige Echtzeit-Bestandserfassung von genehmigten und nicht genehmigten KI-Ressourcen sowie deren Nutzung. Die Grundlage der Cloudflare AI Security Suite bildet die kontinuierliche Überwachung und automatisierte Identifizierung, damit sämtliche KI-Modelle, Assistenten, Agenten und Schatten-KI-Implementierungen in jeder Art von Umgebung ermittelt werden – ob öffentlich zugänglich, privat oder intern.

Proaktives KI-Risikomanagement

Mit der Cloudflare AI Security Suite werden KI-spezifische Schwachstellen, Fehlkonfigurationen und Angriffspfade, einschließlich derer, die in der OWASP Top 10 für LLM aufgeführt sind, erkannt und neutralisiert. Das unterstützt Unternehmen bei der Angriffsabwehr. Die Bewertung der Anwendungssicherheit hilft bei der Priorisierung der Problembehebung, sodass sich Mitarbeitende den größten Risiken zuerst widmen können.

Anwendungssicherheit für KI-gestützte Anwendungen

Um SecOps dabei zu unterstützen, bezüglich der neuesten KI-Bedrohungsvektoren auf dem Laufenden zu bleiben, beinhaltet die Cloudflare AI Security Suite eine proaktive Bedrohungserkennung und -abwehr für KI-spezifische Schwachstellen, Fehlkonfigurationen und Angriffspfade innerhalb von KI-Pipelines. Dies umfasst den Schutz vor Prompt Injection, Data Poisoning und Modellmissbrauch.

Spezielle KI-Firewalls spüren generative oder agentenbasierte KI und entsprechende API-Endpunkte auf und kennzeichnen diese. Sie erkennen außerdem Versuche, personenbezogene Daten auszuschleusen, und blockieren schädliche Prompts, bevor diese die Performance von KI-Modellen beeinträchtigen oder ein Modell mit schädlichen Inhalten oder Falschinformationen „vergiften“.

Zero Trust-Zugriff für Workflows im Zusammenhang mit generativer und agentenbasierter KI

Zero Trust-Prinzipien wie das der Erteilung geringstmöglicher Rechte werden sowohl auf Mitarbeitende als auch auf KI-Agenten angewandt. Die Cloudflare AI Security Suite kann ZTNA (Zero Trust Network Access)-Richtlinien für Interaktionen zwischen Mensch und KI sowie für solche zwischen verschiedenen künstlichen Intelligenzen durchsetzen. Protokollierung und Kontrollen für MCP-Server an zentraler Stelle sorgen dafür, dass agentenbasierte KI nur auf das zugreift, wofür sie auch autorisiert ist. Davon sind auch bedarfsgerechte Workflows abgedeckt.

KI-bezogener Datenschutz

Wirksame KI-Sicherheit erfordert eine Funktion zum Schutz vor Datenverlust (Data Loss Prevention –DLP), die mehrere Sprachmodelle nutzt, um sowohl den Inhalt als auch die Absicht eines Prompts zu verstehen. In der Cloudflare AI Security Suite ist dieses Feature sowohl für das Training als auch für Prompts und Antworten integriert. Dies verhindert die Offenlegung personenbezogener Daten, Datenverluste und unbefugten Zugriff innerhalb von KI-Modellen und -Pipelines. Die intern eingesetzte, API-zentrierte Laufzeitsicherheit fungiert als schnelle und einfache erste Verteidigungslinie. Damit wird der durch eine CNAPP (Cloud-Native Application Protection Platform) ermöglichte „Shift Left“-Ansatz ergänzt.

Datenlokalisierung

LLM und KI-Anwendungen unterliegen den gleichen Vorschriften wie jede andere Art von Datenumgebung. Mithilfe der Cloudflare AI Security Suite können Unternehmen sicherstellen, dass KI-Workloads geografische und rechtliche Grenzen einhalten. Dafür werden Richtlinien angewandt, die dafür sorgen, dass Trainingsdaten und Inferenzanfragen bestimmte, zuvor festgelegte Weltregionen nicht verlassen.

Sicherheit als Grundbestandteil der KI-Entwicklung

Wie bei jeder Software sollte Sicherheit auch im Fall von KI-Anwendungen bei der Entwicklung von Anfang an mitgedacht und nicht nur nachträglich ergänzt werden. Die Cloudflare AI Security Suite bietet Tools und Frameworks zur Entwicklung KI-gestützter Anwendungen, die von vornherein sicher sind.

Der geschäftliche Effekt der Cloudflare AI Security Suite

Der Siegeszug der KI ist mehr als eine Weiterentwicklung – er stellt eine tiefgreifende Umwälzung dar, die sich mit der Industrialisierung oder Computerisierung messen kann. Eine schnelle und erfolgreiche Einführung dieser Technologie ist deshalb nicht nur für die Wachstumsförderung, sondern auch für die Rentabilität von Bedeutung. Eine langsame oder schlecht abgesicherte Implementierung stellt heute eine existenzielle Bedrohung für Unternehmen in ganzen Branchen und Sektoren dar.

Die Cloudflare AI Security Suite ermöglicht eine sichere, kontrollierte und effiziente Transformation, mit der man den wachsenden Kundenerwartungen und Marktanforderungen in einem wettbewerbsintensiven Umfeld gerecht werden kann.

- **Schnellere KI-Innovation:** Wenn Sicherheit eine geschützte Nutzung ermöglicht, anstatt die Einführung von KI zu verhindern, können Mitarbeitende und Teams neue KI-Anwendungen erproben, die sie in ihrem Arbeitsalltag produktiver machen. Geeignete Sicherheitsframeworks geben das nötige Vertrauen, um anspruchsvolle KI-Funktionen zu entwickeln, ohne dabei sensible Daten oder Systeme zu gefährden.
- **Verringertes KI-bezogenes Risiko:** Unternehmen können das gesamte Spektrum an Risiken verwalten, die mit der Einführung von KI verbunden sind – einschließlich solcher, die naturgemäß mit KI-fähigen Webanwendungen einhergehen. KI-Applikationen, die sich in der Entwicklung befinden, können abgesichert werden, damit durch Mitarbeitende keine sensiblen Daten verloren gehen oder in KI-Trainingsdatensätzen offengelegt werden. Die SecOps-Abteilung kann KI-spezifische Bedrohungen und Schwachstellen proaktiv ermitteln und neutralisieren, wodurch die Angriffsfläche verringert und kritische Daten und Modelle geschützt werden.
- **Optimierte Sicherheitsabläufe:** Zentrale Übersicht und Kontrolle über das KI-Sicherheitsniveau vereinfachen die Verwaltung und ermöglichen eine optimierte Vorfallreaktion. So kann sich die SecOps-Abteilung auf strategisch bedeutsame Initiativen konzentrieren, anstatt sich ständig mit Vorfällen im Verbindung mit KI herumzuschlagen.
- **Robuste Data Governance und Compliance:** KI-spezifische Datenschutzmechanismen helfen dabei, sensible Daten zu schützen und die sich wandelnden gesetzlichen Anforderungen während des gesamten KI-Lebenszyklus zu erfüllen.
- **Niedrigere Gesamtbetriebskosten:** Die Nutzung einer übergreifenden Plattform, in die bestehende Sicherheitsinvestitionen integriert werden, ist kostensparender als die Implementierung gesonderter Insellösungen für jede Herausforderung im Bereich KI-Sicherheit.

Modellanwendungsfälle für die Cloudflare AI Security Suite

Die Cloudflare AI Security Suite ist perfekt dafür geeignet, grundlegende Anforderungen im Zusammenhang mit der Einführung von KI zu erfüllen.

- **Absicherung der Nutzung von KI-Tools durch Mitarbeitende:** Für den Zugriff der Belegschaft auf öffentlich zugängliche Tools mit generativer KI wie ChatGPT und intern entwickelte, KI-gestützte Anwendungen werden Zero Trust-Richtlinien durchgesetzt.
- **Schutz öffentlich zugänglicher, KI-gestützter Anwendungen:** Webanwendungen und API, die KI-Modelle beinhalten – wie Chatbots und Empfehlungs-Engines – werden vor Angriffen geschützt, in deren Rahmen sensible Daten offengelegt oder Modelle missbraucht werden könnten.
- **Handhabung von Schatten-KI:** Nicht genehmigte KI-Tools im Unternehmen werden automatisch erkannt. Um fortgesetzte Innovation innerhalb festgelegter Risikoparameter zu ermöglichen, werden geeignete Kontrollen angewandt.
- **KI-gestützter DLP für KI-Interaktionen:** Durch das Unterbinden der Offenlegung sensibler Daten in KI-Prompts und -Ergebnissen wird der Schutz personenbezogener Daten gewährleistet.
- **KI-Entwicklungssicherheit:** Engineering-Abteilungen werden Frameworks bereitgestellt, die eine sichere Entwicklung zum Standardansatz machen und es ihnen ermöglichen, KI-Funktionen schnell und ohne Beeinträchtigung der Sicherheit zu erschaffen.

Überlegungen zur Implementierung

Als grundlegende Sicherheitsstrategie sollte man KI-Sicherheit wohlüberlegt angehen.

Mit der aktuellen Infrastruktur beginnen	Die erfolgreichsten KI-Sicherheitsimplementierungen bauen auf vorhandenen Tools für SASE und Anwendungssicherheit auf, anstatt sie zu ersetzen. Bei dieser Herangehensweise werden aktuelle Investitionen genutzt. Gleichzeitig erweitert man den Schutz, damit auch KI-spezifische Risiken abgedeckt werden.
Einheitliche interne Schutzmaßnahmen anwenden	Die Fähigkeit, schädliche Aktivitäten am Netzwerkrand in Echtzeit zu blockieren, ist entscheidend für die KI-Sicherheit. Interne Echtzeit-Kontrollen sollten durch API-basierte Überwachung ergänzt werden.
Vollständige Abdeckung sicherstellen	Die KI-Sicherheitsstrategie sollte das gesamte Anforderungsspektrum abdecken: die Nutzung von Tools mit generativer KI durch Mitarbeitende, den Schutz für KI-gestützte Anwendungen und Workflows, die Absicherung von Workflows im Zusammenhang mit agentenbasierter KI und die Sicherheit für die Arbeitsabläufe bei der KI-Entwicklung.
In großem Maßstab planen	Es sollten Lösungen gewählt werden, die mit der jeweiligen KI-Einführung wachsen können. Was bei einem Pilotprogramm funktioniert, muss bei zunehmender KI-Nutzung im gesamten Unternehmen auch skalierbar sein.
Erfolgskriterien überprüfen	Vor der endgültigen Entscheidung sollte die Lösung in realistischen Szenarien auf die Probe gestellt werden. Es sollte eine Plattform gewählt werden, deren Enterprise-Funktionen kostenlos und in Eigenregie aktiviert werden können. So kann die gesamte Sicherheitssuite in kleinem Maßstab getestet werden – beispielsweise für ein einzelnes Team oder eine einzelne Anwendung –, um schnell ihren Nutzen zu belegen und sicherzustellen, dass sie die festgelegten Erfolgskriterien erfüllt.

Mit der Cloudflare AI Security Suite den nächsten Schritt machen

Angesichts der rasanten Verbreitung von KI auf der ganzen Welt werden die Unternehmen, die diese Technologie auf sichere Weise skalieren können, über einen großen Wettbewerbsvorteil verfügen. Der springende Punkt ist: Bei KI-Sicherheit geht es nicht darum, die Nutzung von KI zu verhindern, sondern sie auf smarte Weise zu ermöglichen.

Die **Cloudflare AI Security Suite** gibt Unternehmen die nötige Sicherheit, um Innovationen voranzutreiben. Sie baut auf unseren weit verbreiteten Plattformen für SASE und Anwendungssicherheit auf und erweitert diese. Geboten werden integrierte KI-Erkennung, Zero Trust-Zugriffskontrollen, eine informationsgetriebene, proaktive Bedrohungsabwehr und solide Data Governance. Wenn Unternehmen die Nutzung von KI-Modellen und die Modelle selbst rundum absichern, können sie ihre Produkte schneller entwickeln und die Produktivität ihrer Mitarbeitenden erhöhen, ohne dabei die Endnutzererfahrung zu beeinträchtigen.

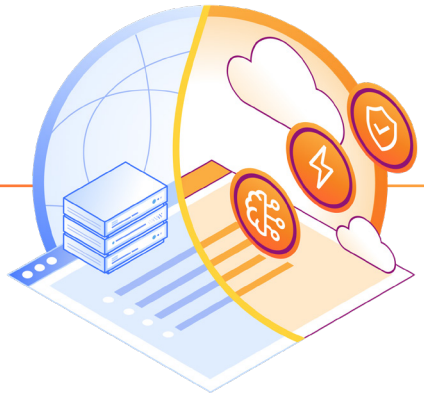
Beratungstermin vereinbaren

Erfahren Sie, wie die Cloudflare AI Security Suite die Herangehensweise Ihres Unternehmens an die sichere KI-Einführung grundlegend verändern kann.

+49 89 2555 2276

enterprise@cloudflare.com

www.cloudflare.com/de-de/



1. ManageEngine. „The Shadow AI Surge in Enterprises: Insights from the US & Canada".
2. Check Point Research. „Q1 2025 Global Cyber Attack Report from Check Point Software: An Almost 50% Surge in Cyber Threats Worldwide, with a Rise of 126% in Ransomware Attacks".