netskope

Cybersecurity INSIDERS

2026

AI Risk and Readiness Report

Übersicht

Für die meisten Unternehmen ist dieses Jahr der Zeitpunkt gekommen, an dem KI zur Infrastruktur wird. Agenten führen Aktionen bereits selbstständig aus: Sie ändern Datensätze, erstellen Konten und übertragen Code über API-Aufrufe, die abgeschlossen sind, bevor sie von einem Menschen überprüft werden können. Dadurch wird jede KI-Implementierung zu einem Sicherheitsrisiko – auch wenn Unternehmen sie nicht unbedingt als solches betrachtet.

Die Sicherheitssysteme der meisten Unternehmen wurden für eine andere Umgebung entwickelt. Darin waren Menschen die einzigen Akteure, Prozesse liefen deterministisch ab, Daten hatten stets eine erkennbare Form und die Vertrauenswürdigkeit wurde auf Browserebene überprüft. Diese Umgebung existiert jedoch nicht mehr.

Dieser Bericht basiert auf einer umfassenden Umfrage unter 1.253 Cybersicherheitsexperten. Darin wird untersucht, wie Unternehmen ihre KI-Systeme im Hinblick auf Governance, Transparenz, Datenschutz und Agentenkontrolle absichern.

Wichtige Erkenntnisse:

- **Die KI-Einführung hat die Security Governance** überholt: 73 % der befragten Unternehmen setzen bereits KI-Tools ein, doch nur bei 7 % gibt es Governance-Systeme, die Sicherheitsmechanismen und Richtlinien in Echtzeit durchsetzen. Daraus ergibt sich ein strukturelles Defizit von 66 Prozentpunkten – und dieses Defizit wird immer größer, da sich die Einführung von KI schneller vollzieht als die Entwicklung von Kontrollmechanismen.
- **Die Ausgaben steigen, doch das Vertrauen nimmt ab**
  90 % erhöhten ihre Budgets für KI-Sicherheit in diesem Jahr, doch 29 % fühlen sich weniger sicher als noch vor zwölf Monaten. Das Problem wächst schneller als die Investitionen.
- **Die meisten KI-Aktivitäten sind für Sicherheitskräfte unsichtbar**
  94 % der Befragten berichten von Lücken bei der Transparenz über KI-Aktivitäten. 88 % können private KI-Konten nicht von Unternehmensinstanzen unterscheiden. Nur 6 % geben an, den gesamten Umfang der KI-Pipeline ihres Unternehmens überblicken zu können.
- **KI macht herkömmliche Data Loss Prevention (DLP) wirkungslos**
  DLP gleicht Muster ab, während KI Inhalte umdeutet. Nur 8 % verfügen über Kontrollmechanismen, die Inhalte semantisch auswerten, unabhängig davon, wie diese umgeschrieben wurden.
- **Agenten handeln ohne Leitplanken**
  KI-Agenten verfügen über Schreibzugriff auf Collaboration-Tools (53 %), E-Mail-Programme (40 %), Code-Repositorys (25 %) und Identitätsanbieter (8 %). 91 % der Unternehmen erfahren erst, was ein Agent getan hat, nachdem die Aktion bereits ausgeführt wurde.
- **Zu viele KI-Sicherheitsmechanismen basieren auf Vertrauen**
  31 % der Unternehmen verlassen sich hauptsächlich auf schriftliche Richtlinien sowie darauf, dass Mitarbeiter diese einhalten. Weitere 11 % haben nichts dergleichen. Nur 23 % geben an, dass sie KI-Sicherheitsfunktionen inline, also direkt am Ort der Ausführung, durchsetzen.

KI-bedingte Risiken weiten sich heute vom Missbrauch der Technologie durch den Menschen auf die Autonomie von Maschinen aus. Die bestehenden Kontrollmechanismen sind jedoch nach wie vor nur auf die Bewältigung der ersten Herausforderung ausgerichtet. Die Umfrage weist auf vier Prioritäten in Bezug auf die Architektur hin: durchgängige Transparenz über alle KI-Aktivitäten, einschließlich des Agenten- und M2M-Datenverkehrs, Inline-Durchsetzung ohne Reibungsverluste und Latenz, semantisch orientierte Datenkontrollen, die Inhalte statt Mustern bewerten, sowie die Ausweitung des Zero-Trust-Ansatzes auf nicht menschliche Identitäten (NHIs). Die folgenden Kapitel befassen sich mit der Messung der Reife von Unternehmen und der Frage, wie sich die Lücke bei der Ausführung schließen lässt.

KI-Governance bleibt weit hinter der Einführungsrate zurück

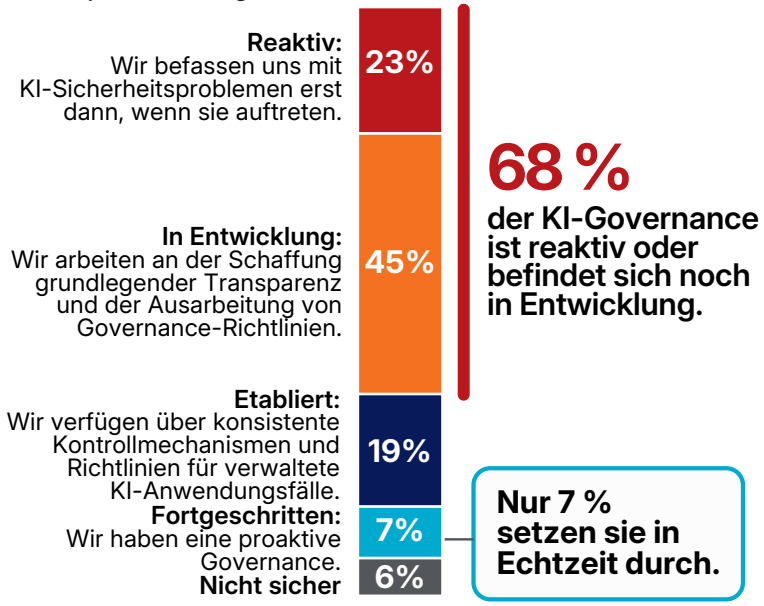
Vor zwölf Monaten betrachteten die meisten Unternehmen KI-Governance als ein Thema für die Zukunft: etwas, das erst dann formalisiert werden muss, wenn die Nutzung der Technologie eine gewisse Stabilität erreicht hat. Dieser Prozess vollzog sich jedoch schneller als erwartet. Copilots, Tools zur Code-Vervollständigung und Inhaltsgeneratoren wurden in unterschiedlichen Abteilungen in die Produktion integriert, und als die Sicherheitsabteilung endlich ein Framework eingerichtet hatte, war die KI-Infrastruktur bereits in Betrieb.

Heute bezeichnen 68 % der Unternehmen ihre KI-Governance als reaktiv oder noch in der Entwicklung befindlich. Nur 7 % haben bisher einen hohen Reifegrad erreicht und setzen Richtlinien in Echtzeit durch. Die Differenz von 66 Prozentpunkten zwischen den 73 % der Unternehmen, die KI-Tools einsetzen, und den 7 % mit Governance in Echtzeit ist ein strukturelles Missverhältnis. Unternehmen setzen mit hoher Geschwindigkeit Technologien ein, obwohl sie kaum eine Sicherheits- und Compliance-Grundlage dafür haben. Die Folgen dieser Vorgehensweise zeichnen sich bereits ab: 39 % der Unternehmen haben schon einen KI-bedingten Vorfall mit einer unbeabsichtigten Offenlegung von Daten erlebt. 17 % dieser Unternehmen nahmen danach keine Veränderungen vor.

Es mag zunächst nicht alarmierend klingen, dass die Governance bei 68 % der Unternehmen reaktiv ist oder sich noch in der Entwicklung befindet. Allerdings verbirgt sich hinter dieser Zahl ein problematisches, unternehmensinternes Muster. Mehr als ein Drittel der Befragten berichtete von einer fragmentierten KI-Einführung: Mehrere Teams setzten Tools unabhängig voneinander ein, wobei es weder ein gemeinsames Framework noch einheitliche Standards oder Sicherheitsrichtlinien gab. Ein Geschäftsbereich betreibt zum Beispiel autonome Agenten auf Basis informeller Richtlinien, während ein anderer gar nicht dokumentiert hat, welche KI-Tools die Mitarbeiter überhaupt nutzen. Das Gespräch über Governance findet nicht nur viel zu spät statt, es hat bei vielen Unternehmen erst gar nicht begonnen. 48 % glauben, dass Mängel in Bezug auf Governance – speziell Schatten-KI und allzu großzügige Zugriffsrechte – den nächsten großen KI-bezogenen Sicherheitsvorfall auslösen werden. Die Fachleute, die sich am besten mit diesem Problem auskennen, wissen bereits, dass die größten Risiken von den Tools ausgehen, die ihre eigenen Mitarbeiter im letzten Quartal unbemerkt eingesetzt haben. Dieses Governance-Defizit hat Auswirkungen auf alle weiteren Sicherheitsebenen.

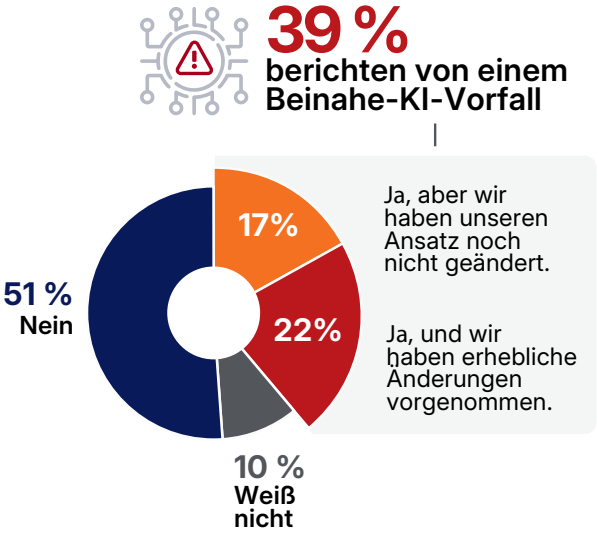
Echtzeit-Governance gibt es selten

Welche Aussage beschreibt den Reifegrad Ihres Unternehmens hinsichtlich der Governance und Sicherheit von KI-Implementierungen und Daten am treffendsten?



Beinahe-Vorfälle sind bereits Realität

Haben Sie schon einmal einen Beinahe-Vorfall im Zusammenhang mit KI erlebt, bei dem es zu einer unbeabsichtigten Offenlegung oder dem Verlust sensibler Daten kam und der Sie dazu veranlasst hat, Ihren Sicherheitsansatz grundlegend zu überdenken?



Die einfachste Lösung wäre, die drei risikoreichsten KI-Anwendungsfälle in Ihrer Umgebung zu ermitteln, durchsetzbare Richtlinien in technische Kontrollmaßnahmen für alle drei zu integrieren und jedem einen Verantwortlichen zuzuweisen.

Mehr Budget, weniger Vertrauen

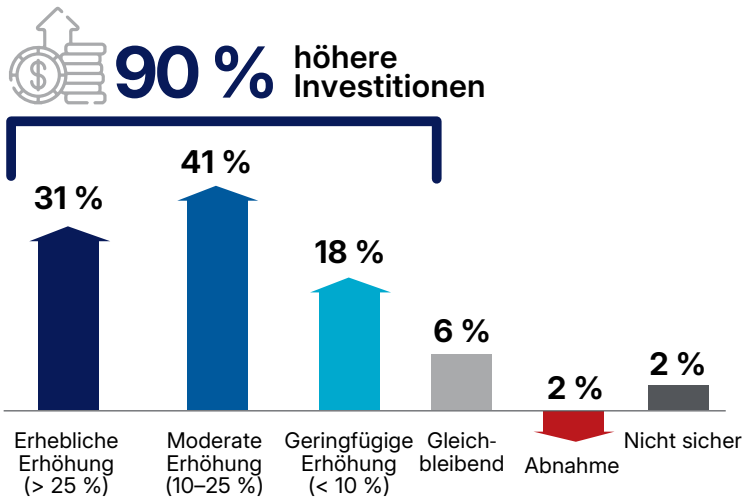
Paradoxerweise gibt es diese Lücke in der KI-Governance, obwohl Unternehmen mehr denn je in Sicherheit investieren. 90 % gaben in diesem Jahr mehr für KI-Sicherheit aus als in der Vergangenheit, wobei fast ein Drittel das Budget um mehr als 25 % aufstockte. Dennoch fühlen sich 29 % weniger sicher als vor zwölf Monaten. Die Investitionen nehmen zu – das Vertrauen nimmt ab.

Die Studienteilnehmer erklärten, warum: Für 34 % liegt das größte Problem im wirtschaftlichen Druck, KI schneller einzuführen, als die Sicherheitsstrukturen daran angepasst werden können. An zweiter Stelle standen Kompetenzlücken mit 25 %, und an dritter Stelle rangierten mit 21 % veraltete Tools, die KI-spezifische Bedrohungen nicht erkennen können. Finanzielle Herausforderungen wurden mit 14 % an vierter Stelle genannt. Das Budget steht zwar für viele bereits zur Verfügung, doch die Architektur, die damit finanziert wird, basiert noch immer auf einem Bedrohungsmodell aus der Zeit vor der KI.

Bestehende Sicherheitstools wurden für bekannte Dateiformate, vorhersehbare Datenflüsse und menschliche Interaktionen entwickelt. Ein höheres Budget für diese Sicherheitsstrukturen bedeutet lediglich, dass mehr Mittel in etwas fließen, das sich gegenüber KI-bedingten Risiken ohnehin schon als unwirksam erwiesen hat. Die Vertrauenskrise in der IT bestätigt dies: 51 % halten ihre technischen Kontrollmechanismen, 50 % die Transparenz und 61 % die NHI-Governance für unzureichend.

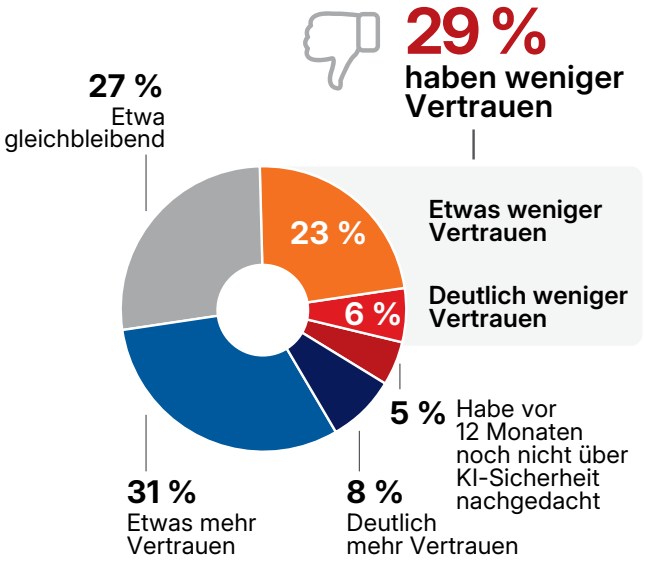
Die Ausgaben nehmen rapide zu

Wie werden sich die Investitionen Ihres Unternehmens in KI-spezifische Sicherheitskontrollen in den nächsten 12 Monaten verändern?



Das Vertrauen nimmt ab

Wie hat sich Ihr Vertrauen in die Fähigkeit Ihres Unternehmens, KI-Systeme vor Angriffen und Datenlecks zu schützen, in den letzten 12 Monaten verändert?



Zur Umverteilung des Budgets sollten die aktuellen KI-Sicherheitsausgaben zunächst diesen drei Sicherheitslücken (Transparenz, technische Kontrollmechanismen und NHI-Governance) zugeordnet werden. Anschließend können die Investitionen dort konzentriert werden, wo die größte Diskrepanz zwischen Ausgaben und Leistungsfähigkeit besteht.

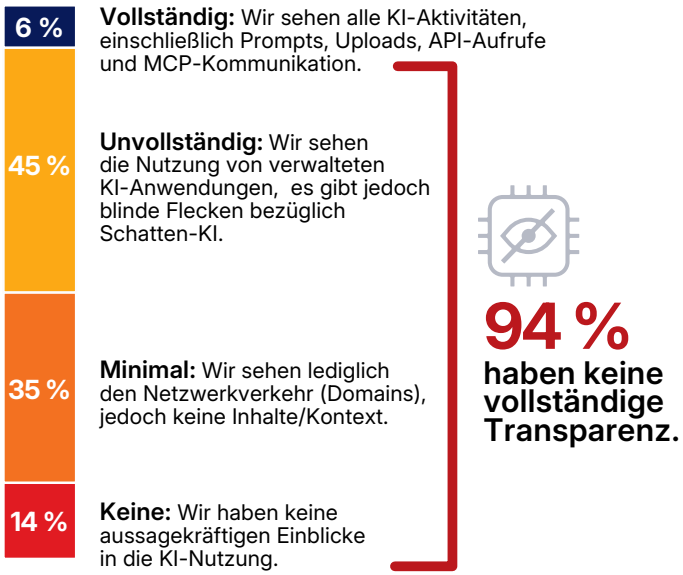
Die meisten KI-Aktivitäten sind für Sicherheitssysteme unsichtbar

Was man nicht sieht, kann man auch nicht schützen. Nur 6 % der Unternehmen geben an, vollständige Transparenz über die KI-Nutzung in ihrer gesamten Umgebung zu haben. 45 % haben eine eingeschränkte Transparenz, die sich auf verwaltete Anwendungen beschränkt. Alles, was außerhalb der autorisierten Tools liegt, ist für sie unsichtbar. 35 % erkennen nur die Datenverkehrsmuster auf Netzwerkebene – genug, um zu wissen, dass etwas passiert, aber nicht, was. Die restlichen 14 % haben überhaupt keine Transparenz. 94 % treffen Entscheidungen zur KI-Sicherheit ohne vollständige Informationen – für die meisten ist das der Normalzustand.

Wenngleich Erkennungsmöglichkeiten vorhanden sind, ist es schwierig, das Wesentliche herauszufiltern. 88 % können persönliche KI-Konten nicht zuverlässig von Unternehmenskonten auf derselben Plattform unterscheiden – das ist laut Umfrage der größte technische Schwachpunkt. Wenn ein Sicherheitsteam nicht erkennen kann, ob ein Mitarbeiter einen autorisierten KI-Mandanten oder ein privates Konto ohne Daten-Governance verwendet, ist die Zuverlässigkeit von DLP-Richtlinien, Zugriffskontrollen und Prüfpfaden nicht gewährleistet. Die Schatten-KI fügt diesen Blindspots noch eine weitere Dimension hinzu: 31 % verlassen sich bei der Suche nach nicht autorisierten KI-Tools auf Protokolle, 21 % können Schatten-KI überhaupt nicht erkennen, und nur 27 % nutzen eine CASB- oder SWG-Lösung zur Echtzeiterkennung.

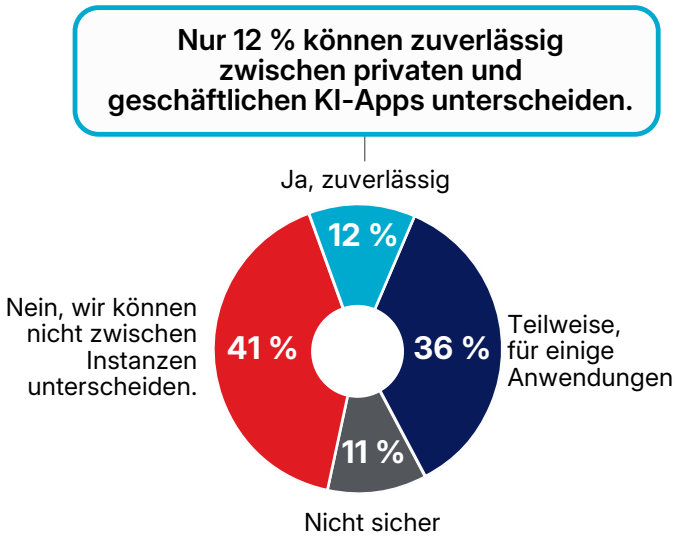
Transparenz ist meist unzureichend

Wie viel Transparenz hat Ihr Sicherheitsteam über die KI-Nutzung?



Die Identität von Instanzen ist unklar

Können Ihre Sicherheitstools zwischen privaten und geschäftlichen Instanzen von KI-Anwendungen unterscheiden (z. B. privates oder geschäftliches ChatGPT)?



Transparenz ist die grundlegende Voraussetzung, von der alle anderen Kontrollmechanismen abhängen. Bei DLP, Zugriffsrichtlinien und der Durchsetzung von Nutzungsrichtlinien wird stets davon ausgegangen, dass das Unternehmen die Aktivitäten sehen kann, die es kontrollieren möchte. Am schwierigsten ist das Monitoring der am schnellsten zunehmenden Interaktionen: API-Integrationen, MCP-basierte Agentenverbindungen (Model Context Protocol) und M2M-Kommunikation stehen bei der Einstufung nach Schwierigkeitsgrad an erster Stelle, während der direkte Chat zwischen Benutzer und KI mit 6 % den letzten Platz einnimmt.

> Zur Schließung dieser Transparenzlücke muss die Überwachung von Aktivitäten auf diese Kanäle ausgeweitet werden. Die Unterscheidung zwischen privaten und geschäftlichen KI-Konten sollte dabei als Grundlage dienen, auf die alle nachfolgenden Maßnahmen aufbauen.

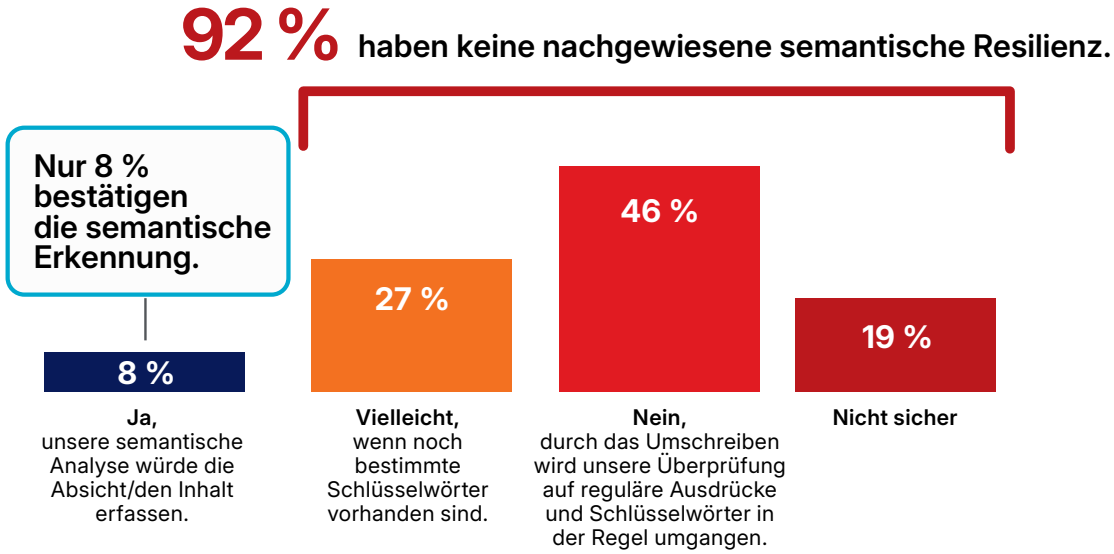
KI macht ältere DLP-Technologie unwirksam

Auch wenn Unternehmen KI-Aktivitäten erkennen können: Das wichtigste Tool zur Identifizierung von übertragenen Daten wurde für eine völlig andere Art von Bewegung entwickelt. Die DLP ist auf die Erkennung bestimmter Muster ausgelegt, wie Kreditkartenformate, Sozialversicherungsnummern, Übereinstimmungen von regulären Ausdrücken mit bekannten sensiblen Inhalten. Die DLP kann das Hochladen oder Kopieren/Einfügen sensibler Daten in Eingabefelder durch die Suche nach diesen Mustern verhindern. Sobald jedoch die KI Zugriff auf die Daten erhält, formuliert sie sensible Inhalte unter Beibehaltung ihrer Bedeutung um und verwischt dabei ihren ursprünglichen digitalen Fingerabdruck.

Der Unterschied liegt in der Architektur. Die DLP agiert auf syntaktischer Ebene und gleicht Zeichenfolgen mit vordefinierten Regeln ab. KI operiert auf semantischer Ebene. Sie wandelt Inhalte um, ohne ihre Intention zu verändern. Ein einfacher Transformationstest verdeutlicht dies: Wenn ein Mitarbeiter eine geheime Projektbeschreibung mitnimmt und eine KI bittet, das „in einer professionellen E-Mail zusammenzufassen", ersetzt die KI möglicherweise „Projekt X" durch „unsere bevorstehende strategische Initiative". Für einen Regex-Filter erscheint „strategische Initiative" vollkommen unbedenklich, obwohl der semantische Wert (das Geheimnis) identisch bleibt. Ähnliche Probleme treten auf, wenn Geheimnisse aus dem Englischen in eine andere Sprache und dann wieder zurück übersetzt werden. Die KI generiert eine Version der Daten, in der jedes ursprüngliche Schlüsselwort ersetzt wurde, doch das zugrunde liegende Risiko bleibt bestehen. Herkömmliche DLP versagt auch bei der Inferenz. Die DLP kann zum Beispiel zwei separate Listen scannen – eine mit Namen und eine mit Erkrankungen – und diese einzeln als harmlos einstufen; eine KI kann das Dokument jedoch so umformulieren, dass beide Listen explizit miteinander verknüpft sind. Dadurch entsteht ein Verstoß gegen die HIPAA-Vorschriften, den ein musterbasierter DLP-Filter nicht erkennen kann. 46 % der Befragten gaben an, dass ihre Kontrollmechanismen diese Art von Richtlinienverstößen übersehen, da beim Umschreiben die Überprüfung auf reguläre Ausdrücke und Schlüsselwörter in der Regel umgangen wird. 27 % meinten, dass die Erkennung nur dann funktioniert, wenn bestimmte Schlüsselwörter die Transformation überstehen – eine Kontrollmaßnahme, die nur dann greift, wenn der Widerpart kooperiert. Wenn man die 19 % hinzuzählt, die sich über ihren Schutz nicht sicher sind, fehlt es 92 % der Unternehmen an einer DLP-Lösung, die garantiert funktioniert, nachdem Inhalte durch KI umformuliert wurden.

Umschreiben umgeht Musterkontrollen

Der Transformationstest: Wenn ein Mitarbeiter eine KI auffordert, „dieses Dokument als allgemeinen Blogbeitrag umzuschreiben", können Ihre Sicherheitsmaßnahmen dann die sensiblen Daten im Output erkennen?



Die DLP erkennt zwar immer noch musterbasierte Verstöße, doch die von der KI veränderten Inhalte gelangen unerkannt in das System. Das Problem hat sich auf eine Ebene verlagert, für deren Überprüfung die DLP nie vorgesehen war. Die Methode zur Ermittlung des Risikos ist eindeutig: Den oben beschriebenen Transformationstest an Ihrem eigenen Stack durchführen, die KI dazu auffordern, ein vertrauliches Dokument umzuformulieren, und dann prüfen, ob Ihre Kontrollmechanismen das Ergebnis als unzulässig kennzeichnen. Dieses Ergebnis dient dann als Grundlage für die Bereitstellung einer inhaltsbezogenen Überprüfung, die die Bedeutung zum Zeitpunkt der Übertragung bewertet.

KI-Agenten agieren unbeaufsichtigt

Obwohl die meisten Unternehmen das Risiko von Datenlecks durch KI-Tools erkennen, besteht ein noch größeres Problem darin, dass KI-Systeme mittlerweile eigenständig agieren – viele davon im Hintergrund, wo sie vom Sicherheitspersonal nicht wahrgenommen werden.

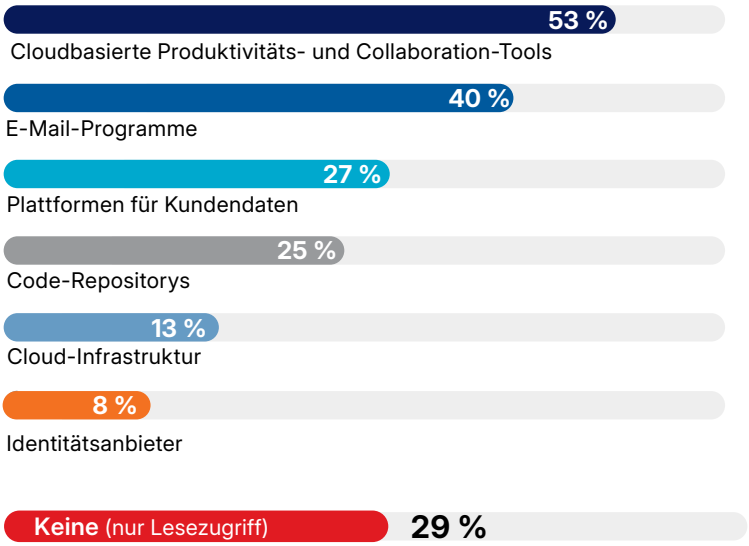
Aus der Umfrage geht hervor, in welchem Ausmaß sich dieses Problem verbreitet hat. 56 % berichten von einer realen Gefährdung durch agentische KI: 24 % in der begrenzten Produktion, 9 % im großen Maßstab beim Umgang mit Kerngeschäftslogik und 23 % durch Schatten-IT-Implementierungen, von denen die IT-Abteilung nichts weiß. 32 % haben keinerlei Transparenz über die Aktionen ihrer Agenten, und 36 % wissen überhaupt nichts über den M2M-KI-Datenverkehr.

Unternehmen, die ihre Agenten nicht sehen können, wissen auch nicht, ob sie Schattenagenten haben. 10 % geben an, agentische KI im Unternehmen verboten zu haben, dennoch berichten 23 % aller Befragten von ihrer inoffiziellen Nutzung. In der Praxis bewirken Verbote oft, dass die jeweiligen Aktivitäten nur noch im Verborgenen stattfinden. Dadurch ist es schwieriger, sie zu regulieren oder gar einzudämmen, wenn Probleme auftreten.

Die mit dem Schreibzugriff verbundenen Sicherheitsrisiken sind weitreichender, als den meisten Sicherheitsteams bewusst ist. 53 % gewähren KI-Tools Schreibzugriff auf Produktivitäts- und Collaboration-Suites in der Cloud, 40 % auf E-Mail-Programme, 25 % auf Code-Repositorys und 13 % auf die Cloud-Infrastruktur. Daneben gibt es noch Faktoren, die die Art des Risikos verändern: 8 % gewähren Identitätsanbietern Schreibzugriff. Ein Agent, der Schreibzugriff auf die Identitätsschicht hat, kann Dienstkonten erstellen, Berechtigungen auf verbundenen Systemen erweitern und sich selbst über API-Aufrufe externe Zugriffsrechte erteilen, die den Netzwerkperimeter nie überschreiten.

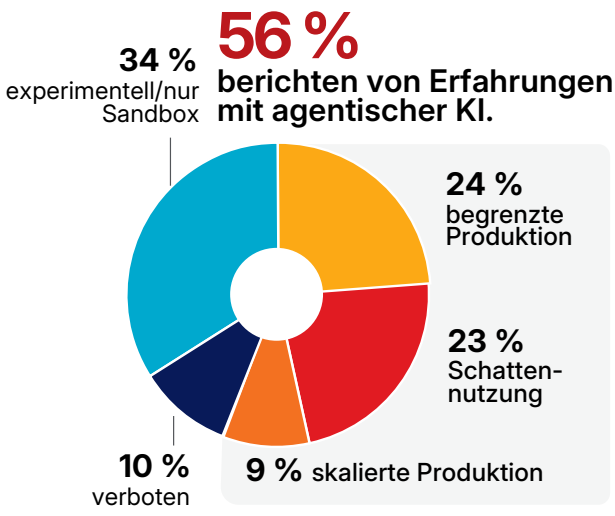
Agenten haben bereits Schreibbefugnis

Auf welche internen Systeme haben Ihre KI-Tools/Agenten Schreibzugriff?



Schattenbereitstellungen sind allgemein üblich

Wie würden Sie Ihre Nutzung von „agentischer KI“ beschreiben (KI, die selbstständig Ziele verfolgt)?



Nur 29 % der Unternehmen gewähren KI-Tools ausschließlich Lesezugriff. Für die restlichen 71 % ist die Vorgehensweise klar: Prüfen, welche KI-Tools aktuell Schreibzugriff haben, und Genehmigungsverfahren für alle Aktionen einrichten, bei denen Konten angelegt, Berechtigungen geändert oder Daten nach außen übertragen werden.

Wenn Agenten in Aktion treten, kann sie niemand aufhalten

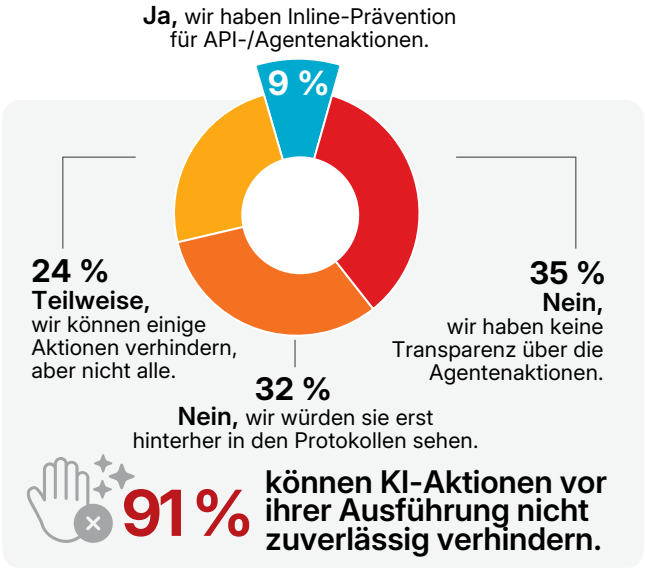
Agenten haben weitreichenden Zugriff auf Unternehmenssysteme, der kaum unterbunden werden kann. Wenn ein Agent eine schädliche Aktion in Gang setzt, können nur 9 % der Unternehmen eingreifen, bevor sie abgeschlossen ist. Die übrigen 91 % sind mehr oder weniger hilflos: 24 % können einige, aber nicht alle Agenten-Aktionen stoppen, 35 % würden erst später durch Protokolle davon erfahren und 32 % haben überhaupt keine Einblicke in Agenten-Aktionen. Unter zehn Unternehmen, die agentische KI einsetzen, gibt es nicht einmal eines, das einen Agenten daran hindern kann, ein Repository zu löschen, einen Kundendatensatz zu ändern oder eine Berechtigung zu erweitern, bevor die Aktion ausgeführt wurde.

Die Folgen zeichnen sich schon jetzt ab. Bei 37 % traten in den letzten zwölf Monaten durch KI-Agenten verursachte betriebliche Probleme auf, von denen 8 % so schwerwiegend waren, dass sie zu Systemausfällen oder Datenkorruption führten. Für 38 % ist die größte Sorge, dass ein Agent eigenmächtig Daten an einen nicht vertrauenswürdigen Ort übertragen könnte, während 24 % befürchten, dass ein Agent wichtige Konfigurationen oder Code löscht. Diese Bedenken spiegeln sich in unabhängig voneinander gemeldeten Ereignissen aus den Jahren 2025–2026 wider. Mitte 2025 führte die EchoLeak-Schwachstelle (CVE-2025-32711, CVSS 9.3) zu einer Zero-Click-Prompt-Injection bei Microsoft 365 Copilot, was die Exfiltration von Unternehmensdaten ohne Benutzereingriff ermöglichte. Anfang 2026 berichteten Sicherheitsforscher von einer neuen Angriffsmethode namens „Reprompt", bei der drei Techniken ineinandergreifen, um den KI-Assistenten Copilot Personal mit einem einzigen Klick in einen Kanal für Datenexfiltration zu verwandeln.

Irgendwo in dieser Gruppe ohne KI-Agenten-Transparenz (32 %) gibt es einen SOC-Analysten, der am Montagmorgen zur Arbeit kommt und feststellt, dass eine ungewöhnliche Änderung der Zugriffsrechte auf ein Dienstkonto vorgenommen wurde. Er verfolgt diese Änderung zu einem Agenten zurück, der dieses Konto 72 Stunden zuvor erstellt hatte, und entdeckt, dass der Agent während des ganzen Wochenendes auf Produktionssysteme geschrieben hat. Jede Aktion ist in den Protokollen ersichtlich. Es wurde keine Warnmeldung ausgelöst, da es keine Erkennungsregel für von Agenten ausgelöste Vorgänge gab.

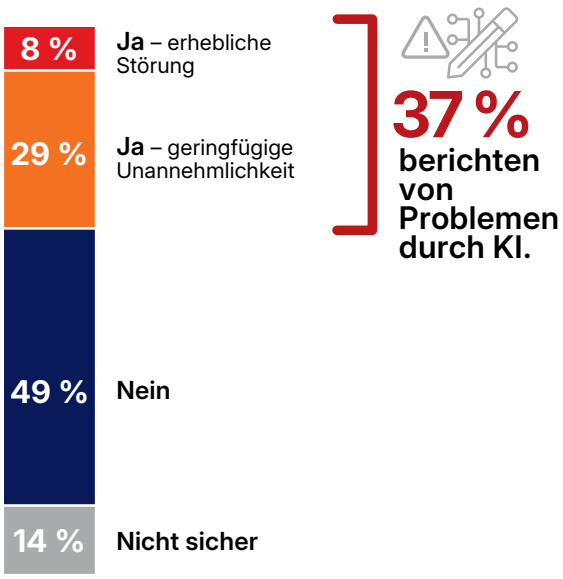
Prävention ist die Ausnahme

Kann Ihr Unternehmen eine riskante, KI-gesteuerte Aktion (wie das Löschen eines Repositorys durch einen Agenten) vorab verhindern?



Betriebliche Auswirkungen werden deutlich

Hat ein KI-Tool im vergangenen Jahr zu einer Betriebsstörung geführt?



Dieses Problem kann gelöst werden, indem Sie festlegen, welche Aktionen von Agenten in Ihrer Umgebung als anormal zu betrachten sind. Sie können Erkennungsregeln für diese Muster erstellen und für risikoreiche Aktionen von Agenten – wie die Erstellung von Konten, Änderungen an Berechtigungen und externe Datenübertragungen – eine manuelle Genehmigung durch einen Mitarbeiter vorschreiben. Wenn die Tools ausgereifter sind, sollte ein automatisiertes Abfangen auf Anforderungsebene angestrebt werden.

Zero Trust hört beim Gerät auf

91 % der Unternehmen sind nicht in der Lage, Agenten zu stoppen, bevor sie Aktionen ausführen. Der Grund ist die Architektur: Zero Trust wurde für einen Benutzer mit einem Gerät, einem Standort, einem Verhaltensmuster und einer Risikobewertung entwickelt. Ein KI-Agent hat eine Berechtigung, einen Anwendungsbereich und eine Aufgabe. 62 % wenden in irgendeiner Form Zero-Trust-Prinzipien zur Gewährleistung der KI-Sicherheit an, was der gängigste Ansatz ist. Gleichzeitig geben 65 % an, dass ihre derzeitigen Zero-Trust-Kontrollen nicht in der Lage sind, mit nicht menschlichen Identitäten (NHI) umzugehen.

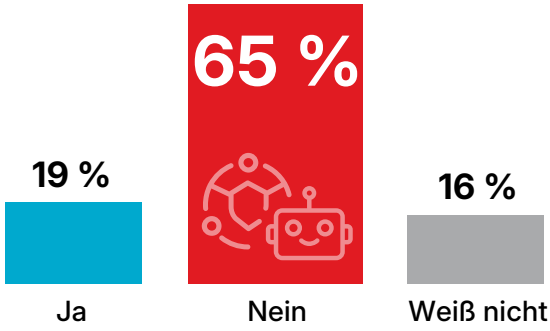
NHI-Governance schneidet in allen bewerteten Bereichen am schlechtesten ab: 61 % stufen sie als unzureichend ein, doch 78 % rechnen damit, dass die Zahl der NHI im kommenden Jahr schneller zunehmen wird als die der menschlichen Identitäten. Jeder neue Agent, Microservice und Automatisierungs-Workflow erstellt Dienstkonten und API-Schlüssel, für die eine herkömmliche Identitätsverwaltung nie vorgesehen war. Diese Frameworks wurden für Entitäten entwickelt, die mehrere Geschäftsquartale lang bestehen bleiben. Eine Agenten-Identität kann mehrere Minuten existieren.

Durch die von Agenten verwendeten Kommunikationsprotokolle entsteht eine zweite Lücke. MCP ist zu einer gängigen Schnittstelle zwischen KI-Agenten und Unternehmenstools geworden. Bei vielen derzeitigen Implementierungen hat die Interoperabilität Vorrang vor der integrierten Identitätsprüfung, der Durchsetzung des Prinzips der geringsten Berechtigungen oder der Transparenz unabhängiger Audits. Die Umfrage zeigt, dass nur 8 % der Unternehmen MCP-Richtlinien haben. Die übrigen 92 % überwachen diese entweder gar nicht oder haben noch nie davon gehört. 36 % haben keine Transparenz über den M2M-KI-Datenverkehr, und weitere 28 % verlassen sich vollständig auf die Sicherheitsmechanismen der Anbieterplattform, ohne eigene Überprüfungen durchzuführen. Nur 14 % kontrollieren den API-Datenverkehr genauso streng wie den Benutzerdatenverkehr.

Nicht menschliche Identitäten sind nicht abgedeckt

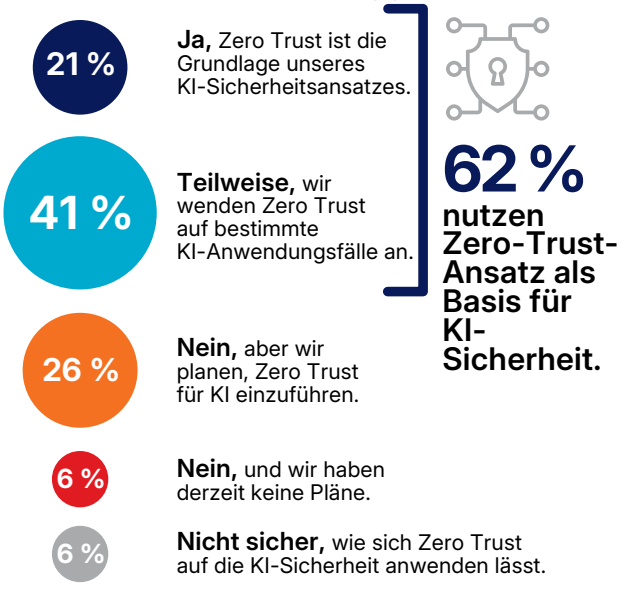
Kann Ihr Unternehmen mit Ihren derzeitigen Zero-Trust-Zugriffskontrollen auch mit nicht menschlichen Identitäten umgehen?

Die meisten Zero-Trust-Programme lassen sich nicht auf nicht menschliche Identitäten anwenden.



Zero Trust ist nach wie vor die Strategie

Basiert Ihre KI-Sicherheitsstrategie auf Zero-Trust-Prinzipien (jede Anfrage prüfen, jeden Datenfluss überwachen, Zugriff anhand einer dynamischen Risikobewertung gewähren)?



Um diesen Mangel zu beseitigen, müssen Unternehmen die Protokollebene mit der Identitätsebene zusammenführen, damit die Anmeldedaten, Anwendungsbereiche und Berechtigungen von Agenten mit derselben Sorgfalt geprüft werden wie die der menschlichen Identitäten.

Ein Großteil der KI-Sicherheit basiert auf Vertrauen

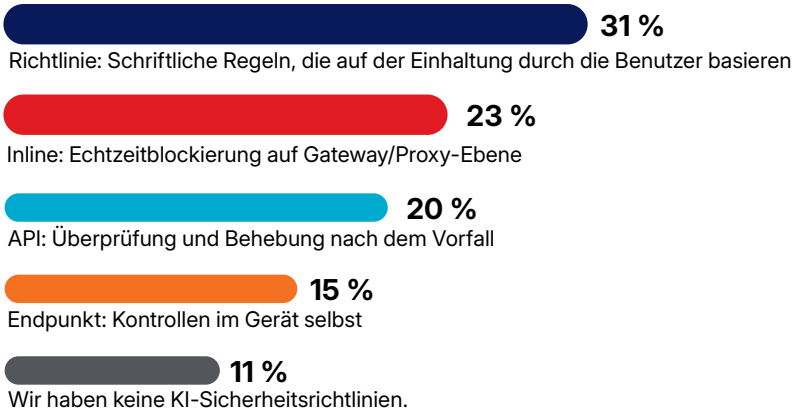
Zero Trust deckt menschliche Identitäten ab, KI-Agenten und NHI bleiben jedoch weitgehend unreguliert. Hier drängt sich eine praktische Frage auf: Wenn ein KI-Tool gegen eine Richtlinie verstößt oder ein Agent eine riskante Handlung vornimmt, welcher Durchsetzungsmechanismus greift dann tatsächlich? Die Umfrage beleuchtet das gesamte Durchsetzungsspektrum. 31 % sorgen durch schriftliche Richtlinien und Mitarbeiter-Compliance für KI-Sicherheit. Weitere 20 % verlassen sich auf API-Scans, bei denen Verstöße erst nach Abschluss der Aktion erkannt werden. 15 % führen endpunktbasierte Kontrollen durch und 23 % implementieren Inline-Kontrollen in Echtzeit. Die restlichen 11 % haben überhaupt keine KI-Sicherheitsrichtlinien. Die häufigste Form der Durchsetzung ist das „Ehrensystem". Die zweitwichtigste Methode ist die nachträgliche Überprüfung von Ereignissen.

Eine genauere Betrachtung der Daten lässt vermuten, dass selbst die 23 % der Unternehmen, die eine Inline-Durchsetzung durchführen, möglicherweise mit unzureichenden Tools arbeiten. 42 % überwachen KI-Anwendungen für ganze Plattformen nach dem Prinzip „Blockieren oder zulassen", ohne die Möglichkeit, nur Unternehmenskonten zuzulassen und Privatkonten zu sperren oder eine Suchanfrage zuzulassen und gleichzeitig das Hochladen eines Finanzmodells zu blockieren. Nur 19 % haben differenzierte Kontrollmechanismen auf Aktivitätsebene, die zwischen den Aktionen in einer zugelassenen Anwendung unterscheiden. Sofern Echtzeitkontrollen vorhanden sind, richten sich diese auf menschliche Handlungen: An erster Stelle steht die Blockierung von Datei-Uploads (48 %) sowie die Erkennung von Einfügungen (37 %), während das Veröffentlichen von Inhalten (29 %) und Download-Kontrollen (25 %) weiter hinten liegen. Von Agenten initiierte API-Aufrufe, OAuth-Token-Austausch und M2M-Datenflüsse werden kaum berücksichtigt.

Diese Umsetzungslücke zeugt von mangelnder Kohärenz. Wenn der CASB, die DLP-Engine und die Zugriffsrichtlinie nur ein bruchstückhaftes Bild erhalten, kann keines dieser Tools seine eigentliche Funktion vollständig erfüllen. Ein einfacher Diagnosetest zeigt: Wenn für eine einzige Richtlinienentscheidung Daten von mehr als zwei Konsolen benötigt werden, führt diese Fragmentierung zu einer unzureichenden Durchsetzung. Zur Schließung dieser Lücke müssen diese Ebenen zusammengeführt werden. Dann kann eine einzige Bewertung anhand der Inhaltsklassifizierung, der Benutzeridentität und des KI-Instanztyps durchgeführt werden, bevor eine Aktion abgeschlossen wird.

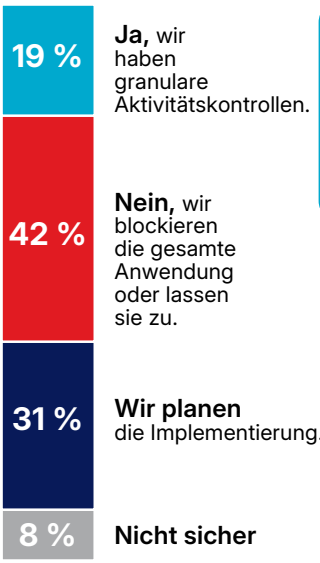
Die Durchsetzung erfolgt meist nachträglich

Wie werden Ihre KI-Sicherheitsrichtlinien überwiegend durchgesetzt?



Kontrollen wirken noch nicht

Setzen Sie unterschiedliche Richtlinien für „Upload" und „Chat" durch?



Nur 19 % haben detaillierte Upload- und Chat-Kontrollfunktionen.

KI-Sicherheit kann nicht auf bestehende perimeterbasierte Modelle aufgesetzt werden. Sie erfordert eine cloudnative Inline-Durchsetzung, bei der Identität, Inhalt und Kontext vor der Ausführung an einem einzigen Entscheidungspunkt geprüft werden.

Wie ausgereift ist Ihre KI-Sicherheit?

Jede der bisher untersuchten Schwachstellen verstärkt die anderen: Eine mangelhafte Transparenz untergräbt die DLP, unkontrollierte Agenten umgehen Zugriffskontrollen und eine fragmentierte Durchsetzung lässt jede Ebene ungeschützt. Das nachstehende Reifegradmodell ordnet sechs zentrale Bereiche der KI-Sicherheit drei Reifegraden zu. In jedem Feld werden die Fähigkeiten in der jeweiligen Phase beschrieben. Suchen Sie in jeder Spalte nach der Beschreibung, die auf Ihr Unternehmen zutrifft. Der Bereich mit dem geringsten Reifegrad ist das schwächste Glied in der Kette und der wahrscheinlichste Schwachpunkt. Dorthin sollten Investitionen zuerst fließen.

BEREICHE DER KI-SICHERHEIT	REAKTIV	VERWALTET	ADAPTIV
Governance und Risiko-ausrichtung	Richtlinien auf dem Papier. Uneinheitliche Durchsetzung.	Richtlinien für verwaltete KI-Tools durchgesetzt. Schatten-KI unkontrolliert.	Richtlinien in technische (Echtzeit-)Kontrollmechanismen integriert.
Transparenz und Situations-bewusstsein	Eingeschränkte oder gar keine Transparenz. Private und geschäftliche Instanzen können nicht unterschieden werden.	Übersicht über verwaltete SaaS und APIs auf Aktivitätsebene. Unterscheidung zwischen privaten und geschäftlichen Instanzen.	Echtzeittransparenz über alle KI-Workflows, einschließlich Agenten und Maschine-zu-Maschine-Interaktionen.
Daten- und Asset-Schutz	Musterbasierte Kontrollmechanismen funktionieren bei der KI-Transformation nicht mehr.	DLP auf KI-Datenverkehr ausgeweitet, einschließlich Chats und Prompts. Semantische Erkennung in der Pilotphase.	Semantische Überprüfung aller KI-Datenflüsse.
Zugriffs- und Ausführungs-kontrolle	Durchsetzung nach der Ausführung. Ehrensystem zur Durchsetzung von Richtlinien.	Inline-Durchsetzung für von Menschen initiierte KI. Agentenaktionen protokolliert, nicht blockiert.	Dynamische Durchsetzung vor der Ausführung (menschlich und nicht menschlich).
Erkennung und Reaktion	Protokollbasierte Überwachung. Es gibt keine spezielle Erkennungslogik für KI-gesteuertes Verhalten.	Erkennungsregeln für bekannte KI-Missbrauchsmuster. Manuelle Eindämmung.	Kontinuierliche Überwachung mit automatisierter Eindämmung für alle KI-Aktivitäten.
Architektur-integration und betriebliche Resilienz	Fragmentiert. Transparenz, Schutz und Durchsetzung in Silos.	Kontrollen für verwaltete SaaS integriert. Lücken im API- und Agentendatenverkehr.	Einheitliches Framework, das auch bei Automatisierung und Skalierung robust ist.

Auf die Frage, was bei der KI-Einführung falsch gelaufen war, wünschten sich 38 %, dass Governance-Maßnahmen vor der umfassenden Einführung von KI getroffen worden wären, und 25 % hätten gern früher in Maßnahmen zur Transparenzkontrolle investiert. Lediglich 7 % sind mit ihrem derzeitigen Ansatz zufrieden – dies ist der niedrigste Vertrauensindikator in der gesamten Umfrage.

Veränderungen werden forciert

52 %
sagen, dass Vorschriften Veränderungen forcieren

47 %
sagen, dass Verstöße Veränderungen forcieren

Die Umsetzungslücke schließen

Das Reifegradmodell verdeutlicht, wo es Lücken gibt. Im Folgenden werden die wichtigsten Verbesserungsmaßnahmen für jeden Risikovektor zusammengefasst, beginnend mit der Transparenz – der Grundlage für alle weiteren Kontrollmaßnahmen.

1. **Die KI-Transparenzlücken schließen:** 94 % berichten von Lücken; 88 % können nicht zwischen privaten und geschäftlichen Konten unterscheiden.
   Weiten Sie die Überwachung auf Aktivitätsebene auf den SaaS-, API- und M2M-Datenverkehr aus – beginnend mit der Unterscheidung zwischen privaten und geschäftlichen KI-Konten. Dies ist die Voraussetzung für zuverlässige DLP, Zugriffskontrollen und Prüfpfade.

2. **Richtlinien in durchsetzbare Leitplanken umsetzen:** 68 % agieren reaktiv; nur 7 % setzen Maßnahmen in Echtzeit durch.
   Ermitteln Sie die drei risikoreichsten KI-Anwendungsfälle in Ihrer Umgebung, integrieren Sie durchsetzbare Richtlinien in technische Kontrollmaßnahmen dafür und weisen Sie jedem einen Verantwortlichen zu, bevor Sie die Abdeckung auf alle verbleibenden KI-Anwendungsfälle ausweiten.

3. **Semantischen Datenschutz implementieren:** 46 % scheitern am Inhaltsumwandlungstest.
   Führen Sie den Inhaltsumwandlungstest mit Ihrem eigenen DLP-System durch: Nehmen Sie ein vertrauliches Dokument, lassen Sie es von einem KI-Tool umformulieren und prüfen Sie, ob Ihre Kontrollmechanismen das Ergebnis als verdächtig markieren. Dieses Ergebnis dient dann als Grundlage für die Bereitstellung einer kontextbezogenen Überprüfung, die die Bedeutung zum Zeitpunkt der Übertragung bewertet.

4. **Vor der Ausführung durchsetzen:** 23 % setzen Maßnahmen inline durch; 9 % können eine riskante Aktion eines Agenten im Voraus verhindern.
   Prüfen Sie, welche KI-Agenten aktuell Schreibzugriff haben, und richten Sie Genehmigungsverfahren für alle Aktionen ein, bei denen Konten angelegt, Berechtigungen geändert oder Daten nach außen übertragen werden.

5. **Erkennung und Eindämmung modernisieren:** 67 % verlassen sich auf Protokolle oder haben keine Transparenz über die Aktionen der Agenten; 37 % haben bereits KI-bedingte Betriebsprobleme erlebt.
   Bestimmen Sie, welche Verhaltensweisen bei Agenten in Ihrer Umgebung als anormal zu betrachten sind, und erstellen Sie Erkennungsregeln für diese Muster. Erstellen Sie Playbooks zur Eindämmung von Sicherheitsvorfällen, die auf Anforderungsebene eingreifen können, bevor eine Aktion abgeschlossen ist, anstatt erst dann ein Ticket zu erstellen, wenn der Schaden bereits entstanden ist.

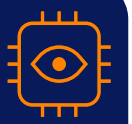
6. **Fragmentierung der Kontrolle verringern:** 42 % nutzen binäre Kontrollen nach dem Prinzip „Blockieren oder Zulassen“ ohne Unterscheidung nach Aktivitätsebene; nur 14 % überprüfen den API-Datenverkehr mit derselben Sorgfalt wie den Benutzerdatenverkehr.
   Vereinheitlichen Sie CASB, DLP und Zugriffsrichtlinien, damit eine einzige Bewertung auf Inhaltsklassifizierung, Benutzeridentität und den Typ der KI-Instanz zurückgreifen kann. Wenn für eine einzige Richtlinienentscheidung Daten von mehr als zwei Konsolen benötigt werden, führt diese Fragmentierung zu einer unzureichenden Durchsetzung.

KI-Sicherheit ist heute eine eigene betriebliche Disziplin. Reifegrade werden bestimmt, die Abhängigkeiten sind klar und konkrete Maßnahmen werden festgelegt. Nun bleibt nur noch die Entscheidung, darauf aufbauen.

Governance hinkt hinterher

68 %
reagieren anhand von oder entwickeln Governance- und Sicherheitsrichtlinien für KI-Implementierungen und Daten

Transparenz ist unzureichend



94 %
haben keine vollständige Transparenz über die KI-Nutzung

Datenkontrollen greifen nach dem Umschreiben nicht

Nur 8 % bestätigen semantische Erkennung von Absicht/Inhalt

Abfangen kommt selten vor

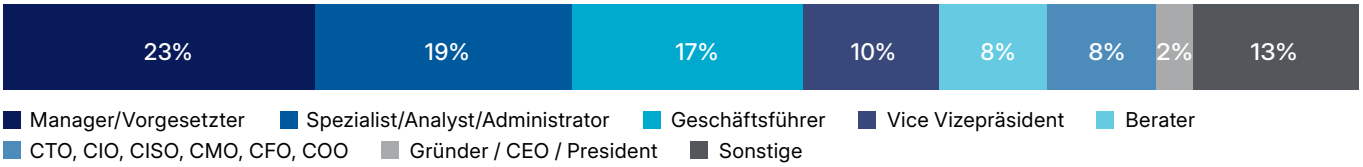
Nur 9 % haben Inline-Prävention für API-/Agentenaktionen

Methodik und demografische Daten

Dieser Bericht basiert auf einer Anfang 2026 unter 1.253 Cybersicherheits- und IT-Fachkräften durchgeführten Umfrage. Befragt wurden Sicherheitsexperten, Architekten und Technologieführer, die für den Schutz von Unternehmensinfrastrukturen, Cloud-Umgebungen und KI-gesteuerten Anwendungen in den unterschiedlichsten Branchen und Unternehmen jeder Größe verantwortlich sind.

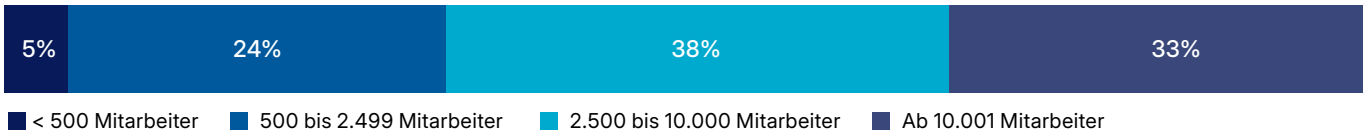
Die Studie untersucht, wie Unternehmen KI-Implementierungen absichern. Dabei liegt der Schwerpunkt auf dem Reifegrad der Governance, der Transparenz von KI-Aktivitäten, dem Datenschutz, der Verwaltung nicht menschlicher Identitäten und der Steuerung autonomer Agenten. Durch die Anwendung eines geschichteten Stichprobenverfahrens erreichte die Umfrage ein Konfidenzniveau von 95 % bei einer Fehlermarge von +/- 2,8 %.

KARRIERESTUFE

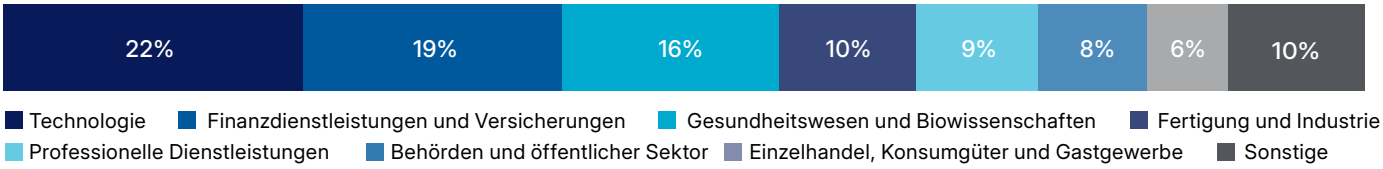


ABTEILUNG

UNTERNEHMENSGRÖSSE



BRANCHE



Über Netskope

Netskope (NASDAQ: NTSK), ein führender Anbieter moderner Sicherheits- und Netzwerklösungen für die Cloud- und KI-Ära, erfüllt die Anforderungen von Sicherheits- und Netzwerkteams gleichermaßen. Das Unternehmen bietet optimierten Zugriff sowie kontextbasierte Sicherheit in Echtzeit für das gesamte KI-Ökosystem – einschließlich Agenten, Anwendungen, Tools, LLMs, Nutzern, Geräten und Daten.

Tausende Kunden, darunter mehr als 30 der Fortune 100, vertrauen auf die Netskope One Plattform, ihre Zero Trust Engine und das leistungsstarke NewEdge-Netzwerk, um Risiken zu reduzieren und vollständige Transparenz sowie Kontrolle über Cloud-, KI-, SaaS-, Web- und private Anwendungen zu gewinnen – und das bei gleichzeitiger Verbesserung der Performance ohne Kompromisse bei der Sicherheit.

Weitere Informationen finden Sie unter

netskope.com/de/products/ai-security

Cybersecurity
INSIDERS

VERGLEICHEN SIE IHREN SICHERHEITSREIFEGRAD

Unabhängige Cybersicherheitsforschung zur Aufdeckung der Schwachstellen in Cybersicherheitsstrategien

Cybersecurity Insiders erstellt unabhängige Studien auf der Grundlage von Umfragen unter führenden Anbietern und Fachleuten für Cybersicherheit weltweit. Unsere Berichte zeigen, wo Sicherheitsstrategien in der Praxis versagen, und helfen Unternehmen dabei, ihren Reifegrad zu bewerten, Kompetenzlücken zu erkennen und Maßnahmen zu ihrer Schließung zu priorisieren.
Weitere Informationen finden Sie unter:

cybersecurity-insiders.com

April 2026 // Version 1.7