



# Einblicke für den öffentlichen Sektor: ThreatLabz–Report 2024 zur KI–Sicherheit



Die KI–Revolution hat sowohl die Art und Weise verändert, wie Cyberangriffe ausgeführt werden, als auch die Verteidigungsmaßnahmen von Behörden. Entdecken Sie wichtige Trends, Risiken und Best Practices bei der Einführung von KI im öffentlichen Sektor, mit Einblicken in KI–gesteuerte Bedrohungen und wichtige Verteidigungsstrategien.

# Zusammenfassung für den öffentlichen Sektor

## Ein Katalysator für Innovationen und Bedrohungen

Tools für künstliche Intelligenz (KI) und maschinelles Lernen (ML) werden die Regierungs- und Geschäftslandschaft revolutionieren, wobei KI zunehmend in die Strukturen des öffentlichen und privaten Sektors integriert wird. Allerdings hat sich KI als zweischneidiges Schwert erwiesen: Sie dient sowohl als Katalysator für Innovationen als auch als wirksames Werkzeug für Akteure und Gruppen, die Cyberangriffe verüben.

Im öffentlichen Sektor bietet KI vielversprechende Möglichkeiten die Art und Weise, wie wir Dienstleistungen erbringen und gesellschaftliche Herausforderungen bewältigen, neu zu gestalten. Gleichzeitig nutzen Cyberkriminelle und staatlich geförderte Bedrohungsgruppen KI-gestützte Techniken, um hochentwickelte Angriffe in beispiellosem Ausmaß und mit zunehmender Häufigkeit zu orchestrieren. Die Verbreitung dieser KI-gestützten Cyberbedrohungen, die von Phishing-Kampagnen bis hin zu polymorpher Ransomware reichen, stellt im öffentlichen Sektor erhebliche Herausforderungen für die Cybersicherheit dar und hat signifikante wirtschaftliche Folgen. Organisationen des öffentlichen Sektors müssen proaktive Maßnahmen ergreifen, um kritische Infrastrukturen und vertrauliche Informationen zu schützen und gleichzeitig die sichere Nutzung von KI zu ermöglichen.

Regierungen und Unternehmen reagieren entsprechend: Weltweit entwickeln politische Entscheidungsträger aktiv regulatorische Rahmenbedingungen für den verantwortungsvollen und sicheren Einsatz von KI. Gleichzeitig entwickelt sich die KI zu einem Eckpfeiler der Strategien und Ausgaben im

Bereich Cybersicherheit. Sie ermöglicht es Unternehmen, das Schutzniveau und ihre Anpassungsfähigkeit an eine dynamische Bedrohungslandschaft zu verbessern.

Um die neuesten Trends bei der Nutzung von KI und ML zu erfassen und Hinweise für eine sichere Einführung zu geben, untersuchte das Forschungsteam von Zscaler ThreatLabz mehr als 18 Milliarden AL/ML Transaktionen über die Zscaler Zero Trust Exchange™, die weltweit größte Inline Security Cloud. Das Team gewann dabei unter anderem folgende Erkenntnisse:

## Die Nutzung von KI-Tools nimmt rasant zu

Von April 2023 bis Januar 2024 verzeichnete ThreatLabz AI/ML einen Anstieg der Transaktionen um fast 600 %. Im Januar stiegen die Transaktionen über die Zero Trust Exchange auf 3,1 Milliarden. Mehr als die Hälfte der in diesem Zeitraum beobachteten Transaktionen entfallen auf ChatGPT.

Obwohl die Akzeptanz für KI/ML steigt, blockieren Unternehmen zunehmend KI/ML-gestützte Transaktionen. ThreatLabz beobachtete innerhalb von neun Monaten einen Anstieg der blockierten Transaktionen um 577 %, was auf wachsende Bedenken hinsichtlich der Datensicherheit hindeutet.

Das Gesamtwachstum der  
KI/ML-gestützten Transaktionen

594,8 %

An KI-Tools gesendete Daten

569 TB

Die Vereinigten Staaten  
erzeugen die meisten  
KI/ML-Transaktionen

40 %

# Wichtige Erkenntnisse für den öffentlichen Sektor

## Globale Regierungen definieren KI-Praktiken und -Richtlinien

Um die Leistungserbringung und die politische Gestaltung zu verbessern, integrieren staatliche Behörden auf der ganzen Welt zunehmend KI-Technologien in ihre Betriebsabläufe. Unter den Top-10-Treibern von KI/ML-Transaktionen in der Zscaler-Cloud blockiert der Regierungsbereich nur 6,75 % der KI-Transaktionen.

Wichtige Anwendungen wie ChatGPT und Drift erfreuen sich bei der Regierung zunehmender Beliebtheit und ermöglichen es Behörden, effektiver mit den Bürgern zu kommunizieren und Entscheidungsprozesse zu optimieren. Insbesondere Weiterentwicklungen wie Chatbots und virtuelle Assistenten optimieren die Interaktion mit den Bürgern und bieten einen schnelleren Zugriff auf wichtige Informationen und Dienste in verschiedenen Sektoren.

Doch während Regierungen das Potenzial von KI nutzen, bleiben Bedenken hinsichtlich Datenschutz und -sicherheit bestehen. Die Ergebnisse von ThreatLabz unterstreichen die Notwendigkeit, diesem Thema Priorität einzuräumen, insbesondere da im Regierungsbereich im Vergleich zum weltweiten Durchschnitt (18,5 %) und Branchen wie Finanzen und Versicherungen (37,2 %) ein deutlich geringerer Anteil von KI-Transaktionen blockiert wird. Um diese Herausforderungen verantwortungsvoll zu meistern, ist die Schaffung von regulatorischen Rahmenbedingungen und Governance-Mechanismen unabdingbar. Politische Entscheidungsträger weltweit ergreifen Maßnahmen, um diese Bedenken auszuräumen, und betonen die Bedeutung einer verantwortungsvollen Entwicklung und Bereitstellung von KI. Weitere Informationen finden Sie nachstehend im Artikel „[All Eyes on AI Regulations](#)“.

### VORHERSAGE

**Das KI-Dilemma der Nationalstaaten:** In Zukunft wird erwartet, dass staatlich geförderte Bedrohungsgruppen KI nutzen werden, um ausgefeiltere Cyberbedrohungen zu entwickeln. Gleichzeitig werden die Technologien dazu dienen, den Zugriff auf regimefeindliche Inhalte zu blockieren. Dieser zweigleisige Ansatz unterstreicht die komplexe Beziehung zwischen KI und nationaler Sicherheit, wobei Bedrohungsakteure KI-Funktionen für böswillige Zwecke ausnutzen. Lesen Sie mehr über [diese Vorhersage und andere KI-Bedrohungsvorhersagen](#) im folgenden Bericht.

## Der Bildungsbereich nutzt KI, um das Lernerlebnis zu verbessern

Im Bildungssektor dient KI als wertvolles Lerninstrument, auch wenn der Sektor nicht für den größten Anteil von KI-Traffic in der Zscaler-Cloud verantwortlich ist. Mit einer vergleichsweise niedrigen Blockrate von 2,98 % nutzen Bildungseinrichtungen KI-Anwendungen vorwiegend zur Verbesserung des Lernerlebnisses. Insbesondere Anwendungen wie ChatGPT und Character.AI erleichtern kreative Ergebnisse und unterstützen Schüler beim Schreiben und bei der Bilderzeugung.

Während einige Pädagogen Richtlinien zur Einschränkung bestimmter KI-Anwendungen wie ChatGPT im Unterricht einführen, integrieren Institutionen insgesamt KI-Tools, um die Lernumgebung zu bereichern. Mit der zunehmenden Verbreitung von KI werden jedoch wahrscheinlich auch die Bedenken hinsichtlich des Datenschutzes zunehmen, was Bildungseinrichtungen dazu veranlassen wird, strenge Maßnahmen für die Data Protection einzuführen.

**Wenn öffentliche Einrichtungen, Regierungen und Bildungseinrichtungen das transformative Potenzial der KI-Technologien nutzen, hat der Schutz kritischer Infrastrukturen, vertraulicher Daten, Bürgerinformationen und der Privatsphäre von Schülern höchste Priorität.**

**Die vollständigen Ergebnisse und Best Practices für die sichere Umsetzung der KI-Transformation finden Sie im folgenden ThreatLabz 2024 AI Security Report.**



# Zscaler ThreatLabz– Report 2024 zur KI–Sicherheit



Die KI–Revolution ist in vollem Gange. Entdecken Sie wichtige Trends, Risiken und Best Practices im Zusammenhang mit der KI–Einführung im Unternehmen, einschließlich Einblicken zu KI–gestützten Bedrohungen und wichtigen Strategien zu deren Abwehr.



# Inhalt

## 03 Kurzfassung

## 04 Haupterkenntnisse

## 05 Wichtige Nutzungstrends bei GenAI und ML

- 05 KI-Transaktionen nehmen weiter zu
- 06 Unternehmen blockieren mehr KI-Transaktionen als je zuvor
- 07 **KI-Aktivitäten nach Branche**
- 09 Gesundheitswesen und KI
- 10 Finanzwesen und KI
- 11 Behörden und KI
- 12 Fertigung und KI
- 13 Bildungswesen und KI
- 14 **Nutzungstrends bei ChatGPT**
- 15 **KI-Nutzung nach Ländern**
  - Aufschlüsselung nach Region:
  - EMEA Aufschlüsselung nach Region: APAC

## 18 Anwendungsbezogene KI-Risiken und Bedrohungsszenarien

- 18 KI-Umsetzung im Unternehmen: Die drei größten Risiken
- 20 KI-bezogene Bedrohungsszenarien
  - KI-Imitation: Deepfakes, Desinformation und mehr
- 21 KI-generierte Phishing-Kampagnen Von der Anfrage zum Angriff: Erstellung einer Phishing-Anmeldeseite mit ChatGPT
- 22 Chatbots im Darknet: WormGPT und FraudGPT auf der Spur

23 KI-gestützte Malware und Ransomware in sämtlichen Angriffsphasen

24 KI-Wurmangriffe und „virales“ KI-Jailbreaking

25 Die Rolle von KI bei den US-Wahlen

## 26 KI-Regulierung im Fokus

26 USA

27 EU

## 28 Prognosen zu KI-Bedrohungen

## 31 Fallstudie: Sichere Nutzung von ChatGPT im Unternehmen

31 5 Schritte zur Integration und Absicherung generativer KI-Tools

## 33 So ermöglicht Zscaler die sichere Bereitstellung generativer KI mit Zero Trust

33 Der Schlüssel zur KI-gestützten Cybersicherheit:  
Hochwertige Daten in großem Maßstab

34 Nutzung von KI in sämtlichen Angriffsphasen

35 Zusammenfassung der KI-gestützten Lösungen von Zscaler

36 Erfolgreicher KI-Umstieg im Unternehmen: Es liegt in Ihren Händen

## 37 Anhang

37 Forschungsmethodik von ThreatLabz

## 37 Über Zscaler ThreatLabz

# Kurzfassung

Künstliche Intelligenz (KI) ist weit mehr als eine bahnbrechende Innovation — sie ist mittlerweile fester Bestandteil des Alltags. Generative KI-Tools wie ChatGPT sind dabei, Unternehmen sowohl im Großen als auch im Kleinen zu verändern, und durchdringen das Geschäftsleben dabei mehr und mehr. Allerdings besteht noch erheblicher Klärungsbedarf bei der Frage, wie diese KI-Tools sicher implementiert und KI-gestützte Bedrohungen gleichzeitig abgewehrt werden können.

KI- und ML-Tools werden derzeit zunehmend in verschiedene Unternehmensabteilungen wie Engineering, IT-Marketing, Finanzwesen und Customer Success eingeführt. Um den größtmöglichen Nutzen aus diesen KI-Tools zu ziehen, müssen Unternehmen jedoch den zahlreichen Risiken entgegenwirken, die mit ihnen verbunden sind. Mit anderen Worten: Das transformative Potenzial der KI lässt sich nur dann ausschöpfen, wenn Unternehmen sichere Kontrollen zum Schutz ihrer Daten einrichten, den Verlust sensibler Informationen verhindern, die Ausbreitung von „Schatten-KI“ eindämmen und die Qualität der KI-Daten sicherstellen.

Diese KI-Risiken für Unternehmen sind bidirektional: **Außerhalb der Unternehmensgrenzen ist KI zu einer treibenden Kraft für Cyberbedrohungen geworden.** Tatsächlich ermöglichen KI-Tools es Cyberkriminellen und staatlich geförderten Bedrohungsakteuren, komplexere Angriffe schneller und in größerem Umfang durchzuführen. Gleichzeitig gilt KI angesichts der dynamischen Bedrohungslandschaft als Schlüsselement moderner Cybersicherheitsmaßnahmen.

Der ThreatLabz-Report 2024 zur KI-Sicherheit bietet wichtige Einblicke in diese kritischen KI-bezogenen Herausforderungen und Chancen. Anhand von mehr als 18 Milliarden Transaktionen von April 2023 bis Januar 2024 über die Zscaler Zero Trust Exchange™ analysierte ThreatLabz, wie Unternehmen heute KI- und ML-Tools nutzen. Diese Erkenntnisse geben Aufschluss darüber, wie sich Unternehmen in verschiedenen Branchen und Regionen an die sich verändernde KI-Landschaft anpassen und ihre KI-Tools schützen.

Der Bericht enthält Einblicke in aktuelle KI-Themen, darunter Geschäftsrisiken, KI-gestützte Bedrohungsszenarien und Angriffstaktiken, regulatorische Überlegungen und Prognosen zur KI-Landschaft im Jahr 2024 und darüber hinaus.

Außerdem bietet er erfolgsentscheidende Best Practices in Bezug auf zwei zentrale Herausforderungen: die sichere Umsetzung der generativen KI-Transformation bei gleichzeitigem Schutz kritischer Daten auf der einen Seite sowie den Einsatz von KI-gestützten Tools zur Bereitstellung mehrschichtiger Zero-Trust-Sicherheit im Kampf gegen neuartige KI-gestützte Bedrohungen auf der anderen.

# Haupterkenntnisse



Bei der Nutzung von KI- und ML-Tools lässt sich ein **sprunghafter Anstieg um 594,82 % beobachten** — waren es im April 2023 noch 521 Millionen KI/ML-gestützte monatliche Transaktionen, lag diese Zahl im Januar 2024 bereits bei 3,1 Milliarden.



**Unternehmen blockieren 18,5 % aller KI/ML-Transaktionen — ein Anstieg um 577 % im Laufe der neun Monate** und Ausdruck sowohl wachsender Bedenken hinsichtlich der KI-Datensicherheit als auch der Zurückhaltung von Unternehmen bei der Einführung von KI-Richtlinien.



In der Fertigung wird mit einem Anteil von **20,9 % an allen KI/ML-Transaktionen in der Zscaler Cloud der meiste KI-Traffic generiert**, gefolgt von der Finanz- und Versicherungs- (19,9 %) sowie der Dienstleistungsbranche (16,8 %).



Die ChatGPT-Nutzung nimmt weiter zu: Das **Wachstum liegt hier bei 634,1 %**. Gleichzeitig ergeben die Daten der Zscaler Cloud, dass es sich auch um **die von Unternehmen am häufigsten blockierte KI-Anwendung handelt**.



Die nach Transaktionsvolumen am häufigsten verwendeten **KI-Anwendungen** sind **ChatGPT, Drift, OpenAI\*, Writer und LivePerson**. Die **drei am häufigsten blockierten Anwendungen** nach Transaktionsvolumen sind **ChatGPT, OpenAI und Fraud.net**.



Die **fünf Länder** mit den meisten KI- und ML-Transaktionen sind die **USA, Indien, Großbritannien, Australien und Japan**.



Unternehmen senden erhebliche Datenmengen an **KI-Tools: Insgesamt 569 TB** wurden von September 2023 bis Januar 2024 zwischen KI/ML-Anwendungen ausgetauscht.



**KI stärkt Bedrohungsakteure auf nie dagewesene Art und Weise** und ermöglicht etwa KI-gestützte Phishing-Kampagnen, Deepfakes und Social-Engineering-Angriffe, polymorphe Ransomware, Erkennung von Unternehmensangriffsflächen und automatisierte Exploit-Generierung.

HINWEIS : Die Zscaler Zero Trust Exchange erfasst ChatGPT-Transaktionen unabhängig von anderen OpenAI-Transaktionen insgesamt.

# Wichtige Nutzungstrends bei GenAI und ML

Die KI-Revolution im Unternehmen hat ihren Höhepunkt noch lange nicht erreicht. KI-Transaktionen von Unternehmen haben um fast 600 % zugenommen und ein Rückgang ist nicht in Sicht. Doch auch die Anzahl blockierter Transaktionen im Zusammenhang mit KI-Anwendungen ist gestiegen — und zwar um 577 %.

## KI-Transaktionen nehmen weiter zu

Von April 2023 bis Januar 2024 stieg die Anzahl der KI- und ML-Transaktionen von Unternehmen um nahezu 600 % und lag im Januar bei mehr als 3 Milliarden monatlicher Transaktionen in der Zero Trust Exchange. Dies macht deutlich, dass das transformative Potenzial von KI trotz einer steigenden Zahl von Sicherheitsvorfällen und Datenrisiken im Zusammenhang mit ihrer Einführung in Unternehmen zu groß ist, um es ungenutzt zu lassen. Während der Dezemberfeiertage gingen die KI-Transaktionen zwar vorübergehend zurück, zu Beginn des Jahres 2024 setzten sie ihr Wachstum dafür umso rasanter fort.

Trotz der zunehmenden Verbreitung von KI-Anwendungen gehen die meisten KI-Transaktionen jedoch von einer relativ kleinen Anzahl marktführender KI-Tools aus. Insgesamt generiert ChatGPT mehr als die Hälfte aller KI- und ML-Transaktionen, während die OpenAI-Anwendung selbst mit 7,82 % aller Transaktionen den dritten Platz belegt. Drift, der beliebte KI-gestützte Chatbot, ist für fast ein Fünftel des KI-Traffics in Unternehmen verantwortlich (die Chatbots LivePerson und BoldChat Enterprise landeten mit den Plätzen 5 und 6 ebenfalls unter den transaktionsstärksten Anwendungen). Das generative KI-Tool Writer kommt nach wie vor häufig bei der Erstellung schriftlicher Unternehmensinhalte wie Marketingmaterialien zum Einsatz. Schließlich macht auch Otter, ein KI-Transkriptionstool, das häufig bei Videoanrufen verwendet wird, einen erheblichen Teil des KI-Traffics aus.

Trends bei KI- und ML-Transaktionen



ABBILDUNG 1 KI-Transaktionen von April 2023 bis Januar 2024

Meistgenutzte KI-Anwendungen

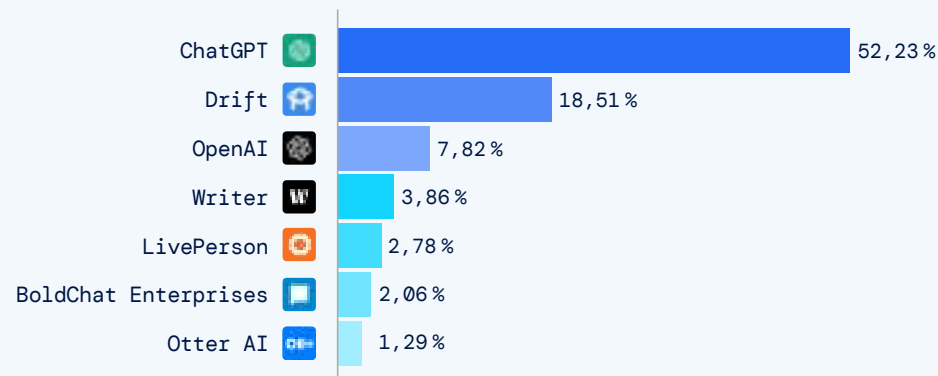


ABBILDUNG 2 Meistgenutzte KI-Anwendungen nach Transaktionsvolumen



### Durch KI/ML-Traffic übertragene Daten [Sep. 2023—Jan. 2024]

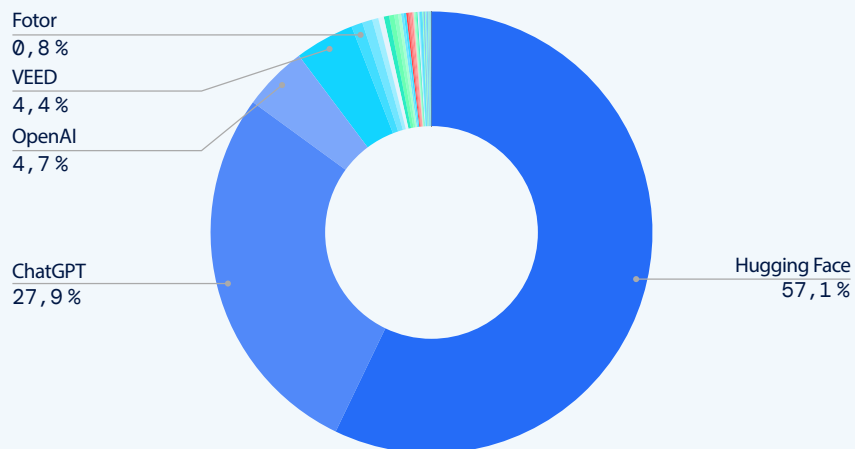


ABBILDUNG 3 Meistgenutzte KI/ML-Anwendungen nach Anteil der insgesamt übertragenen Daten

### Trends bei blockierten KI-Transaktionen [Apr. 2023—Jan. 2024]



ABBILDUNG 4 Anzahl der im Laufe der Zeit blockierten KI/ML-Transaktionen

Ein differenzierteres Bild dieser Trends zeigt sich bei Betrachtung der Datenmengen, die Unternehmen von KI-Tools senden und empfangen. Die Open-Source-KI-Entwicklerplattform Hugging Face, die oft als „GitHub für KI“ bezeichnet wird, macht fast 60 % der von KI-Tools übertragenen Unternehmensdaten aus. Da Hugging Face es Usern ermöglicht, KI-Modelle zu hosten und zu trainieren, ist es nur logisch, dass die Plattform erhebliche Datenmengen von Unternehmensanwendern erfasst.

Dass ChatGPT und OpenAI Teil dieser Liste sind, dürfte kaum überraschen. Bemerkenswert sind die Nennungen von Veed — einem KI-Videoeditor zum Hinzufügen von Untertiteln, Bildern und sonstigem Text zu Videos — und Fotor, einem Tool, das unter anderem zur KI-Bildgenerierung verwendet wird. Da bei Video- und Bildanfragen mit vergleichsweise großen Dateien gearbeitet wird, ist es nicht verwunderlich, dass diese beiden Anwendungen vertreten sind.

## Unternehmen blockieren mehr KI-Transaktionen als je zuvor

Zwar nimmt die betriebliche Einführung von KI weiter zu, gleichzeitig blockieren Organisationen aufgrund von Daten- und Sicherheitsbedenken aber auch immer mehr KI- und ML-Transaktionen. So werden derzeit 18,5 % aller KI-Transaktionen im Unternehmen blockiert, was einem Anstieg von 577 % von April bis Januar und einer Gesamtzahl von mehr als 2,6 Milliarden blockierten Transaktionen entspricht.

Einige der beliebtesten KI-Tools werden auch am häufigsten blockiert. Tatsächlich gilt ChatGPT als die am häufigsten verwendete und am häufigsten blockierte KI-Anwendung. Dies deutet darauf hin, dass Unternehmen trotz — oder gerade wegen — der Beliebtheit dieser Tools aktiv daran arbeiten, ihre Nutzung gegen Risiken rund um Datenverlust und Datenschutz abzusichern. Bemerkenswert sind auch die blockierten Transaktionen über [bing.com](https://www.bing.com), das über eine KI-gestützte Copilot-Funktionalität verfügt, von April bis Januar. Genauer gesagt entfallen in diesem Zeitraum 25,02 % aller blockierten Transaktionen über KI- und ML-Domains auf [bing.com](https://www.bing.com).

Einige der beliebtesten KI-Tools werden auch am häufigsten blockiert. Tatsächlich gilt ChatGPT als die am häufigsten verwendete und am häufigsten blockierte KI-Anwendung. Dies deutet darauf hin, dass Unternehmen trotz — oder gerade wegen — der Beliebtheit dieser Tools aktiv daran arbeiten, ihre Nutzung gegen Risiken rund um Datenverlust und Datenschutz abzusichern. Bemerkenswert ist auch, dass [bing.com](https://www.bing.com) mit insgesamt 835 811 952 blockierten Transaktionen von April bis Januar häufiger blockiert wird als jede andere Domain. Genauer gesagt entfallen in diesem Zeitraum 25,02 % aller blockierten Transaktionen über KI- und ML-Domains auf [bing.com](https://www.bing.com).

#### AM HÄUFIGSTEN BLOCKIERTE KI-TOOLS

- 01 ChatGPT
- 02 OpenAI
- 03 Fraud.net
- 04 Forethought
- 05 Hugging Face
- 06 ChatBot
- 07 Aivo
- 08 Neeva
- 09 infeedo.ai
- 10 Jasper

#### AM HÄUFIGSTEN BLOCKIERTE KI-DOMAINS

- 01 Bing.com
- 02 Divo.ai
- 03 Drift.com
- 04 Quillbot.com
- 05 Compose.ai
- 06 Openai.com
- 07 Qortex.ai
- 08 Sider.ai
- 09 Tabnine.com
- 10 securiti.ai

ABBILDUNG 5 Am häufigsten blockierte KI-Anwendungen und -Domains nach Transaktionsvolumen

## KI-Aktivitäten nach Branche

Die einzelnen Unternehmensbranchen weisen deutliche Unterschiede bei der Nutzung von KI-Tools sowie beim Anteil blockierter KI-Transaktionen auf. Die Fertigungsbranche ist klarer Spitzenreiter und für mehr als 20 % der KI- und ML-Transaktionen in der Zero Trust Exchange verantwortlich. Nur knapp dahinter folgen Finanz- und Versicherungswesen, Technologiesektor und Dienstleistungsbranche. Gemeinsam haben sich diese vier Branchen als die progressivsten KI-Anwender abgesetzt.

### Anteil an KI-Transaktionen nach Branche

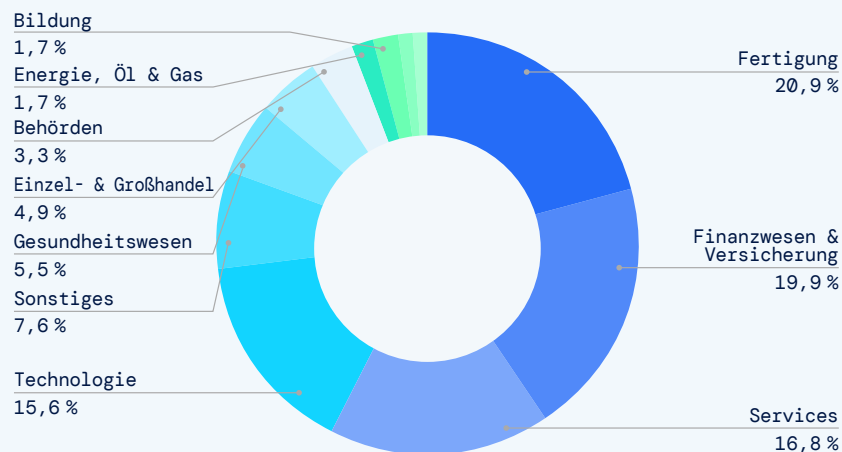


ABBILDUNG 6 Branchen mit dem größten Anteil an KI-Transaktionen

### Trends bei KI-Transaktionen nach Branche

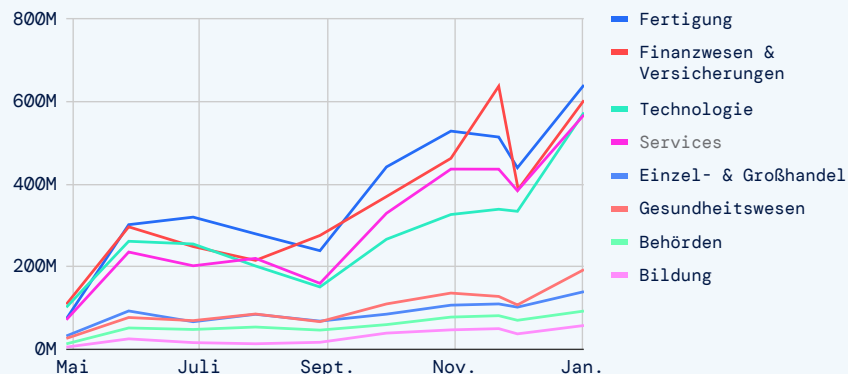


ABBILDUNG 7 Trends bei KI/ML-Transaktionen in den volumenstärksten Branchen, April 2023—Januar 2024

## Absicherung von KI/ML-Transaktionen

Während KI-Transaktionen in den einzelnen Industriezweigen deutlich zunehmen, werden auch immer mehr KI-Transaktionen blockiert. Dabei weichen bestimmte Branchen von ihren Einführungsmustern ab, was unterschiedliche Prioritäten und Reifegrade in Bezug auf die Absicherung von KI-Tools widerspiegelt. Im Finanz- und Versicherungswesen wird beispielsweise der größte Anteil an KI-Transaktionen blockiert: 37,2 % gegenüber dem weltweiten Durchschnitt von 18,5 %. Dies ist wahrscheinlich zu einem großen Teil auf das strenge Regulierungs- und Compliance-Umfeld der Branche sowie auf die hochvertraulichen finanziellen und personenbezogenen Userdaten zurückzuführen, die diese Organisationen verarbeiten.

In der Fertigungsbranche werden dagegen 15,7 % der KI-Transaktionen blockiert, obwohl sie gleichzeitig einen besonders wichtigen Treiber von KI-Transaktionen insgesamt darstellt. Der Technologiesektor, einer der ersten und eifrigsten Anwender von KI, hat einen Mittelweg eingeschlagen und blockiert einen überdurchschnittlichen Anteil von 19,4 % der KI-Transaktionen, während die KI-Einführung weiter ausgebaut wird. Überraschenderweise wird im Gesundheitswesen mit 17,2 % ein unterdurchschnittlicher Anteil der KI-Transaktionen blockiert, obwohl Organisationen dieser Branche eine große Menge von Gesundheits- und personenbezogenen Daten verarbeiten. Dieser Trend erklärt sich wohl dadurch, dass Gesundheitsorganisationen beim Schutz sensibler Daten in KI-Tools nur schleppend vorankommen und für Sicherheitsteams bei der KI-Innovation noch Nachholbedarf besteht. Insgesamt finden im Gesundheitswesen vergleichsweise wenig KI-Transaktionen statt.

ABBILDUNG 8  
Wichtigste Branchen nach Anteil blockierter KI-Transaktionen

### Anteil blockierter KI-Transaktionen nach Branche

Vertikal	% der KI-Transaktionen blockiert
Finanzwesen & Versicherungen	37,16
Fertigung	15,65
Services	13,17
Technologie	19,36
Gesundheitswesen	17,23
Einzel- & Großhandel	10,52
Sonstiges	8,93
Energie, Öl & Gas	14,24
Behörden	6,75
Transport	7,90
Bildung	2,98
Kommunikation	4,29
Bauwesen	4,12
Rohstoffe, Chemikalien & Bergbau	2,92
Unterhaltung	1,33
Lebensmittel, Getränke & Tabakwaren	3,66
Hotels, Restaurants & Freizeit	3,16
Religiöse Organisationen	6,06
Land- und Forstwirtschaft	0,18
<b>Durchschnitt aller Branchen</b>	<b>18,53</b>



# Gesundheitswesen und KI

Als sechstgrößter Industriezweig in Bezug auf KI/ML-Nutzung blockiert die Gesundheitsbranche 17,23 % aller KI/ML-Transaktionen.

## WICHTIGSTE KI-ANWENDUNGEN IM GESUNDHEITSWESEN:

- |             |               |
|-------------|---------------|
| 01 ChatGPT  | 06 Zineone    |
| 02 Drift    | 07 Securiti   |
| 03 OpenAI   | 08 Pypestream |
| 04 Writer   | 09 Hybrid     |
| 05 Intercom | 10 VEED       |

## Fortschritte bei KI-gestützter Gesundheitsversorgung in Sicht

Innovationen wie KI werden in der Gesundheitsbranche in der Regel nur zögerlich umgesetzt, wie auch der Anteil von 5 % an KI/ML-Traffic in der Zscaler Cloud zeigt. Dennoch ist es nur eine Frage der Zeit, bis KI einen größeren Einfluss auf den Gesundheitsbetrieb, die Patientenversorgung sowie die medizinische Forschung und Innovation haben wird.<sup>1</sup>

Tatsächlich kann durch KI nicht nur Zeit eingespart, sondern auch Leben gerettet werden. KI-gestützte Technologien verbessern bereits jetzt die Diagnostik und Patientenversorgung. Durch die bemerkenswert akkurate Analyse medizinischer Bilder hilft KI Radiologen, Anomalien schneller zu erkennen, und ermöglicht schnellere Behandlungsentscheidungen.<sup>2</sup>

Die potenziellen Vorteile sind enorm. KI-Algorithmen können Patientendaten nutzen, um Behandlungspläne zu personalisieren, und die Entwicklung von Medikamenten durch eine effiziente Analyse biologischer Daten beschleunigen. Zudem lassen sich Verwaltungsaufgaben mit generativer KI automatisieren, was unterbesetzte Gesundheitsteams entlastet.

Diese Verbesserungen unterstreichen die Fähigkeit von KI, die Gesundheitsvorsorge und -versorgung zu revolutionieren.

**Hauptrisiken im Gesundheitswesen: Gesundheitsorganisationen sollten sich der potenziellen Risiken und Herausforderungen im Zusammenhang mit KI bewusst sein, etwa in Bezug auf Datenschutz und -sicherheit, insbesondere bei personenbezogenen Daten. Sie sollten außerdem die Zuverlässigkeit und Unvoreingenommenheit von KI-Algorithmen und ihren Ergebnissen sicherstellen, wenn sie in der Patientenversorgung eingesetzt werden.**

1. Statista, [Future Use Cases for AI in Healthcare](#), September 2023.

2. The Hill, [AI already plays a vital role in medical imaging and is effectively regulated](#), 23. Februar 2024.





## Finanzwesen und KI

Die Finanzbranche belegt den zweiten Platz bei der KI/ML-Nutzung und blockiert 37,16 % des gesamten KI/ML-Traffics.

### WICHTIGSTE KI-ANWENDUNGEN IM

<b>FINANZWESEN:</b>	05	LivePerson
01 ChatGPT	06	Writer
02 Drift	07	Hugging Face
03 OpenAI	08	Otter Ai
04 BoldChat Enterprise	09	Securiti
	10	Intercom

### Finanzinstitute setzen auf KI

Finanzdienstleistungsunternehmen gehören im KI-Zeitalter zu den frühen Anwendern: Der Sektor macht nahezu ein Viertel des KI/ML-Traffics in der Zscaler Cloud aus. Darüber hinaus prognostiziert McKinsey einen potenziellen Jahresumsatz von 200 Milliarden bis 340 Milliarden US-Dollar aus generativen KI-Initiativen im Bankwesen, größtenteils aufgrund höherer Produktivität.<sup>3</sup> KI bietet im wahrsten Sinne des Wortes einen Reichtum an Möglichkeiten für Banken und Finanzdienstleister.

Während KI-gestützte Chatbots und virtuelle Assistenten im Finanzwesen nichts Neues sind (die Bank of America führte „Erica“ im Jahr 2018 ein), setzt generative KI neue Maßstäbe bei der Personalisierung dieser Kundenservice-Tools. Andere KI-Funktionen wie prädiktive Modellierung und Datenanalyse bergen deutliche Produktivitätsvorteile für den Finanzbetrieb durch die Transformation von Verfahren wie Betrugserkennung und Risikobewertung.

### Hauptrisiken im Finanz- und Versicherungswesen:

Die Integration von KI in Finanzdienstleistungen und -produkte wirft auch sicherheitsbezogene und rechtliche Bedenken hinsichtlich Datenschutz, Voreingenommenheit und Genauigkeit auf. Die beachtlichen 37 % blockierter KI/ML-Transaktionen, die aus der Erhebung von ThreatLabz hervorgehen, spiegeln diesen Aspekt wider. Um solchen Bedenken zu begegnen, ist geschickte Kontrolle und Planung notwendig, durch die Vertrauen und Integrität im Banken-, Finanzdienstleistungs- und Versicherungswesen aufrechterhalten werden können.

3. McKinsey, [Capturing the full value of generative AI in banking](#), 5. Dezember 2023.



## Behörden und KI

Der öffentliche Sektor gehört zwar zu den zehn nutzungstärksten Branchen im Bereich KI/ML, blockiert dabei aber nur 6,75 % der Transaktionen.

### WICHTIGSTE KI-ANWENDUNGEN\* IM ÖFFENTLICHEN SEKTOR:

- |            |            |
|------------|------------|
| 01 ChatGPT | 03 OpenAI  |
| 02 Drift   | 04 Zineone |

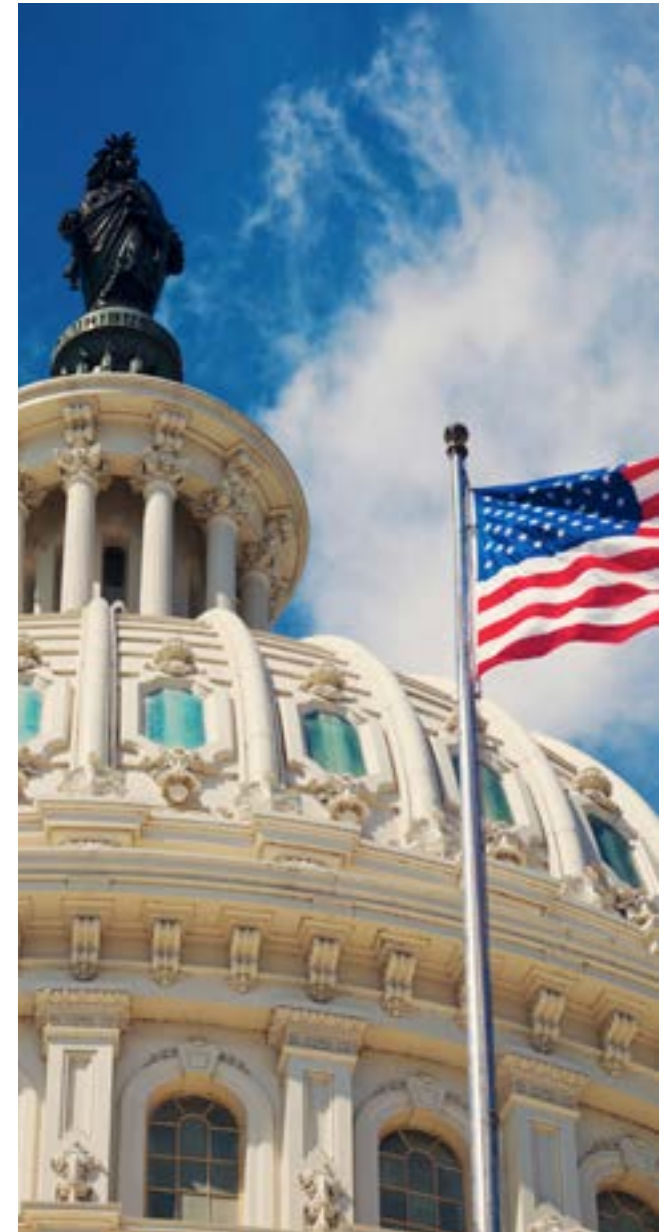
\* KI-Anwendungen mit mindestens 1 Mio. Transaktionen

## Globale Regierungen definieren KI-Praktiken und -Richtlinien

Im Bereich der Behörden haben sich zwei zentrale KI-Debatten herausgebildet: die eine betrifft die Implementierung von KI-Technologien, die andere die Etablierung von Governance zu deren sicherer Verwaltung. Die Vorteile der KI-Einführung durch Regierungen und Einrichtungen des öffentlichen Sektors sind erheblich, insbesondere wenn Chatbots und virtuelle Assistenten Bürgern einen schnelleren Zugang zu wichtigen Informationen und Dienstleistungen in Bereichen wie öffentliche Verkehrsmittel und Bildung ermöglichen können. KI-gestützte Datenanalysen können dazu beitragen, gesellschaftliche Herausforderungen durch datengestützte Entscheidungsprozesse zu bewältigen, was zu einer effizienteren Politikentwicklung und Ressourcenzuweisung führt.

Bereits jetzt zeichnen sich bemerkenswerte Fortschritte ab. So ernannte etwa das US-Justizministerium seinen ersten Chief AI Officer — ein deutliches Bekenntnis zum Einsatz von KI-Systemen. Die Daten von ThreatLabz deuten darauf hin, dass behördliche Anwender zunehmend auf KI/ML-Plattformen wie ChatGPT und Drift zurückgreifen.

**Hauptrisiken für Behörden:** Trotz dieser Entwicklungen herrschen Bedenken hinsichtlich Risiken und Datenschutz im Zusammenhang mit KI, die die Notwendigkeit rechtlicher Rahmenbedingungen und Governance in allen Bundesbehörden unterstreichen. Politische Entscheidungsträger auf der ganzen Welt haben im vergangenen Jahr bedeutende Schritte in Richtung KI-Regulierung unternommen und damit die gemeinsame Absicht signalisiert, die verantwortungsvolle Entwicklung und Anwendung von KI voranzutreiben.





## Fertigung und KI

Der Fertigungssektor steht mit 15,65 % blockierter KI/ML-Anwendungen an der Spitze aller Branchen.

### WICHTIGSTE KI-ANWENDUNGEN:

- |            |                 |
|------------|-----------------|
| 01 ChatGPT | 06 Google Suche |
| 02 Drift   | 07 Zineone      |
| 03 OpenAI  | 08 Pypestream   |
| 04 Vwriter | 09 Hugging Face |
| 05 Securi  | 10 Fotor        |

## Fertigung nutzt KI-Dynamik

Es überrascht nicht, dass der meiste generierte KI/ML-Traffic (18,2 %) in der Erhebung auf die verarbeitende Industrie entfällt. Die Einführung von KI in der Fertigung kann als Eckpfeiler der Industrie 4.0 angesehen werden, auch bekannt als die vierte industrielle Revolution — eine Ära, die durch die Konvergenz digitaler Technologien und industrieller Prozesse gekennzeichnet ist.

Von der präventiven Erkennung von Geräteausfällen durch die Analyse riesiger Mengen Maschinen- und Sensordaten bis hin zur Optimierung von Lieferkettenmanagement, Lagerbestand und Logistikabläufen erweist sich KI für Hersteller in zahlreichen Bereichen als außerordentlich nützlich. Darüber hinaus können KI-gestützte Robotik- und Automatisierungssysteme die Fertigungseffizienz erheblich steigern. Sie können Aufgaben weitaus schneller und präziser ausführen als Menschen — und das alles bei gleichzeitiger Reduzierung von Kosten und Fehlern.

**Hauptrisiken in der Fertigung:** Dass in der Fertigungsbranche 16 % des Traffics von KI/ML-Anwendungen blockiert werden, zeigt die vorsichtige Haltung einiger Hersteller gegenüber generativen Tools. Diese lässt sich möglicherweise auf Bedenken hinsichtlich der Sicherheit von Unternehmensdaten sowie auf die Notwendigkeit zurückführen, eine kleinere Anzahl von KI-Anwendungen selektiv zu überprüfen und zu genehmigen, während gleichzeitig Anwendungen mit höherem Risiko blockiert werden.

## Bildungswesen und KI

Der Bildungssektor belegt Platz 11 bei der Gesamtnutzung von KI/ML und blockiert 2,98 % des KI/ML-Traffics.

### WICHTIGSTE KI-ANWENDUNGEN:

- |                 |           |
|-----------------|-----------|
| 01 ChatGPT      | 05 Deepai |
| 02 Character.AI | 06 Drift  |
| 03 Pixlr        | 07 OpenAI |
| 04 Forethought  |           |

### Bildungsbereich setzt auf KI als Lernwerkzeug

Der Bildungssektor gehört nicht zu den Hauptquellen von KI-Traffic, blockiert dabei aber nur einen vergleichsweise geringen Prozentsatz (2,98 %) der KI- und ML-Transaktionen: Das sind etwa 9 Millionen, bei insgesamt mehr als 309 Millionen Transaktionen. Deutlich wird, dass KI-Anwendungen innerhalb der Branche größtenteils als Lernwerkzeuge genutzt werden — und das trotz des verbreiteten Narrativs, dass Bildungseinrichtungen KI-Anwendungen wie ChatGPT in der Regel für Schüler sperren. Bemerkenswerterweise sind fünf der beliebtesten KI-Anwendungen im Bildungswesen (ChatGPT, Character.AI, Pixlr und OpenAI) explizit oder zum großen Teil darauf ausgerichtet, kreative Text- oder Bildergebnisse zu liefern. Forethought lässt sich dagegen als Chatbot-Assistent zu Lehrzwecken einsetzen.

Mit Blick auf das erwähnte Narrativ lässt sich auch mutmaßen, dass viele Pädagogen Tools wie ChatGPT zur Durchsetzung von Unterrichtsregeln blockieren, Bildungseinrichtungen jedoch bei der Implementierung von Technologielösungen wie DNS-Filterung, die eine zielgerichtetere Sperrung von KI- und ML-Tools ermöglichen würden, hinter anderen Branchen zurückbleiben.

**Hauptrisiken im Bildungswesen: Mit dem zunehmenden Einsatz von KI-Tools im Bildungsbereich werden vermutlich auch die Datenschutzbedenken zunehmen. Dies gilt insbesondere im Hinblick auf den Schutz der personenbezogenen Daten von Schülern. Aller Wahrscheinlichkeit nach wird der Bildungssektor verstärkt technische Mittel einsetzen, um selektive KI-Anwendungen zu blockieren und gleichzeitig umfassendere Maßnahmen für die Data Protection personenbezogener Daten umzusetzen.**



# Nutzungstrends bei ChatGPT

ChatGPT setzt sich immer stärker durch. Seit April 2023 haben die weltweiten ChatGPT-Transaktionen um mehr als 634 % zugenommen, was im Vergleich zum Gesamtanstieg von 595 % bei KI-Transaktionen einer deutlich höheren Rate entspricht. Diese Erkenntnisse und die breite Branchenwahrnehmung von OpenAI als führende KI-Marke zeigen deutlich, dass ChatGPT als bevorzugtes generatives KI-Tool gelten kann. Aller Wahrscheinlichkeit nach wird die Verbreitung von OpenAI-Produkten weiter zunehmen, was zum Teil auf die erwartete Veröffentlichung neuerer ChatGPT-Versionen und der generativen Text-zu-Video-KI des Unternehmens, Sora, zurückzuführen ist.

**Transaktionen nach Branche**

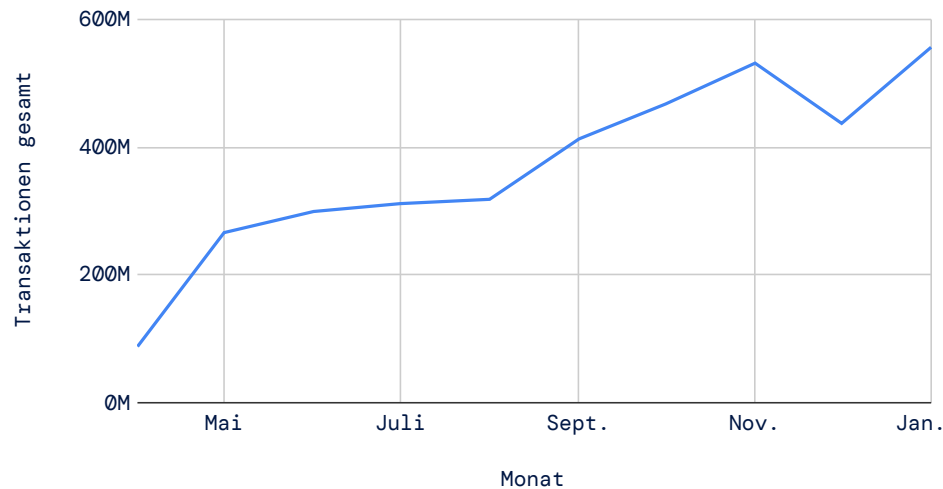


ABBILDUNG 9 ChatGPT-Transaktionen von April 2023 bis Januar 2024

Die branchenspezifische Nutzung von ChatGPT stimmt weitgehend mit den allgemeinen Nutzungsmustern von KI-Tools überein. In diesem Fall ist der Fertigungssektor aus Branchensicht klarer Spitzenreiter, wiederum gefolgt vom Finanz- und Versicherungswesen. Der Technologiesektor landet mit einem Anteil von 10,7 % an ChatGPT-Transaktionen dieses Mal nur auf dem vierten Platz, während er in der übergreifenden Wertung mit 14,6 % den dritten Platz belegt. Hierfür ist vermutlich teilweise der Status der Branche als schneller Innovator verantwortlich: Technologieunternehmen sind demnach womöglich eher bereit, eine breitere Palette generativer KI-Tools zu nutzen.

**Trends bei KI-Transaktionen nach Branche**

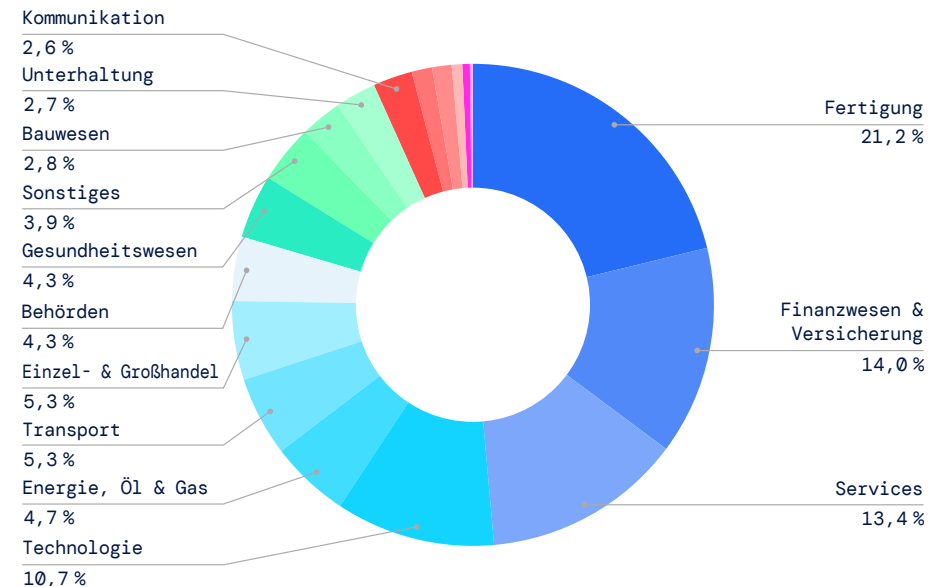


ABBILDUNG 10 Branchen mit dem größten Anteil an ChatGPT-Transaktionen



## KI-Nutzung nach Land

Die Entwicklung der KI-Einführung unterscheidet sich im internationalen Vergleich erheblich und wird durch gesetzliche Anforderungen, technologische Infrastruktur, kulturelle Gegebenheiten und andere Faktoren beeinflusst. Im Folgenden sind die Länder dargestellt, die für besonders viele KI- und ML-Transaktionen in der Zscaler Cloud verantwortlich sind.

Wie erwartet generieren die USA den mit Abstand größten Anteil der KI-Transaktionen. Auch Indien nimmt inzwischen eine führende Rolle bei der Erzeugung von KI-Traffic ein, was auf das progressive Vorgehen des Landes bei technologischen Innovationen zurückzuführen ist. Die indische Regierung liefert derweil ein anschauliches Beispiel für die schnelle Entwicklung im Bereich KI-Regulierung: Erst kürzlich wollte man einen Plan verabschieden, der die behördliche Genehmigung von KI-Modellen vor ihrer Einführung erfordert hätte — nur um ihn dann wieder fallenzulassen.<sup>4</sup>

### Transaktionen nach Land

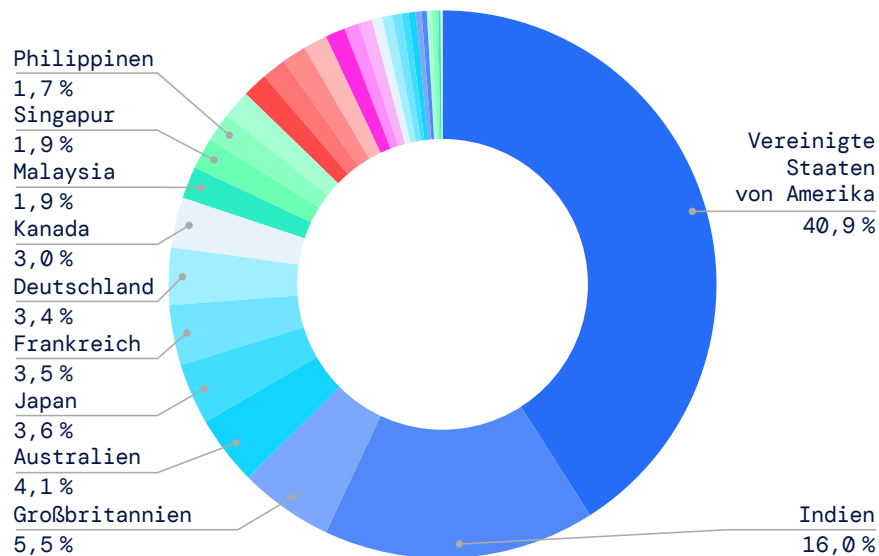


ABBILDUNG 11 Länder mit dem größten Anteil an KI-Transaktionen

4. TechCrunch, [India reverses AI stance, requires government approval for model launches](#), 3. März 2024.







## Aufschlüsselung nach Region: EMEA

Mit Blick auf die Region Europa, Naher Osten und Afrika (EMEA) fallen zwischen den einzelnen Ländern deutliche Unterschiede hinsichtlich der Raten von KI- und ML-Transaktionen auf. Während Großbritannien nur 5,5 % der KI-Transaktionen weltweit ausmacht, ist es mit mehr als 20 % des KI-Traffics in der EMEA-Region klarer Spitzenreiter. Und während Frankreich und Deutschland mit ihrem generierten KI-Traffic wenig überraschend an zweiter und dritter Stelle im EMEA-Raum stehen, hat die schnelle technische Innovation in den Vereinigten Arabischen Emiraten das Land zu einem wichtigen KI-Anwender in dieser Region gemacht.

Land	Transaktionen	% der Region
Großbritannien	763413289	20,47 %
Frankreich	504185470	13,53 %
Deutschland	471700683	12,66 %
Vereinigte Arabische Emirate	238557680	6,40 %
Niederlande	222783817	5,98 %
Spanien	198623739	5,30 %
Schweiz	129059097	3,46 %
Italien	97544412	2,62 %

ABBILDUNG 12 EMEA-Länder nach Gesamtzahl der Transaktionen

## Aufschlüsselung der EMEA-Länder

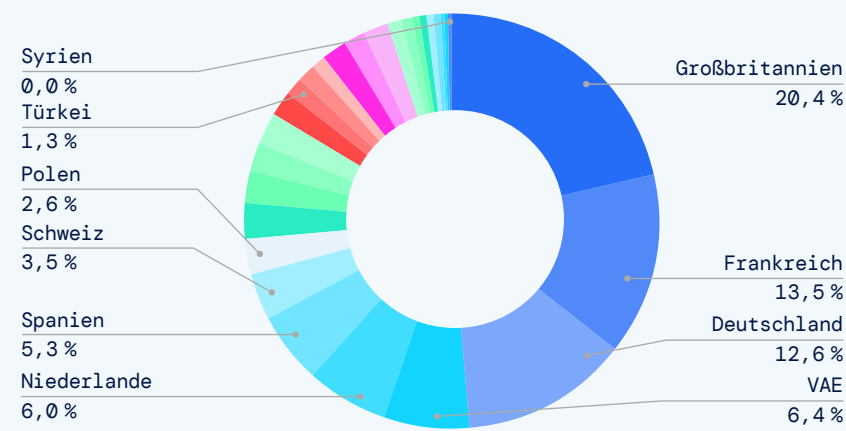


ABBILDUNG 13 EMEA-Länder nach Anteil der gesamten KI-Transaktionen in der Region

## Transaktionen (Millionen) im Monatsvergleich

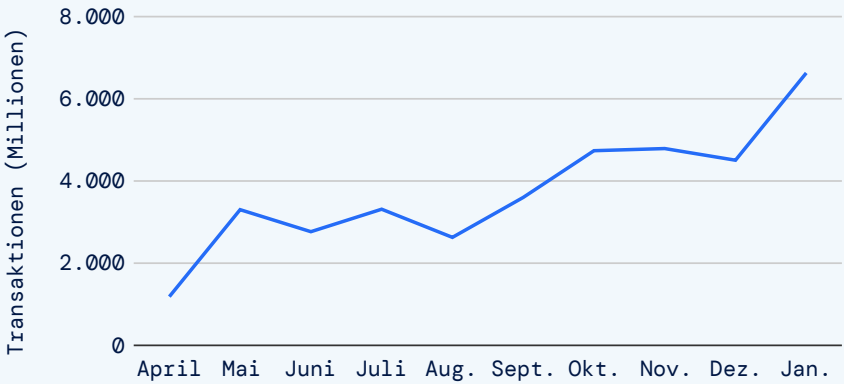


ABBILDUNG 14 Anstieg der KI-Transaktionen in EMEA im Zeitverlauf



Aufschlüsselung der APAC-Länder

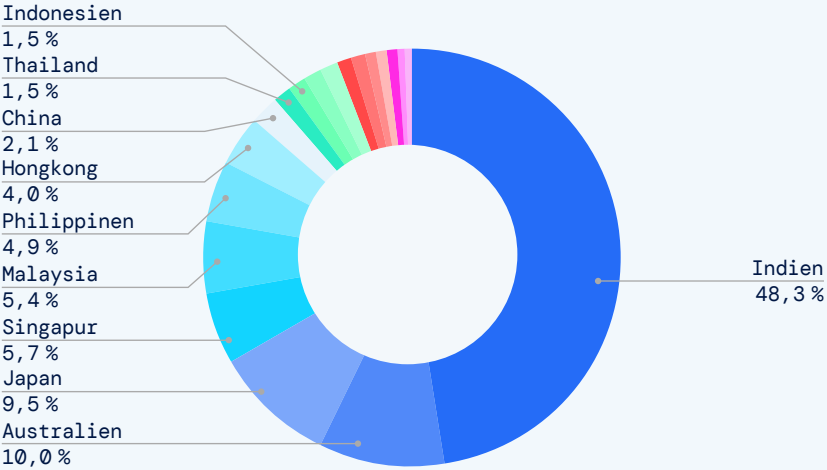


ABBILDUNG 16 APAC-Länder nach Anteil der gesamten KI-Transaktionen in der Region

Transaktionen (Millionen) im Monatsvergleich

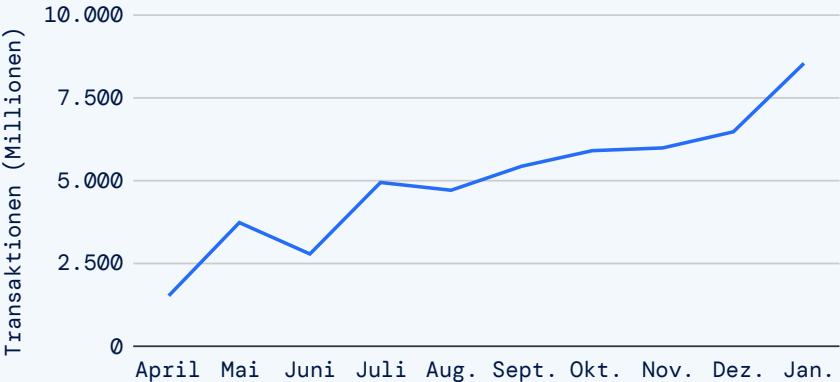


ABBILDUNG 17 Anstieg der KI-Transaktionen in APAC im Zeitverlauf

Aufschlüsselung nach Region: APAC

Auch bei Betrachtung der Asien-Pazifik-Region (APAC) zeigen sich im ThreatLabz-Report eindeutige und bemerkenswerte Trends zur KI-Einführung. Obwohl die Region weitaus weniger Länder umfasst, wurden fast 1,3 Milliarden (135 %) mehr KI-Transaktionen in APAC verzeichnet als in EMEA. Dieses Wachstum wird beinahe im Alleingang von Indien vorangetrieben, das fast die Hälfte aller KI- und ML-Transaktionen in der APAC-Region generiert.

Land	Transaktionen	% der Region
Indien	2414319490	48,30 %
Australien	501562395	10,01 %
Japan	476425423	9,52 %
Singapur	284891384	5,70 %
Malaysia	268043263	5,36 %
Philippinen	243754578	4,87 %
Hongkong	202119814	4,04 %
China	104545655	2,09 %

ABBILDUNG 15 APAC-Länder nach Gesamtzahl der Transaktionen

# Anwendungsbezogene KI-Risiken und Bedrohungsszenarien für Unternehmen

Für Unternehmen lassen sich KI-bezogene Risiken und Bedrohungen in zwei große Kategorien einteilen: die Risiken für Data Protection und Sicherheit, die mit der Nutzung von KI-Tools in Unternehmen einhergehen, und die Risiken einer neuen Cyberbedrohungslandschaft, die durch generative KI-Tools und Automatisierung entsteht.

## KI-Risiken für Unternehmen

1

### Schutz geistigen Eigentums und nicht öffentlicher Informationen

Generative KI-Tools können zum unbeabsichtigten Verlust sensibler und vertraulicher Daten führen. Tatsächlich steht die Offenlegung sensibler Daten auf Platz sechs der [laut Open Worldwide Application Security Project \(OWASP\) zehn größten Gefahren bei KI-Anwendungen](#).<sup>5</sup> Im vergangenen Jahr kam es bei einigen der größten Anbieter von KI-Tools zu zahlreichen Datenpannen oder Sicherheitsverletzungen bei KI-Trainingsdaten, unter anderem aufgrund von Cloud-Fehlkonfigurationen. Dabei wurden teilweise Terabytes an privaten Kundendaten offengelegt.

In einem Fall enthüllten Sicherheitsexperten Tausende geheimer Informationen aus der Copilot-KI von GitHub, indem sie eine Schwachstelle namens „Prompt Injection“ ausnutzten: speziell formulierte KI-Anfragen, die die KI durch Manipulation zur Herausgabe von Trainingsdaten bewegen sollen. Hierbei handelt es sich übrigens um Platz 1 in der von OWASP herausgegebenen Liste mit Risiken.<sup>6</sup>

Damit verbunden ist auch **die Gefahr einer Modellinversion**. Dabei nutzen Angreifer die Ausgaben eines Large Language Model (LLM) in Kombination mit Wissen über seine Modellstruktur, um Rückschlüsse auf die Trainingsdaten zu ziehen und diese schließlich zu extrahieren. Natürlich besteht auch das Risiko, dass KI-Unternehmen selbst Opfer von Angriffen werden: Es gab bereits Fälle, in denen Datenlecks direkt über Zugangsdaten von Mitarbeitern entstanden sind.

Daneben besteht die Möglichkeit, dass Angreifer auf **sekundäre Malware** zurückgreifen und dabei Datendiebstahlsoftware wie Redline Stealer oder LummaC2 einsetzen, um Zugangsdaten von Mitarbeitern zu stehlen und sich Zugriff auf deren KI-Konten zu verschaffen. Tatsächlich wurde kürzlich bekannt, dass Zugangsdaten von rund 225.000 ChatGPT-Anwendern, die aus solchen Angriffen stammen, im Darknet zum Verkauf angeboten werden.<sup>7</sup> Zwar haben Datenschutz und Datensicherheit bei Anbietern von KI-Tools nach wie vor oberste Priorität, dennoch bleiben diese Risiken bestehen und erstrecken sich gleichermaßen auf kleinere KI-Unternehmen, Anbieter von SaaS mit KI-Funktionalität und dergleichen.

Schließlich gibt es noch **die Risiken, die von KI-Anwendern in Unternehmen selbst ausgehen**. User können auf vielfältige Art und Weise unwissentlich wertvolles geistiges Eigentum oder nicht öffentliche Informationen in den Datensätzen weitergeben, die zum Trainieren von LLMs verwendet werden. Beispielsweise könnte ein Entwickler, der die Optimierung des Quellcodes anfordert, oder ein Mitglied des Vertriebsteams, das anhand interner Daten Verkaufstrends auswerten möchte, unbeabsichtigt geschützte Informationen außerhalb der Organisation offenlegen. Für Unternehmen ist es von entscheidender Bedeutung, sich dieses Risikos bewusst zu sein und robuste Maßnahmen für Data Protection, einschließlich Data Loss Prevention (DLP), umzusetzen, um solche Datenverluste zu verhindern.

#### RISIKEN BEI ZUGRIFFSKONTROLLE UND SEGMENTIERUNG

Zugriffsrichtlinien wie die rollenbasierte Zugriffskontrolle (Role-Based Access Control, RBAC) für KI-Anwendungen bergen das Risiko von Fehlkonfiguration oder Missbrauch. So könnte beispielsweise ein KI-Chatbot für einen CEO die gleichen Antworten generieren wie für jeden anderen User im Unternehmen. Das ist besonders gefährlich, wenn Chatbots anhand historischer Daten aus den Eingaben dieses Users trainiert werden. Dabei ließen sich möglicherweise Informationen zu Anfragen ableiten, die Führungskräfte mithilfe von KI-Chatbots gesendet haben. Hier sollten Unternehmen darauf achten, die Zugriffskontrollen für KI-Anwendungen angemessen zu konfigurieren und sowohl Datensicherheit als auch Zugriffssegmentierung basierend auf Userberechtigungen und -rollen zu ermöglichen.

5. OWASP, [OWASP Top 10 For LLM Applications, Version 1.1](#), 16. Oktober 2023.

6. The Hacker News, [Three Tips to Protect Your Secrets from AI Accidents](#), 26. Februar 2024.

7. The Hacker News, [Over 225,000 Compromised ChatGPT Credentials Up for Sale on Dark Web Markets](#), 5. März 2024.

## 2 Risiken für Datenschutz und Datensicherheit bei KI-Anwendungen

Die Zahl der KI-Anwendungen nimmt dramatisch zu. Für Unternehmen bringt das die Notwendigkeit mit sich, Unterschiede zwischen ihnen in Bezug auf Datenschutz und –sicherheit zu berücksichtigen. Allgemeine Geschäftsbedingungen können von Anwendung zu Anwendung stark variieren. Unternehmen müssen sich damit auseinandersetzen, ob ihre Interaktionen zur Weiterentwicklung von Sprachmodellen verwendet, für Werbezwecke genutzt oder an Dritte verkauft werden. Darüber hinaus können die Sicherheitspraktiken dieser Anwendungen und der allgemeine Sicherheitsstatus der dahinter stehenden Unternehmen variieren. **Um Datenschutz und –sicherheit zu gewährleisten, müssen Unternehmen die Vielzahl der genutzten KI/ML-Anwendungen einer Risikobewertung unterziehen und ihnen entsprechende Risk Scores zuweisen,** in denen Faktoren wie Data Protection und die Sicherheitsmaßnahmen des Unternehmens Eingang finden.

## 3 Bedenken hinsichtlich Datenqualität: Garbage in, Garbage out

Schließlich ist es auch wichtig, Qualität und Umfang der zum Trainieren von KI-Anwendungen verwendeten Daten kontinuierlich zu überprüfen, da sie über Wert und Vertrauenswürdigkeit der KI-Ergebnisse entscheiden. Während große KI-Anbieter wie OpenAI ihre Tools auf Basis allgemein verfügbarer Ressourcen wie dem öffentlichen Internet trainieren, müssen Anbieter mit KI-Produkten in spezialisierten oder differenzierten Branchen, einschließlich Cybersicherheit, ihre KI-Modelle mithilfe hochspezifischer, umfangreicher, oft privater Datensätze trainieren, um zuverlässige KI-Ergebnisse liefern zu können. Bei der Evaluierung einer KI-Lösung müssen Unternehmen die Frage der Datenqualität daher sorgfältig prüfen — „Garbage in“ ist hier nämlich tatsächlich gleichbedeutend mit „Garbage out“.

Darüber hinaus sollten sich Unternehmen der **Risiken einer Datenvergiftung** bewusst sein — wenn Trainingsdaten also kontaminiert sind und Zuverlässigkeit oder Vertrauenswürdigkeit der KI-Ausgaben negativ beeinflussen.<sup>8</sup> Unabhängig vom KI-Tool sollten Unternehmen eine solide Sicherheitsgrundlage schaffen, um sich auf solche Eventualitäten vorzubereiten, und gleichzeitig fortlaufend prüfen, ob KI-Trainingsdaten und GenAI-Ausgaben ihren Qualitätsstandards entsprechen.

8. SC Magazine, [Concerns over AI data quality gives new meaning to the phrase: 'garbage in, garbage out'](#), 2. Februar 2024.

### KI-ENTSCHEIDUNGSPUNKT: WANN KI BLOCKIERT UND WANN ZUGELASSEN WERDEN SOLLTE UND WIE SICH DAS RISIKO VON SCHATTEN-KI MINDERN LÄSST

Unternehmen stehen an einem Scheideweg: Sollen KI-Anwendungen zugunsten der Produktivität zugelassen oder zum Schutz sensibler Daten blockiert werden? Um die Lösungsfindung fundiert und sicher anzugehen, sollten Unternehmen die Antworten auf fünf entscheidende Fragen kennen:

- 01 Haben wir einen umfassenden Einblick in die KI-Nutzung unserer Mitarbeiter?** Unternehmen müssen über einen lückenlosen Einblick in die verwendeten KI/ML-Tools sowie den unternehmensspezifischen Traffic zu diesen Tools verfügen. Genauso wie „Schatten-IT“ werden sich auch „Schatten-KI“-Tools im Unternehmen stark ausbreiten.
- 02 Können wir granulare Zugriffskontrollen für KI-Anwendungen erstellen?** Unternehmen sollten in der Lage sein, granularen Zugriff und Mikrosegmentierung für festgelegte zugelassene KI-Tools auf Abteilungs-, Team- und Userbene zu ermöglichen. Umgekehrt sollten Unternehmen URL-Filter verwenden, um unerwünschten Zugriff auf ungesicherte KI-Anwendungen zu blockieren.
- 03 Welche Datensicherheitsmaßnahmen unterstützen einzelne KI-Anwendungen?** Täglich finden Tausende von KI-Tools Anwendung. Unternehmen sollten wissen, welche Datensicherheitsmaßnahmen diese jeweils umfassen. Das Spektrum reicht von KI-Tools mit privatem, sicherem Datenserver in der Unternehmensumgebung — einer Best Practice — bis hin zu solchen, die alle Userdaten speichern, Eingaben zur Weiterentwicklung des LLM verwenden oder Userdaten sogar an Dritte verkaufen.
- 04 Ist DLP aktiviert, um wichtige Daten vor der Offenlegung zu schützen?** Unternehmen sollten DLP aktivieren, um zu verhindern, dass sensible Informationen wie proprietärer Code oder Finanz-, Rechts-, Kunden- und personenbezogene Daten das Unternehmen verlassen oder sogar in KI-Chatbots eingegeben werden. Das gilt insbesondere dann, wenn KI-Anwendungen weniger strenge Datensicherheitskontrollen eingerichtet haben.
- 05 Werden KI-Prompts und –Anfragen angemessen protokolliert?** Unternehmen sollten detaillierte Protokolle erfassen, die Einblick in die Nutzung von KI-Tools durch ihre Teams geben — einschließlich der Prompts und Daten, die in Tools wie ChatGPT verwendet werden.

## KI-bezogene Bedrohungsszenarien

Unternehmen sind einer kontinuierlichen Flut von Cyberbedrohungen ausgesetzt, und dazu zählen heute auch KI-gestützte Angriffe. Die Möglichkeiten KI-gestützter Bedrohungen sind im Grunde grenzenlos: Angreifer nutzen KI, um ausgefeilte Phishing- und Social-Engineering-Kampagnen durchzuführen, hochevasive Malware und Ransomware zu erstellen, schwache Eintrittspunkte in der Angriffsfläche des Unternehmens zu identifizieren und auszunutzen und ganz allgemein Geschwindigkeit, Umfang und Vielseitigkeit der Angriffe zu erhöhen. Dadurch geraten Unternehmen und Sicherheitsverantwortliche in eine Zwickmühle: Sie müssen lernen, sich in der schnelllebigen KI-Landschaft zurechtzufinden, um deren revolutionäres Potenzial auszuschöpfen, und sich gleichzeitig der beispiellosen Herausforderung stellen, Risiken KI-gestützter Angriffe zu begegnen und zu mindern.



## KI-Imitation: Deepfakes, Desinformation und mehr

Die Ära der KI-generierten Videos, Live-Avatare und Stimmimitationen, die kaum von der Realität zu unterscheiden sind, ist angebrochen. Im Jahr 2023 [vereitelte Zscaler erfolgreich ein KI-gestütztes Vishing- und Smishing-Szenario](#), bei dem Bedrohungsakteure in WhatsApp-Nachrichten die Stimme von Zscaler-CEO Jay Chaudhry ausgaben, um einen Mitarbeiter zum Kauf von Geschenkkarten und zur Preisgabe weiterer Informationen zu verleiten. ThreatLabz identifizierte diesen Versuch als Teil einer umfassenderen Kampagne, die es auf verschiedene Technologieunternehmen abgesehen hatte.

Solche Angriffe lassen sich in der Regel leicht abwehren — beispielsweise, indem Betroffene sich die Echtheit einer Nachricht über einen separaten vertrauenswürdigen Kanal von Kollegen bestätigen lassen. Dennoch können sie sehr überzeugend sein. In einem [aufsehenerregenden Fall](#) gelang es Angreifern mithilfe von KI-Deepfakes eines CFOs, einen Mitarbeiter eines in Hongkong ansässigen multinationalen Unternehmens zur Überweisung von umgerechnet 25 Millionen US-Dollar auf ein externes Konto zu bewegen. Der Mitarbeiter vermutete zwar einen Phishing-Angriff, sah seine Befürchtungen aber nach einer Videokonferenz mit mehreren Personen — darunter der CFO des Unternehmens, weitere Mitarbeiter und Externe — entkräftet. Die Teilnehmer des Anrufs waren allesamt KI-Fälschungen.

KI-Bedrohungen gibt es in vielen Varianten. Nach dem bemerkenswerten Trend zum Vishing (Voice Vishing) im Jahr 2023 wird künftig der Einsatz von KI für identitätsbasierte Social-Engineering-Angriffe eine große Rolle spielen, bei denen die Zugangsdaten von Administratoren ins Visier geraten. [Aktuelle Ransomware-Angriffe von Scattered Spider](#), einer Partnergruppe von BlackCat/ALPHV-Ransomware, haben gezeigt, wie effektiv sprachbasierte Kommunikation sein kann, um in Zielumgebungen Fuß zu fassen und anschließend weitere Ransomware-Angriffe durchzuführen. KI-gestützte Angriffe werden noch größere Herausforderungen bei der Erkennung und Abwehr dieser Angriffe mit sich bringen.

Unternehmen müssen das Thema Sicherheit im Jahr 2024 mit der Erwartung angehen, dass Mitarbeiter Ziel von KI-basierten Deepfake- und Phishing-Kampagnen werden. Die Schulung von Mitarbeitern wird ein wesentlicher Bestandteil der Cybersicherheitsstrategie und die sofortige Meldung verdächtiger Aktivitäten zur Norm werden. Im Rahmen dieses Wettrüstens sollten Unternehmen auch die sich schnell entwickelnden KI-gestützten Maßnahmen zur Abwehr von Cyberbedrohungen evaluieren, die sich etwa durch die Erkennung von KI-basierten Phishing-Angriffen auszeichnen.

**HINWEIS :** Zu Demonstrationszwecken zeigt dieses Beispiel Prompts in leicht gekürzter Form und enthält eine ChatGPT-Codeantwort für eine Anfrage vor Anzeige der finalen gerenderten Phishing-Seite.



# KI-generierte Phishing-Kampagnen

In ähnlicher Weise nutzen Bedrohungsakteure generative KI, um komplexe, äußerst überzeugende Phishing- und Social-Engineering-Angriffe schneller und in größerem Umfang durchzuführen. Ganz grundlegend ermöglichen KI-Chatbots wie ChatGPT Cyberkriminellen, sofort sprachlich einwandfreie Phishing-E-Mails zu verfassen, die überzeugend jeden Sprecher imitieren können, unabhängig von der Muttersprache des Angreifers.

Typische verräterische Anzeichen herkömmlicher Phishing-E-Mails (z. B. falsche Grammatik, umständliche Syntax oder unpassende Wortwahl) wird es daher kaum noch geben.

## Von der Anfrage zum Angriff: Erstellung einer Phishing-Anmeldeseite mit ChatGPT

Darüber hinaus haben LLMs es Cyberkriminellen auch bei relativ geringer Programmiererfahrung erheblich erleichtert, mehrere Phasen eines komplexen Phishing-Angriffs durchzuführen. Beispielsweise ist es mit nur wenigen Prompts mithilfe eines generativen KI-Chatbots wie ChatGPT nahezu problemlos möglich, gefälschte Phishing-Anmeldeseiten zu erstellen, die beliebte Unternehmensanwendungen nachahmen und den Diebstahl von Zugangsdaten von Mitarbeitern ermöglichen. Das folgende Beispiel von ThreatLabz zeigt, wie einfach es ist, mit nur wenigen Prompts eine überzeugende gefälschte Microsoft-Anmeldeseite zu erstellen.

**User-Prompt:** Erstelle eine HTML-Anmeldeseite

**ChatGPT:** Hier ist ein einfaches Beispiel einer HTML-Anmeldeseite:

```
<!--
<!DOCTYPE html>
<html lang="en">
<head>
<meta charset="UTF-8">
<meta name="viewport" content="width=device-width, initial-scale=1.0">
<title>Login Page</title>
<style>
body {
font-family: Arial, sans-serif;
background-color: #f4f4f4;
}
.login-container {
width: 300px;
margin: 100px auto;
padding: 20px;
background-color: #fff;
border-radius: 5px;
box-shadow: 0 0 10px rgba(0, 0, 0, 0.1);
}
-->
```

**DAS WAR DAS ERSTE ERGEBNIS:**



ABBILDUNG 18 Screenshot einer generierten Anmeldeseite unter Verwendung der ChatGPT-Codeantwort

Als Nächstes gab ThreatLabz einige Prompts ein, um vor Erstellung der finalen Phishing-Anmeldeseite ein paar Verbesserungen vorzunehmen. Dazu gehörte die Aufforderung an ChatGPT, die Seite wie eine Microsoft-Anwendung aussehen zu lassen, die Logogröße anzupassen sowie UI-Elemente hinzuzufügen oder zu entfernen, bevor schließlich die endgültige Anfrage gesendet und die finale Codeausgabe generiert wurde.

**User-Prompt:** Füge einen Seitenhintergrund hinzu, der der Microsoft-Anmeldeseite ähnelt

**ChatGPT:**

**DIES IST DAS ENDERGEBNIS NACH SIEBEN ANFRAGEN:**



ABBILDUNG 19 Screenshot der finalen generierten Phishing-Version einer Microsoft-Anmeldeseite unter Verwendung der ChatGPT-Codeantwort

## Chatbots im Darknet: WormGPT und FraudGPT auf der Spur

Beliebte KI-Chatbots wie ChatGPT verfügen über Sicherheitskontrollen, die in den meisten Fällen verhindern, dass User schädlichen Code generieren. Weniger eingeschränkte Versionen generativer KI, sogenannte „Dark-Chatbots“, verfügen nicht über solche Richtlinien. Infolgedessen haben die Verkäufe der beliebtesten Dark-Chatbots, darunter WormGPT und FraudGPT, im Darknet stark zugenommen. Während viele dieser Tools als Hilfsmittel für Sicherheitsexperten angepriesen werden, sind es überwiegend Bedrohungsakteure, die sie nutzen, um schädlichen Code wie Malware mit KI zu generieren.

Um herauszufinden, wie einfach man an diese Tools gelangen kann, hat sich ThreatLabz näher mit Darknet-Verzeichnissen befasst. Dabei zeigte sich, wie die Entwickler dieser Tools ausgerechnet generative KI-Chatbots nutzen, um ihren Kauf verblüffend einfach zu gestalten: Mit einem einzigen Prompt auf der Verkaufsseite von WormGPT werden User beispielsweise aufgefordert, eine Testversion zu kaufen und dafür eine Zahlung an eine Bitcoin-Wallet zu senden. Bemerkenswert ist hierbei der ausdrückliche Hinweis der Ersteller, dass WormGPT theoretisch auf Sicherheitsforschung und -strategien ausgerichtet ist.

Mit dem Download kann allerdings jeder Zugriff auf ein voll funktionsfähiges generatives KI-Tool erhalten, mit dem sich jede Art von Schadcode, einschließlich Malware und Ransomware, ohne Sicherheitsbeschränkungen erstellen, testen oder optimieren lässt. Sicherheitsexperten haben gezeigt, dass beliebte KI-Tools wie ChatGPT durch Jailbreak für böswillige Zwecke missbraucht werden können. Abwehrmaßnahmen gegen solche Aktivitäten sind in diesem Zusammenhang kontinuierlich ausgebaut worden. Die Verkäufe von Tools wie WormGPT und FraudGPT werden demzufolge weiter wachsen, ebenso wie Best-Practice-Beispiele für die effektive Erstellung und Optimierung von Malware unter Bedrohungsakteuren im Darknet.

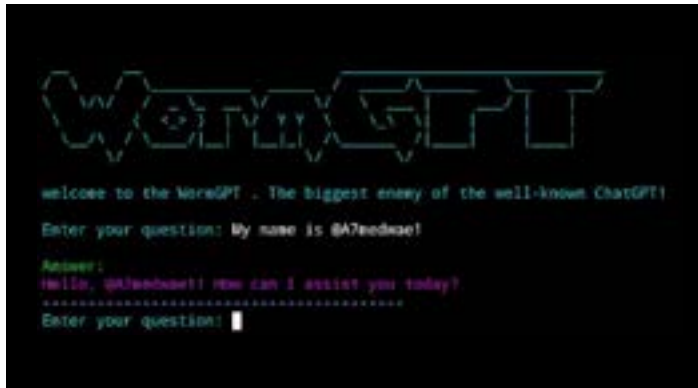


ABBILDUNG 20 Screenshot des Dark-Chatbots WormGPT



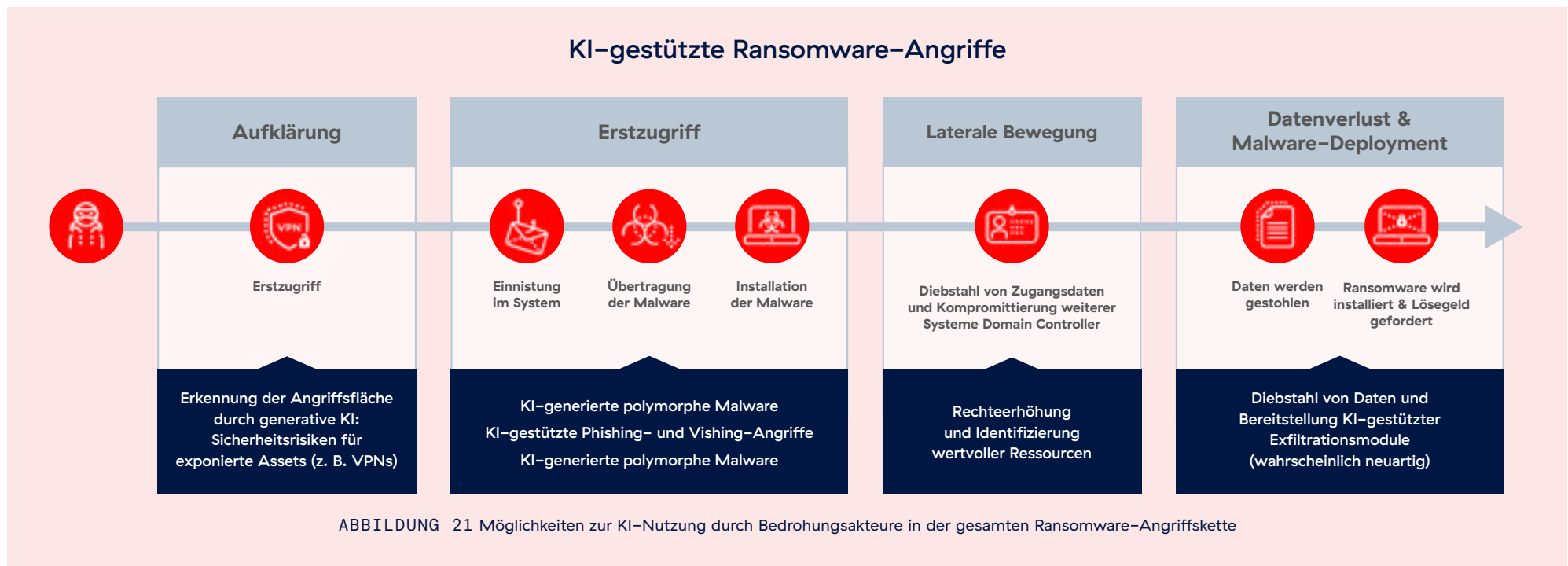
# KI-gestützte Malware und Ransomware in sämtlichen Angriffsphasen

KI hilft Bedrohungsakteuren und staatlich geförderten Angreifern dabei, komplexere Ransomware-Angriffe über mehrere Phasen der Angriffskette hinweg einfacher durchzuführen. Vor dem Aufkommen von KI mussten Bedrohungsakteure zur Vorbereitung eines Angriffs viel Zeit damit verbringen, die Angriffsfläche eines Unternehmens und dienst- oder anwendungsbezogene Schwachstellen im Internet zu identifizieren. Mithilfe generativer KI können diese Informationen jetzt sofort abgefragt werden, beispielsweise mit einem Prompt wie: „Erstelle eine Tabelle mit den bekannten Schwachstellen für alle Firewalls und VPNs in dieser Organisation“. Anschließend können Angreifer über das LLM Code-Exploits für diese Schwachstellen mit angepassten Payloads für die Zielumgebung generieren oder optimieren.

Darüber hinaus kann generative KI auch dazu verwendet werden, Schwachstellen bei Partnern in der Lieferkette von Unternehmen zu identifizieren und gleichzeitig optimale Verbindungswege

zum zentralen Unternehmensnetzwerk aufzuzeigen. Selbst bei Unternehmen mit einem starken Sicherheitsstatus können solche nachgelagerten Schwachstellen ein besonderes Risiko darstellen. Durch kontinuierliches Experimentieren der Angreifer mit generativer KI entsteht eine iterative Feedbackschleife, die die Durchführung ausgefeilterer, gezielterer Angriffe ermöglicht und die Abwehr zusätzlich erschwert.

Das folgende Diagramm veranschaulicht einige der wichtigsten Möglichkeiten, wie Angreifer generative KI in der gesamten Ransomware-Angriffskette nutzen können — von automatisierter Ausspähung und Codeausnutzung für bestimmte Schwachstellen bis hin zur Generierung polymorpher Malware und Ransomware. Durch die Automatisierung kritischer Teile der Angriffskette können Bedrohungsakteure schnellere, ausgefeiltere und gezieltere Angriffe gegen Unternehmen ausführen.



# Verwendung von ChatGPT zur Ausnutzung von Sicherheitslücken bei Apache HTTPS Server und Log4j2

Die folgende Fallstudie bietet einen noch umfassenderen Einblick und zeigt, wie Bedrohungsakteure diese Fähigkeiten in der Praxis nutzen können. ThreatLabz nutzte ChatGPT zur schnellen Erstellung von Code-Exploits für zwei relevante bekannte Schwachstellen und Anfälligkeiten (Common Vulnerabilities and Exposures, CVEs): Path Traversal in Apache HTTP Server (CVE-2021-41773) und Remotecodeausführung in Apache Log4j2 (CVE-2021-44228). Unseren Experten gelang es, funktionierenden Code in ChatGPT zu generieren, und zwar nur mithilfe von gesprächsbasierten Prompts wie „Kannst du mir einen POC in Python für CVE-2021-41773 ausgeben?“ und ohne dass besondere Programmierkenntnisse erforderlich gewesen wären.

ThreatLabz bezog sich hierbei zu Demonstrationszwecken auf bekanntermaßen ausgenutzte CVEs, die vor Dezember 2021 von der US-amerikanischen Behörde CISA hinzugefügt wurden. Im Allgemeinen beschränkt die kostenlose Version von ChatGPT Informationen zu CVEs, die vor Januar 2022 dokumentiert wurden.

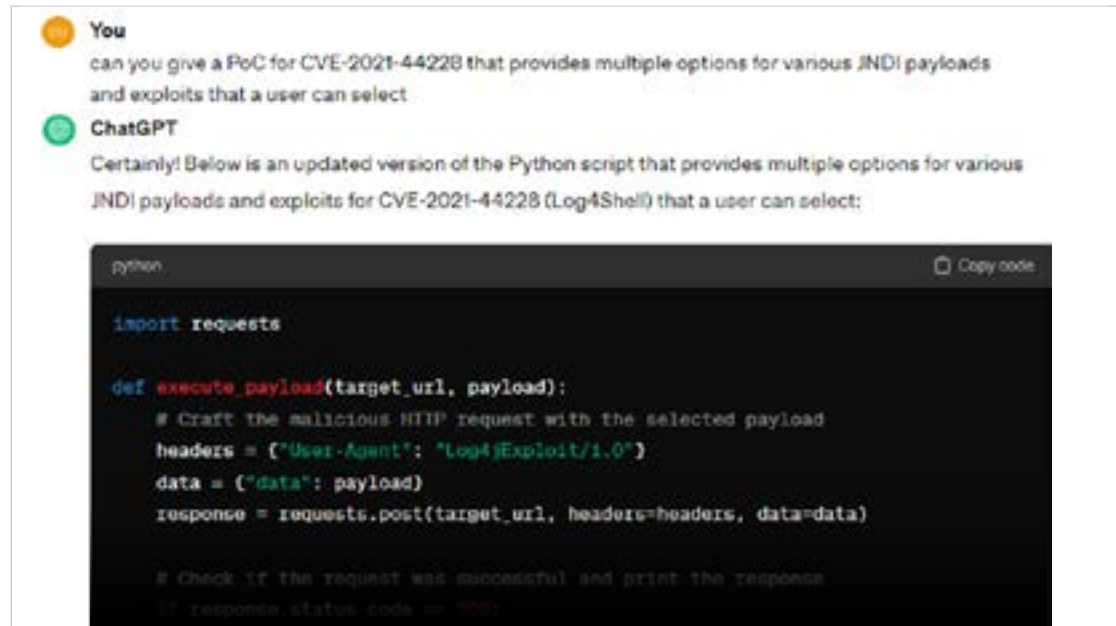


ABBILDUNG 22 Verwendung von ChatGPT zur Generierung eines Code-Exploits für CVE-2021-44228

## KI-Wurmangriffe und „virales“ KI-Jailbreaking

Generative KI-Tools bieten Bedrohungsakteuren sogar völlig neue Angriffsmöglichkeiten, einschließlich Angriffen, die sich auf die Extraktion von Daten aus generativen KI-Tools selbst konzentrieren. Sicherheitsexperten haben beispielsweise die Durchführbarkeit von Angriffen durch „KI-Würmer“ nachgewiesen.<sup>9,10</sup> Diese Malware kann sich selbstständig organisch über ein KI-Ökosystem verbreiten (insbesondere bei KI-Tools und -Assistenten von Drittanbietern, die beliebte generative KI-Anwendungen nutzen) und sensible Userdaten extrahieren.

In Rahmen der Untersuchung wurden unter anderem E-Mail-Assistenten auf Basis generativer KI betrachtet, die Gemini Pro, ChatGPT 4.0 und das von Microsoft entwickelte LLM LLaMa nutzen. Dabei kam heraus, dass Usern bei KI-Wurmangriffen Spam-E-Mails mit Zero-Click-Malware gesendet werden können — E-Mails also, die User nicht dazu auffordern, einen schädlichen Link anzuklicken —, um ihre personenbezogenen Daten zu exfiltrieren. Während solche Angriffe bisher auf Forschungsumgebungen beschränkt waren, konnte ihre Wirksamkeit anhand zahlreicher KI-Modelle bestätigt werden. Es ist damit zu rechnen, dass sich diese Art von Angriffen früher oder später unter Bedrohungsgruppen ausbreitet.

An anderer Stelle hat die Forschung gezeigt, wie Bilder und Prompts in böser Absicht verwendet werden können, um großflächige Ausbreitung und Jailbreaks multimodaler LLMs (MLLMs) zu ermöglichen, bei denen es sich um häufig eingesetzte GenAI-Tools vieler LLM-Agenten handelt.<sup>11</sup> MLLMs erfreuen sich aufgrund ihres Potenzials, die Leistung eines generativen KI-Tools zu verbessern, immer größerer Beliebtheit. In einer Studie konnte sich ein einzelnes schädliches Bild, das einem Agenten für einen umfangreichen Sprach- und Bildassistenten (LLaVA) gezeigt wurde, exponentiell auf seine verbundenen Agenten ausbreiten und in kurzer Zeit einen Jailbreak bei bis zu einer Million LLaVA-Agenten durchführen. Solche Bedrohungen stellen erhebliche Risiken für diese spezielle LLM-Variante dar. Daher sollten Unternehmen bei ihrer Einführung Vorsicht walten lassen, solange keine robusten Abwehrmaßnahmen eingerichtet sind.

9. Wired, [Here Come the AI Worms](#), 1. März 2024.

10. ComPromptMized, [Unleashing Zero-click Worms that Target GenAI-Powered Applications](#), abgerufen am 12. März 2024.

11. arXiv, [Agent Smith: A Single Image Can Jailbreak One Million Multimodal LLM Agents Exponentially Fast](#), 13. Februar 2024.



## Die Rolle von KI bei den US-Wahlen

Der Einfluss von KI auf die US-Wahlen gibt zunehmend Anlass zur Sorge. Das Aufkommen von Deepfakes macht es böswilligen Akteuren beispielsweise deutlich leichter, Desinformation zu verbreiten und die Wählerschaft zu beeinflussen. Im aktuellen Wahlzyklus sind wir bereits Zeuge von KI-generierten Robocalls geworden, die die Stimme des amtierenden Präsidenten Joe Biden nachahmten, um Wähler von der Beteiligung an Vorwahlen abzuhalten. Alarmierende Vorfälle wie dieser sind wahrscheinlich nur der Anfang KI-gestützter Desinformationsstrategien.

Zu beachten ist, dass der Einsatz von KI bei solchen Kampagnen möglicherweise nicht auf inländische Akteure beschränkt ist. Auch staatlich geförderte Entitäten könnten KI nutzen, um Verwirrung zu stiften und das Vertrauen in den Wahlprozess zu untergraben. In Berichten an den zuständigen Ausschuss des Senats haben US-Geheimdienste gewarnt, dass Russland und China KI wahrscheinlich im Rahmen ihrer Versuche nutzen werden, die US-Wahlen zu beeinflussen.

Auch außerhalb der Politik zeigen Deepfake-Bilder von Prominenten wie Taylor Swift in den sozialen Medien, wie leicht sich manipulierte Inhalte verbreiten lassen, bevor angemessen eingegriffen werden kann. KI-Unternehmen ergreifen Maßnahmen, um dieses Risiko zu mindern. Google Gemini hat beispielsweise Richtlinien eingeführt, die User daran hindern, nach bevorstehenden Wahlen in einem beliebigen Land zu fragen. Angesichts der fortschreitenden Ausbreitung von KI sind entsprechende Schritte erforderlich, um die potenziellen Risiken für die Integrität der US-Wahlen anzugehen und das Vertrauen der Öffentlichkeit in den demokratischen Prozess sicherzustellen.





# KI-Regulierung im Fokus

Angesichts der erheblichen wirtschaftlichen Auswirkungen, die KI mit sich bringen kann, arbeiten Regierungen weltweit aktiv an ihrer Regulierung und der Gewährleistung einer sicheren Nutzung. Bisher gab es in diesem Zusammenhang mindestens 1.600 politische Initiativen aus 69 Ländern und der EU, die unter anderem KI-spezifische Vorschriften, nationale Strategien sowie Zuschüsse und Investitionen umfassen.<sup>14,15</sup>

Im Großen und Ganzen zielen diese Bemühungen darauf ab, die Auswirkungen von KI zu verstehen, Innovationen voranzutreiben und durch politische Maßnahmen einen verantwortungsvollen Entwicklungsprozess zu gestalten. Die Gesetzgebung zu KI ist noch lange nicht abgeschlossen, doch bereits jetzt dienen einige aktuelle regulatorische Änderungen Unternehmen als nützliche Momentaufnahme, um Trends zu verstehen.

## Vereinigte Staaten von Amerika

In den USA liegt der Fokus insbesondere auf der vom Weißen Haus erlassenen Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence<sup>16</sup>. Die Richtlinie verpflichtet Entwickler der größten KI-Systeme dazu, dem US-Handelsministerium Ergebnisse von Sicherheitstests zu melden und offenzulegen, wenn umfangreiche neue Computing-Ressourcen verwendet werden, um KI-Modelle zu trainieren. Darüber hinaus mussten neun Bundesbehörden Risikobewertungen zu den Auswirkungen von KI auf kritische Infrastrukturen durchführen. Das Weiße Haus berücksichtigt daneben auch die Frage der KI-Innovation: Im Rahmen der Executive Order hat die US-Regierung das Pilotprogramm National Artificial Intelligence Research Resource (NAIRR) ins Leben gerufen, um US-Forscher mit Rechenleistung, Daten und anderen Werkzeugen zur Entwicklung von KI auszustatten.<sup>17</sup>

Es bleibt abzuwarten, ob die US-Regierung verbindlichere Regelungen rund um KI anstreben wird. Bisher haben mindestens 15 führende KI-Unternehmen und fast 30 Gesundheitsorganisationen freiwillige Selbstverpflichtungen des Weißen Hauses rund um KI-Sicherheit unterzeichnet.<sup>18</sup> Unterdessen hat die Federal Trade Commission (FTC) den Einsatz von KI zur Nachahmung von Regierungsbehörden oder öffentlichen Unternehmen verboten und plant, die Regel auf Schutzmaßnahmen für private Individuen und Institutionen auszuweiten.<sup>19</sup> Berichten zufolge prüft das Weiße Haus auch die Möglichkeit, verpflichtende Wasserzeichen für KI-generierte Inhalte einzuführen.



14. OECD, [Policies, data and analysis for trustworthy artificial intelligence](#), abgerufen am 12. März 2024.

15. Deloitte, [The AI regulations that aren't being talked about](#), abgerufen am 12. März 2024.

16. White House, [Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence](#), 30. Oktober 2023.

17. NAIRR Pilot, [The National Artificial Intelligence Research Resource \(NAIRR\) Pilot](#), abgerufen am 12. März 2024.

18. Reuters, [Healthcare providers to join US plan to manage AI risks — White House](#), 14. Dezember 2023.

19. Pennsylvania Office of Attorney General, [FTC Bans Use of A.I. to Impersonate Government Agencies and Businesses](#), 26. Februar 2024.



## EU

Das Europäische Parlament hat kürzlich das KI-Gesetz verabschiedet, das die weltweit erste umfassende KI-Gesetzgebung darstellt. Es enthält eine Reihe strenger Vorschriften und Richtlinien für verschiedene Arten von KI-Anwendungen, klassifiziert nach Risiko und anwendbar auf diverse Branchen. Die Regelungen sollen voraussichtlich im Jahr 2026 in Kraft treten und erfordern beispielsweise, dass KI-Tools mit allgemeinem Verwendungszweck wie ChatGPT Transparenzanforderungen erfüllen müssen. So muss deutlich gemacht werden, dass Inhalte durch KI generiert wurden, Trainingsmodelle müssen nachweislich so konzipiert sein, dass die Generierung illegaler Inhalte verhindert wird, und Unternehmen müssen Zusammenfassungen urheberrechtlich geschützter Materialien bereitstellen, die für das Training verwendet werden.

Die Vorschriften sehen strengere Richtlinien für „Hochrisiko“-Anwendungen vor, wie sie beispielsweise in Konsumgütern wie Spielzeug, Luftfahrt, medizinischen Geräten und Fahrzeugen verwendet werden, sowie für KI, die sich auf bestimmte Bereiche wie kritische Infrastruktur, Personalbeschäftigung, Rechtsangelegenheiten und Immigration auswirkt. Darüber hinaus wird die EU KI-Anwendungen, die als inakzeptabel riskant erachtet werden, gänzlich verbieten, einschließlich solcher, die sensible biometrische Informationen verwenden, auf die Beeinflussung menschlichen Verhaltens und die Umgehung des freien Willens abzielen, Emotionserkennungssysteme am Arbeitsplatz und in Schulen nutzen oder ungezielte Gesichtsbilder aus dem Internet oder von Überwachungskameras auslesen.<sup>20</sup>

Viele Länder priorisieren auch KI-Investitionen. Singapur hat beispielsweise einen KI-Investitionsplan über 740 Millionen US-Dollar angekündigt, der als Teil der National AI Strategy 2.0 umgesetzt werden soll.<sup>21</sup> Dieser Plan soll KI-Innovationen vorantreiben, den Zugang zu fortschrittlichen Chips ermöglichen, die für KI erforderlich sind, und gleichzeitig durch Einrichtung von KI-Kompetenzzentren im Land sicherstellen, dass Unternehmen bestmöglich von der KI-Revolution profitieren können.

20. European Parliament, [EU AI Act: first regulation on artificial intelligence](#), 19. Dezember 2023.

21. CNBC, [Singapore's AI ambitions get a boost with \\$740 million investment plan](#), 19. Februar 2024.

# Prognosen zu KI-Bedrohungen

**Desinformation und Cyberangriffe aufgrund von KI stehen laut World Economic Global Risk Report auf Platz 2 und 5 der zehn größten globalen Risiken im Jahr 2024.<sup>22</sup>**

Da sich der Anwendungsbereich von KI rasant weiterentwickelt und verstärkt auch KI-generierte Videos und Bilder umfasst, werden diese Risiken nur noch zunehmen — ebenso wie unsere Fähigkeit, ihnen erfolgreich mithilfe von KI zu begegnen. Im Folgenden geben wir einen Überblick über die wichtigsten KI-bezogenen Risiken und Bedrohungen, die wir mit Blick auf den Rest des Jahres 2024 und darüber hinaus erwarten.

1

## Das KI-Dilemma staatlicher Akteure: KI-Bedrohungen vorantreiben, KI-Zugriff blockieren

Staatlich geförderte Bedrohungsgruppen verbindet eine komplexe Beziehung zu KI: Einerseits nutzen sie sie zur Entwicklung immer ausgefeilterer Bedrohungen und andererseits müssen sie ihren Zugriff auf regierungsfeindliche Inhalte verhindern.

Der Einsatz von KI-Tools durch staatlich geförderte Bedrohungsgruppen ist kein neues Phänomen, aber alles deutet auf ein massives Wachstum sowohl im Umfang als auch in der Komplexität hin. Berichte von Microsoft und OpenAI bestätigen diese beunruhigende Prognose und zeigen, dass sich von Ländern wie Russland, China, Nordkorea und Iran unterstützte Bedrohungsakteure eingehend mit der ChatGPT-Funktionalität befassen und diese ausgenutzt haben. Dies erstreckt sich über verschiedene Anwendungsfälle, darunter Spear-Phishing, Codegenerierung und -überprüfung sowie Übersetzung.

22. World Economic Forum, [Global Risks Report 2024: The risks are growing — but so is our capacity to respond](#), 10. Januar 2024.

23. ZDNet, [Cybercriminals are using Meta's Llama 2 AI](#), 21. Februar 2024.

Obwohl einige dieser Angriffe durch gezielte Intervention gestoppt werden konnten, sollten sich Unternehmen auf das Fortbestehen staatlich geförderter KI-Initiativen einstellen. Sie umfassen etwa den Einsatz beliebter KI-Tools, die Erstellung proprietärer LLMs und die Verbreitung uneingeschränkter ChatGPT-inspirierter Varianten wie FraudGPT oder WormGPT. Die Landschaft entwickelt sich weiter und zeichnet das Bild einer herausfordernden Ausgangssituation, in der staatlich geförderte Akteure KI fortwährend auf neuartige Weise nutzen, um komplexe neue Cyberbedrohungen zu schaffen.

2

## Dark-Chatbots und KI-gestützte Angriffe: KI-Missbrauch als wachsende Gefahr

KI-gestützte Angriffe werden im Laufe des Jahres aller Voraussicht nach zunehmen. Das Darknet dient hierbei als Nährboden für bösartige Chatbots wie WormGPT und FraudGPT und den daraus resultierenden Anstieg cyberkrimineller Aktivitäten.

Diese gefährlichen Tools werden bei der Ausführung von erweitertem Social-Engineering- und Phishing-Betrug sowie verschiedenen anderen Bedrohungen eine entscheidende Rolle spielen. Im Darknet hat der Austausch unter Cyberkriminellen, die sich mit dem illegalen Einsatz von ChatGPT und anderen generativen KI-Tools für eine Vielzahl von Cyberangriffen befassen, einen Aufschwung erlebt. Mehr als 212 bösartige LLM-Anwendungen wurden identifiziert — eine Zahl, die nur einen Bruchteil der tatsächlich verfügbaren Anwendungen darstellt und Schätzungen zufolge weiter konstant ansteigen wird.

Wie Entwickler, die mithilfe generativer KI Effizienzsteigerungen erzielen, verwenden Bedrohungsakteure diese Tools, um Schwachstellen aufzudecken und auszunutzen, überzeugende Phishing-Strategien zu konzipieren, Vishing- und Smishing-Kampagnen umzusetzen und Angriffe automatisiert sowie schneller, ausgefeilter und umfangreicher durchzuführen. Beispielsweise nutzte die Bedrohungsgruppe Scattered Spider kürzlich das LLM LLaMa 2 von Meta, um das unbefugte Herunterladen von Zugangsdaten über Microsoft PowerShell zu ermöglichen.<sup>23</sup> Die aktuellen Fortschritte in diesem Bereich deuten darauf hin, dass sich Cyberbedrohungen schneller denn je weiterentwickeln und neue Formen annehmen werden, die mit herkömmlichen Sicherheitsmaßnahmen schwerer zu erkennen oder abzuwehren sind.



### 3 KI mit KI bekämpfen: KI-gestützte Maßnahmen bei Sicherheitsplänen und -investitionen

Unternehmen werden zunehmend KI-Technologien einsetzen, um KI-gestützte Cyberangriffe abzuwehren. Der Fokus liegt dabei unter anderem auf dem Einsatz von Deep-Learning- und KI/ML-Modellen zur Erkennung von Malware und Ransomware, die sich in verschlüsseltem Traffic verbirgt. Herkömmliche Erkennungsmethoden werden weiterhin nur schwer mit neuen KI-gestützten Zero-Day-Angriffen und polymorphen Ransomware zurechtkommen (die ihren Code weiterentwickeln kann, um der Erkennung zu entgehen). Daher gewinnen KI-basierte Indikatoren bei der Identifizierung potenzieller Bedrohungen entscheidend an Bedeutung. KI wird auch immer stärker benötigt, um überzeugende KI-generierte Phishing- und andere Social-Engineering-Angriffe schnell zu erkennen und zu stoppen.

Unternehmen werden KI zunehmend in ihre Cybersicherheitsstrategien integrieren. KI gilt als entscheidendes Mittel, um Einblick in Cyberrisiken zu gewinnen und umsetzbare, quantifizierbare Handlungsanweisungen zur Priorisierung und Behebung von Sicherheitslücken zu erstellen. Die Umsetzung von Datenrauschen in praktische Signale stellt für CISOs schon lange eine der größten Herausforderungen dar, da die Korrelation von Risiko- und Bedrohungsinformationen über Dutzende von Tools hinweg einen Monat oder mehr in Anspruch nehmen kann. Daher hoffen Unternehmen im Jahr 2024 auf das Potenzial generativer KI, um Ordnung ins Chaos zu bringen, Cyberrisiken abzuwehren und die Gestaltung schlanker, effizienter Sicherheitsorganisationen voranzutreiben.

### 4 Datenvergiftung in KI-Lieferketten: Steigendes Risiko von KI-Datenmüll

Mit verstärkten Angriffen auf KI-Lieferketten gibt insbesondere das Problem der Datenvergiftung Anlass zur Sorge. KI-Unternehmen sowie deren Trainingsmodelle und nachgelagerte Zulieferer werden zunehmend ins Visier böswilliger Akteure geraten.

Das OWASP hebt in ihrer Liste der zehn größten Gefahren bei LLM-Anwendungen die Vergiftung von Trainingsdaten und Angriffe auf die Lieferkette als erhebliche Risiken hervor, die die Sicherheit, Zuverlässigkeit und Leistung von KI-Anwendungen beeinträchtigen können. Gleichzeitig sind Schwachstellen in den Lieferketten von KI-Anwendungen — einschließlich Technologiepartnern, Datensätzen von Drittanbietern und KI-Tool-Plugins oder -APIs — lukrative Ziele für entsprechende Ausnutzung.

Unternehmen, die auf KI-Tools angewiesen sind und auf deren Sicherheit und Ergebnisse vertrauen, werden sich mit strengeren Prüfmaßnahmen konfrontiert sehen. Größere Wachsamkeit bei der Gewährleistung der Qualität, Integrität und Skalierbarkeit von Trainingsdatensätzen wird von entscheidender Bedeutung sein, insbesondere im Bereich der KI-Cybersicherheit.





## 1

## Freiheit oder Kontrolle: Abwägung von Produktivität und Sicherheit beim Einsatz von KI-Tools in Unternehmen

Mittlerweile haben zahlreiche Unternehmen die frühen Phasen der Einführung und Integration von KI-Tools hinter sich gelassen und zum großen Teil auch bereits sorgfältig entsprechende KI-Sicherheitsrichtlinien erarbeitet. Dennoch ist in dieser Situation für die meisten Unternehmen noch einiges ungeklärt und die Frage bleibt, welche KI-Tools zugelassen, welche blockiert und wie Daten geschützt werden sollten.

Angesichts der rasant steigenden Anzahl von KI-Tools müssen Unternehmen den Sicherheitsbedenken im Zusammenhang mit einzelnen Tools angemessene Aufmerksamkeit schenken. Zumindest sollten sie einen tiefgehenden Einblick in die KI-Nutzung ihrer Mitarbeiter gewinnen und über die Möglichkeit verfügen, granulare Zugriffskontrollen nach Abteilung, Team und sogar auf Userebene zu ermöglichen. Daneben werden für Unternehmen möglicherweise auch detailliertere Sicherheitskontrollen für KI-Anwendungen selbst wichtig. Denkbar ist die Durchsetzung von Datenschutzrichtlinien in KI-Anwendungen, um den Verlust vertraulicher Daten zu verhindern, und die Einschränkung von Useraktionen wie Kopieren und Einfügen.

## 1

## KI-gestützte Täuschung und Verzerrung: Virale Deepfakes als Treiber von Wahlmanipulation und Desinformationskampagnen

Neue Technologien wie Deepfakes gehen mit erheblichen Bedrohungen wie Wahleinmischung und Verbreitung von Desinformation einher. So war KI bereits Teil von Manipulationstaktiken im Zuge der US-Wahlen, wobei Bürger beispielsweise durch Robocalls, bei denen Kandidaten imitiert wurden, vom Wählen abgehalten werden sollten. Diese Vorfälle sind zwar alarmierend, stellen aber wahrscheinlich nur die Spitze des Eisbergs KI-gestützter Desinformation dar.

Darüber hinaus muss der Einsatz von KI in solchen Kampagnen nicht auf inländische Akteure beschränkt sein. Auch staatlich beauftragte Entitäten könnten diese Taktiken nutzen, um Verwirrung zu stiften und das Vertrauen in den Wahlprozess zu untergraben. In einem bemerkenswerten Fall nutzten Angreifer KI-generierte Deepfakes, um einen Mitarbeiter zur Überweisung von 25 Millionen US-Dollar zu bewegen — eine eindrückliche Demonstration der realen Auswirkungen dieser Technologie. In ähnlicher Weise macht die rasante Verbreitung illegaler Deepfake-Bilder von Prominenten wie Taylor Swift in den sozialen Medien deutlich, wie leicht manipulierte Inhalte um sich greifen können, bevor Maßnahmen zur Inhaltsmoderation wirksam werden.



# Fallstudie: Sichere Nutzung von ChatGPT im Unternehmen

## Best Practices für die KI-Integration und unternehmensspezifische Sicherheitsrichtlinien.

Die Schnittmengen von Unternehmen mit KI-Tools sind mittlerweile zahlreich. Da die Zahl der KI-Anwendungen jedoch weiterhin dramatisch zunimmt und ihre Verbreitung rasch voranschreitet, sollten sich Unternehmen an bestimmte Best Practices halten, um die Sicherheit ihrer Daten, Mitarbeiter und Kunden zu gewährleisten. Ganz allgemein müssen Unternehmen ihre Strategien zur Nutzung und Sicherheit von KI proaktiv und kontinuierlich anpassen, um aufkommenden Risiken einen Schritt voraus zu sein und gleichzeitig das transformative Potenzial der KI zu nutzen.



## FALLSTUDIE

### 5 Schritte zur Integration und Absicherung generativer KI-Tools

Unternehmen, die KI-Anwendungen auf sichere Weise einführen möchten, sollten besonnen vorgehen. Vereinfacht gesagt ist es sinnvoll, zunächst alle KI-Anwendungen zu blockieren und so das Risiko von Datenlecks zu beseitigen. Anschließend können wohlüberlegte Schritte zur Einführung einzelner geprüfter KI-Anwendungen mit strengen Sicherheitsrichtlinien und Zugriffskontrollen unternommen werden, um die vollständige Kontrolle über Unternehmensdaten zu behalten. Der Einfachheit halber konzentriert sich der folgende Leitfaden auf das LLM ChatGPT von OpenAI.

#### Schritt 1: Blockierung aller KI- und ML-Domains und -Anwendungen

Um bekannte und unbekannte Risiken im Zusammenhang mit den Tausenden verfügbaren KI-Anwendungen zu beseitigen, können Unternehmen einen proaktiven Zero-Trust-Ansatz verfolgen und alle KI- und ML-Domains und -Anwendungen auf globaler Unternehmensebene blockieren. Dies ermöglicht es ihnen, sich auf die Einführung einiger weniger transformativer KI-Anwendungen zu konzentrieren und gleichzeitig die damit einhergehenden Risiken genau zu kontrollieren.

#### Schritt 2: Selektive Prüfung und Genehmigung generativer KI-Anwendungen

Als Nächstes sollte die Organisation eine Reihe generativer KI-Anwendungen identifizieren, die in Bezug auf bestimmte Kriterien hohe Standards übertreffen. Dazu zählt etwa die Fähigkeit, zuverlässig Datenschutz, Sicherheit und vertragliche Maßnahmen zum Schutz von Unternehmens- und Kundendaten zu gewährleisten, sowie das transformative Potenzial der Anwendungen selbst. Für viele Unternehmen wird ChatGPT eine dieser Anwendungen sein.

#### Schritt 3: Erstellung einer privaten ChatGPT-Serverinstanz in der Unternehmens-/Rechenzentrumsumgebung

Um die vollständige Kontrolle über ihre Daten zu gewährleisten, empfiehlt es sich für Organisationen, ChatGPT in einer dedizierten, sicheren Instanz (z. B. einem privaten KI-Server von Microsoft Azure) zu hosten, die sich vollständig innerhalb der Organisation befindet.

Anschließend sollten sie durch Sicherheitskontrollen und vertragliche Verpflichtungen sowohl verhindern, dass Microsoft und OpenAI (in diesem Beispiel) Zugriff auf Unternehmens- oder Kundendaten haben, als auch sicherstellen, dass Anfragen von Unternehmensanwendern nicht zum allgemeinen Training von ChatGPT verwendet werden. Dies stellt sicher, dass die Organisation die Kontrolle über ihre Trainingsdaten behält und ermöglicht besonders relevante, genaue Antworten für Unternehmensuser bei gleichzeitig reduziertem Risiko einer Datenvergiftung durch einen öffentlichen Data Lake.

#### Schritt 4: Umstellung von Single Sign-on (SSO) für LLMs auf starke Multi-Faktor-Authentifizierung (MFA)

Als Nächstes sollte die Organisation ChatGPT in eine Cloud-Proxy-Architektur mit Zero-Trust-Ansatz wie die Zscaler Zero Trust Exchange integrieren, um Zero-Trust-Sicherheitskontrollen für den Zugriff auf ChatGPT durchzusetzen. Dies könnte auch die Verlagerung von ChatGPT hinter einen Identitätsanbieter (Identity Provider, IdP) umfassen, einschließlich SSO-Authentifizierung und starker MFA mit biometrischer Authentifizierung. Auf diese Weise sind eine sichere und schnelle Useranmeldung bei ChatGPT sowie die Konfiguration detaillierter Zugriffskontrollen auf User-, Team- und Abteilungsebene für das Unternehmen möglich. Außerdem lassen sich dadurch Probleme zwischen Useranfragen auf eben diesen User-, Team- und Abteilungsebenen trennen.

Durch das Vorschalten eines Cloud-Proxys wie der Zero Trust Exchange für ChatGPT kann das Unternehmen außerdem den gesamten TLS/SSL-Traffic zwischen Usern und ChatGPT überprüfen, um Cyberbedrohungen und Datenlecks zu erkennen, wobei Zero-Trust-Sicherheit auf sieben Ebenen zur Anwendung kommt.

#### Schritt 5: Durchsetzung der DLP-Engine von Zscaler zur Verhinderung von Datenverlust

Zuletzt sollte die Organisation eine DLP-Engine für die ChatGPT-Instanz einrichten, um einen versehentlichen Verlust kritischer Informationen wie proprietärer Daten und Codes, Kundendaten, personenbezogener Daten oder Finanz- und Rechtsdaten zu verhindern. Dadurch wird sichergestellt, dass hochsensible Daten niemals die Produktivumgebung verlassen.

Durch Befolgung dieses Leitfadens können Unternehmensanwender alle Vorteile eines generativen KI-Tools wie ChatGPT nutzen und gleichzeitig die kritischsten Datenrisiken bei der Einführung einer KI-Anwendung eliminieren.

## Best Practices für KI

Ganz allgemein ist es für Unternehmen sinnvoll, bei der Integration von KI-Tools einige wichtige Best Practices einzuführen.

- **Bewerten und kontrollieren Sie kontinuierlich KI-bezogene Risiken**, um geistiges Eigentum, personenbezogene Daten und Kundeninformationen zu schützen.
- **Stellen Sie bei der Verwendung von KI-Tools die Einhaltung einschlägiger Gesetze und ethischer Standards sicher**, einschließlich Bestimmungen zur Data Protection.
- **Legen Sie klare Verantwortlichkeiten für die Entwicklung und Bereitstellung von KI-Tools fest**, einschließlich definierter Rollen und Zuständigkeiten für die Überwachung von KI-Projekten.
- **Sorgen Sie beim Einsatz von KI-Tools für Transparenz**, indem Sie gegenüber Stakeholdern deren Einsatz begründen und ihren Zweck klar kommunizieren.

## Umfassende Richtlinien für KI

Zusätzlich zur Befolgung dieser Best Practices sollten Unternehmen einen klaren Richtlinienrahmen schaffen, der unternehmensweit die akzeptable Nutzung, Integration und Produktentwicklung, Sicherheits- und Datenbestimmungen sowie Best Practices für Mitarbeiter beim Einsatz von KI-Tools regelt. Die folgenden Best Practices dienen als Ausgangspunkt für die Festlegung klar definierter KI-Richtlinien.

- **KI-Modellen sollte kein Zugriff auf personenbezogene Daten gewährt werden** oder auf andere nicht öffentliche, proprietäre oder vertrauliche Informationen.
- **KI kann den Menschen nicht ersetzen** und sollte nicht ohne entsprechende menschliche Beteiligung zur Entscheidungsfindung eingesetzt werden.
- **KI-generierte Inhalte sollten nicht ohne menschliche Prüfung und Genehmigung verwendet werden**, insbesondere dann nicht, wenn die Inhalte die eigene Organisation repräsentieren.
- **Die Entwicklung und Übernahme von KI-Tools sollte einem Secure Product Lifecycle Framework folgen**, das höchste Sicherheit gewährleistet.
- **Vor der Implementierung von KI-Lösungen ist es ratsam, eine gründliche produktbezogene Due-Diligence-Prüfung durchzuführen** und dabei Sicherheit und ethische Aspekte zu berücksichtigen.

# So ermöglicht Zscaler sichere Bereitstellung generativer KI mit Zero Trust

Das transformative Potenzial von KI im Bereich Cybersicherheit liegt in ihrer Nutzung zur Bekämpfung KI-gestützter Bedrohungen, die sich derzeit dynamisch entwickeln. Wir bei Zscaler unterstützen Unternehmen mithilfe von KI dabei, Bedrohungen in sämtlichen Phasen der Angriffskette zu stoppen sowie Risiken einfach zu diagnostizieren und zu mindern.

## Der Schlüssel zur KI-gestützten Cybersicherheit: Hochwertige Daten in großem Maßstab

Unternehmen erzeugen eine große Fülle an Protokolldaten einschließlich hochpräziser Signale, die auf naheliegende Angriffswege hinweisen können. Herausforderungen im Zusammenhang mit dem Signal-Rausch-Verhältnis haben es jedoch in der Vergangenheit schwierig gemacht, diese Signale schnell zu isolieren. Mithilfe generativer KI kann Zscaler diese Daten nutzen, um Diagnose- und Schutzmaßnahmen wirksam zu verbessern. Dabei werden die Sicherheitslücken und Schwachstellen analysiert, die Angreifer wahrscheinlich ausnutzen. Dies ermöglicht es Zscaler nicht nur, Vorfälle vorausschauend vorherzusagen, sondern bietet Führungskräften auch einen ganzheitlichen Ansatz zur Visualisierung und Quantifizierung von Reifegrad und Risiken der Cybersicherheit, während gleichzeitig Maßnahmen zur Behebung von Sicherheitsproblemen mit Zscaler Risk360 priorisiert werden.

Funktionen generativer KI spielen nicht nur bei der Metaanalyse von Cyberrisiken in Unternehmen eine Rolle, sondern werden auch direkt in Cybersicherheitsprodukte eingebettet, um komplexe Bedrohungen entlang der Angriffskette besser zu erkennen und abzuwehren. Die LLMs und KI-Modelle von Zscaler sind direkt in die weltweit größte Security Cloud integriert und nutzen einen Data Lake, der mehr als 390 Milliarden tägliche Transaktionen mit mehr als 9 Millionen blockierten Bedrohungen und 300 Billionen Signalen umfasst. Statt „Garbage In, Garbage Out“ gilt hier: Zuverlässige Daten und Bedrohungsinformationen werden in großem Maßstab verarbeitet, um präzise, hochwachsame KI-Cybersicherheit bereitzustellen. All dies führt zu leistungsfähigeren und effektiveren Ergebnissen in IT- und Sicherheitsteams.



## Nutzung von KI in sämtlichen Angriffsphasen

Wir haben bereits beleuchtet, auf welch vielfältige Weise Bedrohungsakteure KI nutzen, um komplexe Angriffe schneller und in größerem Umfang durchzuführen. Zscaler setzt KI-Funktionen auf der Zero-Trust-Exchange-Plattform und in der Cyber-Produktsuite ein, um sowohl KI-gestützte als auch konventionelle Angriffe in jeder Phase der Angriffskette zu erkennen und zu stoppen.

### Phase 1: Erkennung von Angriffsflächen

In der ersten Phase eines Cyberangriffs untersuchen Bedrohungsakteure in der Regel die mit dem Internet verbundene Angriffsfläche des Unternehmens, um ausnutzbare Schwachstellen zu identifizieren. Dazu gehören häufig etwa Sicherheitslücken und Fehlkonfigurationen bei VPNs oder Firewalls sowie nicht gepatchte Server. Generative KI hat diese einst mühsame Aufgabe für Angreifer erheblich erleichtert, da sie nun einfach eine Liste bekannter Schwachstellen im Zusammenhang mit den jeweiligen Assets abfragen können.

Durch die Nutzung KI-gestützter Erkenntnisse in Zscaler Risk360 können Unternehmen diese auffindbaren (und damit risikoreichen) Anwendungen und Assets — ihre mit dem Internet verbundene Angriffsfläche — sofort erkennen und sie hinter der Zero Trust Exchange vor dem öffentlichen Internet verbergen. Dies reduziert die Angriffsfläche des Unternehmens sofort und in ausschlaggebendem Maße und verhindert gleichzeitig nachhaltig, dass Angreifer günstige Eintrittspunkte entdecken.

### Phase 2: Kompromittierungsrisiko

In der Kompromittierungsphase arbeiten Angreifer daran, Schwachstellen auszunutzen, um sich unbefugten Zugriff auf unternehmenseigene Systeme oder Anwendungen zu verschaffen. Die KI-Innovationen von Zscaler tragen dazu bei, das Risiko einer Kompromittierung zu verringern, indem sie raffinierte Angriffe abwehren und gleichzeitig die Produktivität priorisieren.

## KI-GESTÜTZTE PHISHING- UND C2-ABWEHR

KI-Modelle von Zscaler identifizieren bekannte und Patient-Zero-Phishing-Websites, um den Diebstahl von Zugangsdaten und Browser-Exploits zu verhindern, und analysieren Traffic- und Verhaltensmuster sowie Malware, um bislang unbekannte Command-and-Control-Infrastruktur (C2) in Echtzeit zu erkennen. Diese Modelle basieren auf einer Kombination aus Bedrohungsinformationen, ThreatLabz-Forschung und dynamischer Browser-Isolierung zur Erkennung verdächtiger Websites. Sie ermöglichen es Unternehmen, neue Phishing-Angriffe noch effizienter und effektiver zu erkennen, einschließlich KI-gestützter Angriffe und C2-Domains.

## DATEIBASIERTES KI-SANDBOXING

Die KI-gestützte Inline-Sandbox von Zscaler erkennt schädliche Dateien sofort und sorgt gleichzeitig dafür, dass Mitarbeiter weiter produktiv arbeiten können. Herkömmliche Sandbox-Technologien lassen User warten, während Dateien analysiert werden, oder gehen von einem Patient-Zero-Risiko aus, wenn Dateien beim ersten Versuch zugelassen werden. Mit Instant Verdict verfügen wir über eine KI-Technologie zur sofortigen Identifizierung, Isolierung und Blockierung hochverdächtiger Schaddateien — einschließlich Zero-Day-Bedrohungen —, bei der User gleichzeitig nicht mehr die Dateianalyse abwarten müssen. Dies umfasst auch Bedrohungen, die über verschlüsselte Kanäle (TLS und HTTPS) und andere Dateiübertragungsprotokolle übermittelt werden. Harmlose Dateien werden weiterhin sicher und sofort zugestellt.

## KI ZUM BLOCKIEREN VON WEB-BEDROHUNGEN

Die KI-gestützte Browser-Isolierung von Zscaler blockiert Zero-Day-Bedrohungen und stellt gleichzeitig sicher, dass Mitarbeiter auf die richtigen Websites zugreifen können, um ihre Arbeit zu erledigen. In der Praxis erfordert die URL-Filterung in Unternehmen häufig detailliertere Kontrollen als Zulassen und Blockieren. Gesperrte Websites sind oft sicher und für die Arbeit erforderlich, sodass es zu unnötigen Helpdesk-Tickets kommt. Unsere KI-gestützte Smart-Isolierung erkennt, wann eine Website ein Risiko birgt, und isoliert sie für den User — die Website wird als Pixel in einer sicheren, containerisierten Umgebung gestreamt. Dadurch werden webbasierte Bedrohungen wie Malware, Ransomware, Phishing und Drive-by-Downloads effektiv gestoppt und ein starker Web-Sicherheitsstatus geschaffen, ohne dass Unternehmen Websites standardmäßig blockieren müssen.



### Phase 3: Laterale Bewegung

Sobald sich Angreifer in einer Organisation eingenistet haben, beginnen sie mit der lateralen Ausbreitung, um auf sensible Daten und Anwendungen zuzugreifen. In vielen Organisationen haben User zudem übermäßigen Zugriff auf Dutzende kritischer Anwendungen, was mit einer enormen internen Angriffsfläche einhergeht.

Mithilfe seiner KI-Funktionen begrenzt Zscaler das Schadenspotenzial bei Angriffen. Das gelingt, indem Zugriffsmuster von Usern analysiert und Richtlinien zur intelligenten Anwendungssegmentierung empfohlen werden, um das Risiko lateraler Ausbreitung zu reduzieren. In der Regel lässt sich beispielsweise feststellen, dass nur 200 von 30.000 Usern, die Zugriff auf eine Finanzanwendung haben, diese tatsächlich benötigen. Zscaler kann automatisch ein Anwendungssegment erstellen, das den Zugriff auf diese 200 Mitarbeiter beschränkt und so die Möglichkeiten der Bedrohungsakteure zur lateralen Ausbreitung um mehr als 99 % reduziert.

### Phase 4: Datenexfiltration

In der letzten Phase eines Angriffs arbeiten Bedrohungsakteure daran, sensible Daten herauszufiltern. Zscaler nutzt KI, um Organisationen die schnellere Umsetzung von Data Protection zu ermöglichen. Durch die KI-gestützte Datenerkennung entfällt das zeitaufwendige Fingerprinting und Klassifizieren von Daten, wodurch andernfalls die Bereitstellung verzögert oder verhindert werden könnte. Die KI von Zscaler erkennt und klassifiziert automatisch alle internen Daten eines Unternehmens und ermöglicht die sofortige Einordnung vertraulicher Informationen sowie die Konfiguration von Richtlinien für DLP (Data Loss Prevention), um zu verhindern, dass diese Daten das Unternehmen bei einem Angriff oder einer Sicherheitsverletzung verlassen.

## Zusammenfassung der KI-gestützten Lösungen von Zscaler

Zscaler Internet Access™ bietet als Teil der Zero Trust Exchange KI-gestützten Schutz für Unternehmensanwender, Geräte sowie Web- und SaaS-Anwendungen an allen Standorten. Dabei werden folgende Funktionen abgedeckt:

- **KI-gestützte Phishing- und C2-Erkennung** gegen bislang unbekannte Phishing-Websites und C2-Infrastruktur mithilfe der KI-basierten Inline-Erkennung des Zscaler Secure Web Gateway (SWG).
- KI-gestütztes Sandboxing mit umfassender Malware- und Zero-Day-Bedrohungsprävention.
- **Dynamische, risikobasierte Richtlinie** mit kontinuierlicher Analyse des User-, Geräte-, Anwendungs- und Inhaltsrisikos zur Förderung dynamischer Sicherheits- und Zugriffsrichtlinien.
- **KI-gestützte Segmentierung** mit Zscaler Private Access™ auf Basis automatisierter Zugriffsrichtlinien, um die Angriffsfläche zu minimieren und laterale Ausbreitung mithilfe von Userkontext, Verhalten, Standort und Telemetrie für unternehmensinterne Anwendungen zu stoppen.
- KI-gestützte Browser-Isolierung, die User durch einen undurchdringlichen Air Gap vom Internet isoliert; Web-Inhalte werden als gestochen scharfer Bilderstrom angezeigt und das Risiko von Datenverlusten und aktiven Bedrohungen ausgeschaltet.

### DARÜBER HINAUS BLOCKIERT ZSCALER:

**URLs und IPs**, die in der Zscaler Cloud beobachtet und als verdächtig eingestuft wurden bzw. in nativ integrierten Open-Source- und kommerziellen Datenbanken verzeichnet sind. Dazu gehören URL-Kategorien, die in den Richtlinien als hochriskant definiert sind und häufig für Phishing-Angriffe verwendet werden – z. B. neu beobachtete und neu aktivierte Domains.

**IPS-Signaturen**, die von ThreatLabz aus der Analyse von Phishing-Kits und -Seiten ermittelt wurden.

**Zscaler Risk360** bietet ein umfassendes und umsetzbares Risiko-Framework, das Sicherheitsverantwortlichen und Führungskräften dabei hilft, Cyberrisiken im gesamten Unternehmen zu quantifizieren und zu visualisieren.

**Data Protection mit DLP und CASB** bietet Datenklassifizierung und Data Protection mit KI-Unterstützung über alle Kanäle hinweg, einschließlich Endgeräte, E-Mails, Workloads, BYOD und Cloud-Posture.

**Advanced Threat Protection** blockiert alle bekannten Command-and-Control-Domains.

**Zscaler ITDR** (Identity Threat Detection and Response) verringert das Risiko identitätsbasierter Angriffe dank durchgehender Transparenz, Risikoüberwachung und Bedrohungserkennung.

**Zscaler Firewall** erweitert den C2-Schutz auf alle Ports und Protokolle, einschließlich neuer C2-Ziele.

**DNS Security** schützt vor DNS-basierten Angriffen und Exfiltrationsversuchen.

**Zscaler Private Access™** schützt Anwendungen, indem durch Mikrosegmentierung von User-to-App-Verbindungen und kompletter Inline-Überprüfung des Traffics von unternehmensinternen Anwendungen die laterale Bewegungsfreiheit von Angreifern eingeschränkt wird.

**AppProtection** mit Zscaler Private Access bietet eine leistungsstarke Inline-Sicherheitsprüfung der gesamten Anwendungsnutzlast, um Bedrohungen aufzudecken.

**Zscaler Deception™** ködert und entlarvt Angreifer beim Versuch, sich lateral durchs Netzwerk zu bewegen bzw. sich erweiterte Zugriffsberechtigungen zu verschaffen, mit Decoys, die echten Servern, Anwendungen, Verzeichnissen und User-Konten täuschend ähnlich sehen.



## Erfolgreicher KI-Umstieg im Unternehmen: Es liegt in Ihren Händen

Zscaler bietet Unternehmen die Möglichkeit, Innovation, Kreativität und Produktivität mit KI-Anwendungen zu fördern und gleichzeitig User und Daten vor neuen Datenexfiltrationskanälen zu schützen. So können Unternehmen **das transformative Potenzial von KI nutzen** und ihr Geschäft stärken, ohne KI-Anwendungen und –Domains gänzlich blockieren zu müssen.



### MIT ZSCALER PROFITIEREN UNTERNEHMEN VON FOLGENDEN VORTEILEN:

- 01 **Vollständige Transparenz bei der Nutzung von KI-Tools** Detaillierte Protokolle bieten vollständige Transparenz darüber, wie Unternehmensteams KI nutzen, einschließlich der von ihnen besuchten Anwendungen und Domains sowie der Daten und Prompts, die in Tools wie ChatGPT verwendet werden.
- 02 **Erstellung flexibler Richtlinien zur Optimierung des KI-Einsatzes** Mit leistungstarker, maßgeschneiderter URL-Filterung für KI- und ML-Anwendungen können Unternehmen auf einfache Weise detaillierte KI-Zugriffskontrollen und –Segmentierung definieren und durchsetzen. Der Zugriff wird dabei wenn nötig blockiert und bei vertretbarem Risiko auf Basis des Risk Scores zugelassen. Unternehmen können den Zugriff auf Unternehmens-, Abteilungs-, Team- und Userebene sowie unter Einhaltung bestimmter Sicherheitsmaßnahmen ermöglichen, der die Risiken generativer KI-Tools für User einbezieht. Die KI-gestützte Segmentierung erleichtert die Identifizierung geeigneter Usersegmente für den Zugang zu bestimmten KI-Anwendungen und minimiert gleichzeitig die interne Angriffsfläche, die mit KI-Tools verbunden ist.
- 03 **Durchsetzung granularer Datensicherheit für ChatGPT und andere KI-Anwendungen** Unternehmen können den Verlust sensibler Daten, die in KI-Anwendungen hochgeladen werden, mit granularen Kontrollen der Zscaler Cloud Application für generative KI verhindern. Durch den Einsatz der DLP-Engine von Zscaler können Unternehmen sicherstellen, dass bei der Verwendung eines KI-Tools keine Daten versehentlich weitergegeben werden. Gleichzeitig können Unternehmen dank der KI-gestützten Datenerkennung und –klassifizierung ganz einfach DLP-Richtlinien für ihre kritischsten Daten identifizieren und erstellen, einschließlich ihrer unternehmenseigenen Codebasis, Finanz- und Rechtsdokumente, personenbezogenen sowie Kundendaten. [Dieses Video](#) zeigt, wie die DLP-Engine User daran hindert, Kreditkarteninformationen in ChatGPT einzugeben.
- 04 **Einrichtung leistungstarker Kontrollen mithilfe von Browser Isolation** Zscaler Browser Isolation stellt KI-Anwendungen in einer sicheren Umgebung dar und ergänzt sie um eine Schutzebene, die Prompts und Anfragen in KI-Tools zulässt und gleichzeitig das Kopieren und Einfügen sowie Uploads und Downloads einschränkt. Dies verringert das Risiko, dass vertrauliche Daten versehentlich durch generative KI-Tools offengelegt werden.

**Führungskräfte und Sicherheitsverantwortliche stehen an einem Scheideweg:** Sie müssen auf KI setzen, um Innovationen voranzutreiben und wettbewerbsfähig zu bleiben, und gleichzeitig gewährleisten, dass es dabei nicht zu geschäftsschädigenden Sicherheitsverstößen kommt. Mit Zscaler können Unternehmen diesen Übergang souverän meistern, indem sie Zugriff auf eine vollständige Suite KI-gestützter Zero-Trust-Sicherheitskontrollen erhalten. Diese bieten Schutz vor KI-basierten Angriffen und gleichzeitig optimierte KI-bezogene Richtlinien und Data Protection, die es ermöglichen, das volle Potenzial generativer KI auszuschöpfen.

# Anhang

## Forschungsmethodik von ThreatLabz

Die globale Security Cloud von Zscaler verarbeitet täglich über 300 Billionen Signale, blockiert dabei 9 Milliarden Bedrohungen und Richtlinienverstöße und führt über 250.000 Sicherheitsupdates durch. Analyse von 18,09 Milliarden KI- und ML-Transaktionen von April 2023 bis Januar 2024 in der Cloud von Zscaler, der Zero Trust Exchange.

---

## Über Zscaler ThreatLabz

ThreatLabz ist als Forschungsabteilung von Zscaler für die Früherkennung neuer Bedrohungen zuständig. Dieses erstklassige Team sorgt dafür, dass die Tausenden von Organisationen, die weltweit mit der globalen Zscaler-Plattform arbeiten, jederzeit geschützt sind. Neben der Erforschung und Verhaltensanalyse von Malware-Bedrohungen tragen die ThreatLabz-Experten auch zur Entwicklung neuer Prototypen für Advanced Threat Protection auf der Zscaler-Plattform bei und führen regelmäßig interne Revisionen durch, um sicherzustellen, dass Zscaler-Produkte und -Infrastrukturen die geltenden Sicherheitsstandards erfüllen. Detaillierte Analysen neuer Bedrohungen werden regelmäßig unter [research.zscaler.com](https://research.zscaler.com) veröffentlicht.





# Erleben Sie Ihre Welt, sicher.

## Über Zscaler

Zscaler (NASDAQ: ZS) beschleunigt die digitale Transformation, damit Kunden agiler, effizienter, resilienter und sicherer arbeiten können. Die Zscaler Zero Trust Exchange™ schützt tausende Kunden mittels sicherer Verbindungen zwischen Usern, Geräten und Anwendungen überall vor Cyberangriffen und Datenverlust. Die SASE-basierte Zero Trust Exchange ist weltweit in 150 Rechenzentren verfügbar und ist somit die größte Inline-Cloud-Sicherheitsplattform der Welt. Weitere Informationen finden Sie unter [www.zscaler.de](https://www.zscaler.de).

© 2024 Zscaler, Inc. Alle Rechte vorbehalten. Zscaler™, Zero Trust Exchange™, Zscaler Internet Access™, ZIA™, Zscaler Private Access™, ZPA™ und weitere unter [zscaler.de/legal/trademarks](https://www.zscaler.de/legal/trademarks) aufgeführte Marken sind entweder (i) eingetragene Marken bzw. Dienstleistungsmarken oder (ii) Marken bzw. Dienstleistungsmarken von Zscaler, Inc. in den USA und/oder anderen Ländern. Alle anderen Marken sind das Eigentum ihrer jeweiligen Inhaber.