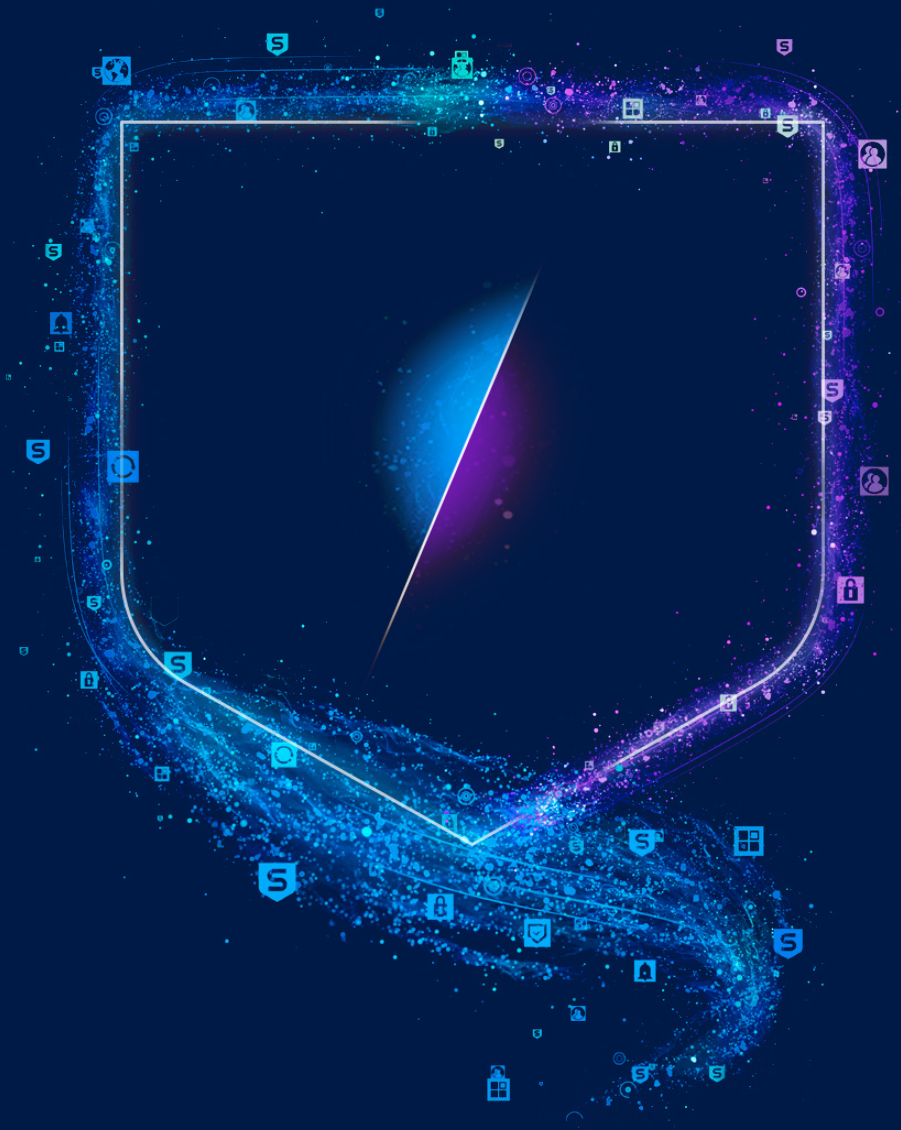
SOPHOS

# KI in der Cybersecurity: Hype oder Allheilmittel?

## So setzen Sie KI optimal und sicher zur Stärkung Ihrer Cyberabwehr ein



# Inhaltsverzeichnis

# Einleitung

Das Thema KI erfährt in der Cybersicherheit aktuell viel Aufmerksamkeit. Unternehmen werden mit verlockenden Versprechungen einer KI-gestützten Transformation der Cybersecurity geradezu bombardiert: mehr Schutz, niedrigere Kosten, ein geringerer Bedarf an Fachkräften. Gleichzeitig wird davor gewarnt, dass KI eine völlig neue Ära von Cyberangriffen einläuten wird.

Dieser Leitfaden soll Unternehmen dabei helfen, den Hype und die Missverständnisse rund um KI in der Cybersicherheit besser einzuschätzen. Sie erfahren, was KI leistet (und was nicht), um die Cyberabwehr in Unternehmen zu optimieren, und welche Risiken KI für Cybersicherheit und Betriebsabläufe mit sich bringt. Dazu erhalten Sie Tipps, wie Sie mögliche Gefahren minimieren und so die Vorteile von KI sicher nutzen können, um sowohl Ihren Cyberschutz als auch Ihren Return on Investment zu verbessern.

Zudem liefert der Guide Einblicke in die Praxis: Wie sieht die KI-Nutzung in der Realität aus? Welche Erwartungen und Bedenken bestehen? Die Erkenntnisse hierzu beruhen auf den Ergebnissen einer unabhängigen, Ende 2024 durchgeführten Befragung von 400 IT-/Cybersecurity-Entscheidern. Diese direkten Erfahrungsberichte bieten eine wertvolle Orientierungshilfe für Unternehmen, die über Einsatzmöglichkeiten von KI nachdenken. Die vollständigen Ergebnisse finden Sie im Artikel KI-gestützte Cybersicherheit: Was Unternehmen wirklich erwarten (können)

Mit oder ohne KI – das Ziel bleibt schlussendlich dasselbe: das erforderliche Maß an Cyberresilienz optimal bereitstellen und gleichzeitig die Gesamtausgaben minimieren. Anders ausgedrückt: Unternehmen möchten das (stets begrenzte) Cybersecurity-Budget bestmöglich einsetzen. Unser Leitfaden unterstützt Sie dabei, dieses Ziel im KI-Zeitalter zu erreichen.

# Die Vorteile von KI für die Cybersecurity

Hinter der Abkürzung KI stecken eine Reihe von Funktionen, die Cybersicherheit in vielerlei Hinsicht unterstützen und beschleunigen können. Die gute Nachricht: KI bietet mehr Vorteile für die Cyberabwehr als für die Bedrohungsakteure. Zwei gängige KI-Konzepte, die in der Cybersecurity eingesetzt werden, sind Deep-Learning-Modelle und generative KI.

## Deep Learning

Deep-Learning-Modelle (DL) können das Gelernte bei der Ausführung von Aufgaben ANWENDEN. Dabei übertrifft die KI menschliche Fähigkeiten in puncto Schnelligkeit. So können entsprechend geschulte DL-Modelle etwa in Sekundenschnelle erkennen, ob eine Datei bösartig oder unbedenklich ist, ohne diese Datei jemals zuvor gesehen zu haben.

DL eignet sich ideal für die Durchführung sich wiederholender Aufgaben in großem Umfang. Dabei erstellt DL ein statistisches Modell, das neue Elemente betrachtet und dabei das aus einem sehr großen Trainingsdatensatz Erlernte berücksichtigt. So können DL-Modelle etwa mühelos Millionen von Dateimustern auf Malware durchsuchen. Aus diesem Grund wird DL häufig verwendet, um Schutzfunktionen von Cybersecurity-Lösungen zu optimieren.

DL-Modelle ermöglichen der Cyberabwehr, das immense Bedrohungsvolumen, das durch Automatisierung und Cybercrime-as-a-Service entstanden ist, erfolgreich zu bewältigen. Zudem lassen sich DL-Modelle aktualisieren und anpassen, wenn sich die Angriffe weiterentwickeln. So halten sie immer mit der aktuellen Bedrohungslage Schritt.

## Generative KI (GenAI)

GenAI-Modelle sind darauf ausgerichtet, anhand von Trainingsdaten neue Inhalte zu ERSTELLEN. Einige Anwendungsbeispiele:

- Bisherige Bedrohungsaktivitäten in natürlicher Sprache zusammenfassen und nächste Schritte für den Analysten empfehlen
- Einblicke in das Verhalten von Angreifern geben, indem Befehle analysiert werden, die Erkennungen auslösen
- Analysten ermöglichen, Suchen in natürlicher Sprache zu nutzen anstelle von codebasierten Abfragen, um verdächtige Erkennungen zu untersuchen
- Das Einspielen von Patches danach zu priorisieren, wie hoch die Wahrscheinlichkeit ist, dass eine Schwachstelle ausgenutzt wird

GenAI ist ein leistungsstarkes Tool, das Security Operations beschleunigt. Da GenAI einen Großteil der Datenauswertung übernehmen kann, können Analysten schnell fundierte Entscheidungen treffen und sich so auf kritische Aufgaben konzentrieren. Auf diese Weise entlastet GenAI Sicherheitsteams und reduziert das Risiko von Burnout und Mitarbeiterfluktuation. GenAI kann auch dazu beitragen, die technologischen Hürden im Bereich der Security Operations zu senken, sodass weniger erfahrene Analysten sich schnell positiv einbringen und ihre Kompetenzen erweitern können.

### Der Weg zu GenAI

Moderne GenAI basiert auf einer Transformer-Architektur. Dabei handelt es sich um ein neuronales Deep-Learning-Netzwerk, das den Kontext und die Beziehung zwischen Eingaben (z. B. die Wörter in einem Satz) lernt und das Gelernte nutzt, um relevante Ergebnisse zu erzeugen. Transformer kommen häufig bei NLP-Aufgaben (Natural Language Processing) zum Einsatz, z. B. bei der Übersetzung von Text und der Beantwortung von Fragen. So steht z. B. das T in ChatGPT für Transformer.

Transformer sind zwar im Bereich GenAI weit verbreitet, doch nicht alle Transformer sind generativ. BERT (Bidirectional Encoder Representations from Transformers) ist beispielsweise ein Open-Source-Machine-Learning-Framework für NLP, das Eingabetext bidirektional lesen kann, d. h. sowohl von links nach rechts als auch von rechts nach links. Auf diese Weise lässt sich das kontextuelle Verständnis von nicht kategorisiertem Text deutlich verbessern. Bei Sophos nutzen wir BERT bereits seit vielen Jahren zur Erkennung und Abwehr von Business-Email-Compromise-Angriffen.

## Keine Einheitslösung

KI-Modelle variieren stark in ihrem Umfang. **Große Modelle** wie Microsoft Copilot und Google Gemini sind große Sprachmodelle (Large Language Models, LLMs), die mit einem umfangreichen Datensatz trainiert werden und unterschiedlichste Aufgaben übernehmen können. **Kleine Modelle** dagegen werden auf Basis eines sehr spezifischen Datensatzes konzipiert und trainiert, um eine konkrete Aufgabe auszuführen, z. B. die Erkennung bösartiger URLs oder ausführbarer Dateien. Kleine Modelle sind zwar in ihrem Umfang begrenzt, bieten im Vergleich zu größeren Modellen jedoch Vorteile hinsichtlich Kosten, Geschwindigkeit und Leistung.

## Die Grenzen der KI

KI allein ist nicht die Lösung – zumindest nicht in absehbarer Zukunft. KI ergänzt menschliche Expertise, kann sie jedoch nicht vollständig ersetzen. Bedrohungen sind äußerst komplex. Effektive Sicherheitsmaßnahmen erfordern sowohl technische Kompetenz als auch die Fähigkeit, Erkenntnisse im Kontext des Unternehmens auszuwerten. KI allein kann Unternehmen keinen Vorsprung vor den fachlich versierten und finanzstarken Cyberkriminellen von heute verschaffen.

### Typ

#### Deep-Learning-KI — *Anwendung*

Erkennt Muster mithilfe künstlicher neuronaler Netzwerke und trifft so Entscheidungen ähnlich wie das menschliche Gehirn. Kann Gelerntes bei der Ausführung von Aufgaben ANWENDEN.

Beispiel: **Erkennung schädlicher URLs**
KI-Modelle werden zur Erkennung schädlicher Websites trainiert, sodass Sicherheitsprodukte den Zugriff darauf blockieren können

#### Generative KI — *ERSTELLUNG*

Nutzt die Struktur und das Muster vorhandener Daten, um vollkommen neue Inhalte zu ERSTELLEN (zu generieren)

Beispiel: **Zusammenfassung des Bedrohungsfalls**
KI-Modelle erstellen eine Zusammenfassung der Bedrohungsaktivitäten und liefern Empfehlungen zu den nächsten Schritten

### Größe

#### Große KI-Modelle

Vielseitig einsetzbare Tools, die mit riesigen Mengen öffentlicher Daten trainiert werden und unterschiedlichste Aufgaben unterstützen können.

Beispiel: **Microsoft Copilot, Google Gemini**

#### Kleine KI-Modelle

Ergebnisorientierte Modelle, die für bestimmte Anwendungsfälle konzipiert und trainiert werden.

Beispiel: **Erkennungsmodell zu Android-Malware**

## Nutzung von KI

KI ist bei den meisten Unternehmen bereits weitreichend in die Cybersecurity-Infrastruktur integriert:

- 73 % geben an, dass ihre Cybersecurity-Lösungen Deep-Learning-Modelle beinhalten
- 65 % geben an, dass ihre Cybersecurity-Lösungen generative KI-Funktionen beinhalten

Die Anwendung von KI für die Cybersecurity beschränkt sich nicht auf externe Anbieter. 34 % der Unternehmen nutzen KI bereits intern, um ihre Cyberabwehr zu stärken, z. B. für Phishing-Test-E-Mails.

Aller Wahrscheinlichkeit nach wird sich KI innerhalb kurzer Zeit durchsetzen: Bei der Auswahl einer Cybersecurity-Plattform stehen KI-Funktionen schon jetzt bei 99 % (gerundet) der Unternehmen auf der Anforderungsliste:

- 57 % halten KI-Funktionen für wesentlich/äußerst wichtig
- 41 % halten KI-Funktionen für wichtig

Angesichts dieser weitreichenden aktuellen und künftigen Nutzung ist es für Unternehmen aller Größen und Branchen unabdingbar, auch die Risiken der KI in der Cybersecurity zu kennen und zu wissen, wie sie diese minimieren können.

**73 %**
**nutzen Cybersecurity-Tools mit Deep-Learning-Modellen**

**65 %**
**nutzen Cybersecurity-Tools mit GenAI-Funktionen**

**99 %**
**verlangen KI-Funktionen bei der Auswahl einer Cybersecurity-Plattform**

## GenAI: Große Erwartungen

Der Hype um GenAI hat große Erwartungen geweckt, wie sich mit dieser Technologie bessere Cybersecurity-Ergebnisse erzielen lassen. Aus der Umfrage geht hervor, welche Vorteile sich Unternehmen vor allem von GenAI-Funktionen in Cybersecurity-Tools erhoffen (siehe Tabelle unten).

| **Wichtigster erhoffter Vorteil von generativer KI**<br>Häufigste Antworten | |
|---|---|
| **20 %** | Mehr Schutz vor Cyberbedrohungen |
| **20 %** | Höhere Rendite (ROI) aus Cybersecurity-Investitionen |
| **17 %** | Mehr Effizienz bei der Sicherheitsanalyse |
| **15 %** | Schritt halten mit Cybersecurity-Innovationen |
| **14 %** | Mehr Gewissheit, dass unser Unternehmen gut vor Angriffen geschützt ist |
| **14 %** | Weniger Burnout-Fälle bei den Mitarbeitern (d. h. Automatisierung von Aufgaben, um Kapazitäten in Sicherheitsteams freizusetzen |

*Welche Vorteile erhoffen Sie sich von generativen KI-Funktionen in Cybersecurity-Tools? Häufigste Antworten (Anzahl=400)*

Die große Bandbreite der Antworten zeigt, dass es keinen einzelnen, herausragenden Vorteil gibt, den sich Unternehmen von GenAI in der Cybersicherheit erhoffen. Die am meisten erhofften Vorteile beziehen sich auf einen besseren Schutz vor Cyberbedrohungen oder eine bessere Unternehmensleistung (sowohl finanziell als auch operativ). Zudem weisen die Daten darauf hin, dass GenAI-Funktionen in Cybersecurity-Lösungen Unternehmen die Zuversicht geben, dass sie mit den neuesten Schutzfunktionen Schritt halten können.

Dass die Reduzierung von Burnouts von Mitarbeitern relativ niedrig eingestuft wird, lässt vermuten, dass Unternehmen sich des Potenzials von GenAI zur Unterstützung der Mitarbeiter weniger bewusst sind oder sich weniger Gedanken darüber machen. Angesichts des Fachkräftemangels in der Cybersecurity ist die Verringerung der Fluktuation jedoch ein wichtiger Einsatzbereich von KI.

Mehr **Schutz** und höherer **ROI** zählen zu den wichtigsten Vorteilen, die sich Unternehmen von GenAI erhoffen

# Die Risiken von KI in der Cybersecurity

KI im Bereich Cybersicherheit hat auch ihre Schattenseiten. Zwar bietet KI enorme Vorteile bei der Abwehr von Cyberangriffen, sie birgt jedoch auch Risiken:

1. **Bedrohungsrisiko:** Einsatz von KI bei Cyberangriffen
2. **Abwehrrisiko:** Schlechte oder schlecht implementierte KI
3. **Betriebliches Risiko:** Zu große Abhängigkeit von KI
4. **Finanzielles Risiko:** Niedriger ROI aus Investitionen in KI
5. **Manipulations-Risiko:** Kompromittierung öffentlich zugänglicher KI-Modelle

## 1. Bedrohungsrisiko: Einsatz von KI bei Cyberangriffen

Zwar ist immer wieder zu hören, dass KI die Bedrohungslandschaft komplett verändern wird, doch die Realität ist weitaus weniger dramatisch. In Foren zur Cyberkriminalität wird nur wenig über KI diskutiert, denn viele Bedrohungsakteure stehen KI weiterhin skeptisch gegenüber. Beobachtete Versuche, Malware, Angriffswerkzeuge und Exploits mithilfe von KI zu entwickeln, sind in der Regel wenig ausgereift und von geringer Qualität.

Genau wie seriöse Unternehmen nutzen auch Angreifer KI in erster Linie, um die Qualität ihrer Inhalte und die Effizienz ihrer Prozesse zu verbessern. Natürlich verfolgen sie dabei sehr unterschiedliche Ziele. Mehr über die aktuelle Bedrohungslandschaft und KI-basierte Angriffe erfahren Sie in unserem Sophos News Blog.

### Qualitativ hochwertigere Inhalte

Besonders schnell und einfach kann KI Cyberkriminellen dabei helfen, die Qualität und Glaubwürdigkeit von Phishing-E-Mails und Scams zu erhöhen, sodass sie ihre Opfer einfacher täuschen können.

Klassische Phishing-Merkmale wie schlechte Grammatik, Rechtschreibfehler und schwache Formatierung lassen sich mit KI-Tools leicht beseitigen. Öffentlich zugängliche LLMs können eine gut geschriebene E-Mail für Phishing-Kampagnen in weniger als einer Minute erstellen. Auch überzeugende und gut geschriebene Texte und Nachrichten in sozialen Medien, die Benutzer zum Klicken auf Links oder zur Weitergabe persönlicher Daten verleiten sollen, sind jetzt leicht in beliebigen Sprachen verfügbar. Außerdem können Angreifer mit LLMs ganz einfach aktuelle Informationen in ihre Angriffe integrieren, wodurch Opfer eher auf den Betrug hereinfallen.

Generative KI-Tools haben auch eine weitere neue Form des Betrugs möglich gemacht: Dabei geben sich Cyberkriminelle als leitende Angestellte aus und bringen ahnungslose Mitarbeiter dazu, Geld zu überweisen. Voice Cloning ist mittlerweile so ausgereift, dass geschulte Angreifer Opfer davon überzeugen können, dass sie mit einer echten Person sprechen. Bei diesen Voice-Phishing- oder „Vishing"-Angriffen gibt sich ein Angreifer oft als hochrangige Führungskraft aus und ruft einen Mitarbeiter an, um ihn zu 'bitten', einen illegalen Geschenkkartenkauf, eine Banküberweisung oder einen Dateitransfer durchzuführen.

Außerdem nutzen Angreifer KI-gestützte Deepfake-Technologien, um bei ihren Angriffen das Aussehen echter Personen zu imitieren. Mit Deepfake-Videos werden ahnungslose Angestellte zu beträchtlichen Zahlungen verleitet. Auch Gesichtserkennungs-Programme für Kreditanträge und Bankkonto-Eröffnungen werden so überlistet.

### Effizientere Abläufe

Genau wie viele seriöse Unternehmen setzen auch Angreifer auf KI-gestützte Chatbots, um das Benutzererlebnis zu optimieren. Manche Bedrohungsakteure erstellen mit LLMs Chatbots und automatische Antworten für Foren. In einem vom Sophos X-Ops-Team untersuchten Beispiel hat das russische Dark-Web-Forum XSS einen speziellen Chatbot entwickelt, der Benutzerfragen beantwortet. Ankündigung des Administrators (aus dem Russischen übersetzt):

> *„In diesem Bereich können Sie mit KI (künstlicher Intelligenz) chatten. Stellen Sie eine Frage – unser KI-Bot antwortet … Dieser Bereich und der KI-Bot dienen zur Lösung einfacher technischer Probleme und bieten unseren Nutzern technisches Entertainment. Außerdem machen sie Nutzer mit den Möglichkeiten von KI vertraut."*

Der Aufbau und das Training von benutzerdefinierten Modellen erfordert umfangreiche KI-Expertise, die jedoch kostspielig und knapp bemessen ist. Zwar verfügen kriminelle Cyberbanden durchaus über Know-how im Bereich KI, doch nutzen Bedrohungsakteure bei ihren Angriffen in der Regel bestehende LLMs, anstatt eigene zu entwickeln.

### Das Toolkit von Cyberkriminellen

Der Einsatz von KI durch Angreifer muss im Kontext betrachtet werden. KI ist nur eines von vielen Werkzeugen im Toolkit der Angreifer. Bereits seit einigen Jahren weiten Bedrohungsakteure ihre Angriffe mit Automatisierung und Cybercrime-as-a-Service-Modellen aus. Diese Vorgehensweisen bergen für viele Unternehmen größere Risiken als KI.

## 2. Abwehrrisiko: Schlechte oder schlecht implementierte KI

Wie zuvor festgestellt, sind KI-Modelle bei den meisten Unternehmen bereits weitreichend in die Cybersecurity-Infrastruktur integriert. Qualitativ schlechte oder schlecht implementierte KI-Modelle können ein großes Risiko für die Cybersicherheit darstellen. Ob dies der Fall ist, hängt unter anderem von den folgenden Faktoren ab:

- **Qualität der Daten, anhand derer die Modelle trainiert werden**. Wie in vielen Bereichen gilt auch in puncto KI die Devise: „Garbage in, garbage out" (GIGO), d.h. die Qualität der Eingabedaten beeinflusst maßgeblich die Qualität der Ergebnisse. Werden Modelle mit qualitativ minderwertigen Daten trainiert, kann dies zu Fehlern führen, und unausgewogene Datensätze können durch die unverhältnismäßige Repräsentation bestimmter Variablen die Ausgaben verzerren. Je mehr qualitativ hochwertige Daten für das Training zur Verfügung stehen, desto besser die Ergebnisse.
- **Expertise der Teams, von denen die Modelle erstellt werden**. Um effektive KI-Modelle für die Cybersecurity zu erstellen, ist ein umfassendes Verständnis zwei separater, sich jedoch ergänzender Bereiche erforderlich:
  - **Bedrohungen:** Um zu bestimmen, was das KI-Modell leisten muss, müssen Sie zunächst verstehen, wie Malware und Bedrohungsakteure arbeiten.
  - **KI:** Wenn Sie ermittelt haben, was die KI leisten soll, geht es darum, das richtige Modell zum Erreichen Ihres Ziels zu erstellen.

  Zur Erstellung effektiver KI-Modelle, die eine deutliche Verbesserung der Cybersecurity ermöglichen, müssen diese beiden Bereiche engmaschig zusammenarbeiten und ihr gegenseitiges Fachwissen nutzen.
- **Qualität von Produktentwicklung und -Rollout**. Mitte 2024 führte ein fehlerhaftes Update in einem Cybersecurity-Produkt weltweit zu Betriebsausfällen. Schlecht getestete, nicht hinreichend qualitätsgeprüfte und mangelhaft bereitgestellte KI-Funktionen können potenziell noch größeren Schaden anrichten. Zudem lassen sich Probleme möglicherweise nicht einfach ermitteln und beheben.

### Vermeintliche (Cyber)Sicherheit

Unternehmen sind sich des Risikos schlecht entwickelter und bereitgestellter KI in Cybersecurity-Lösungen weitgehend bewusst. Die überwiegende Mehrheit (89 %) der befragten IT-/Cybersecurity-Experten macht sich Sorgen darüber, dass Schwächen der generativen KI-Funktionen von Cybersecurity-Tools ihrem Unternehmen schaden könnten. 43 % waren eigenen Angaben nach „sehr besorgt" und 46 % „eher besorgt".

Daher überrascht es kaum, dass 99 % (gerundet) der Unternehmen angeben, bei der Bewertung der GenAI-Fähigkeiten in Cybersicherheits-Lösungen die Qualität der Cybersicherheitsprozesse und -kontrollen zu prüfen und zu beurteilen, die bei der Entwicklung der GenAI eingesetzt werden:

- 73 % geben an, dass sie die Qualität der Cybersicherheitsprozesse und -kontrollen umfassend prüfen und beurteilen
- 27 % geben an, dass sie die Qualität der Cybersicherheitsprozesse und -kontrollen teilweise prüfen und beurteilen

Auf den ersten Blick erscheint es durchaus positiv, dass so viele Unternehmen eigenen Angaben nach eine umfassende Prüfung durchführen. Tatsächlich lassen die Zahlen jedoch auf blinde Flecken bei Unternehmen schließen.

Die Evaluierung der Prozesse und Kontrollen, die zur Entwicklung von GenAI-Funktionen eingesetzt werden, erfordert Transparenz seitens des Anbieters und hinreichende KI-Kompetenz seitens des Bewertungsteams. Leider ist beides knapp bemessen. Nur selten machen Lösungsanbieter ihre vollständigen GenAI-Entwicklungs- und Rolloutprozesse leicht zugänglich. Gleichzeitig verfügen IT-Teams häufig nur über begrenzte Kenntnisse der Best Practices im Bereich der KI-Entwicklung. Für viele Unternehmen bedeutet dies, dass sie „nicht wissen, was sie nicht wissen".

## 3. Betriebliches Risiko: Zu große Abhängigkeit von KI

Im Alltag stoßen wir mittlerweile in fast allen Bereichen auf KI – von der Suche nach dem besten Weg zum Supermarkt bis hin zu Fernsehtipps. Daher verlassen sich viele nicht selten zu sehr auf KI, weil sie davon ausgehen, dass KI bestimmte Aufgaben besser erledigen kann als Menschen. Glücklicherweise sind sich die meisten Unternehmen bewusst und auch besorgt darüber, welche Folgen eine zu große Abhängigkeit von KI für die Cybersicherheit haben kann:

- 84 % sind besorgt über den daraus resultierenden Druck, die Anzahl der Cybersicherheits-Fachkräfte zu reduzieren
- 87 % sind besorgt über einen daraus resultierenden Mangel an Verantwortlichkeit im Bereich Cybersicherheit

Das Problembewusstsein ist der erste Schritt, den Risiken entgegenzuwirken. KI ist nur eine Komponente der Cyberabwehr eines Unternehmens. Zwar leistet sie durchaus gute Dienste, ist jedoch nicht immer der richtige Ansatz und selten das Allheilmittel. Jedes Unternehmen ist anders. Daher sollte die Nutzung von KI im Kontext der individuellen Anforderungen und Geschäftsstruktur stehen.

## 4. Finanzielles Risiko: Niedriger ROI aus Investitionen in KI

Die Entwicklung und Wartung hochkarätiger GenAI-Funktionen in Cybersecurity-Lösungen sind teuer. IT- und Cybersecurity-Entscheider sind sich der Konsequenzen dieser Ausgaben bewusst. 80 % der Befragten gehen davon aus, dass GenAI die Kosten für ihre Cybersecurity-Lösungen deutlich erhöhen wird.

Trotz dieser erwarteten höheren Kosten rechnen die meisten Unternehmen damit, dass sie mit GenAI ihre Gesamtausgaben für Cybersicherheit senken können. 87 % der Befragten äußerten sich zuversichtlich, dass die Kosten für GenAI in Cybersecurity-Tools durch die damit verbundenen Einsparungen vollständig ausgeglichen werden.

Gleichzeitig sind sich die Unternehmen bewusst, dass die Quantifizierung dieser Kosten schwierig ist. Die Kosten für GenAI sind in der Regel in den Gesamtpreis von Cybersecurity-Produkten und -Services integriert. Daher lässt sich nicht leicht ermitteln, wie viel Unternehmen für GenAI für Cybersecurity ausgeben. Angesichts dieses Mangels an Transparenz stimmen 75 % der Befragten zu, dass diese Kosten schwer messbar sind (39 % stimmen voll und ganz zu, 36 % stimmen eher zu).

Ohne effektives Reporting riskieren Unternehmen, dass sich ihre Investitionen in KI für die Cybersicherheit nicht erwartungsgemäß auszahlen. Es kann sogar sein, dass sie Investitionen in KI tätigen, die an anderer Stelle sinnvoller eingesetzt werden könnten.

## 5. Manipulations-Risiko: Kompromittierung öffentlich zugänglicher KI-Modelle

Risiken für die Cybersicherheit durch KI können nicht nur in Cybersecurity-Tools und -Anwendungen entstehen. Der Boom bei der Nutzung öffentlich zugänglicher Large Language Models (LLMs) hat ein neues Einfallstor geschaffen: Finanzstarke, technisch versierte Akteure können nun die KI-Modelle kompromittieren, um sie für ihre Zwecke zu missbrauchen. Hier ein paar Beispiele für mögliche Kompromittierungen:

- **Data Poisoning**. Carlini et. al. haben in ihrem Paper Poisoning Web-Scale Training Datasets is Practical aus dem Jahr 2023 gezeigt, dass Data Poisoning (d. h. die Manipulation der Daten, mit denen das Modell trainiert wird, um die Ergebnisse zu beeinflussen) ein realistisches Bedrohungsrisiko darstellt.
- **Hintertüren für staatliche Akteure**. Viele Staaten verfügen über die Ressourcen, um leistungsstarke LLMs zu erstellen. Indem sie geheime Hintertüren einbauen und die Modelle anschließend öffentlich zugänglich machen, können staatliche Akteure das LLM zu ihrem Vorteil manipulieren.
- **LLM Spoofing**. Böswillige Akteure können seriöse LLMs kompromittieren (z. B. durch Hinzufügen von Hintertüren) und die Änderungen dann als „Verbesserungen" anpreisen. Um Opfer dazu zu bringen, ihr kompromittiertes Tool zu verwenden, fälschen sie den Namen des seriösen Anbieters, indem sie beispielsweise einen Buchstaben weglassen oder den Buchstaben O durch die Zahl 0 ersetzen.

Umfassende Einblicke in LLM-Kompromittierungen erhalten Sie in den aktuellen Beiträgen der KI-Experten von Sophos.

# Praxistipps zum Umgang mit dem KI-Hype

KI birgt definitiv Risiken. Mit einem gut durchdachten Ansatz können Unternehmen diese Risiken jedoch bewältigen und die Vorteile von KI sicher und effektiv nutzen, um ihre Cyberabwehr zu optimieren. Viele dieser Empfehlungen lassen sich auch auf die Implementierung von KI in anderen Bereichen anwenden.

## Bedrohungsrisiko: Rüsten Sie Ihre Cyberabwehr für das KI-Zeitalter auf

Ein Hauptaugenmerk sollte auf mehr Resilienz gegenüber KI-gestützten Bedrohungen liegen. Da Angreifer KI in erster Linie nutzen, um die Qualität und Glaubwürdigkeit von Phishing-E-Mails und Scams zu erhöhen, ist es sinnvoll, sich insbesondere auf diese Bereiche zu konzentrieren. Unsere Empfehlungen:

- **Erhöhen Sie Ihren E-Mail-Schutz**. Suchen Sie nach Lösungen, die KI-generierte Phishing-E-Mails und Betrugsversuche erkennen und so verhindern, dass sie in die Posteingänge Ihrer Mitarbeiter gelangen.
- **Installieren Sie Business Email Compromise (BEC)- und VIP-Schutz**. Wählen Sie E-Mail-Security-Lösungen mit BEC- und VIP-Schutz, die Inhalte beispielsweise auf Tonalität und Stil prüfen, um Betrug zu erkennen.
- **Seien Sie bei sozialen Medien besonders wachsam** – gerade beim entspannten Scrollen durch die sozialen Kanäle können Nutzer eher auf Betrugsversuche hereinfallen
- **Führen Sie Prozesse ein, die das Risiko von Voice Cloning mindern**, z. B. für den Fall, dass Mitarbeiter unerwartet zu Zahlungen oder Datenfreigaben aufgefordert werden. Hier einige Vorschläge:
  - Rückruf, um die Anfrage zu verifizieren
  - Einführung von Passcodes oder Passphrasen

## Abwehrrisiko: Prüfen Sie die Qualität der KI in Cybersecurity-Produkten

Bedenken Sie die Risiken und Auswirkungen von qualitativ minderwertiger KI bei Ihren Sicherheitsinvestitionen. Stellen Sie Anbietern Fragen zu:

- **Trainingsdaten**. Wie gut und umfangreich sind die Daten, mit denen die Modelle trainiert werden, und aus welcher Quelle stammen sie? Bessere Eingabedaten führen zu besseren Ergebnissen.
- **Entwicklerteam**. Informieren Sie sich über die Experten hinter den Modellen. Über welches Maß an KI-Expertise verfügen sie? Wie gut kennen sie Bedrohungen, Angreiferverhalten und Sicherheitsabläufe?
- **Produktentwicklungs- und Rollout-Prozess**. Welche Schritte durchläuft der Anbieter bei der Entwicklung und Bereitstellung von KI-Funktionen in seinen Lösungen? Welche Kontrollmechanismen sind vorhanden?

Fragen Sie sich: Wie sehr vertraue ich der KI-Kompetenz des Unternehmens? Werden die erforderlichen strengen Qualitäts- und Bereitstellungskontrollen entsprechend durchgeführt?

## Betriebliches Risiko: Betrachten Sie KI als Unterstützung zu menschlicher Expertise

Der KI ist es egal, wenn ein Sicherheitsvorfall passiert – Ihren Mitarbeitern nicht. Wenn es zu einem Vorfall kommt, benötigen Sie Experten, die die Situation im Kontext Ihres Unternehmens einordnen und entsprechende Reaktionsmaßnahmen ergreifen können.

- **KI allein ist nicht die Lösung**. KI ist nur eine von vielen Komponenten im Arsenal der Cyberabwehr. Nutzen Sie KI, bedenken Sie dabei jedoch, dass die Verantwortung für die Cybersecurity letztlich beim Menschen liegt.
- **KI als Unterstützung – nicht als Ersatz**. Der anhaltende weltweite Fachkräftemangel in der Cybersecurity ist allgemein bekannt. Erschwerend kommt hinzu, dass viele Mitarbeiter von Burnout betroffen sind. Konzentrieren Sie sich zunächst darauf, wie KI Ihre Mitarbeiter unterstützen kann, anstatt KI zu nutzen, um Personal abzubauen. KI übernimmt viele einfache, sich wiederholende Sicherheitsaufgaben und liefert umsetzbare Erkenntnisse. So kann sie:

- Kapazitäten für strategische, geschäftsrelevante Projekte freisetzen
- die Überlastung durch Warnmeldungen und die Flut irrelevanter Daten reduzieren
- die berufliche Weiterentwicklung von qualifizierten Analysten beschleunigen
- weniger erfahrene Analysten im Bereich Security Operations qualifizieren und eine Ressourcen-Pipeline aufbauen

## Finanzielles Risiko: Gehen Sie bei Investitionen in KI strategisch und strukturiert vor

Das finanzielle Risiko lässt sich von Unternehmen am einfachsten minimieren, da sie viele Faktoren selbst steuern können.

- **Setzen Sie eindeutige Ziele**. Legen Sie klar, spezifisch und detailliert fest, was Sie mit KI erreichen möchten.
  - Ermitteln Sie Ihren Bedarf. Gibt es Lücken? In welchen Bereichen kann KI Abhilfe schaffen?
  - Berücksichtigen Sie Kosten- und Zeitersparnis sowie Vorteile beim Schutz.
- **Quantifizieren Sie den Nutzen**. Ermitteln Sie, wie viel Ihre KI-Investitionen bewirken können.
  - Wenn Sie die Gesamtbetriebskosten für Ihre Cybersecurity senken möchten, berechnen Sie, wie viel Sie mit KI einsparen können.
  - Wenn Sie sich von KI weniger Fluktuation im Bereich IT/Cybersecurity erhoffen, informieren Sie sich darüber, wie genau sich das KI-Tool auf Ihre Mitarbeiter auswirkt. Welche Aufgaben übernimmt KI? Wie viele Stunden sparen Sie ein?
- **Priorisieren Sie Investitionen**. KI kann in vielerlei Hinsicht helfen. Um zu erkennen, welche Bereiche besonders davon profitieren, ermitteln Sie die für Ihr Unternehmen relevanten Kennzahlen. Dazu gehören beispielsweise finanzielle Einsparungen, Auswirkungen auf die Personalfluktuation, Reduzierung der Angriffsfläche usw. Bewerten Sie im Anschluss die unterschiedlichen Optionen anhand dieser Metriken.
- **Bemessen Sie die tatsächliche Leistung**. Investitionsentscheidungen werden mit guten Absichten getroffen. Prüfen Sie, ob Ihre KI auch tatsächlich das leistet, was Sie sich davon versprochen haben. Erzielen Sie die erwünschten Vorteile? Gibt es Vorteile, die Sie nicht erwartet haben? Liegen in manchen Bereichen die Ergebnisse unter den Erwartungen? Nehmen Sie anhand dieser Erkenntnisse gegebenenfalls erforderliche Anpassungen vor.

Fragen Sie sich, ob KI der beste Weg ist, um Ihr Ziel zu erreichen. Könnten Sie mit einer anderen Technologie oder einem anderen Ansatz mehr erreichen?

## Manipulations-Risiko: Seien Sie sich der Gefahren bewusst

Dieses Risiko lässt sich für Unternehmen am schwierigsten eindämmen. Wenn sich Unternehmen jedoch des Risikos bewusst sind, können sie potenzielle Folgen eher minimieren. Achten Sie vor diesem Hintergrund bei der Auswahl von öffentlichen LLMs auf Folgendes:

- **Modelle von bekannten, seriösen Anbietern**. Zwar sind auch diese Modelle nicht immun gegen Data Poisoning, jedoch werden Probleme hinsichtlich der Datenausgabe mit größerer Wahrscheinlichkeit veröffentlicht.
- **Korrekte Anbieternamen**. Angreifer imitieren die Namen seriöser Anbieter, um ihren Opfern vorzutäuschen, dass ihre kompromittierten Modelle legitim sind.

Experten für KI in der Cybersecurity arbeiten aktiv an Strategien zur Beseitigung dieses Risikos.

## Fazit

KI bringt der Cybersecurity enorme Vorteile. Wenn Unternehmen sich nicht vom KI-Hype blenden lassen und einen gut durchdachten, ergebnisorientierten Ansatz verfolgen, können sie die Vorteile dieser Technologie nutzen, um ihre Cyberabwehr zu optimieren und ihre IT- und Cybersecurity-Experten zu entlasten.

## Über die Studie

Quelle: KI-gestützte Cybersicherheit: Was Unternehmen wirklich erwarten (können)

Sophos beauftragte das Marktforschungsunternehmen Vanson Bourne mit der Durchführung einer unabhängigen Befragung von 400 IT- und Cybersecurity-Entscheidern in Unternehmen mit 50 bis 3.000 Mitarbeitern. Die Umfrage, die 13 Branchen umfasste, wurde im November 2024 durchgeführt. Die Umfrageteilnehmer nutzen Endpoint-Security-Lösungen von 19 verschiedenen Anbietern. So wurde eine breite Branchenvertretung gewährleistet.

## Über Sophos

Sophos ist ein weltweit führender Anbieter von vielfach ausgezeichneten Cybersecurity-Produkten und -Services, die von Firewalls, Endpoint Protection und EDR/XDR-Tools bis hin zu Managed Detection and Response (MDR) und Incident Response (IR) Services reichen.

Sophos optimiert seine Cybersecurity-Lösungen seit 2017 mit KI und kombiniert KI-Technologien und menschliche Cybersecurity-Expertise, um Bedrohungen zu stoppen – egal, wo diese auftreten. Deep Learning und generative KI-Funktionen, die zentrale Probleme für unsere Kunden lösen, sind in unsere Produkte und Services integriert und werden über die branchenweit größte KI-native Sicherheitsplattform bereitgestellt. Unsere adaptive KI-Plattform ist auf Daten von Angriffen aus mehr als 600.000 unterschiedlichen Kundenumgebungen trainiert. So bieten wir Kunden branchenweit führenden Schutz vor komplexen Bedrohungen und stärken deren Cyberabwehr.

Sie möchten mehr über das Thema und Sophos-Lösungen erfahren? Besuchen Sie www.sophos.de

2025-01-15 (WP-MP)

SOPHOS