



# AI-native Cyber Attacks and Supply Chain Risks: The New Frontier in AI Security Threats

Posted on September 26, 2025

AKTE-AI-250926-559: 2025 hat das Wettrüsten im Cyberraum eine neue Grenze erreicht – KI ist nicht mehr nur Abwehr, sondern das Schlachtfeld selbst. Sind Unternehmen bereit für die dunkle Seite der KI?

## Die Zeiten ändern sich: Künstliche Intelligenz als Cyberwaffe

Wer heute an IT-Sicherheit denkt, denkt an Firewalls, Intrusion Detection und vielleicht noch an klassische Virens Scanner. Doch diese Instrumente werden immer schneller von einer Flut neuartiger KI-basierter Angriffe überrollt. Eine radikale Verschiebung hat stattgefunden: KI ist nicht mehr nur Werkzeug der Verteidiger, sondern das bevorzugte Arsenal skrupelloser Angreifer weltweit.



## Die Fakten: Zwischen Generative AI und Agentic AI

- 2025 waren 87% der befragten Organisationen laut aktueller Studien mindestens einmal Ziel von KI-basierten Cyberattacken. Die [Branchenberichte](#) sind eindeutig.
- 70% aller bekannt gewordenen KI-Incident-Fälle involvierten generative KI-Modelle - etwa beim Erstellen von Phishing-Nachrichten oder bei Datenextraktionen durch LLM Prompt Injection.
- Den größten Schaden verursachte allerdings "agentic AI": autonome, selbst handelnde KI-Agenten, die sich in Systeme einhacken, Krypto-Guthaben rauben und APIs kompromittieren.

## Vulnerable by Design: Der Supply-Chain-Albtraum

Die zunehmende Verflechtung von KI-Systemen in unternehmenskritischen Supply Chains bringt schwer kontrollierbare Risiken. Laut einem [TrendMicro-Report](#) aus September 2025 haben kompromittierte KI-Trainingsdaten sowie unsichere KI-getriebene Apps in den vergangenen Monaten zu einer Welle von Angriffen geführt, die vom Diebstahl sensibler Nutzerdaten bis zur Manipulation von Produktionsprozessen reichen.

"Mit jeder neuen Verknüpfung in globalen Datenströmen und Lieferketten entstehen Angriffsvektoren, die niemand mehr alleine überblicken kann."

## Beispiel: Der Attacke durch die Hintertür

- **Manipulierte KI-Modelle:** Über Supply Chain-Angriffe landen kompromittierte Checkpoints oder Libraries in betriebskritischen Systemen. In Folge werden Daten abgegriffen oder Produktionsprozesse sabotiert, teils ohne sofortige Entdeckung.
- **Infiltrierte Trainingsdaten:** Der gezielte Einbau toxischer Trainingsdaten ermöglicht Angreifern, Schwachstellen direkt in die Entscheidungslogik der KI zu schreiben, die sich später ausnutzen lassen (sogenannte Data Poisoning-Attacken).
- **Insecure-by-Design KI-Apps:** Fehlerhaft gesicherte AI-powered Apps exponieren persönliche Nutzerdaten durch fehlende Authentifizierung oder ungeschützte APIs - höchst aktuell und global relevant.



## Neue Angriffsvektoren: Hardware, Malware und Pen-Testing als Waffe

2025 erscheinen KI-Angriffe deutlich facettenreicher als Angriffe auf klassische Software. Die Kombination aus Hardware, Software und Netzwerk eröffnet bisher undenkbbare Möglichkeiten.

### RowHammer und die Hardware-Dimension

Einen beunruhigenden Trend markiert die Nutzung von Hardware-Exploits wie RowHammer gegen KI-Infrastruktur, insbesondere bei neuen Speichertechnologien wie DDR5. Dieser Bug ermöglicht es, Bits im Speicher gezielt zu flippen – und dabei etwa Berechtigungen in Rechenzentren zu eskalieren, in denen KI-Modelle verarbeitet werden. Selbst Top-Konzerne sind betroffen, wie The Hacker News [berichtet](#).

### Phishing, Malware, und der KI-Multiplikator

- **KI-gestützte Phishing-Attacken:** Täuschend echte, massenhaft individualisierte Phishing-Mails, generiert und fortlaufend optimiert durch Large Language Models, sind kaum noch erkennbar.
- **Malware der neuen Generation:** Schadsoftware mit eingebauter KI kann in Echtzeit ihr Verhalten an Sicherheitsmaßnahmen anpassen, Signaturen verändern und unauffindbar werden.
- **Pen-Test-Tools im Visier:** Was ursprünglich zur Schwachstellenerkennung diente, wird inzwischen in modifizierten Versionen für gezielte Angriffe genutzt. Die Grenze zwischen Abwehr und Angriff verschwimmt.

### Auswirkungen auf Unternehmen und Behörden

KI-Angriffe schlagen längst in der Breite durch: Von der Manipulation industrieller Steuerungsanlagen bis zur Kompromittierung ganzer Cloud-Infrastrukturen. Besonders alarmierend: Es ist nicht mehr nur das große Unternehmen betroffen, sondern jedes Unternehmen, das KI in irgendeiner Stufe einsetzt – weltweit.

## Statistik: Die Größenordnung der Bedrohung

Faktor

Zahl/Prozentsatz Jahr



Organisationen, die KI-basierte Angriffe erlebten	87%	2025
Organisationen, die weitere Zunahme erwarten	91%	2025
Vorfallanteil mit generativer KI	70%	2025
Dramatischste Schäden durch agentic AI	-	2025
NN Hardware-Exploits (z.B. RowHammer)	Neue Fälle weltweit	2025

(Quellen: [Rod Trent Substack \(2025\)](#), [Adversa AI \(2025\)](#))

## Jenseits klassischer Cybersecurity: Neue Verteidigungsstrategien sind Pflicht

**Die Bausteine klassischer IT-Sicherheitsarchitekturen reichen nicht mehr aus, wo KI selbst das Ziel und Werkzeug der Angreifer ist.**

Die *Verlagerung hin zu agentic AI* und die globale Vernetzung verlangen eine fundamentale Neuausrichtung der Cyberverteidigung. Folgende Prinzipien treten nun in den Mittelpunkt:

- **Behavioral Analytics für KI-Modelle:** Um Angriffe frühzeitig zu identifizieren, müssen Unternehmen adaptive Überwachungsmethoden für „ungewöhnliches“ KI-Verhalten etablieren.
- **Adversarial Robustness:** Widerstandsfähige KI-Systeme benötigen härtere Tests gegen gezielte Manipulation (Adversarial Inputs, Poisoning).
- **Supply Chain Verification:** Die gesamte Lieferkette - von Trainingsdaten über Software-Libraries bis hin zur Hardware - muss transparent, überprüf- und notariell absicherbar sein.
- **Security-by-Design für KI-Apps:** Entwickler stehen in der Pflicht, Sicherheitsmechanismen von Anfang an zu denken und regelmäßig zu kontrollieren - gerade bei Third-Party-Integrationen.
- **Hardware-basierte Schutzmaßnahmen:** Memory Protection, Physische Sicherung und deshalb rigoroses Patch-Management spielen eine neue und zentrale Rolle.

## Globale Perspektive: Gemeinsam gegen die KI-



## Bedrohung

Die Angriffswelle macht nicht an Ländergrenzen halt. Internationale Unternehmen, seien es Tech-Giganten, Banken oder Hersteller, sind Ziele - und Vorreiter zugleich. Die neue Bedrohungslandschaft zwingt zur internationalen Kooperation:

- **Austausch von Angriffsdaten:** Schnelle, globale Transparenz zu neuen KI-Exploits wird essentiell.
- **Standards und Zertifizierungen:** Es braucht neue Benchmarks für KI-Sicherheitsprüfungen, im Code wie im Training und in der Supply Chain.
- **Multinationale Incident Response:** Frühwarnsysteme, abgestimmte Abwehrpläne und kooperative Ermittlungen gewinnen weiter an Bedeutung.

## Blick in die Zukunft: Angriff und Verteidigung im Gleichschritt

Die Zukunft der Cybersicherheit ist ein Wettrennen: Angreifer setzen auf autonome, lernfähige Algorithmen. Verteidiger müssen ebenso KI nativ denken - und dort ansetzen, wo menschliches Mitschreiten nicht mehr ausreicht. Wie lange bleibt die nächste Revolution ein Vorteil für Angreifer, bevor die Verteidigung aufholt?

*Die neue KI-Bedrohung ist global, adaptiv und in ihrer Geschwindigkeit beispiellos - das Zeitfenster für strategische Gegenmaßnahmen schließt sich schnell.*

**KI verschiebt die Regeln der Cyberabwehr grundlegend - nur wer KI-Verstehen, Supply Chain-Sicherheit und internationale Kooperation radikal priorisiert, kann dem neuen Risikouniversum auf Augenhöhe begegnen.**