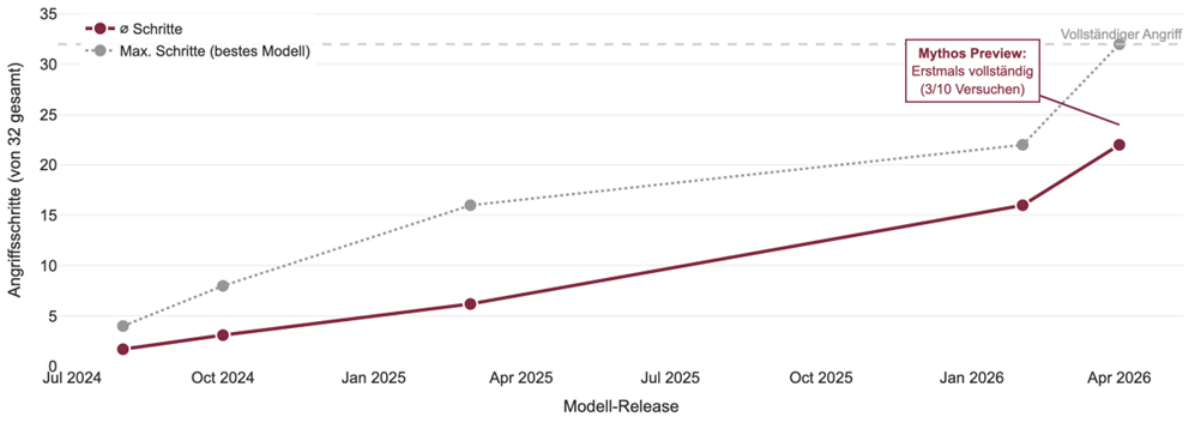
cepExecutive
Was Unternehmen bewegt

Schatten-KI als Geschäftsrisiko

Zuletzt wurde die Welt von einer neuen Generation von KI-Modellen aufgeschreckt: **Mythos von Anthropic und GPT-5.5 Cyber von OpenAI** ermöglichen erstmals autonom durchgeführte Angriffe auf Netzwerke. Die Folgen für kritische Infrastrukturen, aber auch für Unternehmen können verheerend sein. In weniger als zwei Jahren stieg die Angriffstiefe von KI-Modellen in simulierten IT-Netzwerken durchschnittlich von unter 2 auf über 20 eigenständige Schritte. Ein Angriff kostet heute rund 75 Euro. Gleichzeitig nutzen viele Mitarbeiter KI-Tools und KI-Agenten ohne genehmigten Zugang („Schatten-KI"). **Es droht ein massives Cyber-Risiko im Mittelstand.**

KI-Modelle werden exponentiell gefährlicher: autonome Netzwerkangriffe (2024–2026)
Schritte in 32-stufigem Corporate-Network-Angriff (UK AISI, Paper Mar 2026 + Blog Apr 2026).



Executive Insights

Schatten-KI ist mehr als nur ein IT-Problem

- Das Ausmaß an Schatten-KI wird oft unterschätzt
- Jedes nicht genehmigte KI-Tool, das Mitarbeiter im Unternehmen einsetzen, fehlt im Asset-Inventar, das NIS2 für wesentliche und wichtige Einrichtungen vorschreibt, und damit auch im Risikoregister
- Entsteht durch dieses Tool ein Sicherheitsvorfall, fehlt die Grundlage für Erkennung und Meldung

KI-gestützte Angriffe überfordern klassische Abwehr

- KI-gestützte Angriffe agieren autonom, mehrstufig und schneller als klassische Abwehrzyklen
- Regelbasiertes Monitoring erkennt solche Vorfälle oft erst nach Überschreiten der NIS2-Meldefrist
- KI-gestütztes Detection & Response sowie KI-simuliertes Red-Teaming sind notwendige Antworten, technisch wie regulatorisch

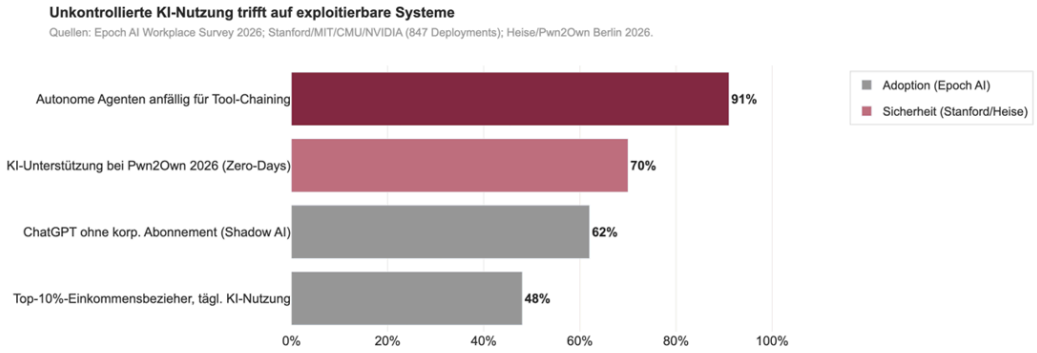
Reaktions-geschwindigkeit strukturell erhöhen

- Das Zeitfenster zwischen Bekanntgabe einer Sicherheitslücke und deren Ausnutzung sinkt durch Cyber-KI-Modelle auf Stunden
- Interne Patch-Management-Prozesse und deren Dokumentation müssen geprüft werden
- Ergänzend: Schwachstellenmeldeverfahren (z.B. Bug-Bounty) als proaktives Frühwarnsystem

Markt & Wettbewerb

Der Business Case

KI-gestützte Angriffe treffen kleine und mittlere Unternehmen häufig dort, wo Zugriffskontrollen lückenhaft und Systemereignisse nicht protokolliert werden. Gerade Bereiche mit unkontrollierter KI-Nutzung durch Mitarbeiter sind prädestiniert. Solche **Schatten-KI umfasst heute vor allem private ChatGPT-Accounts, bald auch persönliche Agenten mit weitreichenden Befugnissen**. Agentische KI bezeichnet KI-Systeme, die ohne menschliche Zwischenschritte Ziele verfolgen und dafür Werkzeuge aufrufen. Solche Systeme können eine Phishing-Mail verfassen und Daten exfiltrieren. Der EU-AI Act erfasst agentische KI aktuell nur unzureichend. **Die neuesten KI-Modelle senken die Kostenschwelle für autonome Netzwerkangriffe signifikant.**



Auswirkungen auf Marktstruktur & Wettbewerbsdynamik

- Cyber-Risiken erhöhen potenziell die **Markteintrittskosten** neuer Unternehmen. Wer heute ein Produkt auf den Markt bringen will, muss neben Entwicklung und Vertrieb auch von Anfang an eine robuste Sicherheitsarchitektur mitfinanzieren.
- Die Kumulrisiken (Schadenskaskaden) von autonomen Cyber-Attacken sind nur bedingt über den Markt versicherbar und nur über **hohe Risikoprämien**.
- Cyber-Risiken erhöhen die Netzwerkexternalitäten entlang von Lieferketten und damit **potenzielle Ausfallrisiken**. Cyber-Sicherheit wird zu einem Auswahlkriterium für Zulieferer. Asymmetrische Informationen können zu Marktversagen führen.
- Cyber-Risiken verändern die **Innovationsanreize** bei Produkten und Prozessen in Richtung Sicherheit. Unternehmen investieren kurzfristig mehr in Absicherung
- Eine umfassende **Cyber-Resilienz wirkt positiv als Signal am Markt und schafft neue Märkte**.

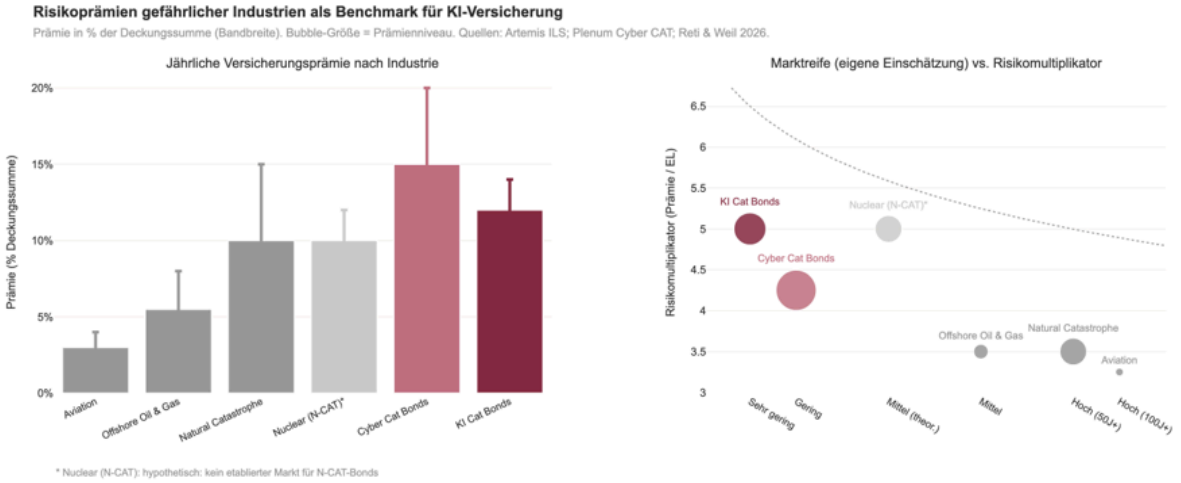
Dimension	Bewertung	Wirkung
Innovationen	🟡 ambivalent	Verzögerte Produktzyklen, Fokus auf Sicherheit
Kosten und Marktstruktur	🔴 kritisch	Fixkosten steigen, Marktkonzentration höher
Markteintrittsbarrieren	🟡 ambivalent	Schutz vor Low-Quality-Wettbewerb, aber auch Eintrittshemmnisse
Vertrauen / Marke	🟢 positiv	Differenzierung durch „Trust“, positives Signalling
Globale Lieferketten	🟡 ambivalent	Auswahl nach Sicherheit, asymmetrische Information

Szenarien & Strategie

Szenarien

Mythos markiert einen Strukturbruch in der Cybersicherheit und stellt ein Risiko dar, das durch Schatten-KI verstärkt wird. Angesichts eines dokumentierten **Rückstands chinesischer offener Modelle von durchschnittlich 3 bis 7 Monaten** ist davon auszugehen, dass eine vergleichbare Cyberangriffsfähigkeit innerhalb des nächsten Jahres einem breiteren Kreis staatlicher und nichtstaatlicher Akteure zugänglich sein wird. **Je nachdem, wie sich das Gleichgewicht zwischen Angreifern oder Verteidigern durch diese Diffusion verschiebt, ergeben sich drei Szenarien.**

- **Szenario I: KI-Sicherheit wird zum Standard im Wettbewerb.** Unternehmen, die heute Regeln für Schatten-KI einführen und KI-gestütztes Red Teaming aufbauen, können das gegenüber Kunden, Versicherern und Aufsichtsbehörden dokumentieren. Das senkt Versicherungsprämien und verkürzt spätere Zertifizierungszeiten. Es könnte auch zu einem Auswahlkriterium in Lieferketten werden.
- **Szenario II: KI-Sicherheitslücken erzeugen Marktineffizienz.** Unternehmen ohne Monitoring für unkontrollierte KI-Nutzung bieten genau die Angriffsfläche, auf die Modelle wie Mythos optimiert sind: offene Zugänge, unsichtbare Tools, lückenhafte Logs. Solche Unternehmen werden daher bis Ende 2027 mit hoher Wahrscheinlichkeit Angriffsziel autonomer KI-Operationen werden.
- **Szenario III: Asymmetrische KI-Lernkurven („Arms Race").** KI-Angriffsysteme überholen die durchschnittliche Unternehmensverteidigung. Aktuelle Schwächen der KI, wie Kontextverlust über Angriffsketten oder fehlerhafte Mehrschrittkoordination, nehmen ab. Nur Unternehmen mit KI-gestütztem Monitoring können standhalten.



Strategische Implikationen

Unternehmen tragen das Risiko immer günstigerer KI-Angriffe. Der Markt hat noch kein funktionierendes Instrument zur Risikoübertragung entwickelt. Klassische Cyberversicherungen greifen nicht, da es zum einen an historischen Schadensstatistiken für KI-spezifische Ereignisse fehlt und zum anderen das Kumulrisiko bei einem koordinierten Agenten-Angriff die Kapazität einzelner Träger übersteigt. Kumulrisiko meint das Risiko, dass ein einzelnes Ereignis (z.B. Cyberangriff) bei einer Vielzahl von Unternehmen oder Anlageklassen gleichzeitig Schäden auslöst.

Executive Outlook

Worauf Sie jetzt achten sollten

In einer KI-basierten, mittelständischen und arbeitsteiligen Industrie wird Cyber-Sicherheit zu einer maßgeblichen Infrastruktur der wettbewerblichen Ordnung. Die exponentiellen Risiken werden durch die aktuelle Gesetzgebung nicht ausreichend adressiert. Die Politik wird zeitnah nachsteuern müssen.

- **EU-AI Act zügig anpassen, denn KI-Agenten sind eine Lücke:** Der EU-AI Act gilt für Agenten dem Grundsatz nach, greift in der Praxis aber nicht. Nur Modellanbieter müssen Missbrauchsrisiken adressieren; Agenten-Anbieter unterliegen lediglich allgemeinen Cybersicherheitspflichten. Die Rechtsabteilung sollte jetzt prüfen, welche eigenen KI-Anwendungen als Hochrisiko-Systeme klassifizierbar sind und ob Prompt-Injection-Risiken im Risikoregister erfasst sind.
- **ENISA, aber auch Unternehmen brauchen Modellzugang:** ENISA, die Agentur der EU für Cybersicherheit, soll zeitnah Zugang zu Mythos und vergleichbaren Modellen erhalten, um die neuen Risiken zu analysieren. Unternehmen, die KI-gestützte Angriffswerkzeuge noch nicht intern getestet haben, sollten das tun, bevor der erste Vorfall es erzwingt.
- **Architekturelle Resilienz ist Board-Thema, kein Technologieprojekt im Silo:** Die deutschen Pläne für ein nationales AI Security & Safety Institute gehen in die richtige Richtung. Mittlerweile sind nicht nur große Konzerne, sondern auch Krankenhäuser, Schulen und kleinere Zulieferer von KI-gestützten Angriffen betroffen: Die gesamte Lieferkette ist Angriffsfläche. Bug-Bounty-Programme, die KI-assistierte Entdeckungen zulassen, lassen sich jetzt noch günstig einführen. Der Zeitvorsprung, bevor Mythos-ähnliche Modelle frei zugänglich werden, muss genutzt werden.

“KI hält immer mehr unkontrolliert Einzug in Unternehmen. Autonome Cyber-Angriffe werden dadurch zu einem massiven Geschäftsrisiko für Unternehmen.”

Dr. Anselm Küsters
Fachbereichsleiter
Digitalisierung und KI
Tel. +49 030 43973746 15
kuesters@cep.eu | www.cep.eu

Weitere Informationen zum Thema finden Sie unter dem QR-Code.

Das **Centrum für Europäische Politik** FREIBURG | BERLIN, das **Centre de Politique Européenne** PARIS, und das **Centro Politiche Europee** ROMA bilden das **Centres for European Policy Network** FREIBURG | BERLIN | PARIS | ROMA.