17.03.2026 – 16:35 Uhr

Zahl der KI-Sicherheitslücken erreicht Rekordhoch

Zahl der KI-Sicherheitslücken erreicht Rekordhoch

Neuer Sicherheitsreport von TrendAI dokumentiert 2.130 neue Schwachstellen in KI-Systemen alleine im vergangenen Jahr

Wallisellen, 17. März 2026 – TrendAI, ein Geschäftsbereich von Trend Micro, einem der weltweit führenden Anbieter von Cybersicherheitslösungen, analysiert in seinem neuesten TrendAI State of AI Security Report, welche Anfälligkeiten das globale KI-Ökosystem gefährden. Der Bericht zeigt: KI-bezogene Schwachstellen

haben in den letzten Jahren deutlich zugenommen. Sowohl ihre Anzahl als auch ihre Kritikalität wachsen – und damit auch die Diskrepanz zwischen der schnellen Einführung von KI in Unternehmen und der tatsächlichen Sicherheitsreife.

Die Zahlen sprechen für sich: Von 2018 bis Ende 2025 wurden insgesamt 6.086 Software-Schwachstellen identifiziert, die direkt KI-Systeme betreffen – aus einer Gesamtdatenbank von über 330.000 CVEs (Common Vulnerabilities and Exposures). Allein im Jahr 2025 wurden 2.130 KI-bezogene Sicherheitslücken gemeldet. Das entspricht einem Anstieg von 34,6 Prozent gegenüber dem Vorjahr. Zum Vergleich: Die Gesamtzunahme von CVEs betrug lediglich 17,9 Prozent – ein deutliches Signal, dass Angreifer es zunehmend auf KI-Systeme abgesehen haben. Der Anteil KI-bezogener Schwachstellen an allen CVEs erreichte 2025 mit 4,42 Prozent den höchsten je gemessenen Wert.

Fast jede zweite Schwachstelle mit hohem oder kritischem Schweregrad

Nicht nur die Quantität, auch die Qualität der Lücken gibt Anlass zur Sorge: Von den 3.257 mit einem Schweregrad bewerteten KI-Schwachstellen fallen 48,9 Prozent in die Kategorien „hoch" oder „kritisch" (Common Vulnerability Scoring System, CVSS ab 7,0). Besonders exponiert sind dabei die aktuell besonders schnell wachsenden Bereiche des KI-Ökosystems: In der KI-Lieferkette weisen 46,5 Prozent der Schwachstellen hohe oder kritische Schweregrade auf, gefolgt von dem LLM-Ökosystem (Large Language Models, 45,1 Prozent), Agentischer KI (40,5 Prozent) und MCP-Servern (Model Context Protocol, 38,2 Prozent).

Gerade in diesen gefährdeten aufstrebenden Bereichen fehlt es häufig an ausgereiften Sicherheitspraktiken, was zu einem gefährlichen Missverhältnis zwischen Innovationsgeschwindigkeit und Absicherung führt. Im Gegensatz dazu weisen ausgereifte Kategorien wie Frameworks für Maschinelles Lernen (32,9 Prozent), KI-Datenpipelines (27 Prozent) und GPUs / KI-Hardware (18 Prozent) eine geringere Konzentration schwerwiegender Schwachstellen auf.

Risikofaktoren MCP-Server und agentische KI

Zwei Entwicklungen stechen besonders hervor: Das plötzliche Auftreten von Schwachstellen in MCP-Servern, die bisher gänzlich unbekannt waren, bietet Bedrohungsakteuren eine grosse, neuartige Angriffsfläche. Zudem ist eine explosionsartige Zunahme von Schwachstellen in agentischen KI-Systemen zu beobachten. MCP-Server, Schnittstellen, die es KI-Agenten ermöglichen, eigenständig auf externe Tools und Dienste zuzugreifen, verzeichneten 2025 direkt 95 neue CVEs (gegenüber nahezu null in den Vorjahren). Über 60 Prozent dieser Schwachstellen betreffen Injection-Angriffe, die es ermöglichen, schädliche Befehle direkt in KI-Systeme einzuschleusen. Agentische KI, also autonome Systeme, die eigenständig Entscheidungen treffen und Aktionen ausführen, verzeichnete den grössten prozentualen Anstieg aller Kategorien: von 74 CVEs im Jahr 2024 auf 263 im Jahr 2025, ein Zuwachs von 255 Prozent.

Die Lage wird sich weiter verschärfen

TrendAI Research prognostiziert für 2026 zwischen 2.800 und 3.600 neue KI-bezogene Schwachstellen, was einen weiteren Anstieg von bis zu 69 Prozent gegenüber 2025 darstellt. Besonders MCP-Server (plus 89–195 Prozent) und agentische KI (plus 33–109 Prozent) dürften das stärkste Wachstum verzeichnen.

„KI ist keine neue Angriffsfläche mehr, sondern hat sich mittlerweile etabliert", erklärt Richard Werner, Security Advisor bei TrendAI. „Unsere Untersuchung zeigt, dass die Schwachstellen in KI-Systemen schneller zunehmen als im gesamten Software-Ökosystem und dass die grössten Risiken in gemeinsam genutzten Komponenten wie Modell-Frameworks und Lieferketten liegen. Wenn Unternehmen KI produktiv einsetzen, müssen sie Cybersicherheit als Grundlage jedes Projekts berücksichtigen. KI erfordert dieselbe Transparenz und dasselbe Risikomanagement wie jedes andere kritische Geschäftssystem."

Weitere Informationen

Den vollständigen Bericht *Fault Lines in the AI Ecosystem: TrendAI State of AI Security Report* finden Sie in englischer Sprache hier: https://www.trendmicro.com/vinfo/de/security/news/threat-landscape/fault-lines-in-the-ai-ecosystem-trendai-state-of-ai-security-report

Einen deutschsprachigen Überblick über die wichtigsten Ergebnisse finden Sie unter: https://www.trendmicro.com/de_de/research/26/c/bruchstellen-im-ki-oekosystem-durch-schwachstellen.html

Der TrendAI-Report basiert auf einen umfangreichen CVE-Datensatz, mehrstufigen KI-Klassifizierungsworkflows und strengen Bewertungs- und Verifizierungsstandards. So stellt er einen klaren und zuverlässigen Überblick dar, wie sich KI-bezogene Schwachstellen im Laufe der Zeit entwickeln.

Pressestelle Trend Micro Schweiz
c/o BRAND AFFAIRS AG

Mischa Keller / MSc Business Administration
Partner

Telefon: +41 44 254 80 00
E-Mail: trendmicro-media@brandaffairs.ch
Mühlebachstrasse 8 / 8008 Zürich / Switzerland

Medieninhalte

Richard Werner, Security Advisor bei TrendAI.