



Sicherheit: KI als Angreifer und Verteidiger

13.10.26

Martin Seeger

Aspekte der KI-Security

Wer sind wir?

- NetUSE ist ein Managed Service Provider aus Kiel
- Gegründet 1992
- Schwerpunkt IT-Sicherheit
- KI-Aktivitäten:
 - Eigenes Produkt für KI-gestützte Dokumentenverwaltung und Audits
 - KI-gestützte interne Prozesse, z.B. Zeiterfassung
 - KI-Arbeitsgruppe und -Workshops seit 2 Jahren
 - Aufsetzen lokaler KI in vielen Spielarten
 - Prüfung von KI-Sicherheit, KI-Sicherheitslösungen

Aspekte der KI-Security

Wird uns die KI alle töten?

Künstliche Intelligenz

Und dann war es das mit der Menschheit

30. September 2026, 11:49 Uhr | Lesezeit: 7 Min. |  179 Kommentare



How AI could put humanity at risk: Nuclear war, bioweapons and a runaway internet



A worker passes by robots from the Chinese company Lens at the China International Supply Chain Expo in Beijing. (Ng Han Guan / Associated Press)

Subscribers are Reading >

Newsom signs fire insurance reforms but vetoes two bills addressing complaints by wildfire victims

Voices

Contributor: Arts hubs like the Kennedy Center sustain the nation

Why Neon put a Korean-language thriller in 1,500 U.S. theaters

How the top 25 high school football teams in the Southland fared this week

Voices

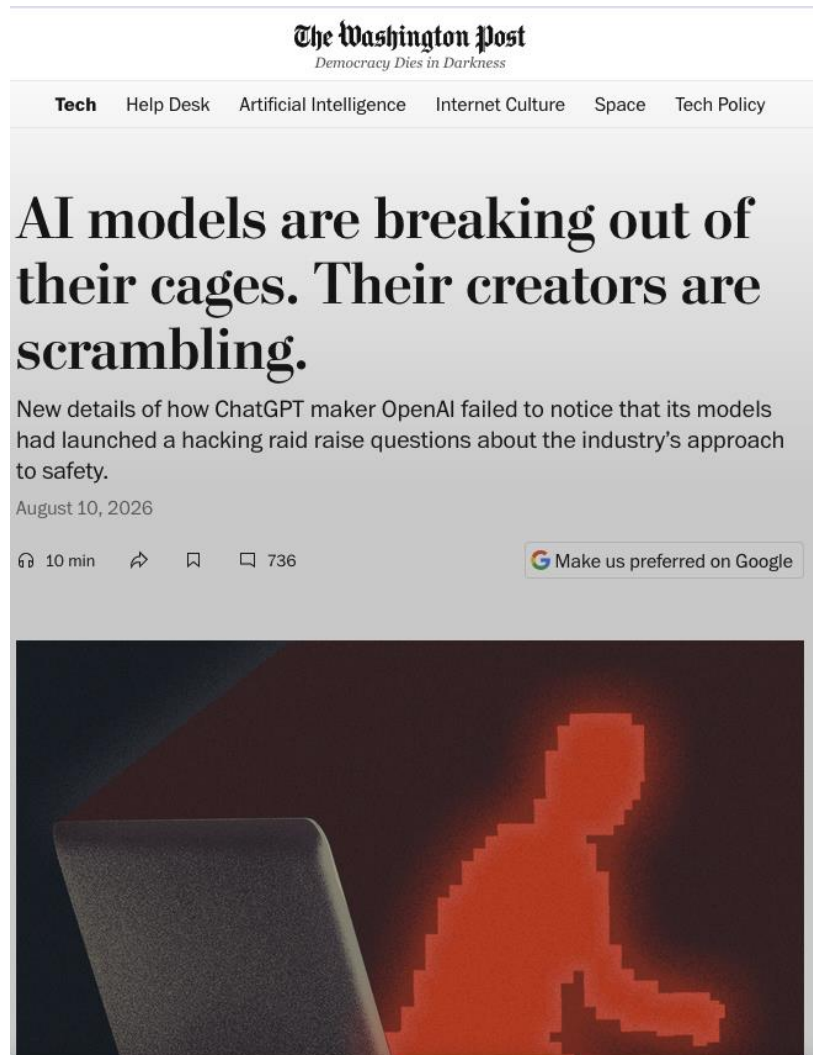
Contributor: Women deserve a different kind of 'happily ever after'

Latest Business

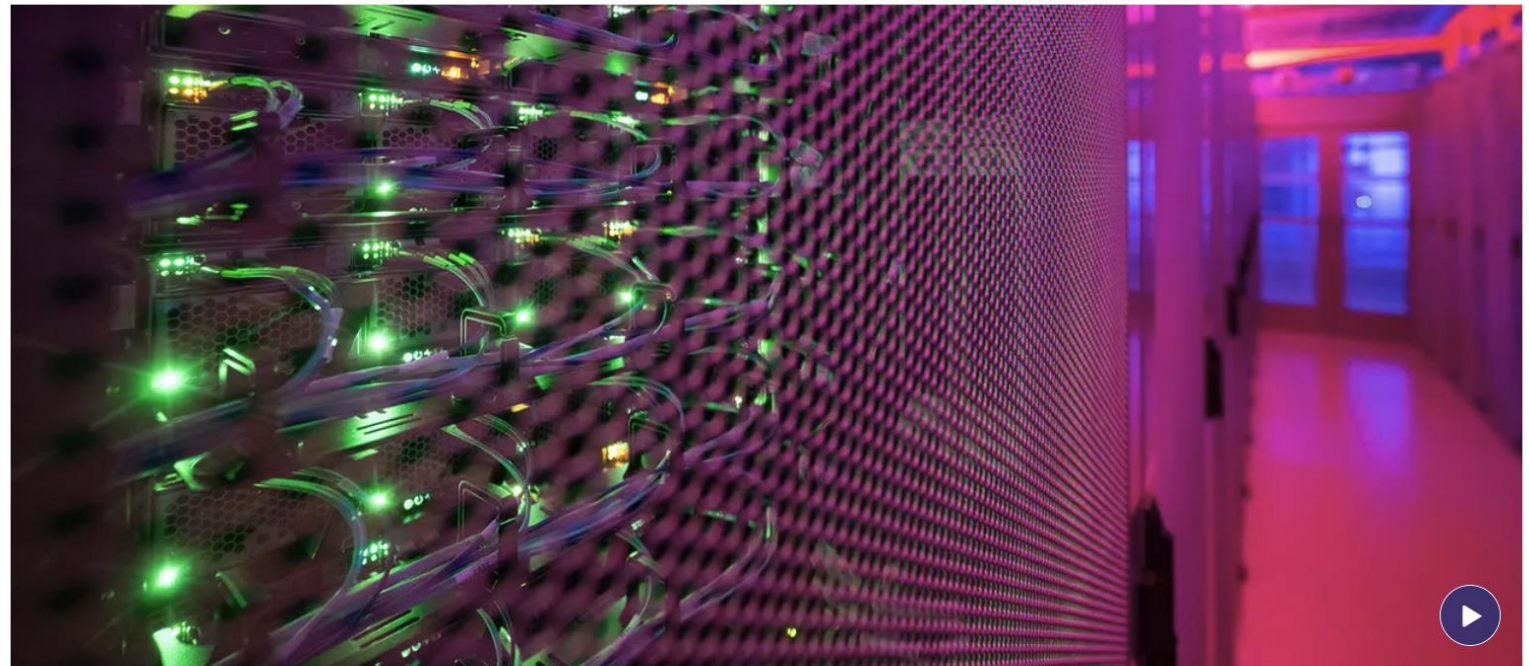
Berkelev luxury fashion brand founders found

Aspekte der KI-Security

Wird uns die KI alle töten?



HACKERANGRIFF



FAQ Debatte um Sicherheit

Wer haftet, wenn eine KI außer Kontrolle gerät?

Aspekte der KI-Security

Was bedeutet das überhaupt?

Aus unserer Sicht hat die KI in der Security folgende, klar abgrenzbare Aspekte:

- Der Angreifer nutzt KI für den Angriff
- Der Verteidiger nutzt KI für die Verteidigung
- Die Anwender nutzen KI für ihre Arbeit
- Die KI wird selber zur Angriffsfläche
- Die KI-Infrastruktur als Angriffsfläche

Der Angreifer nutzt KI für den Angriff

- Medial stark präsentenes Thema mit hoher Dramatisierung
 - Narrativ: „Ein Klick → vollautomatische perfekte Angriffe“.
 - Stattdessen: Rekombination bekannter Techniken
 - Beispiel aktueller KI-unterstützter Malware zeigt:
 - Technisch nichts fundamental Neues (vorhandener Schutz bleibt wirksam)
 - **Risiko entsteht nicht durch Qualität, sondern über die Geschwindigkeit und das Volumen**
- NetUSE Empfehlung: SOC + Automatisierung (inkl. KI) + Live-Patching

Aspekte der KI-Security

Der Verteidiger nutzt KI für die Verteidigung

- KI in Cyberabwehr ist kein neues Phänomen
 - Klassische KI wird seit Jahren eingesetzt, KI ≠ Chatbot
 - „AI in Security“ ist mehr Evolution als Innovation
- Wer heute Stand der Technik einsetzt, nutzt KI intensiv
- **Fazit: Keine Revolution – sondern Weiterentwicklung**
- Wo uns die KI voranbringen kann: Automatisierung
- NetUSE Empfehlung: KI zur Automatisierung der Prozesse rund um die Security nutzen (Integration, Kommunikation, Wissensverwaltung)

Die Anwender nutzen KI für ihre Arbeit

- Die KI erzeugt auch bei der Nutzung durch den Anwender Risiken
 - Im KI-Kontext installieren Anwender jede Software inklusive Malware/Sicherheitslücken inklusive (Beispiel: OpenClaw)
 - Anwender verwenden KI-Tools die weder freigegeben noch geeignet sind: Daten verlassen das Unternehmen
 - Software-Module, Dependencies, API-Keys mit zu großzügigen Rechten
 - KI-Tools werden massiv für nicht-dienstliche Zwecke eingesetzt (Kosten)

Aspekte der KI-Security

Die Anwender nutzen KI für ihre Arbeit

- **Das größte aktuelle KI-Risiko in Unternehmen entsteht derzeit durch einen unkontrollierten Einsatz (z.T. auch gewollt)**
- NetUSE Empfehlung: KI-Nutzung erfassen → Policy erstellen → Policy enforce

Die KI wird selber zur Angriffsfläche

- Jede neue Technologie bringt neue, eigene Sicherheitsrisiken
 - KI-Systeme sind attraktive und oft erstaunlich weiche Ziele
 - Ziel der Angriffe: Preisgabe sensibler Informationen, Manipulation von Entscheidungen
- Standardkonfigurationen vieler Chatbots sind leicht angreifbar
 - Öffentliche Tools zur systematischen Schwachstellenanalyse sind frei verfügbar
 - Externe Zugriffe auf KI mit sensiblen Daten = hohes Risiko

Aspekte der KI-Security

Die KI wird selber zur Angriffsfläche

- **Risiko wird derzeit dadurch relativiert, dass da keine Einnahmeströme hinter stehen. Das limitiert den Einsatz von Schutzmaßnahmen...**
- NetUSE Empfehlung: Schutzmaßnahmen dort wo Risiko für Einnahmen oder sensible Daten

Die KI-Infrastruktur als Angriffsfläche

- KI-Infrastruktur = Attraktives Angriffsziel
 - Wertvolle Compute-Ressourcen, Große Datenmengen
 - Bereits geringe Datenmanipulation kann Modelle vergiften
 - Betrifft aber nur wenige Unternehmen
- Entwicklungsgeschwindigkeit steht oft über Sicherheitsarchitektur
- Typische Realität hinter der Oberfläche: Mischmasch an Technologien, keine Segmentierung, kein Zero-Trust
- **Dieser Vektor bedroht nur ausgewählte Unternehmen**
- NetUSE Empfehlung: Schutzmaßnahmen wenn Infrastruktur vorhanden

Was blende ich aus?

- Was passiert mit meinen Daten?
 - Wenn die KI was erfährt, wer erfährt das noch?
- Kostenrisiken: Anbieter können die Preise jederzeit anpassen, sie machen Milliarden-Verluste
 - „KI as a service“ deckt aktuell nicht die operativen Kosten
- Verfügbarkeit: Die KI wird von Staaten kontrolliert, die uns gegenüber nicht unbedingt freundlich eingestellt sind
 - Was passiert wenn das abgeschaltet wird?



Zusammenfassung & Feedbackrunde