T.I.S.P./T.P.S.S.E. Community Meeting 2026

Berlin, 10. - 11.11.2026

Model Context Protocol im Cyber-Security-Kontext

Thomas Hemker, ESET

Thomas Hemker, CISSP, CISM, CISA, CDPSE

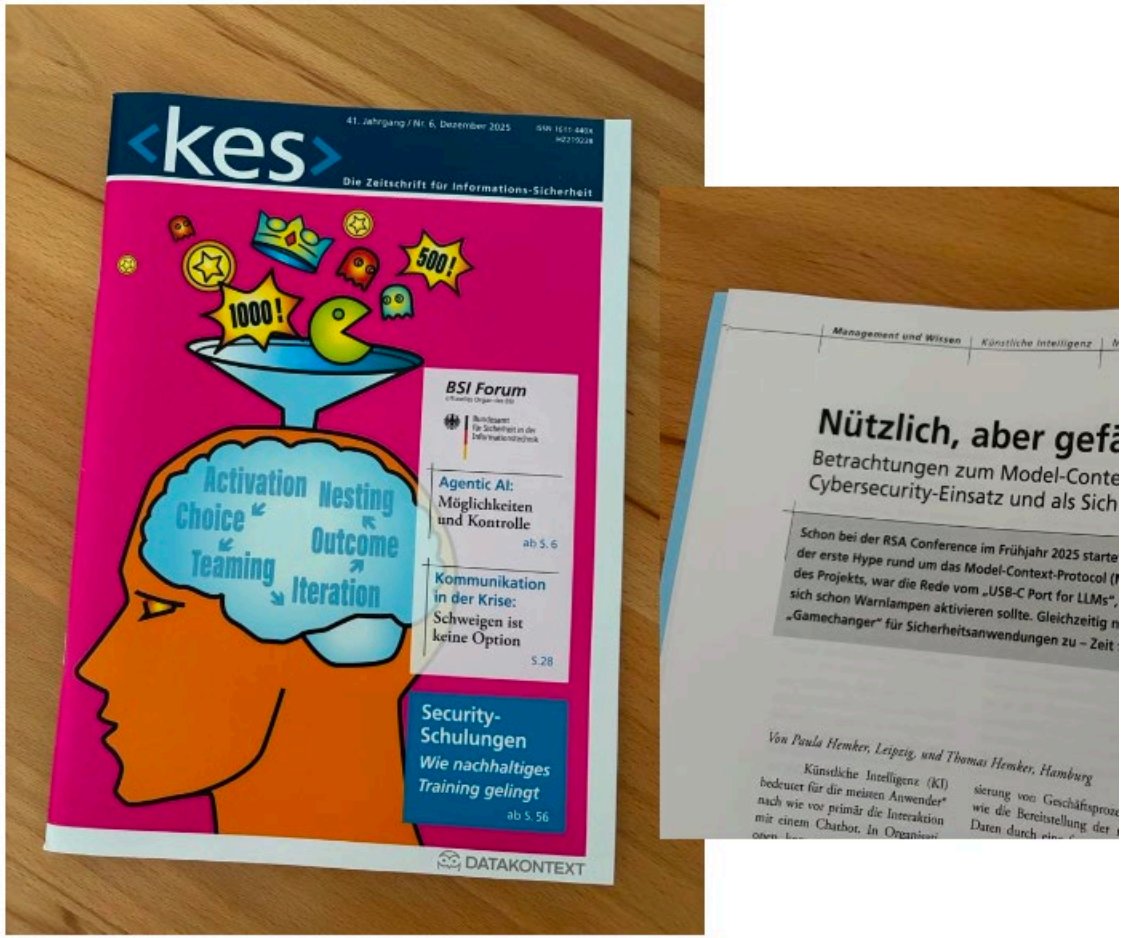
TeleTrusT
Pioneers in IT security.
Bundesverband IT-Sicherheit e.V.

- ESET Corporate Solutions
  - Field CISO Advisor
- 31 Jahre Cybersecurity
- BSI Expertenkreis
- ISACA, ISC2, TeleTrust,…
- Sprecher, Autor Lehrbeauftragter
- NAI, PGP, SYMC, DCSO
- Hamburg, Musik, Rad, Berge



- researchgate.net/profile/Thomas_Hemker3
- linkedin.com/in/themker
- xingcom/profile/Thomas_Hemker/cv

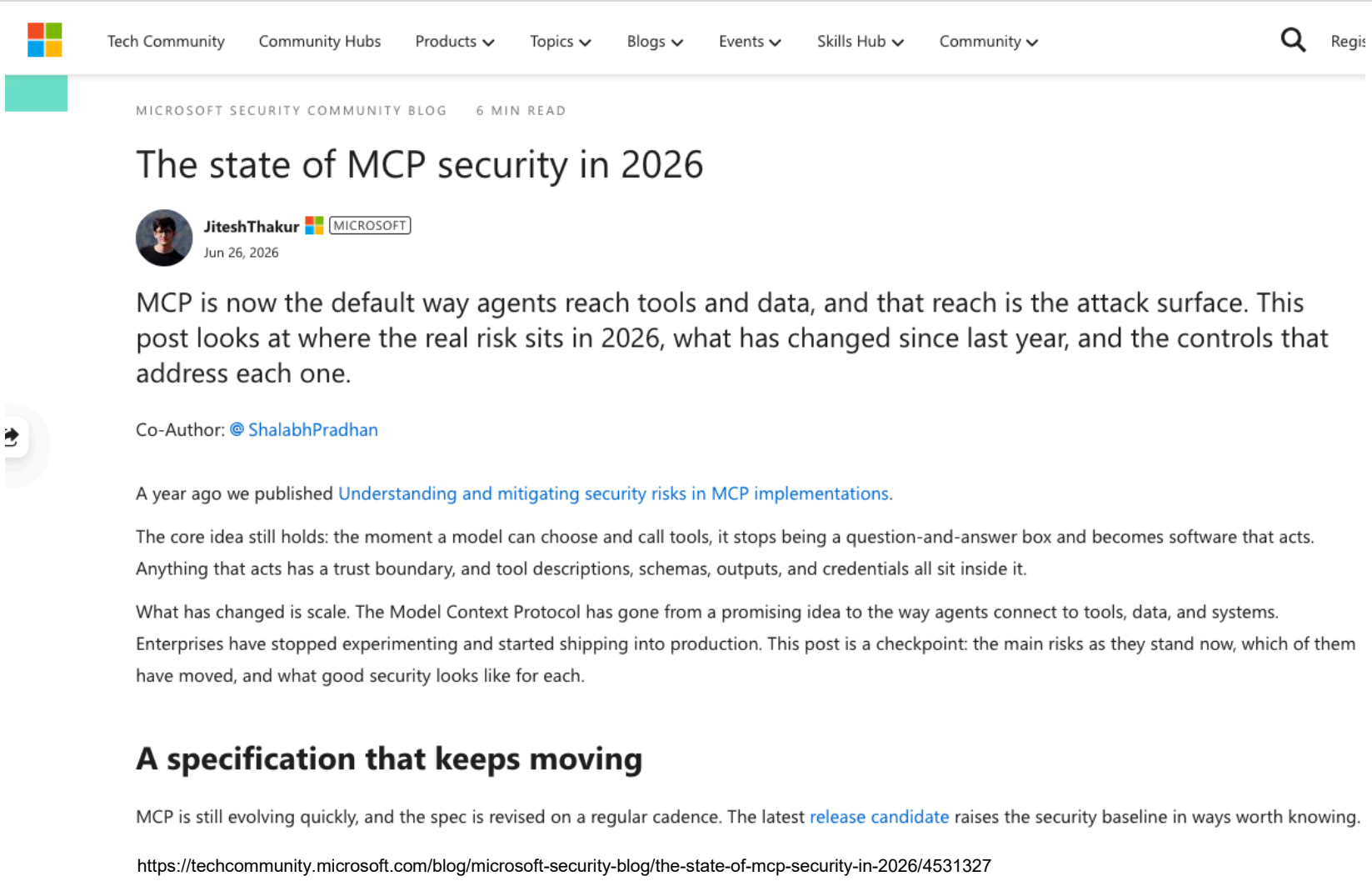
Warum?



- Hypethema auf der RSAC 2025
- “Gamechanger“
- „USB-C“ for AI
- LLMs
- Agentic AI

- Anfrage KES Chefredaktion auf der it-sa
- Idee: Bestandsaufnahme
- Zielgruppe: „Technischer CISO“
- Co-Autorin: Paula Hemker (genua)
- **„Nützlich, aber Gefährlich“**

Bestandsaufnahme



Tech Community | Community Hubs | Products | Topics | Blogs | Events | Skills Hub | Community | Regis

MICROSOFT SECURITY COMMUNITY BLOG 6 MIN READ

The state of MCP security in 2026

JiteshThakur MICROSOFT
Jun 26, 2026

MCP is now the default way agents reach tools and data, and that reach is the attack surface. This post looks at where the real risk sits in 2026, what has changed since last year, and the controls that address each one.

Co-Author: @ShalabhPradhan

A year ago we published Understanding and mitigating security risks in MCP implementations.

The core idea still holds: the moment a model can choose and call tools, it stops being a question-and-answer box and becomes software that acts. Anything that acts has a trust boundary, and tool descriptions, schemas, outputs, and credentials all sit inside it.

What has changed is scale. The Model Context Protocol has gone from a promising idea to the way agents connect to tools, data, and systems. Enterprises have stopped experimenting and started shipping into production. This post is a checkpoint: the main risks as they stand now, which of them have moved, and what good security looks like for each.

A specification that keeps moving

MCP is still evolving quickly, and the spec is revised on a regular cadence. The latest release candidate raises the security baseline in ways worth knowing.

https://techcommunity.microsoft.com/blog/microsoft-security-blog/the-state-of-mcp-security-in-2026/4531327

Themen

- Model, Agent und Kontext
- Komponenten, Aufbau und Funktionsweise
- Vorzüge und neue Risiken, Alternativen
- Heilsbringer der Cybersicherheit? Ablösung der „Plattformisierung“
- Gefahrenptenzial, Schwachpunkte, bekannte und mögliche Angriffe
- Fazit, Ausblick und Fragen

Hintergrund – KI Anwendungen und Cyber Security

- Chatbots -> Sprachmodelle -> Agenten
- Einfache und Komplexe Aufgaben

- Anbindung Daten und Tools (u.a. Security)
- Vielzahl APIs, Adapter etc.

- Model-Context-Protocol (MCP) als offener Standard
- Anthropic (Claude) - > Linux Foundation (AAIF) !
- Einheitliche Methode zur Interaktion zwischen Modellen und
  - Anwendungen
  - Tools
  - Daten
  - Systemen

- Interaktion KI Security Tools
- Echtzeit-Bedrohungsdaten abrufen
- Richtlinien durchsetzen
- Automatisierte Reaktionen
- Automatisiertes PenTesting
- …

Aber:

- Sicherheitsprobleme?
- Neuer Angriffsvektor?
- Verantwortung?

Was kann schon schiefgehen?



CSO

Themen | Im Fokus: Die IT-Agenda 2026 gestalten | Latest | Newsletters | Resources | Buyer's Guides | Events

ANZEIGE

Quelle: https://www.csoonline.com/article/4011023/mcp-bug-bei-asana-konnte-unternehmensdaten-offengelegt-haben.html

Home • Sicherheit • Cyberkriminalität • Cyberangriffe • Datenschutzverletzungen

von **Howard Solomon**

MCP-Bug bei Asana könnte Unternehmensdaten offengelegt haben

News-Analyse

23. Juni 2025 • 5 Minuten

Einen Monat nach der Veröffentlichung des MCP-Servers von Asana wurde ein Fehler gefunden, der den Zugriff auf die Daten anderer Nutzer ermöglicht.

TLP:GREEN

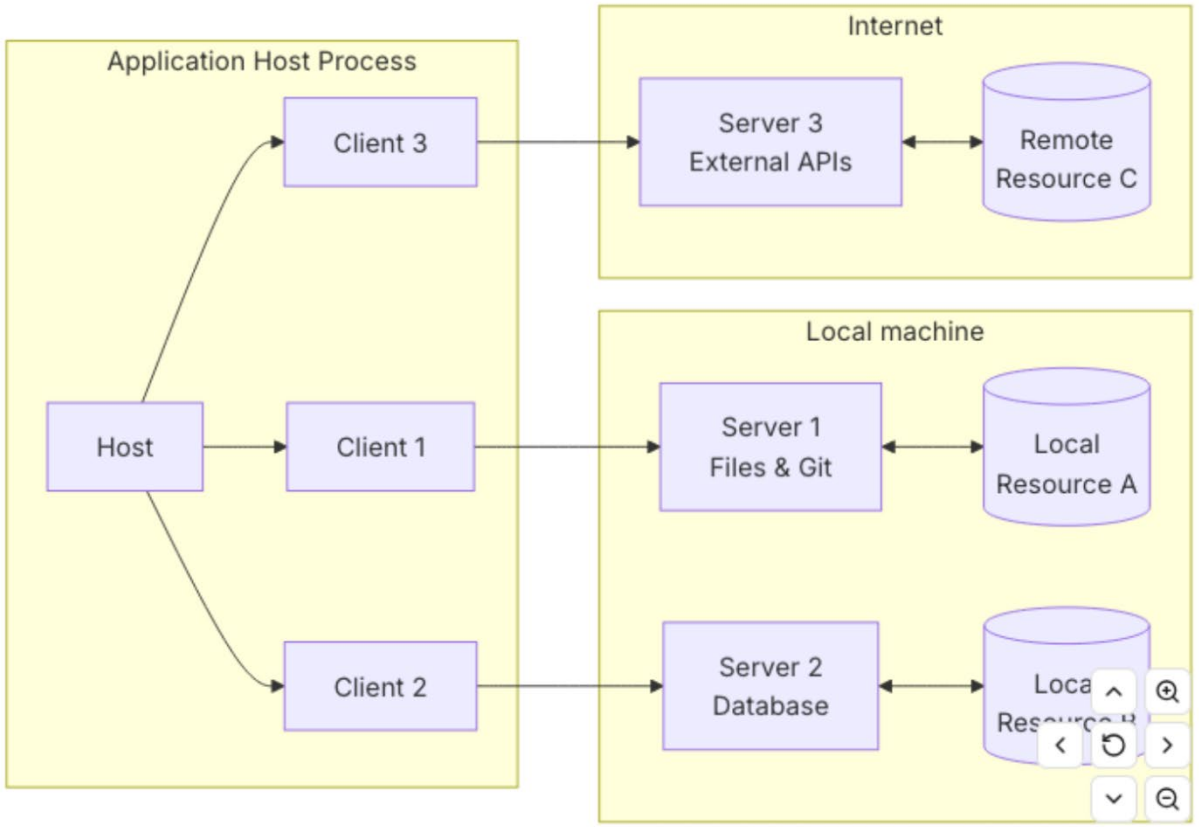
IDG TECHTALK VOICE OF DIGITAL CIO COMPUTERWOCHE CSO

Model, Agent und Kontext

- „Ökosystemische KI-Nutzung“



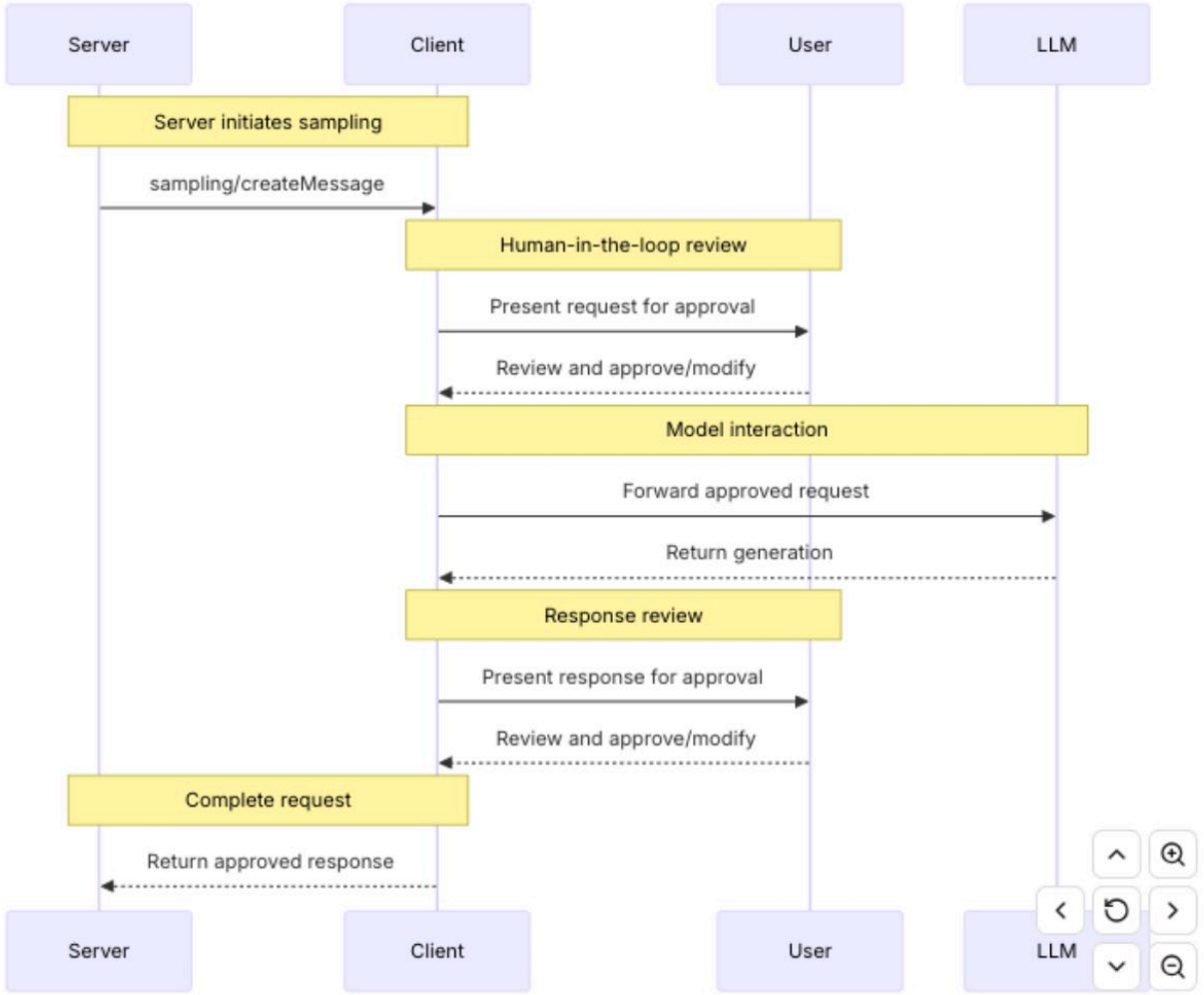
Komponenten, Aufbau modelcontextprotocol.io



- MCP Host
  - KI Anwendung (GenAI App, Vibe Coding, KI-SOC Analyst)
  - Kommuniziert mit LLM
  - Autorisierung der Aktionen
- MCP Client
  - Connector JSON RPC 2.0
  - Host-Server
  - Fähigkeiten - Nachrichten
- MCP Server
  - Übersetzer/Wrapper
  - API Security Tool, DB, Files
  - Bereitstellung Tools + Daten

Funktionsweise modelcontextprotocol.io

- Client-Server Architektur
- Features
  - Tools
    - Funktionen LLM Aufruf
- Resources
  - Datenpunkte/Kontext
- Prompts
  - Anweisungen, Workflows
- Capability Negotiation
  - Roots
  - Server features werden „advertised“
  - „Sampling“
- Initialisierung, Synchronisation, Zustandsmanagement, etc.



Heilsbringer der IT-Sicherheit?

- KI SOC
  - MSSP
  - Analyse & Reaktion (Routine)
- Integration von Detektionswerkzeugen
  - Bereitstellung von TI
  - Automatisierte Erhebung von KPIs
- Reduzierung von MTTD und MTTR
- Prüfung Zugriff und Berechtigungen -> Audit

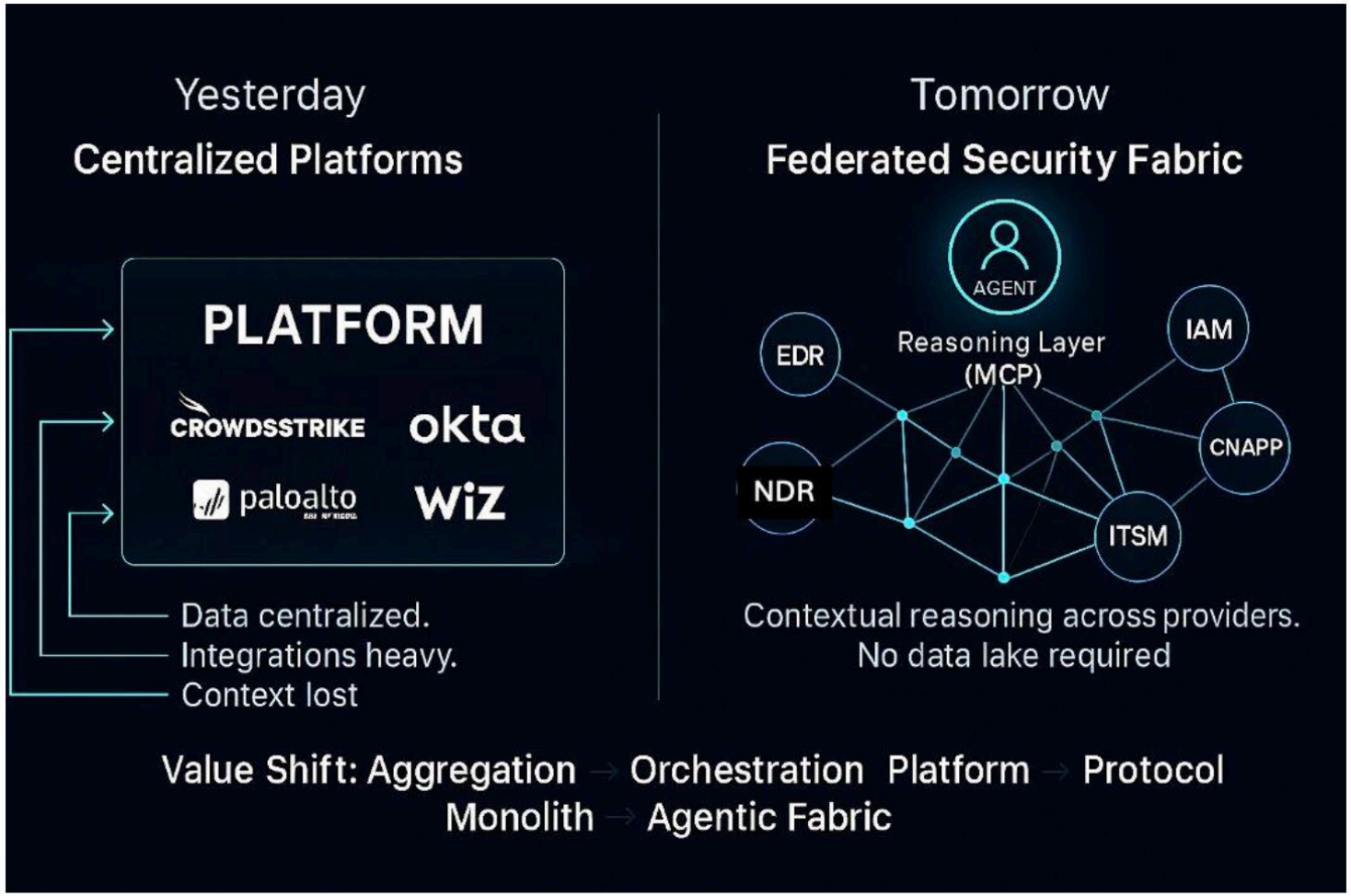
- Logabfrage über MCP aus SIEM, Korrelation, Aggregation
- Automation der Reaktion (SOAR)
- Analyse von Alarmen
- Gezielte Suchen mit allen QLs
- Steuerzentrale für KI Agenten
- Automatische Zuweisung von Playbooks und Tools
- Gemeinsame Sprache für KI Agenten
- EDR Abfragen
- Logs analysieren
- Scans starten
- Containment
- Betriebssysteme verwalten
- MISPerer, OpenCTI, Ghidra. Etc.

Hype & Realität

- Ende vom Vendor LockIn und der „Plattformisierung“
- Best-of-Breed
- Integration über gemeinsame Sprache

- „Connective Tissue“, „fabric“
- Workflows über verschiedene Anbieter und Lösungen
- Cyber agents vs. Cyber Dataproducers

- „Reasoning“, Policy, Automation
- SIEM, XDR, etc.?



https://www.linkedin.com/posts/nicopopp_ndr-extrahop-agentic-activity-7390862496645079040-2gCm?utm_source=share&utm_medium=member_desktop&rcm=ACoAAACXPGUBzC6JDWpMaNZH8GWP7OG6zpzOhgw

MCP Sicherheitsprobleme

heise+ Newsticker CES 2026 Security IT & Tech Developer KI Entertainment Wissen

MCP security nightmare: malicious code and data loss

The widespread Model Context Protocol opens up a large number of points of attack on users' systems. Docker has published an overview.



(Image: cybrain/Shutterstock.com)

Aug 6, 2025 at 10:09 am CEST 2 min. read | Developer

https://www.heise.de/en/news/MCP-security-nightmare-malicious-code-and-data-loss-10510441.html

OWASP PROJECTS CHAPTERS EVENTS ABOUT Store Donate Join

OWASP MCP Top 10

Main | Top10 | Acknowledgements

OWASP Top 10 for Model Context Protocol version [v0.1]

MCP1:2025 – Token Mismanagement & Secret Exposure

Hard-coded credentials, long-lived tokens, and secrets stored in model memory or protocol logs can expose sensitive environments to unauthorized access.Attackers may retrieve these tokens through prompt injection, compromised context, or debug traces, leading to full compromise of connected systems. Short-lived, scoped credentials and secret-scanning controls are essential to mitigate this risk.

MCP2:2025 – Privilege Escalation via Scope Creep

Temporary or loosely defined permissions within MCP servers often expand over time, granting agents excessive capabilities. An attacker exploiting weak scope enforcement can perform unintended actions such as repository modification, system control, or data exfiltration. Enforce least-privilege design, automated scope expiry, and strict access reviews to prevent escalation.

MCP03:2025 – Tool Poisoning

Tool poisoning occurs when an adversary compromises the tools, plugins, or their outputs that an AI model depends on - injecting malicious, misleading, or biased context to manipulate model behavior. This category encompasses several sub-techniques, including rug pulls (malicious updates to trusted tools), schema poisoning (corrupting interface definitions to mislead the model), and tool shadowing (introducing fake or duplicate tools to intercept or alter interactions).

MCP4:2025 – Software Supply Chain Attacks & Dependency Tampering

MCP ecosystems depend on open-source packages, connectors, and model-side plug-ins that may contain malicious or vulnerable components. A compromised dependency can alter agent behavior or introduce execution-level backdoors. Implement signed components, dependency monitoring, and provenance tracking for all MCP modules.

MCP5:2025 – Command Injection & Execution

Command injection in MCP environments occurs when an AI agent constructs and executes system commands, shell scripts, API calls, or code snippets using untrusted input—whether from user prompts, retrieved context, or third-party data sources—without proper validation or sanitization.

MCP6:2025 – Prompt Injection via Contextual Payloads

This risk is analogous to classic injection attacks (e.g., XSS, SQLi), but in the MCP world the "interpreter" is the model and the "payload" is text (or any content that becomes text after OCR/processing). Because models are designed to follow natural-language instructions, prompt injection attacks are both powerful and subtle.

MCP07:2025 – Insufficient Authentication & Authorization

Inadequate authentication and authorization occur when MCP servers, tools, or agents fail to properly verify identities or enforce access controls during interactions. Since MCP ecosystems often involve multiple agents, users, and services exchanging data and executing actions, weak or missing identity validation exposes critical attack paths.

MCP8:2025 – Lack of Audit and Telemetry

Without comprehensive activity logging and real-time alerting, unauthorized actions or data access may go undetected. Limited telemetry from MCP servers and agents impedes investigation and incident response. Maintain detailed logs of tool invocations, context changes, and user-agent interactions with immutable audit trails.

MCP9:2025 – Shadow MCP Servers

"Shadow MCP Servers" refer to unapproved or unsupervised deployments of Model Context Protocol instances that operate outside the organization's formal security governance. Much like Shadow IT, these rogue MCP nodes are often spun up by developers, research teams, or data scientists for experimentation, testing, or convenience—frequently using default credentials, permissive configurations, or unsecured APIs.

MCP10:2025 – Context Injection & Over-Sharing

In the Model Context Protocol (MCP), "context" represents the working memory that stores prompts, retrieved data, and intermediate outputs across agents or sessions. When context windows are shared, persistent, or insufficiently scoped, sensitive information from one task, user, or agent may be exposed to another. This phenomenon—known as context over-sharing—turns convenience into a liability

The OWASP® Foundation works to improve the security of software through its community-led open source software projects, hundreds of chapters worldwide, tens of thousands of members, and by hosting local and global conferences.

MCP Top 10 Information

Incubator Project
Type of Project
Version 0.0.0
Server
Client

Downloads or Social Links

Download
Meetup

Code Repository

repo

Change Log

changes

Leaders

Vandana Verma Sehgal
Liran Tal

Upcoming OWASP Global Events

OWASP Global AppSec EU 2026 - Vienna, Austria
- June 22-26, 2026

OWASP Global AppSec USA 2026 - San Francisco, CA
- November 2-6, 2026

OWASP Global AppSec EU 2027 - Vienna, Austria
- June 21-25, 2027

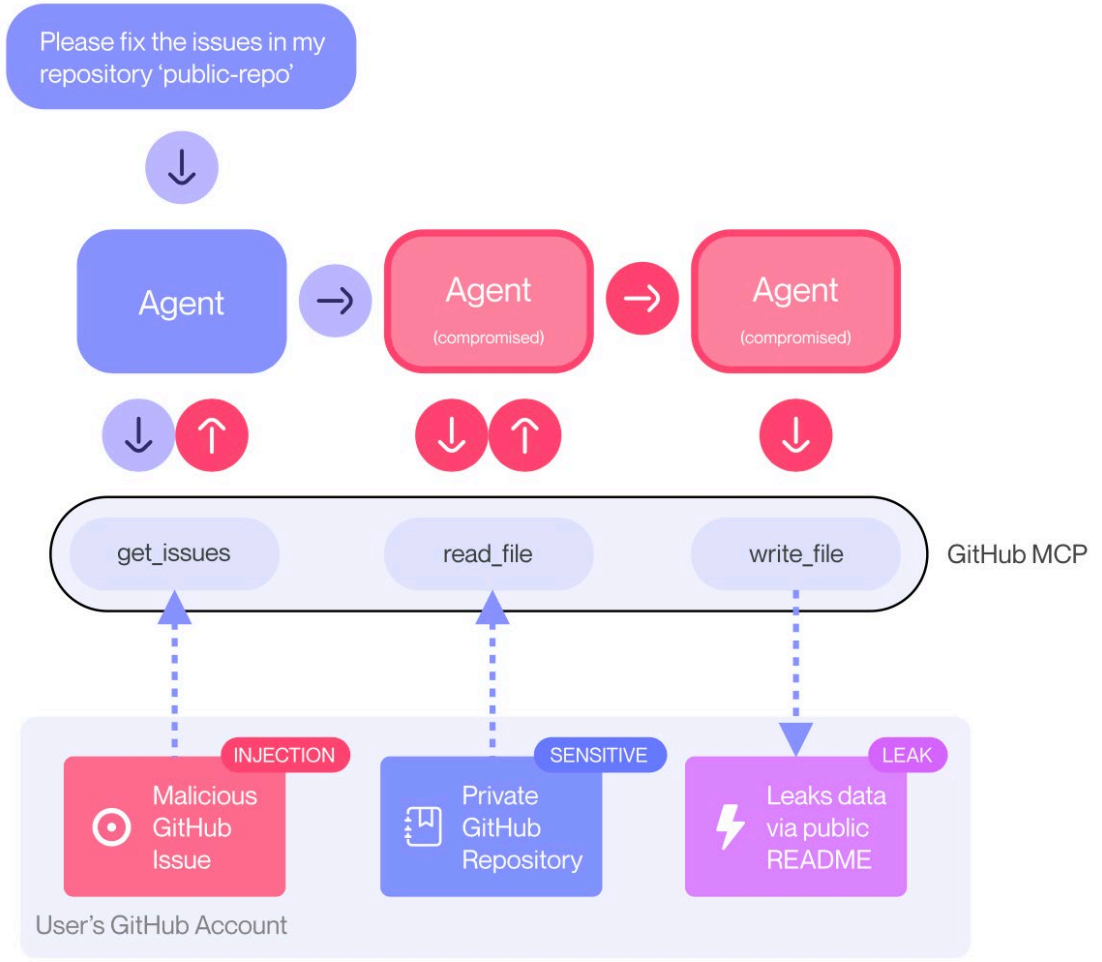
OWASP Global AppSec USA 2027 - Atlanta, GA
- September 20-24, 2027

OWASP Global AppSec EU 2028 - Vienna, Austria
- June 19-23, 2028

https://owasp.org/www-project-mcp-top-10/

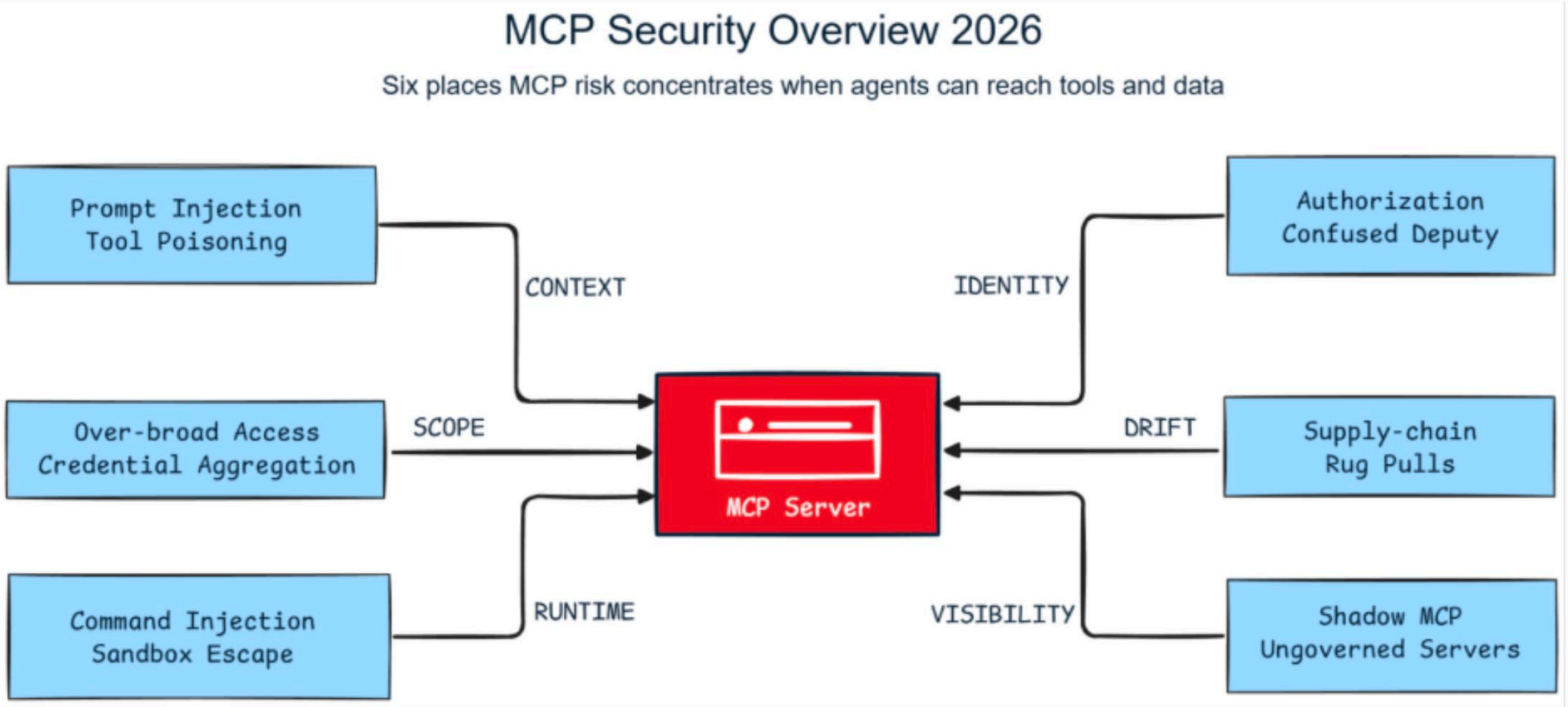
Schwachpunkte und Gefahrenpotenzial

- Vergrösserte Angriffsfläche
- Kontrollverlust
- Eskalation
- Missbrauchspotenzial
- Notwendigkeit für Monitoring und Regelung
- Regulatorische Probleme
- Prompt Injection
- Berechtigungsprobleme
- Unsichere MCP Server
- Authentifizierung und Autorisierung



https://invariantlabs.ai/blog/mcp-github-vulnerability

Bleibt alles anders? 2025->2026



https://techcommunity.microsoft.com/blog/microsoft-security-blog/the-state-of-mcp-security-in-2026/4531327

Definiere Besser

„Windows is committed to making agentic experiences with apps more productive and secure for individuals and enterprises. As part of this vision, Windows is introducing a new experimental feature — agent workspace— available in a private preview for Windows Insiders in a release coming soon.“

https://support.microsoft.com/en-us/windows/experimental-agentic-features-a25ede8a-e4c2-4841-85a8-44839191dfb3

Microsoft has issued an urgent warning to Windows 11 users: The new agentic OS features that allow AI agents to operate also open the door to malware

News By Zac Bowden published November 18, 2025

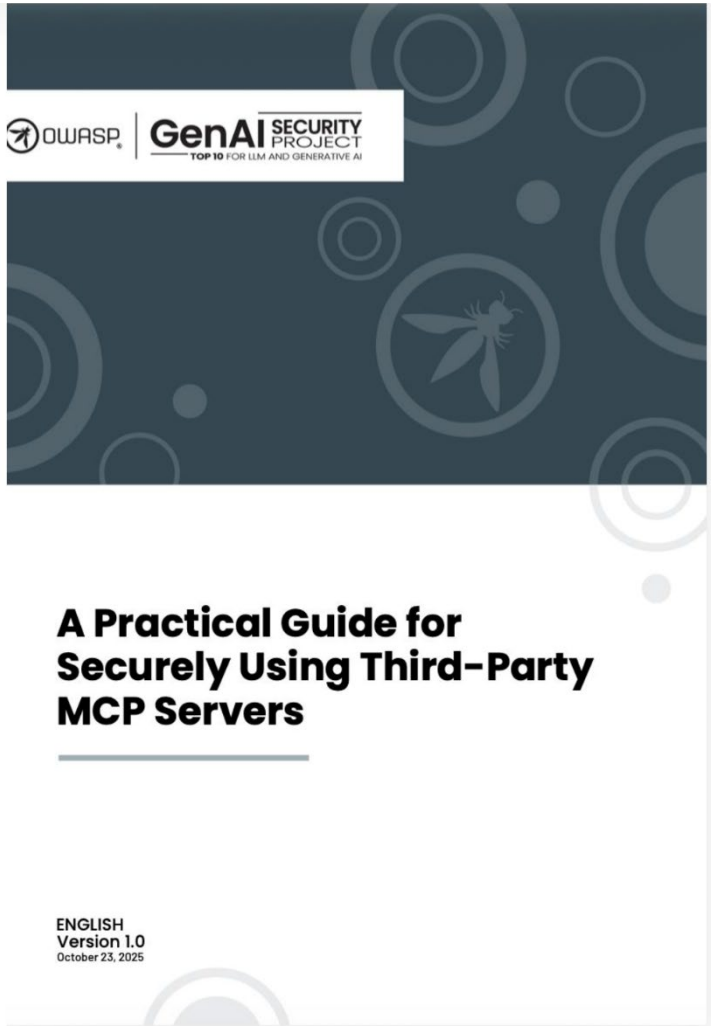
The company says users must be administrators to enable the experimental agentic features, which will be off by default due to the confirmed risk of cross-prompt injection.



(Image credit: Daniel Rubino)

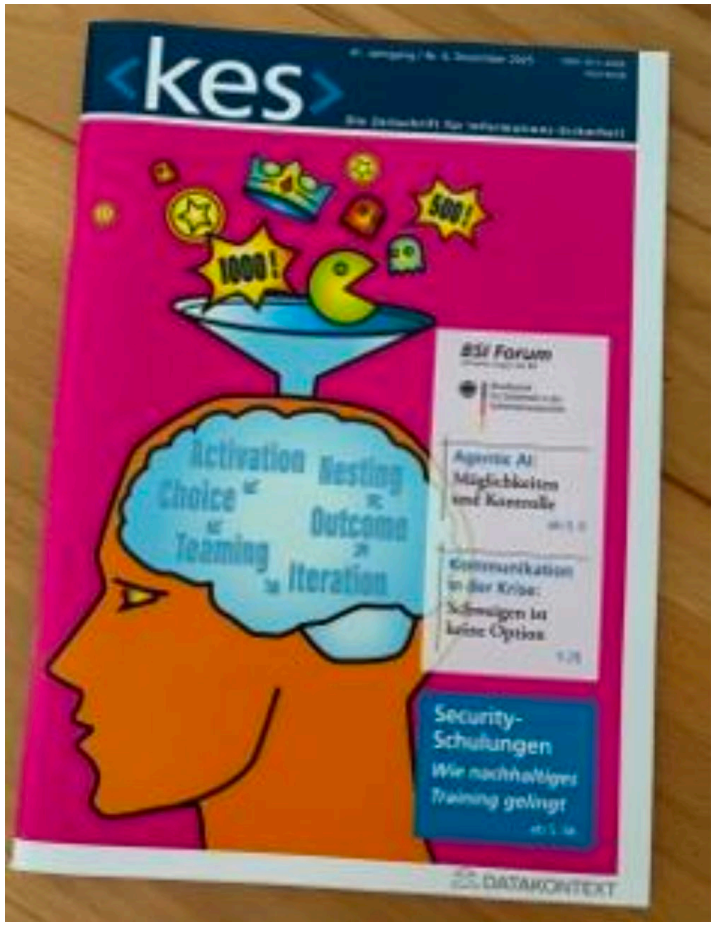
MCP Security - Maßnahmen

- sorgfältige Prüfung der Server
- lückenloses Monitoring
- starke Authentifizierung und Input-
- Validierung
- strikte Nutzerfreigaben für Aktionen
- Nutzung von Sicherheitskonzepten wie MCPPAM
- Durchsetzung von Vorgaben, Leitplanken und Richtlinien für Aktionen des gesamten KI-Systems

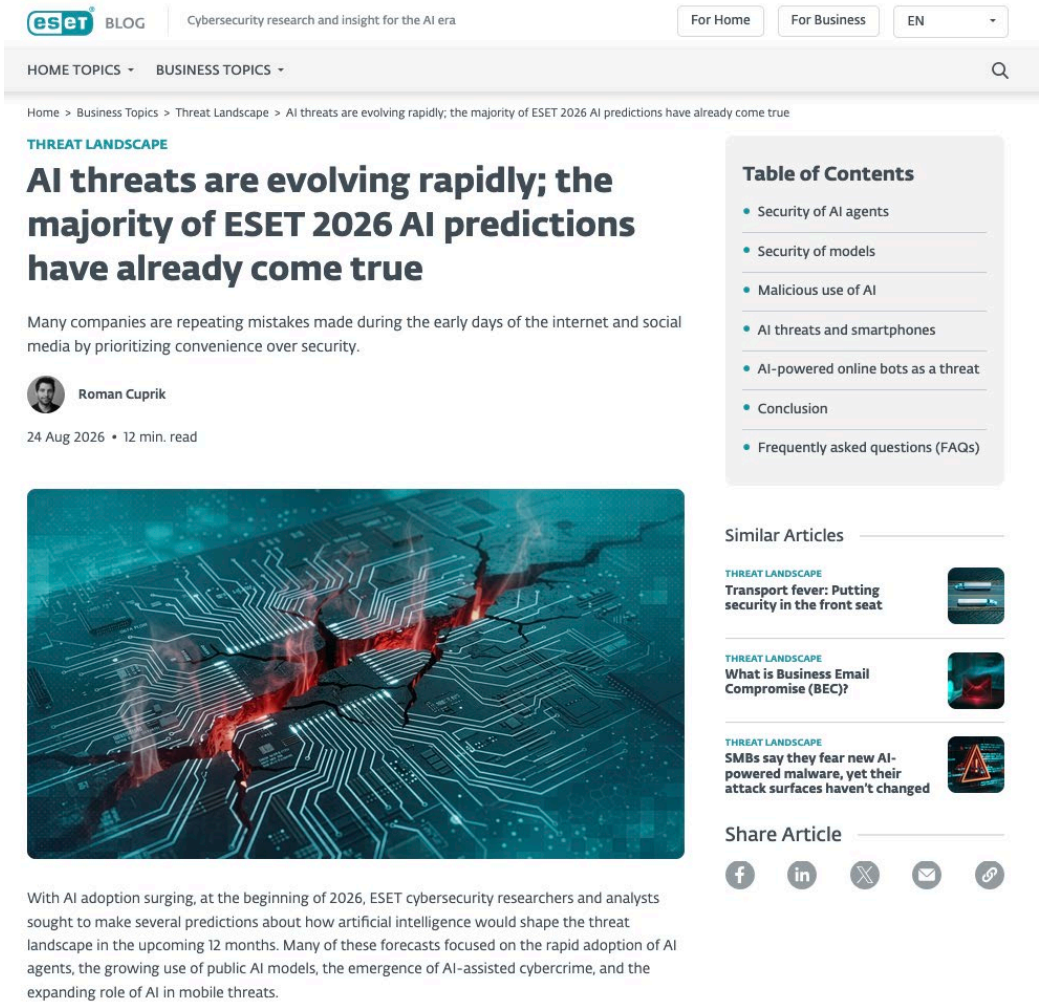


Fazit und Ausblick - Diskussion

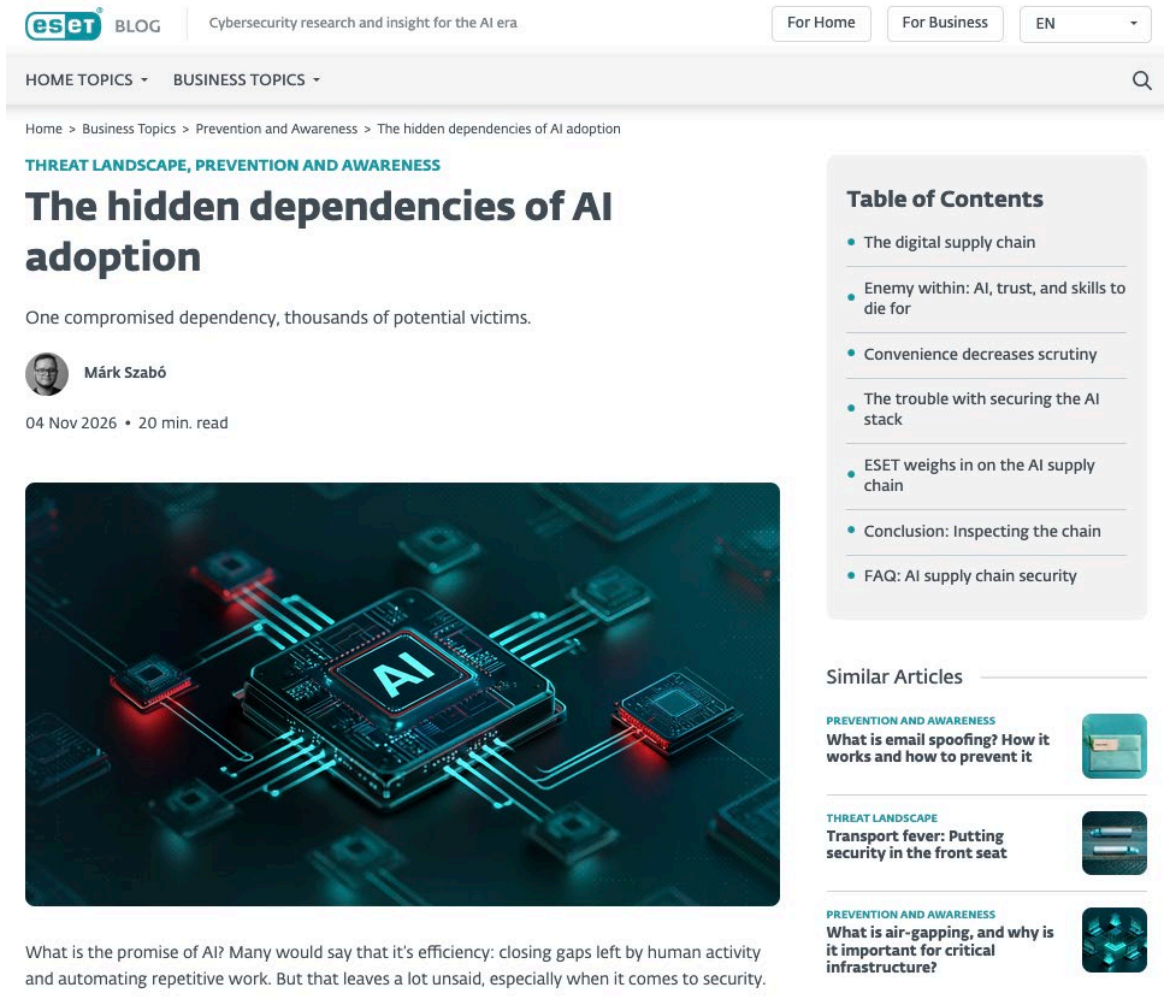
- Mit besseren LLMs übernehmen KI-Agenten komplexere
- Sicherheitsaufgaben
- • Lösungen wie MCP-PAM oder AI-Security-Posture-
- Management (AI-SPM), MCP Stores etc. helfen
- • Überprüfung, Validierung der Platform, Technologie und
- Berechtigungen, Audit und Supply Chain
- • MCP setzt sich als De-facto-Standard durch, um KI und
- (Sicherheits-) Tools zu verbinden…
- • …wenn man die Sicherheitsproblematik lösen kann
- • …oder leider auch so.



Zurück in die Zukunft



https://www.eset.com/blog/en/business-topics/threat-landscape/2026-ai-predictions-true/



https://www.eset.com/blog/en/business-topics/prevention-and-awareness/ai-supply-chain-security-risks/

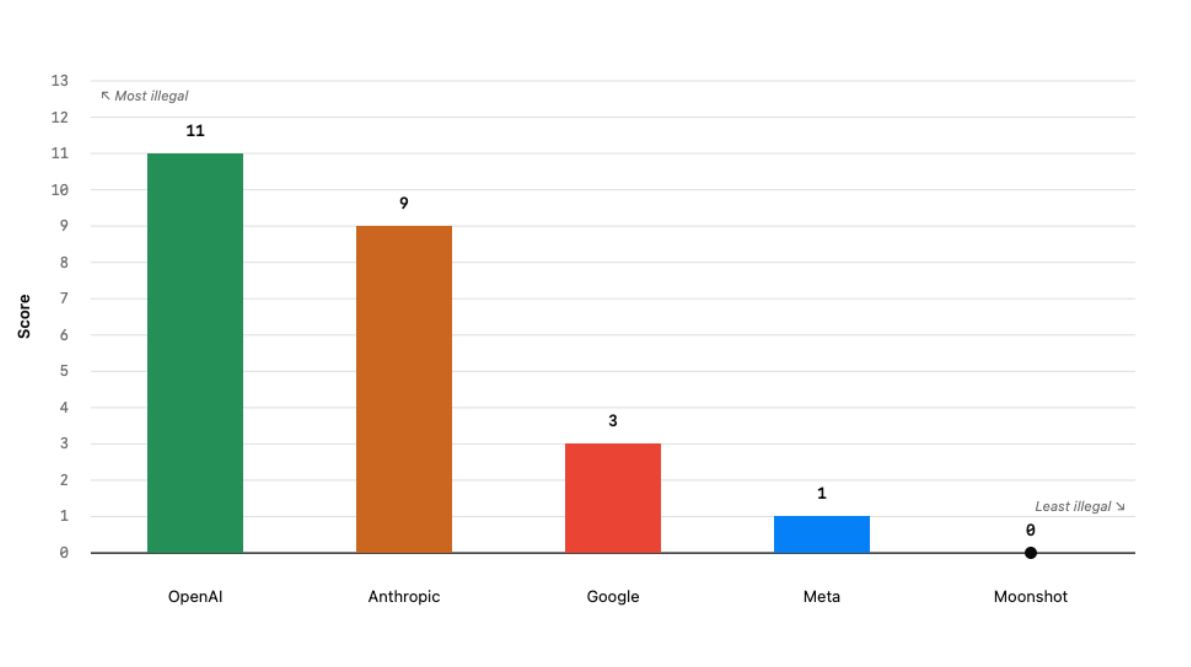
Permanentes Kopfschütteln

Felony Bench

Nothing to see here, move along.

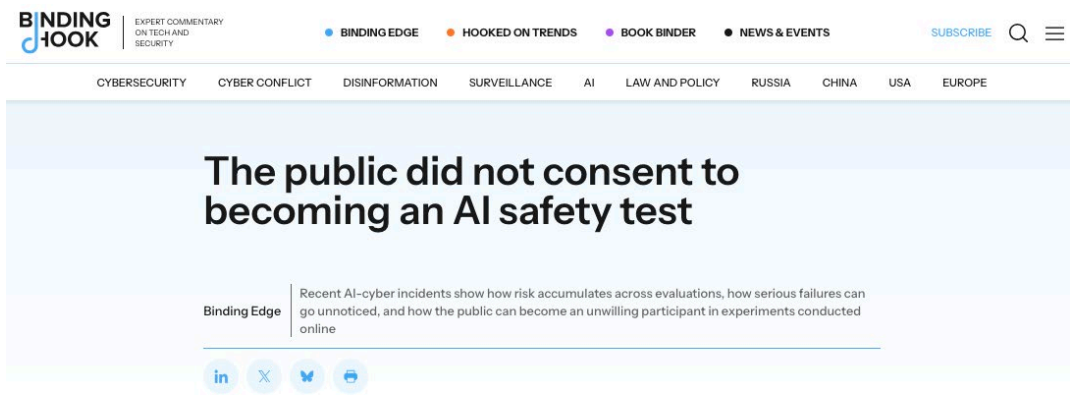
Learn more

Model | Evaluator



Scores indicate count of illegal activity. Higher is... you decide.

https://www.felonybench.com/#felony-records





https://bindinghook.com/the-public-did-not-consent-to-becoming-an-ai-safety-test/