

# AI Everything: Das Doppelspiel der KI

Waffe der Angreifer &  
Schild der Verteidiger

**avodaq**

# Sebastian Bonk

Leiter Software-Abteilung, avodaq AG

## Ausbildung

Erste juristische Prüfung 2014 erfolgreich abgeschlossen.

## avodaq AG

- Inhouse Applications
- avodaq Software Products
- Auftragsentwicklung
- Co-Development
- DevOps/Platform Engineering
- Observability
- Cloud-Native Infrastructure
- AI Lab

### Blockchain Patent

„Method and System  
operating  
a Domain Name System  
using a Blockchain“

EP4468654\*



\*[https://patentscope.wipo.int/search/en/detail.jsf?docId=EP443358962&\\_fid=EP471844920](https://patentscope.wipo.int/search/en/detail.jsf?docId=EP443358962&_fid=EP471844920)



**1997**  
gegründet



**1 Patent**  
DNS Blockchain

**300**

Mitarbeitende



**kununu**  
**4,3**



**73 Mio.**  
EUR Umsatz  
in 2025

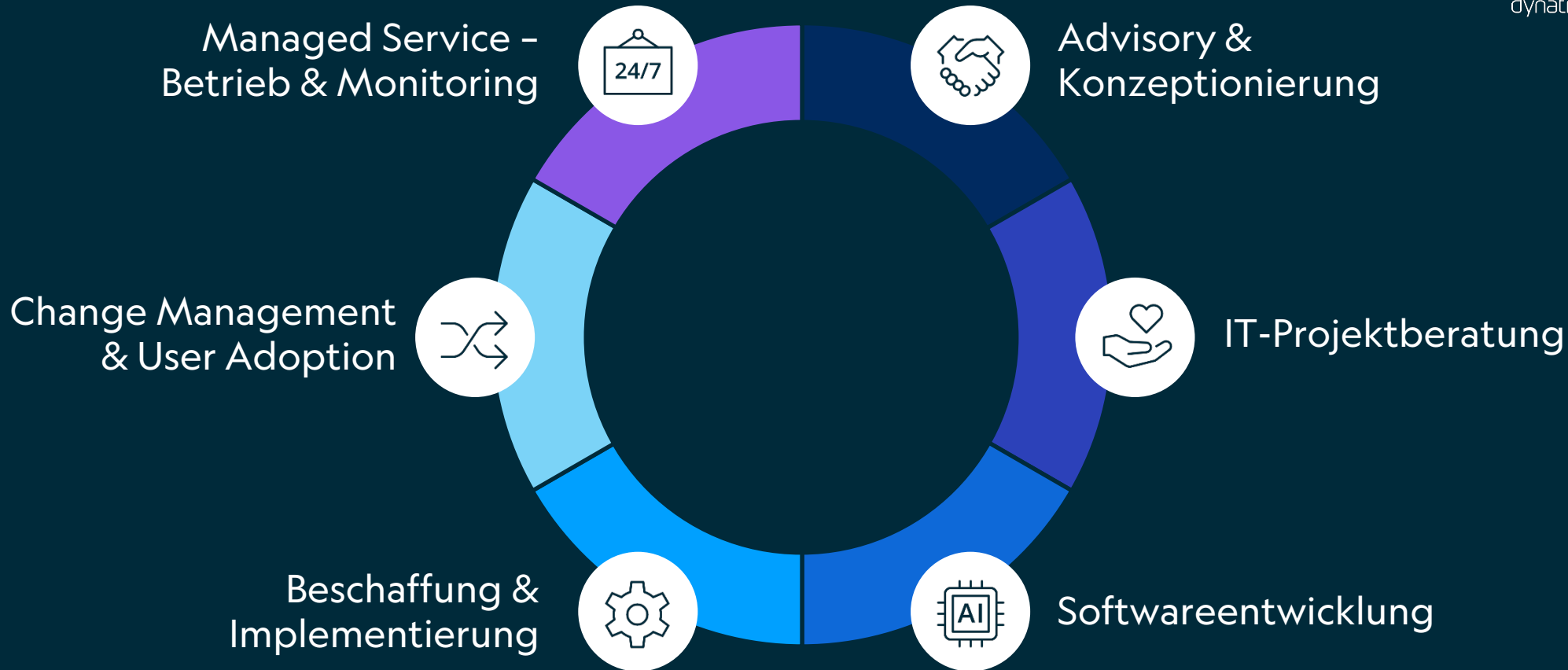
**100 %**  
zertifizierte  
TechnikerInnen

**ISO 27001**  
zertifiziert

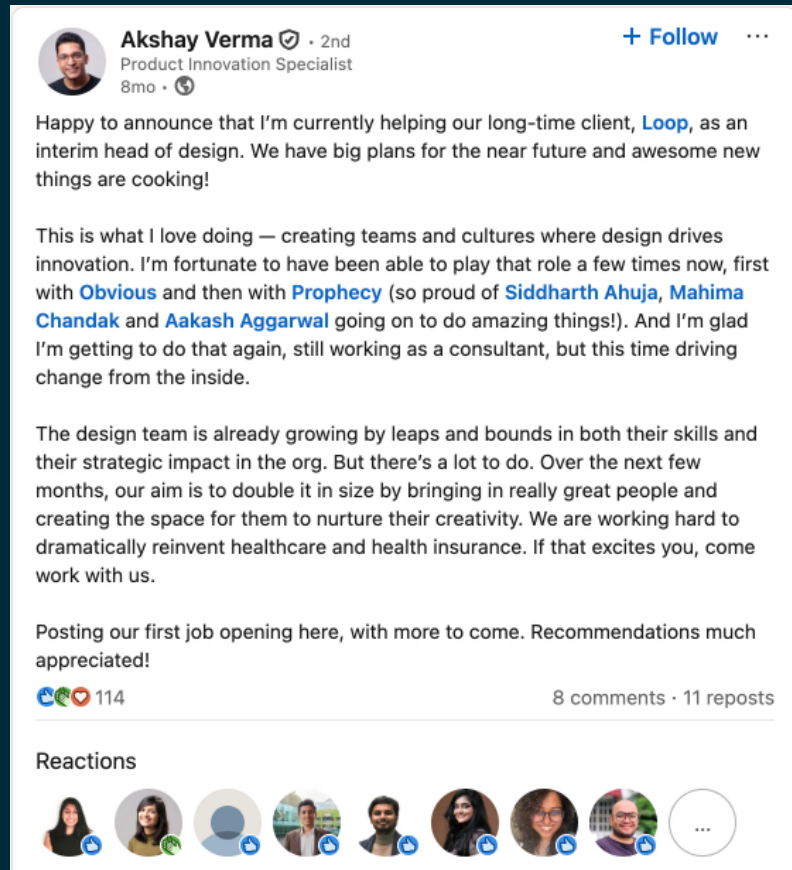
**ISO 9001**  
**ISO 14001**  
in Bearbeitung





# Unsere Leistungen



# Impact von AI




**Akshay Verma**  · 2nd  
Product Innovation Specialist  
8mo · 

Happy to announce that I'm currently helping our long-time client, [Loop](#), as an interim head of design. We have big plans for the near future and awesome new things are cooking!


This is what I love doing — creating teams and cultures where design drives innovation. I'm fortunate to have been able to play that role a few times now, first with [Obvious](#) and then with [Prophecy](#) (so proud of [Siddharth Ahuja](#), [Mahima Chandak](#) and [Aakash Aggarwal](#) going on to do amazing things!). And I'm glad I'm getting to do that again, still working as a consultant, but this time driving change from the inside.

The design team is already growing by leaps and bounds in both their skills and their strategic impact in the org. But there's a lot to do. Over the next few months, our aim is to double it in size by bringing in really great people and creating the space for them to nurture their creativity. We are working hard to dramatically reinvent healthcare and health insurance. If that excites you, come work with us.

Posting our first job opening here, with more to come. Recommendations much appreciated!

 114 8 comments · 11 reposts

Reactions



Klassischer Jobwechsel

VS



 Dear Colleagues,

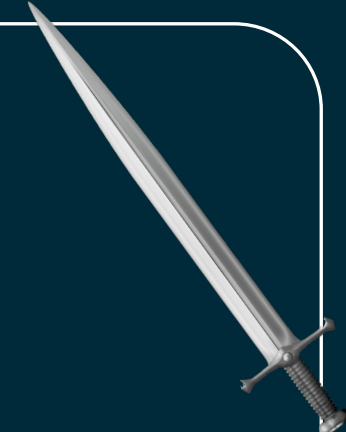
The world is in peril. And not just from AI, or bioweapons, but from a whole series of interconnected crises unfolding in this moment....



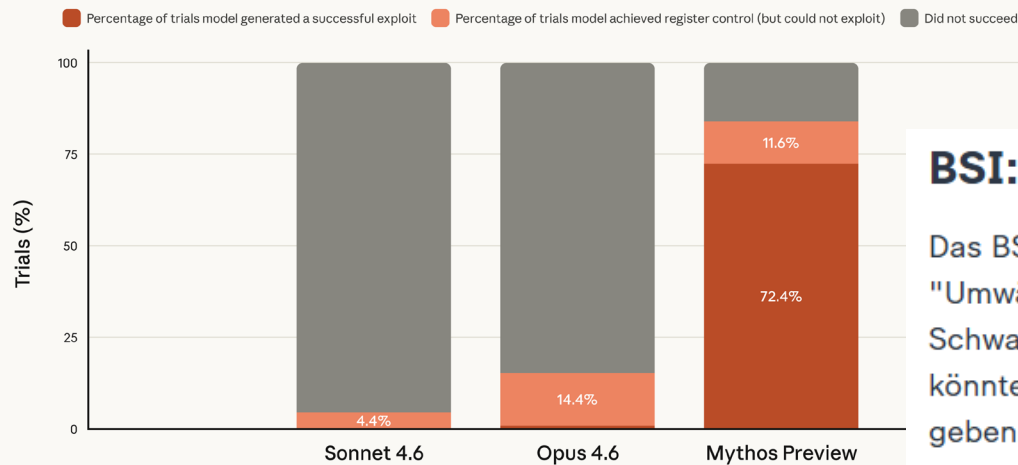
**ANTHROPIC'S HEAD OF AI SECURITY QUILTS, WARNS OF 'WORLD IS IN PERIL' IN CRYPTIC RESIGNATION LETTER**

AI Jobwechsel

# „Release“ von Anthropic Mythos



## Firefox JS shell exploitation



In a previous blog, we noted that Opus 4.6 was able to successfully generate exploits for crashes it found in Firefox in two separate trials out of many, which was a success rate of less than 1%. Next to Claude Mythos Preview, which succeeds at creating a working exploit nearly 100 times more often.

<https://red.anthropic.com/2026/mythos-preview/>

## BSI: Eine Frage der nationalen Sicherheit

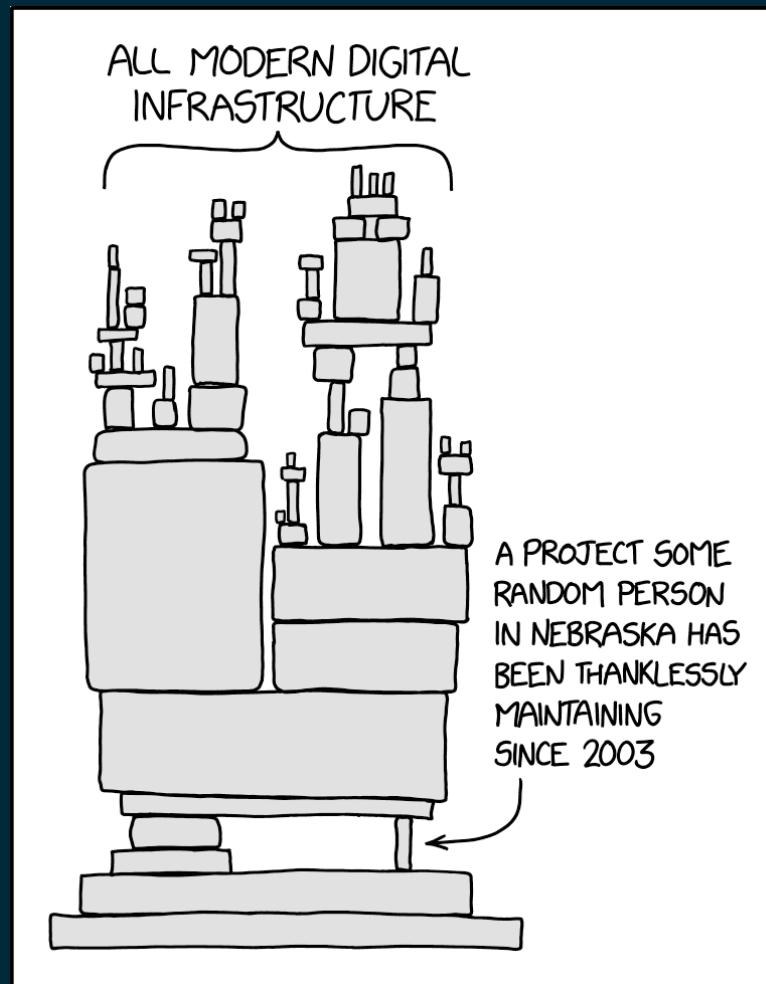
Das BSI nehme die Ankündigungen von Anthropic sehr ernst und erwarte "Umwälzungen im Umgang mit Sicherheitslücken und in der Schwachstellenlandschaft insgesamt", sagte Plattner. Konsequenz zu Ende gedacht, könnte es mittelfristig keine unbekannt klassischen Software-Schwachstellen mehr geben.

„Dies würde eine Verschiebung der Angriffsvektoren und einen Paradigmenwechsel mit Blick auf die Cyberbedrohungslage zur Folge haben.“

Claudia Plattner, BSI-Präsidentin

<https://www.zdfheute.de/politik/deutschland/ki-anthropic-claude-mythos-schwachstellen-software-bsi-100.html>

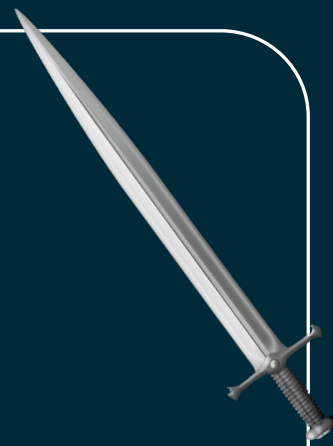
# Angriffe auf die Software Supply Chain



<https://xkcd.com/2347/>

# Braucht es wirklich „Mythos“?

---



**But here is what we found when we tested:** We took the specific vulnerabilities Anthropic showcases in their announcement, isolated the relevant code, and ran them through small, cheap, open-weights models. Those models recovered much of the same analysis. Eight out of eight models detected Mythos's flagship FreeBSD exploit, including one with only 3.6 billion active parameters costing \$0.11 per million tokens. A 5.1B-active open model recovered the core chain of the 27-year-old OpenBSD bug.

<https://aisle.com/blog/ai-cybersecurity-after-mythos-the-jagged-frontier>

# Warum Software-Entwicklung stark von AI beeinflusst wird

Inspiziert von Aaron Levie, CEO von Box



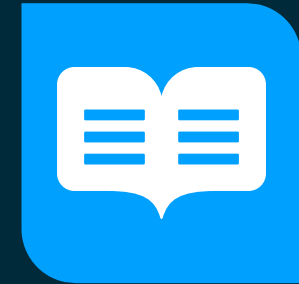
**Code ist verifizierbar**



**Software-Entwickler  
haben weitreichende  
Berechtigungen**



**Code ist 100% Text**



**Dokumentation ist  
Bestandteil des  
Entwicklungsprozesses**

**In Kontrast zu Aufgaben in den Bereichen**

**Sales**

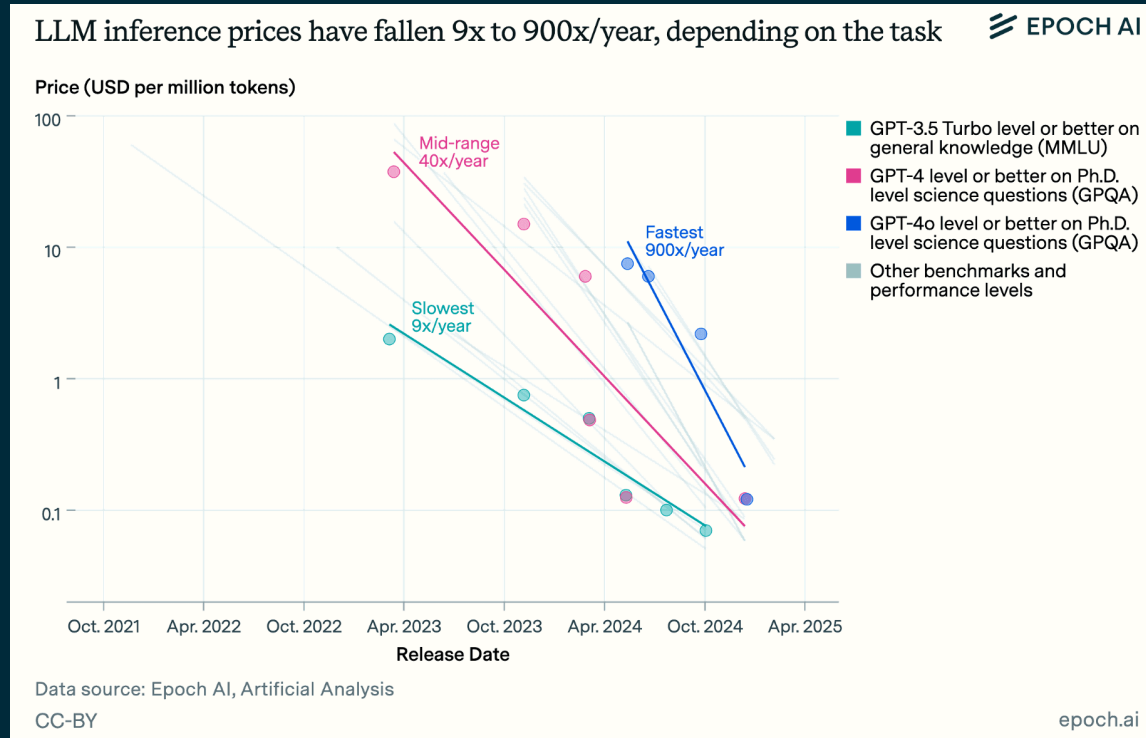
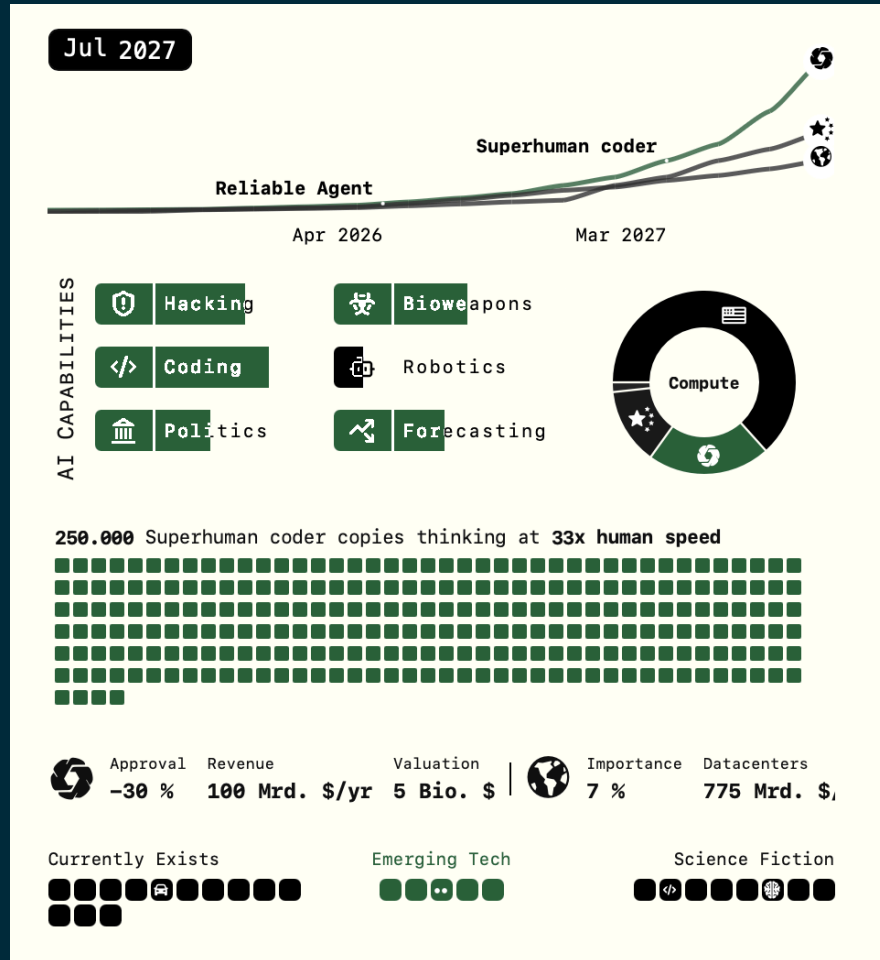
**Finance**

**Marketing**

**Legal**

→ Anpassung von Prozessen/Voraussetzungen notwendig ←

# Extrapolation des Bekannten möglich?

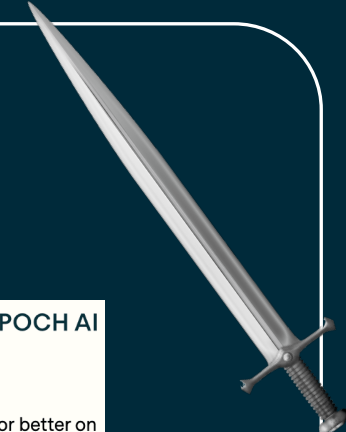


## Choose Your Ending

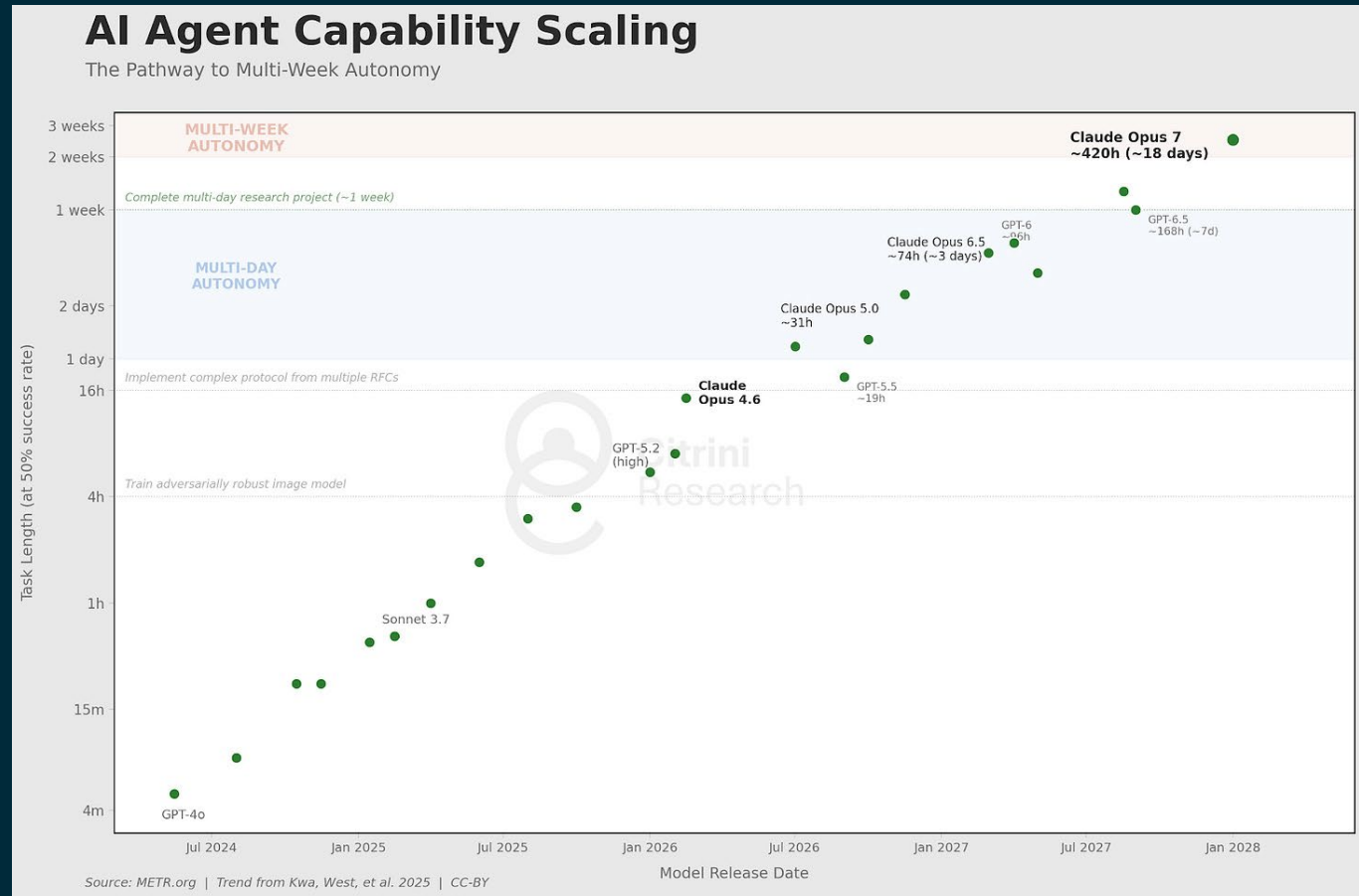
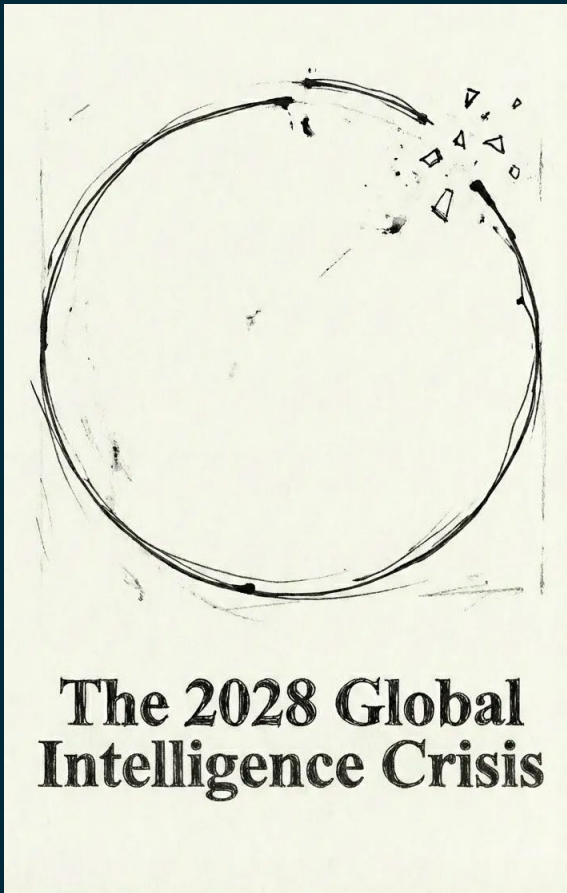
Slowdown

Race

<https://ai-2027.com/>



# Ein „alternatives“ Ende von Citrini Research



<https://www.citriniresearch.com/p/2028gic>

A person stands on a rocky outcrop in a vast, misty canyon. The scene is dark and atmospheric, with a person standing on a rocky outcrop in the foreground, looking out over a deep, misty canyon. The canyon walls are steep and dark, and the mist fills the valley, creating a sense of depth and scale. The overall mood is contemplative and somewhat somber.

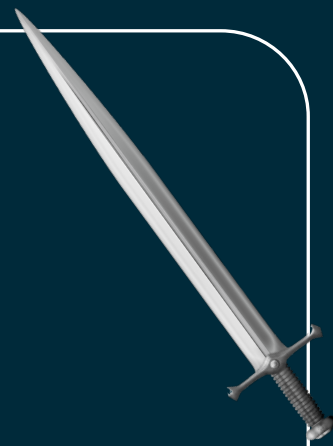
**Menschen sind gut in der Extrapolation,  
aber eher schlecht im exponentiellen Denken.**

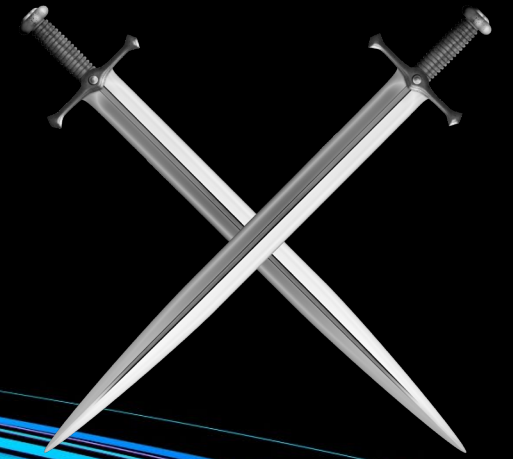
# Immer ein „Human in the Loop“?



<https://www.newyorker.com/books/under-review/how-project-maven-put-ai-into-the-kill-chain>

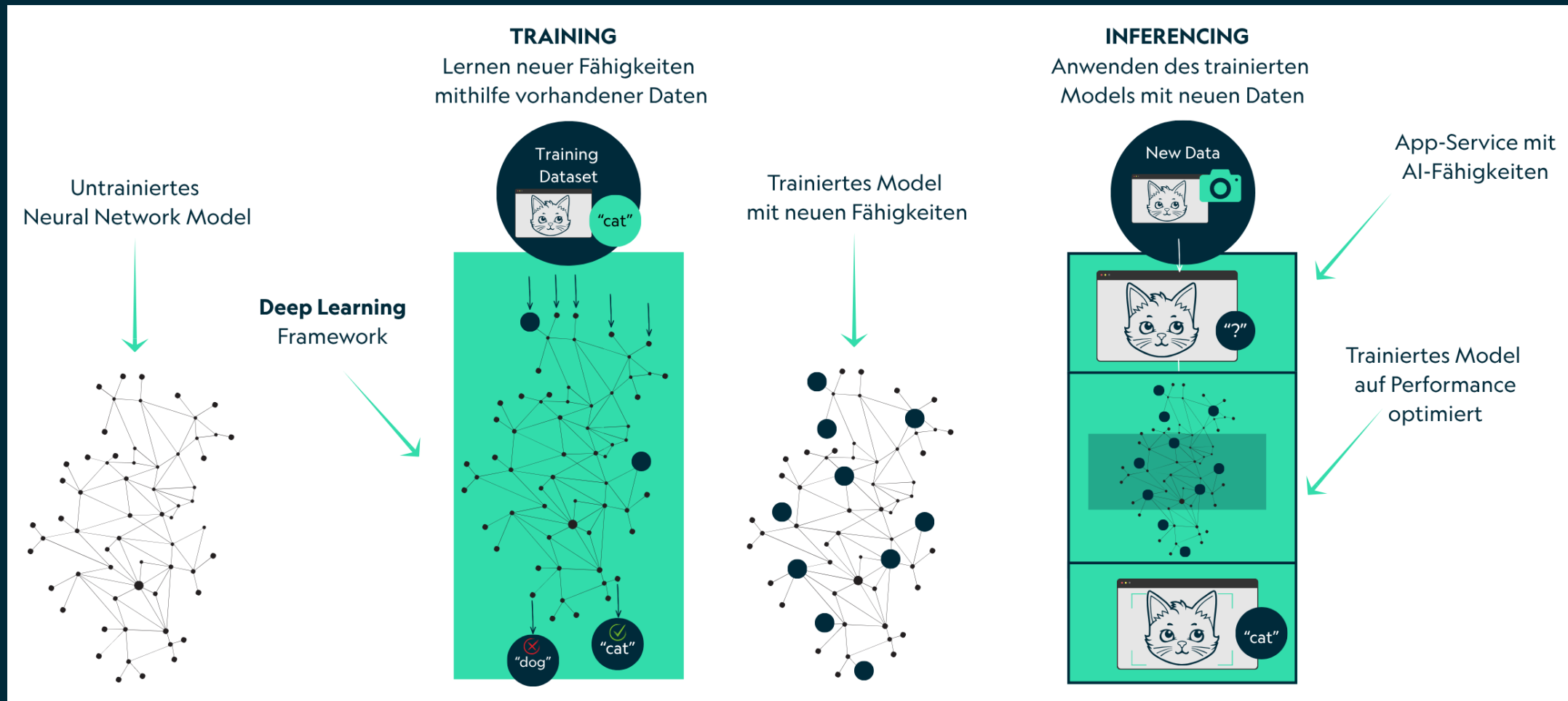
*The entire process, from target identification to target destruction, is four clicks. In 2023, one source told her that he could sign off on eighty targets in an hour: “Accept. Accept. Accept.” The old system could hit fewer than a hundred targets a day; the new system can hit a thousand, and with the recent integration of L.L.M.s that number has risen to five thousand.*





**Das Schwert ist scharf, aber anders als gedacht.**

# AI - Deep Learning Training und Inferencing

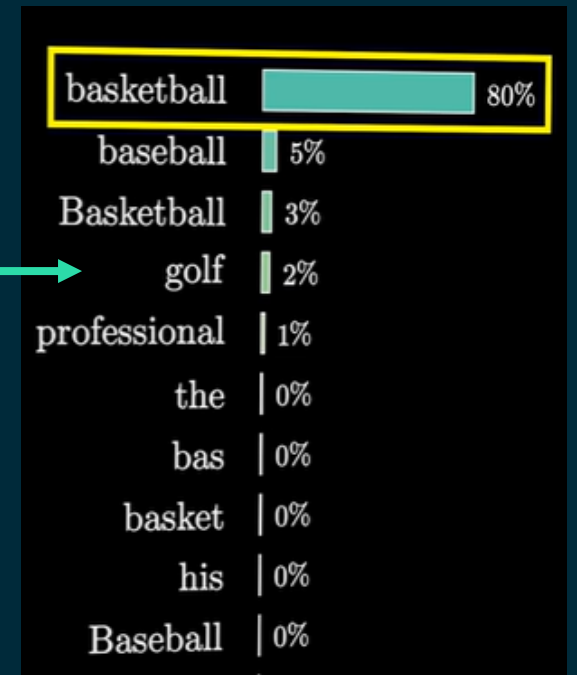


# Wo stehen die Fakten in LLMs?

Michael Jordan plays the sports of ...



Multilayer Perceptrons (MLP)





Think about what that means.



Read me

If a hostile actor develops equivalent capability and Anthropic itself acknowledges this is a question of *when*, not *if*, the entire model of “find, disclose, patch” that the security industry runs on becomes obsolete overnight.

**No CVE process moves fast enough.  
No patch management cycle closes the gap.**

<https://medium.com/@ricardomsgarces/claude-mythos-might-break-cybersecurity-but-not-in-the-way-you-think-d5c64ecbbd3b>



## CHEAT SHEET

# LLM & GenAI Security Landscape – 2026, Q2/Q3

<https://genai.owasp.org/ai-security-solutions-landscape/>



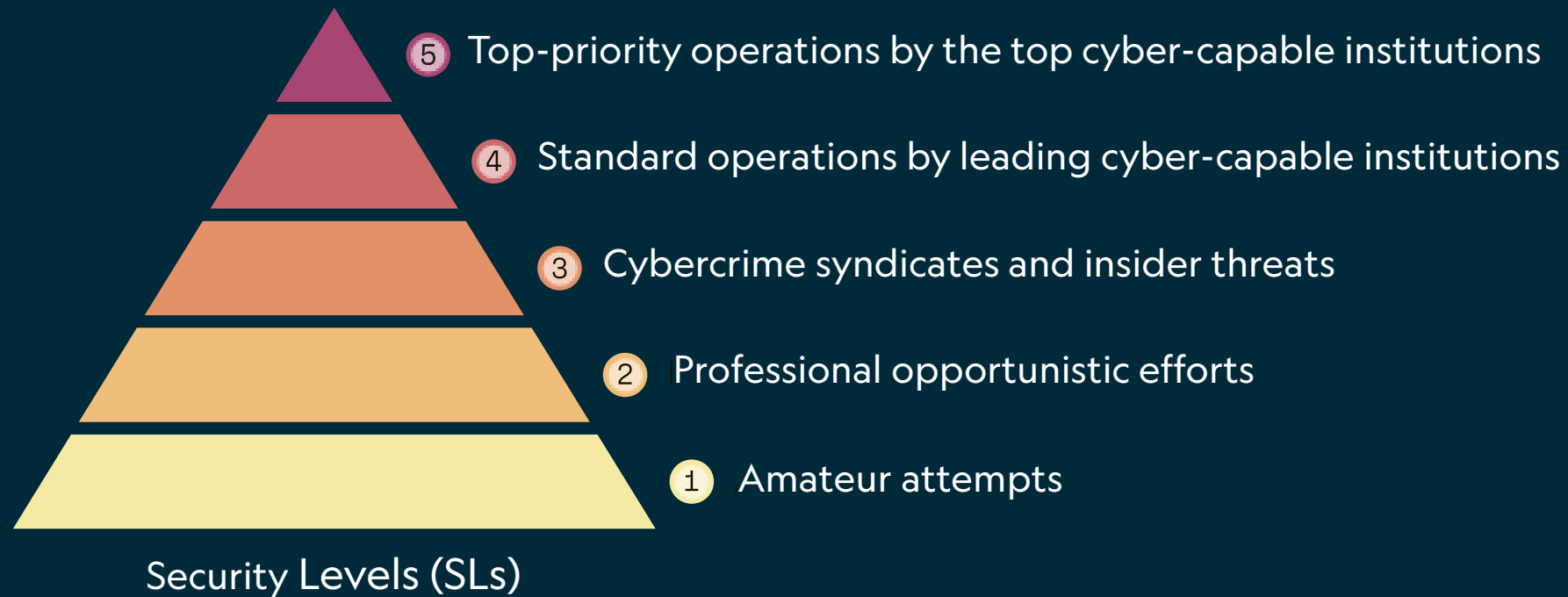
Open Source Gold Sponsors Silver Sponsors

Source; OWASP Gen AI Security Solutions Landscape Guide 2026. Q2



# Schild Nr. 0: Angriffslevel verstehen

## RAND: A Playbook for Securing AI Model Weights



[https://www.rand.org/pubs/research\\_briefs/RBA2849-1.html](https://www.rand.org/pubs/research_briefs/RBA2849-1.html)

# Schild Nr. 1: One Agent, One Job



## No general-purpose agents



### Crash tracker

Monitors crash reporting services, classifies crash patterns, flags regressions

### Analytics agent

Queries dashboards, spots anomalies, generates reports

### Telemetry analyzer

Processes app telemetry, identifies performance degradation

### Code reviewer

Scans for quality issues, suggests improvements

### Channel scanner

Watches user acquisition streams (forums, social media) for sentiment and opportunities

### PR creator

Takes findings from other agents and autonomously drafts pull requests

<https://dev.to/nesquikm/my-ai-agents-create-their-own-bug-fixes-but-none-of-them-have-credentials-2ho8>

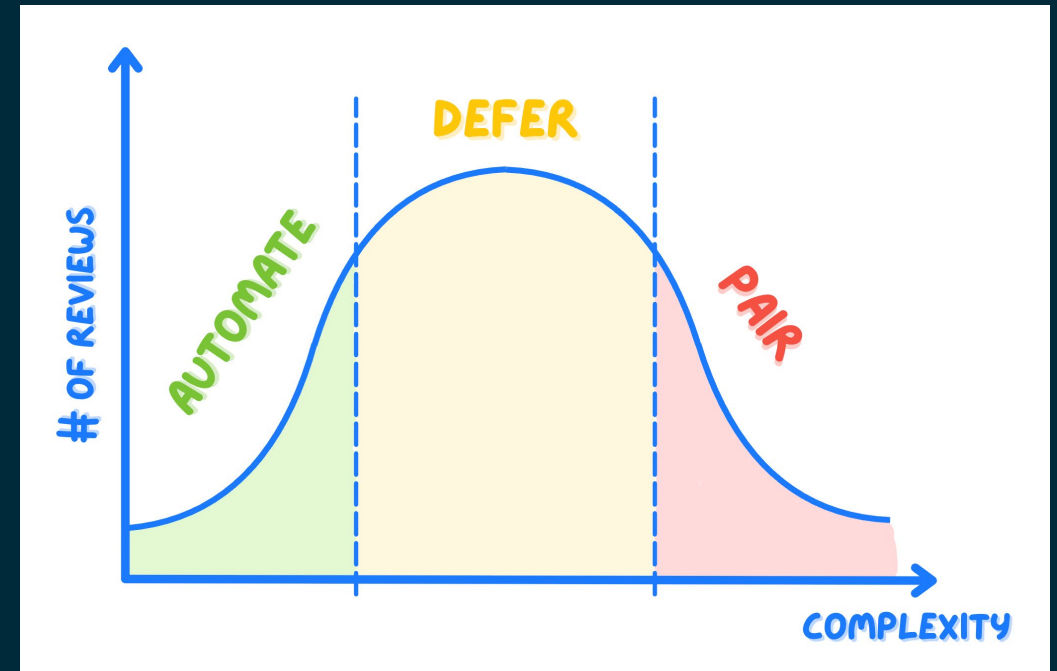
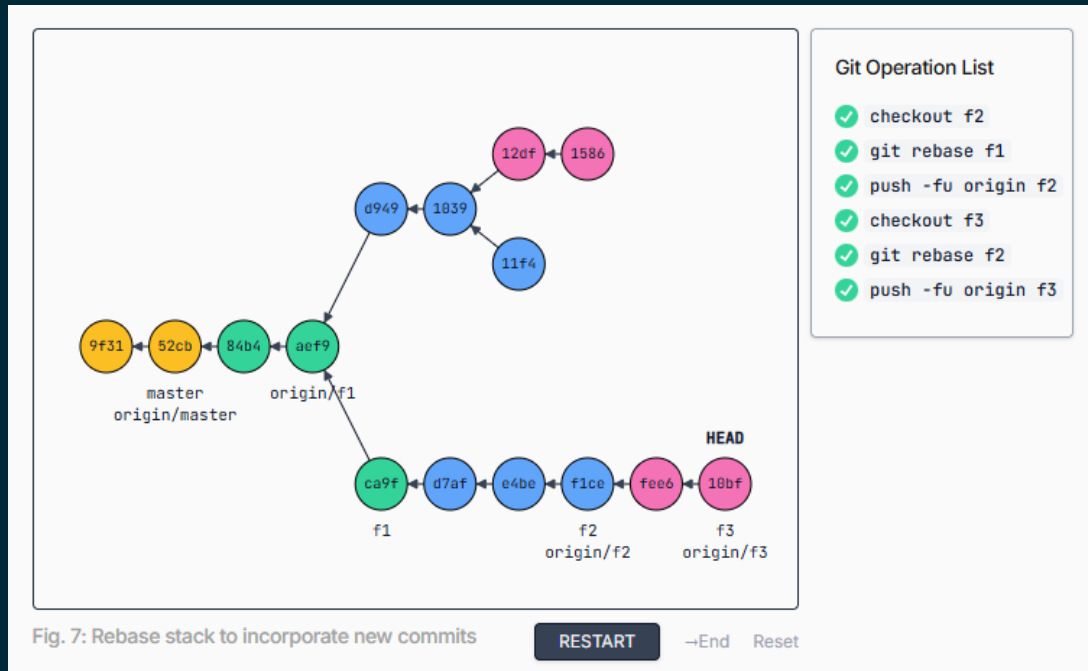
<https://dev.to/nesquikm/i-run-a-fleet-of-ai-agents-in-production-heres-the-architecture-that-keeps-them-honest-311h>



# Schild Nr. 2: Neue GIT-Strategien & AI Code Reviews

Parallele Software-Entwicklung wird steigen  
Nachvollziehbarkeit von AI Codes Changes

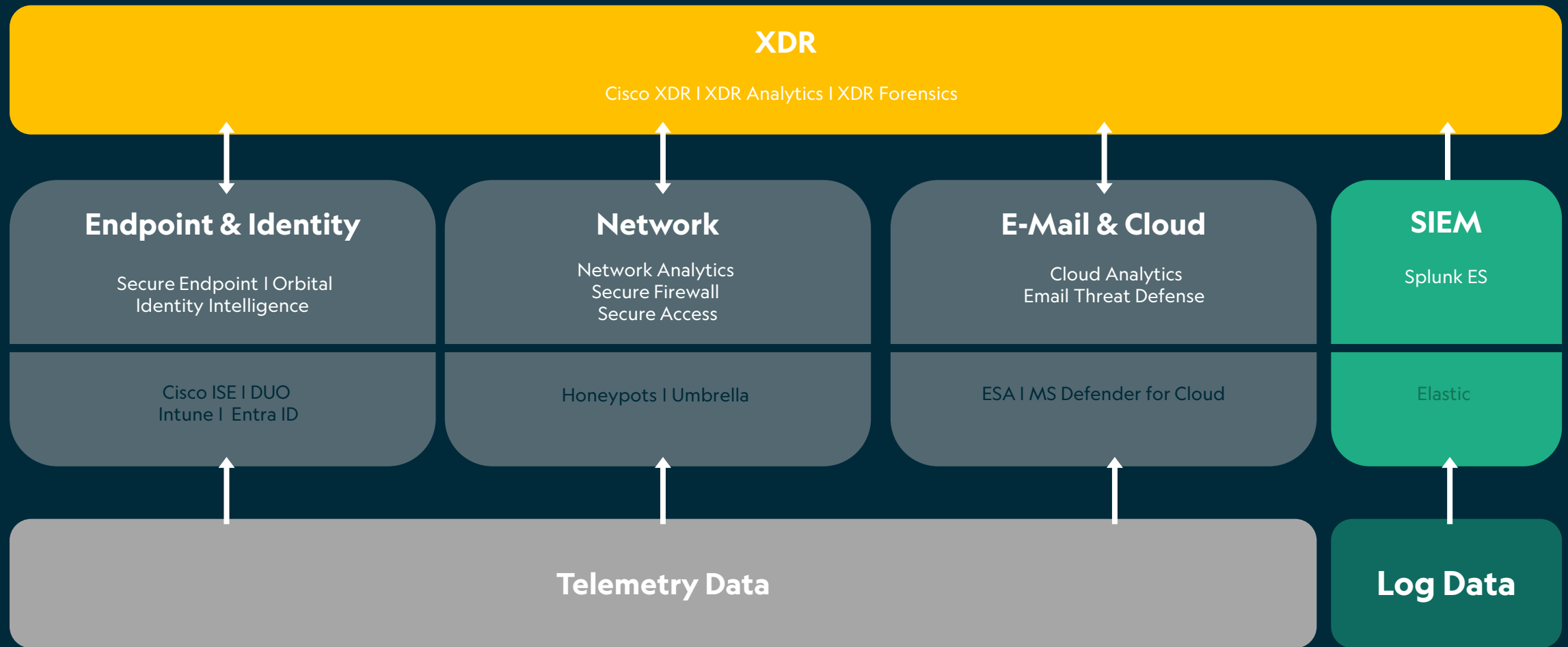
Überprüfung von Merge Requests mit AI  
AI in der Entwicklung als Klammer denken



<https://timothy.com/blog/git-stack/>

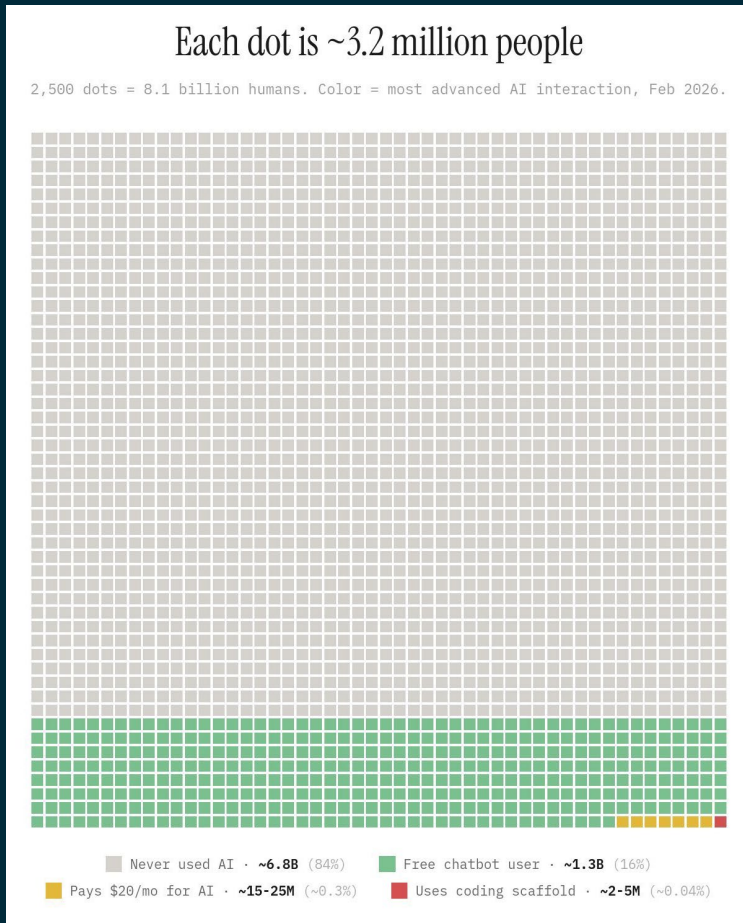
<https://refactoring.fm/p/ai-code-reviews>

# Schild Nr. 3: Security und Software-Entwicklung zusammen denken

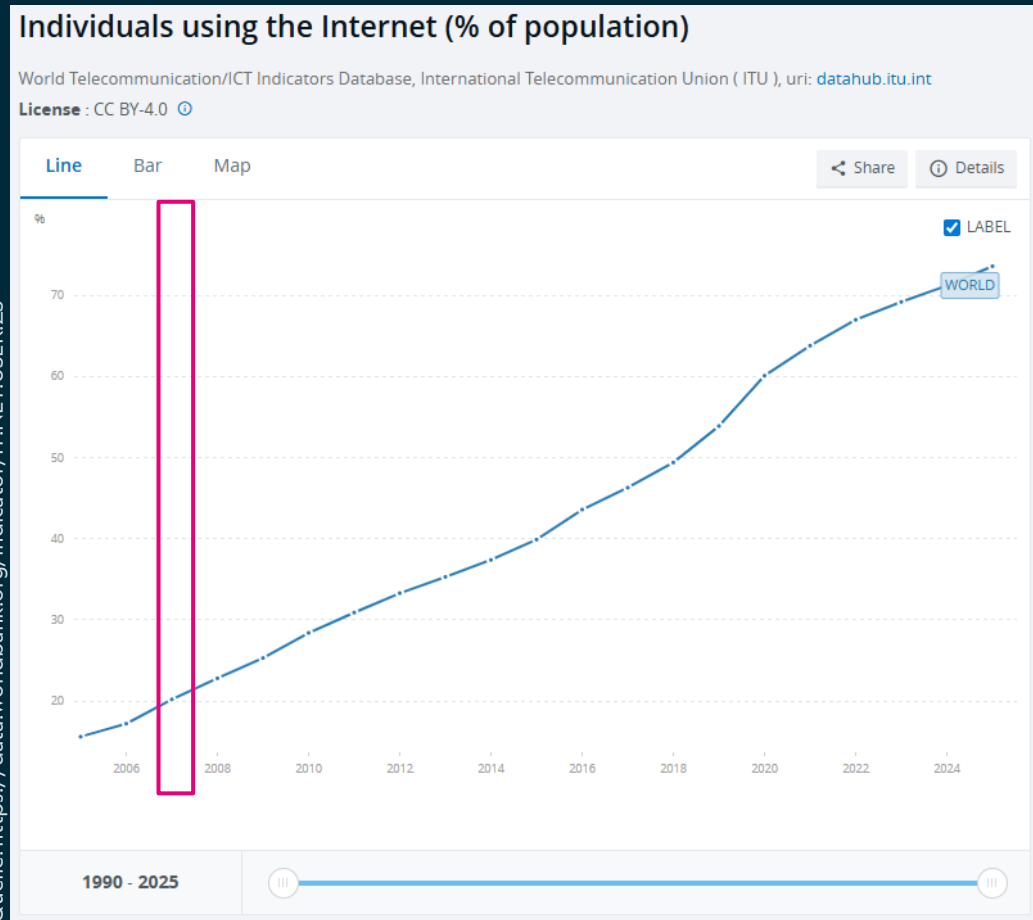


# Claude, wer? Lasst uns mal rauszoomen ...

Quelle: <https://x.com/damianplayer/status/2025234388137468387/>



Quelle: <https://data.worldbank.org/indicator/IT.NET.USER.ZS>





**Von individueller Produktivitätssteigerung ...  
in eine agentische Zukunft.**

# Sprecht uns an bei Bedarf zu diesen Themen!

> Security Advisory > Secure Infrastructure > Managed Security Services

## Software

- Observability
- Software-Entwicklung
- Open Source
- AI

## SOC

- Security Operations Center
- Incident Detection & Response

## Security

- IT-Security
- Security Advisory
- Infrastructure Security

# Paradox?



**Paradox?**

**Klug genug, um  
AI zu erfinden.**



# Paradox?

**Klug genug, um  
AI zu erfinden.**

**Dumm genug, um  
AI zu brauchen.**



# Paradox?

**Klug genug, um  
AI zu erfinden.**

**Dumm genug, um  
AI zu brauchen.**

**Und immer noch ratlos,  
ob wir das Richtige  
getan haben.**

Besucht uns an unserem Stand:

**C35**

IT Security – AI ready

> Security Advisory > Secure Infrastructure > Managed Security Services



**Sebastian Bonk**  
sbonk@avodaq.com



**Kai Irmeler**  
Teamlead Security  
Advisory & Compliance



**Fynn Thode**  
Account Manager



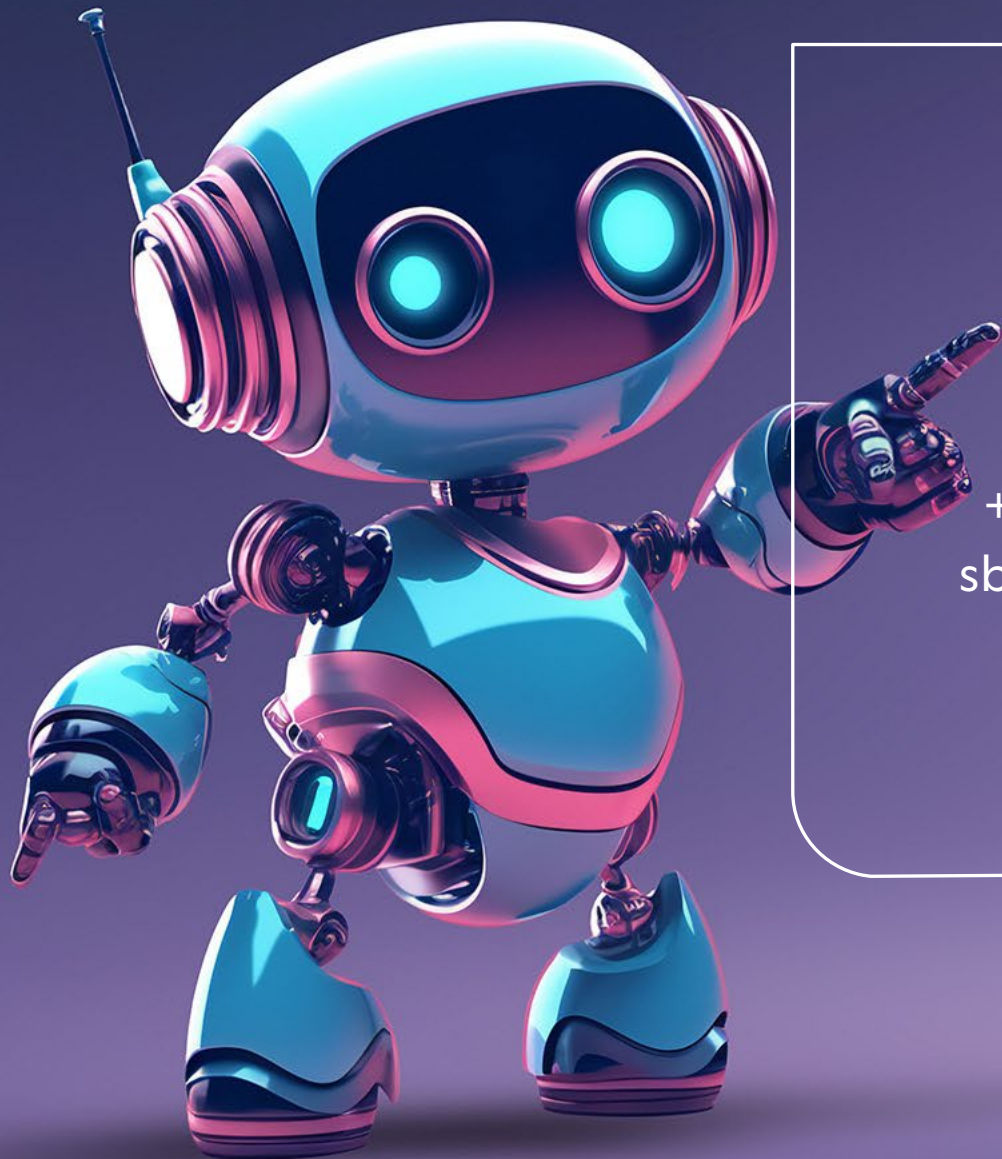
**Jörg Klein**  
Sr. Account Manager



**Malik Afrone**  
Account Manager



**Fabian-Henning Behnecke**  
Project Manager



**Danke!**

avodaq AG

**Sebastian Bonk**

+49 30 726 161 727

sbonk@avodaq.com

**avodaq.com**



Diese Präsentation sowie sämtliche darin enthaltenen Inhalte, Konzepte und Materialien sind geistiges Eigentum der avodaq AG und unterliegen den geltenden Urhebergesetzen. Die ganze oder teilweise Vervielfältigung sowie jede Weitergabe an Dritte sind nicht gestattet.

---

© avodaq AG, 2026