
Bild erzeugt mit Qwen3 30B A3B Instruct 2507
Prompt im PDF-Kommentar
KI und
Security Awareness
TU BERGAKADEMIE FREIBERG
INFORMATIONS
SICHERHEIT

Teil 1

Spicy Autocomplete

Foto von Vu Nghi Thai auf Unsplash

Foto von Vu Nghi Thai auf Unsplash

Künstliche neuronale Netze

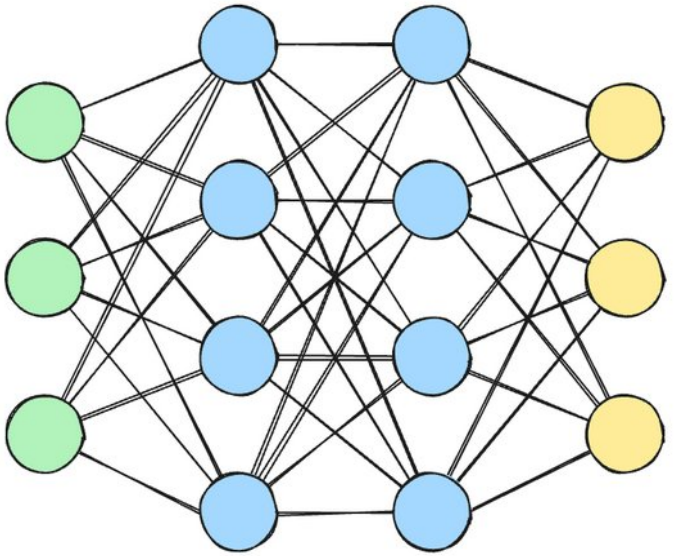


Illustration: weaviate.io

Foto von Vu Nghi Thai auf Unsplash

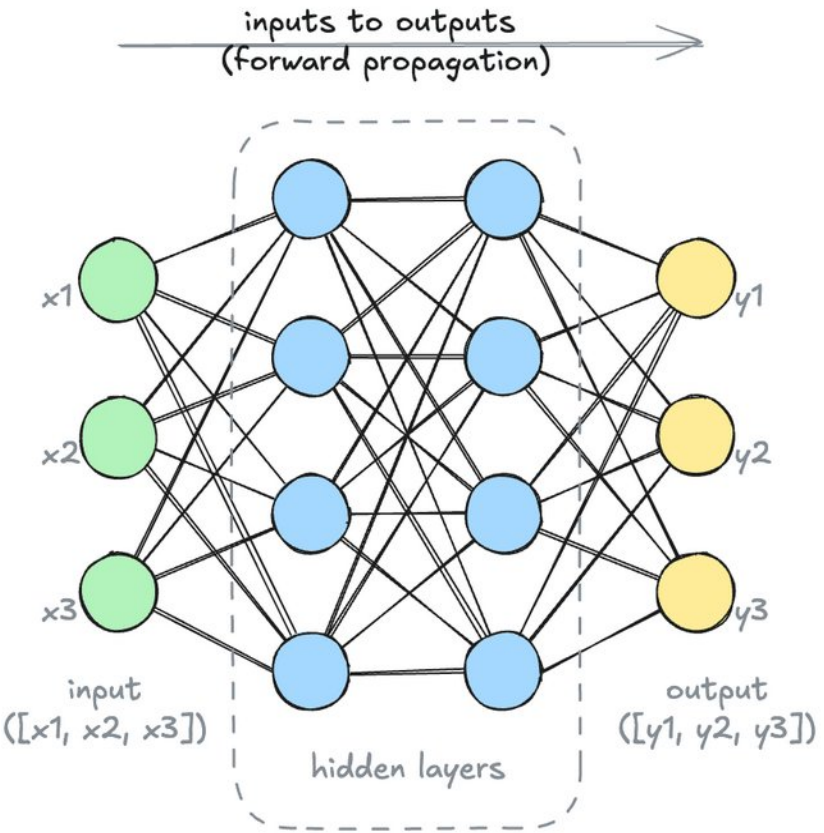


Illustration: weaviate.io

Foto von Vu Nghi Thai auf Unsplash

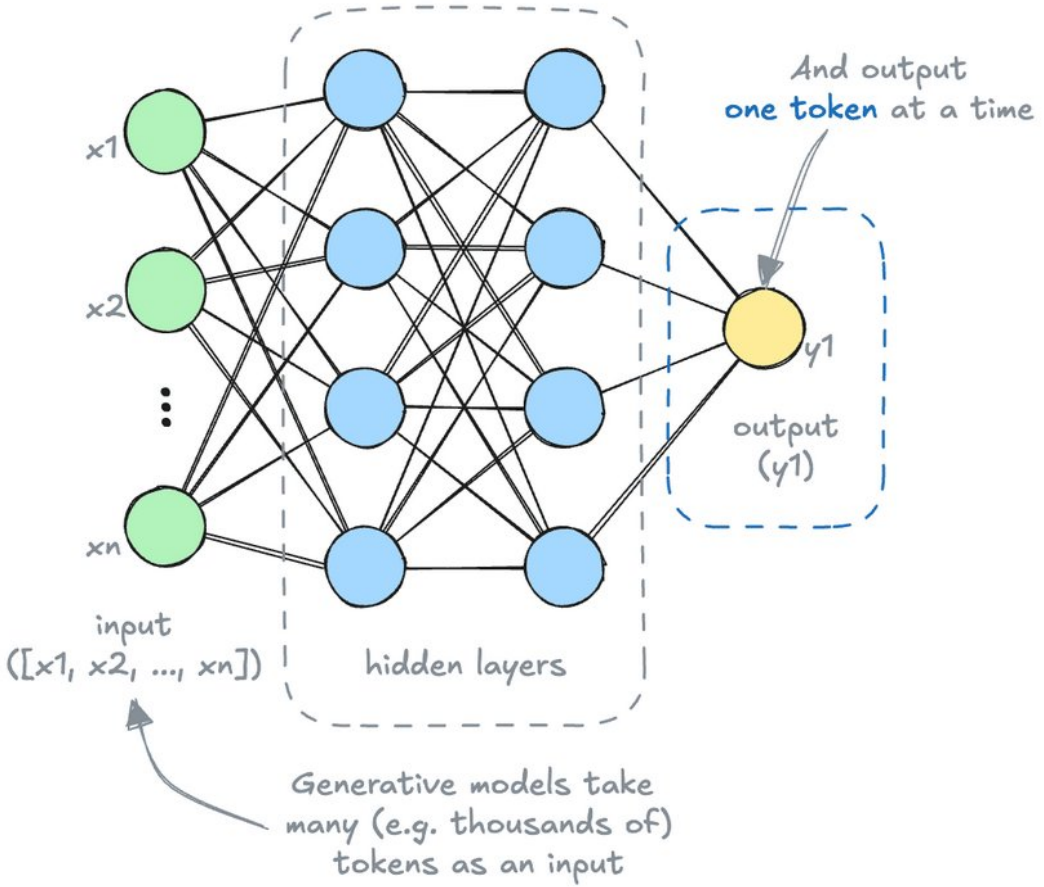


Illustration: weaviate.io

Foto von Vu Nghi Thai auf Unsplash

Intelligenz und LLMs

LLMs wandeln Input Token numerisch in Output um

Das Bild von der “smarten KI”, die gelegentlich “halluziniert” ist falsch!

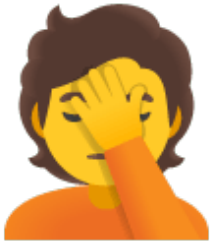
LLMs generieren **plausibel assoziierten** Text, bei dem korrekte Informationen **zufällig** auftreten

Screenshot: Wes Davis / The Verge

Foto von Vu Nghi Thai auf Unsplash

Beispiel Apple Intelligence

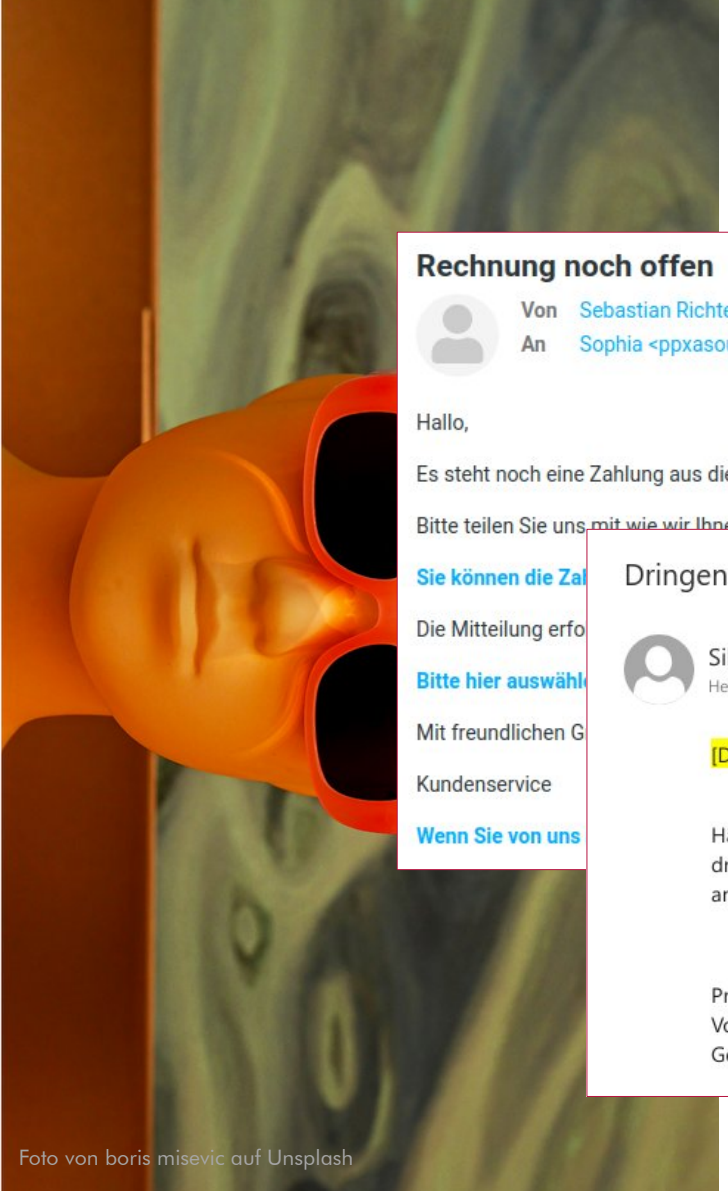
```
"{{ specialToken.chat.role.system }}You are an
assistant which helps the user respond to their
mails. Please draft a concise and natural reply
based on the provided reply snippet. Please limit
the answer within 50 words. Do not hallucinate. Do
not make up factual information. Preserve the
input mail tone.{{
```



Screenshot: Wes Davis / The Verge

Foto von boris misevic auf Unsplash

Cyberangriffe zielen oft auf Menschen



Rechnung noch offen

Von Sebastian Richter <hoangdk@vitductphuongnam.vn>
An Sophia <ppxasou@gmx.de>

Hallo,

Es steht noch eine Zahlung aus die bisher noch nicht gesendet wurde.

Bitte teilen Sie uns mit wie wir Ihnen die Zahlung senden sollen

Sie können die Zal

Die Mitteilung erfo

Bitte hier auswähl

Mit freundlichen G

Kundenservice

Wenn Sie von uns

Dringend

Silvia Rogler <officeonline258y@gmail
Heute 7:53

[Dies ist eine externe e-Mail, welche nicht v

Hätten Sie einen Moment Zeit? Ich bin gera
dringende Anfrage, die ich nur per E-Mail b
antworten Sie mir so schnell wie möglich.

Prof. Dr. Silvia Rogler
Vorsitzende
Gesendet von meinem E-Mail

IPHONE 17 PRO MAX PROMO 📱✨

Sehr geehrter Deutsche Telekom Kunde 👋,

Wir freuen uns, Ihnen eine exklusive Gelegenheit zu bieten, das neueste **iPhone 17 Pro Max** 📱🔥 zu gewinnen, das nur über Deutsche Telekom erhältlich ist.

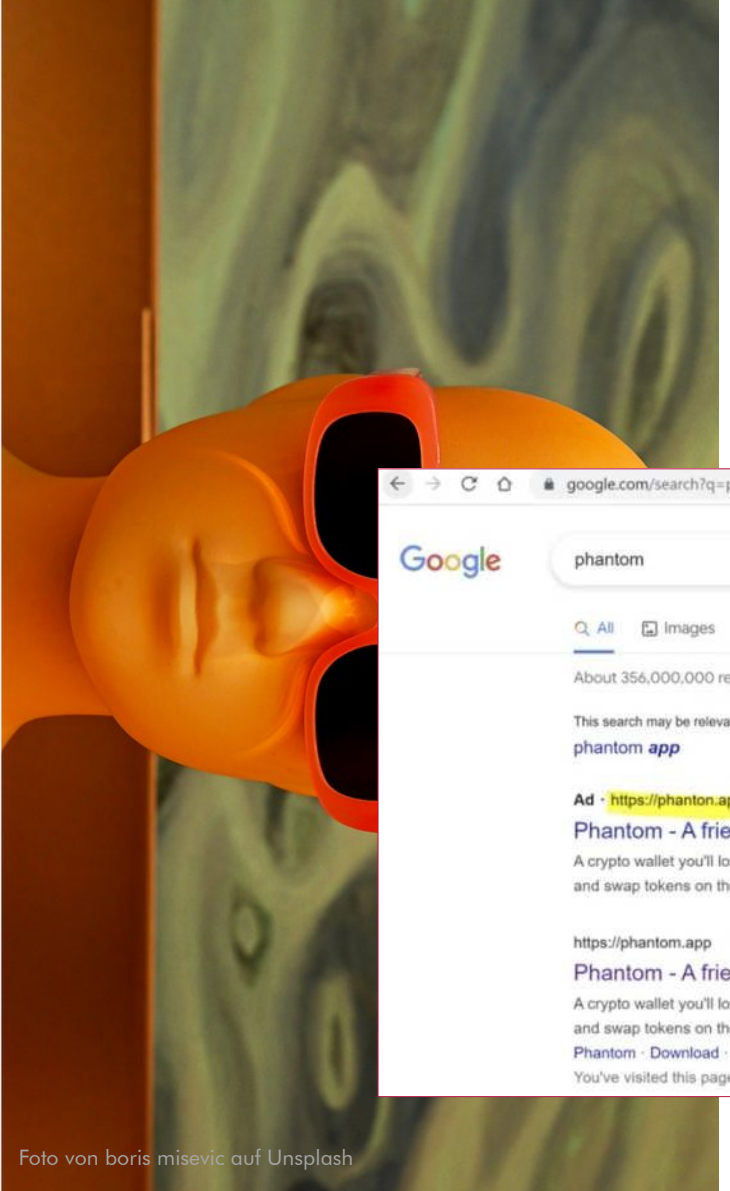
Unser Ziel ist es, unsere treuesten Kunden zu belohnen 💎 und den Wert zu betonen, den wir Ihrem Vertrauen und Ihrer mobilen Reise mit uns beimessen.

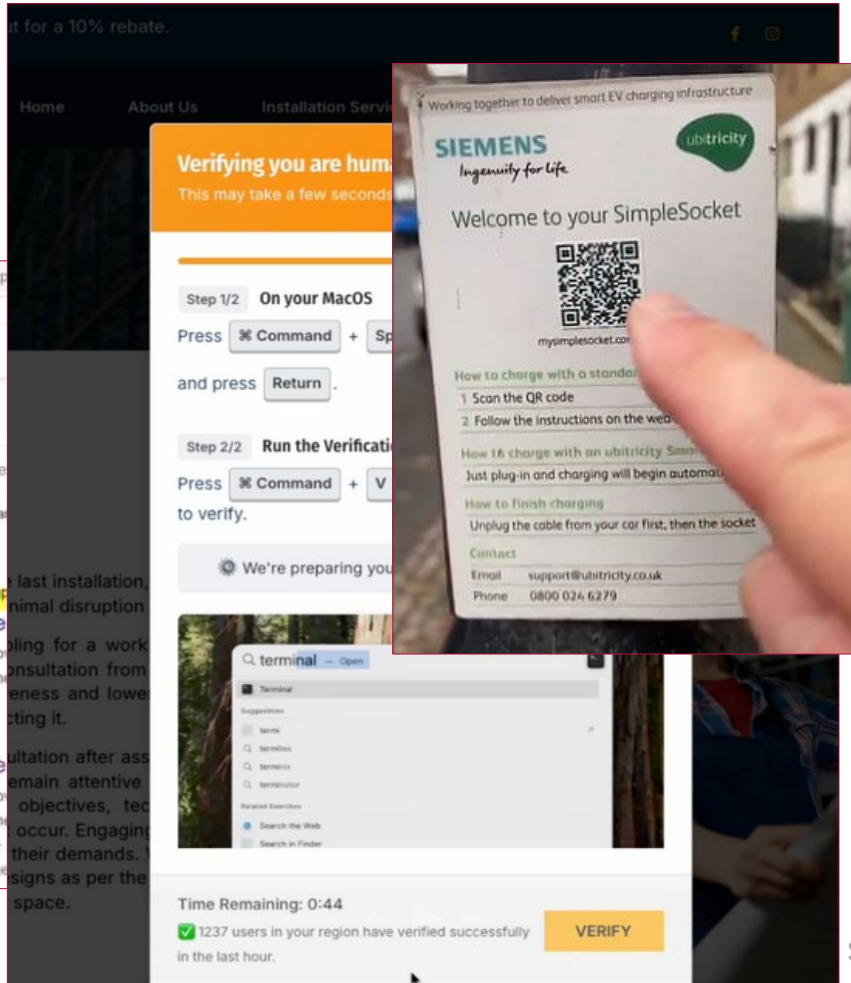
Sie wurden speziell ausgewählt ✨, um ein **iPhone 17 Pro Max** zu erhalten – mit modernster Leistung ⚡ und außergewöhnlichen Kamerafunktionen 📸. Vielen Dank, dass Sie Teil der Deutschen Telekom Familie sind.

FORDERN SIE IHR IPHONE 17 PRO MAX AN →

Foto von boris misevic auf Unsplash

Cyberangriffe zielen oft auf Menschen





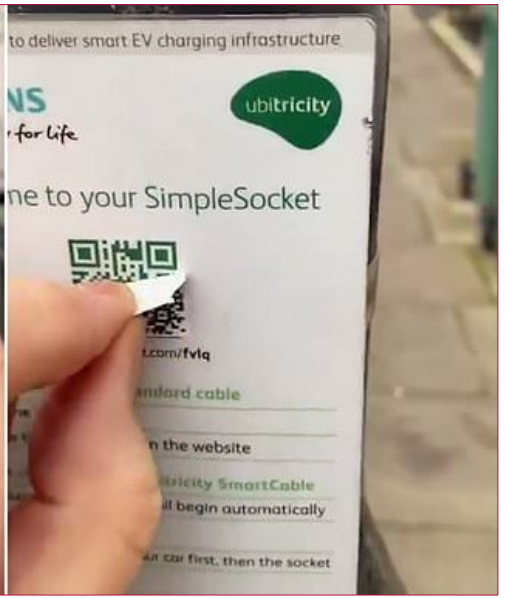


Foto von boris misevic auf Unsplash

Screenshot: Push Security

Bild: Daily Mail

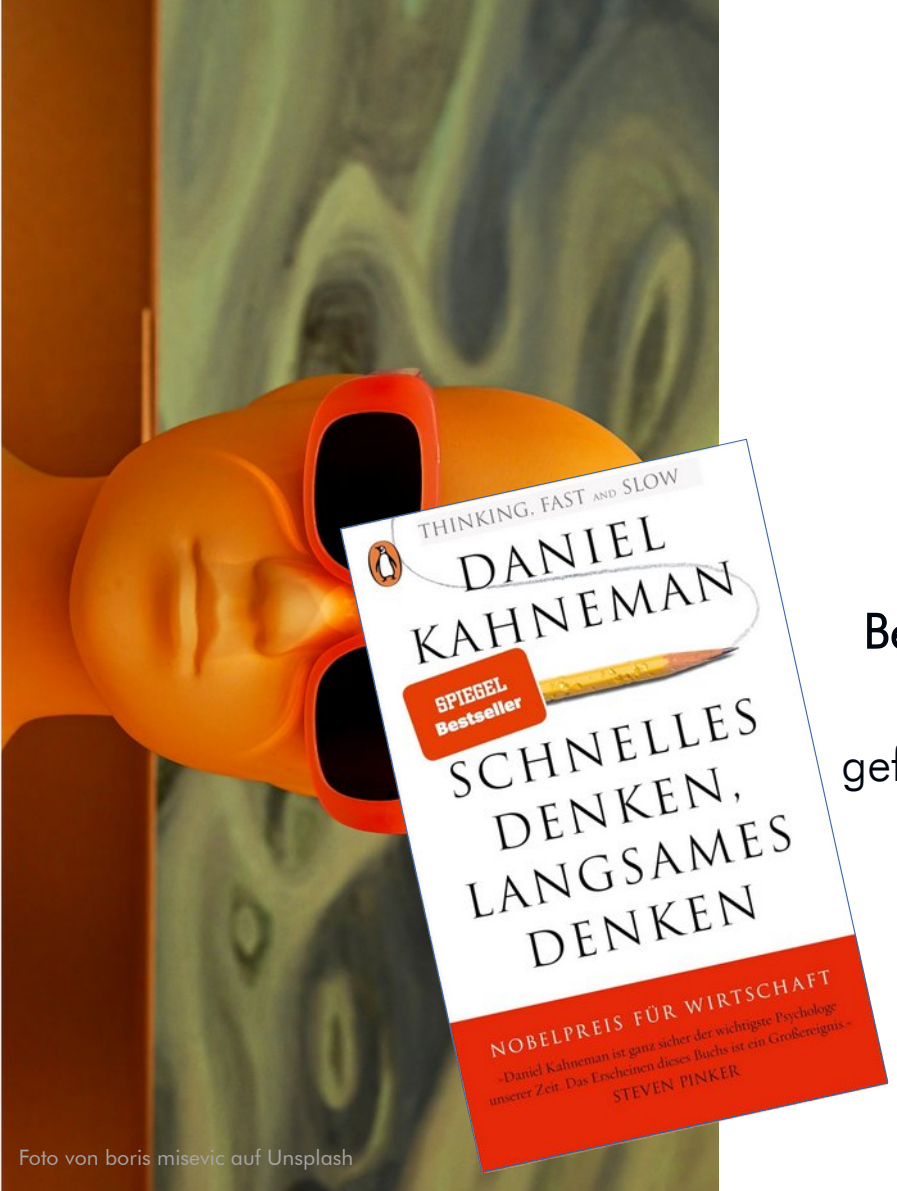


Foto von boris misevic auf Unsplash

Cyberangriffe zielen oft auf Menschen

Ziele von Security Awareness Maßnahmen

Bewusstsein schaffen

gefährliche Situationen erkennen

und in **Handeln** überführen

richtig auf gefährliche Situationen reagieren

Foto von boris misevic auf Unsplash

Das funktioniert... so "mittel"

Despite the growing reliance on artificial intelligence in cybersecurity, there is a striking gap in security awareness among both end users and organizations, leaving critical systems vulnerable to exploitation.

(Smith et al., 2023)

The integration of AI into cybersecurity systems has outpaced the development of user education, making individuals more susceptible to advanced phishing attacks and social engineering tactics.

(Jones & Lee, 2022)

As AI tools become more sophisticated, the lack of proper cybersecurity training for AI developers and users exacerbates the risk of accidental or intentional misuse, further compounding security challenges.

(Nguyen et al., 2021)

While AI can enhance threat detection, its rapid deployment in security systems often overlooks the need for ongoing security awareness, making it a double-edged sword in protecting against both human and machine-driven threats.

(Martínez & Rodriguez, 2024)

ChatGPT

Anmelden | Kostenlos registrieren

give me four quotes from academic articles showing the dire state of security awareness in the age of ai (1-2 sentences each)

format:
- quote text (author et.al., YYYY)

Here are four quotes from academic articles discussing the challenges and risks of security awareness in the age of AI:

1. *"Despite the growing reliance on artificial intelligence in cybersecurity, there is a striking gap in security awareness among both end users and organizations, leaving critical systems vulnerable to exploitation."* (Smith et al., 2023)
2. *"The integration of AI into cybersecurity systems has outpaced the development of user education, making individuals more susceptible to advanced phishing attacks and social engineering tactics."* (Jones & Lee, 2022)
3. *"As AI tools become more sophisticated, the lack of proper cybersecurity training for AI developers and users exacerbates the risk of accidental or intentional misuse, further compounding security challenges."* (Nguyen et al., 2021)
4. *"While AI can enhance threat detection, its rapid deployment in security systems often overlooks the need for ongoing security awareness, making it a double-edged sword in protecting against both human and machine-driven threats."* (Martínez & Rodriguez, 2024)

These quotes reflect the ongoing struggle to keep pace with both technological advances and the necessity for comprehensive security training.

Stelle irgendeine Frage

Anhängen | Suchen | Lernen | Bild erstellen | Audio

ChatGPT kann Fehler machen. Überprüfe wichtige Informationen. Siehe Cookie-Voreinstellungen.

Jedes Zitat ist frei erfunden

...hittel"

...ersecurity
...d use...s
...itation.
...et al., 2023)

...ty systems has outpaced
...making individuals more
...acks and social

(Jones & Lee, 2022)

...cybersecurity
...f accidental or
...es.
...uyen et al., 2021)

...loyment in
...g security
...tecting

...uez, 2024)

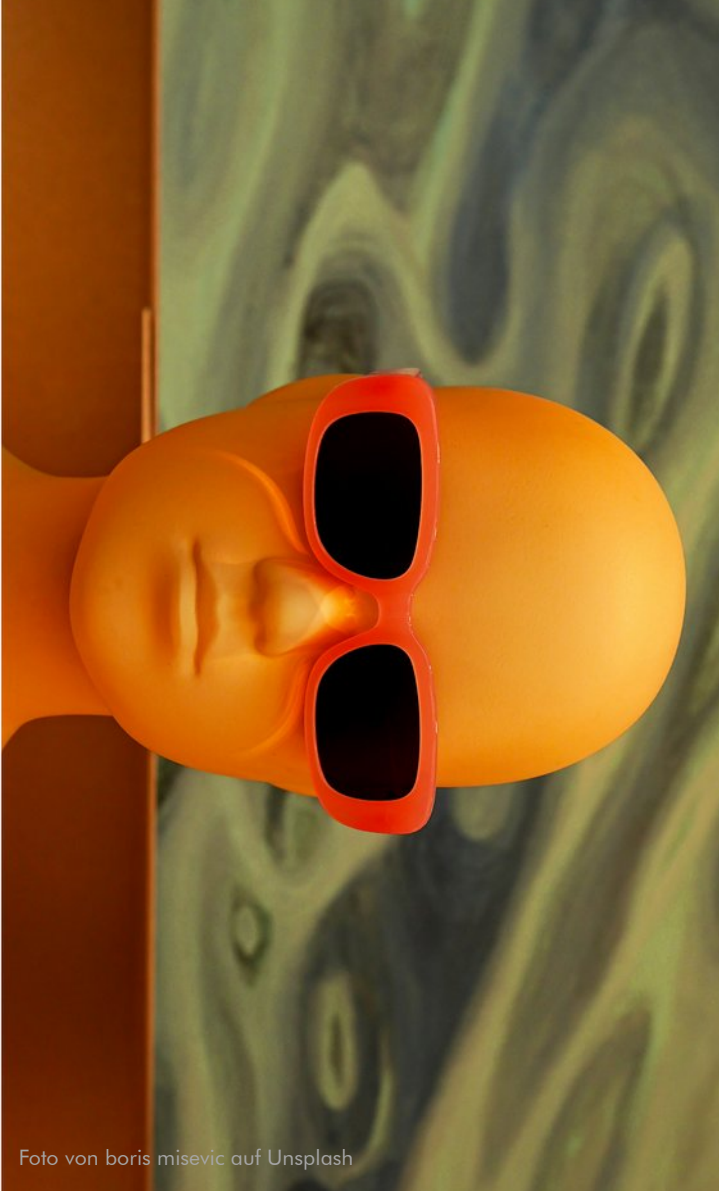


Foto von boris misevic auf Unsplash

Das funktioniert... so "mittel"

With respect to the performance over time, we found that after six and eight months participants' skills were no longer significantly better than before participating in the tutorial.

(Reinheimer et al., 2020)

Participants are "tired," "turned off," and "overwhelmed" by it all. This overarching fatigue factors into their cost-benefit analysis, and the result is that many reject security advice or practices that they realize might protect them more [...]

(Stanton et al., 2016)

Taken together [...] our results offer a sobering picture on the efficacy of existing phishing training. As currently designed and deployed in practice, training is unlikely to offer significant value relative to its considerable expense in time and effort.

(Ho et al., 2025)

[...] we found that the combination of simulated phishing exercises and voluntary embedded training [...] not only failed to improve employee's phishing resilience, but it actually even the made employees more susceptible to phishing.

(Lain et al., 2022)

Teil 3

Düstere Zeiten?

Foto von Greg Johnson auf Unsplash

Foto von Greg Johnson auf Unsplash

Generative KI stellt Awareness auf den Kopf :/

alles wie immer…

- Phishing Mails sind ein alter Hut

  Jetzt aber noch einfacher zu erzeugen und mit perfekter Grammatik und Rechtschreibung?

- Webseiten und Anhänge mit Schadcode

  Automatisch und passgenau erzeugt?

- Vortäuschung falscher Identitäten

  Jetzt mit hochrealistischen Deepfakes (Telefon, Video, …)

nur schlimmer?

Foto von Greg Johnson auf Unsplash

Generative KI stellt Awareness auf den Kopf?

[...] the only modality we identified offering significant improvements in outcomes was interactive training taken to completion.

(Ho et al., 2025)

We find a significant interaction between the type of administered feedback type and the amount of reported emails.

(Lain et al., 2022)

28% of our respondents learned to practice a behavior due to a negative experience or a story told about a negative experience.

(Redmiles et al., 2016)

Foto von Greg Johnson auf Unsplash

Was sie heute schon kaufen können

(und vielleicht in wenigen Jahren funktioniert)

- Automatische Erzeugung von auf konkrete Anlässe und Organisationen zugeschnittenes Schulungsmaterial
- Integrierte “Phishing-Melde-Buttons”, inklusive KI-generiertem, zielgerichtetem Feedback (Prompt Injection?)
- Klassifizierung von Phishing-Mails mit KI (aber unklar wie gut für neuartige Maschen, machine learning für Spamerkennung ist alte Technologie: bayes filter)

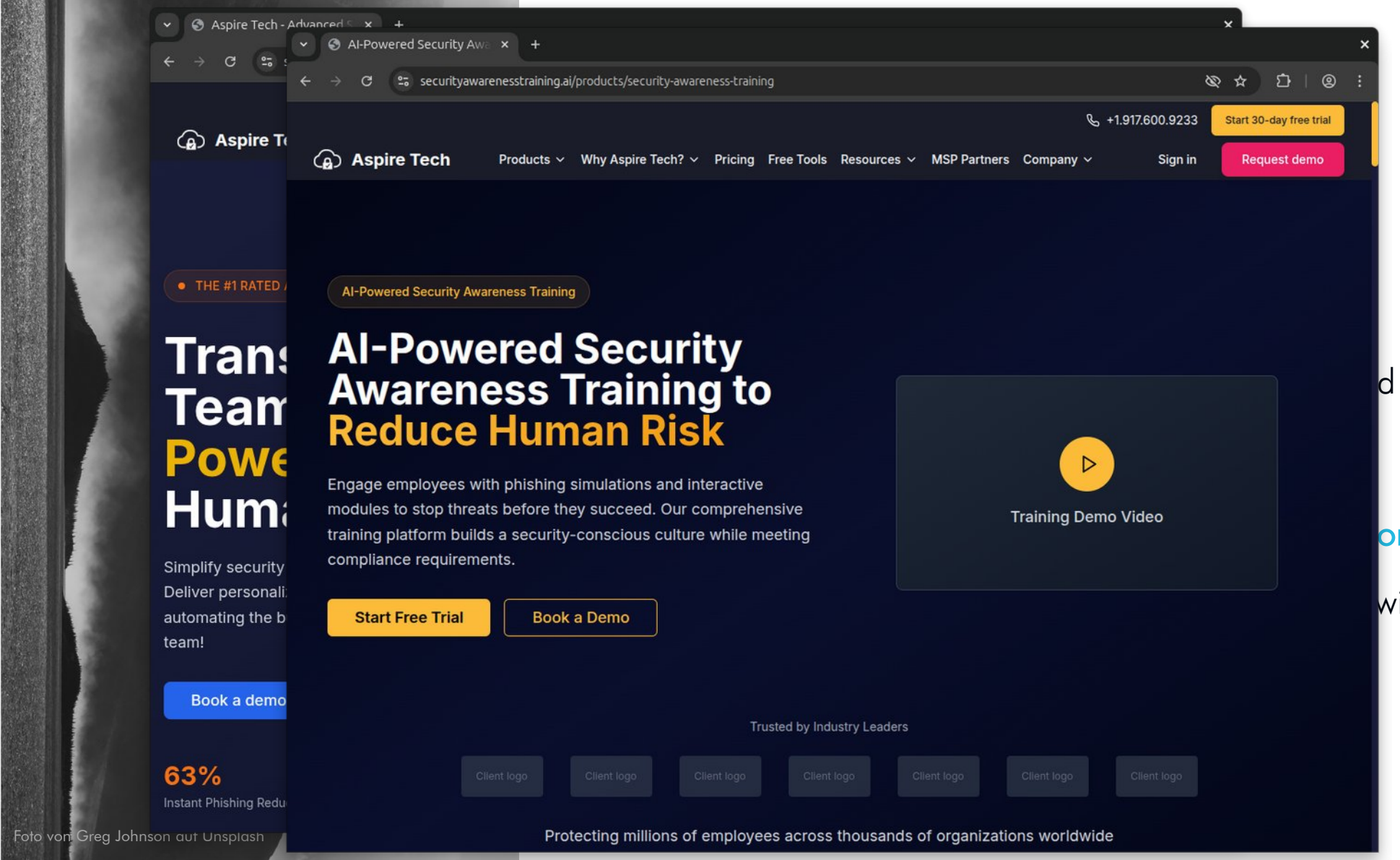


Foto von Greg Johnson auf Unsplash

d

on?)

wie

Foto von Greg Johnson auf Unsplash

Alles wie immer

- Entscheidend ist der Moment der Skepsis
- An echten Mails und Webseiten üben
  - Wo steht Absender(-adresse) / Domain
  - An welche Mailadresse antworte ich
  - Wie bestimme ich (echtes) Linkziel
- Bei unklaren E-Mails beim Absender über bekannten Kontaktweg nachfragen (Telefon, bekannte Mailadresse, Sekretariat, …)
- Mit Kollegen sprechen
- Im Zweifel: informationssicherheit@tu-freiberg.de

INFORMATIONS
SICHERHEIT

informationssicherheit@tu-freiberg.de

https://tu-freiberg.de/informationssicherheit