18. MÄRZ 2026 / ANNE KATRIN OBERDORF

ki x desinformation

VORTRAG

SWR Aktuell
8 Std. ·

Im Video auf Facebook wirbt eine KI-generierte Figur für finanzielle Gewinne: https://1.ard.de/deepfake-facebook. Das Unternehmen Würth warnt aber: Das Video sei fake, ... Mehr anzeigen

zdfheute

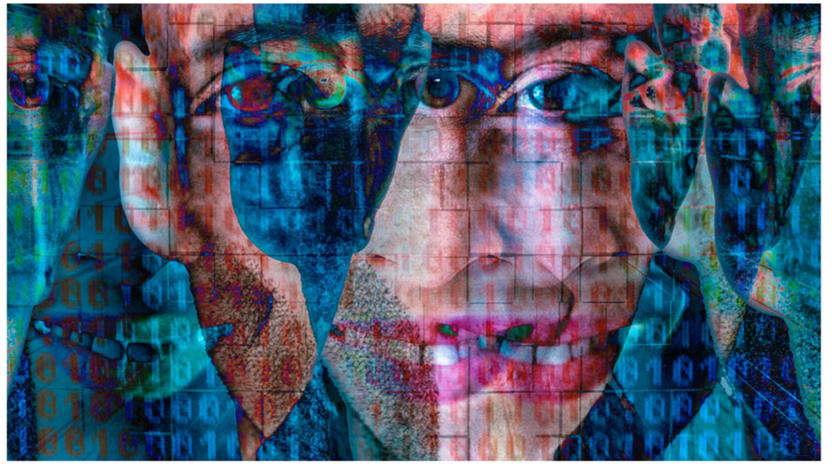
Interview

Expertin zu Internetkriminalität

Wie Deepfakes neue Dimensionen des Online-Betrugs eröffnen

15.11.2025 | 08:22

Künstliche Intelligenz und Deepfakes eröffnen Kriminellen neue Dimensionen des Online-Betrugs. Die Forscherin Julia Krickl erklärt, wie Betrüger vorgehen - und was uns schützt.

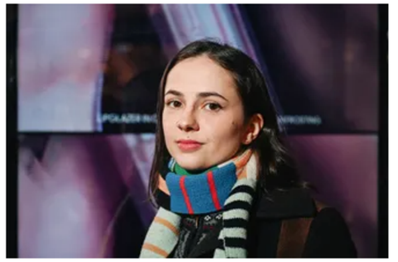


Z+ Deepfakes

Ihr Gesicht, aber der Rest ist Deepfake

Wegen Elon Musks KI Grok kann wirklich jeder Nacktbilder fälschen. Die Gefahr ist nicht neu, doch ein Gesetz dagegen fehlt. Den Schaden haben Frauen wie Theresia Crone.

15. Januar 2026



Was ist das eigentlich?

Über was reden wir hier?

Desinformation

Falsche Informationen, die absichtlich verbreitet werden, um Schaden anzurichten oder die falsch kontextualisiert werden, um Schaden zu verursachen

Misinformation

Falsche Informationen, die unabsichtlich geteilt werden.

! “Fake News” wird immer mehr zur Diffamierung politischer Gegner verwendet (vgl. “Lügenpresse”), daher nutzen Fachleute präferiert “Desinformation”.

[Q14, Q16]

Was macht Desinformationen so gefährlich?

- Soziale Medien erlauben schnellere Verbreitung
- Digitale Plattformen und künstliche Intelligenz sorgen für leichtere Erstellung
- In privaten Chatgruppen bleibt öffentliche Diskussion unterdrückt
- Urheber bleiben anonym
- Bots bringen Desinformationen automatisiert in Umlauf
- Autoritäre Staaten wie Russland nutzen Desinformation gezielt als strategisches Instrument

> Desinformationen untergraben die Meinungsbildung von Bürgerinnen und Bürgern zu gesellschaftlich relevanten Themen, also eine zentrale Säule unserer Demokratie.
>
> **Dr. Jan-Hinrik Schmidt, Leibniz-Institut für Medienforschung**

[Q14, Q16, Q17]

FIMI

= Ausländische Informationsmanipulation und Einmischung

„Die SDA verfolgt bei ihrem Vorgehen sehr konkrete Ziele, die sie in sogenannten KPIs (Abkürzung für Key Performance Indicator) festhält. KPI nennt man in der Wirtschaft Kennzahlen, die als erste Erfolgsziele gelten. Die SDA hat sich KPIs für die Länder gegeben, die sie mit ihrem Infokrieg ins Visier nimmt. Dazu gehören Deutschland, Frankreich, Israel und die Ukraine.

Für Deutschland formuliert die SDA in einer Präsentation das Ziel: "Erhöhung des Stimmenanteils der AfD auf 20 Prozent" - gemessen werden sollte das Ziel am Abschneiden der Partei in bundesweiten Wahlumfragen. Die Partei hatte damals noch deutlich weniger Prozente, befand sich jedoch bereits im Aufwärtstrend und überwand im Juli 2023 erstmals die 20-Prozent-Marke.“

Tagesschauartikel ***„Tiefe Einblicke in Putins Lügenmaschine“,*** Stand: 16.09.2024
https://www.tagesschau.de/investigativ/ndr-wdr/russland-propaganda-fakenews-sda-deutschland-100.html

[Q17]

Was hat das mit KI zu tun?

DEMO

Deepfakes

sind Bild-, Ton- oder Videoaufnahmen, die echt wirken, jedoch mit KI geziehlt manipuliert wurden, um Menschen zu täuschen

Beispiel:

- Desinformation: die Russen marschieren in Paris ein
- Pornografische Deepfakes zur Erpressung von Politikern
- Gefälschte Produkte und Onlineshops

Viele kommerzielle LLMs sind unterliegen ethischen Guidelines und Sicherheitsfiltern. Diese verhindern z.b. das erstellen von Pornografischen Inhalten.

Jailbraking: Profis können mit schlauen Prompts und Techniken diese Sicherheitsmechanismen umgehen.

[Q7, Q8]

Problem

Viele Menschen denken fälschlicherweise, dass sie KI-generierte Inhalte identifizieren können:

40-50% der LLM werden selbst dann für einen Menschen gehalten, selbst wenn die Nutzer aktiv versuchen herauszufinden, ob es KI ist oder nicht.

51% haben in einer Studie in den UK synthetisch generierte Medien erkannt.

Q7, Q9

Können wir KI entdecken?

Echt oder KI?

https://www.zeit.de/digital/internet/2025-07/kuenstliche-intelligenz-video-faelschung-unterschiede-test

Künstliche Intelligenz: Erkennen Sie noch, was echt ist?

KI-Videos werden in beängstigendem Tempo besser. Können Sie Ihren Augen noch trauen? Testen Sie Ihren Deepfake-Detektor in unserem Quiz.

Z DIE ZEIT / Jul 20, 2025

Verschiedene Arten von Deepfakes

Synthese

Frei erfunden ohne “Basisbild”.

⊕ Lizenzfreie Bilder für Webseiten oder Werbung

⊖ Desinformation

Replacement

Man ersetzt das Gesicht mit einem anderen Gesicht.

⊕ Auf Fotos anonym bleiben anstatt Bilder oder Personen zu verpixeln.

⊖ Verunglimpfung durch z.B. Revenge Porn

Editing

Verändern oder Weglassen von Attributen.

⊕ Man kann neue Frisuren ausprobieren oder Brillen

⊖ Man kann jemanden gesünder aussehen lassen oder ausziehen

Reanactment

Man folgt der Bewegung eines realen Menschens.

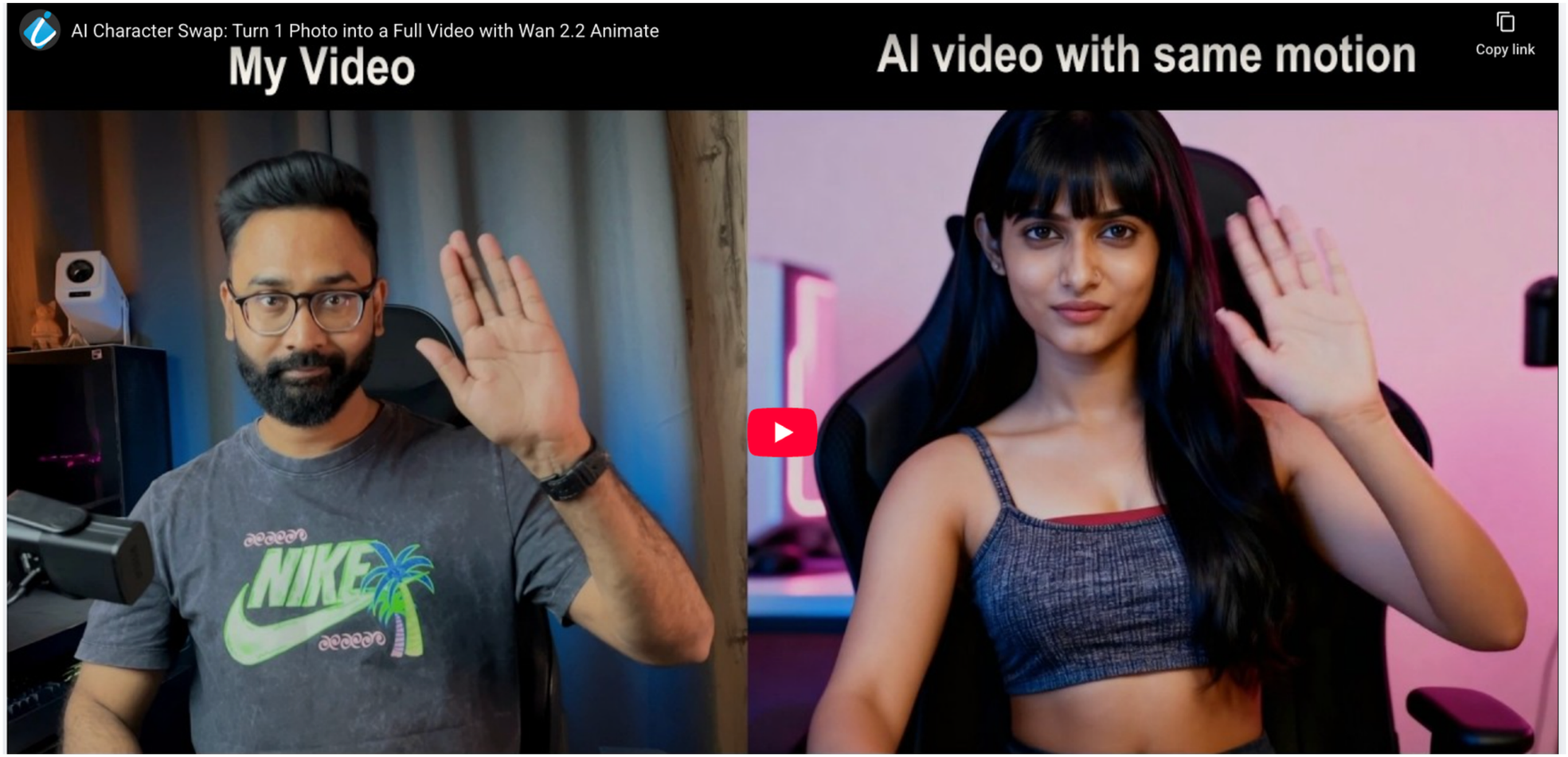
⊕ Simultanübersetzungen oder Synchronisation mit anpassen der Mundbewegungen

⊖ Identitätsdiebstahl

[Q12]

Beispiel für Reanactment Deepfakes:

https://www.youtube.com/watch?v=ZgxeqTusq8k



Wie erkenne ich ein KI generiertes Video oder Bild?

KI in Bild und Video erkennen

- **gar nicht mehr :(**

Was trotzdem (machmal) noch hilft:

- “zu **perfekt**” - z.B. perfektes Aussehen in einer Krisensituation
- **Metadaten** z.B. wo wurde das Bild mit welchem Gerät aufgenommen, gibt es ein Wasserzeichen
- Unterschiedliche Schatten und Linien, die nicht passen
- **Gesichtsausdrücke** passen oft nicht zum Inhalt
- **Aktiv** konsumieren, vorallem bei skandalösen Bildern/Videos

[Q10]

KI in Texten erkennen

- Immer das gleiche **Schema**: Einleitung, Liste mit Punkten, Zusammenfassung, Warnung.
- **Perfektion**: Keine Tippfehler, keine Stilbrüche, keine persönlichen Einschübe. Das wirkt steril.
- **Redundanz**: KI wiederholt sich, produziert mehr Text als nötig, sagt dasselbe mit anderen Worten.
- **Englische Spuren**: LLMS »denken« auf Englisch und übersetzen danach
- **Inhaltliche Fehler**: KI erfindet Fakten, verwechselt Zusammenhänge, halluziniert Quellen.

[Q10]

KI in Ton erkennen

- **Sound** oft (noch) blechern
- Nutzt keine **personen-typischen** Formulierung
- **Kontext** passt nicht: Anruf von fremder Nummer → rufe auf einer dir bekannten Nummer zurück
- Bei engen Familienmitgliedern z.B. **Safeword** ausmachen oder Sachen fragen, die nur die echte Person kennen kann

[Q10]

Diskussion:
Wie verändert KI unsere Haltung gegen über den Medien und Medienschaffenden?

Fazit

Was habt ihr gelernt?

Was können wir tun?

1. **Hinterfrage die Nachricht:**
    a. Von wem kommt die Info?
    b. Welche Absicht steht wohlmöglich dahinter?
2. **Überprüfe die Quelle:**
    a. Ist die Quelle **seriös** z.B. hat die Website ein Impressum?
    b. Berichten **mehrere** unabhängige Medien von dem gleichen Ereignis und zeigen ähnliche Aufnahmen?
3. **Kritische Haltung einnehmen:**
    a. Kann das, was ich sehe, überhaupt stimmen?
    b. Nutze die **Bilderrückwärzssuche** um in z.B. Google Lens, da oft bereits existierende Fotos als z.B. Hintergrund verwendet werden.
    c. Schaue nach, ob die Aufnahme schon in einem **Faktenchecker** überprüft wurde
    d. Optional: **KI Erkennungstoolsempfehlung**

[Q5, Q6, Q8, Q12, Q15]

Faktencheckerempfehlung

https://correctiv.org/faktencheck
https://www.dpa.com/de/faktencheck
https://www.mimikama.org/
https://www.br.de/nachrichten/faktenfuchs-faktencheck,QzSIzl3
https://www.volksverpetzer.de/
https://faktencheck.afp.com/list
https://euvsdisinfo.eu/disinformation-cases/

KI-basierte KI-Detektoren

Übersicht: https://www.heise.de/tipps-tricks/KI-generierte-Texte-erkennen-so-klappt-s-10244739.html

Gretchen AI aus Deutschland: https://www.dfki.de/en/web/news/nun-sag-wie-hast-dus-mit-der-wahrheit-gretchen-ai-revolutioniert-verifikation-von-digitaler-information
Frankreich: https://sightengine.com/detect-ai-generated-images
Serbia: https://wasitai.com/
USA: https://www.zerogpt.com/ai-image-detector
Chrome Plugin zur Erkennung von KI Stimmen: hiya-deepfake-voice-detec
Griechenland: https://mever.iti.gr/forensics/

Danke fürs Zuhören!

Quellen

Quellen und weitere Informationen:

[Q1] https://www.ibm.com/de-de/think/topics/artificial-intelligence
[Q2] https://www.ibm.com/de-de/think/topics/machine-learning#7281535
[Q3] https://www.ibm.com/de-de/think/topics/generative-ai#257779831
[Q4] https://www.turing.com/resources/generative-ai-tools
[Q5] https://sightengine.com/detect-ai-generated-images
[Q6] https://blog.wasitai.com/2025/10/19/refund-fraud-2-0-how-ai-images-are-changing-the-game/
[Q7] Are You Human? An Adversarial Benchmark to Expose LLMs. Gressel et al (2024) https://arxiv.org/pdf/2410.09569
[Q8] https://www.bpb.de/lernen/digitale-bildung/werkstatt/542670/deepfakes-wenn-man-augen-und-ohren-nicht-mehr-trauen-kann/
[Q9] As Good as a Coin Toss Human Detection of AI-Generated Images, Video, Audio, and Audiovisual Stimuli, Di Cooke et al. (2024) https://arxiv.org/pdf/2403.16760
[Q10] https://bildungssprache.net/ki-texte-erkennen-detektoren/
[Q11] https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai
[Q12] Deutsche Welle: Wie erkenne ich KI-Videos von Sora? https://www.youtube.com/watch?v=FFyfqrII9Yg
[Q13] The Creation and Detection of Deepfakes: A Survey. Mirskey et al (2020) https://dl.acm.org/doi/epdf/10.1145/3425780
[Q14] https://www.bpb.de/shop/zeitschriften/izpb/medienkompetenz-355/539986/fake-news-misinformation-desinformation/
[Q15] https://www.bpb.de/themen/medien-journalismus/stopfakenews/
[Q16] https://www.bmftr.bund.de/SharedDocs/Publikationen/DE/L/31723_Forschung_gegen_Fake_News.pdf?__blob=publicationFile&v=4
[Q17] https://www.tagesschau.de/investigativ/ndr-wdr/russland-propaganda-fakenews-sda-deutschland-100.html