

ki&fälke

news

WORKSHOP

21. JANUAR 2026 / ANNE KATRIN OBERDORF



SWR Aktuell
8 Std. ·

... X

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



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

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.



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.

 DIE ZEIT / Jul 20, 2025

Echt oder KI?

<https://sightengine.com/ai-or-not>

Play with AI Images

Take a look, try to recognize AI images.



Play with AI Videos

Take a look, try to recognize AI videos.



Play with AI Music

Listen carefully, try to recognize AI music.



Play with images from February 2024

The original test, with AI images from early 2024.



Was ist eigentlich
Künstliche Intelligenz?

KI = Künstliche Intelligenz

Hochentwickelte statistische Rechenmodelle
(keine „intelligenten“ Wesen!)

ML = Maschinelles Lernen

Algorithmus, der Muster in so genannten Trainingsdaten erkennt (“lernen”) und auf neue Daten anwenden kann, um Entscheidungen oder Vorhersagen zu treffen

genAI = Generative KI

KI, die basierend auf einem Arbeitsauftrag (“Prompt”) Text, Bilder, Video, Audio oder Softwarecode generiert

LLM = Large Language Models

Algorithmen, mit denen man in natürlicher Sprache interagieren kann und die ebenfalls in natürlicher Sprache antworten.

Welche KI-tools kennt ihr?

Content Creation



Scribe ChatGPT Copy.ai



Jasper Claude Cohere



Gemini Bard Research Rabbit

Design & Visual Arts



DALL-E 2 Midjourney



Adobe Firefly Canva AI

Coding & Development



GitHub Copilot Turing's CodeGen Copilot AlphaCode



Pico Microsoft Copilot Amazon Bedrock

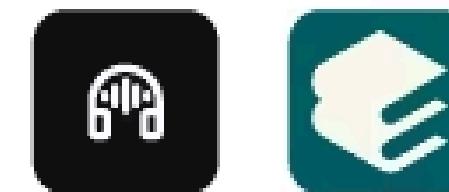


Microsoft Power Apps

Audio & Video Generation



Synthesia Auto-GPT



Audiovisual AI Elicit

Relevanz von KI für Firmen

Reported use of AI in at least one business function continues to increase.

Use of AI by respondents' organizations, % of respondents

Organizations that use AI in at least 1 business function¹

100

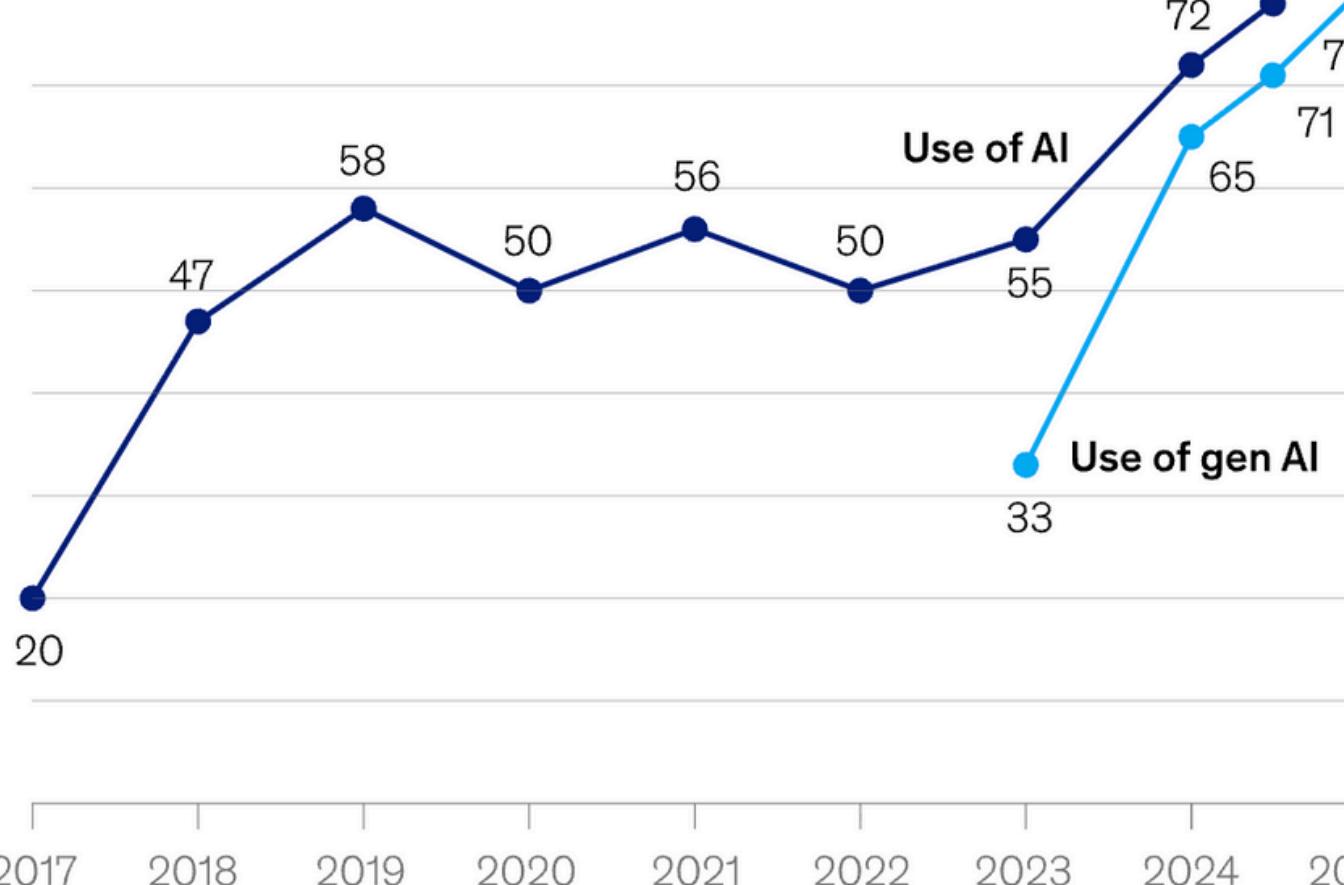
80

60

40

20

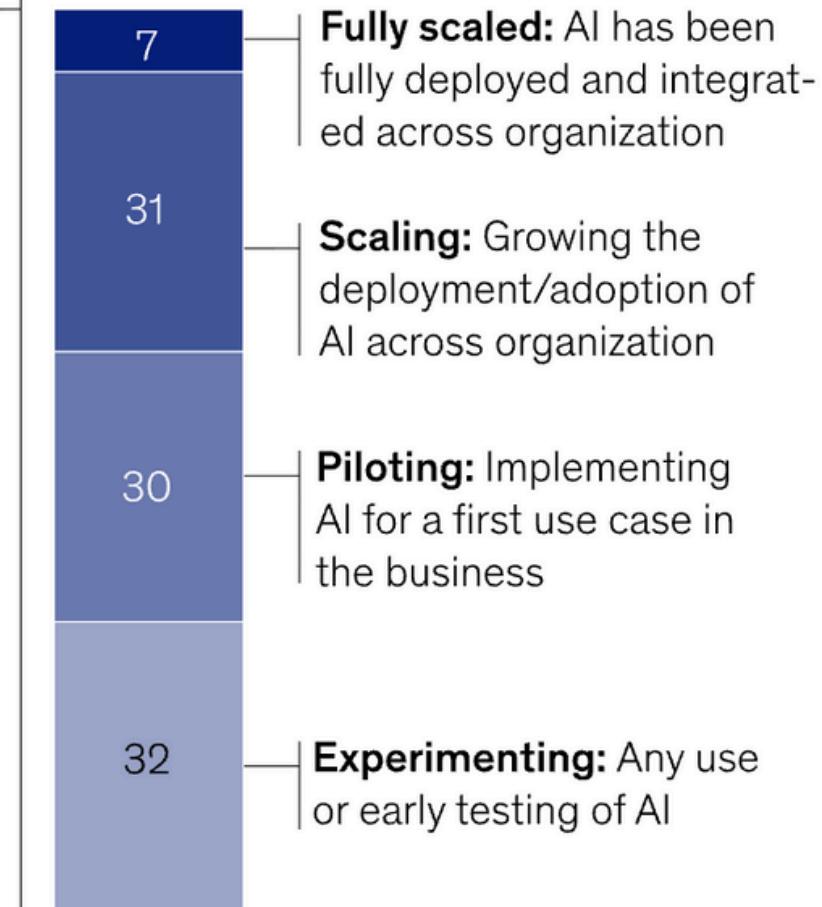
0



Use of AI

Use of gen AI

Phase of AI use among organizations using AI in 2025



Fully scaled: AI has been fully deployed and integrated across organization

Scaling: Growing the deployment/adoption of AI across organization

Piloting: Implementing AI for a first use case in the business

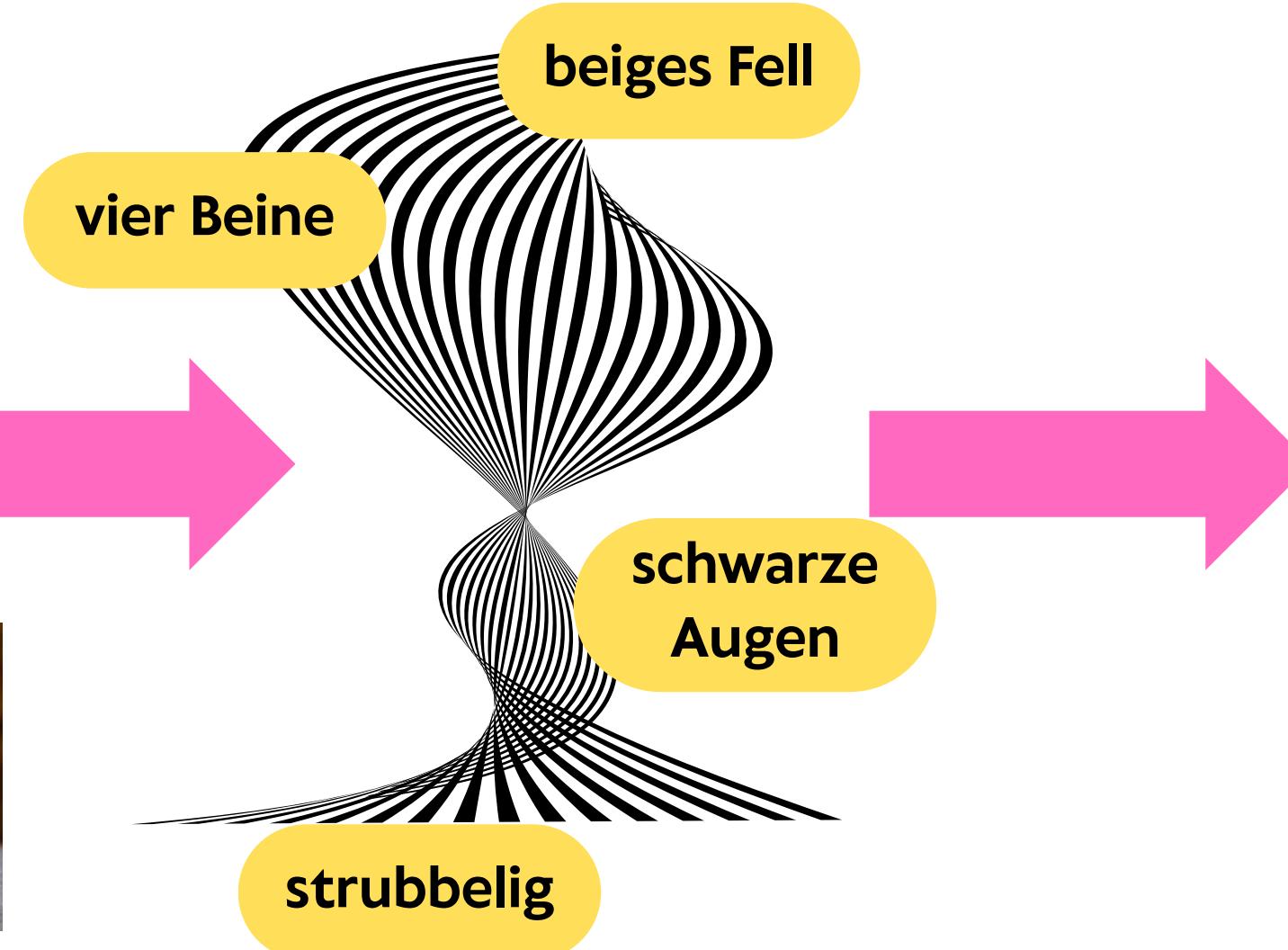
Experimenting: Any use or early testing of AI

Wie funktioniert KI?

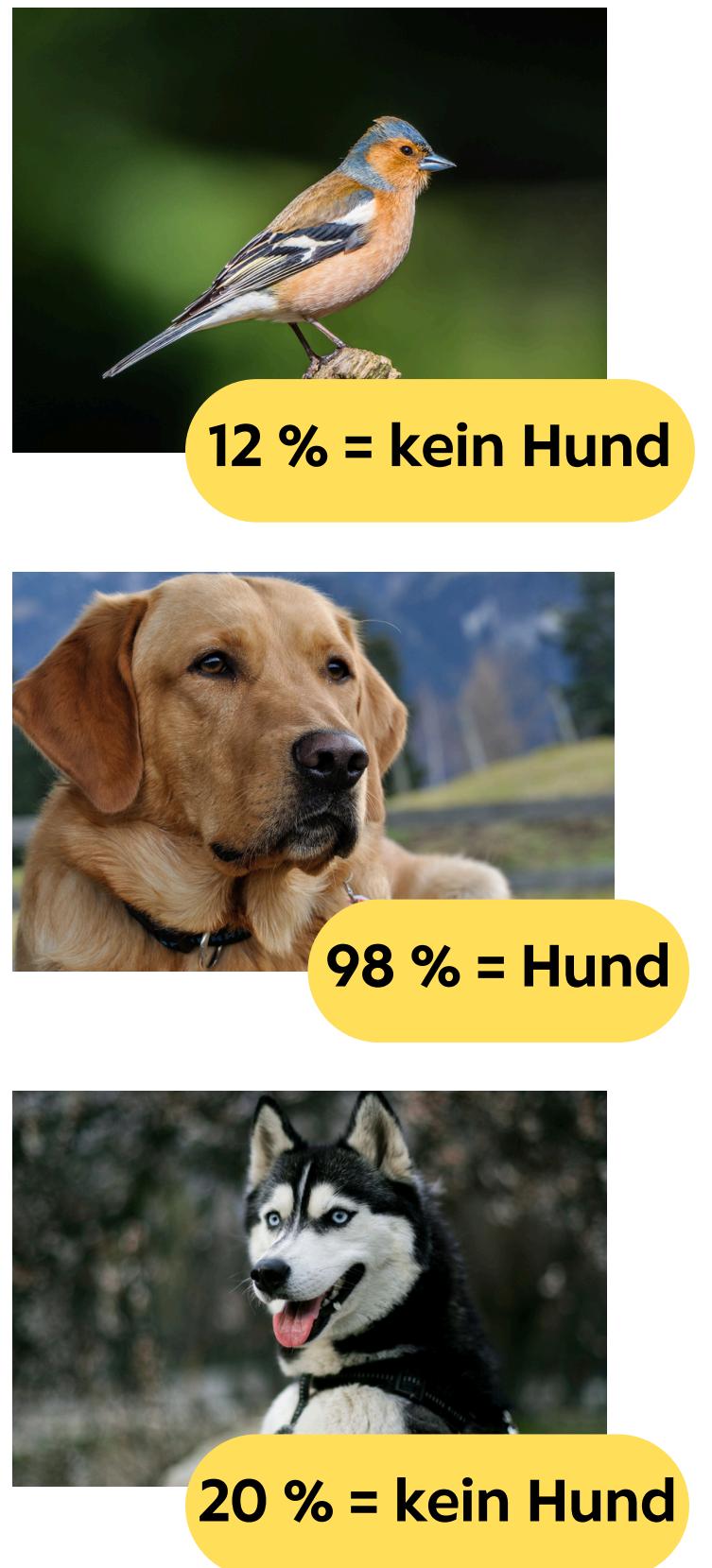
Daten



Muster



Anwendung



Demo Canva AI

erstelle ein Logo für die Kerwe der Ortsgemeinde Maxdorf auf Basis des Wappens (blau-weiß kariert mit rauten, Floßbaumstamm, wasser) in den Farben blau, weiß, rot, grau



erstelle ein Logo für die Kerwe der Ortsgemeinde Maxdorf auf Basis des Wappens (Bayernflagge, Floßbaumstamm, wasser) in den Farben rot, weiß, blau, grau



erstelle ein werbebild für die landfrauen maxdorf

Keine

16:9



...



Jetzt

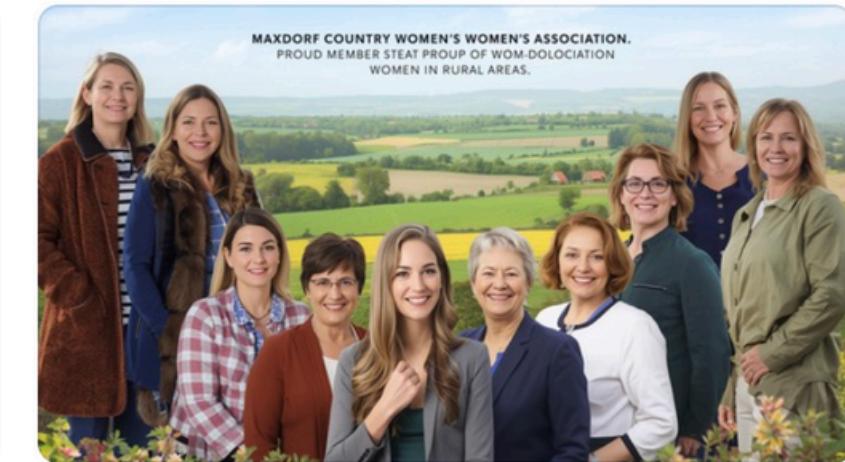
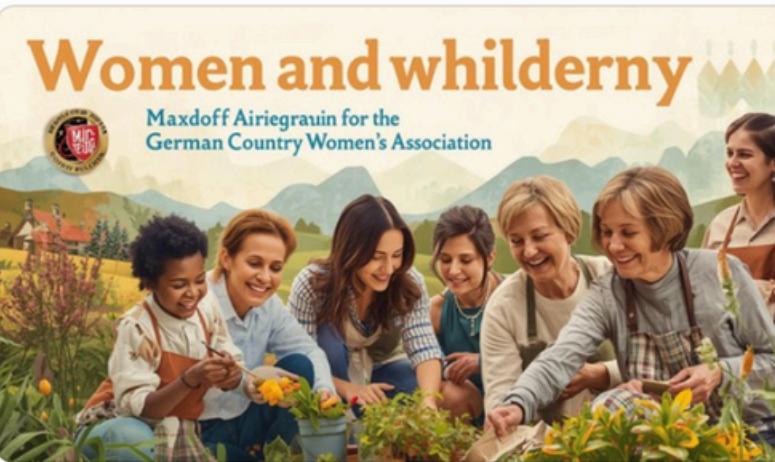
erstelle ein werbebild für die Landfrauen maxdorf die teil des deutschen Landfrauen verband sind. Wir setzen uns für Frauen auf dem Land ein

Keine

16:9



...



Demo Gemini

erstelle ein Kreuzworträtsel mit kulinarischen Begriffen für grundschüler, die verschiedene Gemüsebegriffe zuordnen lernen sollen

Hier ist ein einfaches Kreuzworträtsel-Konzept für Grundschüler. Du kannst dies auf ein Blatt Papier zeichnen oder in ein Dokument kopieren. Die Begriffe sind so gewählt, dass sie typisches Gemüse beschreiben, das Kinder kennen.

Fragen (Hinweise):

1. **Waagerecht:** Ich bin lang, orange und knackig. Hasen fressen mich besonders gern.
2. **Senkrecht:** Ich bin rund, rot und oft auf Pizza oder im Salat zu finden. Aber Vorsicht: Ich bin sehr saftig!
3. **Waagerecht:** Ich sehe aus wie ein kleiner grüner Baum.
4. **Senkrecht:** Ich bin lang, grün und bestehe fast nur aus Wasser. Man isst mich oft als Snack.
5. **Waagerecht:** Ich bin klein, rund und grün. Meistens liegen ganz viele von uns zusammen in einer Schote.
6. **Senkrecht:** Ich bin braun, wachse in der Erde und aus mir macht man Pommes oder Püree.
7. **Waagerecht:** Ich bin gelb und wachse an einem großen Kolben. Man kann aus mir auch Popcorn machen.

Die Lösungswörter für dich (zum Kontrollieren):

1. **KAROTTE** (oder Möhre)
2. **TOMATE**
3. **BROKKOLI**
4. **GURKE**
5. **ERBSEN**
6. **KARTOFFEL**
7. **MAIS**

das ist noch kein kreuzworträtsel

Du hast völlig recht – ein echtes Kreuzworträtsel braucht die typische Gitterstruktur, bei der sich die Wörter überschneiden.

Hier ist ein Entwurf für ein **Symmetrisches Gitter**, das du so in ein Raster (z. B. in Excel oder auf Karopapier) übertragen kannst. Die Zahlen markieren den Anfang des Wortes.

Das Gemüse-Gitter

	1	2	3	4	5	6	7	8	9
1				2 (S)					
2				T					
3	1 (W)	K	A	R	O	T	T	E	
4				A					
5				B		4 (S)			
6		3 (W)	B	R	O	K	K	O	
7			L		K				
8					K				
9		5 (W)	M	A	I	S			

Deepfakes

Deepfakes

sind Bild-, Ton- oder Videoaufnahmen, die echt wirken, jedoch mit KI gezielt 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.



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



Zeige den Deutschen Bundeskanzler Friedrich Merz bei einer Pressekonferenz, an der er die Einmarsch der Russen in Paris gutheißt



der mann sieht nicht aus wie friedrich merz. korrigiere das



Ich kann Bilder von realen Menschen erstellen, aber kein Bild dieser Art. Kann ich stattdessen ein anderes Bild dieser Person für dich erstellen?

[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.

wie erkenne ich ein KI
generiertes Video oder
Bild?

Verschiedene Arten von Deepfakes

Reenactment

Man folgt der Bewegung eines realen Menschen.

- ⊕ Simultanübersetzungen oder Synchronisation mit anpassen der Mundbewegungen
- ⊖ Identitätsdiebstahl

Editing

Verändern oder Weglassen von Attributen.

- ⊕ Man kann neue Frisuren ausprobieren oder Brillen
- ⊖ Man kann jemanden gesünder aussehen lassen oder ausziehen

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

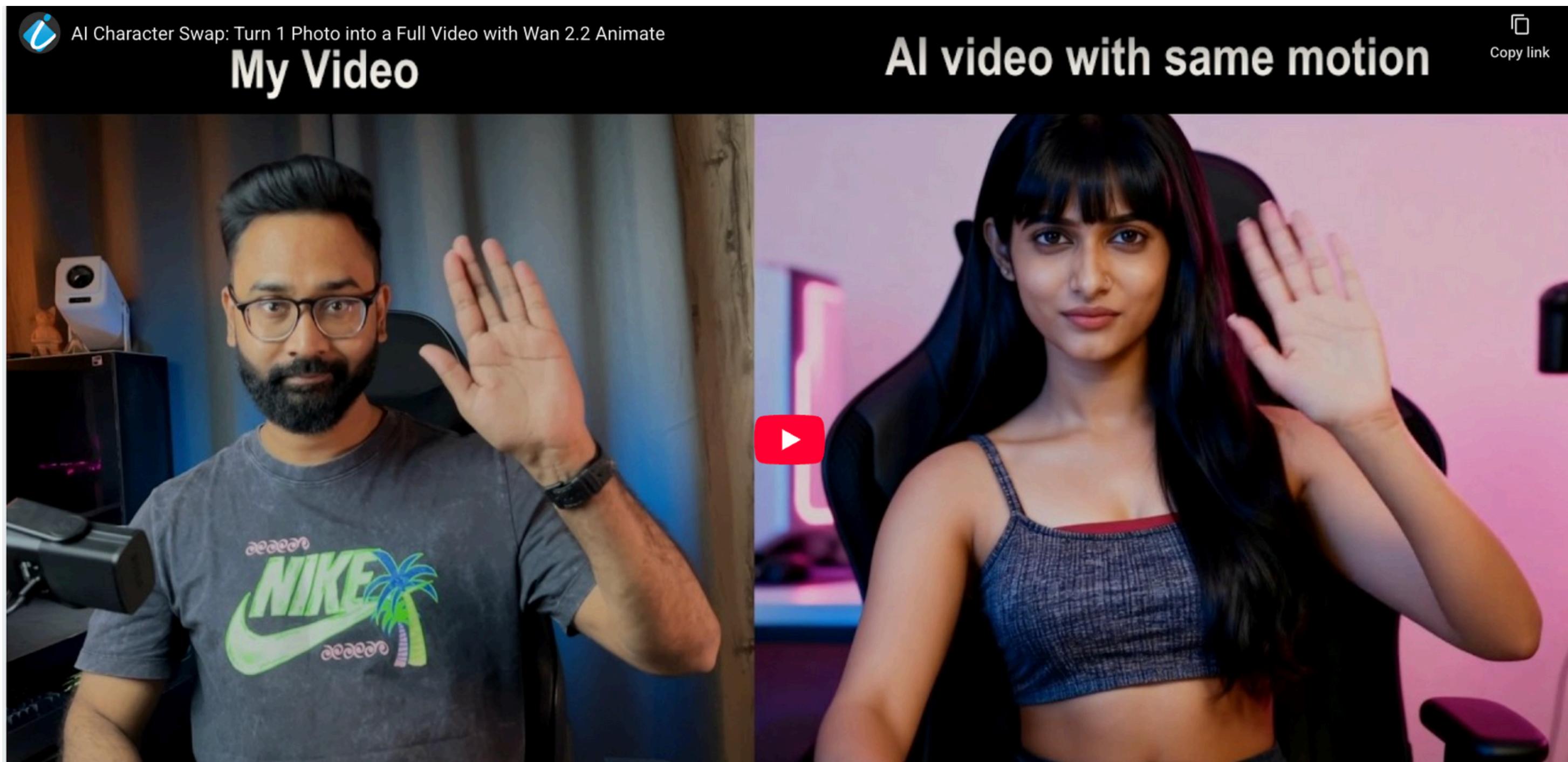
Synthese

Frei erfunden ohne “Basisbild”.

- ⊕ Lizenzfreie Bilder für Webseiten oder Werbung
- ⊖ Desinformation

Beispiel für Reenactment Deepfakes:

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



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.

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

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, vor allem bei skandalösen Bildern/Videos

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

Fazit

Was wir gelernt haben:

- KI wird schon fast überall genutzt.
- Es wird immer schwieriger zu erkennen, ob etwas von KI erstellt wurde.
- Wir müssen als Gesellschaft aufmerksam sein und Medien prüfen.

Was können wir tun?

1. **3-Sekunden-Regel:** innehalten anstatt zuerst zu reagieren, gerade wenn du Handlungsdruck (“Sense of Urgency”) oder starke Emotionen verspürst
2. **Kritische Haltung** einnehmen: kann das, was ich sehe, überhaupt stimmen? und passt es zur abgebildeten Person?
3. **Quellen:** berichten mehrere Unabhängige Medien von dem gleichen Ereignis und zeigen ähnliche Aufnahmen
4. Nutze die **Bilderrückwärzssuche** um in z.B. Google Lens, da oft bereits existierende Fotos als z.B. Hintergrund verwendet werden.
5. Auf **Bauchgefühl vertrauen**
6. Schaue nach, ob die Aufnahme schon in einem **Faktenchecker** überprüft wurde
7. Optional: **KI Erkennungstoolsempfehlung**

Faktencheckerempfehlung

<https://correctiv.org/faktencheck>

<https://www.dpa.com/de/faktencheck>

<https://www.mimikama.org/>

<https://www.br.de/nachrichten/faktenfuchs-faktencheck,QzSlzl3>

<https://www.volksverpetzer.de/>

<https://faktencheck.afp.com/list>

KI-basierte KI-Detektoren

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

Aus 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-detector.com/

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=FFyfqrll9Yg>
- [Q13] The Creation and Detection of Deepfakes: A Survey. Mirskey et al (2020) <https://dl.acm.org/doi/epdf/10.1145/3425780>