Deepfakes

Deepfakes können eine Gefahr für Unternehmen, Vereine, Verbände, Geschäftsstellen, Vorstände sein!

- Was sind Deepfakes?
- Wie erkennt man sie?
- eigenes Risiko bewerten
- Checkliste 10 Regeln für NGOs
- KI-Kennzeichnung
- hilfreiche Seiten



Deepfakes

AI Act der EU (KI-Verordnung) definiert **Deepfake** als:

- Bild-, Ton- oder Videoinhalt durch KI erzeugt oder manipuliert
- realen Personen, Gegenständen, Orten, Einrichtungen oder Ereignissen ähnelt
- fälschlicherweise als echt oder wahrheitsgemäß erscheint

Verordnung (EU) 2024/1689



(Deep-)Fakes *… nicht ganz so deep ;-)*

kennen wir alle!

- *Fake SMS, WhatsApps, „Enkeltrick“*
- *Paketdienste oder Telekommunikationsdienste*
- *Vorstands-Mails*

  immer prüfen:
  Sprache, Absenderadresse







Bekannte Deepfakes

Beispiele:

- Prominente machen Werbung für Finanz- oder Gesundheitsprodukte
- „25-Mio-$-Betrug“ in einer Videokonferenz
- Fall Collien Fernandes
- kürzlich viral:
  Lieferwagen schiebt Auto in letzter Sekunde von den Gleisen



https://www.mimikama.org/ki-video-bahnuebergang-lieferwagen

Deepfakes erkennen?

Woran kann man KI-Bilder und KI-Videos erkennen?

1) Nicht im Bild beginnen, sondern davor:

(Begleitinformationen zum Bild einbeziehen)

- ✓ KI-Label vorhanden
- ✓ Hinweise in Beschreibung, Hashtags oder Profil?
- ✓ Postet der Account regelmäßig ähnliche Inhalte?

Manche Hinweise stehen nicht im Bild, sondern drumherum.

2) Das Bild selbst prüfen

- ✓ Unmögliche Physik oder falsche Perspektiven
- ✓ Schrift und Logos
- ✓ Körperteile auffällig
- ✓ Übertrieben glatter, künstlicher Look

KI scheitert selten am Eindruck – aber oft an Details.



3) Bewegung entlarvt vieles

- ✓ Körperteile verformen sich
- ✓ Nicht Lippensynchron
- ✓ Objekte tauchen auf oder verschwinden
- ✓ Bewegungen weich, aber unlogisch

Videos verraten KI schneller als Standbilder.

4) Woher kommt das Bild?

- ✓ Rückwärtssuche
- ✓ unabhängige Berichte?
- ✓ Passt Ort, Zeit, Kontext zusammen?

Was real ist, hinterlässt Spuren.

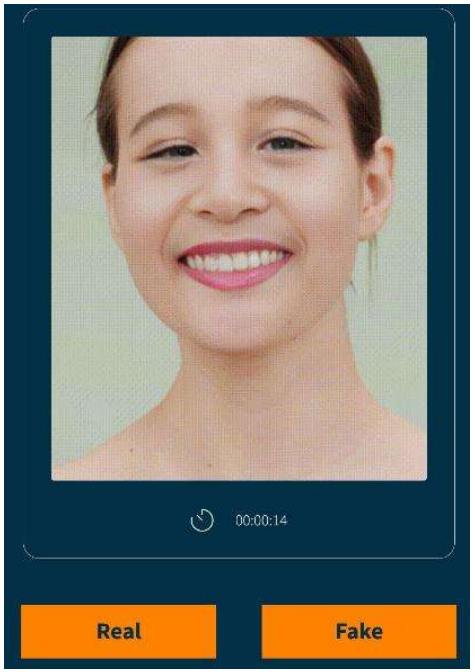
Ein einzelnes Merkmal reicht nicht.
Entscheidend ist die Kombination.

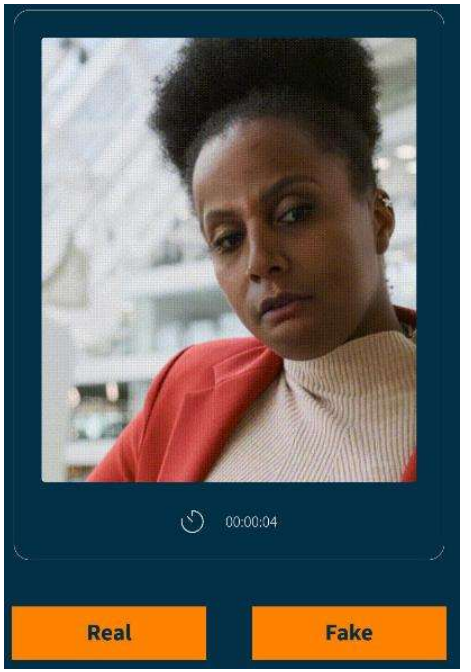
https://www.mimikama.org/woran-erkennst-du-ki-bilder-und-ki-videos

Deepfakes erkennen

… ist gar nicht so leicht!



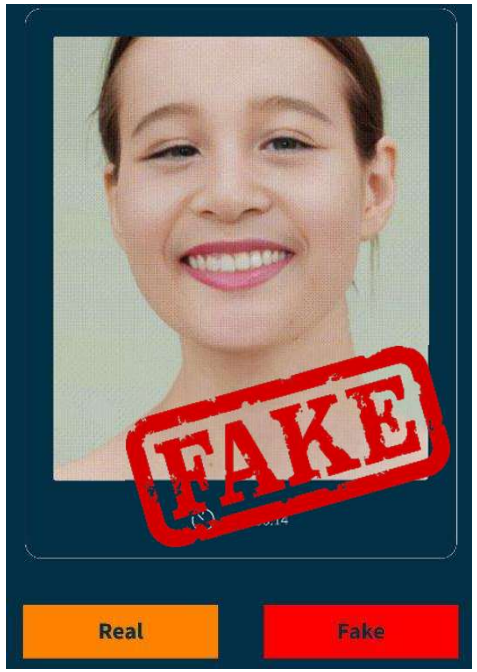


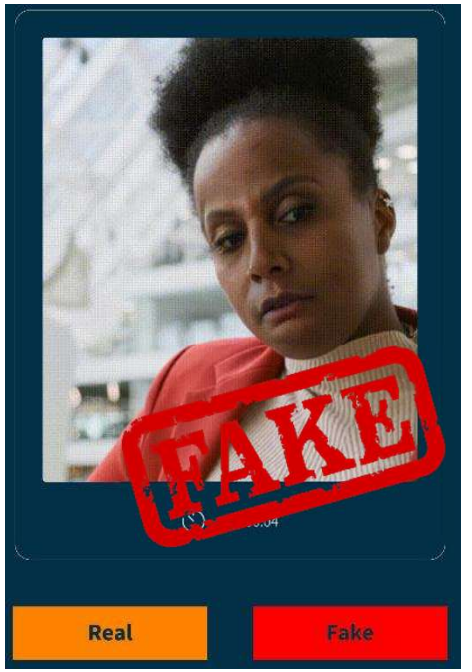


Deepfakes erkennen

… ist gar nicht so leicht!







gestochen scharf oder verpixelt, mit oder ohne Hintergrund,
Emotionen – alles lässt sich mit KI darstellen

2 Deepfake-Quiz:
https://quiz.iproov.com/quiz/
https://detectfakes.kellogg.northwestern.edu/

Deepfakes - das eigene Risiko?

- Falsche Mails / Sprachnachrichten / Videos
  von eigenem Geschäftsführer/in / Vorstand mit Aufträgen (€)
- Falsche Mails / Sprachnachrichten / Videos
  von Mitgliedsorganisationen, Partnern, Ministerium, Parteien etc.
- (polit.) Aussagen im Namen des Verbandes etc.
- Fake-Shops
- …

Das besondere **Risiko bei Organisationen der Selbsthilfe**:
Wir arbeiten mit Vertrauen, Spenden, Öffentlichkeit
und oft sensiblen Personendaten.
Das sind besonders verletzliche Bereiche!



Deepfake → Risiko - Checkliste & Lösungen!



https://secure-ngo-shield.lovable.app/

Fragen zum Risikocheck?

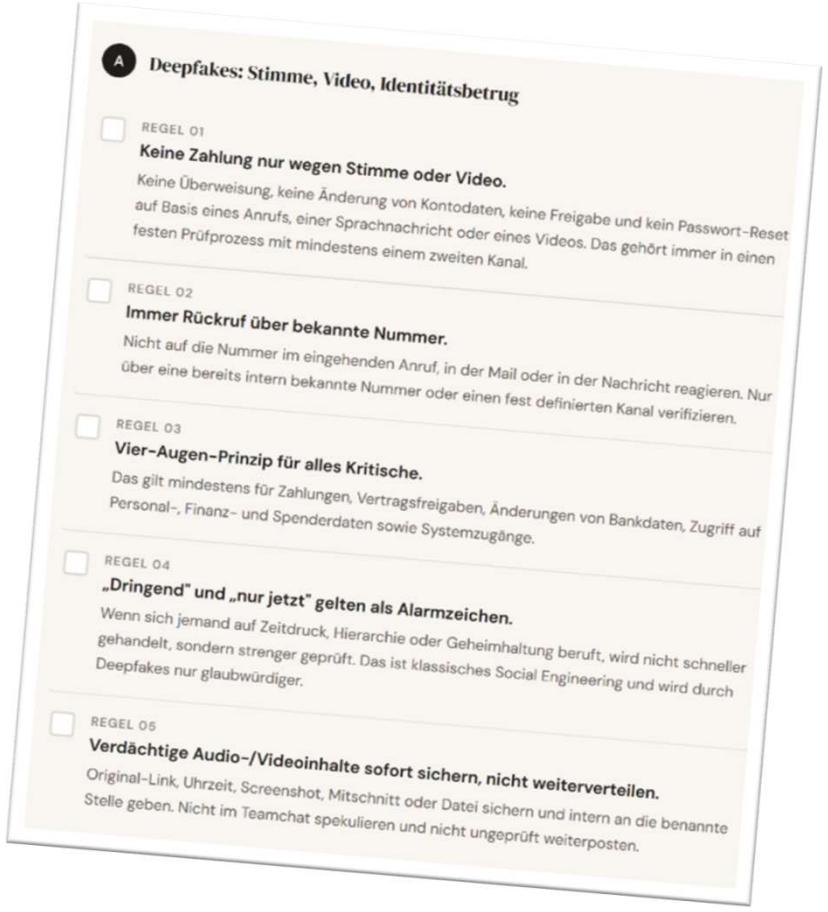
1) Wie viele Mitarbeitende (…) ?
2) Habt ihr einen öffentlichen YouTube-Kanal (…) ?
3) Nutzt ihr WhatsApp (…) ?
4) Habt ihr eine öffentliche Spendenpage (…) ?
5) Verschickt ihr Newsletter (…) ?
6) Ist eure Geschäftsführung / Vorstand öffentlich sichtbar (…) ?
7) Nutzt ihr KI-Tools (…) ?
8) Habt ihr festgelegt, wer im Krisenfall kommuniziert?

Googelt mal euren Organisationsnamen plus 'Spenden' und prüft, ob es Fake-Seiten gibt!

Deepfakes & Desinformation - 10 Regeln für NGO-Teams

(A) Deepfakes: Stimme, Video, Identitätsbetrug

1) Keine Zahlung nur wegen Stimme oder Video.
2) Immer Rückruf über bekannte Nummer.
3) Vier-Augen-Prinzip für alles Kritische.
4) „Dringend" und „nur jetzt" gelten als Alarmzeichen.
5) Verdächtige Audio-/Videoinhalte sofort sichern.



10 Regeln für NGO-Teams Gesprächsleitfaden & Notfallcheckliste unter:

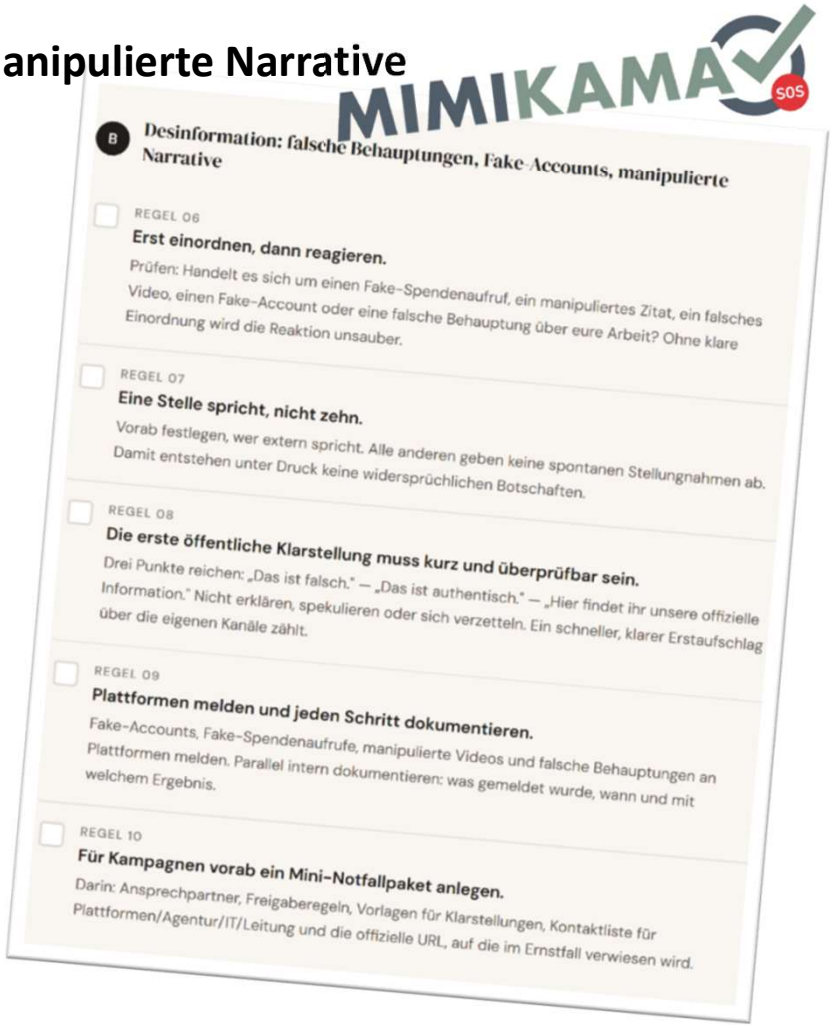
https://claude.ai/public/artifacts/4d623520-bd1d-489a-9969-1cf571b64e28

Autorin: Inken Paland, vely.ai — Responsible AI Innovation

Deepfakes & Desinformation - 10 Regeln für NGO-Teams

(B) Desinformation: falsche Behauptungen, Fake-Accounts, manipulierte Narrative

6) Erst einordnen, dann reagieren.
7) Eine Stelle spricht, nicht zehn.
8) Die erste öffentliche Klarstellung ist kurz & überprüfbar.
9) Plattformen melden und jeden Schritt dokumentieren.
10) Für Kampagnen vorab ein Mini-Notfallpaket anlegen.



10 Regeln für NGO-Teams Gesprächsleitfaden & Notfallcheckliste unter:

https://claude.ai/public/artifacts/4d623520-bd1d-489a-9969-1cf571b64e28

Autorin: Inken Paland, vely.ai — Responsible AI Innovation

Kennzeichnung von KI-Inhalten

Seit 01.August gilt Artikel 50 des EU AI Act: Chatbots, Deepfakes und ungeprüft veröffentlichte KI-Texte zu Themen von öffentlichem Interesse müssen gekennzeichnet werden!

Inhalt	Kennzeichnung nötig?
Chatbot auf der Website Hinweis im Chat, außer es ist ohnehin offensichtlich	Ja
Täuschend echtes Bild, Video oder Audio realer Personen oder Ereignisse Deepfakes, gilt ohne Ausnahme	Ja, immer
Komplett KI-erstellter News- oder Infotext, ungeprüft veröffentlicht	Ja
KI-Text mit menschlicher Prüfung und redaktioneller Verantwortung	Nein
KI als Werkzeug (Korrektur, Übersetzung, Recherche, Entwurf)	Nein
Erkennbar stilisierte Illustrationen und Grafiken	Nein
Werbetexte, Produktbeschreibungen, interne Dokumente	Nein
Fotorealistische KI-Bilder ohne realen Bezug Kennzeichnung ist der sichere Weg	Grauzone

https://www.seven-bytes.de/blog/ki-inhalte-kennzeichnen-pflicht-2026/



https://digital-strategy.ec.europa.eu/en/policies/eu-icons-labelling-ai-generated-content



Hilfreiche Seiten …

Nicht sicher ob eine E-Mail Betrug, eine Webseite oder Nachrichtenmeldung echt ist?

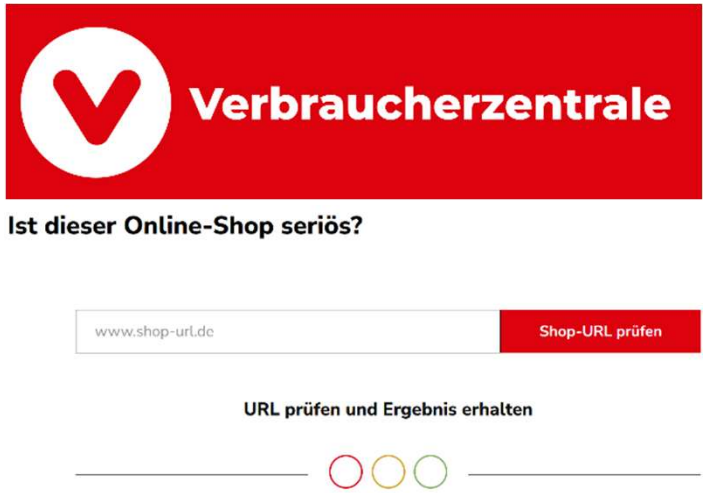
Der Faktencheck bei MIMIKAMA kann helfen!



https://www.mimikama.org

Ist ein Online Shop seriös?

Der Fakeshop-Finder der Verbraucherzentrale kann helfen!

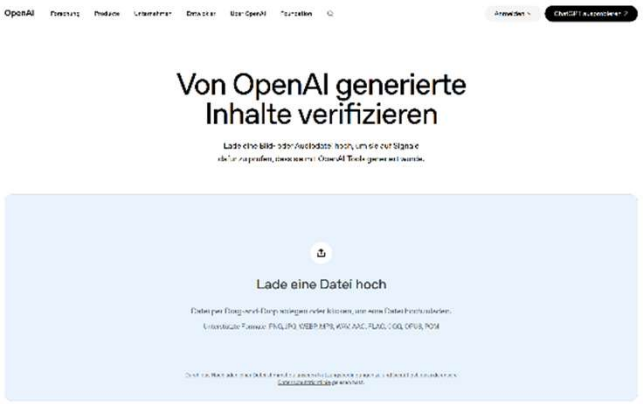


https://www.verbraucherzentrale.de/fakeshopfinder

Hilfreiche Seiten 2 …

Prüfen, ob Inhalte von einer Open AI KI (ChatGPT)generiert wurden?

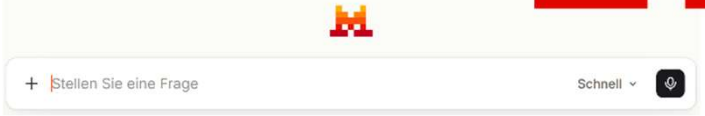
https://openai.com/de-DE/research/verify



Alternative KIs (Datenschutzkonform)

(zur eigenen weiteren Recherche)

MISTRAL (Vibe, früher Le Chat)
Europäische KI aus Frankreich
https://chat.mistral.ai/chat



VELY.AI
KI Lokal und Offline
https://vely.ai



Alle Links aus dieser Präsentation

Deepfake-Beispiel:
- https://www.mimikama.org/ki-video-bahnuebergang-lieferwagen

Deepfake Erkennen:
- https://www.mimikama.org/woran-erkennst-du-ki-bilder-und-ki-videos

Deepfake-Quiz:
- https://quiz.iproov.com/quiz/
- https://detectfakes.kellogg.northwestern.edu/

Risiko Checkliste:
- https://secure-ngo-shield.lovable.app/

Kennzeichnung von KI-Inhalten
- https://digital-strategy.ec.europa.eu/en/policies/eu-icons-labelling-ai-generated-content
- https://www.seven-bytes.de/blog/ki-inhalte-kennzeichnen-pflicht-2026/

10 Regeln für NGO-Teams:
- https://claude.ai/public/artifacts/4d623520-bd1d-489a-9969-1cf571b64e28

10 Regeln für NGO-Teams:
Handliches Word-Dok zum ausdrucken und im Büro aufhängen:
- https://urlr.me/Cu7A3Q

Faktencheck, Notfall-Check (SOS) & Fakeshop-Finder:
- https://www.mimikama.org
- https://www.verbraucherzentrale.de/fakeshopfinder

KI-Inhalte prüfen (nur OpenAI)
- https://openai.com/de-DE/research/verify

Alternative KIs (zur eigenen, weiteren Recherche)
- https://chat.mistral.ai/chat
- https://vely.ai

Alle Links überprüft am 14.10.2026