Deepfakes

Wenn Künstliche Intelligenz täuschend echte Inhalte schafft.

Deepfakes sind Medieninhalte wie Bilder, Videos oder Tonaufnahmen, die mit Hilfe von Künstlicher Intelligenz erzeugt werden und realistisch wirken, obwohl sie es nicht sind.

Grob kann zwischen **manipulierten Inhalten** (bestehende Fotos oder Bilder werden mit KI bearbeitet) und vollständig **künstlichen Inhalten** (vollständig von KI erstellte Bilder oder Videos) unterschieden werden.

Manipulierte Inhalte

Scannen Sie die QR-Codes, um sich die Beispiele anzusehen.

Face Swap – „Gesichter-Tausch“: In einem Bild oder einem Video wird das Gesicht einer Person mit dem Gesicht einer anderen Person ausgetauscht.

Lipsync - Lippensynchronisation: Diese Technik wird vor allem verwendet, um bekannten Personen Worte in den Mund zu legen, die sie nie gesagt haben.

Face Reenactment – „Gesichtsanimation“: In einem Video wird die Mimik einer Person auf das Gesicht der Person im Zielvideo übertragen.

Face Morph – „Gesichter-Verschmelzung“: Bilder oder Videos von zwei verschiedenen Personen werden übereinandergelegt und so kombiniert, dass ein neues Gesicht entsteht.

Künstliche Inhalte

Voice Cloning – „Stimme klonen“: Beim Voice Cloning wird die menschliche Stimme digital nachgebildet. Damit können fremde Personen so klingen wie ein Familienmitglied.

Generative KI-Bild- und Videogenerierung: Auf Basis von realen Bildern und Videos wird eine KI trainiert, die mittels Texteingaben (Prompts) neue, noch nicht existierende Bilder und Videos erzeugt. Rechts sehen Sie ein Beispiel in dem das Kolosseum, das eigentlich in Rom steht, in der Wüste platziert wurde.

Abb.: KI-generiertes Bild (Fooocus)

Bundesministerium Soziales, Gesundheit, Pflege und Konsumentenschutz
Bundeskanzleramt
oiat
digitaleSenior:innen
Stand: 12/2024

Deepfakes

Verantwortung, Regulierung und Potenziale.

Gefahren von Deepfakes

Manipulation und Täuschung: Deepfakes können genutzt werden, um Personen zu verleumden, politische Prozesse zu beeinflussen oder Falschinformationen zu verbreiten. Prominente oder Politiker sind besonders häufige Ziele, da es von ihnen viele Aufnahmen im Internet gibt. Deepfakes können das Vertrauen in mediale Inhalte erschüttern.

Beeinträchtigung der Autonomie: Täuschende Inhalte erschweren es den Menschen, zwischen Wahrheit und Fiktion zu unterscheiden, was eine informierte Entscheidungsfindung beeinträchtigt.

Missbrauchspotenzial: Deepfakes bergen Risiken wie Cybermobbing und Identitätsdiebstahl, die nicht nur individuelle Datenschutz- und Persönlichkeitsrechte verletzen, sondern auch psychische Belastungen, soziale Folgen und finanzielle Schäden nach sich ziehen können.

Die Erstellung von Deepfakes mit Aufnahmen des Gesichts, der Stimme oder des Körpers einer Person, verletzt deren Persönlichkeitsrechte und deren Datenschutzrecht. Die Internet Ombudsstelle berät Sie gerne bei rechtlichen Fragen: www.ombudsstelle.at.

Rechtliche Einordnung

Die EU hat mit dem Digital Services Act und dem AI-Act Maßnahmen ergriffen, um die Verbreitung von Desinformationen einzudämmen und KI-Systeme zu regulieren. Online-Plattformen wie YouTube sind zum Beispiel verpflichtet, KI-generierte Inhalte, einschließlich Deepfakes, klar zu kennzeichnen.

Abb.: KI-generiertes Bild (ChatGPT)

Positive Potenziale von Deepfakes

Nicht alle Anwendungen der Deepfake-Technologie sind negativ. Sie kann auch eingesetzt werden, um Menschen zu helfen, z.B. durch die Rekonstruktion von Stimmen, die durch Krankheit verloren gegangen sind.

Deepfakes können für realistisches Sprachtraining und medizinische Simulationen eingesetzt werden. Auch in der Spieleindustrie und der Kunstwelt eröffnen sie neue kreative Möglichkeiten. Alte Filme und Fotos können mit Deepfakes restauriert und digital verbessert werden.

Bundesministerium Soziales, Gesundheit, Pflege und Konsumentenschutz
Bundeskanzleramt
oiat
digitaleSenior:innen
Stand: 12/2024

Deepfakes

Wissen, was echt ist: Digitale Täuschungen erkennen.

Mit Medienkompetenz gegen Deepfakes

- **Bleiben Sie kritisch:** Hinterfragen Sie ungewöhnliche oder schockierende Inhalte.
- **Prüfen Sie Quellen:** Verlassen Sie sich nicht auf einzelne Videos oder Bilder. Suchen Sie nach vertrauenswürdigen Quellen, die die Inhalte bestätigen.
- **Bleiben Sie informiert:** Halten Sie sich über neue Technologien und aktuelle Entwicklungen zu Deepfakes auf dem Laufenden und tauschen Sie sich auch mit anderen dazu aus.
- **Achten Sie auf Details:** Waren Deepfakes bis vor kurzem in der Regel noch mit bloßem Auge zu erkennen, wird dies durch die rasant fortschreitende Professionalisierung von KI-Tools immer schwieriger. Es gibt jedoch einige Tipps, mit denen viele der Deepfakes derzeit (noch) erkannt werden können:

1. Unscharfe Übergänge
2. Unnatürliche Mimik und Bewegungen
3. Fehlendes Blinzeln
4. Unterschiedliche Qualitäten
5. Fehlerhafte Hände, Zähne oder Haare
6. Unstimmige Stimme

Beispiel: Ein Hinweis ist etwa die Partie um die Augen bzw. Brille. Die Brille auf dem linken KI-generierten Bild hat unten keinen Rand und geht über in die Haut. Auch der Schatten der Brille auf der rechten Wange passt nicht zur Form der Brille. Dem falschen Papst fehlt außerdem ein für den echten Franziskus charakteristischer Leberfleck unter dem rechten Auge.

Abb.: Deepfake Beispiel vom Papst, Quelle: BR24

DeepFake-o-meter oder Deepware Scanner sind kostenlose Tools, die es ermöglichen, Videos und Bilder auf Deepfakes zu überprüfen. Bitte beachten Sie jedoch, dass kein Tool absolute Sicherheit bieten kann. Bleiben Sie stets kritisch gegenüber verdächtigen Inhalten!

Bundesministerium Soziales, Gesundheit, Pflege und Konsumentenschutz
Bundeskanzleramt
oiat
digitaleSenior:innen
Stand: 12/2024

Deepfakes

Begriffserklärungen & hilfreiche Links

Begriffserklärungen

Cybermobbing bezeichnet das absichtliche Beleidigen, Bedrohen, Bloßstellen oder Belästigen von Personen über digitale Kommunikationsmittel wie soziale Medien oder E-Mails.

Medienkompetenz ist die Fähigkeit, Medieninhalte kritisch zu bewerten und verantwortungsvoll zu nutzen. Sie umfasst das Verständnis, wie verschiedene Medien funktionieren, wie man Informationen richtig interpretiert und wie man eigene Inhalte sicher erstellt und teilt.

Online-Plattformen sind Internetseiten oder Anwendungen am Smartphone, auf denen Sie mit anderen kommunizieren, Informationen austauschen oder Dienste nutzen können, wie z.B. Facebook oder YouTube.

Tools sind digitale Werkzeuge oder Programme, die Ihnen helfen, bestimmte Aufgaben am Computer oder Smartphone zu erledigen, wie Bildbearbeitung oder das Schreiben von Texten.

Links

Infoblatt „Künstliche Intelligenz“: https://www.digitaleseniorinnen.at/fileadmin/redakteure/Downloads/Infoblatt_KI.pdf

Infoblatt „Generative Künstliche Intelligenz“: https://www.digitaleseniorinnen.at/fileadmin/redakteure/Downloads/Infoblatt_Generative-KI.pdf

Infoblatt „Der AI-Act“: https://www.digitaleseniorinnen.at/fileadmin/redakteure/Downloads/Infoblatt_AI-Act.pdf

DeepFake-o-meter: https://zinc.cse.buffalo.edu/ubmdfl/deep-o-meter/landing_page

Deepware Scanner: https://scanner.deepware.ai/

KI-Avatare in der Lehre (erwachsenenbildung.at): https://www.youtube.com/watch?v=lMX8Y-qUyUvU&t=754s

Digital Services Act: https://www.rtr.at/medien/was_wir_tun/DigitaleDienste/DSA/DSA.de.html

KI-Tool „Voice Engine“: https://openai.com/index/navigating-the-challenges-and-opportunities-of-synthetic-voices/

Welches Gesicht ist echt: https://www.whichfaceisreal.com/

Bundesministerium Soziales, Gesundheit, Pflege und Konsumentenschutz
Bundeskanzleramt
oiat
digitaleSenior:innen
Stand: 12/2024