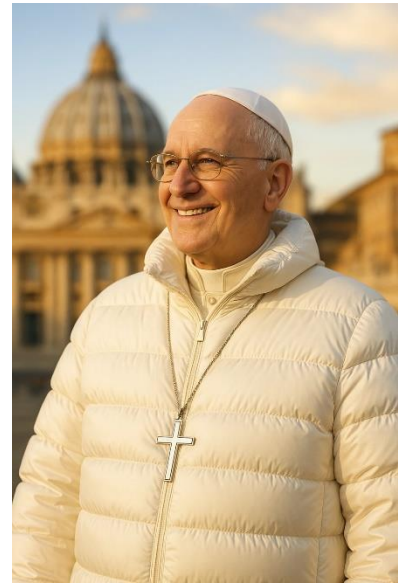


M6-S Deepfakes

Täuschend echte Videos, Bilder und Tonaufnahmen, die vollständig von KI-Systemen generiert oder manipuliert werden, sogenannte Deepfakes¹, eröffnen eine neue Dimension digitaler Täuschung.

Schon mit einfachen Prompts lassen sich Deepfakes erzeugen: Wer einem Bildgenerator präzise Vorgaben zu Motiv, Stil und Stimmung macht, erhält fotorealistische Ergebnisse, die echte Aufnahmen vortäuschen können.

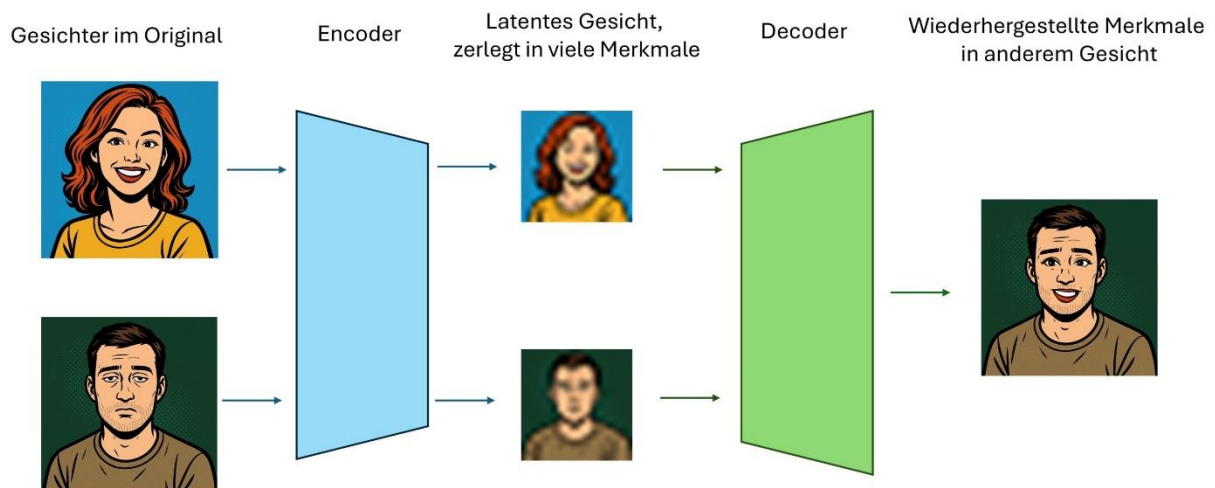
Das hier gezeigte Bild von Papst Leo XIV. im weißen Daunenmantel wurde kreiert mit dem Prompt: „Erstelle eine Aufnahme von Papst Leo XIV. in einem weißen Daunenmantel mit einer Kette mit einem Kreuzanhänger um den Hals. Er steht im Sonnenlicht vor dem Vatikan, der Hintergrund ist verschwommen, fotorealistischer Stil.“



Bei der Herstellung dieses Bildes wurde am 16.05.2025 ChatGPT o3 eingesetzt.

Wie entsteht ein Deepfake?

Deepfakes entstehen durch Computerprogramme, die Methoden Künstlicher Intelligenz nutzen. Zuerst „lernt“ das Programm das Gesicht oder die Stimme einer Person kennen. Dafür bekommt es viele Bilder, Videos oder Tonaufnahmen. Ein Teil des KI-Systems (der „Encoder“) zerlegt das Gesicht oder die Stimme in wichtige Merkmale. Ein anderer Teil (der „Decoder“) setzt diese Merkmale später wieder zusammen – aber zum Beispiel in einem neuen Video.



Bei der Herstellung dieses Bildes wurde am 17.05.2025 ChatGPT o3 eingesetzt.

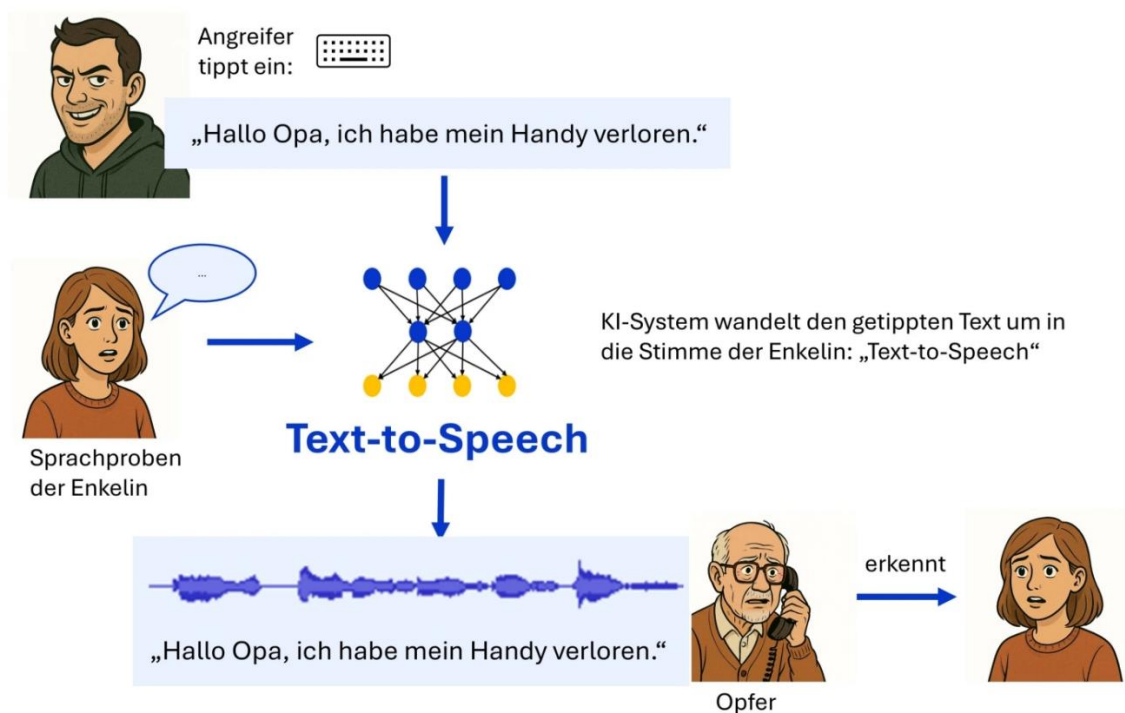
So entsteht beim „Face-Swap“ ein neues Bild, in dem das ursprüngliche Gesicht durch das einer anderen Person oder Teilen davon ersetzt ist. Das Programm vergleicht dabei das Ergebnis ständig

¹ Der Begriff „Deepfake“ setzt sich aus „Deep Learning“ - einer Methode des maschinellen Lernens - und dem Wort „Fake“ zusammen.

mit dem Original und verbessert sich in vielen Durchläufen (Iterationen). Auch Stimmen können gefälscht werden. Dabei lernt das KI-System aus vielen Sprachaufnahmen, wie eine bestimmte Stimme klingt. Anschließend kann es neue Sätze mit genau dieser Stimme sprechen lassen. So hören sich Deepfake-Stimmen täuschend echt an (DFKI, 2025).

Deepfakes basieren nicht auf einer einzigen Methode, sondern auf verschiedenen Verfahren, die je nach Anwendungsfall unterschiedliche Schwerpunkte setzen – von der Audiomanipulation über Gesichtsaustausch bis hin zur vollständigen Körperanimation. Nachfolgend ein Beispiel, die Text-to-Speech-Manipulation aus dem Bereich der Audiomanipulation:

Bei *Text-to-Speech* wandeln KI-Systeme geschriebenen Text in gesprochene Sprache um. Für eine authentische Stimmimitation genügen oft schon wenige kurze Sprachproben, die sich leicht aus sozialen Netzwerken oder anderen Onlineportalen beschaffen lassen. Ein anschauliches Beispiel dafür sind sogenannte „Schockanrufe“. Dabei geben sich Betrügerinnen oder Betrüger am Telefon als Vertrauenspersonen (z. B. Angehörige) aus und nutzen dazu KI-generierte Stimmsamples, um ihre Opfer zu täuschen (Häusler, 2024). Die Angreifer wandeln zuvor festgelegte Texte in scheinbar von der Vertrauensperson gesprochene Sprache um. So können sie nahezu jede gewünschte Aussage in der Stimme einer beliebigen Person erzeugen (BSI, 2025).



Bei der Herstellung dieses Bildes wurde am 15.04.2025 ChatGPT 4o eingesetzt (nach Bundesamt für Sicherheit in der Informationstechnik, 2025).

Angriff auf die Demokratie

Herausforderung

Deepfakes lassen sich heute mithilfe von KI-Methoden in kürzester Zeit erstellen und sind visuell wie akustisch so überzeugend, dass selbst Fachleute sie nur schwer von authentischem Material unterscheiden können. Jedem Fortschritt bei Deepfake-Detektionssystemen steht dabei ein ebenso großer Fortschritt neuer Manipulationstechniken gegenüber, sodass das Entlarven zu einem dauerhaften Katz-und-Maus-Spiel wird. Selbst spezialisierte Prüftools verfehlen unbekannte Varianten häufig, produzieren Fehlalarme und erreichen in der Praxis keine hundertprozentige Trefferquote (Müller, 2024).

Weil audiovisuelle Inhalte traditionell hohes Vertrauen genießen und sich viral verbreiten, erreichen Deepfakes die Öffentlichkeit oft lange vor Faktenchecks. Das nährt Befürchtungen vor politischem Missbrauch und einer Gefährdung demokratischer Prozesse (Pawelec, 2024).

Wahlbeeinflussung

Weltweit häufen sich Fälle, in denen Deepfakes in Wahlkämpfen eingesetzt werden – etwa diffamierende Videos in Argentinien (2023), ein Audio-Fake gegen den slowakischen Spitzenkandidaten Michal Šimečka kurz vor der Wahl (2023) oder sexualisierte Deepfakes, die den türkischen Präsidentschaftskandidaten Muharrem İnce zum Rückzug bewegten (2023). Auch bei den US-Wahlen 2024 kursierten Biden-Robocalls, Fake-Videos des Republican National Committee und KI-generierte Montagen von Kamala Harris. Obwohl nicht bekannt ist, wie viel Einfluss solche Deepfakes auf die tatsächlichen Wahlentscheidungen der Bürgerinnen und Bürger hatten, zeigen die Beispiele, wie leicht sich die Stimmung in den letzten Wahlkampftagen beeinflussen lässt (Pawelec, 2024).

Vertrauensverlust

Der dauerhafte Schaden entsteht weniger durch einzelne erfolgreiche Fälschungen als durch den schleichenden Vertrauensverlust: Je öfter Deepfakes bekannt werden, desto schneller zweifeln Bürgerinnen und Bürger Fotos, Videos oder Tonaufnahmen an – selbst, wenn sie echt sind. Gleichzeitig können Machthabende belastendes Material als angebliche Fälschung abtun („The Liar’s Dividend“), wie etwa Donald Trump 2024, als er ein Foto mit vielen pro-Harris-Anhängerinnen und -Anhängern als manipuliert bezeichnete. So wird die journalistische Überprüfbarkeit von Informationen und die Rechenschaftspflicht politischer Akteure massiv erschwert (Pawelec, 2024).

Ausländische Einflussnahme

Wiederholt aus dem Ausland verbreitete Deepfakes verunsichern Wählerinnen und Wähler und schwächen das Vertrauen in die Parteien. Seit dem russischen Angriff auf die Ukraine 2022 kursieren manipulierte Videos und Audiodateien in der militärischen wie geopolitischen Kommunikation; sowohl pro-russische Propagandanetzwerke als auch ukrainische Akteure nutzen diese Mittel. Ein vergleichbares Muster zeigt sich im Gaza-Konflikt, wo Deepfakes die ohnehin polarisierten Informationsräume zusätzlich destabilisieren (Pawelec, 2024).

Rufschädigung von Opposition und Zivilgesellschaft

Deepfakes dienen gezielt dem Rufmord in Politik und Journalismus. Ein bekanntes Beispiel ist die indische Journalistin Rana Ayyub, die 2018 durch manipuliertes Material diffamiert wurde. Solche Angriffe können ganze Bewegungen einschüchtern und kritische Stimmen verstummen lassen (Pawelec, 2024).

Kompromittierende Deepfakes zielen oft auf die Existenz von Frauen im öffentlichen Raum ab und darauf, sie aus dem öffentlichen Diskurs zurückzudrängen (Silencing). Dies führt zu einer geringeren Meinungsvielfalt im Netz und dazu, dass Frauen ihr Recht auf freie Meinungsäußerung nicht voll wahrnehmen können (Klicksafe 2025, S. 24).

Wirkung & Ausblick

Bislang lässt sich nur selten ein direkter Einfluss einzelner Deepfakes auf Wahlergebnisse nachweisen, doch ihre Gesamtwirkung schwächt das Vertrauen in Medien, Institutionen und demokratische Entscheidungsprozesse. Mit zunehmender Qualität und Verbreitung KI-basierter Werkzeuge geraten insbesondere regionale Wahlen stärker in Gefahr – vor allem dort, wo journalistische Korrekturmöglichkeiten begrenzt sind (Pawelec, 2024).

Quellen:

- Bundesamt für Sicherheit in der Informationstechnik (BSI). (15.4.2025). Deepfakes - Gefahren und Gegenmaßnahmen. Deutschland Digitale Sicher BSI. Verfügbar unter: https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Informationen-und-Empfehlungen/Kuenstliche-Intelligenz/Deepfakes/deepfakes_node.html, abgerufen am 23.03.2026
- Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI). (15.4.2025). *Was sind eigentlich DeepFakes, Herr Prof. Krüger?*. dfki ai. Verfügbar unter: <https://www.dfki.de/web/news-media/news-events/interview-was-sind-eigentlich-deepfakes>, abgerufen am 23.03.2026
- Müller, N. (5.12.2024). bpb: Bundeszentrale für politische Bildung. Technische Ansätze zur Deepfake-Erkennung und Prävention. Verfügbar unter: <https://www.bpb.de/lernen/bewegt-bild-und-politische-bildung/556855/technische-ansaeetze-zur-deepfake-erkennung-und-praevention/>, abgerufen am 23.03.2026
- Pawelec, M. (5.12.2024). Politische Manipulation und Desinformation. bpb: Bundeszentrale für politische Bildung. Verfügbar unter: <https://www.bpb.de/lernen/bewegt-bild-und-politische-bildung/556305/politische-manipulation-und-desinformation/>, abgerufen am 23.03.2026
- Klicksafe, Deep Fake. Deep Impact. Wie Jugendliche Deepfakes erkennen und ihre Folgen kritisch hinterfragen lernen (2025) https://www.klicksafe.de/fileadmin/cms/download/Material/P%C3%A4d._Praxis/Material_Paed-Praxis_Deep-Fake-Deep-Impact_klicksafe.pdf, abgerufen am 23.03.2026

Weiterführende und vertiefende Informationen:

- AK Bildung in der digitalen vernetzten Welt, in: klickpunkt.schule (2025). KI | Deepfakes. Verfügbar unter: <https://klickpunktschule.bycs.de/beitrag/ki-deepfakes>