Factsheet

Thema: Deepfakes

Was sind Deepfakes?

Deepfakes sind KI-generierte Bild-, Video- oder Audioaufnahmen, bei denen menschliche Gesichter, Körper oder Stimmen täuschend echt manipuliert oder synthetisch erzeugt werden. Der Begriff „Deepfake“ setzt sich aus „Deep Learning“ und „Fake“ zusammen und wurde Ende 2017 bekannt, als ein anonymer Reddit-Nutzer manipulierte pornografische Videos von Schauspielerinnen sowie den dazugehörigen Erstellungscode veröffentlichte. Deepfakes ermöglichen es, Personen in beliebigen Situationen darzustellen, beispielsweise in politischen, humorvollen oder pornografischen Inhalten, ohne deren Wissen oder Zustimmung (Pawelec & Bieß, 2021; Westerlund, 2019).

Abb. 1: https://spotintelligence.com/2025/05/08/deepfake/ (04.05.2026)

Wie entstehen Deepfakes?

Die technologische Grundlage von Deepfakes sind sogenannte Generative Adversarial Networks (GANs). Das sind zwei künstliche neuronale Netze, die in einem Wettbewerb miteinander trainiert werden. Dabei erstellt das eine Netz (Generator) gefälschte Inhalte, während das andere (Diskriminator) versucht, diese als unecht zu erkennen. Durch dieses Wechselspiel verbessern sich beide kontinuierlich.

Für die Erstellung von Deepfake-Bildern und -Videos werden große Mengen an Fotos oder Videos einer Person als Trainingsdaten benötigt. Je besser diese Datenbasis ist, desto realistischer ist anschließend auch das Ergebnis. Dank frei verfügbarer Software sinken die technischen Hürden zur Deepfake-Erstellung zunehmend, so dass Deepfakes heute auch ohne spezielles Fachwissen erstellt werden können (Pawelec & Bieß, 2021; Westerlund, 2019).

Chancen & Risiken

Deepfakes bieten durchaus legitime und innovative Einsatzmöglichkeiten. In der Filmbranche ermöglichen sie beispielsweise realistische Spezialeffekte oder die Darstellung historischer Persönlichkeiten. Auch im Bildungsbereich können sie Lerninhalte anschaulicher vermitteln, etwa durch interaktive Lernvideos. In Assistenztechnologien können synthetische Stimmen oder digitale Avatare Menschen unterstützen, die ihre Stimme verloren haben oder Sprachbeeinträchtigungen haben.

Gleichzeitig bringen Deepfakes aber auch erhebliche Risiken mit sich. Sie erleichtern Desinformationskampagnen und können das Vertrauen in Medien und den öffentlichen Diskurs schädigen. Das kann langfristig auch dazu führen, dass selbst echte Inhalte zunehmend angezweifelt werden.

Besonders problematisch ist der Missbrauch von Deepfakes für pornografische Inhalte (Deepnudes) sowie für Identitätsbetrug, bei dem Gesichter oder Stimmen von Personen genutzt werden, um sich online als diese auszugeben oder andere zu täuschen.

Studien zeigen, dass die große Mehrheit der Deepfakes im Internet pornografische Inhalte sind: Bereits eine frühe Analyse kommt zum Ergebnis, dass rund 96% aller Deepfake-Videos online pornografisch sind und überwiegend Frauen betreffen (Ajder et al., 2019).

Ein weiteres Problem ist, dass viele Menschen Deepfakes nur schwer von echten Inhalten unterscheiden können, besonders bei hoher Qualität. Obwohl ein großer Teil der Bevölkerung bereits mit Deepfakes in Kontakt gekommen ist, kennt nur rund die Hälfte den Begriff. Dies verstärkt die Risiken zusätzlich und zeigt, wie wichtig Medienkompetenz ist.

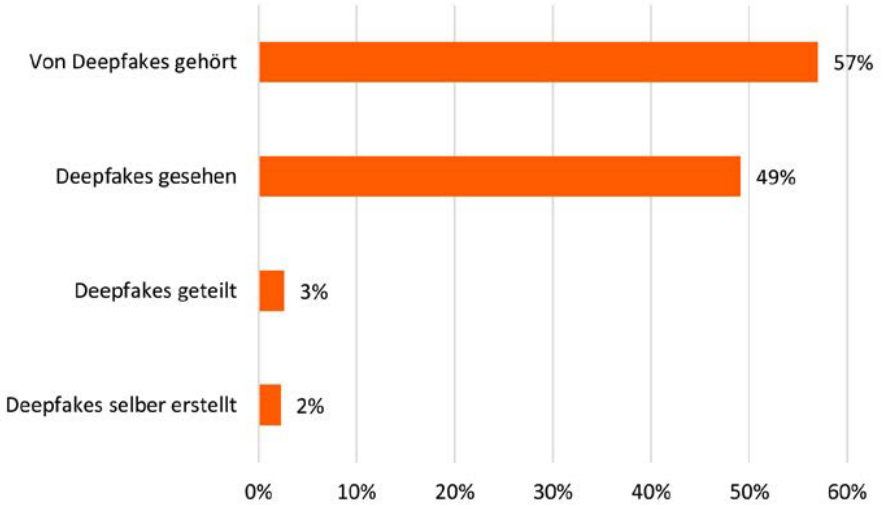

Abb. 2: Erfahrungen mit Deepfakes (prozentualer Anteil der Befragten, die dem jeweiligen Item zugestimmt haben). Vogler et al., 2024.

Deepfakes erkennen

Deepfakes zu erkennen wird zunehmend schwieriger, da sich die Technologie ständig weiterentwickelt. Hinweise auf manipulierte Inhalte können unter anderem sein:

- Unnatürliche Augenbewegungen/Mimik
- Unschärfen an Gesichtskonturen
- Inkonsistente Beleuchtung und Schatten
- Wackeln oder Verzerrungen der Haut
- Unstimmigkeiten zwischen Lippenbewegungen und Sprache

Auch KI-basierte Detektionstools können bei der Erkennung von Deepfakes unterstützen: Sie analysieren Videos Bild für Bild und werten z. B. biologische Merkmale wie Blutfluss-Veränderungen im Gesicht aus. Gleichzeitig verbessern Deepfake-Entwickler:innen ihre Methoden laufend weiter, wodurch die Erkennung von Deepfakes zunehmend anspruchsvoller wird. Zusätzlich ist es wichtig, Quellen zu überprüfen und Informationen mit vertrauenswürdigen Medien abzugleichen (Pawelec & Bieß, 2021; Westerlund, 2019).

Weitere Informationen in der Lernumgebung „Fake News“ und in der Lernumgebung & Webinar zu „Faktencheck im Internet“.

Regeln & Verantwortung

Die Regulierung von Deepfakes steht noch am Anfang. Auf EU-Ebene verpflichtet der „AI-Act“ zur Kennzeichnung KI-generierter Inhalte, während der „Digital Services Act“ insbesondere die Verantwortung von Plattformen betont. Plattformbetreiber gehen dabei oft strenger vor als gesetzlich vorgeschrieben und löschen Inhalte auf Grundlage eigener Ethikkodizes. Ein vollständiges Verbot der Technologie wäre jedoch problematisch, da es legitime Anwendungen, beispielsweise in Kunst, Forschung und Bildung, unterbinden würde. Wirksame Governance erfordert daher ein Zusammenspiel aus klarer gesetzlicher Regelung, Plattformverantwortung, technischer Deepfake-Erkennung, der Förderung professioneller Faktenprüfung sowie der Stärkung von Medienkompetenz (Pawelec & Bieß, 2021; Westerlund, 2019).

Quellen:

Ajder, H., Patrini, G., Cavalli, F., & Cullen, L. (2019). The state of deepfakes: Landscape, threats, and impact. Deeptrace.

Pawelec, M., & Bieß, C. (2021). Deepfakes: Technikfolgen und Regulierungsfragen aus ethischer und sozialwissenschaftlicher Perspektive. Nomos Verlagsgesellschaft mbH & Co. KG. https://doi.org/10.5771/9783748928072

Vogler, D., Rauchfleisch, A., & de Seta, G. (2024). Wahrnehmung von Deepfakes in der Schweizer Bevölkerung. In M. Karaboga, N. Frei, M. Puppis, D. Vogler, P. Raemy, F. Ebbers, G. Runge, A. Rauchfleisch, G. de Seta, G. Gurr, M. Friedewald, & S. Rovelli (Eds.), Deepfakes und manipulierte Realitäten: Technologiefolgenabschätzung und Handlungsempfehlungen für die Schweiz (pp. 125–151). vdf Hochschulverlag. https://vdf.ch/deepfakes-und-manipulierte-realitaten-e-book.html

Westerlund, M. (2019). The Emergence of Deepfake Technology: A Review. Technology Innovation Management Review, 9(11), 39–52. https://doi.org/10.22215/timreview/1282