



UNIVERSITÄT
LEIPZIG

Das **Deepfake** Projekt

**Zwölf Kernaussagen über die Wirkung
und Wahrnehmung von Deepfake-
Technologie aus fünf Jahren Forschung**

Prof. Dr. Alexander Godulla
Prof. Dr. Christian P. Hoffmann
Daniel Seibert, M.A.

deepfake-project.com

Studie 1: Systematische Literaturanalyse und Status Quo der Deepfake-Forschung

1.

Forschung zum Thema Deepfakes ist vorrangig geprägt von Studien aus der Informatik, die sich mit der Entwicklung von Deepfake-Erkennungstools beschäftigen, und Arbeiten aus der Rechtswissenschaft, die sich der Entwicklung von Regulierungen zum Umgang mit schädlichen Deepfakes widmen.

2.

Mit steigender Bekanntheit und Verbreitung der Technologie befassen sich auch mehr geistes- und sozialwissenschaftliche Studien mit dem Phänomen und untersuchen beispielsweise die Auswirkungen von Deepfakes auf das Publikum sowie dessen Fähigkeiten, Deepfakes zu erkennen.

Studie 2: Quantitative Bevölkerungsbefragung zu Deepfakes in Deutschland

3.

Zum Zeitpunkt der Studie (05.-11.10.2022) gaben die meisten Befragten (n = 1.421) an, noch keinen Kontakt mit Deepfakes gehabt zu haben.

75,5 % der Befragten glauben nicht, dass sie in den letzten 12 Monaten auf ein Deepfake gestoßen sind;
24,5 % der Befragten geben an, dass sie bereits einem Deepfake online begegnet sind.

4.

Der Großteil der Befragten weist kein oder lediglich geringes Wissen über Deepfakes auf und ist sich der negativen, aber auch positiven Einsatzmöglichkeiten der Technologie nicht bewusst.

Variable		Anteil in %
Deepfake Wissensstand	Kein Vorwissen	67,1
	Niedrig	11,3
	Mittel	17,4
	Hoch	4,2

Hinweis: Der Wissensstand über Deepfakes wurde in Form von 13 Wissensfragen erfasst. Kein Vorwissen über Deepfakes umfasst Befragte, die im Fragebogen angaben, nicht zu wissen, was ein Deepfake ist sowie Befragte, die keine der Quizfragen richtig beantworteten. Die richtige Beantwortung von ein bis vier Fragen wurde als niedriger Wissensstand, fünf bis sieben richtige Antworten als mittlerer Wissensstand und acht oder mehr richtige Antworten als hoher Wissensstand kodiert.

5.

Es zeigen sich Charakteristika *Digitaler Ungleichheiten* im Wissen über Deepfake-Technologie und deren Einschätzung.

- (a) Unsere Studie deutet auf soziodemografische Unterschiede hin: Jüngere Personen, jene mit einem hohen Bildungsstand und männlichen Geschlechts weisen eher ein Verständnis von Deepfakes, ihrer Potenziale und Einsatzbereiche auf.
- (b) Die Ergebnisse unserer Studie zeigen außerdem, dass die digitalen Fähigkeiten („digital skills“) einer Person einen starken Zusammenhang mit dem Wissen über die Technologie haben.
- (c) Der Kenntnisstand über Deepfakes hat im Umkehrschluss einen erheblichen Zusammenhang mit der Wahrnehmung von Risiken und Potenzialen der Technologie. Ein differenziertes Verständnis der Deepfake-Technologie könnte demnach die Wahrscheinlichkeit erhöhen, persönliche Vorteile aus deren Nutzung zu ziehen, während das Verständnis ihrer Gefahren die Fähigkeit stärken könnte, persönliche Risiken der Technologie zu mindern.

6.

Die meisten Befragten empfinden das Aufkommen von Deepfakes als große Gefahr.

Variable		Anteile in %
Wahrgenommene Risiken	Niedrig	5,8
	Mittel	24,3
	Hoch	69,9
Wahrgenommene Potenziale	Niedrig	26,3
	Mittel	41,9
	Hoch	31,7

Hinweis: Die Einstellungen der Befragten zu Deepfakes wurden mittels 21 Aussagen (zwölf Risiken, neun Chancen) über Einsatzbereiche der Technologie gemessen. Die Befragten wurden gebeten, ihre Zustimmung zu diesen Aussagen zu bewerten („stimme überhaupt nicht zu“ bis „stimme voll und ganz zu“). Eine niedrige Gefahren-Wahrnehmung umfasst Befragte, die keinen der Risiko-Aussagen zustimmten, eine mittlere Gefahren-Wahrnehmung umfasst die Zustimmung von einer bis sechs dieser Aussagen und eine hohe Gefahren-Wahrnehmung die Zustimmung von sieben bis zwölf dieser Aussagen. Eine niedrige Potenzial-Wahrnehmung umfasst Befragte, die keinem der Potenziale zustimmten, eine mittlere Potenzial-Wahrnehmung umfasst die Zustimmung von einer bis vier der Potenzial-Aussagen und eine hohe Potenzial-Wahrnehmung die Zustimmung von fünf bis neun dieser Aussagen.

Ich mache mir Sorgen, ...	Stimme eher nicht oder überhaupt nicht zu (%)	Teils/teils (%)	Stimme eher oder voll und ganz zu (%)
auf ein Deepfake hereinzufallen.	6,6	26,1	67,3
selbst Opfer eines schädlichen Deepfakes zu werden (z.B. pornografischer Deepfake über mich, Deepfake-Betrug am Telefon).	18	31,3	50,7
dass Deepfakes die öffentliche Meinung beeinflussen (z.B. indem sie in Wahlkampagnen genutzt werden).	3,6	16,9	79,5
dass ich bald keinen Videoinhalten mehr trauen kann.	6,7	25,8	67,5
dass die Grenze zwischen Realität und Fiktion verschwimmt.	3,9	18,6	77,5
dass Deepfakes unserer Demokratie schaden (z.B. durch die Verbreitung von Desinformationen/Fake News).	4,2	18,7	77,1
dass jede Person in der Lage ist, Deepfakes zu erstellen.	12,9	29,4	57,7
dass die Manipulation meiner persönlichen Daten immer einfacher wird (z.B. von Audio-, Bild- oder Videodateien von mir).	6,1	20,6	70,4
meine persönlichen Daten künftig noch stärker schützen zu müssen (z.B. weil diese zu betrügerischen Zwecken manipuliert werden könnten).	5,2	20,7	74,1
selbst aus Versehen Deepfakes zu verbreiten (z.B. in Sozialen Medien).	25,8	31,9	42,3
dass das Erkennen von Deepfakes immer schwieriger wird.	2,7	10,7	86,6
dass die Verbreitung von Deepfakes immer weiter zunimmt.	3	16,7	80,3

7.

Vorteile und Potenziale von Deepfakes werden von den Befragten nur bedingt erkannt.

Aussage	Stimme eher nicht oder überhaupt nicht zu (%)	Teils/teils (%)	Stimme eher oder voll und ganz zu (%)
Ich bin der Meinung, dass Deepfakes auch für sinnvolle Zwecke genutzt werden können.	18,1	37,1	44,8
Meiner Meinung nach können mir Deepfakes auch Vorteile bieten (z.B. indem ich lustige Videos von mir und meinen Freunden/Freundinnen erstellen kann).	27,2	39,9	32,9
Ich sehe Vorteile darin, dass jede Person in der Lage ist, Deepfakes zu erstellen.	44,5	38,0	17,5
Der Einsatz von Deepfakes kann Vorteile für die Filmindustrie bieten (z.B. indem Action-Szenen realistischer dargestellt oder bereits verstorbene Personen in Filme integriert werden können).	8,5	28,5	63
Es hat auch Vorteile, wenn mit Deepfakes unter Zustimmung der Beteiligten pornografische Inhalte erstellt werden können (z.B. indem Nutzer/Nutzerinnen freiwillig Fotos von sich in einer App hochladen und zur Erstellung von pornografischen Inhalten freigeben).	30,9	41	28,1
Der Einsatz von Deepfakes kann Vorteile für die Kunstbranche bieten (z.B. für die Kreierung neuer Kunstwerke im Stil berühmter Künstler/Künstlerinnen).	19	39,5	41,5
Der Einsatz von Deepfakes kann Vorteile für die Modeindustrie bieten (z.B. für die virtuelle Anprobe von Kleidung).	12,8	35,6	51,6
Der Einsatz von Deepfakes kann Vorteile für die Videospelindustrie bieten (z.B. durch den Einsatz personalisierter Avatare im Spiel).	12	32,6	55,4
Ich finde, Deepfakes können die Kreativität in der Produktion von Medieninhalten fördern.	20,5	41,5	38

8.

Die Mehrheit der Befragten fühlt sich nicht für den Umgang mit Deepfakes gewappnet.

- (a) Die meisten Befragten schätzen ihre eigenen Fähigkeiten, Deepfakes erkennen und von realen Videos unterscheiden zu können, als sehr gering ein.
- (b) Die Mehrheit der Befragten befürchtet, auf den Inhalt eines Deepfakes hereinfallen zu können.
- (c) Lediglich ein Drittel der Befragten ist sich sicher, das Internet und Fact-checking-Seiten nutzen zu können, um ein Deepfake als solches zu entlarven.
- (d) Befragte, die sich durch eine häufige Nutzung des Internets und Sozialer Medien auszeichnen und ein hohes Interesse an neuen Technologien haben, schätzen Ihre Fähigkeiten, Deepfakes erkennen zu können, höher ein als Befragte, die weniger aktiv in der digitalen Welt sind und sich weniger für neue Technologien interessieren.

Selbstwirksamkeit der Erkennung von Deepfakes		Häufigkeit (%)
Ich bin zuversichtlich, dass ich in der Lage bin, einen Deepfake selbst zu erkennen.	Stimme eher nicht oder überhaupt nicht zu	62,9
	Teils/teils	27,2
	Stimme eher oder voll und ganz zu	9,9
Ich bin zuversichtlich, dass ich zwischen einem echten Video und einem Deepfake-Video unterscheiden kann.	Stimme eher nicht oder überhaupt nicht zu	60,8
	Teils/teils	27,0
	Stimme eher oder voll und ganz zu	12,2
Ich bin zuversichtlich, dass ich ein Deepfake-Video erkennen kann, indem ich im Internet nach Informationen suche (z.B. Fact-Checking-Seiten).	Stimme eher nicht oder überhaupt nicht zu	35,3
	Teils/teils	32,7
	Stimme eher oder voll und ganz zu	32,0
Ich bin zuversichtlich, dass ich nicht auf den Inhalt eines Deepfake-Videos hereinfallen würde.	Stimme eher nicht oder überhaupt nicht zu	53,8
	Teils/teils	29,2
	Stimme eher oder voll und ganz zu	17,0

9.

Die Wahrnehmung der Befragten, vermehrt online Deepfakes ausgesetzt zu sein, verstärkt bei ihnen ein Gefühl von Medienezynismus – eine unsachliche, überzogene Medienkritik mit verschwörungstheoretischen Tendenzen.

Befragte, die ihre Fähigkeiten, Deepfakes erkennen zu können, höher einschätzen, entwickeln eher ein Gefühl von Medienezynismus, wenn sie auf Deepfakes stoßen. Es ist möglich, dass diejenigen, die glauben, die neue Deepfake-Technologie besser zu verstehen, sich ihrer potenziellen (negativen) Auswirkungen eher bewusst sind oder sich mehr Sorgen machen (z.B. über Missbrauch der Technologie durch mächtige – politische oder mediale – Institutionen).

Studie 3: Qualitative Experteninterviews in den Feldern Journalismus und Technologie in Deutschland

10.

Die Wahrnehmung der Deepfake-Technologie unter Expertinnen und Experten in den Feldern Journalismus und Technologie (n = 20) ist überwiegend von Risiken für die journalistische Praxis geprägt (Befragungszeitraum 09/2022-05/2023).

Herausforderungen im Umgang mit der Technologie beziehen sich auf zukünftig notwendige intensivere Factchecking-Prozesse und sind von Diskussionen über die Vereinbarkeit der Technologie mit journalistischen Grundprinzipien gekennzeichnet (Genauigkeit, Wahrhaftigkeit, Transparenz).

11.

Die Potenziale der Deepfake-Technologie für die journalistische Praxis umfassen aus Sicht der Befragten in erster Linie Möglichkeiten der Personalisierung von Inhalten sowie einfache(re) Formen der Erstellung und Bearbeitung von journalistischen Inhalten.

Deepfakes werden überwiegend als unterstützende Technologie wahrgenommen, die alltägliche Arbeitsprozesse, wie z.B. die Bearbeitung und Visualisierung von Inhalten beschleunigen kann. Die Technologie erhält weniger Zuspruch hinsichtlich der Entwicklung neuer, komplexer, journalistischer Formate und Produkte. Das Potenzial der Technologie wird in der Automatisierung personalisierter, Standort- und Interessenbezogener Inhalte verortet (z.B. Verkehr, Wetter, Sport).

12.

Bislang mangelt es in Medienorganisationen und in der journalistischen Praxis an Leitlinien und strategischen Ansätzen für den Umgang mit den Risiken und Potenzialen von Deepfake-Technologie.

- (a) Wissen über Deepfakes und Kompetenzen im Umgang mit Technologien, die auf Künstlicher Intelligenz basieren, müssen gefördert werden, um ein breites Verständnis für die Vorteile und Nachteile der Deepfake-Technologie zu etablieren.
- (b) Leitlinien zum Umgang mit Deepfakes sollten ein Verständnis der Grenzen von Realität und Fiktion in der Darstellung (insbesondere visueller) Inhalte, die mittels Deepfake-Technologie generiert wurden, etablieren. So kann erkannt werden, welche synthetischen Inhalte und Darstellungsformen sich unter Berücksichtigung journalistischer Grundprinzipien überhaupt für die Praxis eignen.
- (c) Es sollten Abläufe entwickelt und Verantwortlichkeiten geklärt werden, die bei der Erkennung potenziell schädlicher Deepfakes berücksichtigt werden.