fög

Forschungszentrum Öffentlichkeit und Gesellschaft

Jahrbuch Qualität der Medien Studie 3/2024

Deepfakes: Medienberichterstattung und Wahrnehmung in der Schweizer Bevölkerung

Universität Zürich[UZH]

fög – Forschungszentrum Öffentlichkeit und Gesellschaft/Universität Zürich

Inhalt:

Deepfakes: Medienberichterstattung und Wahrnehmung in der Schweizer Bevölkerung

Daniel Vogler, Adrian Rauchfleisch

Zusammenfassung

In unserer Studie untersuchten wir die Medienberichterstattung über Deepfakes und wie die Nutzung von journalistischen Medien mit der Wahrnehmung von Deepfakes in der Schweizer Bevölkerung zusammenhängt. Damit konnten wir die Frage beantworten, wie Medien und Gesellschaft mit einem neuartigen Kommunikationsphänomen umgehen, das von Künstlicher Intelligenz (KI) geprägt ist. In einem ersten Schritt führten wir eine Inhaltsanalyse der Berichterstattung über Deepfakes in Schweizer Onlinemedien durch. In einem zweiten Schritt analysierten wir anhand von Daten einer repräsentativen Befragung unter Schweizer:innen, ob die Wahrnehmung von Deepfakes mit der Nutzung journalistischer Medien, Social Media und Videoplattformen zusammenhängt. Die Resultate der Inhaltsanalyse zeigen, dass das Thema Deepfakes in Schweizer Medien im Jahr 2023 zwar deutlich an Resonanz gewonnen hat, aber nach wie vor ein Nischenthema ist. Die Medienberichterstattung fokussiert stark auf Risiken wie Desinformation und weniger auf Chancen wie Anwendungen im Film- und Musikbereich. Die Befragung hat gezeigt, dass die Nutzung journalistischer Medien positiv mit der Bekanntheit von Deepfakes korreliert. Menschen, die häufig journalistische Medien konsumieren, kennen den Begriff also eher als Menschen, die dies wenig tun. Die Nutzung von Videoplattformen korreliert hingegen mit der direkten Erfahrung, sprich dem Sehen von Deepfakes, sowie dem Wissen über Deepfakes. Wer journalistische Medien nutzt, geht auch eher davon aus, dass Deepfakes die Meinung der Schweizer Bevölkerung beeinflussen. Dies ist möglicherweise ein Effekt der stark negativen Berichterstattung in Schweizer Medien mit dem Fokus auf Desinformation. Allerdings gehen Menschen mit hohem Medienvertrauen von einem geringeren Effekt von Deepfakes auf die Bevölkerungsmeinung aus. Auch wenn wir mit unserem Untersuchungsdesign keinen direkten Einfluss der Medienberichterstattung über Deepfakes auf die Wahrnehmung von Deepfakes in der Schweizer Bevölkerung nachweisen können, verdeutlichen die Ergebnisse der Studie trotzdem die ambivalente Rolle journalistischer Medien: Einerseits sorgen Informationen und Einordnungen von Journalist:innen dafür, dass Menschen über Anwendungen, Funktionsweisen, Chancen und Risiken von Technologien wie KI informiert sind. Andererseits kann eine eher einseitig auf Risiken fokussierte Berichterstattung auch Unsicherheit in der Bevölkerung wecken und die Nutzung von Potenzialen neuer Technologien reduzieren.

1 Einleitung

Deepfakes sind synthetische Bild-, Audio- oder Videoinhalte, die mit KI hergestellt oder verändert wurden (Westerlund, 2019). Das Wort Deepfake setzt sich aus den Begriffen Deep Learning, einem Teilbereich der KI, und dem Wort Fake zusammen. Der Begriff Deepfake tauchte Ende 2017 erstmals in einem Forum auf dem sozialen Netzwerk Reddit auf. Ein anonymer Nutzer mit dem Pseudonym Deepfake hatte mittels KI-Verfahren pornografische Videos mit den Bildern von prominenten Schauspielerinnen versehen und veröffentlicht. Pornografie war und ist vermutlich der grösste Anwendungsbereich von Deepfake-Technologie (Gosse & Burkell, 2020). Eine Studie der auf Deepfake-Erkennung spezialisierten Firma Deeptrace aus dem Jahr 2019 zeigt etwa, dass es sich bei 96% der Deepfakes um Pornografie handelt (Ajder et al., 2019).

Der Innovationsschub bei der KI-Technologie im Bereich der Large Language Models (LLM) wie GPT-3 verschärft auch die Problematik von Deepfakes (Sison et al., 2023). Mit KI-Anwendungen wie Midjourney und Dall-E können Nutzer:innen über Programmierbefehle, sogenannte Prompts, praktisch jedes beliebige Bild erstellen. Zwei der bekanntesten Beispiele von Deepfakes, die mit dem Tool Midjourney produziert wurden, zeigen Papst Franziskus in einer weissen Daunenjacke und die angebliche Verhaftung von Donald Trump (Klaus, 2023). Mit Sora, dem neusten Tool von OpenAI, können über Prompts sogar Videos kreiert werden. Mit diesen

Anwendungen wird KI-Technologie für mehr Menschen zugänglich und es wird einfacher, qualitativ gute Deepfakes herzustellen. Damit verbunden ist die Sorge, dass mit Deepfakes ein grosses Manipulationspotenzial einhergeht. Vogler et al. (2024) haben jüngst in einem Onlineexperiment gezeigt, dass Menschen in der Schweiz kaum mehr in der Lage sind, gut gemachte Deepfakes von realen Videos zu unterscheiden. Auch eine kurze Hilfestellung zur Erkennung von Deepfakes, die der Hälfte der Proband:innen gezeigt wurde, hatte keinen Effekt auf das Erkennen von Deepfakes. Allerdings führte diese Intervention auch nicht dazu, dass die Befragten die realen Videos überproportional oft als Deepfakes bezeichneten. Kompetenz im Umgang mit Social Media und wenn die Teilnehmer:innen wussten, wer die Person im gezeigten Video war, hatten einen Einfluss auf die Erkennungskompetenz.

Folglich sind Deepfakes in jüngerer Zeit mit Risiken in Verbindung gebracht und insbesondere im Kontext von Desinformation diskutiert worden (Ahmed, 2021; Hameleers et al., 2022; Vaccari & Chadwick, 2020). Deepfakes werden auch für kriminelle Aktivitäten wie Betrug oder Diebstahl eingesetzt. Kriminelle verwenden insbesondere Audio-Deepfakes, um das Handeln von Menschen zu beeinflussen, auch in der Schweiz (Bundesamt für Cybersicherheit, 2024). Beispielsweise werden Anrufe von Vorgesetzten oder Verwandten vorgetäuscht, um Personen zur Überweisung von Geld oder vertraulichen Informationen zu bewegen. Deepfake-Technologie bietet aber auch Vorteile. Deepfakes können als Kunstform oder zur Unterhaltung kreativ eingesetzt werden. Chancen bieten sich insbesondere auch im Bildungsbereich, in der Film-, Musik- und Werbeindustrie (Karaboga et al., 2024).

Vieldiskutierte Beispiele von Deepfakes sind jüngst im Kontext der Kriege in der Ukraine und in Gaza aufgetaucht (Ho Tran, 2023; Rubiera, 2023). Auch in der politischen Kommunikation spielen Deepfakes eine Rolle, insbesondere während Wahlen und Abstimmungen. Viele bekannte Beispiele zeigen die ehemaligen US-Präsidenten Barack Obama oder Donald Trump. In der Schweiz fielen ebenfalls erste Deepfakes im Kontext des Wahljahres 2023 auf. Das bekannteste und meistdiskutierte Beispiel war ein Deepfake-Video der grünen Nationalrätin Sibel Arslan, das durch eine Agentur erstellt und von SVP-Nationalrat Andreas Glarner veröffentlicht worden war. Im Video äusserte sich die gefälschte Sibel Arslan negativ über Migrant:innen und forderte die Ausschaffung von kriminellen Ausländer:innen, was der politischen Positionierung der echten Sibel Arslan diametral widerspricht. Das Video wurde aufgrund des unplausiblen Inhalts und der schlechten Qualität schnell als Deepfake erkannt. Gemäss Nationalrat Glarner sei die Aktion ein Scherz gewesen (Rosch, 2024). Auch wenn Deepfakes in den Schweizer Wahlen kaum eine Rolle spielten, zeigt das Beispiel exemplarisch die Risiken der Technologie für den politischen Prozess.

Ob eine Technologie als Risiko oder Chance wahrgenommen wird, hängt massgeblich davon ab, wie in der Öffentlichkeit über sie diskutiert wird. Dabei spielen Debatten in Social-Media- oder auf Videoplattformen eine wichtige Rolle, weil viele Menschen auf diesen Kanälen mit Deepfakes in Kontakt kommen, z. B. wenn sie Videos von Bekannten erhalten und diese dann auch selber weiterverbreiten (Hameleers et al., 2022; Vaccari & Chadwick, 2020). Von grosser Bedeutung in der Wahrnehmung von neuen Technologien sind aber nach wie vor journalistische Medien (Gosse & Burkell, 2020). Mit ihren Informationen und Einordnungen können sie dazu beitragen, dass Menschen über Chancen, Risiken und Funktionsweisen von Deepfakes Bescheid wissen. Wenn in der Berichterstattung zu Deepfakes hingegen vornehmlich Risiken wie Desinformation oder Kriminalität thematisiert werden, ist es wahrscheinlich, dass auch beim Publikum eine negative Wahrnehmung überwiegt. Tatsächlich zeigen internationale Studien, dass Desinformation stark im Fokus der Medienberichterstattung über Deepfakes steht (Gosse & Burkell, 2020; Wahl-Jorgensen & Carlson, 2021). Auch Journalist:innen in der Schweiz verbinden Deepfakes vor allem mit Gefahren im Zusammenhang mit Desinformation. Sie sehen deshalb ihre Aufgabe bezüglich Deepfakes insbesondere in der Sensibilisierung des Publikums und der Förderung von Medienkompetenz (Raemy et al., 2024).

Noch wissen wir aber wenig darüber, wie journalistische Medien in der Schweiz über Deepfakes berichten und welche Zusammenhänge zwischen Mediennutzung und Wahrnehmung von Deepfakes bestehen. In unserer Studie analysierten wir zuerst mit einer quantitativen Inhaltsanalyse, wie Medien

in der Schweiz über Deepfakes berichten, welche Beachtung Deepfakes erhalten, wie sie bewertet werden, in welchem thematischen Kontext über sie berichtet wird und ob in den Beiträgen eine, zumindest rudimentäre, Definition des Phänomens enthalten ist. In einem zweiten Schritt ermittelten wir auf Basis einer repräsentativen Befragung, ob in der Schweizer Bevölkerung die Nutzung von journalistischen Medien, Social Media und Videoplattformen mit der Wahrnehmung von Deepfakes zusammenhängen. Konkret untersuchten wir den Zusammenhang zwischen der Mediennutzung und der Bekanntheit sowie dem Sehen von Deepfakes, dem Wissen über Deepfakes und dem wahrgenommenen Einfluss von Deepfakes auf die Meinung der Bevölkerung.

2 Methode

Für die vorliegende Studie wurden eine quantitative Inhaltsanalyse von Schweizer Medien und eine repräsentative Befragung der Schweizer Bevölkerung durchgeführt.

2.1 Quantitative Inhaltsanalyse

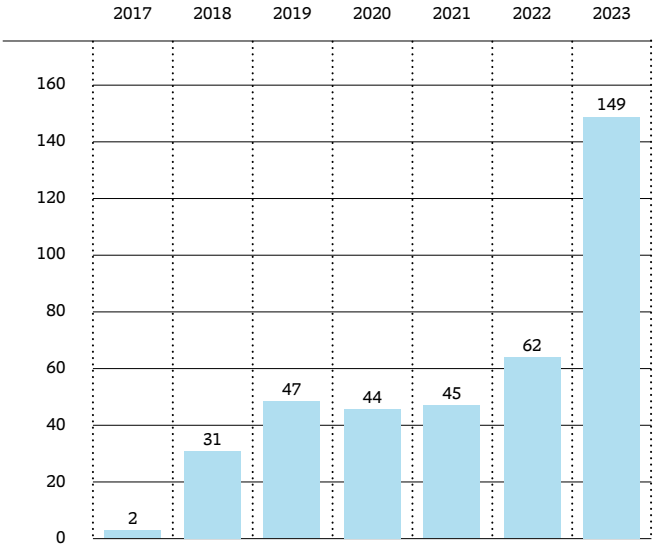
In einem ersten Schritt wurde die Berichterstattung über Deepfakes von elf Schweizer Onlinemedien aus der Deutschschweiz und der Suisse romande im Zeitraum vom 13. Dezember 2017 bis 31. Dezember 2023 mit einer quantitativen Inhaltsanalyse untersucht. Das Mediensample bestand für die Deutschschweiz aus 20minuten.ch, aargauerzeitung.ch, blick.ch, nzz.ch, srf.ch/news und tagesanzeiger.ch. Für die Suisse romande wurden 20minutes.ch, 24heures.ch, lematin.ch, letemps.ch und rts.ch/news berücksichtigt. Insgesamt wurden 380 Medienbeiträge ermittelt, die den Begriff Deepfake – in unterschiedlichen Schreibweisen – enthielten. Als Startpunkt der Untersuchung wurde das Erscheinungsdatum des ersten Artikels zum Thema Deepfake in den untersuchten Medien gewählt. Für die Beiträge wurde erhoben, in welchem thematischen Kontext über Deepfakes berichtet wurde, wie Deepfakes bewertet wurden und ob die Beiträge eine rudimentäre Definition des Phänomens Deepfakes beinhalten. Um als Definition zu gelten, musste explizit oder aus dem Kontext des Beitrags ersichtlich sein, dass es sich bei Deepfakes um Bild-, Audio- oder Videoinhalte handelt, die mit Verfahren der Künstlichen Intelligenz hergestellt oder verändert wurden.

2.2 Repräsentative Befragung

In einem zweiten Schritt wurden Daten einer repräsentativen Befragung ausgewertet. Diese waren im Rahmen einer Studie des fög erhoben worden, die von der Schweizerischen Stiftung für Technologiefolgenabschätzung TA-Swiss gefördert wurde (Vogler et al., 2024). Dazu wurde im September 2023 eine Onlineumfrage unter 1359 Schweizer:innen durchgeführt. Grundgesamtheit war die Schweizer Wohnbevölkerung aus der Deutschschweiz und der Suisse romande, die das Internet nutzte und zwischen 16 und 74 Jahre alt war. Das Ziel bestand darin, die Wahrnehmung von Deepfakes in der Schweizer Bevölkerung zu ermitteln: ob die Menschen den Begriff kennen und ob sie schon ein Deepfake gesehen hatten. Weiter fragten wir, ob sie Deepfakes definieren können – eine Antwort war korrekt, wenn aus ihr hervorging, dass es sich bei Deepfakes um Bild-, Audio- oder Videoinhalte handelt, die mit KI erstellt oder verändert wurden. Ausserdem wollten wir von den Menschen wissen, welchen Effekt Deepfakes ihnen zufolge auf die Meinung der Schweizer Bevölkerung haben. Zusätzlich wurden die Mediennutzung, das Vertrauen in die Medien sowie soziodemografische Angaben zu Alter, Geschlecht und Bildung erfasst.

3 Resultate

Im folgenden Kapitel werden die Resultate der Inhaltsanalyse der Medienberichterstattung über Deepfakes sowie Befunde aus der repräsentativen Befragung zur Wahrnehmung von Deepfakes präsentiert.

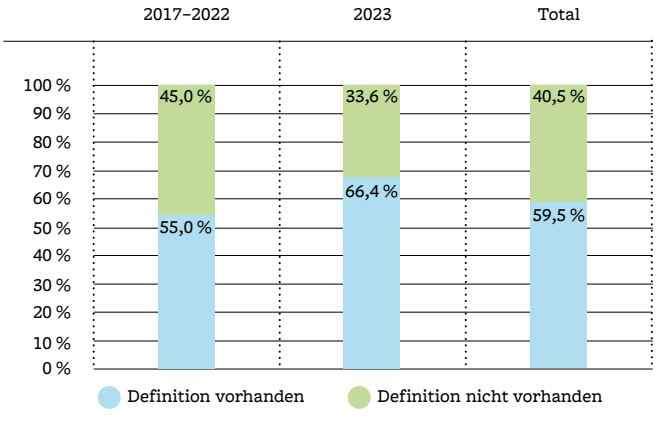


Darstellung 1: Anzahl Medienbeiträge zu Deepfakes zwischen 2017 und 2023

Die Darstellung zeigt die Anzahl der Medienbeiträge (Inhaltsanalyse von elf Schweizer Onlinemedien; n = 380) zum Thema Deepfakes, die pro Jahr erschienen.
Lesebeispiel: 2023 erschienen 149 Medienbeiträge zu Deepfakes. 2018 waren es 31 Beiträge gewesen.

3.1 Berichterstattung über Deepfakes in Schweizer Medien

Zunächst werteten wir die Anzahl der Medienbeiträge über die Zeit aus (vgl. Darstellung 1): Im Jahr 2023 erschienen mehr als doppelt so viele Beiträge wie jeweils in den Jahren 2017–2022. Auch für die Berichterstattung über Deepfakes scheint der Innovationsschub in der KI-Technologie im Bereich der Large Language Models (LLM), für die exemplarisch die Lancierung des Chatbots GPT-3 Ende 2022 steht, eine Zäsur darzustellen. Für die weiteren Analysen der Medienberichterstattung unterschieden wir deshalb zwei Zeiträume: Beiträge, die zwischen 2017 und 2022 (n = 231 Beiträge), und Beiträge, die 2023 (n = 149 Beiträge) publiziert wurden. Die insgesamt eher tiefe Resonanz zeigt aber auch, dass es sich bei Deepfakes noch um ein Nischenthema handelt.



Darstellung 2: Anzahl Medienbeiträge mit Definition von Deepfakes

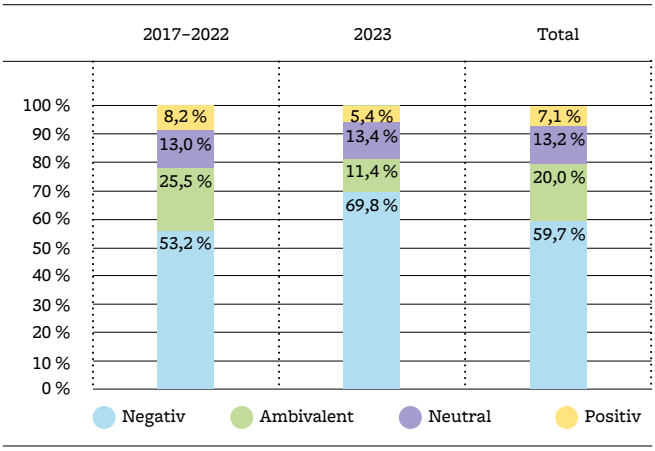
Die Darstellung zeigt den Anteil der Medienbeiträge (Inhaltsanalyse von elf Schweizer Onlinemedien; n = 380), die mindestens eine rudimentäre Definition von Deepfakes enthalten, für die Jahre 2017–2022, das Jahr 2023 und den gesamten Zeitraum.
Lesebeispiel: Von 2017 bis 2022 enthielten 55,0 % der Beiträge eine Definition von Deepfakes. 2023 waren es 66,4 %.

3.1.1 Definitionen von Deepfakes

Journalistische Medien übernehmen eine wichtige Rolle bei der Vermittlung von Wissen über neue Technologien wie beispielsweise KI. Deshalb ist es wichtig, dass Medien nicht nur über Chancen und Risiken berichten, sondern auch Informationen zu den Technologien und den Prozessen vermitteln. Wir untersuchten, ob Medien in ihrer Berichterstattung über Deepfakes mindestens eine rudimentäre Definition des Phänomens anbieten. In über der Hälfte der Medienbeiträge (59,5 %) wurde erwähnt, dass es sich bei Deepfakes um Bild-, Audio- oder Videoinhalte handelt, die mittels KI erstellt oder verändert wurden (vgl. Darstellung 2). Dies wurde entweder explizit so benannt, oder es ging aus dem Kontext hervor. Der Anteil von Beiträgen mit einer solchen Definition nahm 2023 von 55,0 % auf 66,4 % zu.

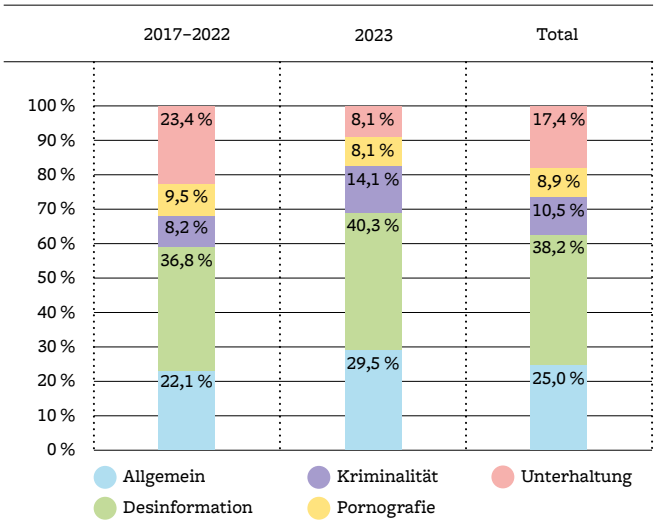
3.1.2 Bewertung von Deepfakes

Wie Technologien von der Bevölkerung wahrgenommen werden, hängt auch davon ab, wie über diese in der Öffentlichkeit gesprochen wird.



Darstellung 3: Bewertung von Deepfakes in der Medienberichterstattung

Die Darstellung zeigt, wie Deepfakes in der Medienberichterstattung (Inhaltsanalyse von elf Schweizer Onlinemedien; n = 380) bewertet werden, und zwar für die Jahre 2017–2022, das Jahr 2023 und den gesamten Zeitraum.
Lesebeispiel: Von 2017 bis 2022 enthielten 53,2 % der Beiträge eine negative Bewertung von Deepfakes. 2023 waren es 69,8 %.



Darstellung 4: Thematischer Kontext in der Medienberichterstattung über Deepfakes

Die Darstellung zeigt den thematischen Kontext in der Medienberichterstattung (Inhaltsanalyse von elf Schweizer Onlinemedien; n = 380) über Deepfakes für die Jahre 2017–2022, das Jahr 2023 und den gesamten Zeitraum.
Lesebeispiel: Von 2017 bis 2022 fokussierten 36,8 % der Medienbeiträge über Deepfakes auf das Thema Desinformation. 2023 waren es 40,3 %.

Wir untersuchten, wie Deepfakes in der Medienberichterstattung in der Schweiz bewertet werden. Wenig überraschend stehen bei Deepfakes primär Risiken im Vordergrund und es dominiert eine klar negative Perspektive auf die Technologie (vgl. Darstellung 3). Mehr als die Hälfte (59,7 %) der Beiträge zu Deepfakes besitzen eine negative Tonalität. In jedem fünften Beitrag (20,0 %) wird ambivalent über Deepfakes berichtet, sprich positive und negative Aspekte werden im Beitrag etwa gleichgewichtig thematisiert. In 13,2 % der Berichterstattung werden Deepfakes neutral dargestellt. Lediglich 7,2 % der Beiträge zeigen Deepfakes in einem positiven Licht. Die primär negative Berichterstattung mag sicherlich auch mit der Bezeichnung Deepfake selber zusammenhängen, die vermutlich über das Wort «Fake» per se negative Assoziationen weckt. Die Risiken, sprich eine negative Tonalität, stehen in jüngerer Zeit verstärkt im Zentrum. Im Vergleich zu den Jahren 2017–2022 stieg der Anteil negativer Beiträge im Jahr 2023 von 53,2 % auf 69,8 %, während der Anteil der Beiträge mit positiver Tonalität leicht von 8,2 % auf 5,4 % sank. Es wird auch deutlich weniger mit ambivalenter Tonalität über Deepfakes berichtet. Der Anteil nahm deutlich von 25,5 % auf 11,4 % ab. In der Berichterstattung scheint es einen wachsenden Konsens zu geben, dass Deepfakes insbesondere mit Risiken behaftet sind.

3.1.3 Thematischer Kontext

Die Risikofokussierung in der Berichterstattung zu Deepfakes zeigt sich auch an den erfassten Themen (vgl. Darstellung 4). Im Zentrum steht der Aspekt der Desinformation (38,4 %), insbesondere mit Blick auf das Ausland (beispielsweise auf die US-Politik und die Kriege in der Ukraine und Gaza). Der Anteil von Beiträgen mit Fokus auf Desinformation stieg 2023 leicht von 36,8 % auf 40,3 %. Berichte über Deepfakes im Kontext von Kriminalität machten 10,5 % der Berichterstattung aus. Der Anteil stieg 2023 von 8,2 % auf 14,1 %. Pornografie, die am Anfang der Entwicklung von Deepfakes im Zentrum stand, machte insgesamt 8,9 % der Beiträge aus und blieb relativ konstant (9,5 % versus 8,1 %). 17,4 % der

Medienberichterstattung waren Beiträge über Deepfakes in der Unterhaltung. Der Anteil sank deutlich von 23,4% auf 8,1%. Allgemeine bzw. unspezifische Beiträge nahmen von 22,1% auf 29,5% zu und machten genau ein Viertel (25,0%) der Berichterstattung aus. Dieser Befund deutet an, dass sich das Thema bzw. das Konzept Deepfakes in der Öffentlichkeit grundsätzlich zunehmend etabliert.

Wenn man die Bewertung nach Themenbereichen auswertet, zeigt sich, dass lediglich in der Unterhaltung Deepfake-Technologie zu grossen Teilen positiv bewertet wird. Insgesamt 34,8% der Beiträge zu Film, Musik oder Gaming weisen eine positive Tonalität auf. In allen anderen Bereichen dominieren (fast) ausschliesslich negative Bewertungen. Dies ist zum einen wenig überraschend, weil es sich bei Desinformation, Kriminalität und Pornografie um negativ konnotierte Bereiche handelt. Allerdings wurden im Laufe der Erhebung neben Unterhaltung keine weiteren Bereiche identifiziert, in denen positiv über Deepfakes berichtet wurde.

3.2 Wahrnehmung von Deepfakes in der Schweizer Bevölkerung

Anhand einer repräsentativen Befragung der Schweizer Bevölkerung untersuchten wir, welche Rolle die Mediennutzung bei der Wahrnehmung von Deepfakes spielt. Wir fragten: Korreliert die Nutzung von journalistischen Medien, Social Media sowie Videoplattformen erstens mit der Bekanntheit des Begriffs Deepfakes, zweitens mit dem Sehen von Deepfakes, drittens mit dem Wissen über Deepfakes sowie viertens mit dem wahrgenommenen Einfluss von Deepfakes auf die Meinung der Schweizer Bevölkerung?

3.2.1 Bekanntheit von Deepfakes

Auf die Frage, ob sie den Begriff Deepfake bereits kannten, antworteten 57,0% der Teilnehmenden, dass sie schon einmal von Deepfakes gehört hatten. Daraufhin prüften wir mit einem binär logistischen Regressionsmodell, welche Faktoren mit der Bekanntheit (1 = davon gehört, 0 = noch nicht davon gehört) von Deepfakes korrelieren (vgl. Tabelle 1).

Es zeigt sich, dass der Konsum von journalistischen Medien positiv mit der Bekanntheit von Deepfakes korreliert (OR = 1,12; $p \leq 0{,}001$). Wer also oft News nutzt, kennt auch eher den Begriff Deepfake. Für die Nutzung von Social Media (OR = 1,01; $p = 0{,}795$) und Videoplattformen (B = 0,97; $p = 0{,}409$) besteht kein solcher Zusammenhang. Weiter gaben weniger Frauen als Männer an, den Begriff Deepfakes zu kennen (OR = 0,74; $p = 0{,}015$). Weder das Alter (OR = 1,00; $p = 0{,}449$) noch die Bildung (OR = 1,07; $p = 0{,}680$) korrelieren mit der Bekanntheit von Deepfakes. Menschen aus der Suisse romande gaben eher als Deutschschweizer:innen an, mit dem Begriff vertraut zu sein (OR = 2,02; $p \leq 0{,}001$).

3.2.2 Sehen von Deepfakes

Sodann eruierten wir, ob die Befragten nach eigenen Angaben bereits ein Deepfake gesehen hatten. Knapp die Hälfte (49,2%) bejahte dies. Daraufhin prüften wir, welche Faktoren mit dem Sehen von Deepfakes zusammenhängen (vgl. Tabelle 1).

Anders als bei der Bekanntheit korreliert der Konsum von journalistischen Medien nicht mit dem Sehen von Deepfakes (OR = 0,99; $p = 0{,}680$), die Nutzung von Videoplattformen hingegen schon. Wer also viel Zeit beispielsweise auf YouTube verbringt, hat eher schon einmal ein Deepfake gesehen als Menschen, die solche Plattformen wenig nutzen (OR = 1,09; $p = 0{,}016$). Für die Nutzung von Social Media zeigt sich kein solcher Zusammenhang (OR = 1,06; $p = 0{,}107$). Ausserdem gilt: Je jünger die Menschen sind, desto eher haben sie schon einmal ein Deepfake gesehen (OR = 0,98; $p \leq 0{,}001$). Männer haben schon eher Deepfakes gesehen als Frauen (OR = 0,72; $p = 0{,}008$). Kein Zusammenhang besteht zwischen der Bildung und dem Sehen von Deepfakes (OR = 1,25; $p = 0{,}136$). Deutschschweizer:innen gaben eher an, Deepfakes gesehen zu haben, als Menschen aus der Suisse romande (OR = 0,73; $p = 0{,}009$).

3.2.3 Wissen über Deepfakes

Zu Beginn der Erhebung hatten wir ungestützt, d.h. ohne Vorinformationen zu geben, gefragt, ob die Menschen wissen, was Deepfakes sind (vgl. Ta-

Dimensionen	Variablen	Bekanntheit (OR)	Sehen (OR)	Wissen (OR)
Soziodemografie	Alter	1,00	0,98***	0,97***
	Geschlecht[a]	0,74*	0,72**	0,44***
	Bildung[b]	1,07	1,25	2,31***
	Sprachregion[c]	2,02***	0,73**	0,49***
Mediennutzung	Journalistische Medien	1,12***	0,99	1,02
	Social Media	1,01	1,06	0,95
	Videoplattformen	0,97	1,09*	1,24***
	R^2 Tjur	0,047	0,048	0,099

Tabelle 1: Binär logistische Regressionsmodelle für die Bekanntheit von Deepfakes, das Sehen von Deepfakes und das Wissen über Deepfakes

Die Tabelle stellt die Resultate von binär logistischen Regressionsanalysen dar. Sie zeigen, wie die Bekanntheit von Deepfakes, das Sehen von Deepfakes und das Wissen über Deepfakes mit der Mediennutzung und soziodemografischen Faktoren korrelieren (repräsentative Befragung in der Deutschschweiz und der Suisse romande; n = 1359). [a] Männlich ist die Referenzkategorie; [b] keine Hochschulbildung ist die Referenzkategorie; [c] Deutschschweiz ist die Referenzkategorie. Signifikante Zusammenhänge sind mit Sternchen markiert (* p < 0,05; ** p < 0,01; *** p < 0,001). OR = Odds Ratio.
Lesebeispiel: Die Nutzung journalistischer Medien hängt positiv mit der Bekanntheit von Deepfakes zusammen. Die Nutzung von Videoplattformen korreliert mit dem Wissen über Deepfakes.

belle 1). Wir werten die Antworten dann als korrekt, wenn der Gegenstand (Bild-, Audio- oder Videoinhalte) und das technologische Verfahren (KI) genannt wurden, also analog zum Vorgehen bei der Inhaltsanalyse. Insgesamt 18,7% der Befragten waren in der Lage, Deepfakes zumindest rudimentär zu definieren. Anders gesagt: 18,7% der Teilnehmenden wussten, um was es sich bei Deepfakes handelt. Die grosse Mehrheit (81,3%) hatte demnach kein spezifisches Verständnis von Deepfakes.

Wiederum prüften wir mit einem binär logistischen Regressionsmodell, welche Faktoren mit grundlegendem Wissen über Deepfakes zusammenhängen. Die Nutzung von Videoplattformen korreliert positiv mit dem Wissen über Deepfakes (OR = 1,24; p ≤ 0,001). Kein Zusammenhang besteht zwischen der Nutzung von journalistischen Medien und dem Wissen über Deepfakes (OR = 1,02; p = 0,602). Auch zwischen der Social-Media-Nutzung und dem Wissen über Deepfakes besteht kein signifikanter Zusammenhang (OR = 0,95; p = 0,319). Weiter zeigen die Daten, dass jüngere Menschen (OR = 0,97; p ≤ 0,001) und Befragte mit Hochschulbildung (OR = 2,31; p ≤ 0,001) eher korrekt antworteten. Männer gaben ebenfalls eher eine korrekte Definition als Frauen (OR = 0,44; p ≤ 0,001). Deutschschweizer:innen verfügten eher über grundlegendes Wissen als Menschen aus der Suisse romande (OR = 0,49; p ≤ 0,001).

3.2.4 Vermuteter Einfluss von Deepfakes auf Meinungen in der Gesellschaft

Weiterhin untersuchten wir mit einem linearen Regressionsmodell, ob ein Zusammenhang zwischen der Nutzung von journalistischen Medien und dem vermuteten Einfluss von Deepfakes auf die Meinung der Schweizer Bevölkerung besteht (vgl. Tabelle 2).

Die Daten zeigen, dass, je höher die Nutzung von journalistischen Medien ist, desto höher ist auch die Risikowahrnehmung (B = 0,079; p = 0,009). Menschen, die viel journalistische Medien konsumieren, gehen also eher davon aus, dass Deepfakes einen Einfluss auf die Meinung der Schweizer Bevölkerung haben. Dies ist möglicherweise ein Effekt der starken Risikofokussierung in der Medienberichterstattung. Der gleiche Effekt wurde für die Nutzung von Videoplattformen festgestellt (B = 0,092; p = 0,005): Je höher deren Nutzung ist, desto eher gehen Menschen davon aus, dass Deepfakes einen Einfluss auf die Meinung der Schweizer:innen haben. Die Nutzung von Social Media korreliert nicht mit dem vermuteten Einfluss auf die Gesellschaft (B = 0,022; p = 0,482).

Abschliessend untersuchten wir das Vertrauen in die Medien: Menschen mit einem höheren Medienvertrauen gehen weniger davon aus, dass sich Deepfakes negativ auf die Einstellungen der Schweizer:in-

Dimensionen	Variablen	Vermuteter Einfluss (B)
Soziodemografie	Alter	0,027
	Geschlecht[a]	**0,011**
	Bildung[b]	0,062*
	Sprachregion[c]	−0,109***
Mediennutzung	Journalistische Medien	0,079**
	Social Media	0,022
	Videoplattformen	0,092**
Vertrauen	Medienvertrauen	−0,103***
	korr. R^2	0,025

Tabelle 2: Lineares Regressionsmodell für den vermuteten Einfluss von Deepfakes auf die Meinung der Schweizer Bevölkerung

Die Tabelle stellt die Resultate eines linearen Regressionsmodells dar. Es zeigt, wie der vermutete Einfluss von Deepfakes auf die Meinung der Schweizer Bevölkerung mit der Mediennutzung und soziodemografischen Faktoren korreliert (repräsentative Befragung in der Deutschschweiz und der Suisse romande; n = 1359). [a] Männlich ist die Referenzkategorie; [b] keine Hochschulbildung ist die Referenzkategorie; [c] Deutschschweiz ist die Referenzkategorie. Signifikante Zusammenhänge sind mit Sternchen markiert (* p < 0,05; ** p < 0,01; *** p < 0,001).
Lesebeispiel: Wer oft journalistische Medien nutzt, geht eher von einem Einfluss von Deepfakes auf die Meinung der Schweizer Bevölkerung aus.

nen auswirken (B = −0,103, p ≤ 0,001). Menschen mit einem Hochschulabschluss denken eher, dass Deepfakes die Meinung der Schweizer Bevölkerung beeinflussen (B = 0,236; p = 0,025). Weder Alter (B = 0,027; p = 0,418) noch Geschlecht (B = 0,011; p = 0,710) hängen mit einer solchen Wahrnehmung zusammen. Menschen aus der Suisse romande gehen weniger von einem Einfluss von Deepfakes auf die Bevölkerung aus als Deutschschweizer:innen (B = −0,337; p ≤ 0,001).

4 Fazit

In dieser Studie untersuchten wir, wie journalistische Medien über Deepfakes berichten und wie die Mediennutzung mit unterschiedlichen Dimensionen der Wahrnehmung von Deepfakes zusammenhängt. Unsere Inhaltsanalyse der Berichterstattung zeigte, dass es sich bei Deepfakes noch um ein Nischenthema in den Medien handelt. Die Resonanz von Deepfakes hat aber seit dem Innovationsschub in der KI-Technologie zugenommen. Deepfakes werden vornehmlich mit Risiken in Verbindung gebracht: insbesondere mit Desinformation, aber auch mit Kriminalität und Pornografie. Folglich herrscht in der Berichterstattung eine überwiegend negative Tonalität. Positive Thematisierungen im Bereich der Unterhaltung, etwa in Film und Musik, treten zunehmend in den Hintergrund. Kritisch auf Technologien wie Deepfake zu blicken und vor deren Risiken zu warnen, ist eine zentrale Aufgabe des Journalismus, und dies wird auch von Medienschaffenden in der Schweiz so gesehen (Raemy et al., 2024). Werden aber KI-Technologien in der Medienberichterstattung überwiegend und einseitig mit Risiken in Verbindung gebracht, besteht die Gefahr, dass auch Innovationspotenziale reduziert werden. Eine kürzlich erschienene Studie zeigt beispielsweise, dass Schweizer:innen gegenüber KI in den letzten Jahren kritischer geworden sind (Christen, 2024).

In punkto Mediennutzung zeigen sich interessante Zusammenhänge zwischen der Wahrnehmung von Deepfakes und der Nutzung von journalistischen Medien: Je mehr journalistische Medien Menschen nutzen, desto eher kennen sie den Begriff Deepfake. Dieser Befund deutet an, dass journalistische Medien dazu beitragen können, neue Technologien wie Deepfakes in der breiten Bevölkerung bekannt zu machen. Diese entspricht auch den Erwartungen und dem Rollenverständnis der Schweizer Journalist:innen im Kontext von Deepfakes (Raemy et al., 2024). Allerdings hängt der Konsum von journalistischen Medien weder mit den direkten Erfahrungen mit Deepfakes noch mit dem Wissen über Deepfakes zusammen. Je mehr journalistische Medien Schweizer:innen konsumieren, desto eher gehen sie davon aus, dass Deepfakes einen negativen Effekt auf die Gesellschaft haben. Möglicherweise könnte diese erhöhte Risikowahrnehmung mit dem starken Fokus

auf Risiken in der Berichterstattung zusammenhängen. Unser Studiendesign erlaubt aber keine Aussagen zu einem kausalen Zusammenhang zwischen den genutzten Medieninhalten und der Wahrnehmung von Deepfakes.

Interessant sind auch die Zusammenhänge zwischen der Nutzung von Videoplattformen und der Wahrnehmung von Deepfakes. Wenn über Deepfakes gesprochen wird, stehen meist Videoinhalte im Zentrum. Es ist deshalb wenig überraschend, dass die Nutzung von beispielsweise YouTube positiv mit direkten Erfahrungen mit Deepfakes, in unserem Fall dem Sehen von Deepfakes, zusammenhängt. Menschen, die häufig Videoplattformen nutzen, haben aber auch ein höheres Wissen über Deepfakes. Auch wenn wir nur sehr grundlegendes Wissen abgefragt haben, scheinen Menschen, die sich für audiovisuelle Medieninhalte interessieren, besser über Deepfakes Bescheid zu wissen. Gerade bei Deepfakes zeigt sich, dass Videoplattformen eine zentrale Rolle spielen. Die Beschäftigung mit dem Anwendungsbereich von Deepfakes hängt offenbar positiv mit dem Wissen über die Technologie zusammen.

Bezüglich der soziodemografischen Merkmale zeigten sich vor allem erwartete und in der Forschung bekannte Zusammenhänge: Je jünger die Menschen sind, desto eher haben sie schon einmal ein Deepfake gesehen und desto eher wissen sie, was Deepfakes sind. Allerdings gaben ältere Menschen eher an, Deepfakes zu kennen als jüngere. Ältere Menschen scheinen das Phänomen also durchaus zu kennen, haben aber keine direkten Erfahrungen und auch weniger Wissen zu Deepfakes. Typisch für die Forschung zu Technologie ist auch der Befund, dass Männer im Vergleich zu Frauen mehr Erfahrungen mit Deepfakes haben und auch eher in der Lage waren, den Begriff korrekt zu definieren.

Limitationen

Die Studie hat einige Limitationen. Sie erlaubt keine kausalen Aussagen zu Effekten der Medienberichterstattung auf die Wahrnehmung der Befragten. Wir wissen lediglich, welche Medien die Befragten nutzten, aber nicht, ob sie tatsächlich dort auch etwas über Deepfakes erfuhren. Auch wenn sie plausibel sind, stellten wir mögliche Effekte der Medieninhalte auf die Wahrnehmung von Deepfakes nur interpretativ her. Darum wären fortführende Studien notwendig, die eine direktere Verbindung zwischen genutzten Inhalten und ihren Effekten auf die Wahrnehmung von Technologien wie KI und Deepfakes ermöglichen. Weiterführende Studien könnten zudem nicht nur die Berichterstattung von journalistischen Medien erfassen, sondern auch Inhalte von Social Media und Videoplattformen berücksichtigen.

Literatur

Ahmed, S. (2023). Navigating the Maze: Deepfakes, Cognitive Ability, and Social Media News Skepticism. *New Media & Society, 25*(5), 1108–1129. https://doi.org/10.1177/14614448211019198

Ajder, H., Patrini, G., Cavalli, F. & Cullen, L. (2019). *The State of Deepfakes: Landscape, Threats, and Impact*. https://regmedia.co.uk/2019/10/08/deepfake_report.pdf

Bundesamt für Cybersicherheit – BACS (2024). *Woche 14: Online-Meeting mit Deep-Fake-Chef: CEO-Betrug 2.0*. https://www.ncsc.admin.ch/ncsc/de/home/aktuell/im-fokus/2024/wochenrueckblick_14.html

Christen, M. (2024). *ChatGPT erhöht die Skepsis gegenüber KI*. https://www.news.uzh.ch/de/articles/news/2024/dsi-insight-christen.html

Gosse, C. & Burkell, J. (2020). Politics and Porn: How News Media Characterizes Problems Presented by Deepfakes. *Critical Studies in Media Communication, 37*(5), 497–511. https://doi.org/10.1080/15295036.2020.1832697

Hameleers, M., van der Meer, T. G. L. A. & Dobber, T. (2022). You Won't Believe What They Just Said! The Effects of Political Deepfakes Embedded as Vox Populi on Social Media. *Social Media + Society, 8*(3), 1–12. https://doi.org/10.1177/20563051221116346

Ho Tran, T. (2023, 26. Oktober). AI Deepfakes Are Making War in Ukraine Even More Chaotic and Confusing. *The Daily Beast.* https://www.thedailybeast.com/ai-deepfakes-are-making-war-in-ukraine-and-israel-even-more-chaotic

Karaboga, M., Frei, N., Puppis, M., Vogler, D., Raemy, P., Ebbers, F., Runge, G., Rauchfleisch, A., de Seta, G., Gurr, G., Friedewald, M. & Rovelli, S. (2024). *Deepfakes und manipulierte Realitäten. Technologiefolgenabschätzung und Handlungsempfehlungen für die Schweiz*. vdf.

Klaus, J. (2023, 27. März). Warum das Papst-Foto nicht nur witzig ist. *zdf heute.* https://www.zdf.de/nachrichten/panorama/prominente/papst-daunenjacke-fake-ki-kuenstliche-intelligenz-100.html

Raemy, P., Puppis, M. & Gurr, G. (2024): Deepfakes im Journalismus. 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 (Hg.). *Deepfakes und manipulierte Realitäten* (S. 205–252). vdf.

Rubiero, L. R. (2023, 26. Oktober). Deepfakes in Warfare: New Concerns Emerge from Their Use around the Russian Invasion of Ukraine. *The Conversation.* https://theconversation.com/deepfakes-in-warfare-new-concerns-emerge-from-their-use-around-the-russian-invasion-of-ukraine-216393

Sison, A. J. G., Daza, M. T., Gozalo-Brizuela, R. & Garrido-Merchán, E. C. (2023). ChatGPT: More Than a «Weapon of Mass Deception» Ethical Challenges and Responses from the Human-Centered Artificial Intelligence (HCAI) Perspective. *International Journal of Human-Computer Interaction*, 1–20. https://doi.org/10.1080/10447318.2023.2225931

Vaccari, C. & Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society, 6*(1), 1–13. https://doi.org/10.1177/2056305120903408

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 (Hg.). Deepfakes und manipulierte Realitäten (S. 125–151). vdf.

Wahl-Jorgensen, K. & Carlson, M. (2021). Conjecturing Fearful Futures: Journalistic Discourses on Deepfakes. *Journalism Practice, 15*(6), 803–820. https://doi.org/10.1080/17512786.2021.1908838

Westerlund, M. (2019). The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review, 9*(11), 40–53. https://doi.org/10.22215/timreview/1282

Kontakt

fög / Universität Zürich
Andreasstrasse 15
CH – 8050 Zürich

kontakt@foeg.uzh.ch
+41 (0)44 635 21 11
www.foeg.uzh.ch