

Greta Runge und Murat Karaboga

Deepfakes als kulturelle Praxis und gesellschaftliche Herausforderung: Zu Potentialen und Wirkungsweisen der Technologie

1. Einleitung

Deepfake-Technologie bietet die Möglichkeit, mit vergleichsweise geringem Aufwand das Gesicht, die Stimme und die Schreibweise anderer Personen zu imitieren. Deepfakes stellen synthetische audiovisuelle Medien von Gesichtern, Körpern oder Stimmen dar, die mittels Künstlicher Intelligenz (KI) produziert oder verändert wurden. Die technische Basis hierfür bilden Generative Adversarial Networks (GANs), künstliche Neuronale Netze, die eine synthetische Datei erzeugen (Pawelec und Bieß 2021). Die Entwicklungen im Bereich „generativer KI“ schreiten schnell und großflächig voran, was sich an dem Aufkommen und der breiten Verwendung von Systemen wie ChatGPT und DALL-E zeigt. Auf Grundlage von GANs können diese Systeme Texte, Bilder und Audiodateien durch Befehle (Prompts) erstellen. Deepfakes bilden in der Regel Personen ab, die etwas zu sagen oder zu tun scheinen, was sie nie gesagt oder getan haben (van Huijstee et al. 2021). Mit dem Aufkommen von Bild-, Audio- und Textgeneratoren können neuerdings nicht mehr nur Personen, sondern in zunehmender Qualität auch jedwede Gegenstände, Situationen und Geräusche unter KI-Einsatz synthetisiert werden.

Der vorliegende Beitrag basiert auf einer von TA-SWISS in Auftrag gegebenen Studie zum Thema „Deepfakes und manipulierte Realitäten“.

Seit der Existenz von Fotomontagen ist die technikbasierte Täuschung von Lebenswirklichkeiten kein neues Phänomen: „Photoshoppen“ ist zum integralen Bestandteil der *digitalen Kultur* geworden. An der Schnittstelle zwischen Realität und öffentlicher, medialer Darstellung formt die Bildbearbeitungstechnik die Wahrnehmung von (optischer) Wirklichkeit. Ferner führt dies zu einer Manipulation des Informationsgehalts und stellt die Authentizität und Aussagekraft von Bildmaterial in Frage. Die Fortschritte der KI-Technologien haben die Vielfalt an Manipulationstechniken des Informationsgehalts von Selbst- und Fremddarstellung seither vor allem in Richtung Bewegtbild ausgeweitet, doch auch Audio-inhalte sind betroffen (Paris und Donovan 2019).

Zahlreiche Artikel und Studien haben sich in den vergangenen Jahren mit den Herausforderungen auseinandergesetzt, die sich aus der Verbreitung von Deepfakes ergeben. Neben Risiken für Betroffene, z. B. durch Verleumdung oder Deepfake-Pornographie, und Risiken für Organisationen, etwa Reputationsschäden für Unternehmen, verweisen zahlreiche Beiträge insbesondere auf die gesellschaftlichen Risiken. Demnach seien Deepfakes beispielsweise in der Lage, durch Informationsverschmutzung, dass also zunehmend unklarer wird, was Tatsachen und was manipulierte Inhalte sind, zu einer Erosion des Vertrauens, etwa in wichtigen gesellschaftlichen Bereichen (politisches, Wissenschafts- oder Justizsystem) beizutragen. In der einschlägigen Literatur ist hingegen gar nicht bis selten Thema, welche Chancen sich durch Deepfakes ergeben könnten (Godulla et al. 2021). In den Gesprächen mit der projektbezogenen TA-SWISS-Begleitgruppe äusserte sich daher die Notwendigkeit der Berücksichtigung der Chancen, um das Themenfeld Deepfakes in ausgewogener Weise betrachten zu können.

Angesichts dieses Erkenntnisinteresses ist es das erste Ziel des vorliegenden Beitrags zu untersuchen, welche Chancen mit Deepfakes und synthetischen Medien verknüpft werden. Dazu wurde eine qualitative Medieninhaltsanalyse durchgeführt, um einen Einblick in die öffentlichen Diskussionen, Anwendungsfelder sowie die hiermit verbundenen Chancen von Deepfakes zu erhalten.

Das zweite Ziel des Beitrags ist es, die im ersten Schritt identifizierten Chancen im Hinblick darauf zu untersuchen, ob und inwiefern sich aus diesen Chancen wiederum neue Herausforderungen ergeben, die in der weiteren Diskussion um Deepfakes berücksichtigt werden müssten.

Die Struktur des Kapitels spiegelt die Beantwortung der Forschungsfragen wider: In Kapitel zwei werden zunächst die in der Literatur diskutierten Herausforderungen benannt. In Kapitel drei werden die empirischen Ergebnisse zur Untersuchung der mit Deepfakes verknüpften Chancen diskutiert und Kapitel vier dient schließlich der Identifizierung neuer Herausforderungen, die aus den zuvor identifizierten Chancen resultieren. Der Beitrag endet mit einem Schlusskapitel, in dem die Ergebnisse in den Diskurs um Deepfakes eingeordnet werden.

2. Herausforderungen von Deepfakes und synthetischen Medien

Zahlreiche Beiträge widmeten sich in den vergangenen Jahren den aus Deepfakes resultierenden Herausforderungen. Typologisierungen unterteilen die aus Deepfakes resultierenden Herausforderungen meist in drei Kategorien, die sich an der Kritikalität der Nutzung, an den durch Deepfakes verursachten Schadens-

typen und der Ebene der Betroffenen oder des Impacts orientieren. Im Folgenden orientieren wir uns insbesondere an der Kategorisierung aus Bateman (2020) sowie Chesney und Citron (2019), während die konkreten Schadenstypen aus van Huijstee et al. (2021, S. 29) für die Befüllung der Dimensionen instruktiv sind (vgl. Tabelle 1). Demnach lässt sich das breite Spektrum möglicher Risiken in die drei Schadenskategorien *individuelle, institutionelle und gesellschaftliche Risiken* unterteilen. Individuelle Risiken betreffen die von einem Deepfake betroffenen Personen und reichen von Mobbing, über Verleumdung und Einschüchterung bis hin zu Erpressung und Rufschädigung. Institutionelle Risiken entstehen etwa aus Erpressung oder Identitätsdiebstahl. Risiken für Institutionen können auch aus Erpressung oder Identitätsbetrug resultieren, außerdem aber auch durch Reputationsschädigung oder die Manipulation von Aktienkursen mittels Deepfakes. Deepfakes haben zudem das Potenzial, übergreifende gesellschaftliche Schäden mit sich zu bringen, etwa im Falle von Wahlmanipulationen oder der generellen Beschädigung des demokratischen Diskurses.

Mikro	Meso	Makro
<i>Individuelle Risiken</i>	<i>Institutionelle Risiken</i>	<i>Gesellschaftliche Risiken</i>
Erpressung	Erpressung	Manipulation der Berichterstattung
Verleumdung	Identitätsdiebstahl	Beeinträchtigung der wirtschaftlichen Stabilität
Einschüchterung	Betrug (z. B. Versicherung/Zahlung)	Schaden für das Justizsystem
Mobbing	Manipulation von Aktienkursen	Schädigung des Wissenschaftssystems
Rufschädigung	Markenschaden	Erosion des Vertrauens
Untergrabung des Vertrauens	Reputationsschaden	Schädigung der Demokratie
Unterdrückung der Freiheit der Meinungsäußerung		Manipulation von Wahlen
		Beeinträchtigung internationaler Beziehungen
		Beeinträchtigung der nationalen Sicherheit

Tabelle 1: Überblick zu Deepfakes-Risiken (in Anlehnung an: van Huijstee et al. 2021, 29, eigene Übersetzung)

Ausgehend von dem dargestellten Forschungsstand zu den Risiken von Deepfake-Technologien in exemplarischen Anwendungsbereichen wird im nächsten

Abschnitt die in der Literatur unterrepräsentierte Antwort auf die Frage nach den Chancen, die sich durch den Einsatz und die Verbreitung von Deepfake-Technologien ergeben, beantwortet.

3. Chancen von Deepfakes

3.1. Grundlagen der empirischen Medieninhaltsanalyse

Zur Beantwortung der ersten Forschungsfrage, die sich auf die Chancen von Deepfakes und synthetischen Medien bezieht, wurde eine Medieninhaltsanalyse durchgeführt. Insbesondere Tageszeitungen sowie überregionale Medien decken aufgrund ihrer Reichweite ein grosses Spektrum an gesellschaftlichen Debatten und der öffentlichen Meinung ab (Mayring 2010). Die inhaltsanalytische Betrachtung von Zeitungsartikeln und Online-Nachrichten ist ein guter Indikator dafür, „was eine Gesellschaft bewegt und worüber sie diskutiert“ (Taddicken, 2019). Der Analyse geht es um die gesamtgesellschaftliche Sicht auf den Technologieeinsatz und nicht um das Einfangen von partikularen Interessen und Meinungen, sodass meinungsbetonte Texte wie Kommentare nicht analysiert worden sind.

Für die vorliegende Untersuchung wurden daher zum einen die reichweitenstärksten Schweizer-Tageszeitungen¹, internationale überregionale Medien sowie Online-News-Portale mit einem thematischen Schwerpunkt auf *Technik/Technologie* ausgewählt. Dabei wurde darauf geachtet, dass regionale und überregionale Medien, Print- und Online-Formate sowie die Abbildung eines diversen politischen Meinungsspektrums Berücksichtigung finden. Das Untersuchungsmaterial wurde zum einen über das Online-Recherche-Tool für Presseerzeugnisse *Lexis-Nexis* und zum anderen händisch über die Webseiten der betreffenden News-Portale erhoben. Als Untersuchungszeitraum wurde das Erscheinungsjahr des ersten Deepfakes 01.01.2017 bis 11.08.2022 festgelegt. Insgesamt wurden 445 Artikel erhoben, die sich direkt auf den Untersuchungsgegenstand *Deepfakes* beziehen.² Das Untersuchungsmaterial verdichtete sich auf insgesamt 238 Artikel, die in Bezug auf Chancen und Risiken von Deepfake-Technologien analysiert worden sind.

1 Dieser Fokus auf die Schweiz ergibt sich aus der Beauftragung der Studie durch TA-SWISS, in deren Rahmen dieser Beitrag entstanden ist.

2 Im Rahmen der Erhebung wurden verschiedene Schreibweisen von Deepfakes (Deepfake(s), Deep Fake(s), Deep-Fake(s)) sowie das Synonym synthetische Medien berücksichtigt.

Als Grundlage für die Selektion und Analyse des Untersuchungsmaterials ($n=238$) dienten die drei vorgestellten Risiko-Kategorien für Deepfakes. Diese deduktiven Kategorien gelten für die Untersuchung als anschlussfähig, da sie nicht nur auf Ebene des Individuums Risiken von Deepfakes tangieren, sondern auch institutionelle Aspekte auf der Meso-Ebene sowie gesellschaftliche Auswirkungen mitberücksichtigen.

3.2. Chancen von Deepfakes

Im Rahmen der empirischen Betrachtung wurden für den Einsatz von Deepfakes vier Anwendungsfelder identifiziert: *Entertainment, Bildung, Politische Prozesse sowie Werbung und kommerzieller Nutzen*. Im Folgenden werden beispielhaft Chancen in den vier Anwendungsfeldern vorgestellt.³

3.2.1. Unterhaltung

Deepfake-Technologie ist schon heute im Bereich der Unterhaltungsindustrie anzutreffen, und es wird ihr zugetraut, weitere enorme Anwendungspotentiale zu erschließen. Im Bereich der Filmindustrie wird dies besonders deutlich. Als 2016 im Kinofilm „Rogue One: A Star Wars Story“ sichtbar der bereits 1994 verstorbenen Peter Cushing mittels CGI digital *wiederbelebt* wurde, waren Zuschauer und Kritikerinnen hin- und hergerissen. Einige fanden die 3-D-Computergrafik beeindruckend, andere hingegen bemängelten die Qualität der unter enormen Kosten erstellten Animation (Walsh 2016). Nur wenige Jahre später wurde dieselbe Szene neu und in weitaus besserer Qualität produziert. Dem Youtuber Shamhook, einer Einzelperson, war es gelungen unter Einsatz von Deepfake-Technologie der CGI-Produktion eines Hollywood-Studios nicht nur Konkurrenz zu machen, sondern nach Ansicht vieler Fans deren Arbeit auch eindeutig zu übertreffen (Chichizola 2020). Nachdem der Youtuber seine Qualitäten auch bei anderen Szenen unter Beweis gestellt hatte (Aschenbrenner 2020), wurde er schließlich von Lucasfilm eingestellt, um die Technik weiter zu entwickeln und künftig CGI-Technik ggf. durch Deepfakes zu ersetzen (Bastian 2021). Inzwischen ist bei Disney ein neues Deep-Learning-basiertes Verfahren im Einsatz (Zoss et al. 2022).

Auch im Rahmen eines Dokumentarfilms über den 2018 verstorbenen Autor und Moderator Anthony Bourdain wurde die Technik eingesetzt, um den TV-Star mediensynthetisch auferstehen zu lassen. Ein Deepfake-Audio liess die Stimme

3 Eine vollständige Auflistung und Erläuterung von Beispielen ist in der TA-SWISS Studie „Deepfakes und manipulierte Realitäten“ nachzulesen.

von Bourdain in der Dokumentation „Roadrunner“ erneut erklingen (Jacobs 2021).

Auch in der Musikindustrie finden sich vielfältige Anwendungsmöglichkeiten. „Deepfake music“ nennt sich ein Trend, dessen Software *JukeBox* von der Firma OpenAI entwickelt wurde⁴. Die Entwickler von OpenAI haben ihre Software mit bekannten Audio- und Textdaten gefüllt, auf deren Basis computergenerierte Songs erzeugt werden können, die bspw. nach Queen oder Mozart klingen. Das Programm zerteile die Audiosignale in Fragmente und setze sie nach Input, Stil oder Künstler neu zusammen. Für die Musikindustrie ergeben sich hierbei viele wirtschaftliche Potentiale, wie die Beschleunigung und Vereinfachung von Produktionsabläufen sowie die Senkung von Produktionskosten. Aufwändige Motion-Capturing-Verfahren könnten in Musikvideos (aber auch in der Filmproduktion) durch Deepfakes ersetzt werden. Ein Beispiel für den Einsatz von Deepfakes und synthetischen Inhalten in der Musikindustrie ist die Veröffentlichung eines Musikvideos der K-Pop-Band «Etern!ty», welches ausschliesslich aus bereits bestehendem Bildmaterial der elf Bandmitglieder entwickelt worden ist und in der Fangemeinschaft positive Resonanz erhielt (New York Times 2021). Die Band Limp Bizkit produzierte 2023 ein Musikvideo und inszenierte Wladimir Putin, Joe Biden und Wolodymyr Selenskyj in einem Video (t3n 2023).

In einem engen Verhältnis zu dem Einsatz von Deepfakes in der Film- und Musikindustrie stehen Deepfakes als Werkzeug für die Realisierung von Fan-Art. Besonderer Beliebtheit erfreuen sich veränderte Varianten ikonischer Filmszenen, in denen die Gesichter der Protagonisten mittels Deepfake-Technologie meist durch die Gesichter anderer Schauspieler und Schauspielerinnen ausgetauscht werden (Foley 2023). Möglich ist aber auch das Einfügen des eigenen Gesichts. Sogenannte Face Swap-Apps, die eine solche Nutzung ermöglichen, sind bereits seit vielen Jahren populär, und entsprechende Face Swap-Funktionen sind inzwischen nicht mehr nur in spezialisierten Apps wie Reface oder Snapchat, sondern auch in Instagram integriert (Harasewich 2022).

Darüber hinaus finden Deepfakes bereits heute Anwendung für komödiantische, satirische Zwecke. Etwa der ehemalige brasiliianische Präsident Bolsonaro, der die Bevölkerung vermeintlich über die Wichtigkeit des Händewaschens während einer globalen Pandemie aufklärt (Ajder et al. 2019).

Die Erschaffung von Jenseits-Charakteren wird auch auf Ebene des Individuums durch Unternehmen der sogenannten Digital Afterlife Industry (DAI) vor-

4 <https://openai.com/research/jukebox>

angetrieben (Murray 2020)⁵. Werden Deepfakes von Verstorbenen zur Trauerbewältigung eingesetzt, können sie Angehörigen bei der Verarbeitung des Verlustes dienen. Die dahinterliegende Geschäftsidee vereint die Produktion synthetischer Medien mit der Bestattungsbranche. So trafen in einer südkoreanischen TV-Show Eltern auf ihre verstorbene siebenjährige Tochter, die mittels Deepfake-Technologie und KI inszeniert worden war. An künstlicher Kommunikation mit Jenseits-Charakteren via Chatbots, die die Sprache von Verstorbenen nachahmen, arbeitet auch Microsoft (Harinja et al. 2023).

3.2.2. Werbung und kommerzieller Nutzen

Der zweite große Bereich, in dem Deepfakes wirtschaftliches Potential zugeschrieben wird, ist der Werbesektor. Das Startup „Rephrase.AI“ hat mit den Anwendungen „Rephrase Personalized“ und „Rephrase Studio“ Deepfake-Software auf den Markt gebracht, die es Unternehmen ermöglicht, den Trend der Personalisierung und Individualisierung via Deepfakes in Absatzstrategien zu integrieren. Avatare können Konsumentinnen ein eigenes, einzigartiges Produkt nach individuellem Geschmack präsentieren und zum Aufbau nachhaltiger Kundenbeziehungen beitragen (hy 2022). Weitere Chancen bietet die Deepfake-Technologie für den kommerziellen Bereich bei der Formulierung von personalisierten Werbebotschaften. Die Software Synthesia erstellt aus Texten Videos beispielsweise für Nachrichten oder Erinnerungen an abgebrochene Einkäufe (Synthesia 2023). Deepfakes bieten hier neue kommerzielle Anwendungen in der Unternehmenskommunikation.

Darüber hinaus kann die synthetische Produktion von Medieninhalten durch Algorithmen Unternehmen einen wesentlichen Kostenvorteil in der Präsentation von Mode verschaffen und sie als innovative Unternehmen im Markt positionieren (t3n 2022). Ein Beispiel hierfür ist der Stockfoto-Provider „Generated Photos“, der mittels KI-Fotos von Menschen auf Basis von bestehendem Bildmaterial generiert. Neben der vereinfachten Erstellung von Werbematerialien könnten insbesondere Kostensenkungen im Bereich des Bookings von Models die Folge sein (t3n 2019).

Weitere Potenziale bieten virtuelle Influencer, die schon heute in den sozialen Medien wirken und denen Millionen Menschen folgen.⁶ Nicht nur ergeben sich für Unternehmen hierbei reichweitenstarke Kooperationspartnerschaften, sondern auch in der Präsentation der Produkte effiziente Werbe-Produktionen:

5 Vgl. hierzu auch die Studie von TA-SWISS zum Tod im digitalen Zeitalter.

6 <https://www.virtualhumans.org/article/the-most-followed-virtual-influencers-of-2022>

Avatare benötigen nicht wie andere Models mehrere Takes und Shootingtage. Unternehmen könnten auch ihre ganz eigenen Avatare erschaffen. Denkbar ist auch, bekannte Persönlichkeiten oder Prominente virtuell erscheinen zu lassen und somit bekannte Testimonials einsetzen zu können (t3n 2021).

Warner Bros. setze 2021 Deepfake-Technologie als Werbemassnahme ein, sodass interessierte Personen sich in den Trailer des Films „Reminiscence“ projizieren konnten. Auf einer Website des Unternehmens konnte ein Foto hochgeladen werden, das dann durch KI-Systeme in den Trailer eingebaut wurde. Das persönliche Deepfake-Video konnte danach heruntergeladen und mit Kontakten geteilt werden. Dies verdeutlicht, dass Deepfakes als Hebel in der Fanökonomie wirken können, indem positive Emotionen, Identifikation mit dem Produkt und zugleich Netzwerkeffekte erzeugt werden (Heise Online 2021).

Deepfakes können auch für öffentlichkeitswirksame Kampagnen, die beispielsweise auf Gesundheitsförderung oder Prävention abzielen, eingesetzt werden. Ein Beispiel hierfür ist die Kampagne «Malaria must die» (Zero Malaria Britain 2019) mit dem ehemaligen Fussballstar David Beckham. Eine synthetische Version des Fussballers spricht darin in neun verschiedenen Sprachen.⁷ Das Video konnte aufgrund der Multilingualität in verschiedenen Ländern gezeigt werden, erzielte damit eine hohe Reichweite und konnte viele Spenden einbringen (Langhart 2021).

3.2.3. Bildung

Im Bildungsbereich birgt Deepfake-Technologie das Potenzial, Bildungsangebote innovativer zu gestalten und z. B. Schülerinnen und Schülern Lerninhalte durch neue visuelle und mediale Formate näher zu bringen. Synthetische Medien können historische Persönlichkeiten (bspw. Politiker und Wissenschaftlerinnen) wieder zum Leben erwecken und so den Unterricht ansprechender und interaktiver gestalten (Westling 2019). Der Einsatz von neuen Technologien wie Deepfakes kann das Interesse in Bildung seitens der Schülerinnen insgesamt steigern. In Bezug auf Fernunterrichtssysteme können synthetische Medien die Lernumgebung für Lernende interessanter machen. Diese Systeme sind zudem fähig, das Lernverhalten von Schülerinnen und Schülern zu analysieren, um für die Vermittlung von Inhalten personalisierte Strategien zu finden. Letztlich sollen die Leistung der Schülerinnen und Schüler verbessert und ihre Motivation gesteigert werden (Temir 2020).

⁷ <https://www.youtube.com/watch?v=QiiSAvKJIHo>

Prominente Beispiele zeigen, dass Deepfake-Technologie als Instrument der Medienpädagogik zur Sensibilisierung der Öffentlichkeit im Hinblick auf Desinformation verwendet werden kann. Zum einen wird die Technologie damit in der Öffentlichkeit bekannter; zum anderen werden Rezipientinnen und Rezipienten an die kritische Reflexion von konsumierten Informationen und Medieninhalten herangeführt. Prominente Beispiele hierfür sind die populäre Deepfake-Rede des ehemaligen US-Präsidenten oder die Weihnachtsansprache der verstorbenen britischen Queen (t3n 2020). In einem anderen Fall wurde ein Deepfake des Facebook-Gründers Mark Zuckerberg dafür genutzt, die Öffentlichkeit über die Gefahren der „digitalen Beeinflussungsindustrie“ aufmerksam zu machen (Kuhn 2019).

3.2.4. Politik

Deepfakes können im Kontext von Wahlen deliberationsfördernd und für die kollektive Willens- sowie Meinungsbildung eingesetzt werden, indem Humor und Unterhaltung den Fokus auf gesellschaftspolitische Gegebenheiten und Herausforderungen richten (Borucki et al. 2020). So könnte eine Beschäftigung mit Deepfakes in medienpädagogischen Kontexten hilfreich sein, um Hintergründe von politischen Themen zu analysieren und darüber zu diskutieren, welche Aussagen durch Deepfakes vermittelt werden können (Appel und Prietzel 2022). Dadurch könnte nicht nur erfahrbar werden, was hinter dem Phänomen Deepfakes und der Technologie steckt, sondern auch, welche Rolle manipulierte Inhalte im politischen Kontext und für die Bildung der öffentlichen Meinung spielen können.

Politische Akteure können auf humoristische Deepfakes zurückgreifen, um diese als reichweitenstarkes Mittel der Wähleransprache zu nutzen – denn Humor und Witz gelten als wirkungsvolle Katalysatoren für öffentliche Aufmerksamkeit (Dörner und Porzelt 2016). Unabhängig vom Bildungsniveau und dem politischen Interesse der Rezipientinnen und Rezipienten könnten sich so nicht nur Chancen für die Sympathiegenerierung für die Politiker selbst eröffnen, sondern auch für wichtige politische Themen (Porzelt 2015).

Im Bereich des politischen Aktivismus kann die Deepfake-Technologie auch dafür verwendet werden, Identitäten und Minderheiten zu schützen. Dies wird an dem Beispiel des Films „Welcome to Chechnya“ von David France deutlich, in dem Deepfakes eingesetzt worden sind, um die Verfolgung homosexueller Personen zu dokumentieren, ohne dabei Identitäten von realen Personen offenzulegen (New York Times 2020).

Über die verschiedenen Anwendungsbereiche hinweg konnten im Rahmen der Inhaltsanalyse eine Reihe von Potenzialen des Technologieeinsatzes identifiziert werden, die sich in die Analyseebenen individuelle, institutionelle und gesellschaftliche Chancen einordnen lassen:

Individuelle Chancen	Institutionelle Chancen	Gesellschaftliche Chancen
Neue Unterhaltungsangebote	Digitale Wiedergänger bei Filmproduktionen	Bewusstseinsbildung und Sensibilität für Falschinformationen
Identitätsschutz	Innovative Unterrichtsgestaltung durch Medienunterstützung	Stärkung der Medienkompetenz
Trauerhilfe (Unterstützung in Trauерprozessen)	Positionierung am Markt und Generierung von Aufmerksamkeit für eigene Produkte	Kulturelle Annäherung durch multilinguale Ansprachen
Emotionalität und positive Gefühle (Fan-Kultur)	Stärkung der Kundenbindung	Schutz von Minderheiten und Identitäten
Ausleben von Fan-Phantasien	Kosten- und zeiteffiziente Produktionen	
Personalisierte Nachhilfe durch Avatare für bessere Lernergebnisse		
Steigerung der Lernmotivation		
Bewusstseinsbildung für Falschinformationen		

Tabelle 2 Zusammenfassende Darstellung von Chancen durch den Einsatz und die Verbreitung von Deepfakes

4. Neue Herausforderungen aus den Chancen des Technologieeinsatzes

In den vorangegangenen Kapiteln wurden die mit Deepfakes einhergehenden Chancen aufgezeigt, die sich für Individuen, Unternehmen und die Gesellschaft ergeben. Aus diesen Potenzialen des Technologieeinsatzes ergeben sich neue Herausforderungen auf allen drei Analyseebenen in den aufgezeigten Anwendungsfeldern.

4.1. Unterhaltung

Die Steigerung der Kosten- und Zeiteffizienz bei Produktionen im Unterhaltungsbereich kann zu materiellen Umverteilungsprozessen und zur Entwertung von Arbeit führen. Etwa, wenn aufgrund des Rückgriffs auf kostengünstige Deepfake-Produktionen Fachkräfte im Bereich digitaler, computergenerierter Bildefekte

(CGI) abgelöst werden (Fergusson et al. 2023). Im Bereich der Film- und Musikproduktion ergeben sich Fragen im Bereich des Persönlichkeits- und des Datenschutzes, da bestehendes Stimm- und Sprachmaterial zweckentfremdet wird. Darüber hinaus stellen sich Fragen des urheberrechtlichen Schutzes (Chesney und Citron 2019).

Deepfakes, die für die Belebung bereits verstorbener Personen eingesetzt werden, benötigen einen Rahmen für den Umgang mit eventuellen rechtlichen und ethischen Implikationen. Im Kontext des Diskurses um die Digital Afterlife Industry (DAI), definiert als „any activity of production of commercial goods or services that involves online usage of digital remains“ (Öhman und Floridi 2017), werden offene Fragen zu den sozialen, wirtschaftlichen und regulatorischen Aspekten, die eine respektvolle Haltung gegenüber den Verstorbenen seitens der technikentwickelnden Industrie fokussieren, gegenwärtig und zukünftig diskutiert werden müssen. Umstritten ist die Produktion von posthumen Deepfakes deshalb, weil es keine Möglichkeit gibt, die Zustimmung der bereits verstorbenen Person einzuholen. Im Falle von Persönlichkeiten des öffentlichen Lebens besteht ein offensichtlicher finanzieller Anreiz, ihr digitales Abbild zu erstellen, weswegen diese durch posthume Publizitätsrechte geschützt sind. Jedoch kann das Bildmaterial durch die Zustimmung von Angehörigen verwendet werden, sodass hierbei gegebenenfalls ein Spannungsfeld zwischen emotionalen und finanziellen Beweggründen entsteht (ebd.).

4.2. Werbung und kommerzieller Nutzen

Der Einsatz von Deepfakes im Bereich der Werbung und des kommerziellen Nutzens wirft Fragen im Bereich des Persönlichkeitsrechts auf, weil Personen des öffentlichen Lebens zunehmend unfreiwillig in Werbeproduktionen auftreten könnten – ohne informiert oder um Erlaubnis gefragt worden zu sein. Darüber hinaus sind synthetische Personen und Avatare im Bereich der Werbung in der Lage, Konsumentinnen und Konsumenten in die Irre zu führen und zu täuschen. Insgesamt ergibt sich in diesem Anwendungsbereich ein Spannungsfeld zwischen wirtschaftlichen, ethischen und sozialen Aspekten.

4.3. Bildung

Im Bereich der Bildung stellt sich die Grundsatzfrage, welche Gefahren insbesondere für Kinder und Jugendliche von realistisch wirkenden Video-Manipulationen und Deepfakes ausgehen. Unklar ist bislang, welche psychologischen

Auswirkungen (denkbar wäre Selbstisolation) die Technologie auf das Verhalten von Schülerinnen und Schüler haben kann. Je nach Qualität des Deepfakes kann die Bewusstseinsbildung über bzw. die Erkennung von falschen Informationen (Informationskompetenz) nicht gelingen. Ferner stellen sich grundlegende Fragen nach den Risiken für die Privatsphäre von Individuen (bspw. bezüglich des Lernverhaltens) und dem Datenschutz bei der Verwendung von KI-Systemen zur Erstellung von Deepfakes im Bildungsbereich.

4.4. Politische Prozesse

Der Versuch, Deepfakes im Kontext von Wahlen zur Wähleransprache oder satirisch deliberationsfördernd und für die kollektive Willens- sowie Meinungsbildung einzusetzen, birgt die Gefahr, dass der demokratische Diskurs sukzessive durch die Verbreitung von irreführenden Inhalten manipuliert wird. Wenn die entsprechenden Deepfakes nicht als solche gekennzeichnet sind, könnten Betrachterinnen und Betrachter diese für bare Münze nehmen und die Inhalte oder eine Kommentierung darüber als Misinformation weiterverbreiten, was z.B. Wahlentscheidungen beeinflussen könnte (Heesen et al. 2021). Zudem kann die satirische Verwendung von Deepfakes als Entschuldigung herangezogen werden, um Deepfakes zu verbreiten, die schädliche Effekte haben (Ajder und Glick 2021).

Eine andere Gefahr resultiert aus der Normalisierung der Verbreitung von Deepfake-Inhalten. Wenn Menschen nicht mehr unterscheiden können, welche Inhalte echt sind und welche mittels Deepfakes erstellt oder manipuliert wurden, könnte das Vertrauen in Informationen generell, aber auch in politische Institutionen und den Journalismus erodieren. Dies kann schließlich zum sog. «Dividende des Lügners»-Phänomen beitragen, sodass echten Ton- und Videoaufnahmen, die Politikerinnen und Politiker während korrupter Handlungen zeigen, nicht geglaubt wird (Fallis 2021). Selbst besonders skandalöse (und deshalb vielleicht unglaublich) klingende oder aussehende Inhalte, die in schlechter Qualität vorliegen (bspw. die Ibiza-Affäre, um den österreichischen Politiker Heinz-Christian Strache), könnten einfacher gelegnet werden, indem das belastende Material als Deepfake bezeichnet wird.

Mikro	Meso	Makro
Individuelle Risiken	Institutionelle Risiken	Gesellschaftliche Risiken
<ul style="list-style-type: none"> Erpressung Verleumdung Einschüchterung/ Demütigung Mobbing Rufschädigung Untergrabung des Vertrauens Unterdrückung der Freiheit der nungsäußerung 	<ul style="list-style-type: none"> Erpressung Identitätsdiebstahl Betrug (z. B. Versicherung/ Zahlung) Manipulation von Aktienkursen Markenschaden Reputationsbeschädigen Manipulation von Aktienkursen Brandschäden Reputationsbeschädigen 	<ul style="list-style-type: none"> Manipulation der Berichterstattung Beeinträchtigung der wirtschaftlichen Stabilität Schaden für das Justiz- system Schädigung des Wissenschaftssystems Erosion des Vertrauens Schädigung der Demo- kratie Wahlen Manipulation von nationaler Beziehungen Beeinträchtigung der nationalen Sicherheit

Neue Risiken	Urheberrechtliche Fragen von Bildmaterial und Künstlerschem/geistigen Eigentum	Unklare Folgen für das Sozialverhalten von Schülerinnen und Schülern
<ul style="list-style-type: none"> Diskriminierung Verlust der digitalen Selbstbestimmung Täuschungen (bspw. Wahrheit als Grund- satz der Werbung) Unzureichende Medien- und Informations- kompetenz Gefahr für geistige Gesund- heit Privatsphäre und Daten- schutz (Analyse Lernverhalten) 	<ul style="list-style-type: none"> Urheberrechtliche Fragen von Bildmaterial und Künstlerschem/geis- tigen Eigentum Spannungsfeld zwi- schen wirtschaftlichen, ethischen und sozialen Aspekten 	<ul style="list-style-type: none"> Sozialverhalten von Schülerinnen und Schülern

Tabelle 3 Chancen und (neue) Risiken von Deepfakes auf verschiedenen Ebenen (in Anlehnung an die Risikokategorien nach Huijstee et al. 2021, S. IV)

5. Schluss

Der Beitrag zeigt auf, dass der Einsatz von Deepfakes in verschiedenen Anwendungsbereichen mit positiven Nutzungszwecken verbunden ist. Diese Potenziale des Technologieeinsatzes bleiben in der gegenwärtigen, auf Risiken fokussierten Deepfakes-Debatte weitgehend unberücksichtigt. Darüber hinaus verstellt eine einseitige Fokussierung auf gesellschaftliche und politische Risiken den Blick auf weitere Herausforderungen, die sich aus den technologiebedingten Chancen ergeben. Insgesamt liefert der Beitrag (aus der TA-SWISS Studie: „Deepfakes und manipulierte Realitäten“) empirische Anhaltspunkte zu einem holistischen Verständnis über die Deepfake-Technologie. Aufgrund der Kürze des Beitrags bildet die Diskussion der neuen Herausforderungen allerdings nur einen ersten Ausblick. Gänzlich offen bleibt schließlich die Frage, wie die im Bereich von Chancen erkannten neuen Herausforderungen zu adressieren sind. Künftige Forschung könnte sich daher der vertieften Untersuchung beider Dimensionen widmen.

Das Technologiefeld „generative KI“, dessen Teil Deepfakes sind, charakterisiert sich durch eine hohe Entwicklungsdynamik. Derzeit ist etwa zu beobachten, dass durch neue Anwendungen zur Generierung synthetischer Medien die im Deepfake-Diskurs bislang übliche Fokussierung auf die Manipulation des Aussehens und Ausdrucks von Personen in den Hintergrund rückt. Stattdessen treten neue Anwendungsmöglichkeiten in den Vordergrund, die von der Erschaffung neuer Welten mittels Bild- und Videogeneratoren bis hin zur Alltags- oder beruflichen Unterstützung mittels LLM-basierter Chatbots reichen. Die Herausforderungen generativer KI sind in der Diskussion neuen Anwendungen zwar weiterhin präsent und beziehen sich teils auf ganz neue Problemfelder. Zugleich nimmt die Diskussion von Potentialen allerdings eine grösitere Rolle ein, als im engeren Deepfakes-Diskurs zuvor. Insofern bleibt abzuwarten, wie sich die Debatte um generative KI weiter entwickeln wird – an weiteren Entwicklungsschüben wird es aller Voraussicht nach jedenfalls nicht mangeln.

Literaturverzeichnis

- Ajder, Henry; Patrini, Giorgio, Cavalli, Francesco; Cullen, Laurence (2019): The state of deepfakes: Landscape, threats, and impact. Amsterdam: Deeptrace, 27. Online verfügbar unter https://regmedia.co.uk/2019/10/08/deepfake_report.pdf, zuletzt geprüft am 30.06.2023.
- Ajder, Henry; Glick, Joshua (2021): Just joking! Deepfakes, satire and the politics of synthetic media. WITNESS and MIT Open Documentary Lab. Retrieved April, 13, 2022.
- Appel, Markus; Prietzel, Fabian (2022): The detection of political deepfakes. In: Journal of Computer-Mediated Communication 27(4), Artikel zmac008. DOI: 10.1093/jcmc/zmac008.

- Aschenbrenner, Vali (2020): The Mandalorian: Deepfake-KI schlägt Disney-CGI um Längen. In: GameStar. Online verfügbar unter <https://www.gamestar.de/artikel/the-mandalorian-staffel-2-finale-deepfake-ki,3365494.html>, zuletzt geprüft am 30.06.2023.
- Bastian, Matthias (2021): Mehr Retro Star Wars? Lucasfilm schnappt sich Deepfaker. In: MIXED. Online verfügbar unter <https://mixed.de/mehr-retro-star-wars-lucasfilm-schnappt-sich-youtuber-deepfaker/>, zuletzt geprüft am 30.06.2023.
- Bateman, Jon (2020): Deepfakes and Synthetic Media in the Financial System: Assessing Threat Scenarios. In: Cyber Policy Initiative Working Paper Series. Online verfügbar unter https://carnegieendowment.org/files/Bateman_FinCyber_Deepfakes_final.pdf, zuletzt geprüft am 25.04.2023.
- Borucki, Isabelle; Michels, Dennis; Marschall, Stefan (2020): Die Zukunft digitalisierter Demokratie – Perspektiven für die Forschung. In: Zeitschrift für Politikwissenschaft, 30(2), 359–378 DOI: 10.1007/s41358-020-00234-3.
- Gosse, Chandell; Burkell, Jacquelyn (2020): Politics and porn: how news media characterizes problems presented by deepfakes. Critical Studies in Media Communication, 37(5), 497–511.
- Chesney, Bobby; Citron, Danielle (2019): Deep fakes: A looming challenge for privacy, democracy, and national security. Calif. L. Rev. 107, 1753.
- Chichizola, Corey (2020): Rogue One Deepfake Makes Star Wars' Leia And Grand Moff Tarkin Look Even More Lifelike. In: CINEMABLEND. Online verfügbar unter <https://www.cinemabrend.com/news/2559935/rogue-one-deepfake-makes-star-wars-leia-and-grand-moff-tarkin-looks-even-more-lifelike>, zuletzt geprüft am 30.06.2023.
- Dörner, Andreas; Porzelt, Benedikt (2016): Politisches Gelächter. Rahmen, Rahmungen und Rollen bei Auftritten politischer Akteure in satirischen Interviews des deutschen Fernsehens. In: M&K 64(3), S. 339–358. DOI: 10.5771/1615-634X-2016-3-339.
- Fallis, Don (2021): The Epistemic Threat of Deepfakes. In: Philosophy & technology 34(4), S. 623–643. DOI: 10.1007/s13347-020-00419-2.
- Fergusson, Grant; Fitzgerald, Caitriona; Frascella, Chris; Iorio, Megan; McBrien, Tom; Schroeder, Calli; Winters, Ben; Zhou, Enid (2023): Generating Harms: Generative AI's Impact & Paths Forward. Online verfügbar unter: <https://epic.org/documents/generating-harms-generative-ai-s-impact-paths-forward/>, zuletzt geprüft am 30.06.2023.
- Foley, Joseph (2023): 20 of the best deepfake examples that terrified and amused the internet. In: Creative Bloq. Online verfügbar unter <https://www.creativebloq.com/features/deepfake-examples>, zuletzt geprüft am 27.06.2023.
- Frei, N., Zlatescu, P. E., Naedele, R. (2024): Tod im digitalen Zeitalter: rechtliche Ergänzung. TA-SWISS Publikationsreihe (Hrsg.): TA 82.1/2024. Zollikon: vdf.
- Godulla, Alexander; Hoffmann, Christian P.; Seibert, Daniel (2021): Dealing with deepfakes – an interdisciplinary examination of the state of research and implications for communication studies. In: Studies in Communication and Media 10(1), S. 72–96. DOI: 10.5771/2192-4007-2021-1-72.
- Harasewich, Tasha (2022): The 7 Best Face Swap Apps. In: MUO. Online verfügbar unter <https://www.makeuseof.com/tag/best-face-swap-apps/>, zuletzt geprüft am 30.06.2023.
- Harbinja, Edina; Edwards, Lilian; McVey, Marisa (2023): Governing ghostbots. In: Computer Law & Security Review 48, S. 105791. DOI: 10.1016/j.clsr.2023.105791.

- Heesen, Jessica; Bieber, Christoph; Grunwald, Armin; Matzner, Tobias; Roßnagel, Alexander (2021): KI-Systeme und die individuelle Wahlentscheidung. DOI: 10.48669/pls_2021-1.
- Heise Online (2021): Deepfake: Warner Bros. personalisiert Filmwerbung. Online verfügbar unter <https://www.heise.de/news/Deepfake-Warner-Bros-personalisiert-Filmwerbung-6166228.html>, zuletzt geprüft am 30.06.2023.
- hy (2022): Deep Fake (for good) – Risiken und Potenziale für Unternehmen. Online verfügbar unter: <https://hy.co/2022/04/28/deep-fake-for-good/>, zuletzt geprüft am 30.06.2023.
- Jacobs, Julia (2021): Bourdain Documentary's Use of A.I. to Mimic Voice Draws Questions. Online verfügbar unter: <https://www.nytimes.com/2021/07/16/movies/anthony-bourdain-ai-voice.html>, zuletzt geprüft am 30.06.2023.
- 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. TA-SWISS Publikationsreihe (Hrsg.): TA 81/2024. Zollikon: vdf.
- Kuhn, Johannes (2019): Zuckerbergs böser Zwilling bleibt auf Instagram. In: Süddeutsche Zeitung, 12.06.2019. Online verfügbar unter <https://www.sueddeutsche.de/digital/deepfake-facebook-zuckerberg-instagram-1.4483936>, zuletzt geprüft am 26.04.2023.
- Langhart, Nicole (2021): Deepfakes: Wie die AI-Technologie eindrückliche Marketing Massnahmen ermöglicht. Hg. v. Marketing.ch. Online verfügbar unter <https://marketing.ch/deepfakes-wie-die-ai-technologie-eindrueckliche-marketing-massnahmen-ermoeglicht/>, zuletzt aktualisiert am 17.05.2021, zuletzt geprüft am 17.05.2022.
- Lenk, Hartmut; Vesalainen, Marjo (2012): Der Kommentar als persuasiver Text. Vergleichende Untersuchungen zu einer meinungsbetonten Textsorte in europäischen Massenkommunikationsmedien. In: Persuasionsstile in Europa. Methodologie und Empirie kontrastiver Untersuchungen zur Textsorte Kommentar, S. 8–23.
- Mayring, Philipp (2010): Qualitative Inhaltsanalyse. Grundlagen und Techniken. 11. Neuauflage. Weinheim: Beltz (Beltz Pädagogik).
- Murray, Sarah (2020): Postdigital cultural studies. In: International Journal of Cultural Studies 23(4), S. 441–450. DOI: 10.1177/1367877920918599.
- Taddicken, Monika (2019): Analyse von Zeitungsartikeln und Online-Nachrichten. In: Handbuch Methoden der empirischen Sozialforschung, S. II157–II164.
- The New York Times (2020): Deepfake Technology Enters the Documentary World. Online verfügbar unter: <https://www.nytimes.com/interactive/2020/11/21/science/artificial-intelligence-fake-people-faces.html>, zuletzt geprüft am 30.06.2023.
- The New York Times (2021): What's even Real? Online verfügbar unter: <https://www.nytimes.com/interactive/2020/11/21/science/artificial-intelligence-fake-people-faces.html>, zuletzt geprüft am 30.06.2023.
- Öhman, Carl; Floridi, Luciano (2017): The Political Economy of Death in the Age of Information: A Critical Approach to the Digital Afterlife Industry. In: Minds & Machines 27(4), S. 639–662. DOI: 10.1007/s11023-017-9445-2.
- Paris, Britt; Donovan, Joan (2019): Deepfakes and cheap fakes. The manipulation of audio and visual evidence. Data & Society. Online verfügbar unter: https://datasociety.net/wp-content/uploads/2018/09/DS_Deepfakes, zuletzt geprüft am 30.06.2023.

- Pawelec, Maria; Bieß, Cora (2021). Deepfakes: Technikfolgen und Regulierungsfragen aus ethischer und sozialwissenschaftlicher Perspektive. Baden-Baden: Nomos Verlag.
- Porzelt, Benedikt (2015): Humor – Wie „heute show“ und Co. unseren Blick auf Politik verändert. Online verfügbar unter https://www.boell.de/sites/default/files/uploads/2015/04/benedikt_porzelt_humor_und_politik.pdf, zuletzt geprüft am 16.05.2023.
- Strub, J.-D., Bosisio, F., Jox, R. J., Rochel, J., Sterie, A.-C. (2024): La mort à l'ère numérique. Chances et risques du Digital Afterlife. TA-SWISS Publikationsreihe (Hrsg.): TA 82/2024. Zollikon: vdf.
- Synthesia (2023): The Future of (Synthetic) Media. Online verfügbar unter <https://www.synthesia.io/post/the-future-of-synthetic-media>, zuletzt geprüft am 30.06.2023.
- t3n (2019): Diese KI hat 100.000 Porträts nicht existenter Personen generiert 2019. Online verfügbar unter: <https://t3n.de/news/ki-stylegan-generated-photos-1201068/>, zuletzt geprüft am 30.06.2023.
- t3n (2020): Deepfake: Falsche Queen mit falscher Weihnachtsansprache auf Channel 4 2020. Online verfügbar unter: <https://t3n.de/news/bisschen-deepfake-falsche-queen-1347172/>, zuletzt geprüft am 30.06.2023.
- t3n (2021): Angela Merkel in der Werbung_ Wie Tibber mit Deepfake-Elementen spielt. Online verfügbar unter: <https://t3n.de/news/angela-merkel-werbung-startup-deepfake-1378618/>, zuletzt geprüft am 30.06.2023.
- t3n (2022): Synthetische Medien: Laut Meta beginnt „neues Kapitel für Influencer_innen-Marketing“. Online verfügbar unter: <https://t3n.de/news/meta-synthetische-medien-marketing-influencer-1444243/>, zuletzt geprüft am 30.06.2023.
- t3n (2023): Deepfake: Limp Bizkit dreht Musikvideo mit Putin, Biden und Selenskyj. Online verfügbar unter: <https://t3n.de/news/deepfake-limp-bizkit-musikvideo-putin-biden-selenskyj-1533827/>, zuletzt geprüft am 30.06.2023.
- Temir, Erkam (2020): Deepfake: New Era in The Age of Disinformation & End of Reliable Journalism. In: SELÇUK İLETİŞİM DERGİSİ 13(2). Online verfügbar unter 0.18094/JOSC.685338.
- van Huijstee, Mariëtte; van Boheemen, Pieter; Das, Djurre; Nierling, Linda; Jahn, Jutta; Karaboga, Murat; Fatun, Martin (2021): Tackling deepfakes in European policy. Online verfügbar unter [https://www.europarl.europa.eu/stoa/en/document/EPRS_STU \(2021\)690039](https://www.europarl.europa.eu/stoa/en/document/EPRS_STU (2021)690039), zuletzt geprüft am 30.06.2023.
- Walsh, Joseph (2016): Rogue One: the CGI resurrection of Peter Cushing is thrilling – but is it right? Online abrufbar unter <https://www.theguardian.com/film/filmblog/2016/dec/16/rogue-one-star-wars-cgi-resurrection-peter-cushing>, zuletzt geprüft am 30.06.2023.
- Westling, Jeffrey (2019): Are Deep Fakes a Shallow Concern? A Critical Analysis of the Likely Societal Reaction. TPRC47: The 47th Research Conference on Communication, Information and Internet Policy 2019. Online verfügbar unter https://papers.ssrn.com/sol3/papers.cfm?abst_ract_id=3426174.
- Zero Malaria Britain (2019): David Beckham speaks nine languages to launch Malaria Must Die Voice Petition. Hg. v. YouTube. Online verfügbar unter <https://www.youtube.com/watch?v=QiiSAvKJHo>, zuletzt aktualisiert am 26.04.2023, zuletzt geprüft am 26.04.2023.
- Zoss, Gaspard; Chandran, Prashanth; Sifakis, Eftychios; Gross, Markus; Go-tardo, Paulo; Bradley, Derek (2022): Production-Ready Face Re-Aging for Visual Effects. In: ACM Trans. Graph. 41(6), S. 1–12. DOI: 10.1145/3550454.3555520.