Dieser Beitrag ist in ähnlicher Form erschienen in Westerkamp, M. (2021): Deepfakes – „Sie glauben diese Geschichte ist wahr? Da muss ich Sie leider enttäuschen, sie ist frei erfunden.“, Janßen, S./ Kirstges, T./ Kull, S./ Neumann, M./Schmoll, E. (Hrsg): Jahresband 2021 des Fachbereichs Wirtschaft – Gesammelte Erkenntnisse aus Lehre und Forschung, S. 337-335, ISBN 978-3-643-14961-9.

Markus Westerkamp

Deepfakes - „Sie glauben diese Geschichte ist wahr? Da muss ich Sie leider enttäuschen, sie ist frei erfunden.“

1 Einleitung

Das Internet ist eine unendliche Informationsquelle. Zu beachten ist jedoch, dass nicht sämtliche Inhalte auf nachweisbaren Fakten, sondern auch auf Fake News (gezielten Falschinformationen) beruhen, die über das World Wide Web publiziert werden können. Heutzutage gestaltet sich die Differenzierung zwischen seriösen und falschen Meldungen (Deepfakes) schwer.

Deepfakes sind im Grunde genommen ähnlich wie der atavistische Gag aus Kindheitstagen. Tatsächlich, und hier liegt die Polarität, sind Deepfakes professioneller. Bei Deepfakes werden nicht Schere und Klebstoff eingesetzt, sondern Künstliche Intelligenz (KI).

Deepfakes, also mit KI hergestellte Fake-Videos, u. ä., fluten das Internet. Sie lassen sich stets leichter herstellen, wobei Sie wie realistische Aufnahmen wirken, aber gefälscht sind. Mittlerweile kann jede Person mit geeigneter Software vermeintlich wirklichkeitsnahe Videos herstellen, in denen Gesichter substituiert werden, Personen Äußerungen tätigen und Handlungen vornehmen, die sie nie getätigt bzw. vollzogen haben.

Was hat dies also für Auswirkungen auf unsere Gesellschaft, wenn Video- oder Audioaufnahmen nicht länger als eindeutige Beweise für die Realität gelten können? Deepfakes stehen daher durchaus in der Kritik.

Aus diesen Gründen wird der Artikel neben der Begriffsdefinition von Deepfakes, eine Einführung in deren Funktionsweisen, technologischen Grundlagen, Gefahrenpotenziale und Schutzmaßnahmen vornehmen**.** Des Weiteren werden diverse Anwendungsbereiche von Deepfakes transparent dargestellt und kritisch hinterfragt. Dieser Auszug von Anwendungsbereichen mit konkreten Praxisbeispielen zeigt die aktuelle und zukünftige Bedeutsamkeit der Thematik Deepfakes auf und gibt Beweggründe sich mit dem Hintergrund, technischen Möglichkeiten und weiteren Aspekten wie der Manipulationsgefahr von und Schutzmechanismen gegen Deepfakes auseinanderzusetzen.

2 Begriffsdefinition von Deepfakes

Als Deepfakes werden realistisch wirkende Bild-, Audio oder Videomaterialien definiert, die mit Hilfe von KI manipuliert wurden.[1] Hierbei kommen primär Methodiken aus dem Machine Learning (ML) zum Einsatz, vorrangig dem Deep Learning (DL). Der Name Deepfake ist daher eine kombinatorische Wortschmelzung der englischen Wörter DL und Fake.[2]

Deepfakes werden, insofern Sie dazu eingesetzt werden anderen zu schädigen, zu einer weiteren monetären und gesellschaftlichen Herausforderung für Privatpersonen und Unternehmen. Die Deepfakes-Qualität hat sich in den letzten Jahren, angesichts des technologischen Fortschritts, stets weiterentwickelt. Mit geeignetem technischem Equipment speziell Graphics Processor Units (GPUs)[3] lassen sich Bild-, Audio- und Videoinhalte modellieren, die auf den ersten Blick kaum noch als Fake erkennbar sind.[4] Dahin gehend sind auch diverse Deepfakes-Anleitungen und -Software im Internet unentgeltlich abrufbar. Auf Grund dessen beschäftigen seither die Polizei und andere staatliche Behörden bei der Sicherstellung von Beweismitteln Experten zur Deepfakes-Erkennung. Zugleich steigt die Relevanz

[1] Vgl. Trend Micro, 2021, S. 24.
[2] Vgl. Walorska, A., 2020, S. 9.
[3] Definition GPU: Der Begriff GPU (Graphics Processing Unit) bedeutet zu Deutsch Grafikprozessor. Die GPU ist für die Bildberechnung und die Bildschirmausgabe bei Spielekonsolen und Computern zuständig. Der Grafik-prozessor ist dafür zuständig alle Aufgaben zur Berechnung von 2D- oder 3D-Grafiken zu übernehmen.
[4] Vgl. Heller, M./Porup, J. M., 2021, o. S.

von Deepfakes, dass belegt das enorme Informationsausmaß, welches mittlerweile zur Generierung von manipuliertem Material verfügbar ist.[5]

Deepfakes stellen also prinzipiell in Frage, ob und welche Medieninhalten vertraut werden können. Ergo ist es bei ungewissen Handlungsaufforderungen empfehlenswert mit der natürlichen (Ziel-)Person in Kontakt zu treten und die Anfrage zu verifizieren. Demzufolge erfordert es auch einer Sensibilisierung bzw. Aufklärung für die positiven und negativen Möglichkeiten von Deepfakes.[6]

3 Entstehung von Deepfakes

Jeder Deepfake wird durch eine KI generiert. Die Grundlage für eine Generierung ist das zur Verfügung gestellte Quellmaterial. Dem Zielmaterial, ob Bild-, Audio oder Videomaterial, wird dabei der Inhalt, der nicht im real existierten Material verfügbar ist, manipulativ hinzugefügt. Die KI untersucht das vorhandene Material mithilfe von verschiedener biometrischer Variablen. Im Code des neuronalen Netzes introduziert ein sog. Autoencoder, der trainiert werden möchte. Die Zielsetzung des Autoencoders ist es, eine wirksame Darstellung - Encoding - für einen Datensatz zu lernen und folglich wichtigste Kriterien zu extrahieren. Anschließend werden diese Daten von einem Decoder dekomprimiert. Bei der Dekomprimierung ist der Zweck der KI, dass das Resultat so nah wie möglich am Original ist. Je präziser der Auswahlprozess ist respektive je mehr Inhalte die KI als Trainingsmaterial zur Verfügung gestellt bekommen hat, umso zutreffender ist das Manipulationsergebnis.[7]

4 Anwendungsbereiche

Deepfakes sind zu diversen Zwecken praktikabel. Die Anwendungsmöglichkeiten können bei der Satire und Unterhaltungsbranche anfangen, gehen über zur beabsichtigten Desinformation und enden bei der systematischen Demütigung einzelner Persönlichkeiten. Summa summarum werden

[5] Vgl. Müller, N., 2021, S. 56 f.
[6] Vgl. ebd.
[7] Vgl. Kaspersky, 2021.

Deepfakes für vielfältige legale oder illegale Praktiken eingesetzt. Typische Anwendungsbereiche sind: Kunst, Medizin, Politik und/oder Filmbranche.

Im Kunstbereich können Deepfakes beansprucht werden. Ein Beispiel hierfür ist die Videoarbeit *Mosaic Virus*. Die britische Forscherin und Künstlerin Anna Ridler nutzt Deepfakes für diese Arbeit. Es ist ein Projekt von 2019, bei welcher drei Bildschirme verschiedene durch Deepfakes generierte Tulpen darstellen. Das Aussehen der Tulpe wird durch den Preis eines Bitcoins gesteuert. Es soll ein Stillleben des 21. Jahrhunderts darstellen und belegen, dass Deepfakes auch in der Kunstszene einen Zusatznutzen generieren können.[8]

Außerdem ist das *Project Revoice* vom kanadischen KI-Start-up Lyrebird in Kooperation mit der ALS Association (ALSA) ein weiteres positives Beispiel für den Einsatz von Deepfakes. Hierbei handelt es sich um ein medizinisches Projekt. Die Deepfake Voice-Technologie wird genutzt, um krankheitsbedingten Menschen, die einen Stimmenverlust zu verzeichnen haben, einen authentischen Ersatz für ihre Stimme zurückzugeben.[9]

Ein weiteres positives Beispiel ist der Einsatz von Deepfakes im E-Commerce: *„Die FIA, Superpersonal und HANGER nutzten die Deepfake-Technologie, um Kunden direkt in Markenvideos und Kampagnen einzubinden. Mittels AR-Techniken auf dem Smartphone kann man beispielsweise Schuhe anprobieren. Mit Deepfakes geht das noch einen Schritt weiter: User können ganze Outfits an verschiedenen Models ansehen, mit verschiedenen Hautfarben, Kleidergrößen und Aussehen, oder sie können die Outfits an sich selbst ausprobieren.“*[10]

Eine Gefahr, die im politischen Umfeld von Deepfakes entstammen kann, ist die Skepsis, an bis dato glaubwürdige Quellen, wie Audio- oder Videoaufnahmen. Im Jahr 2018 gab es in Gabun Spekulationen über den Gesundheitszustand des Präsidenten Ali Bongo. Um Spekulationen zu beenden, hielt Ali Bongo seine traditionelle Neujahrsansprache, welche die Regierung durch ein Youtube-Video *„Gabon President Ali Bongo 2019 New Years Message“* veröffentlichte. Viele Interessierte hielten dies für ein

[8] Vgl. Hinterleitner, B., 2020.
[9] Vgl. Sidyuk, A./Titova, D., 2021.
[10] ebd.

Deepfake, was wiederum für Verwirrung sorgte. Eine Woche nach dem benannten Video startete das gabunische Militär einen erfolglosen Putschversuch. Forensische Analysen fanden allerdings kein Anzeichen für eine Videomanipulation. Ali Bongo ist danach wieder aufgetreten und erklärte seine Abwesenheit aufgrund medizinischer Behandlungen.[11]

Des Weiteren gibt es in der Filmbranche unterschiedliche Einsatzmöglichkeiten von Deepfakes. Als Beispiel sei die chinesische Deepfake App *ZAO* genannt. Mithilfe der App war es möglich anhand eines Selfies das eigene Gesicht auf unterschiedlichste Videos per Deepfakes darzustellen. Für Nutzer ist es eine einfache Option erste Interaktionen mit Deepfakes vorzunehmen.[12]

5 Schutzmaßnahmen gegen Deepfakes

Die Gefahr, die durch Deepfakes hervorgerufen werden können, lassen sich auf drei Kernbereiche eingrenzen: Zum einen das Manipulieren von Meinungen durch falsche Inhalte, was vor allem für Social Media eine Gefahr darstellt. Ein wichtiger Ansatzpunkt sind also die Verbreitungsmechanismen, da die gesellschaftliche Wirkung von Manipulationen mit wachsendem Verbreitungsgrad zunimmt. Ausgehend vom ersten Kernbereich der Gefahr hat Facebook zusammen mit einigen Partnern, wie Microsoft oder Amazon die *„Deepfake Detection Challenge“* (DFDC) ins Leben gerufen.[13] Des Weiteren haben Plattformen wie Facebook, YouTube und Twitter bereits Anfang des Jahres 2020 ihr Regelwerk für Anti-Deepfake-Paragraphen erweitert. Demnach sind satirische und parodistische Inhalte weiterhin erlaubt, manipulative Inhalte jedoch verboten.[14] Zudem sollten durch entsprechende gesetzliche Regelungen Social Media-Betreiber in die Pflicht genommen werden, stärker auf Manipulationen zu kontrollieren.

Ein weiterer denkbarer Lösungsansatz wäre die verpflichtende Einbindung eines digitalen Wasserzeichens, welches Rückschlüsse über die Produktion des Videos und die Verbreitungswege, beispielsweise durch die Blockchain Technologie, ermöglichen könnte. Hiermit lassen sich nachträgliche

[11] Vgl. Reischl, G., 2020, o. S.
[12] Vgl. Arlt, F./Arlt, H.-J., 2020, S. 129.
[13] Vgl. Gupta, B./ Nedjah, N., 2021, S. 59.
[14] Vgl. Velten, A.-K., 2020.

Manipulationen an Dokumenten aller Art, mithin auch Videos, ausschließen.[15]

6 Fazit

Eine Schlussfolgerung ist, dass nicht nur kritische Auffassungen von Deepfakes existieren, wie es oftmals von den Medien deskribiert wird. Diese Deepfakes-Auffassungen können einen positiven Effekt für die unterschiedlichsten Anwendungsbereiche fördern. Simultan bestätigt sich, dass mit fortschreitender Technik die Entwicklung von Deepfakes ansteigt. Heutzutage und zukünftig wird der Arbeitsaufwand, um Deepfakes zu modellieren stets weniger und weiterhin wird die Technik permanent leichter zugänglich für (Privat-)Personen und Unternehmen. Zuverlässig kann jedoch niemand prognostizieren, wie sich die (Deepfake-)Technik in Zukunft fortentwickelt und wie sie dann eingesetzt wird. Weiterhin ist die Tendenz festzustellen, dass die Qualität fortwährend besser wird. Primär im Bereich der Filmbranche bedeutet die Qualitätsverbesserung auf Full High Definition bzw. Ultra High Definition einen bedeutsamen Innovationsschritt, da aufwändige Effekte schneller und kostensparender realisiert werden können.

Trotz allem dürfen die Risiken in der Social Media Umgebung nicht ignoriert werden. Es ist elementar, dass kontinuierlich an Erkennungsprozeduren geforscht wird, um die Nutzer zu schützen. Überdies ist als Schutz bzw. Absicherung anzuraten, dass eine (Deepfakes-)Gesetzeslage in der Bundesrepublik Deutschland redigiert wird, weil es hier an einer einheitlichen gesetzlichen Regelung fehlt.

Grundsätzlich sollte auch in der Gesellschaft und in den Unternehmen eine Skepsis gegenüber Bild-, Audio- und Videomaterialien größer werden. Dennoch lässt sich auch beurteilen, dass die offenbar größte Gefahr von Deepfakes nicht darin besteht, dass Menschen gefälschte Aufnahmen für real halten könnten, sondern kurioserweise im genauen Gegenteil - nämlich darin, dass sie echte Aufnahmen für gefälscht halten.

[15] Vgl. Lossau, N., 2020, S. 6.

Schlussendlich ist es relevant sich die notwendigen Kompetenzen anzueignen, um wahre von falschen Informationen differenzieren zu können. Dabei sollen nicht nur die fiktiv unendlichen Möglichkeiten der Manipulation verdeutlicht werden, sondern auch die Grenzen des Möglichen. Es muss zudem aufgezeigt werden, wie man mittels eigener Expertise oder Webangeboten wie Faktencheck-Seiten Deepfakes erkennen kann.

Quellenverzeichnis

Arlt, F./Arlt, H.-J. (2020): Spielen ist unwahrscheinlich - Eine Theorie der ludischen Aktion, 1. Aufl., Wiesbaden, Springer Verlag.

Gupta, B./Nedjah, N. (2021): Safety, Security, and Reliability of Robotic Systems - Algorithms, Applications, and Technologies, 1. Aufl., Broken Sound Parkway, CRC Press.

Heller, M./Porup, J. M. (2021): Was sind Deepfakes, URL: https://www.computerwoche.de/a/was-sind-deepfakes,3549772, Zugriff: 31.05.2021.

Hinterleitner, B. (2020): KI ist nicht künstlich, sondern menschlich. Ars Electronica. URL: https://ars.electronica.art/aeblog/de/2020/03/26/ai-is-human/**,** Zugriff: 16.06.2021.

Kaspersky (2021): Was Sie über Deepfakes wissen sollten? URL: https://www.kaspersky.de/resource-center/definitions/deepfakes, Zugriff: 12.06.2021.

Lossau, N. (2020): Deep Fake: Gefahren, Herausforderungen und Lösungswege, Berlin, Konrad Adenauer Stiftung, Analysen & Argumente. In: Digitale Gesellschaft, Nr. 382.

Müller, N. (2021): Gesundes Misstrauen. In: Protector.

Reischl, G. (2020): Internet of Crimes - Warum wir alle Angst vor Hackern haben sollten, 1. Aufl., München, Redline Verlag.

Sidyuk, A./Titova, D. (2021): Deepfake-Technologie positive Einsatzmöglichkeiten. Softeq Development Corp. URL: https://www.softeq.com/de/blog/deepfake-technologie-positive-einsatzmoeglichkeiten, **Zugriff: 16.06.2021.**

Trend, M. (2021): Kriminelle nutzen KI - nicht nur für Deepfakes. In: it&t business.

Walorska, A. M. (2020): Deepfakes & Desinformation, Potsdam: Friedrich-Naumann-Stiftung für die Freiheit, 05/2020.

Velten, A.-K. (2020): TikToks Maßnahmen zum Schutz der US-Präsidentschaftswahl, URL: https://www.absatzwirtschaft.de/tiktoks-massnahmen-zum-schutz-der-us-praesidentschaftswahl-174566/, Zugriff: 21.06.2021.