Aufsatz

Deepfakes als Gefahr für die Demokratie — eine rechtliche Einordnung

Eva Beute & Anna-Katharina Dhungel

Open Peer Review

Dieser Beitrag wurde lektoriert von:
Hannah Wissler & Hendrik Scheja

Sind Sie es leid, dass Politiker die Öffentlichkeit jedes Mal anlügen, wenn sie den Mund aufmachen? Möchten Sie in der Lage sein, das Gesicht einer Person auf den Körper einer anderen Person zu bearbeiten? Stellen Sie sich vor, jeder hätte die Möglichkeit, das Gesicht von jedem, den er will, so aussehen zu lassen, wie er will. Dieser Tag ist da. Diese Technologie wird ‚Deepfake' genannt. In einem Deepfake-Video werden die Form, die Größe und das Gewicht einer Person verwendet, um ein überzeugendes, dreidimensionales Abbild des Gesichts der Zielperson zu erstellen. Das bedeutet, dass ein gutes Deepfake über genügend Informationen verfügt, um eine überzeugende Nachbildung der Gesichtszüge einer Person zu erstellen.

Deepfakes sind aus Gründen, die man sich gut vorstellen kann, gefährlich. Vielleicht wird es benutzt, um Politiker zu erpressen. Vielleicht wird es genutzt, um politische Gegner zu schikanieren. Vielleicht wird es von Kriminellen verwendet, um ihre Opfer zu terrorisieren oder, wenn sie clever genug sind, um den Opfern vorzugaukeln, dass sie es mit anderen Personen zu tun haben. Vielleicht wird es eingesetzt, um die Öffentlichkeit zu täuschen oder sie etwas Falsches glauben zu lassen.

Der Text, den Du soeben gelesen hast, ist nicht ‚echt', sondern wurde von einer Maschine über die Plattform ***InferKit*** erzeugt.[1] Lediglich das Schlagwort ‚Deepfake' wurde eingegeben, der restliche Inhalt ist von einem System der künstlichen Intelligenz (KI) geschrieben worden. Anschließend wurde der Text – ebenfalls von einer KI, der Anwendung ***DeepL*** – vom Englischen in das Deutsche übersetzt.[2] Ähnlich ist es bei den Autorenbildern[3], hierbei handelt es sich nicht um die Autorinnen, sondern vielmehr existieren die abgebildeten Personen gar nicht. Es sind von künstlichen neuronalen Netzen hergestellte Fotos von nicht existierenden Personen, die über eine Website abgerufen werden können.[4]

A. Aufbau dieses Beitrags

Neben Texten und Bildern, die von KI-Systemen generiert werden, entsteht momentan eine neue Dimension der künstlich erzeugten Medien: Deepfakes. Ebenso wie von Maschinen erzeugte Texte und Bilder für echt gehalten werden können, ist es mittlerweile möglich, Videos zu generieren, die nie geschehene Sachverhalte täuschend echt inszenieren. Ziel dieses Beitrags ist es, die Erstellung und Verbreitung von Deepfakes rechtlich zu analysieren und zu bewerten. Der Fokus liegt dabei auf solchen Deepfakes, die zur Manipulation der öffentlichen Meinung und zur gezielten Beeinflussung von politischen Prozessen eingesetzt werden. Zu diesem Zweck wird zunächst eine Definition von Deepfakes und deren Funktionsweise vorgestellt sowie der aktuelle Stand der Forschung erörtert (vgl. Kapitel B). Hierauf aufbauend wird analysiert, welche Gefahren sich für eine Demokratie aus der Verbreitung von Deepfakes ergeben, unter welche Straftatbestände solche Fälle subsumiert werden können und welche weiteren Lösungsansätze für diese neue Herausforderung in Betracht kommen (siehe Kapitel C). Der Beitrag schließt mit einem Ausblick (siehe Kapitel D).

> „Dass online gerne Wirklichkeiten verdreht, aus dem Kontext gerissen oder verfälscht dargestellt werden, ist bekannt."

B. Ein Überblick

I. Kann man noch glauben, was man sieht?

Auf ***TikTok*** führt ein vermeintlicher ***Tom Cruise*** einen Zaubertrick vor oder zeigt, wie man seine Hände korrekt wäscht.[5] Im Film ***Star Wars: Rogue One*** taucht die verstorbene Schauspielerin ***Carrie Fisher*** plötzlich in einer Szene auf und in einem Video auf der Plattform ***Vimeo*** erklärt eine Person, die wie ***Kim Kardashian*** aussieht: „***I genuinly love the process of manipulating people online for money***."[6]

1 Hier abrufbar (Stand: 01.08.2022).
2 Hier abrufbar (Stand: 01.08.2022).
3 Zum Zwecke der besseren Lesbarkeit wird bei den personenbezogenen Hauptwörtern nur die männliche Form verwendet. Diese Begriffe sollen für alle Geschlechter gelten.
4 Hier abrufbar (Stand: 01.08.2022).
5 Hier abrufbar (Stand: 01.08.2022).
6 Black/Fullerton, Digital Deceit: Fake News, Artificial Intelligence, and Censorship in Educational Research, in: Open Journal of Social Sciences 08 (07): 71-88, hier abrufbar (Stand: 01.09.2022); Posters, 'When there's so many haters…', hier abrufbar (Stand: 19.08.2022). Der Künstler hat diverse Deepfake Videos erstellt, neben denen von Kim Kardashian auch u.a. von Mark Zuckerberg oder Boris Johnson, hier abrufbar (Stand: 01.09.2022).

Nicht immer entspricht das, was in einem Video suggeriert wird, der Wahrheit. Dass online gerne Wirklichkeiten verdreht, aus dem Kontext gerissen oder verfälscht dargestellt werden, ist mittlerweile bekannt. Der ***Duden*** nahm die Begrifflichkeit „***Fake News***" 2017 auf und versteht darunter: „***in den Medien und im Internet, besonders in sozialen Netzwerken, in manipulativer Absicht verbreitete Falschmeldungen***".[7] Welche gesellschaftlichen Auswirkungen diese Fake News haben können, zeigt etwa der Fall ‚***Pizzagate***': Während der US-Präsidentschaftswahl 2016 verbreitete sich online der Verschwörungsmythos[8], dass ***Hillary Clinton*** aus dem Keller einer Pizzeria in Washington D.C. einen internationalen Pädophilen-Ring leiten würde. Nachrichten rund um diese Theorie verbreiteten sich in den sozialen Medien rasant und führten schließlich dazu, dass sich ein bewaffneter Mann Zugang zu der besagten Pizzeria verschaffte, um die vermeintlich dort anzutreffenden Kinder zu retten. Er feuerte Schüsse auf Türen und einen Computer ab, musste dann jedoch feststellen, dass die Pizzeria noch nicht einmal über einen Keller verfügte.[9]

Eine neue Art der Fake News sind Deepfakes, deren gesellschaftliche Folgen sich bisher nur erahnen lassen. Was passiert, wenn in den oben genannten Fake-Videos nicht ***Tom Cruise***, ***Carrie Fisher*** oder ***Kim Kardashian*** zu sehen wären, sondern ein Afroamerikaner, der von einem weißen Polizisten brutal zusammengeschlagen wird oder ein Soldat, der einen Gefangenen foltert? Dass solche Videos enorme gesellschaftliche Auswirkungen haben können, zeigen vergleichbare (reale) Videos, etwa das der Tötung von ***George Floyd*** – welches in den USA landesweite Demonstrationen hervorrief und kürzlich mit dem Pulitzerpreis ausgezeichnet wurde[10] – oder die Foto- und Videoaufnahmen rund um die Folterungen durch US-Soldaten im irakischen Abu Ghraib, welche international für Entsetzen sorgten.[11]

Doch wie ist es nun ersichtlich, wann der Inhalt eines Videos ‚echt' ist, welche Folgen hat es, wenn Inhalte generell angezweifelt werden und welche rechtlichen Möglichkeiten hat ein Staat, um einer negativen gesellschaftlichen Entwicklung entgegenzuwirken?

7 Hier abrufbar (Stand: 17.08.2022).
8 Information zur Unterscheidung zwischen Verschwörungsmythos und Verschwörungstheorie hier abrufbar (Stand: 13.10.2022).
9 Bis heute gibt es Anhänger dieser Theorie, siehe beispielsweise Kalenberg, Wieso Menschen weiterhin an „Pizzagate" glauben, hier abrufbar (Stand: 17.08.2022).
10 Spanhel, Der Kampf gegen Rassismus geht weiter, hier abrufbar (Stand: 18.08.2022); Häntzschel, George-Floyd-Video bei Pulitzer-Preisen gewürdigt, hier abrufbar (Stand: 17.08.2022).
11 Wittwer, Abu Ghraib: Es kann jeden Tag wieder passieren, hier abrufbar (Stand: 19.08.2022).

II. Definition Deepfake

Westerlund definiert Deepfakes als *„hyper-realistic videos digitally manipulated to depict people saying and doing things that never actually happened."*[12] Zum Teil wird dies näher konkretisiert: *„In a deepfake video, a person's face, emotion or speech are replaced by someone else's face, different emotion or speech, using deep learning technology."*[13] Der Wörter-Zusammenschluss entsteht aus *„deep learning"* und *„fake"*, wobei ersteres eine spezielle Form der künstlichen Intelligenz meint: Deep Learning umschreibt künstliche neuronale Netze, die mittlerweile in vielen Anwendungen zum Einsatz kommen, etwa bei der Bild- und Spracherkennung, bei Zeitreihenanalysen oder beim Entdecken von Betrugsfällen. Im Rahmen von Deepfakes lernt ein künstliches neuronales Netz anhand großer Datenmengen die Mimik, Stimme, Tonlage und Eigenarten von zwei Personen, um das Gesicht der einen Person in einem Video für das originäre Gesicht auszutauschen.[14]

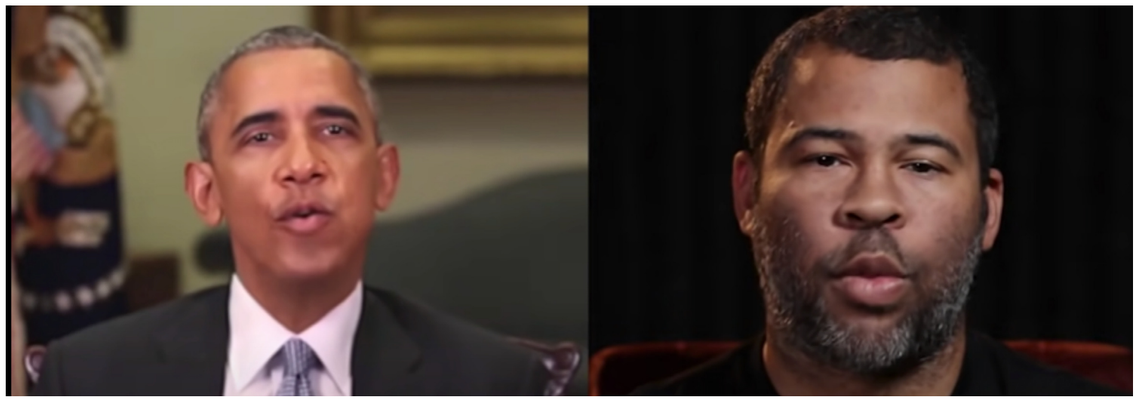
Teilweise werden Deepfakes zusätzlich kategorisiert in die Klassen *face-swap*, *lip-sync* und *puppet-master*.[15] Beim *Face-Swap* wird das Gesicht der einen Person auf den Kopf einer anderen Person transferiert. Eine besonders bekannte App für *Face-Swaps* ist beispielsweise *Snapchat*, aber auch andere Anbieter wie etwa *Instagram* bieten den Tausch und Transfer von Gesichtern an. Beim *Lip-Sync* wird in einem Video der gesprochene Inhalt einer Person mit einem neuen Audio hinterlegt und anschließend werden die Lippen- und Mundbewegungen dementsprechend angepasst. Bei dieser Methode lässt es sich in der Regel leichter erkennen, dass es sich um eine Fälschung handelt, weil die Lippenbewegungen nicht immer zur restlichen Gestik und Mimik der Person passen. Bei der *Puppet-Master-Methode* wird das Gesicht und der Körper der Zielperson im Video beibehalten, die Gesichtsbewegungen können jedoch komplett manipuliert werden. Sowohl die Stimme, als auch die Mimik können somit in Einklang gebracht werden.[16] *Wagner* und *Blewer* betonen, dass ein Deepfake mehr ist als ein Video, welches mittels Software entsteht. Vielmehr ist gerade das Besondere an Deepfakes, dass ein KI-System die Proportionen der Gesichter und deren Ausdrücke lernt und es somit möglich ist, neuen Inhalt mit diesen Gesichtern zu erstellen bzw. das Gesicht der einen Person durch das der anderen Person in beliebig vielen neuen Situationen auszutauschen.[17]

> „In a deepfake video, a person's face, emotion or speech are replaced by someone else's face, different emotion or speech, using deep learning technology."

Erstmals kamen Deepfakes in das öffentliche Bewusstsein, als ein Nutzer auf der Plattform *Reddit* 2017 ein Video teilte, in dem er das Gesicht der Schauspielerin *Gal Gadot* in pornografische Videos transferiert hatte.[18] Die Erstellung von gefälschter Pornografie ist eine der häufigsten Einsatzszenarien.[19] Es wird geschätzt, dass es sich bei rund 96 % aller Deepfakes um nicht einvernehmlich erstellte Pornografie handelt.[20] Daneben werden Deepfakes häufig in sozialen Netzwerken veröffentlicht, weil sie sich dort rasant verbreiten können und Nutzer dazu neigen, häufig geteilten Inhalten mehr zu glauben. Parallel dazu führt eine andauernde Informa-

12 Westerlund, 2019, The Emergence of Deepfake Technology: A Review, in: Technology Innovation Management Review 9 (11), 39 (40).
13 Mitra, Mohanty u.a., A Machine Learning based Approach for DeepFake Detection in Social Media through Key Video Frame Extraction, in: SN Computer Science, 2, 1.
14 Westerlund, 2019, The Emergence of Deepfake Technology: A Review, in: Technology Innovation Management Review 9 (11), 39 (40).
15 Agarwal, Farid u.a., Detecting Deepfake Videos from Appearance and Behavior, in: IEEE International Workshop on Information Forensics and Security (WIFS), New York, 1 (3), hier abrufbar (Stand: 17.08.2022).
16 Semaan, Die Demokratisierung von Deepfakes, hier abrufbar (Stand: 17.08.2022).
17 Wagner/Blewer, "The Word Real Is No Longer Real": Deepfakes, Gender, and the Challenges of AI-Altered Video, in: Open Information Science 3, 36.
18 Hier abrufbar (Stand: 13.10.2022).
19 *Hancock/Bailenson,* The Social Impact of Deepfakes, in: Cyberpsychology, Behaviour, and Social Networking 24 (3), 149 (150).
20 *Ajder, Patrini u.a.,* The State of Deepfakes: Landscape, Threats, and Impact, hier abrufbar (Stand 13.10.2022).

tionsüberflutung dazu, dass Nutzer quellenunabhängig alle Informationen anzweifeln und das Vertrauen in Medien signifikant sinkt – es sei denn, die Inhalte bestätigen eigene Meinungen und Weltansichten. In diesem Fall sind Personen sogar bereit, Inhalten zu glauben, selbst wenn es deutliche Hinweise auf einen Fake gibt.[21]



Screenshot eines Deep Fakes von Barack Obama / Jordan Peele

Das in der Abbildung referenzierte Video „***You Won't Believe What Obama Says In This Video!***" wurde von dem US-amerikanischen Medienunternehmen ***BuzzFeed*** 2018 auf ***YouTube*** veröffentlicht und verbreitete sich über die sozialen Netzwerke innerhalb kürzester Zeit.[22] Es zeigt zunächst nur ***Barack Obama***, der einige ungewöhnliche und überraschende Sätze sagt. Nach 35 Sekunden wird neben ihm der Schauspieler ***Jordan Peele*** eingeblendet und es wird deutlich, dass er die Stimme des ehemaligen US-Präsidenten ist. Das Video wurde auf ***YouTube*** über 9,1 Millionen Mal aufgerufen (Stand September 2022) und es war bereits Grundlage für wissenschaftliche Forschung.[23]

21 Westerlund, The Emergence of Deepfake Technology: A Review, in: Technology Innovation Management Review 9 (11), 39 (40).
22 Hier abrufbar (Stand: 17.08.2022).
23 Vaccari, Cristian; Chadwick, Andrew, Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News, in: Social Media + Society January-March, 1-13.

III. Funktionsweise und Erstellung

Bei den für Deepfakes genutzten künstlichen neuronalen Netzen handelt es sich konkret um sogenannte ***Generative Adversarial Networks*** (***GAN***). Diese bestehen aus zwei miteinander agierenden Systemen: dem ***Generator*** und dem ***Discriminator***. Beide trainieren mit denselben Datensätzen. Der Generator versucht ein Video zu erstellen, das so gut ist, dass der Discriminator es nicht als Fake erkennt. Wenn dieser den Fake erkennt, gibt er diese Information zurück und der Generator versucht erneut ein besseres Deepfake zu erstellen. Dieser Vorgang wird fortgesetzt, bis optimale Ergebnisse erreicht werden und der Discriminator das gefälschte Video als real einstuft.[24] Je mehr Zeit und Rechenkapazitäten vorhanden sind und je besser die Datenqualität ist, desto realistischer kann das Deepfake aussehen. Das System erstellt beim Training 3D-Modelle der Personen und ist dadurch in der Lage, auch Gesichtsausdrücke der Personen abzubilden, die vorher nicht im Datenbestand gespeichert waren. Basierend auf diesen Modellen wird für jedes Einzelbild des originalen Videos der Ausdruck der originären Person analysiert und dementsprechend wird das Gesicht der anderen Person angepasst und in das Video integriert.[25]

Das Angebot an Software, um qualitativ hochwertige Deepfakes zu erstellen, erhöht sich fortlaufend.[26] Teilweise wird eine leistungsfähige Grafikkarte benötigt, ansonsten liegen aber keine technischen Hindernisse vor – eine Entwicklung ist mittlerweile auf einem handelsüblichen Laptop möglich. Es gibt bereits Anwendungen wie die chinesische App ***ZAO,*** mit der man selbst auf einem Smartphone Deepfakes erzeugen kann – und das mit nur einem Foto und mit einer Schnelligkeit, die selbst Experten überrascht.[27] Häufig sind die Anwendungen frei verfügbar und ermöglichen es

24 Westerlund, The Emergence of Deepfake Technology: A Review, in: Technology Innovation Management Review 9 (11), 39 (41).
25 Wagner/Blewer, "The Word Real Is No Longer Real": Deepfakes, Gender, and the Challenges of AI-Altered Video, in: Open Information Science 3, 32 (36).
26 Eine anschauliche Übersicht der aktuell verfügbaren Software zur Erstellung von Deepfakes findet sich bei Nguyen, Nguyen u.a., Deep Learning for Deepfakes Creation and Detection: A Survey, 1 (3).
27 Der Standard, Zao: Aufregung um neue Deepfake-App, die erschreckend gute Resultate erzeugt, hier abrufbar (Stand: 16.08.2022).

auch Nutzern mit geringen technischen Fähigkeiten, Videos zu erstellen, bei denen die ausgetauschte Person äußerst überzeugend in der Gestik, den Gesichtszügen und der Stimme ist.[28]

IV. Status zur Identifikation von Deepfakes

Weltweit wird nicht nur an der Erstellung, sondern auch an der Identifikation von Deepfakes gearbeitet und geforscht. Allerdings wird das Verhältnis der Anzahl an Personen, die an der Erstellung von Deepfakes arbeiten, gegenüber der Anzahl an Personen, welche an deren Erkennung und Identifizierung arbeiten, als 100 zu eins geschätzt.[29] Ein Experte beschreibt die Situation wie folgt: *„We are witnessing an arms race between digital manipulations and the ability to detect those, and the advancements of AI-based algorithms are catalyzing both sides."*[30]

	2018	2019	2020	2021	2022*
Science Direct	2	2	5	21	28
Scopus	0	27	151	329	212
IEEE Xplore	3	18	82	126	16
JSTOR	3	20	75	21	20
arXiv	4	11	62	97	100

*Stand September 2022

Anzahl wissenschaftlicher Veröffentlichungen über Deepfakes pro Jahr

Im Bereich der digitalen Multimediaforensik wird seit etwa Mitte der 2000er Jahre intensiv geforscht, wie man die Echtheit medialer Inhalte verifizieren kann. Häufig geht es dabei jedoch um vergleichsweise einfache Manipulationen medialer Daten. Die per künstlichen neuronalen Netzen erzeugten Deepfakes stellen eine neue Herausforderung dar, weil sie sich nicht mit bisherigen Methoden identifizieren lassen.[31] Derzeit entsteht ein eigener Forschungszweig, in welchem ebenfalls mithilfe solcher Netze nach Möglichkeiten gesucht wird, Deepfakes zu erkennen. In Tabelle 1 wird die Anzahl relevanter Studien seit 2018 dargestellt, die sich mit dem Thema Deepfakes beschäftigen.[32]

> „We are witnessing an arms race between digital manipulations and the ability to detect those, and the advancements of AI-based algorithms are catalyzing both sides."

Schaut man sich die Zahlen rund um wissenschaftliche Veröffentlichungen zu KI an, so kann bei Deepfakes als Teilanwendungsbereich von KI die Kritik geäußert werden, dass das Thema wissenschaftlich vernachlässigt oder sogar ignoriert wird. Eine der Schwierigkeiten im Bereich Deepfakes ist aus Forschungssicht die schwache Datenlage. Während es für Bildmanipulationen zahlreiche Datensätze gibt, sind größere Deepfake-Datensätze bisher rar. Nennenswert sind bis dato nur zwei Datensätze von ***Google*** und ***Facebook's*** Mutterkonzern ***Meta Platforms***.[33] ***Google*** hat gemeinsam mit der konzerneigenen Tochtergesellschaft ***Jigsaw*** einen

28 Westerlund, The Emergence of Deepfake Technology: A Review, in: Technology Innovation Management Review 9 (11), 39 (40).
29 Galston, Is seeing still believing? The deepfake challenge to truth in politics, hier abrufbar (Stand: 17.08.2022).
30 Hao Li, zitiert nach Knight, Will, A New Deepfake Detection Tool Should Keep World Leaders Safe—for Now, hier abrufbar (Stand: 17.08.2022).
31 Mitra/ Mohanty u.a., A Machine Learning based Approach for DeepFake Detection in Social Media through Key Video Frame Extraction, in: SN Computer Science, 2, 1 (4).
32 Anschauliche Übersichten zu aktuellen Studien finden sich bei: Nguyen/Nguyen u.a., Deep Learning for Deepfakes Creation and Detection: A Survey, 1 (2); Mitra/Mohanty u.a., A Machine Learning based Approach for DeepFake Detection in Social Media through Key Video Frame Extraction, in: SN Computer Science, 2, 1 (9).
33 Mitra/Mohanty u.a., A Machine Learning based Approach for DeepFake Detection in Social Media through Key Video Frame Extraction, in: SN Computer Science, 2, 1 (10).

großen Datensatz an authentischen und manipulierten Videos zusammengestellt.[34] ***Meta Platforms*** hat in Zusammenarbeit mit ***Microsoft*** und mehreren Universitäten aus den USA, dem Vereinigten Königreich, Deutschland und Italien im Jahr 2019 die ***Deepfake-Detection-Challenge*** gestartet und stellte im Rahmen davon einen Datensatz mit über 100.000 Deepfake-Videos bereit. Über 2.000 Forschende reichten ihre Ergebnisse ein, wobei die besten Lösungen circa 80 % der Deepfakes in dem Datensatz identifizieren konnten.[35] Daneben gibt es etliche einzelne Lösungsansätze; einige dieser Ansätze werden im Folgenden kurz vorgestellt.

Auf der ***IEEE Conference on Computer Vision and Pattern Recognition*** stellten ***Agarwal, Farid u.a.*** eine Möglichkeit vor, hochrangige Politiker vor Deepfakes zu schützen.[36] Sie erstellten Pseudo-Modelle über die Art und Weise, wie diese Personen sprechen und erzeugten damit eine Art Fingerabdruck über die individuelle Gestik und Mimik. Grundsätzlich liefert das Modell gute Ergebnisse, die Fehlerrate nimmt allerdings zu, wenn die Person in dem Video nicht direkt in die Kamera blickt. Außerdem ist dieser Detektor nur für Deepfakes von Personen relevant, die besonders schutzwürdige Positionen besetzen. Es fehlen ferner Angaben darüber, wie aufwändig es ist, diese Pseudo-Modelle zu erstellen.[37] ***Microsoft*** stellte im September 2020 seinen ***Video-Authenticator*** vor. Dieser gibt eine Wahrscheinlichkeit an, nach der es sich bei dem vorliegenden Video um künstlich manipulierten Inhalt handelt.

Auch wenn das Thema in der Wissenschaft langsam an Bedeutung zunimmt, wird es politisch bisher noch größtenteils vernachlässigt, obwohl es bereits eine Studie gibt, in der ein politischer Einfluss von Deepfakes nachgewiesen wurde.[38] Innerhalb der ***Europäischen Union*** sind Projekte rund um das Thema Fake News personell stark unterbesetzt und der ***EU*** wird regelmäßig vorgeworfen, dass sie sich zwar mit Fake News beschäftige, aber kaum mit dem Thema Deepfakes.[39] Ein Bericht im Auftrag des ***Europäischen Parlaments*** lässt sich jedoch finden, in welcher der aktuelle Status, die Chancen und Risiken sowie politische Handlungsmöglichkeiten zusammengefasst werden.[40] ***Europol*** hat in einem aktuellen Bericht die Herausforderungen von Deepfakes ausführlich beschrieben. Danach würde der Einsatz von Deepfakes bei Kriminellen immer beliebter. Eines der Ziele sei es dabei, die öffentliche Meinung zu manipulieren und falsche Informationen zu verbreiten.[41] ***Europol*** stellt fest: „***The increase in use of deepfakes will require legislation to set guidelines and enforce compliance***."[42]

34 Dufour/Gully, Contributing Data to Deepfake Detection Research, hier abrufbar (Stand: 18.08.2022).
35 Skibba, Accuracy Eludes Competitors in Facebook Deepfake Detection Challenge, in: Engineering, 6 (12), 1339 (1339-1340).
36 Agarwal/Farid u.a., Detecting Deep-Fake Videos from Appearance and Behavior, in: IEEE International Workshop on Information Forensics and Security (WIFS), New York, 1 (3), hier abrufbar (Stand: 17.08.2022).
37 Solsman, Deepfake Debunking Tool May Protect Presidential Candidates. For Now. Sometimes, hier abrufbar (Stand: 18.08.2022).
38 *Dobber, Metoui u.a.,* Do (Microtargeted) Deepfakes Have Real Effects on Political Attitudes?, The International Journal of Press/Politics, 26 (1), (69–91).
39 *Bressan*, Can the EU Prevent Deepfakes From Threatening Peace?, hier abrufbar (Stand: 18.08.2022).
40 *EPRS (European Parliamentary Research Service), Scientific Foresight Unit (STOA),* "Tackling deepfakes in European policy, hier abrufbar (13.10.2022).
41 Europol, Facing reality? Law enforcement and the challenge of deepfakes. An Observatory Report from the Europol Innovation Lab. Publications Office of the European Union, 1 (10), hier abrufbar (Stand: 16.08.2022).
42 Europol, Facing reality? Law enforcement and the challenge of deepfakes. An Observatory Report from the Europol Innovation Lab. Publications Office of the European Union, 1 (22), hier abrufbar (Stand: 16.08.2022).

C. Deepfakes und Rechtswissenschaften

Die Entwicklung und Verbreitung von immer hochwertigeren Deepfakes führt zwangsläufig zu der Frage, wie das Recht diesem Phänomen begegnen kann. Was ist, wenn Deepfakes eingesetzt werden, um parlamentarische Entscheidungen zu manipulieren oder Wahlen zu beeinflussen, indem man versucht, Politiker zu diskreditieren? Was ist, wenn sie Mittel zum Zweck werden, um Menschen gegeneinander aufzuhetzen oder populistische Strömungen zu verstärken? Ist das Recht für solche Szenarien gewappnet?

I. Gefahren für die Staatssicherheit und die Demokratie

Aktuelle Beispiele zeigen, dass Deepfakes inzwischen zur politischen Destabilisierung und Manipulation der Meinungsbildung eingesetzt werden. Im Russland-Ukraine-Krieg wird derzeit nicht nur mit echten, sondern auch mit medialen Waffen gekämpft. Ende März 2022 tauchte in den sozialen Medien ein Video auf, in dem der ukrainische Präsident ***Wolodymyr Selenskyj*** ukrainische Soldaten dazu aufruft, sich zu ergeben: Der Krieg sei verloren. Das Video ist eine Fälschung, was der ***Facebook***-Konzern ***Meta Platforms*** schnell bemerkte.[43] Selbst wenn das Video nur kurze Zeit online war, dürfte es in der angespannten Kriegssituation für weitere Verunsicherungen bei einigen Ukrainern gesorgt haben.

Auch Wahlen könnten zukünftig durch verfälschte Videos beeinflusst werden. Angenommen, kurz vor einer Wahl taucht ein Deepfake eines Spitzenkandidaten auf, in welchem er sich vermeintlich rassistisch oder sexistisch äußert. Klar ist, dass die Folgen für den Spitzenkandidaten verheerend wären. Eine Rehabilitation innerhalb kürzester Zeit erscheint fast aussichtslos. Zumindest für die betreffende Wahl wäre der Kandidat de facto ausgeschlossen. Deepfakes können damit zu einer großen Bedrohung für die Demokratie und die Gesellschaft werden.

43 Metzger/Schneider, Wie Deepfakes im Ukraine-Krieg genutzt werden, hier abrufbar (Stand: 16.08.2022).

II. Rechtliche Herausforderungen

1. Status Quo: Welche Gesetze finden Anwendung?

Die gesamte Materie der Künstlichen Intelligenz unterliegt bislang noch keinem eigenen Regelungsregime. Daher ist es nicht verwunderlich, dass Deepfakes als spezielle Form von Künstlicher Intelligenz dem deutschen Recht völlig unbekannt sind. Natürlich gibt es Vorschriften im deutschen Recht, die auch ‚Deepfake-Fälle' erfassen. Speziell auf Künstliche Intelligenz und damit auch Deepfakes zugeschnittene Regelungen fehlen jedoch bislang.

Die Bundesregierung hält vor allem die Stärkung der Medienkompetenz, insbesondere der Nachrichten- und der digitalen Informationskompetenz, für entscheidend, um gegen Desinformation im Allgemeinen und Deepfakes im Besonderen gewappnet zu sein.[44] Nationale Gesetze zur Regulierung von Deepfakes scheinen daher nicht geplant zu sein. Problematisch ist dies vor allem für Deepfakes, die zu politischen Zwecken eingesetzt werden. Denn während pornografische und vermögensschädigende Deepfakes unter mehrere Straftatbestände subsumiert werden könnten,[45] ist dies bei politisch motivierten Deepfakes häufig nicht der Fall.

a) § 201a Abs. 2 StGB

In Betracht kommt zunächst § 201a Abs. 2 StGB.

> **§ 201a Abs. 2 StGB:** „Ebenso wird bestraft, wer unbefugt von einer anderen Person eine Bildaufnahme, die geeignet ist, dem Ansehen der abgebildeten Person erheblich zu schaden, einer dritten Person zugänglich macht."

44 Antwort der Bundesregierung auf die Kleine Anfrage der Abgeordneten Manuel Höferlin, Frank Sitta, Grigorios Aggelidis, weiterer Abgeordneter und der Fraktion der FDP – Drs. 19/15210 Beschäftigung der Bundesregierung mit Deepfakes, 5, hier abrufbar (Stand: 23.08.2022).
45 Nähere Ausführungen hierzu in: Lantwin, MMR 2020, 78 ff.

Hier könnte bereits fraglich sein, ob ein Deepfake eine Bildaufnahme im Sinne des § 201a Abs. 2 StGB darstellt. Unter den Begriff der „Bildaufnahme" fallen hauptsächlich Fotos und Videos. Erforderlich ist, dass eine andere Person aufgenommen wird. Daher sind Karikaturen, Zeichnungen oder auch rein computergenerierte Bilder – wie beispielsweise die Fake-Bilder unserer Autorinnen – nicht erfasst.[46] Bei Deepfakes wird eine Person nicht im klassischen Sinne aufgenommen. Ein Deepfake erschafft aber gerade eine solche Bildaufnahme. Denn im Ergebnis erscheint es so, als sei die betroffene Person aufgenommen worden, da Deepfakes es ermöglichen, täuschend echt nie geschehene Sachverhalte zu inszenieren. Da der Gesetzgeber mit der Einführung des § 201a StGB der technischen Entwicklung und der damit einhergehenden Bedrohung des allgemeinen Persönlichkeitsrechts entgegentreten wollte, und ein Deepfake im besonderen Maße eine Gefahr für das von § 201a StGB geschützte Rechtsgut darstellt, wird man Deepfakes daher unter den Begriff subsumieren können. Weiter setzt der Tatbestand eine Ansehensschädigung voraus. Es geht um solche Aufnahmen, welche die abgebildete Person in einer peinlichen, ihre Würde verletzenden Situation zeigen. Im Hinblick auf die Zugänglichmachung von pornografischen Deepfakes wird der Tatbestand regelmäßig zu bejahen sein. Nicht einvernehmlich erstellte pornografische Inhalte dürften geeignet sein, dem Ansehen der abgebildeten Person erheblich zu schaden.[47] Das kann für politische Deepfakes aber nicht so pauschal bejaht werden. Denn nach der Gesetzesbegründung sollen insbesondere Aufnahmen erfasst werden, bei denen nach allgemeiner gesellschaftlicher Bewertung angenommen werden kann, dass ein Interesse daran besteht, die Aufnahmen nicht Dritten zugänglich zu machen.[48] Damit wollte der Gesetzgeber unter anderem ein Signal gegen das immer stärker um sich greifende Cybermobbing setzen und hatte ganz offensichtlich einen anderen Anwendungsfall als politische Deepfakes im Blick.

46 Eisele, in: Schönke/Schröder, Strafgesetzbuch, 30. Aufl. 2019, § 201a Rn. 6.
47 ebd., 79.
48 Entwurf eines Gesetzes zur Änderung des Strafgesetzbuches – Umsetzung europäischer Vorgaben zum Sexualstrafrecht, BT-Drucks. 18/2601, 37.

Deepfakes, die mit dem Ziel eingesetzt werden, die Meinungsbildung zu beeinflussen oder für politische Unruhen zu sorgen, werden sich daher nur selten unter den Tatbestand subsumieren lassen. Bei dem Deepfake des ukrainischen Präsidenten beispielsweise geht mit der Äußerung, der Krieg sei verloren, keine Ansehensschädigung einher. § 201a Abs. 2 StGB schützt nämlich nicht das Ansehen selbst. *„Denn es gibt […] kein Recht des Einzelnen auf einen bestimmten Grad der Wertschätzung seiner Person durch Dritte."*[49] Gleiches gilt etwa für den vermeintlichen Aufruf, nicht wählen zu gehen. Allein diese Äußerungen zeigen eine Person nicht in einer peinlichen, ihre Würde verletzenden Situation. *„Die Eignung zur Schädigung muss nach dem Wortlaut des § 201a Abs. 2 StGB aber allein aus der Bildaufnahme resultieren; es genügt nicht, dass das Ansehen erst durch die Art und Weise des Zugänglichmachens, etwa durch hämische Kommentare, geschädigt werden kann."*[50] Trotz der Schädigung des politischen Ansehens scheidet in solchen Fällen eine Strafbarkeit nach § 201a Abs. 2 StGB aus. Die politische Integrität und Karriere des Geschädigten werden von § 201a Abs. 2 StGB somit nicht an sich geschützt.

b) §§ 185 ff. StGB

Die Straftatbestände der §§ 185, 186 und 187 StGB bezwecken den Schutz der persönlichen Ehre. Politische Deepfakes enthalten jedoch häufig keine ehrverletzenden Äußerungen, da sie regelmäßig nicht den Zweck verfolgen, einzelne Personen zu diffamieren, sondern vielmehr eine politische Destabilisierung herbeizuführen. Das tatbestandsmäßige Verhalten in § 185 StGB wird als *„Beleidigung"* beschrieben, ohne diesen Begriff näher zu erläutern. Nach ständiger Rechtsprechung des BGH liegt eine solche bei einem Angriff auf die Ehre einer Person durch Kundgabe von Missachtung oder Nichtachtung vor.[51] *„Als Äußerungsdelikt erfordert die Beleidigung also die Kundgabe der ehrverletzenden Tatsachenbehauptung bzw. des herabwürdigenden Werturteils."*[52] Bei politischen Deepfakes, wie dem des ukrainischen

49 Altenhain, in: Matt/Renzikowski, 2. Aufl. 2020, StGB, § 201a Rn. 21.
50 Eisele/Sieber, Strafverteidiger 2015, 312 (315).
51 BGHSt 11, 67; BGHSt 36, 145.
52 Valerius, in: BeckOK, StGB, 52. Ed., § 185 Rn. 17.

Präsidenten, fehlt es an einer solchen Kundgabe. Der Ersteller dieses Deepfakes bringt weder seine Miss- oder Nichtachtung zum Ausdruck, noch nutzt er ihn, um seine Missachtung gegenüber einer anderen Person kundzutun. Vielmehr bezweckt der Täter, den Rezipienten zu täuschen und damit deren Willensbildung zu manipulieren. Ein ehrverletzender Sinn lässt sich der vermeintlichen Äußerung des Präsidenten, die Soldaten sollen ihre Waffen niederlegen, nicht beimessen.

Die Straftatbestände der §§ 186, 187 StGB setzen voraus, dass Tatsachen geäußert werden, die geeignet sind, den Betroffenen verächtlich zu machen oder in der öffentlichen Meinung herabzuwürdigen. Hierbei kann auf die Rechtsprechung zu ‚Fake News' zurückgegriffen werden. Danach ist das ‚In-den-Mund-Legen' von politischen Statements nur dann strafbar, wenn die Zuschreibung des Zitats ehrverletzenden Charakter hat. „*Die bisherige Rechtsprechung deutet darauf hin, dass Äußerungen, die zwar in Teilen der Bevölkerung Empörung gegen den Betroffenen auslösen, aber im Einklang mit der Rechtsordnung stehen, in der Regel nicht die Grenze zur Ehrverletzung überschreiten. Damit ist die Beeinträchtigung etwa der öffentlichen Reputation eines Politikers nicht geschützt – trotz möglicherweise gravierender Schäden für die persönliche politische Karriere oder den Wahlerfolg der betroffenen Partei.*"[53]

> „Für den Wähler ist es aufgrund der Qualität der Deepfakes aber schlicht unmöglich, ein solches Video als Fake zu identifizieren."

53 Hoven/Krause, JuS 2017, 1167 (1169).

c) § 108a StGB

Besonderes Gefahrenpotenzial entfalten Deepfakes kurz vor einer Wahl. Denn hier kann die Manipulation unmittelbar Wirkung entfalten. Sehen Wähler in dieser Situation Deepfakes, wie die von ***Boris Johnsen*** und ***Jeremy Corbyn***[54], in denen die beiden den jeweils anderen als nächsten Premierminister empfehlen, kann sich dies auf die Willensbildung der Wähler auswirken. Sie könnten hierdurch zu einer Wahlentscheidung bewegt werden, welche sie sonst nicht getroffen hätten.

Für den Wähler ist es aufgrund der Qualität der Deepfakes aber schlicht unmöglich, ein solches Video als Fake zu identifizieren. Strafrechtliche Sanktionen für den Ersteller ergeben sich allerdings auch aus § 108a StGB nicht.

Nach § 108a StGB wird bestraft, wer durch Täuschung bewirkt, dass jemand bei der Stimmabgabe über den Inhalt seiner Erklärung irrt oder gegen seinen Willen nicht oder ungültig wählt. § 108a StGB schützt jedoch nur vor einer Täuschung beim Akt der Wahl selbst, das heißt bei der Stimmabgabe.[55] Dadurch wird die Entscheidungsfreiheit des Wählers, nicht aber seine Willensbildungsfreiheit vor Täuschung geschützt. Unwahre Wahlpropaganda wird folglich nicht vom Tatbestand erfasst.[56]

d) Strafrechtliche Nebengesetze

§ 33 Kunsturhebergesetz (KUG) stellt die Verbreitung und öffentliche Zurschaustellung eines Bildnisses entgegen den §§ 22, 23 KUG unter Strafe. Fraglich ist bereits, ob ein Deepfake ein Bildnis in diesem Sinne darstellt. Ein Bildnis im Sinne des KUG ist ein Personenbildnis. Das heißt die Darstellung einer oder mehrerer Personen, welche die äußere Erscheinung der Abgebildeten in einer für Dritte erkennbaren Weise wiedergibt.[57] Schutzgut des KUG ist das Selbstbestimmungsrecht der abgebildeten Person. „*Erfasst werden soll die Freiheit des Menschen, ausschließlich selbst über*

54 Hier abrufbar (Stand: 09.09.22).
55 Kühl, in: Lackner/Kühl, Kommentar zum Strafgesetzbuch, 29. Aufl. , § 108a Rn.1.
56 v. Heintschel-Heinegg, in: BeckOK, StGB, 52. Ed., § 108a Rn. 1.
57 Götting, in: Schricker/Loewenheim, Kommentar zum Urheberrecht, 6. Aufl., § 22 Rn. 14.

sein dem höchstpersönlichen Lebensbereich zuzuordnendes Erscheinungsbild zu bestimmen."[58] Nach Ansicht von ***Hartmann*** soll dem Einzelnen aber kein allgemeines Verfügungsrecht über die eigene Darstellung zustehen. Deepfakes sind aus seiner Sicht „***keine Derivate des Selbstdarstellungsrechts, sondern originäre Fremddarstellung***." Damit unterliegen sie seiner Ansicht nach nicht dem KUG.[59]

„Eine Ausgestaltung als Antragsdelikt ist bei politisch motivierten Deepfakes zudem wenig zweckmäßig."

Soweit man Deepfakes entgegen dieser Ansicht die Eigenschaft als Bildnis zuerkennt, dürfte diese Norm politisch motivierte Deepfakes erfassen. Denn eine Einwilligung in die Verbreitung und öffentliche Zurschaustellung wird regelmäßig nicht vorliegen und eine Ausnahme im Sinne des § 23 KUG dürfte zumindest aufgrund von Abs. 2 ebenfalls zu verneinen sein. Gemäß § 23 Abs. 2 KUG erstreckt sich die Befugnis nämlich nicht auf eine Verbreitung und Schaustellung, durch die ein berechtigtes Interesse des Abgebildeten verletzt wird. So kann insbesondere die Veröffentlichung von manipulierten Aufnahmen unzulässig sein, wenn der Aussagegehalt der Abbildung verfälscht worden ist.[60] „***Insoweit kann es an dem legitimen Informationsinteresse der Öffentlichkeit fehlen, weil unrichtige Informationen grundsätzlich nicht als schützenswertes Gut anzusehen sind***."[61] Die als Antrags- und Privatklagedelikt ausgestaltete Norm hat in der Praxis bislang jedoch so gut wie keine Bedeutung.[62]

58 Herrmann, in: BeckOK, InfoMedienR, 35. Ed., KunstUrhG § 22 Rn. 3.
59 Hartmann, Kommunikation & Recht 2020, 350 (353).
60 BVerfG, NJW 2005, 3271 (3273).
61 LG Frankfurt am Main, ZUM-RD 2020, 329 (335).
62 Specht-Riemenschneider, in: Dreier/Schulze, Kommentar zum Urheberrechtsgesetz, 7. Aufl., KUG, §§ 33-50 Rn. 3.

Der Schutz des KUG kommt zudem nur für solche Deepfakes in Betracht, die uneingeschränkt sichtbar für alle sind. Das heißt nur in solchen Fällen, in denen das Deepfake-Video auf einer öffentlich zugänglichen Plattform hochgeladen wird und damit einem nicht begrenzten Personenkreis zugänglich gemacht wird. Wird das Video hingegen in einer Benutzergruppe geteilt, unterfällt es der Norm nicht. Unsicherheiten bestehen zudem in Hinblick auf den Begriff des Bildnisses. „***Als adäquate Sanktionsfolgen für die Ahndung von Persönlichkeitsrechtsverletzungen eignen sich diese Instrumentarien daher nur teilweise.***"[63]

Im Hinblick auf §§ 106, 108 Urheberrechtsgesetz (UrhG) und § 42 Bundesdatenschutzgesetz (BDSG) dürften politische Deepfakes nicht anders zu behandeln sein als pornografische und vermögensschädigende Deepfakes, sodass eine Strafbarkeit zu bejahen ist. Gerade für politische Deepfakes wird der Ersteller auf fremdes Bild und Videomaterial zurückgreifen müssen, sodass die §§ 106, 108 UrhG einschlägig sind. „***Zudem stellen die Gesichtszüge einer Person personenbezogene Daten i.S.d. Art. 4 Nr. 1 DS-GVO dar, weswegen sich eine Strafbarkeit in Einzelfällen aus Datenschutzstrafrecht ergeben kann.***"[64] Zu beachten ist allerdings, dass insbesondere bei politisch motivierten Deepfakes häufig auf allgemein zugängliches Bildmaterial zurückgegriffen wird, sodass eine Strafbarkeit aus § 42 Abs. 2 Nr. 1 BDSG nur selten in Betracht kommt.

Folglich eignen sich diese Vorschriften aus dem Nebenstrafrecht regelmäßig nicht als Instrumentarien für die Ahndung von Persönlichkeitsrechtsverletzungen zur Verhütung von Manipulation und politischer Destabilisierung. Straftaten aus dem Urheberrecht sind ebenfalls ‚nur' Antragsdelikte, § 109 UrhG. Es sind nur dann Offizialdelikte, wenn gewerbsmäßiges Handeln vorliegt, § 108 a, § 108 b Abs. 2, 3 UrhG. § 42 BDSG ist sogar ein absolutes Antragsdelikt. Damit unterliegen sie einer kurzen Frist im Hinblick auf die Strafantragstellung und ein Strafantrag darf nur von der verletzten Person gestellt werden. Trotz der Gefahren für die Demokratie hat der

63 Heuchemer, in: BeckOK, 52. Ed., Der strafrechtliche Schutz des Persönlichkeitsrechts, StGB, Rn. 18.
64 Insoweit wird auf den bereits zitierten Aufsatz von Lantwin verwiesen.

Staat damit nur sehr eingeschränkt die Möglichkeit, diese Taten zu sanktionieren. Eine Ausgestaltung als Antragsdelikt ist bei politisch motivierten Deepfakes zudem wenig zweckmäßig. Straftatbestände, die sich auf Eingriffe in die Privatsphäre des

§ 106 Abs. 1 UrhG: „Wer in anderen als den gesetzlich zugelassenen Fällen ohne Einwilligung des Berechtigten ein Werk oder eine Bearbeitung oder Umgestaltung eines Werkes vervielfältigt, verbreitet oder öffentlich wiedergibt, wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft."

Verletzten oder Verletzung von Rechtsgütern mit ausgeprägtem Persönlichkeitsbezug beziehen, sind in der Regel als Antragsdelikte ausgestaltet, weil der Verletzte oftmals ein Interesse daran hat, dass der Fall nicht in einem Strafverfahren erörtert wird. Dadurch wird der auf der Tat beruhende Verletzungseffekt häufig nur noch verstärkt.[65] Bei politisch motivierten Deepfakes besteht ein solches Interesse des Verletzten in der Regel jedoch gerade nicht. Solche Deepfakes wurden bereits regelmäßig der breiten Öffentlichkeit zugänglich gemacht. Durch ein anschließendes Strafverfahren kommt es daher nicht mehr zu einer Verstärkung des Verletzungseffekts. Vielmehr wird ein gerichtliches Verfahren die einzige Möglichkeit für die Rehabilitation der Verletzten sein.

2. Mögliche Lösungsansätze

Das vorige Kapitel zeigt, dass für die Zukunft noch Regelungsbedarf für Deepfakes besteht. Allerdings wird eine strafrechtliche Sanktionierung allein nicht ausreichen, um die Erstellung und Verbreitung von Deepfakes zu verhindern. Um den freien Diskurs und die Meinungsbildung zu schützen, sind weitere Maßnahmen erforderlich, die nachfolgend diskutiert werden.Da die Verbreitung von Falschnachrichten in der Regel kein schützenswertes Verhalten darstellt, kommt eine weitergehende Pönalisierung grundsätzlich in Betracht.[66] Vom Schutzbereich des Art. 5 Abs. 1 Satz 1 GG ausgeschlossen sind nach ständiger Rechtsprechung[67] bewusst unwahre Tatsachenbehauptungen („***bewusste Lüge***") und solche, deren Unwahrheit bereits im Zeitpunkt der Äußerung unzweifelhaft feststeht.[68] Dennoch ist es für den Gesetzgeber eine schmale Gratwanderung zwischen rechtsstaatlicher Selbstbehauptung und totalitären Tendenzen. Es wird nur schwer möglich sein, den gesamten öffentlichen Meinungsbildungsprozess zu schützen, ohne den Bürger in seiner Meinungsfreiheit einzuschränken. Im Hinblick auf den Schutz des staatlichen Willensbildungsprozesses könnte aber der Vorschlag von ***Mafi-Gudarzi*** ein geeignetes Mittel darstellen.[69] So wäre eine Ergänzung von § 108a StGB (Wählertäuschung) oder die Schaffung eines neuen § 108f StGB in Bezug auf die Verbreitung falscher Tatsachen, die geeignet sind, den Wählerwillen zu beeinflussen, zu erwägen.[70]

> „Dennoch ist es für den Gesetzgeber eine schmale Gratwanderung zwischen rechtsstaatlicher Selbstbehauptung und totalitären Tendenzen."

In der ***Europäischen Union*** wird ebenfalls nicht über ein Verbot nachgedacht. Der Vorschlag der EU-Kommission zur ***Festlegung harmonisierter Vorschriften für***

65 Mitsch, JA 2014, 1 (2).

66 Mafi-Gudarzi, ZRP 2019, 65 (67).
67 BVerfGE 61, 1 (8); 99, 185 (197).
68 Holznagel, MMR 2018, 18 (20).
69 Mafi-Gudarzi, ZRP 2019, 65 (68).
70 Mafi-Gudarzi verweist diesbezüglich auf § 264 Abs.1 des österreichischen StGB gegen die Verbreitung falscher Nachrichten bei Wahlen: „Wer öffentlich eine falsche Nachricht über einen Umstand, der geeignet ist, Wahl- oder Stimmberechtigte von der Stimmabgabe abzuhalten oder zur Ausübung des Wahl- oder Stimmrechts in einem bestimmten Sinn zu veranlassen, zu einer Zeit verbreitet, da eine Gegenäußerung nicht mehr wirksam verbreitet werden kann, ist mit Freiheitsstrafe bis zu sechs Monaten oder mit Geldstrafe bis zu 360 Tagessätzen zu bestrafen."

Künstliche Intelligenz sieht für den Einsatz von Deepfakes sogar nur minimale Transparenzpflichten vor.[71] Der Ersteller des Deepfakes muss lediglich darauf hinweisen, dass es sich um ein Deepfake handelt. Vielversprechender erscheint da der ***Gestärkte Verhaltenskodex für Desinformation***.[72] Vertreter von Online-Plattformen, führenden Technologieunternehmen und Akteuren der Werbebranche haben sich erstmals weltweit und auf freiwilliger Basis auf Selbstregulierungsstandards zur Bekämpfung von Desinformation geeinigt. Unterzeichner des Kodex müssen danach Maßnahmen zur Eindämmung von Desinformation ergreifen. Der Verhaltenskodex erfasst explizit auch neue manipulative Verhaltensweisen wie Deepfakes. Der ***Verhaltenskodex für Desinformation*** soll mit dem ***Gesetz über digitale Dienste*** (***Digital Services Act***)[73] verknüpft werden. Unternehmen, die ihren Verpflichtungen im Rahmen des aktualisierten Kodex nicht nachkommen, müssten dann mit Geldstrafen von bis zu 6 % ihres weltweiten Umsatzes rechnen.

Eine weitere Idee wäre, die Verbreitung von Deepfakes mit technologischen Mitteln einzuschränken, etwa durch effektive Erkennungsprogramme, welche manipulierte Videos identifizieren und diese sichtbar als solche markieren. Wie jedoch bereits in Kapitel B. IV. angesprochen, liefern sich die Entwickler von Deepfakes mit denen, die an deren Identifizierung forschen, ein unausgeglichenes Wettrennen, bei dem Erstere klar im Vorsprung sind. Der KI-Forscher und Unternehmer ***Hao Li*** geht davon aus, dass man zeitnah in der Lage sein wird Deepfakes zu erzeugen, die weder von Menschen noch von Maschinen erkannt werden können.[74] ***David Doermann***, Professor für Medienforensik an der ***Buffalo Universität***, bezeichnet die Entwicklung als „***cat-and-mouse game***".[75] Dies hat zur Folge, dass aktuelle Lösungen zur Identifikation von Deepfakes mittelfristig weniger zuverlässig sind. Da es sich bei Deepfakes um ein verhältnismäßig neues Phänomen handelt, gibt es viele Menschen, denen diese Technik noch völlig unbekannt ist. Eine Stärkung der Medienkompetenz – insbesondere mit Blick auf die kritische Reflektion von Online-Informationen und der Erkennung von manipulierten Videos – ist zwar keine rechtliche Lösung, kann jedoch dazu beitragen, den möglichen Schaden von Deepfakes zu begrenzen.

„Deepfakes haben ein beispielloses Potenzial zu manipulieren, da sie erfundene Informationen leicht mit Autoritätsquellen kombinieren können."

D. Ausblick

Deepfakes haben ein beispielloses Potenzial zu manipulieren, da sie erfundene Erzählungen und Informationen leicht mit Autoritätsquellen kombinieren können. Das macht die Enttarnung der Fehlinformation für den Rezipienten sehr schwer. Zugleich vermuten die wenigstens Menschen bei der Betrachtung eines Videos eine Falschmeldung. Videos und Audioaufnahmen halten die meisten Menschen noch immer für manipulationsfest und erkennen sie als unumstößliche Beweise an.

Neben dem großen Potenzial für Falschinformationen besteht zudem die Gefahr, dass die Menschen das Interesse an der Wahrheit verlieren. Menschen bevorzugen als Quelle der Gewissheit das, was sie selbst gesehen haben vor dem, was sie bloß von anderen gehört oder irgendwo gelesen haben.[76] Dieses Vertrauen kann

71 Vorschlag für eine Verordnung des Europäischen Parlamentes und des Rates zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz, Titel IV, KOM(2021)206 final, hier abrufbar (Stand: 01.08.2022).
72 The Strengthened Code of Practice on Disinformation 2022, hier abrufbar (Stand: 17.08.2022).
73 Mehr Informationen hier abrufbar (Stand: 18.08.2022); Duda, Der Digital Services Act – EU zwischen Innovation und Informationskrise, CTRL 2/22, 10 ff.
74 Laaff, Deepfakes: Hello, Adele – bist du's wirklich?, hier abrufbar (Stand: 18.08.2022).
75 Solsman, Deepfake Debunking Tool May Protect Presidential Candidates. For Now. Sometimes, hier abrufbar (Stand: 18.08.2022).
76 Rini, Deepfakes Are Coming. We Can No Longer Believe What We See., hier abrufbar (Stand: 17.10.2022).

durch Deepfakes stark erschüttert werden. Dies werden sich Menschen in Zukunft vermutlich zunutze machen. Es ist damit zu rechnen, dass bei Gerichtsverfahren immer häufiger der Einwand erhoben wird, die vorgelegte Video-, Bild- oder Tondatei sei ein Deepfake . Bei entsprechend substantiierten Vortrag wird regelmäßig die Bestellung eines Sachverständigen (mit entsprechender Expertise) erforderlich sein.[77]

Um den Gefahren, die mit der Verwendung von Deepfakes einhergehen, in Zukunft Einhalt gebieten zu können, ist es wichtig, dass dieses Thema mehr in den Fokus politischer Aufmerksamkeit gerückt wird. Die Verwendung von Deepfakes wird zunehmen; vor allem im politischen Kontext. Hierauf müssen sich auch die Bürger, die Plattformen und der Gesetzgeber einstellen. Die rasante Entwicklung von Deepfake-Software darf nicht unterschätzt werden. Der Gesetzgeber muss sich überlegen, wie er diese Technologie regulieren will. Bislang ist er überwiegend untätig geblieben, doch damit wird sich das Problem der missbräuchlichen Verwendung von Deepfakes sicher nicht in den Griff bekommen lassen. Begrüßenswert ist, dass das Thema Deepfakes bei der Frühjahrskonferenz der Justizminister im Juni 2021 auf der Tagesordnung stand. Es bleibt abzuwarten, ob der Vorschlag des bayerischen Justizministers zur Schaffung einer Regelung in einem neuen § 141 des StGB bei der Bundesjustizministerin Gehör finden wird.[78]

Eva ist wissenschaftliche Mitarbeiterin am Lorenz-von-Stein-Institut für Verwaltungswissenschaften an der CAU Kiel (gf.) und am Institut für Multimediale und Interaktive Systeme an der Universität zu Lübeck. Sie beschäftigt sich vor allem mit der Wirkung neuer Technologien auf das Recht und die Gesellschaft.

Anna ist wissenschaftliche Mitarbeiterin am Institut für Multimediale und Interaktive Systeme an der Universität zu Lübeck. In ihrer Forschung beschäftigt sie sich mit dem Einsatz von intelligenten Systemen für Richterinnen und Richter im Strafprozess.

77 Kuhlmann, Realität, Fiktion und das Problem, sie vor Gericht zu kriegen, hier abrufbar (01.09.22).
78 Pressemitteilung der Bayerischen Staatsregierung vom 16. Juni 2021, hier abrufbar (Stand: 01.08.2022).

Folge 25

Künstliche Intelligenz – was ist das eigentlich, Manuela Lenzen?

Folge 28

Regulierung & Innovation – wie lässt sich beides vereinbaren, Martin Ebers?

Folge 40

CTRL-KI als Rechtssubjekt, Transitional Justice & Legal Tech und das Internet der Dinge – Was ist die Cologne Technology Review & Law?

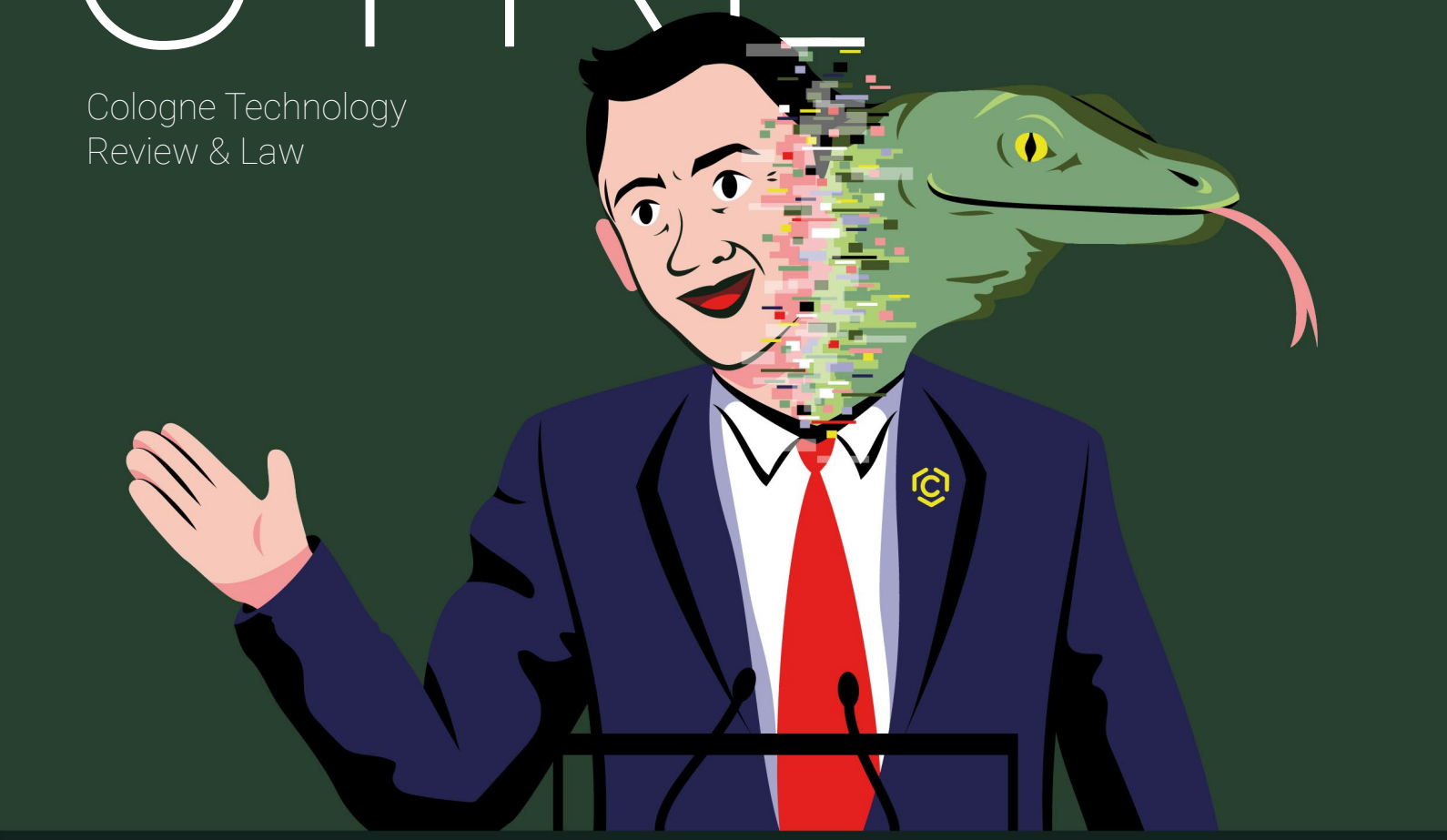
Zurück zum Inhaltsverzeichnis

CTRL

1/23

3. Jahrgang,1. Ausgabe
www.legaltechcologne.de/ctrl

Cologne Technology
Review & Law



Hier geht‘s zur ganzen Ausgabe!

Was das BGB mit Data Science und das StGB
mit Deepfakes zu tun hat und noch vieles mehr
in 12 spannenden Beiträgen!

LEGAL TECH LAB
COLOGNE

Cologne Technology
Review & Law