Büro für Technikfolgen-Abschätzung
beim Deutschen Bundestag

TAB

Rechtliche und gesellschaftliche Herausforderungen und Potenziale von Deepfakes

Octavia Madeira
Steffen Albrecht

unter Mitarbeit von
Jan-Henning Hansen

Dezember 2025
TA-Kompakt Nr. 3

Inhalt

Das Wichtigste in Kürze

Neun Fragen – neun Antworten

Worin unterscheiden sich Deepfakes von anderen Formen der Bildmanipulation bzw. KI-generierter Inhalte?

- KI-generierte Inhalte in Form von Text, Bild, Audio und/oder Video werden als synthetische Medien bezeichnet.
- Von Deepfakes spricht man, wenn solche Inhalte fälschlicherweise den Eindruck erwecken, dass sich bestimmte Sachverhalte zugetragen hätten oder Menschen bestimmte Aussagen treffen würden bzw. in Handlungen involviert wären.
- Daneben gibt es weitere Formen der Desinformation oder der Bildmanipulation, die ohne Unterstützung künstlicher Intelligenz (KI) auskommen. Diese werden zum Teil als Cheapfakes bezeichnet.

Welche Technologien werden für die Erstellung von Deepfakes genutzt?

- Für die Erstellung von Deepfakes werden vor allem vier KI-Grundmodelle genutzt: autoregressive Modelle, Autoencoder, Generative Adversarial Networks (GAN) und Diffusionsmodelle. Sie unterscheiden sich hinsichtlich der Menge an Trainingsdaten, die benötigt werden, der erreichbaren Qualität und dem benötigten Rechenaufwand sowie dem Grad an Spezialisierung für bestimmte Aufgaben.
- Die Entwicklung der Technologien ist dadurch gekennzeichnet, dass mit immer weniger Aufwand immer bessere Ergebnisse erzielt werden können.

Was ist über das Ausmaß der Nutzung von Deepfaketechnologien bekannt?

- Bereits heute lassen sich Deepfakes sowohl mit Blick auf die Soft- als auch die Hardwareanforderungen mit verhältnismäßig wenig Aufwand erstellen und veröffentlichen. Es wird davon ausgegangen, dass die Verbreitung von Deepfakes zukünftig noch zunehmen wird und die Echtheit von Onlineinhalten selbst in Livesituationen nicht mehr ohne Weiteres angenommen werden kann.
- Genaue Daten zum Ausmaß der Nutzung von Deepfaketechnologien liegen nicht vor, bei den wenigen dazu existierenden Studien ist nicht klar, wie belastbar sie sind. Insgesamt fehlt es an empirischer Forschung zum Ausmaß der Nutzung von Deepfaketechnologien.

Welchen Nutzen bieten Deepfaketechnologien in welchen Anwendungsbereichen?

- Deepfaketechnologien bieten Potenziale für innovative Anwendungen insbesondere in den Bereichen Kunst und Unterhaltung, Bildung und Wissenschaft sowie Werbung und Marketing. Beispiele sind die lippensynchrone Übersetzung von Filmen, der didaktische Einsatz in Museen sowie satirische Darstellungen.

- Allerdings gehen auch mit diesen Nutzungen indirekte Gefahren wie eine zunehmende Geringschätzung menschlicher Arbeit oder ein Verschwimmen der Grenzen von Wirklichkeit und Erfindung einher, wie sie für generative KI-Anwendungen allgemein beschrieben wurden.
- Deepfakes können außerdem für die Strafverfolgung genutzt werden, etwa indem mit ihrer Hilfe Fahndungsaufrufe gestaltet werden oder der Zugang zu kriminellen Netzwerken ermöglicht wird.

Welche gesellschaftlichen Risiken ergeben sich aus der Möglichkeit, Deepfakes zu erstellen und zu verbreiten?

- Deepfakes können bewusst mit dem Ziel produziert und verbreitet werden, die öffentliche Debatte oder den politischen Prozess zu beeinflussen. Bereits der Verdacht einer Fälschung kann ausreichen, die Glaubwürdigkeit von Nachrichten und das Vertrauen auf ein gemeinsames Verständnis der Wirklichkeit infrage zu stellen.
- Deepfakes werden zudem gezielt gegen Menschen angewendet, um deren Reputation für betrügerische Zwecke auszunutzen oder gar zu zerstören. Sie können Identitätsdiebstahl und die Überwindung von Identifizierungsverfahren erleichtern. Pornografische Deepfakes werden als Form der sexuellen Gewalt insbesondere gegen Frauen eingesetzt und können auf gesellschaftlicher Ebene eine Zurückdrängung von Frauen aus dem öffentlichen Raum bewirken.

Welche Chancen und Herausforderungen stellen Deepfakes speziell für das Rechtssystem dar?

- Deepfakes können für eine Vielzahl von Anwendungen eingesetzt werden, um Menschen, Unternehmen und Institutionen zu schaden. Eine explizite Regulierung von Deepfakes existiert im deutschen Recht bisher nicht. Auf zivilrechtlicher Ebene können der Bildnisschutz, das allgemeine Persönlichkeitsrecht sowie das Urheber- und Datenschutzrecht herangezogen werden, zudem sind verschiedene strafrechtliche Normen anwendbar. Allerdings ergeben sich Rechtslücken insbesondere bei sexualisierten Deepfakes von Erwachsenen, außerdem in Bezug auf den Schutz vor Diskriminierung sowie beim Einsatz von Deepfakes durch Strafverfolgungsbehörden.
- Ein besonderes Problem stellt die Rechtsdurchsetzung bei Deepfakes dar. Häufig werden Deepfakes im Ausland produziert und publiziert, zudem ist die Beweisführung aufgrund der fehlenden Transparenz erschwert und Betroffene berichten, dass seitens der Behörden häufig nur geringe Bereitschaft zu einer strafrechtlichen Verfolgung besteht.
- In Deutschland wurde 2024 ein Gesetzesantrag zur Ergänzung des Strafgesetzbuchs in den Bundesrat eingebracht. Vor dem Hintergrund der grundsätzlich sowohl nutzbringenden als auch schädlichen Anwendungsmöglichkeiten von Deepfaketechnologien sowie der Sensibilität von staatlichen Eingriffen in den öffentlichen Diskurs erscheint eine sorgfältige Abwägung von Regulierungsoptionen unabdingbar. Perspektivisch bieten Deepfakes für das Rechtssystem auch Chancen, etwa im Bereich der Strafverfolgung oder beim Schutz von Zeugen in Gerichtsverfahren. Allerdings stellt die Möglichkeit, Beweismittel mit einfachen Mitteln in überzeugender Qualität fälschen zu können, eine Herausforderung dar.

Wie haben die Rechtssysteme anderer Länder und der EU auf Deepfakes reagiert?

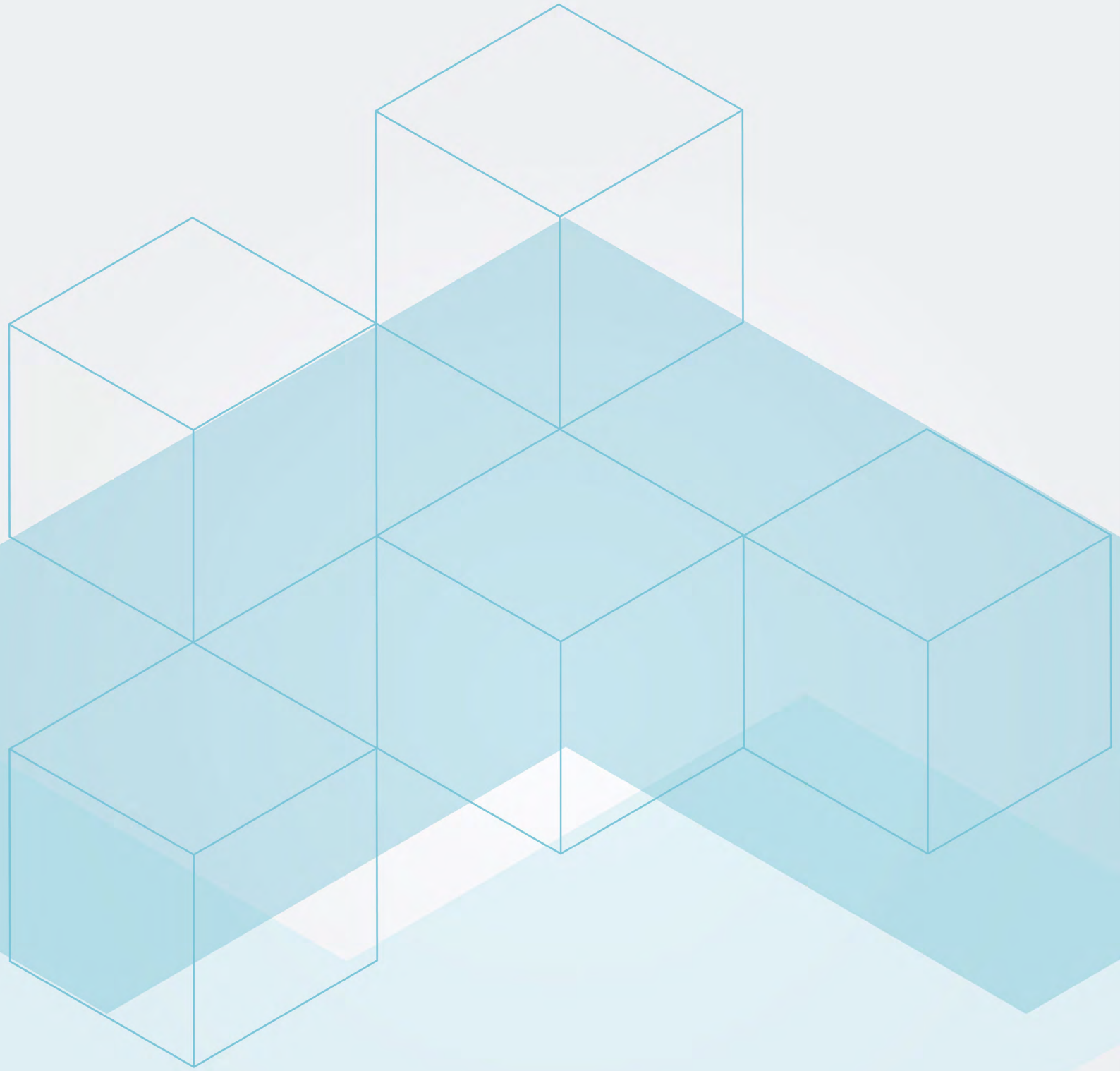
- In einigen Ländern wurden bereits Gesetze erlassen, um bestimmte Auswirkungen von Deepfakes, insbesondere im Bereich der nicht einvernehmlichen Pornografie, gezielt zu erfassen. In den USA besteht seit Mai 2025 eine landesweite Regelung zu pornografischen Deepfakes, in mehreren Bundesstaaten sind politische Deepfakes reguliert. In China werden Deepfakes wie allgemein synthetische Medieninhalte durch ein eigenes Gesetz erfasst, Kernpunkte dabei sind ein umfassendes Missbrauchsverbot sowie eine Klarnamen- und Kennzeichnungspflicht. Ausnahmen für die Meinungs- oder Kunstfreiheit sind nicht vorgesehen. Auf europäischer Ebene legt die KI-Verordnung den Entwicklern von Deepfaketechnologien Transparenzpflichten auf, der Digital Services Act reguliert die Verbreitung von Inhalten insbesondere über reichweitenstarke Plattformen.

Durch welche technischen Maßnahmen lässt sich einer missbräuchlichen Verwendung von Deepfaketechnologien begegnen?

- Die Entwicklung von Detektoren, mit denen sich Deepfakes automatisiert erkennen lassen, wird in mehreren deutschen Forschungsprojekten gefördert. Sie sind allerdings häufig auf bestimmte Deepfakearten spezialisiert. Ein grundsätzliches Problem von Detektoren besteht darin, dass sie auch für die Verbesserung von Deepfakes genutzt werden können und somit selbst dazu beitragen, die Hürden für die Erkennung zu erhöhen.
- Alternative Ansätze richten sich auf die Kennzeichnung und Verifizierung authentischer Inhalte und die Rückverfolgung des Ursprungs der Daten, zum Beispiel durch digitale Wasserzeichen oder kryptografische Signierung. Auch diese können jedoch von böswilligen Akteuren manipuliert bzw. umgangen werden.

Welche weiteren Maßnahmen werden seitens der Wissenschaft empfohlen?

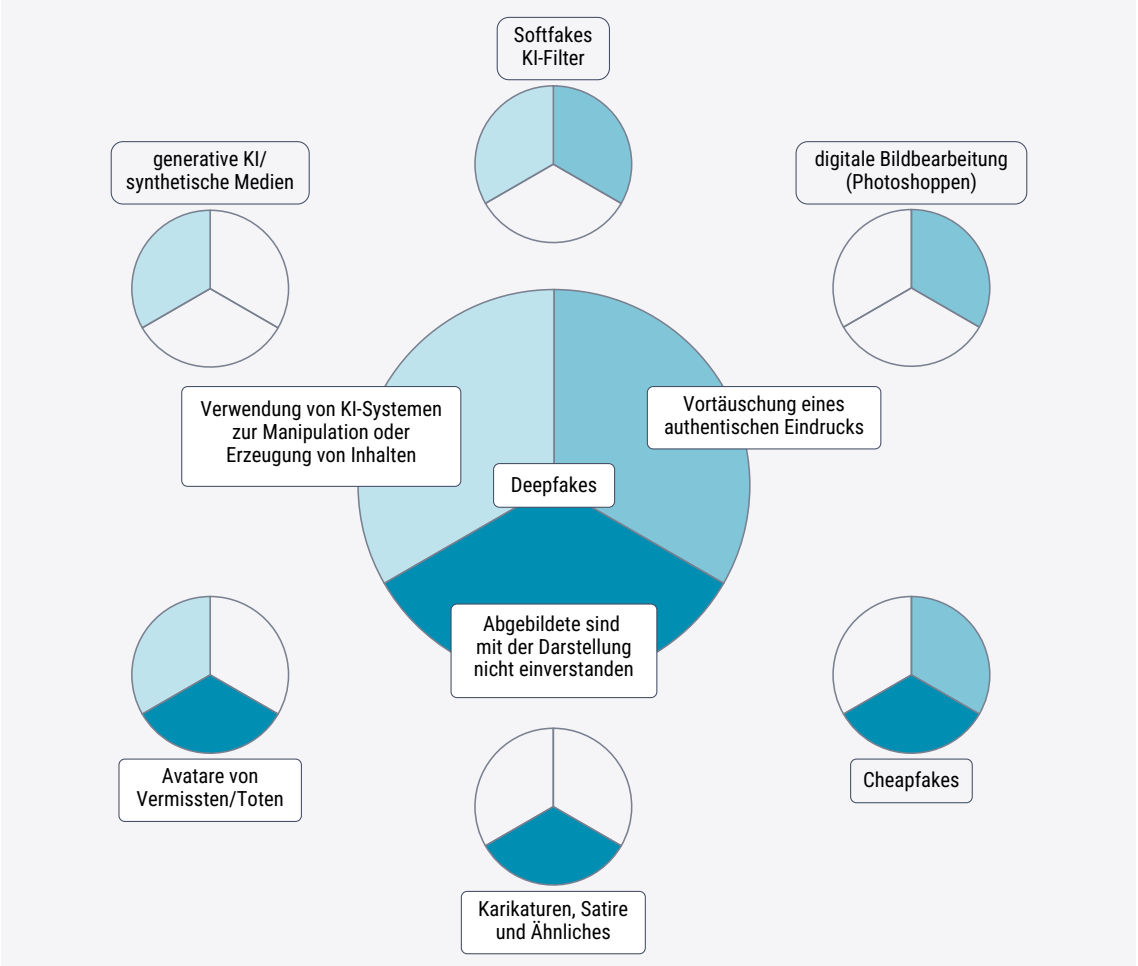
- Neben der Entwicklung von technischen Maßnahmen zur Authentifizierung von Inhalten und der Schließung von Regelungslücken wird vorgeschlagen, durch Aufklärung und transparente Berichterstattung über die möglichen Gefahren von Deepfakes zu informieren und die Wachsamkeit und Medienkompetenz zu erhöhen. Selbstverpflichtungen etwa von Parteien können zum Schutz der öffentlichen Debatte beitragen. Sowohl die Ausbreitung von Deepfakes als auch ihre Auswirkungen auf Individuen und die Gesellschaft und mögliche Reaktionen durch staatliche wie zivilgesellschaftliche Institutionen darauf sind noch wenig erforscht.



1 Einleitung

Mit dem Begriff Deepfake werden manipulierte oder synthetisch erzeugte Medien in Form von Ton, Bild oder Video bezeichnet, die in ihrer Qualität überzeugend authentisch wirken und somit den Eindruck erwecken, dass sich bestimmte Sachverhalte zugetragen hätten oder Menschen bestimmte Aussagen treffen würden bzw. in Handlungen involviert wären, obwohl dies nicht der Fall ist.[1] Grundlage zur Herstellung von Deepfakes sind Techniken der künstlichen Intelligenz (KI) wie Machine Learning und Deep Learning (Łabuz/Nehring 2024; Vaccari/Chadwick 2020). Auf KI-Basis erzeugte Medieninhalte werden auch allgemein als synthetische Medien bezeichnet.[2]

Abbildung 1.1 Abgrenzung von Deepfakes zu anderen Formen der Medienmanipulation



Eigene Darstellung

1 Artikel 3 Nr. 60 der KI-Verordnung definiert einen Deepfake als „einen durch KI erzeugten oder manipulierten Bild-, Ton- oder Videoinhalt, der wirklichen Personen, Gegenständen, Orten, Einrichtungen oder Ereignissen ähnelt und einer Person fälschlicherweise als echt oder wahrheitsgemäß erscheinen würde". Darüber hinaus gibt es eine Vielzahl unterschiedlicher Definitionen, die sich unter anderem hinsichtlich der zur Erzeugung genutzten (KI-)Technologien, der Medienart sowie des abgebildeten Inhalts (Personen bzw. auch Objekte) unterscheiden (Birrer/Just 2024; Whittaker et al. 2023)

2 Neben Ton, Bild und Video zählt hierzu auch künstlich generierter Text, der zum Beispiel von großen Sprachmodellen wie ChatGPT erzeugt werden kann. Im Folgenden bleibt jedoch der Blick auf synthetische visuelle, audiovisuelle oder auch nur auditive Medieninhalte im Sinne der Definition der KI-Verordnung gerichtet.

Die neutrale begriffliche Beschreibung als synthetische Medien steht durchaus in einem Widerspruch zur Bezeichnung Deepfakes, da durch den Bestandteil Fake eine Täuschungsidee oder -absicht oder gar eine schädigende Anwendung implizit nahegelegt wird. Die Entstehungsgeschichte von Deepfakes ist tatsächlich auf solche missbräuchlichen Ansätze zurückzuführen, die auch die hauptsächliche gesellschaftliche und politische Relevanz bedingen. Dennoch existieren auch Anwendungsmöglichkeiten, die eine authentisch wirkende Medienbearbeitung ohne Schadensabsicht vorsehen, zum Beispiel im Unterhaltungssektor. Da sich die Bezeichnung Deepfake in der Öffentlichkeit und der medialen Debatte etabliert hat und auch von der Forschung aufgenommen wurde (STOA 2021), wird im Folgenden der Begriff Deepfake für die technische Herstellung authentisch wirkender Medieninhalte verwendet, die sowohl missbräuchlich als auch nicht missbräuchlich angewendet werden können (Abbildung 1.1). Auch Softfakes, also mit KI generierte Bildmanipulationen, die ein/e Akteur/in von sich selbst erstellt bzw. erstellen lässt, um zum Beispiel in einem besseren Licht dazustehen (Chowdhury 2024), fallen in diese Kategorie, nicht aber Cheapfakes, also Manipulationen mit Schadensabsicht, die sich aber keiner technisch avancierten Mittel bedienen (wohl aber täuschend echt wirken können).

Zielsetzung und Vorgehensweise

Auf Initiative des Ausschusses für Digitales wurde das Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag (TAB) vom Ausschuss für Bildung, Forschung und Technikfolgenabschätzung mit einer Kurzstudie beauftragt, um einen konzentrierten Überblick über den Stand der technologischen Entwicklung, typischer Anwendungsmöglichkeiten und -beispiele sowie der gesellschaftlichen und politischen, insbesondere der rechtlichen Herausforderungen durch Deepfakes zu geben. Um die Rechtslage fundiert darlegen und diskutieren zu können, wurde ein rechtswissenschaftliches Gutachten ausgeschrieben und vergeben:

- Rechtliche Aspekte von Deepfakes. Prof. Dr. Christoph Busch, Prof. Dr. Jan Oster, Osnabrück/Ober-Olm

Aufgrund personeller Wechsel in der Projektbearbeitung konnte das Gutachten, das mit Stand Juni 2023 vorgelegt wurde, erst 2024 im Zuge der Berichtserstellung durch das TAB ausgewertet werden. Um eine größtmögliche Aktualität zu gewährleisten, wurde Ende 2024 ein Kommentargutachten zu dem Berichtsentwurf an Johannes Härtlein, wissenschaftlicher Mitarbeiter am Lehrstuhl für Strafrecht, Strafprozessrecht, Informationsrecht und Rechtsinformatik an der Julius-Maximilians-Universität Würzburg, vergeben. Den drei Experten sei vonseiten des TAB sehr für ihre engagierte Kooperation gedankt. Darüber hinaus geht ein besonderer Dank an Dr. Alma Kolleck, die das Projekt bis zum Sommer 2023 betreute, sowie an Dr. Arnold Sauter, der seitdem die Projektorganisation mit übernommen hat, und nicht zuletzt an Brigitta-Ulrike Goelsdorf für die Durchsicht des Manuskripts und die Gestaltung des Endlayouts. Das TA-Kompakt gibt den juristischen Sachstand zum Zeitpunkt Ende 2024 wieder und wurde im August 2025 aktualisiert, um zentrale Entwicklungen im Themenfeld zu berücksichtigen.

Aufbau des Berichts

In Kapitel 2 werden Methoden zur Erzeugung und zum Nachweis von Deepfakes beschrieben sowie ein Überblick zur Verbreitung und zum Bedeutungszuwachs gegeben, bevor in Kapitel 3 Anwendungsmöglichkeiten und -beispiele von Deepfaketechnologien mit den aus ihnen erwachsenden wirtschaftlichen und gesellschaftlichen Potenzialen oder aber Risiken vorgestellt werden. In Kapitel 4 werden zunächst überblicksartig rechtliche Regelungen in den USA und in China vorgestellt, bevor dann die einschlägigen Regelungen auf EU-Ebene, also vor allem die Verordnung (EU) 2024/1689[3] (KI-Verordnung) und die Verordnung (EU) 2022/2065[4] (Digital Services Act – DSA), sowie die Rechtslage in Deutschland in zivil- und strafrechtlicher Hinsicht vertieft diskutiert werden. Im abschließenden Kapitel 5 werden dann Regelungslücken und Herausforderungen der Rechtsdurchsetzung beschrieben und Handlungsoptionen im Bereich der Technikentwicklung sowie in Bezug auf das Rechtssystem und die Sicherung demokratischer Prozesse benannt.

3 Verordnung (EU) 2024/1689 vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG) Nr. 300/2008, (EU) Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828 (Verordnung über künstliche Intelligenz)

4 Verordnung (EU) 2022/2065 vom 19. Oktober 2022 über einen Binnenmarkt für digitale Dienste und zur Änderung der Richtlinie 2000/31/EG (Gesetz über digitale Dienste)

2 Technische Grundlagen von Deepfakes

2.1 Methoden zur Erzeugung von Deepfakes

Die allgemeine technische Grundlage zur Erzeugung von Deepfakes bilden *Deep Neural Networks*, wobei unterschiedliche KI-Grundmodelle zum Einsatz kommen: *autoregressive Modelle*, *Autoencoder*, *Generative Adversarial Networks* (GAN/erzeugendes gegnerisches Netzwerk) und *Diffusionsmodelle* (Seow et al. 2022).

Ein *autoregressives Modell* erstellt anhand einer statistischen Auswertung von Trainingsdaten synthetische Dateien. Solche Modelle liegen Transformern zugrunde, die als Grundlage von ChatGPT besondere Bekanntheit erlangt haben. Während bei der Generierung von Texten Wortfolgen mithilfe bedingter Wahrscheinlichkeiten generiert werden, können autoregressive Modelle auch zur Bildgenerierung genutzt werden. Dabei werden Bilder auf der Ebene einzelner Pixel als lineare Informationsketten interpretiert und neue Pixel in Abhängigkeit vorangegangener Werte erzeugt. Auf diesem Wege können Bilder in hoher Qualität produziert werden, wofür jedoch wegen der Erzeugungsweise auf Pixelebene sehr zeitintensive Rechenleistung benötigt wird (Seow et al. 2022; Yang et al. 2023).

Autoencoder werden häufig genutzt, um die Mimik zu verändern oder Gesichter auszutauschen. Hierbei wird mit zwei neuronalen Netzen gearbeitet. Der *Encoder* extrahiert anhand des gegebenen Bildmaterials Gesichtsausdrücke, zum Beispiel die relationale Veränderung von Mund und Augen im Ruhezustand zu einem Lächeln, indem eine mathematische Repräsentation erstellt wird. Diese Repräsentation wird wiederum vom *Decoder* genutzt, um in Bildmaterial Gesichter nahtlos einzufügen oder den Gesichtsausdruck nach gewünschter Absicht zu bearbeiten. Je vielfältiger die Mimik und Aufnahmeperspektiven des Gesichts einer Zielperson im Trainingsmaterial enthalten sind, desto realistischer wird die Umsetzung durch den Autoencoder ausfallen (Seow et al. 2022; STOA 2021).

Das oftmals genutzte Grundmodell zur Herstellung von Deepfakes ist das *GAN* (Karaboga et al. 2024, S. 87). Im Vergleich zu Autoencodern und autoregressiven Modellen erreichen die synthetischen Erzeugnisse eines GAN eine höhere Qualität, die sich aber zugleich in einer längeren Trainingszeit und einer umfangreichen Trainingsdatensammlung niederschlägt (Seow et al. 2022). Die Idee eines GAN zur Entwicklung von Deepfakes liegt in einem Wechselspiel zwischen zwei neuronalen Netzwerken. Ein neuronales Netzwerk, der Generator, erzeugt eine synthetische (Bild-)Datei, die an das andere Netzwerk, den Diskriminator, übergeben wird. Der Diskriminator überprüft auf Basis seines Trainingsmaterials, ob es sich hierbei um eine Fälschung handelt oder nicht. Die Einschätzung des Diskriminators wird wiederum an den Generator zurückgemeldet, sodass ein Prozess in Gang gesetzt wird, der zu einer kontinuierlichen Verbesserung der synthetisch erzeugten Dateien führt (Pawelec/Bieß 2021, S. 37). Grundsätzlich können GAN genutzt werden, um beliebige audiovisuelle Medieninhalte zu erzeugen. Für bestimmte Anwendungsaufgaben, zum Beispiel für die Erzeugung von Portraits, werden GAN-Systeme in der Regel optimiert, sodass mittlerweile eine Vielzahl spezifischer Entwicklungstools vorzufinden ist (Seow et al. 2022; STOA 2021).

Bei *Diffusionsmodellen* verweist der Name auf den Trainingsprozess des neuronalen Netzwerks, bei dem Bildern (oder anderen Daten) schrittweise Rauschen hinzugefügt wird, bis schließlich

nichts mehr zu erkennen ist. Die erlernten Schritte kann das Modell anschließend wieder rückwärts vollziehen, sodass aus einem zufälligen Rauschen – je nach Umfang und Qualität des Trainingsmaterials – prinzipiell beliebige Bilder generiert werden können, und dies in fotorealistischer Qualität (Croitoru et al. 2023, 2024; Merkert 2022).[5] Diffusionsmodelle bilden die Grundlage vieler KI-Bildgeneratoren. Solche Modelle können mittels Feinabstimmung auf die Erzeugung von Bildern bestimmter Personen, zum Beispiel Prominenter, angepasst und so zur Erzeugung von Deepfakes dieser Personen genutzt werden (Hawkins et al. 2025).

Große Aufmerksamkeit liegt bei Deepfakes auf Videomaterial. Dazu muss Bild- und Tonmaterial erzeugt oder verändert werden, wobei bei ersterem häufig die visuelle Erscheinung von Personen bzw. Gesichtern betroffen ist, bei letzterem geht es vor allem um synthetisch erzeugte Stimmen. Für die Manipulation von Gesichtern und Körperbewegungen lassen sich vier Typen von Deepfakeverfahren unterscheiden (Seow et al. 2022):

- *Entire Face Synthesis*: Es wird ein vollständig synthetisches Gesicht anhand allgemeiner menschlicher Merkmale generiert, das kein Abbild einer realen Person darstellt.
- *Reenactment*: Es werden die Mimik und/oder die Bewegungen einer Person auf eine andere übertragen, das heißt, eine schauspielende Person vollzieht die Gesichtsausdrücke und Bewegungen, die auf die (virtuelle) Zielperson übertragen werden. Neben einer rein synthetischen Erzeugung lässt sich auf diesem Wege auch die Lippensynchronisation für Videos realistisch generieren, nachdem etwa die Tonspur geändert wurde.
- *Facial Attributes Manipulation*: Nur bestimmte Gesichtsmerkmale werden modifiziert, zum Beispiel Augenfarbe, Frisur, Alter etc.
- *Face Swap*: Das Gesicht einer Person wird mit dem Gesicht einer anderen Person ausgetauscht, ohne eine Änderung in Mimik oder anderen Merkmalen.

Die Erzeugung von Stimmen nach dem Vorbild bestimmter Personen lässt sich ebenso wie die Erzeugung von Bild- und Videomaterial mithilfe eines GAN bewerkstelligen. Zusammen mit den technischen Möglichkeiten der Lippensynchronisation lassen sich beliebige Worte in überzeugender Weise in den Stimmen beliebiger Personen generieren. Insbesondere können die Stimmen von bekannten Personen glaubhaft imitiert werden, sodass die Überzeugungskraft von Videomaterial deutlich erhöht wird (STOA 2021). So werden zum Teil auch manipulierte Stimmen zusammen mit originalem Videomaterial, bei dem die Mimik zum Gesagten passt, verwendet, um gefälschte Botschaften zu vermitteln. Für die Erzeugung und Manipulation von Videos stehen nicht zuletzt zunehmend leistungsfähigere Videogeneratoren (zum Beispiel Gen-3 Alpha von Runway, Veo von Google Deepmind, Sora von OpenAI) bereit, die wie die beliebten Bildgeneratoren auf Diffusionsmodellen basieren und anhand von textlichen Beschreibungen oder vorgegebenen Bildern Videomaterial produzieren können (Karaboga et al. 2024, S. 93 f.).

Für die Qualität der Deepfakeerzeugung ist die Menge an verwendetem Trainingsdatenmaterial entscheidend. Da die Produktion in hoher Qualität mithilfe eines GAN eine umfangreiche Sammlung erfordert, beziehen sich viele bekannte Beispiele für Deepfakes auf Personen, die eine entsprechend große mediale Präsenz haben und für die daher eine große Menge an Bilddaten zugänglich ist, zum Beispiel Spitzenpolitiker wie Joe Biden oder Olaf Scholz. Allerdings geht die Entwicklung in

5 Die meisten bekannten KI-Bildgeneratoren wie Dall-E, Midjourney, Stable Diffusion oder Flux.1 beruhen Expert/innen zufolge auf Diffusionsmodellen (allerdings machen die Unternehmen dies in der Regel nicht transparent). Bekannte Fotodeepfakes wie das Bild von Papst Franziskus in einer Daunenjacke oder Bilder einer vermeintlichen Verhaftung Donald Trumps durch Polizeibeamte wurden mit Midjourney erstellt.

die Richtung, dass für die Imitation von Stimmen bereits wenige Minuten Datenmaterial ausreichen (Pawelec/Bieß 2021, S. 40) und auch der Forschungstrend für den Bild- und Videobereich zielt auf eine möglichst geringe Trainingsdatengrundlage ab (Seow et al. 2022). Die Personengruppe, die Ziel eines Deepfakes werden kann, weitet sich mit diesen technologischen Fortschritten enorm aus.

2.2 Verbreitung und Bedeutungszuwachs von Deepfakes

Bild- und Videomanipulationen sind grundsätzlich kein neuartiges Phänomen und inzwischen eine Alltagserfahrung, zum Beispiel als Filter in der Handykamera oder der Hintergrundgestaltung in Videokonferenzen. Die Verfälschung von Fotos zu Manipulationszwecken fand bereits mit nicht digitalen Entwicklungstechniken statt, zum Beispiel ließen Stalin, Hitler und Castro Fotos bearbeiten, um eine für sie vorteilhafte Darstellung von Ereignissen bzw. der Geschichte scheinbar glaubhaft belegen zu können (Farid/Schindler 2020). Mit dem Aufkommen der Digitaltechnik ergaben sich einfachere und leistungsstärkere Bildbearbeitungsmöglichkeiten, die für Verfälschungen von Fotos seit der Jahrtausendwende zunehmend zum Einsatz kamen. So wurden zum Beispiel auf Fotos von Kriegsschauplätzen zusätzlicher Rauch eingefügt, Politiker/innen auf Kundgebungen mit umstrittenen Aktivist/innen gezeigt, um sie zu diskreditieren, oder scheinbar der erfolgreiche Test militärischer Raketen gezeigt (Farid/Schindler 2020).

Deepfakes lassen sich zweifelsfrei dieser technischen Entwicklungsreihe und den genannten Anwendungsbereichen zuordnen. Allerdings muss für die Einschätzung der Wirkungsweise von Deepfakes, das heißt, wie sie sich im heutigen gesellschaftlichen Gefüge entfalten können, insbesondere die Entwicklung der jüngeren Vergangenheit betrachtet werden (STOA 2021). Seit dem Aufkommen des Internets durchlief die Medienlandschaft einen Wandlungsprozess, der den gesellschaftlichen Informationsfluss grundlegend umformte. Neben den etablierten Nachrichten- und Medienorganisationen entstanden weitere Angebote, die durch ihre Themensetzung und Informationsauswahl Einfluss auf den öffentlichen Meinungsbildungsprozess ausüben, zum Beispiel Content Creator oder Social-Media-Plattformen. Entsprechend hat sich die Bandbreite der genutzten Informationsquellen innerhalb der Gesellschaft erweitert (TAB 2022a, S. 75 f.). Insbesondere scheint es nicht einfach, eine verbindliche ethische Richtlinie oder einen gemeinsamen Verhaltenskodex unter Medienschaffenden bzw. Distributoren auf den unterschiedlichen privatwirtschaftlich organisierten Plattformen festzulegen, sodass in Konsequenz die Qualität der Informationen sowie die Verbreitung von Falsch- und Desinformationen durch die sozialen Medien eine besondere Herausforderung darstellt (STOA 2021).

Medieninhalte erreichen über das Internet potenziell eine enorme Reichweite und sind in der Regel eng an Plattformen, soziale Medien und Messengerdienste gebunden. Visuelle Kommunikationsformen, insbesondere Videoformate, haben dabei in den letzten Jahren immer mehr an Bedeutung gewonnen als besonders effektive Art, Eindrücke zu vermitteln (STOA 2021). Die besondere Suggestivkraft audiovisueller Medien hat das Bundesverfassungsgericht (BVerfG) in seinen Rundfunkurteilen aus dem zeitgleichen Zusammenspiel von Bild und Ton und dem „dadurch vermittelten Anschein der Authentizität und des Miterlebens“ (BVerfGE 97, 228 [256]) abgeleitet. Sie wird sogar dann wirksam, wenn die Möglichkeit der Manipulation der Bild- und Tonfolgen bekannt bzw. bewusst ist (DLM 2003, S. 7). Inwieweit allerdings auch für Deepfakes eine solche Suggestivkraft angenommen werden kann, ist zumindest für Deepfakes in der möglichen Ausprägung von Desinformation fraglich (Birrer/Just 2024, S. 7 f.).

Die Verbreitung von Inhalten wird auf den sozialen Medien zu großen Teilen über algorithmische Mechanismen in die Wege geleitet (TAB 2022a). Neben der algorithmischen Verstärkung und Sichtbarmachung bestimmter Inhalte ist auch das Verhalten der Nutzer/innen wichtig, da sie häufiger dazu neigen, negative oder kontroverse Inhalte zu teilen und so zu deren stärkeren Verbreitung beitragen. Zudem werden ihnen von den Plattformalgorithmen vor allem solche Inhalte angezeigt, die ihren Präferenzen entsprechen, sodass unabhängig vom Wahrheitsgehalt der Informationen eine Bestätigung bzw. Verstärkung der eigenen Meinungen gefördert wird (Pawelec/Bieß 2021, S. 77 f.).

Neben der Verbreitung von Deepfakes selbst sind mittlerweile auch Entwicklungstools oder Angebote entsprechender Dienstleistungen im Internet leicht erreichbar (Sensity B. V. 2024). Sowohl Forschende als auch enthusiastische Entwickler/innen stellen ihre Tools häufig als Open-Source-Software zur Verfügung und gestalten sie so anwendungsfreundlich, dass auch Nutzer/innen mit geringem technischen Wissen Deepfakes erzeugen können (Seow et al. 2022; Westerlund 2019). Des Weiteren sinken die Anforderungen an die benötigte Hardware stetig, sodass für die Herstellung von Deepfakes prinzipiell keine besonders anspruchsvolle technische Ausstattung nötig ist. Die Erzeugung und Veröffentlichung von Deepfakes, sowohl mit Blick auf die Soft- als auch die Hardwareanforderungen, sind daher – je nach angestrebter Qualität und Komplexität des Ergebnisses – mit verhältnismäßig wenig Aufwand verbunden.

Die technische Entwicklung von Deepfakes lässt inzwischen auch eine Echtzeiterzeugung zu, das heißt, Livestreams oder Videokonferenzen können in Bild und Ton so manipuliert werden, dass das vermeintliche Gegenüber zwar überzeugend authentisch wirkt, jedoch ein nahezu beliebiges Aussehen haben kann (Wolf 2024). Die Annahme eines Echtheitsnachweises aufgrund einer authentischen *Livesituation*, wie beispielsweise ein Telefonat in geschäftlichen Zusammenhängen, wird daher zu überdenken sein (Seow et al. 2022; STOA 2021).

2.3 Detektion von Deepfakes

Deepfakes tragen dazu bei, dass die Unterscheidbarkeit zwischen echten und künstlich generierten Medieninhalten für die menschliche Wahrnehmung abnimmt. Ihre technische Verfeinerung und die Weiterentwicklung von GAN-Systemen haben in den vergangenen Jahren zu enormen Qualitätssteigerungen geführt, sodass sich für visuelle wie auch auditive Inhalte nicht mehr ohne Weiteres sagen lässt, ob sie authentisch oder generiert bzw. manipuliert sind (Mai et al. 2023; Westerlund 2019). Einerseits ist dies eine bemerkenswerte technische Leistung, andererseits stellt genau diese technische Möglichkeit die zentrale gesellschaftliche Problematik von Deepfakes dar: Wie kann die Glaubwürdigkeit von Medieninhalten noch beurteilt werden? Welche sicheren Möglichkeiten gibt es, zu überprüfen, ob es sich bei einem Inhalt um einen Deepfake oder eine Abbildung von Tatsachen handelt?

Im Falle einer geringen technischen Qualität manipulierter oder generierter Inhalte lassen charakteristische Merkmale wie verschwommene Bildbereiche, eine unstimmige Sprechweise bzw. nicht zueinander passende Bewegungen eine menschliche Beurteilung zu (BSI o. J.). Bereits in diesem Fall stellt jedoch die große Menge an verfügbaren Medieninhalten eine Herausforderung dar (STOA 2021). Daher wird aufgrund der Quantität und auch der zunehmend hohen Qualität von Deepfakes für die Detektion auf automatisierte Erkennungsverfahren gesetzt, bei denen zunehmend KI-Techniken zum Einsatz kommen (Pawelec/Bieß 2021, S. 136).

Die Ansatzpunkte für technische Detektoren können unterschiedlicher Art sein und beziehen sich zum Beispiel auf subtile Hautfarbenveränderungen aufgrund der Pulsfrequenz, Inkonsistenzen in der Datenstruktur oder die Pixelkorrelation in Bildern (Pawelec/Bieß 2021, S. 136; Seow et al. 2022). Die Erfolgsquoten einiger Detektoren belegen eine relativ zuverlässige Einschätzung (teilweise mit 90 % oder mehr), ob es sich bei vorliegendem Datenmaterial um einen Deepfake handelt oder nicht. Dennoch sind solche Zuverlässigkeitsaussagen in ihrem Kontext zu verstehen. So beziehen sich Detektoren zum einen auf bestimmte Arten der Deepfakeerzeugung, zum anderen wird die Erfolgsquote anhand eines bestimmten Trainingsmaterials bemessen. In der Regel erweisen sich die Detektoren in ihrer Leistung als sehr eingeschränkt, wenn sich die Datengrundlage ändert (zum Beispiel andere Bilder oder eine andere Datenkompression) oder die zugrunde liegenden KI-Methoden nicht übereinstimmen, sodass es derzeit nicht möglich ist, beliebig erstellte Medieninhalte zuverlässig durch einen Detektor als Deepfake einzustufen (Karaboga et al. 2024, S. 113 f.; Pawelec/Bieß 2021, S. 14 f.; Seow et al. 2022; STOA 2021). Aus diesem Grund wird die Entwicklung generellerer Detektoren, das heißt unabhängig von speziellen Deepfaketechnologien, von großer Hoffnung begleitet, doch sie befindet sich erst noch in ihren Anfängen und wird vorerst als ein ungelöstes Problem betrachtet (Pawelec/Bieß 2021, S. 151; Seow et al. 2022). Dennoch spielen spezialisierte Detektoren eine wichtige Rolle. Sie dienen als Hürde, die dazu beiträgt, die Verbreitung von Deepfakes zu erschweren (Pawelec/Bieß 2021, S. 150 f.).

Eine grundlegende Problematik der Detektion besteht im – häufig so beschriebenen – „Katz-und-Maus-Spiel“. Neben der eigentlichen Aufgabe einer zuverlässigen Erkennung werden Detektionsalgorithmen dazu genutzt, um Deepfakes weiterzuentwickeln. Angestrebt wird damit nicht nur eine täuschend echte Wirkung für die menschliche Wahrnehmung, sondern auch eine erschwerte Detektion. Ist ein Detektionsalgorithmus bekannt, so kann er in einem GAN für den Diskriminator genutzt werden, also das neuronale Netzwerk, als dessen Gegenspieler der Generator trainiert wird. Detektionsalgorithmen können auf diesem Weg zur Verbesserung von Deepfakes beitragen. Detektoren erweisen sich daher in ihrer Effektivität als zeitlich begrenzt und laufen den Weiterentwicklungen von Deepfakes stets hinterher (Pawelec/Bieß 2021, S. 147 f.; Seow et al. 2022; STOA 2021). Zentral ist hierbei der offene Zugang zu den Detektoren, sei es open source auf Plattformen oder in Forschungszusammenhängen, der aus ethischer Sicht und Transparenzgründen gefordert wird, aber in diesem Kontext einen abträglichen Effekt mit sich bringt (Pawelec/Bieß 2021, S. 148).

3 Anwendungsmöglichkeiten und -beispiele von Deepfaketechnologien

Während der Begriff Deepfake definitionsgemäß immer eine Täuschung impliziert, können Deepfaketechnologien grundsätzlich sowohl für böswillige als auch für wohlmeinende Zwecke angewendet werden. Sie können zudem mit oder ohne transparente Kennzeichnung sowie mit oder ohne Einverständnis der dargestellten Personen genutzt werden. Dabei ist zwischen der Anwendung und ihren Auswirkungen zu unterscheiden, denn neben den beabsichtigten Effekten können sich weitere, unbeabsichtigte Wirkungen ergeben. Die folgenden Beispiele kommen insbesondere aus den Anwendungsbereichen Kunst/Unterhaltung, Bildung/Forschung/Wissenschaft, Werbung/Marketing, sowie Politik/öffentliche Meinungsbildung, Polizeiarbeit/Strafverfolgung. Diese sollen einen Eindruck von der Bandbreite möglicher sowie bereits erfolgter Anwendungen von Deepfaketechnologien geben. Weiterhin wird auch die Rolle von Deepfakes im Zusammenhang mit falschen Identitäten und sexualisierter Gewalt dargestellt.

3.1 Kunst, Unterhaltung

Deepfakes sind ein neues Werkzeug, um der eigenen Kreativität Ausdruck zu verleihen und Vorstellungswelten zu visualisieren. In der Film- und Videospielbranche wird ein großes Potenzial für Deepfakes ausgemacht (Karaboga et al. 2024, S. 300 ff.; Lees 2024; TAB 2024). Sie können zur Erstellung von 3D-Modellen verwendet werden, ermöglichen eine nachträgliche Bearbeitung von Filmmaterial, zum Beispiel aufgrund einer Skriptänderung oder von Versprechern, oder erlauben eine realistische Synchronisation in andere Sprachen, die sogar die Originalstimme der Schauspieler/innen verwendet. Des Weiteren könnten Schauspieler/innen durch die Bereitstellung eigener digitaler Abbilder mit einem gealterten oder verjüngten Erscheinungsbild auftreten oder müssten für die Aufnahme von Szenen nicht vor Ort sein. Denkbar wäre auch die Wiederbelebung verstorbener Schauspieler/innen oder von Popikonen, die erneut Konzerte geben (STOA 2021).

Kasten 3.1 Deepfake vs. klassische Computergrafik im Film (2020)

In der an die Star-Wars-Geschichte anknüpfenden Serie „The Mandalorian“ ließ das zu Disney gehörende Filmstudio Lucasfilm die Figur des Luke Skywalker als jungen Menschen auftreten. Das Besondere war, dass Skywalker wie in den früheren Star-Wars-Filmen vom Schauspieler Mark Hamill gespielt wurde, dessen Erscheinung mithilfe von Computergrafik verjüngt wurde (Abbildung 3.1 linke Seite). Ein Youtuber zeigte, dass ein ähnliches (laut Kommentatoren sogar besseres) Ergebnis mit vergleichsweise geringem Aufwand durch Deepfaketechnologie erreicht werden kann (Abbildung 3.1 rechte Seite). Dies wirft allerdings ethische Fragen auf, etwa ob auch Filmszenen mit bereits verstorbenen Schauspieler/innen realisiert werden dürfen (Kain 2020).

Abbildung 3.1 **Gegenüberstellung von Kinoversion (mit Visual Effects) und Deepfakeversion von Luke Skywalker in „The Mandalorian“**

Quelle: Shamook (Screenshot www.youtube.com/watch?v=wrHXA2cSpNU; 12.9.2025)

Kasten 3.2 **Weihnachtsansprache von Queen Elizabeth II (2020)**

Während die britische Königin traditionell ihre Weihnachtsansprache über die Sender BBC und ITV ausstrahlen lässt, sendete Channel 4 eine als Deepfake gekennzeichnete Weihnachtsansprache, die aufgrund der geringen Qualität und des unrealistischen Inhalts leicht als Parodie zu erkennen war und zudem die Gefahren von Deepfakes thematisierte (STOA 2021, S. 23). Das Video (Abbildung 3.2) lässt sich anhand von Merkmalen wie der Belichtung bereits als künstlich bearbeitet erkennen, am Ende des Clips wird zudem die Herstellung explizit thematisiert und die Schauspielerin, deren Gesichtszüge und Lippenbewegungen für die Deepfakeerzeugung genutzt wurden, gezeigt (Abbildung 3.3).

Abbildung 3.2 Weihnachtsansprache 2020 der Queen auf Channel 4 (Deepfake)

Quelle: Channel 4 Entertainment (Screenshot www.youtube.com/watch?v=IvY-Abd2FfM; 12.9.2025)

Abbildung 3.3 Making-of der Weihnachtsansprache 2020 der Queen auf Channel 4

Quelle: Channel 4 Entertainment (Screenshot www.youtube.com/watch?v=IvY-Abd2FfM; 12.9.2025)

In Dokumentarfilmen bzw. im Journalismus können sie zum Schutz der Identität von Quellen eingesetzt werden. Viele Anwendungen zu Unterhaltungszwecken reduzieren die Produktionskosten erheblich, sodass auch kleinere Entwicklungsstudios in dieser Branche Fuß fassen können. Für bestimmte Berufsgruppen in der Unterhaltungsindustrie werden Deepfakes jedoch eine bedrohliche Entwicklung sein, zum Beispiel für Synchronsprecher/innen oder Maskenbildner/innen (Pawelec/Bieß 2021, S. 62).

Neben einer künstlerischen Anwendung bieten Deepfakes auch für Satire und Parodie neue Möglichkeiten (STOA 2021). Im Rahmen von Satire wird Bezug auf politische Vorgänge, Politiker/innen oder allgemein Personen des öffentlichen Lebens genommen, um durch humoristische Kritik Debatten anzustoßen. In diesem Zusammenhang handelt es sich um offensichtlich oder zumindest nachweislich falsche Informationen, die nicht irreführen oder schädigen, sondern der Unterhaltung und Kritik dienen sollen.

Außerhalb des entsprechenden Rahmens, durch den die künstlerische Absicht deutlich wird, können satirische Deepfakes allerdings zu nicht intendierten Folgen führen. Das Erkennen und das Verstehen von Satire erfordern oftmals eine subtile sprachliche bzw. kulturelle Kennerschaft, die fundiertes (wenn auch nicht unbedingt greifbares) Wissen der verwendeten Codes voraussetzt. Fehlt der Kontext, zum Beispiel beim Teilen in den sozialen Medien, und sind die verwendeten satirischen Stilmittel nicht dechiffrierbar, so können die Darstellungen entgegen der Absicht als wahr aufgefasst werden (Pawelec/Bieß 2021, S. 51 f.). Die satirische Verwendung von Deepfakes lässt sich häufig nicht klar von einem dezidiert politisch motivierten Einsatz abgrenzen, die Einordnung als Satire kann je nach Zeitpunkt und Rezipientengruppe unterschiedlich ausfallen.

Kasten 3.3 Kampagne des Zentrums für Politische Schönheit (2023)

Im Rahmen einer Kampagne für ein Verbot der AfD nutzten Aktionskünstler des Zentrums für Politische Schönheit unter anderem ein Deepfakevideo, in dem Olaf Scholz scheinbar ein Verbot der Partei fordert. Die Verbreitung des Videos, die auch über Social-Media-Plattformen erfolgte, wurde auf Antrag der Bundesregierung gerichtlich untersagt, unter anderem weil es nicht in ausreichendem Maß als Satire zu erkennen war (LTO 2024). Das Gericht begründete seine Entscheidung durch eine Verletzung des Namensrechts (§ 12 BGB[6]) und die Gefahr einer Zuordnungsverwirrung.[7] Die Künstlergruppe kündigte an, Rechtsmittel gegen die Entscheidung einzulegen.

6 Bürgerliches Gesetzbuch in der Fassung der Bekanntmachung vom 2. Januar 2002, zuletzt am 7. April 2025 geändert
7 LG Berlin II, Beschluss vom 13. Februar 2024, Az. 15 O 579/23

3.2 Bildung, Forschung, Wissenschaft

Deepfakes können für Bildungszwecke eingesetzt werden, zum Beispiel im schulischen Bereich oder in Museen.[8] Das nahezu immersive Erlebnis, das eine realistische audiovisuelle Darstellung historischer Persönlichkeiten oder von Ereignissen bietet, macht gesellschaftliche Zusammenhänge besonders eindrücklich erfahrbar. Dies betrifft nicht nur die Aufbereitung der Vergangenheit, sondern kann ebenso als Zugang zum gegenwärtigen politischen Geschehen oder auch möglichen Zukünften genutzt werden (STOA 2021). Die Darstellung von Zukunftsszenarien könnte zum Beispiel in einer Visualisierung unterschiedlicher Entwicklungsszenarien des Klimawandels bestehen und so die Auswirkungen heutiger gesellschaftlicher Entscheidungen und Handlungen verdeutlichen. Auch für realistische Simulationen beispielsweise gefährlicher Situationen können Deepfakes in der beruflichen Bildung genutzt werden (Erduran 2024).

Besonders im Bildungskontext ist die Reflexion der bewusst eingegangenen Täuschung durch ein Deepfake enorm wichtig, da alternative Geschichtsverläufe oder Darstellungen von Zukünften einen Denkprozess anregen sollen und keine Tatsachen abbilden.

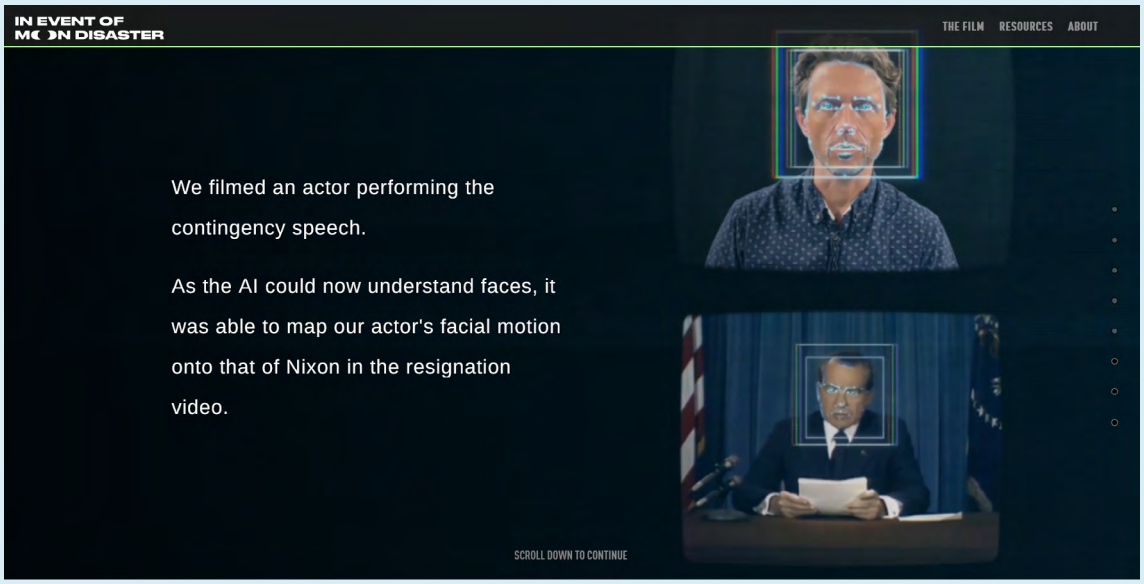
Im Bereich der Forschung können Deepfakes außer für Szenarien auch für Experimente etwa in der Sozialforschung genutzt werden (Eberl et al. 2022), ebenso wird ein therapeutischer Einsatz bei bestimmten psychischen Traumata diskutiert und wissenschaftlich untersucht (Van Minnen et al. 2022; Wiederhold 2021). Allerdings sind für therapeutische Anwendungen noch medizinische Studien durchzuführen und eine Reihe von ethischen Fragen zu klären (Hoek et al. 2024). Weniger problematisch erscheint dagegen die Nutzung von Deepfaketechnologien zur Erzeugung bzw. zum Klonen von Stimmen von Menschen, die aufgrund einer Erkrankung oder Operation ihre Stimme verloren haben (AP 2024). Hier bieten sich durch die Technologie einige Vorteile gegenüber den früher genutzten Sprachsynthesizern.

Kasten 3.4 Alternative Vergangenheit am Beispiel der Mondlandung (1974/2019)

In einem Videodeepfake wurde für eine Multimediadokumentation die Ansprache des US-Präsidenten Richard Nixon aufbereitet, die für den Fall eines Scheiterns der Apollo-11-Mission vorbereitet war (Lees 2024, S. 112 ff.; Abbildung 3.4). Dadurch werden sowohl die historische Erinnerung in eine neue Perspektive gerückt als auch die Möglichkeiten von Deepfaketechnologien plastisch illustriert. Ein ähnliches Beispiel aus den USA ist ein Audiodeepfake der Rede des US-Präsidenten John F. Kennedy, die er am Tag seiner Ermordung hätte halten sollen (Pawelec/Bieß 2021, S. 67).

8 Für das Dalí Museum in Florida wurde bereits 2019 mit Deepfaketechnologie eine virtuelle Präsentation des Künstlers in Form eines lebensgroßen Avatars erstellt, mit der die Besucher/innen auf mehreren Bildschirmen interagieren können (Mihailova 2021; https://the dali.org/exhibit/dali-lives/; 14.1.2025).

Abbildung 3.4 Making-of des Deepfakes der Rede Nixons zur Mondlandung



Quelle: Francesca Panetta, Halsey Burgund, MIT Center for Advanced Virtuality (CC-BY-NC-ND 4.0; Screenshot https://moondisaster.org/explainer; 3.9.2025)

3.3 Werbung, Marketing

Bei der Nutzung von Deepfakes im Bereich des Marketings soll gezielt der Einfluss genutzt werden, den diese auf die Meinung, Einstellungen und das Verhalten von Menschen haben. Die Technologie kann bei der Produktion von Werbeclips den Aufwand verringern und Anpassungen des Materials an die jeweilige Zielgruppe bis hin zur persönlichen Ansprache ermöglichen. Analog zu Anwendungen in der Filmindustrie können in Deepfakewerbeclips Prominente auftreten, ohne selbst vor der Kamera stehen zu müssen. Werbedeepfakes mit Prominenten wie Taylor Swift wurden allerdings auch bereits in betrügerischer Absicht eingesetzt, um an persönliche Daten oder das Geld von arglosen Kund/innen zu gelangen (Hsu/Lu 2024; Medler 2024), Wissenschaftler/innen wurden ohne ihr Wissen und Zutun für Onlinebetrugsversuche (Scams) missbraucht (Nordling 2024). Eine nicht betrügerische Anwendungsmöglichkeit dagegen sind Avatare im Kundendienst, die mithilfe von Deepfaketechnologie verbessert werden könnten.

Kasten 3.5 David-Beckham-Werbeclip in neun Sprachen (2019)

In einem Videostatement der britischen Organisation Malaria No More UK rief David Beckham zum Kampf gegen Malaria auf. Die Botschaft wurde in neun Sprachen übersetzt und die Gesichtsbewegungen des Fußballstars (nicht aber seine Stimme) wurden mithilfe von Deepfaketechnologie so angepasst, dass sie jeweils dem Gesagten entsprechen. Auf diese Weise sollten mehr Menschen in unterschiedlichen Ländern erreicht werden (Karaboga et al. 2024, S. 308).

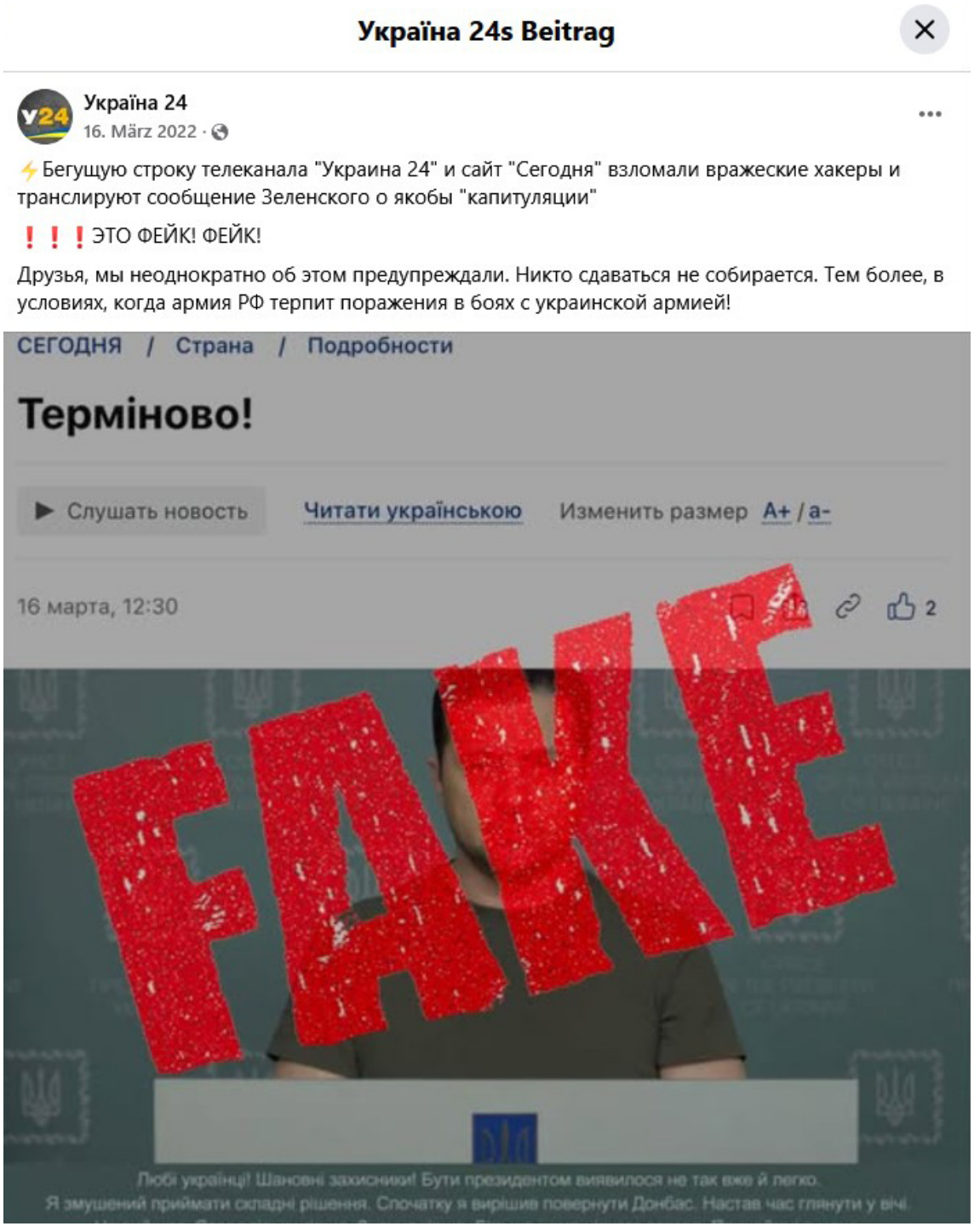
3.4 Politik und Meinungsbildungsprozesse

Deepfakes können bewusst mit dem Ziel produziert und verbreitet werden, die öffentliche Debatte oder den politischen Prozess zu beeinflussen. In demokratischen Gesellschaften ist dies insbesondere in Hinblick auf Wahlen mit einem großen Schadenspotenzial verbunden, da zum einen konkrete Entscheidungen beeinflusst, zum anderen indirekte Auswirkungen wie eine Störung des Vertrauens in eine faktenbasierte gesellschaftliche Debatte herbeigeführt werden können. Es steht zu befürchten, dass zunehmend ein gemeinsames Verständnis der Wirklichkeit aufgelöst und der Ausbildung extremer Ansichten Vorschub geleistet wird. Dies kann unabhängig von der tatsächlichen Fälschung von Darstellungen bereits durch den bloßen Verdacht bzw. die Unsicherheit darüber, ob es sich bei einer Nachricht um einen Deepfake handelt oder nicht, geschehen und den gesellschaftlichen Zusammenhalt destabilisieren.

In der Tat wurde in den letzten Jahren eine große Bandbreite von Deepfakes mit politischem Bezug auf nahezu der ganzen Welt beobachtet, insbesondere im Kontext von Wahlen (Allen/Nehring 2025; Schwieter/Gandhi 2024; Sensity B. V. 2024). Bislang wurden erst wenige Analysen zum tatsächlichen Einfluss auf die öffentliche Meinung veröffentlicht, nach diesen ist unklar bis zweifelhaft, ob es zu nennenswerten Effekten kam, insbesondere, weil die Qualität der Deepfakes gering war und sie kaum nennenswerte Verbreitung fanden (Adam 2024; CeMAS 2025, S. 6; Łabuz/Nehring 2024; Mukherjee 2024; Stockwell 2024). Beispiele von Deepfakes im Bereich von Politik und Meinungsbildung umfassen die Nutzung, um Kandidat/innen bei Wahlen (mit deren Einverständnis) in Bild- oder Videodarstellungen in einem guten Licht dastehen zu lassen und um die Wähler/innen in unterschiedlichen Sprachen anzusprechen (Bieber et al. 2024, S. 6 u. 12; Chowdhury 2024). Sie wurden aber auch genutzt, um die Gegenseite in dystopisch anmutenden Videoclips zu diskreditieren oder durch vorgebliche Audiomitschnitte missgünstiger Äußerungen zu diffamieren (Łabuz/Nehring 2024, S. 7; Schwieter/Gandhi 2024, S. 13 f.). Ganze Staaten sollten mithilfe von KI-generierten Nachrichtensendungen positiv dargestellt werden (Schwieter/Gandhi 2024, S. 13), in anderen Fällen wurden staatlichen Repräsentanten falsche, der Wirklichkeit diametral entgegenstehende Äußerungen unterstellt, um internationale Konflikte zu beeinflussen, wie im Fall eines Videos von 2022, in dem scheinbar Wolodymyr Selenskyj die Kapitulation erklärt (Abbildung 3.5; Kleemann 2023, S. 3). Weitere Deepfakes sollten Unsicherheit in der Bevölkerung erzeugen, sei es gezielt, um sie durch Deepfakeanrufe vom Wählen abzuhalten (Schwieter/Gandhi 2024, S. 9), sei es ungezielt durch Verbreitung von Bildern von Katastrophen wie einer vorgeblichen Explosion nahe des Pentagons (Schwieter/Gandhi 2024, S. 14). In letzterem Sinn wirkten auch Audiodeepfakes, bei denen der Anschein seriöser Medien wie der ARD-Tagesschau genutzt wurde, um Falschinformationen zu

verbreiten (Siggelkow/Reveland 2023). Zugleich kann mit Deepfakes bekannter Journalist/innen deren Reputation (siehe der Fall Rana Ayyub; Kapitel 3.7) bzw. die Seriosität von Medienorganisationen untergraben werden (Bieber et al. 2024, S. 12).

Abbildung 3.5 Warnung vor Selenskyj-Deepfake beim TV-Sender Ukraine 24 (2022)



Quelle: Ukraine 24 (Screenshot www.facebook.com/www.ukraine24.ua/posts/1847515155441880; 4.9.2025)

3.5 Polizeiarbeit und Justiz

Die polizeiliche Suche nach vermissten oder getöteten Personen kann durch Deepfakes, die eine gealterte Erscheinung gesuchter Personen realistisch darstellen oder Situationen überzeugend nachbilden, unterstützt werden (Busch/Oster 2023, S. 6). Auf diese Weise kann auch emotional besonders eindringlich für Aufmerksamkeit geworben und zur Unterstützung der Ermittlungsarbeit aufgerufen werden. Als emotional eindringlich wurde auch ein Deepfake beschrieben, das als Victim Impact Statement in einem Gerichtsverfahren in Arizona gezeigt wurde (Neff 2025). Die Schwester des Opfers eines Tötungsdelikts nutzte diese im angelsächsischen Rechtssystem vorgesehene Möglichkeit, die Opferperspektive im Verfahren deutlich zu machen, indem sie ihr Statement so formulierte – und mithilfe von Deepfaketechnologie visualisierte –, dass es schien, als würde der Tote selbst zum Angeklagten sprechen.

Kasten 3.6 Zeugenaufruf durch das Mordopfer (2022)

Die niederländische Polizei veröffentlichte ein mit KI generiertes Video eines fast 20 Jahre zuvor getöteten Jugendlichen als Zeugenaufruf (Abbildung 3.6). In dem Video ist der Getötete mit seiner Familie und Freunden zu sehen und fordert am Ende selbst Zeugen dazu auf, zur Aufklärung des Verbrechens beizutragen (Wilkens 2022).

Abbildung 3.6 Zeugenaufruf durch einen von der Polizei verbreiteten Deepfake des Getöteten

Quelle: Polizei Rotterdam (Screenshot www.youtube.com/watch?v=GgPCt2kzB2I; 2.9.2025)

Ethisch umstritten, aber seit 2020 nach § 184b Abs. 6 StGB[9] für die Kriminalpolizei zulässig, ist die rein synthetische Herstellung von Kinderpornografie zu Ermittlungszwecken. Diese wird verwendet, um Zutritt zu Foren und Netzwerken zu erhalten, der in der Regel nur bei einer Weitergabe kinderpornografischen Materials gewährt wird. Die eigentlichen Ermittlungen können oftmals erst nach einem solchen Zugang durchgeführt werden und schließlich zu einer konkreten Strafverfolgung führen.

Strafverfolgungsbehörden diskutieren zudem, ob bzw. unter welchen rechtlichen Voraussetzungen Technologien wie Voice Cloning durch Ermittlungsbehörden eingesetzt werden können (Margerie 2024; Margerie/Hartmann 2025). Anstelle von verdeckten Ermittler/innen oder Vertrauenspersonen könnten künstlich geschaffene Identitäten, deren Stimmen nach Vorbildern aus dem Umfeld von Verdächtigen generiert wurden, in der Ermittlungsarbeit eingesetzt werden und die damit verbundenen persönlichen Risiken und den Ressourcenaufwand verringern. Nach derzeitiger Rechtslage ist ein solches Vorgehen in Deutschland allerdings nicht möglich, es liegen zwar einzelne Ermächtigungen vor, die aber nicht kombiniert werden dürfen (Härtlein 2025, S. 5).

Deepfakes sind aber nicht nur ein Werkzeug der polizeilichen Arbeit, sondern stellen grundlegend die Glaubwürdigkeit von Bild-, Ton- und Videodokumenten in juristischen Verfahren infrage. Dies erschwert zunehmend die Arbeit von Strafverfolgungsbehörden, da Echtheitsnachweise von Medien schwierig zu führen sind (Pawelec/Bieß 2021, S. 59 f.). Manipuliertes „Beweismaterial" könnte unberechtigte Verdächtigungen zur Folge haben oder auch die Verschleierung krimineller Handlungen erleichtern.

3.6 Falsche Identitäten

Deepfakes ermöglichen es Menschen, andere Identitäten anzunehmen bzw. fremde Identitäten vorzuspielen. Dies wird immer wieder für betrügerische Zwecke genutzt, wenn etwa Angestellte von Unternehmen vorgeblich von ihren Führungskräften am Telefon oder per Videoanruf aufgefordert werden, Dinge zu tun, die in Wirklichkeit den Betrüger/innen nützen, zum Beispiel Geld auf deren Konto zu überweisen.

Kasten 3.7 Betrug per CEO-Deepfake (2019)

Frühe Bekanntheit erlangte der Fall einer Führungskraft in einem Energieunternehmen in Großbritannien, die erfolgreich dazu gebracht wurde, 220.000 Euro auf ein ungarisches Konto zu überweisen. Für den Betrug wurde ein Telefonanruf inszeniert, in dem die Stimme des Vorgesetzten aus dem deutschen Mutterunternehmen per Voice Cloning imitiert wurde (Mai et al. 2023; STOA 2021).

9 Strafgesetzbuch in der Fassung der Bekanntmachung vom 13. November 1998, zuletzt am 7. November 2024 geändert

Ebenso sind Erpressungsmöglichkeiten denkbar, die auf Rufschädigungsdrohungen basieren, oder auch Versuche, Einfluss auf die Entwicklung des Aktienkurses eines Unternehmens auszuüben, indem schädigende Inhalte zur Unternehmensentwicklung bzw. dessen Zustand verbreitet werden. Eine naheliegende Umsetzung besteht zum Beispiel in der verfälschten Darstellung von Führungskräften, denen per Deepfake moralisch fragwürdiges bzw. illegales Verhalten unterstellt wird oder die scheinbar falsche Informationen zur eigenen wirtschaftlichen Lage verbreiten (STOA 2021). Auch im privaten (zum Beispiel als neue Form des Enkeltricks) bzw. Arbeitskontext können Deepfakes zu betrügerischen oder rufschädigenden Zwecken eingesetzt werden (Spiegel Online 2024).

Die Verwendung von Deepfakes für kriminelle Absichten wird als ein zunehmendes Sicherheitsrisiko eingeschätzt (Pawelec/Bieß 2021). Aufgrund der hohen Qualität ist die Nachahmung von Menschen oder nur bestimmten Identitätsmerkmalen so überzeugend, dass dadurch nicht nur Menschen beeinflusst werden können, geschützte Informationen preiszugeben, sondern auch Identifizierungsverfahren, zum Beispiel beim Onlinebanking, überwunden werden können.

Umgekehrt werden Deepfaketechnologien auch genutzt, um persönliche Informationen zu schützen, etwa wenn in Videokonferenzen der Hintergrund durch einen virtuellen ersetzt wird, um die Privatsphäre der Nutzer/innen zu wahren, oder wenn KI-generierte Avatare die eigene Person vertreten und so zum Beispiel vor Nachstellung schützen (Belghaus/Cohrs 2024). Auch für Zeugenaussagen vor Gericht ist ein solcher Einsatz denkbar, gegenüber bisherigen in der Strafprozessordnung vorgesehenen Schutzoptionen wie etwa einer Verhüllung hätte dies den Vorteil, dass das Verhalten und die Emotionen der befragten Person erkennbar bleiben (Härtlein 2025, S. 6). Zudem wurden Chatbots mit künstlichen Identitäten entwickelt, um Betrüger/innen auf Social-Media-Plattformen in Interaktionen zu verwickeln und zur Preisgabe von Informationen über ihre Vorgehensweise und ihre Identität zu verleiten (Acharya et al. 2024).

3.7 Sexualisierte Gewalt, (Rache-)Pornografie, individuelle Verleumdung, Erpressung und Einschüchterung

Die ersten Deepfakes wurden 2017 zu pornografischen Zwecken entwickelt, indem in vorhandener Pornografie die Gesichter der Darstellerinnen durch Gesichter prominenter Frauen ersetzt wurden. Bearbeitung und Verbreitung geschahen auf einer nicht einvernehmlichen Basis und stellten eine schwere Schädigung der persönlichen Autonomie und Unversehrtheit der Angegriffenen dar (STOA 2021). Heutige Apps und Onlinedienste machen die Erzeugung realitätsnaher sexualisierter Deepfakes sehr leicht und zugänglich und benötigen nur wenige Vorgaben (Meineck 2024), sodass die Zahl der potenziellen Produzierenden wie der Opfer enorm groß ist.[10] In mehreren Studien wird davon ausgegangen, dass ein Großteil der Deepfakes (96 bzw. 98 %) pornografischer Art ist (Ajder et al. 2019; Security Hero 2023), allerdings ist nicht klar, wie belastbar diese Ergebnisse sind, und es fehlt an empirischer Forschung dazu (Birrer/Just 2024, S. 9).

Ähnlich wie bei dem Phänomen der konventionellen Rachepornografie (ohne Bildmanipulation) wird Deepfakepornografie auch zu Zwecken der Bloßstellung, Herabwürdigung, Einschüchterung

10 Das Unternehmen Security Hero (2023; ebenso Lakatos 2023) geht davon aus, dass zur Erzeugung von Deepfakes ein regelrechtes Ökosystem aus 15 Onlineforen und -communities entstanden ist, von dem ein „bedeutender Teil" auf „adult content" ausgerichtet ist und dem über 600.000 Personen zugerechnet werden. Auch in einem Bericht des für investigative Recherchen bekannten Kollektivs bellingcat (Koltai 2024) wird ein Netzwerk von Unternehmen beschrieben, dem eine Reihe von Onlinediensten und Apps zur Erstellung und Verbreitung von sexualisierten Deepfakes zugeordnet wird und das Zahlungsdienstleistungen unter anderem der Gamesbranche zur diskreten Bezahlung von Deepfakes nutzt.

oder Erpressung eingesetzt. Diese Fälle zielen fast ausschließlich auf Frauen (Ajder et al. 2019; Han et al. 2025; Hawkins et al. 2025; Security Hero 2023), sodass sie nicht allein als individuelle Übergriffe zu werten sind, sondern auch gesellschaftliche Auswirkungen haben, indem sie die Gleichstellung von Frauen im öffentlichen, digitalen Raum angreifen.

Kasten 3.8 Deepfakepornografie zur Einschüchterung und als Druckmittel gegenüber Frauen (2018)

Das Gesicht der indischen Journalistin Rana Ayyub wurde per Deepfake in ein pornografisches Video eingefügt, das viel Aufmerksamkeit in Indien erlangte. Es wird angenommen, dass das Aufkommen des Videos in Zusammenhang mit ihrer kritischen Berichterstattung über die damalige Regierungspartei stand. Als Folge wurde sie Ziel von Vergewaltigungsdrohungen und ihre Wohnortangaben zirkulierten im Internet (Pawelec/Bieß 2021, S. 57). Auch eine ganze Reihe von pornografischen Deepfakes von Schülerinnen in Spanien führte 2023 zunächst zu einer Einschüchterung der Angegriffenen. Allerdings gingen die Mütter mit dem Fall an die Öffentlichkeit und sorgten schließlich für eine nationale Diskussion über die Taten und Möglichkeiten des Schutzes bzw. der Gegenwehr (Köver 2023).[11]

Das Beispiel der Plattform MrDeepFakes verdeutlicht die Probleme, die durch nicht einvernehmliche sexualisierte Deepfakes Angegriffene bei der Durchsetzung ihrer Rechte haben. Die Plattform soll zuletzt nahezu 70.000 Videos und Bilder sowohl von Prominenten als auch – entgegen der selbst gesetzten Regeln – von Personen außerhalb der öffentlichen Wahrnehmung mit insgesamt milliardenfachen Aufrufen angeboten und mit 650.000 Nutzer/innen auch die Erstellung von Deepfakes vermittelt haben (Han et al. 2025; Lang 2025; Szeto et al. 2025). Ermittlungen der deutschen Polizei im Fall einer Bundesministerin waren einem Medienbericht zufolge ergebnislos verlaufen, weil sich die Betreiber der Seite nicht ermitteln ließen (Hoppenstedt et al. 2024). Journalistische Recherchen allerdings konnten mindestens eine in Kanada lebende Person als mutmaßlichen Administrator identifizieren (bellingcat 2025; Hoppenstedt et al. 2024; Szeto et al. 2025), dort werden nicht einvernehmliche sexualisierte Deepfakes allerdings bisher nicht vom Strafrecht erfasst. Dennoch wurde die Plattform im Mai 2025, kurz nachdem der mutmaßliche Betreiber von den Journalist/innen für eine Stellungnahme persönlich kontaktiert worden war, dauerhaft vom Netz genommen (Szeto et al. 2025).[12]

11 Die jugendlichen Täter wurden wegen der Verbreitung von Kinderpornografie zu Bewährungsstrafen von 1 Jahr verurteilt. Für die Betreiber der zur Erstellung der Deepfakes genutzten App hatte das Verfahren „bis heute keine Konsequenzen" (Hoppenstedt/Milatz 2025). Weitere Beispiele und Schilderungen von Angegriffenen finden sich in journalistischen Berichten (Bröckling/Tanner 2024; Compton/Hamlyn 2023; Hoppenstedt et al. 2024). Als Motive der Täter wurden anhand von deren Forenanfragen sexuelle Befriedigung, Machtausübung und ein vorgebliches Interesse der Community an dem gewünschten Deepfake identifiziert (Han et al. 2025, S. 6 f.).

12 Die Plattform Civitai, über die KI-Modelle unter anderem zur Erstellung von Deepfakes von Prominenten verbreitet wurden, kündigte im Mai 2025 an, jegliche Bilder von realen Personen aus ihrem Angebot auszuschließen. Als Grund dafür werden Bedingungen von Kooperationspartnern genannt, die Zahlungsdienstleistungen für die Plattform abwickeln, außerdem wird auf Gesetze wie den TAKE IT DOWN Act in den USA sowie die KI-Verordnung der EU verwiesen (theally 2025) (Kap. 4.1 u. 4.2).

4 Rechtliche Aspekte

Die Produktion und Verbreitung von Deepfakes berühren, ähnlich wie der Einsatz von anderweitig manipulierten Medien, ein weites Feld von Rechtsgrundlagen. Es ist absehbar, dass Deepfakes sowohl in positiv als auch in negativ konnotierten Kontexten in Zukunft verstärkt und mit größerer Reichweite verbreitet werden dürften. Angesichts der rasanten technologischen Entwicklungen, die zu Deepfakes mit deutlich besserer Qualität als bisher führen dürften, hat die Auseinandersetzung mit rechtlichen Rahmenbedingungen bzw. Sanktionsmöglichkeiten bei Verstößen an Bedeutung gewonnen.

Aufgrund der bereits beobachtbaren und erwartbaren Anwendungen von Deepfakes und ihren Auswirkungen sind in rechtlicher Hinsicht Persönlichkeitsrechte, Aspekte des Datenschutzes, der Kinder- und Jugendschutz sowie auch Urheberrechte betroffen (Block 2023, S. 20 ff.; Busch/Oster 2023, S. 5). Weiterhin tangieren Deepfakes den Bereich der Strafverfolgung durch die jeweils zuständigen Behörden. Neben eindeutig schädlichen Aspekten (Manipulation, Herstellung von kompromittierendem Material oder Darstellungen von Kindesmissbrauch) bieten Deepfakes aber auch zahlreiche positive Nutzungsmöglichkeiten, wie zum Beispiel die Unterstützung der polizeilichen Suche nach vermissten Personen, im Bildungssektor oder auch für künstlerische Anwendungen (Kapitel 3). Daher ist eine sorgfältige Abwägung von Regulierungen in Hinblick auf benevolente und malevolente Nutzungen unabdingbar, ein generelles Verbot etwa erscheint vor diesem Hintergrund nicht zielführend (Busch/Oster 2023, S. 42). Vielmehr steht die Auseinandersetzung mit den relevanten Akteuren und ihren individuellen Motiven auf interdisziplinärer Ebene im Vordergrund.

4.1 Regelungen in ausgewählten Ländern

Global betrachtet, gibt es mittlerweile in einigen Ländern Beispiele für deepfakespezifische gesetzliche Regelungen.[13] In den *USA* besteht seit Mai 2025 eine bundesgesetzliche Regelung zu Deepfakes durch den TAKE IT DOWN Act[14]. Dieser richtet sich gegen die nicht einvernehmliche Verbreitung sexualisierter Bilder im Internet und erfasst explizit auch Darstellungen, die mithilfe von maschinellem Lernen oder KI erstellt wurden (Rudl 2025). Vergehen können mit bis zu 2 Jahren Haft geahndet werden (bei minderjährigen Betroffenen bis zu 3 Jahren), Plattformen müssen entsprechendes Material aus ihrem Angebot entfernen, sofern sie einen Hinweis dazu erhalten.

Zuvor hatten bereits mehrere US-Bundesstaaten nicht einvernehmliche pornografische Deepfakes straf- bzw. zivilrechtlich reguliert (Busch/Oster 2023, S. 7 ff.; Graham 2024). In einer Reihe von Staaten bestehen zudem gesetzliche Rahmenbedingungen für die Herstellung und Verbreitung von politischen Deepfakes, beispielsweise in einem bestimmten Zeitraum vor politischen Wahlen, wenn diese darauf abzielen, den Ruf eines Kandidaten zu schädigen oder das Ergebnis einer Wahl

13 Außer in den USA und China wurden in weiteren Ländern Gesetze erlassen bzw. präzisiert, um insbesondere sexualisierte Deepfakes zu regulieren. 2024 wurde in Australien der Criminal Code Amendment (Deepfake Sexual Material) Act 2024 zur Änderung des Strafgesetzbuches verabschiedet, mit dem die Verbreitung von nicht einvernehmlichen sexualisierten Deepfakes unter Strafe gestellt wird (mit Ausnahmen für die Justiz, Medizin und Forschung). Auch in Frankreich wurde 2024 das Strafgesetzbuch durch das Loi n. 2024-449 du 21 mai 2024 visant à sécuriser et à réguler l'espace numérique (Loi SREN) erweitert, um die Verbreitung solcher Deepfakes als strafbar zu erfassen. Eine ähnliche Regelung existiert in Großbritannien mit dem Online Safety Act 2023. Zudem stellt der Data (Use and Access) Act 2025 die Erstellung solchen Materials bzw. die Beauftragung dazu unter Strafe. In Taiwan ist bereits neben der Verbreitung von nicht einvernehmlichen sexualisierten Deepfakes auch die Produktion strafbar (Taipei Times 2023).

14 Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks Act (TAKE IT DOWN Act), Public Law 119–12, vom 19. Mai 2025

zu beeinflussen (Kramer 2024). Davon explizit ausgenommen sind Parodien, Satire bzw. explizit als Deepfake gekennzeichnete Medien. Anzumerken ist, dass sich die Sanktionsfolgen je nach Bundesstaat von Schadensersatzforderungen durch die geschädigte Person bis hin zu strafrechtlich bewehrten Konsequenzen unterscheiden. Eine Besonderheit stellt der 2024 in Tennessee in Kraft getretene ELVIS Act[15] dar, der einer Person das Recht an persönlichen Merkmalen, wie der Stimme oder ihrem Bildnis, zuspricht und dieses schützt – explizit auch vor der Verwendung im Kontext von fälschlich als authentisch erscheinenden Darstellungen, wenn diese ohne Einverständnis der betroffenen Person erfolgt. Zudem ermöglicht die Trade Regulation Rule on Impersonation of Government and Businesses[16] der Federal Trade Commission, gegen Deepfakes im gewerblichen Bereich vorzugehen.

Bereits seit 2023 existiert in *China* mit den Administrative Provisions on Deep Synthesis for Internet-based Information Services (etwa Regelungen zu Deep Synthesis bei Internetinformationsdiensten – Deep-Synthesis-Regelungen) eine umfassende Regulierung von synthetischen Medien. Diese Verordnung bezieht sich nicht nur auf Audio- und Videodateien, sondern auch auf Texte, die mit KI generiert oder verarbeitet werden. Ein umfassendes Missbrauchsverbot, eine Klarnamen- und Kennzeichnungspflicht bilden die drei zentralen Kernpunkte der chinesischen Verordnung. Hervorzuheben ist hierbei, dass sich die Bestimmungen sowohl an die Angebotsseite (Deep Synthesis Services) als auch an die Nutzer/innen richten (Busch/Oster 2023, S. 9 f.). So wird zum Beispiel in Artikel 6 der Deep-Synthesis-Regelungen die missbräuchliche Nutzung entsprechender Dienste in umfassender Weise untersagt. Darunter fällt beispielsweise, Fake News oder auch Inhalte, die die nationale Sicherheit oder die Interessen Chinas gefährden oder dem Ansehen des Landes schaden, zu produzieren und zu verbreiten. Auch aus diesem Grund sind Diensteanbieter/innen zur Identitätsprüfung der Nutzer/innen, zum Beispiel anhand von Ausweisdokumenten (Artikel 9 Deep-Synthesis-Regelungen), sowie zur eindeutigen Kennzeichnung der erstellten oder bearbeiteten Inhalte verpflichtet (Busch/Oster 2023, S. 10). Ausnahmen für die Meinungs- oder Kunstfreiheit sind nicht vorgesehen.

4.2 EU-weite Regelungen

Am 21. Mai 2024 wurde durch den Rat der Europäischen Union die KI-Verordnung endgültig verabschiedet. Ziel ist die Schaffung einheitlicher Regeln für künstliche Intelligenz innerhalb der EU, was bisher global einzigartig ist. Im Einzelnen sollen mit der KI-Verordnung Regeln für die Entwicklung, das Inverkehrbringen und die Nutzung von KI-Systemen geschaffen werden, die mit den demokratischen Werten der EU und der Eindämmung von Gefahren durch missbräuchliche Nutzung vereinbar sind (EP 2023). Ein zentrales Regulierungsprinzip ist die Risikoeinstufung von KI-Systemen nach ihrem Anwendungsbereich.

So müssen KI-Systeme mit hohem Risiko für Gesundheit, Sicherheit oder Grundrechte natürlicher Personen (zum Beispiel Berührung anderer Produktsicherheitsvorschriften, Verwaltung/Betrieb von Kritischen Infrastrukturen [KRITIS], Bildung) vor dem Inverkehrbringen und während ihres gesamten Lebenszyklus fortlaufend auf mit ihnen verbundene Risiken hin evaluiert werden und diese Risiken gemindert oder zumindest kontrolliert werden (Art. 9 KI-Verordnung). Gänzlich verboten sind unter anderem KI-Systeme zur unterschwelligen Beeinflussung und Verhaltensmanipulation

15 Ensuring Likeness Voice and Image Security (ELVIS) Act of 2024, House Bill No, 2092, vom 21. März 2024
16 Trade Regulation Rule on Impersonation of Government and Businesses, 16 CFR Part 461, RIN 3084-AB71, vom 1. März 2024

(durch Manipulations- oder Täuschungstechniken), Ausnutzung von Schwächen, Social Scoring, Kriminalprognose, Auswertung biometrischer Daten und biometrische Echtzeitfernidentifikation (Art. 5 KI-Verordnung). Für letztere sieht die KI-Verordnung einen Strafverfolgungsvorbehalt vor. Für KI-Anwendungen mit geringer Risikoeinstufung werden lediglich minimale Transparenzpflichten vorgeschlagen, darunter fallen auch der Einsatz von Chatbots oder Deepfakes. Dies bedeutet insbesondere eine Kennzeichnungspflicht, wobei auch hier die Verfolgung von Straftaten ausgenommen ist (Art. 50 Abs. 2 und 4 KI-Verordnung).

Die Kennzeichnungspflicht für Deepfakes betrifft zum einen die Anbieter von KI-Systemen (unabhängig davon, ob sie die Systeme entgeltlich oder unentgeltlich anbieten). Sie müssen dafür sorgen, dass die Ausgaben des KI-Systems maschinenlesbar als künstlich erzeugt oder manipuliert gekennzeichnet werden. Das bedeutet nicht, dass menschliche Rezipienten diese als KI-generiert erkennen müssen (Härtlein 2025, S. 7). Zum anderen bestehen Regeln für die Betreiber, also diejenigen, die KI-Systeme anwenden. Diese müssen offenlegen, dass die Inhalte künstlich erzeugt oder manipuliert wurden, und dies in einer Weise, dass Betrachter/innen sie ohne Weiteres als KI-generiert erkennen können (Härtlein 2025, S. 7). Allerdings ist noch unklar, wer genau unter den Begriff „Betreiber" fällt. Natürliche Privatpersonen, die mit der Verwendung von KI ausschließlich persönliche Zwecke verfolgen, sind von der Verpflichtung ausdrücklich ausgenommen (Art. 2 Abs. 10 KI-Verordnung), ebenso gilt sie nicht für diejenigen, die nicht selbst das KI-System verwendet haben, sondern auf anderem Weg ein Deepfake erhalten haben und dies weiterverwenden. Für viele der in Kapitel 3 geschilderten Anwendungen dürfte die Kennzeichnungspflicht aus der KI-Verordnung daher keine Relevanz haben.

Während für die Umsetzung der KI-Verordnung noch Institutionen eingerichtet und Durchführungsbestimmungen geregelt werden müssen, besteht mit dem 2022 in Kraft getretenen und seit Februar 2024 vollumfänglich geltenden DSA bereits eine effektive Möglichkeit der Regulierung, die auf Deepfakes anwendbar ist. Der DSA enthält zwar keinen expliziten Bezug auf synthetische Medien, nimmt aber Anbieter digitaler Dienste sowie Plattformen in die Pflicht, Nutzer/innen vor illegalen Inhalten zu schützen, um ein sicheres Onlineumfeld zu gewährleisten. Von besonderer Bedeutung im Kontext von Deepfakes sind hierbei die Artikel 16 ff. DSA (Melde- und Abhilfeverfahren), die Hostingdienstanbieter dazu verpflichten, ein Meldeverfahren für rechtswidrige Inhalte einzurichten und umgehend Maßnahmen (zum Beispiel Löschung, Sperrung, Kontoschließung; Art. 17 DSA) einzuleiten, wenn gemeldete Inhalte als rechtswidrig erkannt werden. Bei einem Verdacht auf Straftaten sind Anbieter sogar dazu verpflichtet, diese Inhalte unverzüglich den jeweils zuständigen Strafverfolgungs- oder Justizbehörden zu melden (Art. 18 DSA).

Eine weitere Besonderheit des DSA stellen die weitergehenden Regelungen für sehr große Onlineplattformen und sehr große Onlinesuchmaschinen (mindestens 45 Mio. aktive Nutzer/innen pro Monat in der EU) dar (Art. 33 ff. DSA). Diese sind verpflichtet, zusätzlich zu den sonstigen Regelungen im DSA eine Bewertung der Schwere und Wahrscheinlichkeit systemischer Risiken vorzunehmen. Zu solchen zählen nicht nur die Verbreitung rechtswidriger Inhalte, sondern unter anderem auch „etwaige tatsächliche oder vorhersehbare nachteilige Auswirkungen auf die Ausübung der Grundrechte" (Art. 34 Abs. 1 Unterabs. 2 Buchst. b DSA) sowie auf Wahlprozesse und die öffentliche Sicherheit (Art. 34 Abs. 1 Unterabs. 2 Buchst. c DSA). Die Anbieter sollen zudem Maßnahmen zur Minderung der ermittelten Risiken vorsehen (Art. 35 DSA). Dabei wird explizit auf eine gegebenenfalls gebotene Kennzeichnung solcher Inhalte verwiesen, die „fälschlicherweise als echt oder wahrheitsgemäß" erscheinen (Art. 35 Abs. 1 Buchst. k DSA). Nicht zuletzt aufgrund

der Pflichten von Dienstebetreibern in Bezug auf rechtswidrige Inhalte ist davon auszugehen, dass der DSA auch im Kontext von Deepfakes von Bedeutung ist (Busch/Oster 2023, S. 45).

Begleitet wird der DSA von freiwilligen Selbstverpflichtungen, wie zum Beispiel dem Verhaltenskodex gegen Desinformation (EK 2022), den bisher 42 Internetunternehmen und spezialisierte Organisationen unterzeichnet haben und der unter anderem Maßnahmen zur Bekämpfung von Deepfakes wie die Aufdeckung entsprechenden Contents und die Zusammenarbeit zwischen den Plattformbetreibern beinhaltet, die jedoch auf freiwilliger Basis erfolgen. Eine weitere EU-Regulierung von Deepfakes ist durch die Aktualisierung der Richtlinie 2011/93/EU[17] geplant, die auch speziell auf die neuen technologischen Möglichkeiten der künstlichen Intelligenz eingeht. Hierzu hat die EU-Kommission im Februar 2024 einen ersten Entwurf auf den Weg ins Gesetzgebungsverfahren gebracht.[18] Bereits in Kraft ist die Richtlinie (EU) 2024/1385[19], die explizit eine strafrechtliche Regelung von nicht einvernehmlichen pornografischen Deepfakes vorsieht. Sie muss von den Mitgliedstaaten bis 14. Juni 2027 in nationales Recht umgesetzt werden.

4.3 Deutschland: Zivilrecht

Im deutschen Recht gibt es bisher keine explizite Regulierung von Deepfakes (Busch/Oster 2023, S. 67). Allerdings orientiert sich das geltende Recht ohnehin in erster Linie am Schutz von Rechtsgütern und nicht an bestimmten Technologien, sodass grundsätzlich eine Reihe von Regelungen herangezogen werden können, um schädliche Folgen von Deepfakes abzuwehren. Im Bereich des Zivilrechts sind dabei der Bildnisschutz nach dem Kunsturhebergesetz (KunstUrhG),[20] das allgemeine Persönlichkeitsrecht, das Urheberrecht sowie das Datenschutzrecht besonders relevant (Busch/Oster 2023, S. 11).

Das *KunstUrhG* regelt sondergesetzlich das Recht am eigenen Bild. Gemäß § 22 S. 1 dürfen Bildnisse nur verbreitet oder öffentlich zur Schau gestellt werden, wenn der bzw. die Abgebildete eingewilligt hat. Das Recht am eigenen Bild ist betroffen, wenn eine Person begründet annehmen kann, sie könne auf einem Bild identifiziert werden (Specht-Riemenschneider 2018, Rn. 4). Die oft üblichen Maßnahmen wie Verpixelung oder das Einblenden von Augenbalken sind nicht zwangsläufig ausreichend. Es ist daher davon auszugehen, dass die Veröffentlichung eines Deepfakevideos, in dem eine Person erkennbar ist, in das durch § 22 S. 1 KunstUrhG geschützte Recht am eigenen Bild eingreift (Busch/Oster 2023, S. 12 f.).

In diesem Zusammenhang ist jedoch zwischen Herstellung und Verbreitung zu unterscheiden. Denn § 22 KunstUrhG erfasst nur die Verbreitung bzw. öffentliche Zurschaustellung, nicht aber die Herstellung von Bildnissen.[21] Der Gesetzgeber versteht unter „Verbreiten" auch das digitale Versenden im privaten Rahmen, wenn nicht ausgeschlossen werden kann, dass das Risiko einer nicht mehr kontrollierbaren Kenntnisnahme durch weitere Personen besteht (Busch/Oster 2023, S. 13). Dies kann beispielsweise bei der Weiterleitung von WhatsApp-Inhalten der Fall sein. Im

17 Richtlinie 2011/93/EU vom 13. Dezember 2011 zur Bekämpfung des sexuellen Missbrauchs und der sexuellen Ausbeutung von Kindern sowie der Kinderpornografie sowie zur Ersetzung des Rahmenbeschlusses 2004/68/JI des Rates

18 https://oeil.secure.europarl.europa.eu/oeil/en/procedure-file?reference=2024/0035(COD); 12.8.2025

19 Richtlinie (EU) 2024/1385 vom 14. Mai 2024 zur Bekämpfung von Gewalt gegen Frauen und häuslicher Gewalt

20 Gesetz betreffend das Urheberrecht an Werken der bildenden Künste und der Photographie (KunstUrhG) vom 9. Januar 1907, zuletzt am 16. Februar 2001 geändert

21 Bei einem Bildnis handelt es sich um Darstellungen, bei denen eine oder mehrere Personen im Fokus stehen, wohingegen bei Bildern die Abbildung einer Örtlichkeit oder eines Geschehens zentral sind.

Zusammenhang mit dem Begriff der „öffentlichen Zurschaustellung“ ist bisher umstritten gewesen, wie das Hochladen auf Social-Media-Kanälen zu behandeln ist, da hier eine Unterscheidung zwischen öffentlichen und privaten Profilen erforderlich ist. Dabei ist jedoch zu bedenken, dass das (gegebenenfalls auch unberechtigte) Kopieren und Weiterverbreiten von digitalen Inhalten nur geringen Hürden unterliegt und daher sowohl von einer Verbreitung als auch von einer öffentlichen Zurschaustellung im Sinne des KunstUrhG ausgegangen werden kann (Busch/Oster 2023, S. 14).

Für die Frage der Rechtmäßigkeit der Verbreitung bzw. Zurschaustellung kommt es in erster Linie darauf an, ob die Betroffenen ihre Einwilligung erklärt haben. Eine Einwilligung muss sich dabei in der Regel explizit auf die Verwendung des Fotos oder Videos als Deepfake beziehen (Busch/Oster 2023, S. 15). In bestimmten Fällen dürfen Bildnisse auch ohne Einwilligung der Betroffenen verbreitet und zur Schau gestellt werden. Dies gilt etwa für Bildnisse aus dem Bereich der Zeitgeschichte bei einem Informationsinteresse der Öffentlichkeit, für Bilder, auf denen Personen nur als Beiwerk neben einer Landschaft oder sonstigen Örtlichkeit erscheinen, sowie für Bilder von Versammlungen, Aufzügen o. Ä., sofern keine einzelnen Personen besonders hervorgehoben werden. Auch für Zwecke der Rechtspflege und der öffentlichen Sicherheit können Bildnisse ohne Einwilligung veröffentlicht werden, zum Beispiel Fahndungsfotos (§ 24 KunstUrhG).

Im journalistischen Kontext sind bei den erstgenannten Anwendungsfällen verschiedene Motivationen zu unterscheiden, da nach der Rechtsprechung des Bundesverfassungsgerichts ein Informationsinteresse der Öffentlichkeit im Rahmen der Meinungsfreiheit nur bei unveränderten Bildnissen anzunehmen ist. Für unrichtige Informationen müsste der Informationswert jeweils begründet werden. Sollen zum Beispiel mithilfe eines Deepfakes im Rahmen einer Dokumentation die Folgen einer Überflutung einer Küstenstadt mit ertrinkenden Menschen simuliert werden, ohne dass einzelne Personen eindeutig identifiziert werden können, so ist dies – bei entsprechender Kennzeichnung – zulässig (Busch/Oster 2023, S. 18). Dagegen verletzt ein Deepfake einer noch lebenden Person der Zeitgeschichte, von der keine oder nur wenige Bildnisse existieren, das Persönlichkeitsrecht, wenn damit lediglich die Neugier eines bestimmten Publikums befriedigt werden soll. Der erwähnte § 24 KunstUrhG gewährt weitreichende Bildnisrechte (Vervielfältigung, Verbreitung, öffentliches Zurschaustellen), lässt aber offen, ob auch manipulierte Bildnisse darunterfallen. Busch und Oster (2023, S. 19) sehen daher eine Klärung als wünschenswert an, ob die Verwendung von Deepfakes zum Beispiel im Rahmen polizeilicher Fahndungen zulässig ist.

Während das KunstUrhG nur die Verbreitung und öffentliche Zurschaustellung von Bildnissen einzelner Personen sanktioniert, können Deepfakes allgemeiner einen Eingriff in das *allgemeine Persönlichkeitsrecht* darstellen. Dieses umfasst als Grundrecht neben dem Recht am eigenen Bild unter anderem das Recht auf Privatheit und den Schutz der persönlichen Ehre, das Recht am gesprochenen Wort sowie das Recht, von der Unterschiebung nicht getätigter Äußerungen verschont zu bleiben (Busch/Oster 2023, S. 19 f.).[22] Der Schutz der Persönlichkeitsrechte ist immer gegen mögliche andere Rechte abzuwägen, in Abhängigkeit vom Inhalt der Berichterstattung kommen unterschiedliche Abwägungsparameter zum Tragen. So ist zu unterscheiden zwischen wahren, aber die Privatsphäre betreffenden, sowie unwahren und ehrenrührigen Äußerungen. Je nach Schwere der Verletzung gegen das allgemeine Persönlichkeitsrecht können Geschädigte nicht nur einen Anspruch auf Unterlassung (bei unwahren Tatsachenbehauptungen) sowie zudem einen

22 Auch das Recht an der eigenen Stimme ist vom allgemeinen Persönlichkeitsrecht umfasst. Das Landgericht Berlin sprach im August 2025 einem bekannten Synchronsprecher Schadensersatz zu, weil eine von einem KI-System erzeugte Stimme, die seiner täuschend ähnlich war, ohne sein Einverständnis in zwei Youtube-Clips verwendet worden war (LG Berlin II, Urteil vom 20.08.2025, Az. 2 O 202/24; das Urteil ist noch nicht rechtskräftig).

Gegendarstellungsanspruch erwirken, sondern auch Schadensersatz und Entschädigung (Schmerzensgeld) erhalten. Auch deshalb ist zu erwarten, dass dem allgemeinen Persönlichkeitsrecht im Kontext von Deepfakes eine besondere Bedeutung zukommen wird, vor allem weil dieses auch gegen das Unterschieben nicht getätigter Äußerungen schützt, die weder ehrenrührig sind noch das Privatleben betreffen (BVerfGE 54, 148 [155]; Busch/Oster 2023, S. 20 f.).

Abzuwägen bleibt allerdings auch hier zwischen dem Recht der Betroffenen (Schutz der Persönlichkeitsrechte) und der Kommunikationsfreiheit (Busch/Oster 2023, S. 21). Es ist also zuvorderst zu prüfen, inwieweit die mutmaßliche Falschäußerung eine Person des öffentlichen Lebens betrifft oder zu einer Debatte von allgemeinem Interesse beiträgt. Ebenso ist die streitgegenständliche Information selbst zu berücksichtigen. Wer beispielsweise selbst in einem ersten Schritt reputationsmindernde bzw. abwertende Äußerungen tätigt, muss prinzipiell im Gegenzug ebenfalls reputationsmindernde Reaktionen hinnehmen und kann sich dann nicht auf das allgemeine Persönlichkeitsrecht berufen. Einen weiteren wichtigen Punkt bildet die Erlangung des Materials, auf dessen Grundlage der Äußernde (Falsch-)Informationen herstellt. Hat beispielsweise eine Person einer anderen ein Nacktfoto freiwillig zugeschickt, mit dem die andere einen pornografischen Deepfake erstellt, kann es (ausnahmsweise) sein, dass das Persönlichkeitsrecht nicht verletzt ist (wenn etwa mit der Erstellung einer solchen Darstellung zu rechnen war).

Im Zusammenhang mit dem allgemeinen Persönlichkeitsrecht bleibt jedoch der Wahrheitsgehalt als wesentliches Element erhalten. So müssen wahre Tatsachenbehauptungen grundsätzlich hingenommen werden, sofern nicht der Schaden, der von ihnen droht, das Interesse an ihrer Veröffentlichung überwiegt (Busch/Oster 2023, S. 22). Allerdings werden Privatpersonen gegen eine Weitergabe auch richtiger Informationen an die Öffentlichkeit geschützt, insbesondere, wenn die Informationen das Privat- oder gar Intimleben betreffen und die Person willkürlich ausgewählt wurde (Busch/Oster 2023, S. 22). Unrichtige Informationen dagegen werden regelhaft nicht vom verfassungsrechtlich garantierten Recht auf freie Meinungsäußerung gedeckt, weil sie nicht der „Aufgabe zutreffender Meinungsbildung“ dienen können (BVerfGE 54, 208 [219]). Dies betrifft Deepfakes in besonderem Maße, da sie definitionsgemäß auf die Erweckung eines falschen Anscheins abzielen. Etwas anders liegt der Fall, wenn der/dem Äußernden nicht bewusst ist, dass er/sie unwahre Äußerungen verbreitet. Sofern er/sie bestehende Sorgfaltspflichten geachtet hat, ist dies zu seinen/ihren Gunsten zu werten. In jedem Fall ist aber zu berücksichtigen, ob es um Tatsachenbehauptungen oder aber Werturteile etwa im Kontext der Meinungsäußerung geht, denn nur erstere sind dem Wahrheitsbeweis zugänglich (Busch/Oster 2023, S. 24).

Einen Sonderfall stellen Sachaufnahmen (von Gegenständen) dar (Busch/Oster 2023, S. 24 f.). In der deutschen Rechtsordnung existiert kein „Recht am Bild der eigenen Sache“. Hieraus folgt, dass die Herstellung von entsprechenden Bildnissen nicht ausschließlich der/dem Eigentümer/in der Sache zustehen und diese auch ohne dessen/deren Einwilligung genutzt bzw. verbreitet werden dürfen. Allerdings darf dabei nicht gegen andere Regelungen verstoßen werden. Aufnahmen von Gebäuden können zum Beispiel Urheberrechte verletzen, sofern sie nicht durch die Panoramafreiheit (§ 59 UrhG[23]) gedeckt sind. Sachaufnahmen, bei denen ein eindeutiger Personenbezug hergestellt werden kann, können das allgemeine Persönlichkeitsrecht der Eigentümer/innen bzw. Nutzer/innen verletzen (zum Beispiel ein Deepfake einer als brennend dargestellten Villa einer bekannten Fernsehmoderatorin, Busch/Oster 2023, S. 25).

23 Urheberrechtsgesetz vom 9. September 1965, zuletzt am 23. Oktober 2024 geändert

Wenn Unternehmen von Deepfakes betroffen sind, so können sich diese nicht auf das allgemeine Persönlichkeitsrecht berufen, aber gegebenenfalls auf das *Unternehmenspersönlichkeitsrecht* zurückgreifen. Dieses kann verletzt werden, wenn eine Äußerung geeignet ist, das unternehmerische Ansehen in der Öffentlichkeit zu beeinträchtigen. Suggeriert beispielsweise ein Deepfake fälschlicherweise, dass auf einem Biobauernhof Käfighaltung betrieben wird, und in der Folge sinken die Verkaufszahlen, so stehen dem betroffenen Unternehmen unter Umständen Unterlassungs- und gegebenenfalls Schadensersatzansprüche zu (Busch/Oster 2023, S. 25 ff.).

Eine besonders große Bedeutung hinsichtlich der rechtlichen Einordnung von Deepfakes kommt dem *Urheberrecht* zu. Die Erstellung eines Deepfakes kann in bestehendes Urheberrecht eingreifen, wenn dadurch eine geschützte Leistung verändert wird (Busch/Oster 2023, S. 25 ff.). Zum Beispiel kann ein Videoclip, der als Ausgangsmaterial für einen Face-Swap-Deepfake genutzt wird, urheberrechtlich geschützt sein und seine Nutzung einen Verstoß darstellen. Urheberrechtlicher Schutz kann gemäß § 2 Abs. 2 UrhG für persönliche geistige Schöpfungen (Sprachwerke, Werke der Musik, Lichtbildwerke, Filmwerke und Ähnliches) in Anspruch genommen werden. Das UrhG sichert dem/der Urheber/in das Recht auf eine angemessene Vergütung für die Werknutzung zu (= wirtschaftliches Verwertungsrecht etwa durch Vervielfältigung, öffentliche Zugänglichmachung, Nutzung des Senderrechts). Die jeweilige Person wird in ihrer „geistigen und persönlichen Beziehung zum Werk und in dessen Nutzung" geschützt (§ 11 UrhG).

Durch das Urheberpersönlichkeitsrecht ist nicht nur geregelt ob, sondern auch wie ein Werk veröffentlicht werden darf. So stellt die öffentliche Zurschaustellung unveränderter wie auch veränderter urheberrechtlich geschützter Werke ohne Einwilligung des Urhebers/der Urheberin eine Rechtsverletzung dar. Aus dem UrhG leitet sich außerdem eine grundsätzliche Pflicht zur Quellenangabe ab, die ausschließt, dass sich Dritte zu Urheber/innen eines Werkes erklären. Im Fall von Deepfakes kann zudem bedeutsam sein, nicht als Urheber/in eines Werks genannt zu werden, etwa wenn im Deepfake noch das Originalwerk erkennbar ist, dessen Urheber/in aber nicht mit dem Deepfake in Verbindung gebracht werden möchte. Ein solches Recht auf Anerkennung der Nichturheberschaft folgt nicht ohne Weiteres aus dem UrhG, kann sich aber gegebenenfalls aus § 12 BGB und dem allgemeinen Persönlichkeitsrecht ergeben (Busch/Oster 2023, S. 28).

Beispiel 4.1

L nimmt ein Foto mit einem Smartphone auf und schickt es an D. Dieser verändert es mittels eines Deepfakes und lädt es frei zugänglich auf einer Social-Media-Seite hoch. D verstößt dadurch gegen das UrhG, unter anderem durch Eingriff in das Erstveröffentlichungsrecht von L. L hat einerseits Anspruch auf die Urheberschaft am ursprünglichen Foto, andererseits darauf, nicht als Urheberin des resultierenden Deepfakes genannt zu werden.

Quelle: gekürzte Fassung nach Busch und Oster (2023, S. 28 f.)

Bearbeitungen oder andere Umgestaltungen eines Werkes dürfen nur mit Zustimmung des Urhebers/der Urheberin veröffentlicht oder verwertet werden (§ 23 Abs. 1 S. 1 UrhG). Die Herstellung dagegen bleibt zustimmungsfrei, es sei denn, es handelt sich um Verfilmungen, Erstellung von Bauwerken oder Werken der bildenden Kunst nach geschützten Entwürfen oder Bearbeitungen einer Datenbank (§ 23 Abs. 2 UrhG). Insofern verstoßen Deepfakes nicht unbedingt gegen das Urheberrecht, wenn sie ausschließlich für den privaten Gebrauch erstellt werden. Ein Werk gilt zudem nicht als Bearbeitung oder Umgestaltung im Sinn von § 23 Abs. 1 S. 1 UrhG, wenn es einen hinreichenden Abstand zum benutzten Werk wahrt. Dies ist im Einzelfall zu prüfen, dürfte aber bei Deepfakes in der Regel nicht der Fall sein (Busch/Oster 2023, S. 29 f.).[24]

Schrankenregelungen (§§ 44a ff. UrhG) erlauben unter bestimmten Umständen die Vervielfältigung und auch die öffentliche Verbreitung von Werken ohne Zustimmung der Urheber/innen, zum Beispiel wenn die Vervielfältigungen keine eigenständige wirtschaftliche Bedeutung haben, dem Zweck des Data Minings (TAB 2022b, S. 295 ff.) dienen oder wenn Bildnisse zu Zwecken der Rechtspflege und der öffentlichen Sicherheit vervielfältigt oder verbreitet werden sollen. Dabei gilt allerdings ein Änderungsverbot (§ 62 Abs. 1 S. 1 UrhG), sodass die Schrankenregelungen für Deepfakes grundsätzlich nicht in Anspruch genommen werden können – auch nicht in Fällen, in denen die Erstellung eines Deepfakes im öffentlichen Interesse liegt, etwa im Bereich der Rechtspflege oder der öffentlichen Sicherheit (Busch/Oster 2023, S. 30 f.). Der in Kapitel 3.5 beschriebene Fall eines mit KI veränderten Fahndungsbildes wäre daher nur mit Einverständnis des/der Fotograf/in des Originalbildes zulässig. Eine Ausnahme vom Änderungsverbot gilt seit Einführung des § 51a UrhG im Jahr 2021 allerdings für Karikaturen, Parodien oder Pastiches, diese schließt auch Memes ein (Busch/Oster 2023, S. 31; Kreutzer 2022; Lauber-Rönsberg 2024). Sie dürfen jedoch nicht darauf angelegt sein, als Original wahrgenommen zu werden, sondern müssen als Bearbeitung erkennbar sein (zu weiteren Voraussetzungen Kraetzig 2024).

Nicht nur auf deutscher, sondern auch auf europäischer Ebene hat der Gesetzgeber die Pflicht, *Diskriminierungen* aktiv entgegenzutreten. Im Zusammenhang mit KI-Systemen ist zunächst festzuhalten, dass nicht diese selbst diskriminieren, da Diskriminierung eine soziale Praxis darstellt (TAB 2020, S. 31). Allerdings ist seit langem bekannt, dass die Anwendung von KI-Systemen diskriminierende Ergebnisse erzeugen kann. So konnten zahlreiche Befunde zeigen, dass Geschlechterstereotype (zum Beispiel bei Bildern zu bestimmten Berufsgruppen; Naik/Nushi 2023) oder rassistische Grundannahmen (zum Beispiel systematische Benachteiligung von Afroamerikanern bei der Kriminalprognose; Angwin et al. 2016) durch KI-Systeme reproduziert werden. Bereits das Datenmaterial, das zum Training der KI-Systeme verwendet wird, bildet die reale Welt nur stark verzerrt ab (Beuth et al. 2024).

Sofern Deepfakes, die diskriminierend wirken, durch staatliche Stellen erzeugt wurden, greifen die Grundrechte der Gleichberechtigung und der Nichtdiskriminierung, wie sie etwa in Artikel 3 Grundgesetz (GG)[25] und Artikel 21 der Charta der Grundrechte der Europäischen Union[26] niedergelegt sind (Busch/Oster 2023, S. 33). Dagegen sind Diskriminierungen durch Privatpersonen oder Unternehmen nicht per se verboten, da beispielsweise damit verbundene Straftatbestände (wie Beleidigung, Volksverhetzung etc.) nur bei vorsätzlichem Handeln strafrechtlich relevant sind.

24 Wenn jedoch nicht ein konkreter Inhalt als Vorlage eines Deepfakes genutzt wird, sondern ein Deep-Learning-Modell anhand verschiedener Vorbilder trainiert und für die Generierung eines neuen Inhalts genutzt wird, könnte die Text-und-Data-Mining-Schranke des § 44b UrhG greifen und eine Verwendung ermöglichen (Härtlein 2025, S. 9). Ob bzw. inwiefern diese Schranke auf generative KI-Systeme angewendet werden kann, ist jedoch noch Gegenstand der juristischen Debatte (Dornis 2024; Maamar 2023).

25 Grundgesetz für die Bundesrepublik Deutschland vom 23. Mai 1949, zuletzt am 22. März 2025 geändert

26 Charta der Grundrechte der Europäischen Union (2007/C 303/01) vom 14.12.2007

Formen der Diskriminierung können etwa durch bildgenerierende KI-Systeme entstehen, wenn die erzeugten Abbildungen von Personen „zu einer vorrangig weißen und westlich konnotierten Darstellung“ tendieren (Burghard/Hornung 2024, S. 65) oder wenn zum Prompt „terrorist“ die dargestellten Personen „allesamt männlich und mit dunkler Hautfarbe“ sind (Busch/Oster 2023, S. 33). Solche unbewusste Diskriminierung durch Private wird auch in Artikel 10 der KI-Verordnung nur für Hochrisikosysteme erfasst, insofern besteht „eine ganz erhebliche Regelungslücke im Hinblick auf Diskriminierungen mittels Deepfakes durch Privatpersonen“ (Busch/Oster 2023, S. 34). Entsprechende gesetzgeberische Maßnahmen sollten allerdings nach Ansicht der Autoren des Gutachtens (Busch/Oster 2023) unter Berücksichtigung freiheitlich-demokratischer Werte und unter Beachtung des Gebots der Technologieneutralität geprüft werden und in einen breiten gesellschaftlichen Kontext eingebettet werden, der auch Fragen der KI-bezogenen Medienkompetenz einschließt.

Deepfakes weisen auch aus *datenschutzrechtlicher* Perspektive eine hohe Relevanz auf. Sofern in einem Deepfake eine Person identifizierbar ist (personeneigene Merkmale, Name, Stimme etc.), werden personenbezogene Daten verarbeitet. Ob die damit verbreitete Information der Wahrheit entspricht, spielt dabei keine Rolle.

Beispiel 4.2

Eine Podcasterin erstellt mithilfe eines KI-Systems ein Interview, in dem die Stimme einer bekannten Unternehmerin geklont wird. Auch die geklonte Stimme stellt ein personenbezogenes Datum der Unternehmerin dar. Daher können neben Ansprüchen aus dem allgemeinen Persönlichkeitsrecht weitere Ansprüche nach der Verordnung (EU) 2016/679[27] (Datenschutz-Grundverordnung) geltend gemacht werden, zum Beispiel Löschung, Schadensersatz und gegebenenfalls sogar Schmerzensgeld, sofern keine Einwilligung oder ein journalistisches oder anderes Interesse (Art. 6 Abs. 1 Buchst. d bis f Datenschutz-Grundverordnung) vorliegen.

Quelle: gekürzte Fassung nach Busch und Oster (2023, S. 35).

Auch das Sammeln, Erfassen, Organisieren, Speichern, Abfragen, Übermitteln etc. von Daten wird von der Datenschutz-Grundverordnung erfasst. Dieser Teilbereich ist etwa in Bezug auf Trainingsdaten, die zur Erstellung von Deepfakes erhoben werden, nicht trivial. So gibt es bereits Beschwerden gegen KI-Unternehmen, deren Systeme Informationen über bestimmte Personen aus ihrem Datenmaterial generieren (Tagesschau.de 2024). Im obigen Podcastbeispiel stellt etwa das Trainieren eines KI-Systems mit Tonaufnahmen der Unternehmerin genauso wie das Hochladen des Interviews eine Verarbeitung personenbezogener Daten dar und unterliegt damit den Regelungen der Datenschutz-Grundverordnung (Busch/Oster 2023, S. 36). Rechtsfolgen des Datenschutzrechts erstrecken sich auf zivil- und verwaltungsrechtliche Bereiche und können unter Umständen auch bis ins Strafrecht reichen (Busch/Oster 2023, S. 37).

27 Verordnung (EU) 2016/679 vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung)

4.4 Deutschland: Strafrecht

Die Erstellung und Verbreitung von Deepfakes können mit strafbaren Handlungen verbunden sein. Dies beginnt unter Umständen bereits bei der *Vorbereitung eines Deepfakes*. Werden beispielsweise Personen visuell oder akustisch ausgespäht, um Deepfakes zu erstellen, so kommen Straftatbestände wie die Verletzung der Vertraulichkeit des Wortes oder Verletzung des höchstpersönlichen Lebensbereichs und von Persönlichkeitsrechten durch Bildaufnahmen (§ 201 u. 201a StGB) und eventuell das Ausspähen von Daten (§ 202a StGB) in Betracht. Die *Erstellung eines Deepfakes* wird insofern nicht von den genannten Regelungen erfasst, als nur tatsächliche Tonaufzeichnungen bzw. Bildaufnahmen geschützt sind (Busch/Oster 2023, S. 38).[28] Auch weitere Strafnormen, die die Verbreitung pornografischer Inhalte (§ 184 StGB), Beleidigung (§ 185 StGB) oder Nachstellung (§ 238 StGB) zum Gegenstand haben und grundsätzlich auf (pornografische) Deepfakes anwendbar sind, sind an zusätzliche Voraussetzungen wie bestimmte Zielgruppen (zum Beispiel Minderjährige), ehrverletzende oder wiederholte Handlungen geknüpft, die bei Deepfakes nicht unbedingt gegeben sind (Busch/Oster 2023, S. 38). Die Richtlinie (EU) 2024/1385 sieht eine Strafbarkeit unter anderem von sexuellen Deepfakes vor, die vorsätzlich und ohne Einwilligung der betreffenden Person erstellt oder verbreitet werden (Art. 5 Abs. 1 Buchst. b, auch die Drohung damit ist umfasst). Ihre Umsetzung in deutsches Recht würde die genannte Regelungslücke schließen.

Die *Verbreitung rechtswidriger Inhalte* per Deepfakes ist strafbar – unabhängig davon, ob sie sich auf einem Speichermedium befinden oder per Informations- oder Kommunikationstechnik übertragen werden. Hierzu zählen zuvorderst sämtliche Varianten von kinder/jugendpornografischen Inhalten (§§ 184b u. c StGB), bei denen schon allein die Verbreitung, der Erwerb, der Besitz und auch der Abruf strafrechtlich bewehrt sind.[29] Es ist dabei unerheblich, ob ein reales Geschehnis abgebildet wird, sofern ein entsprechender Anschein erweckt wird (Busch/Oster 2023, S. 39).[30] Weiterhin können auch Inhalte mit Volksverhetzung, Gewaltdarstellungen, öffentlichen Aufforderungen zu Straftaten, das Vortäuschen einer Straftat, Verunglimpfungen eines Staats- oder Verfassungsorgans sowie Beschimpfungen von beispielsweise Religionsgemeinschaften (§§ 90 ff., 111, 131, 166, 145d, 184 ff., 184b, 184c StGB) strafrechtlich verfolgt werden.

Auch Verstöße gegen das KunstUrhG, das Urheber- und Datenschutzrecht können strafrechtliche Konsequenzen nach sich ziehen. Besonders hervorzuheben sind in diesem Kontext Ehrverletzungsdelikte (§ 185 ff StGB), da für einen Deepfake in diesen Fällen genau zwischen der durch den Deepfake geschaffenen Situation und der im Deepfake kommunizierten Information zu unterscheiden ist (Busch/Oster 2023, S. 39). Beispielsweise könnte ein Deepfake eine Ministerin zeigen, die scheinbar zugibt, regelmäßig Betäubungsmittel zu konsumieren. Dies trifft in Wirklichkeit aber nicht zu und der/die Urheber/in des Deepfakes weiß das auch. Ein solcher Deepfake könnte nicht als Verleumdung geahndet werden, weil dafür eine andere Person als die Betroffene selbst eine Unwahrheit verbreiten müsste. Allerdings könnte der Deepfake als Beleidigung geahndet werden,

28 Fraglich ist, inwiefern § 201a StGB die Verbreitung von Deepfakes erfasst (Härtlein 2025, S. 11 f.). Einige Autor/innen halten § 201a Abs. 2 StGB für anwendbar, der das Zugänglichmachen von Bildaufnahmen unter Strafe stellt, wenn diese geeignet sind, dem Ansehen der abgebildeten Person erheblich zu schaden (Lantwin 2020, S. 80; Thiel 2021, S. 204). Auch die frühere Bundesregierung hatte sich entsprechend geäußert (Bundesrat 2024, S. 21). Der Deutsche Juristinnenbund (DJB 2023, S. 7) weist darauf hin, dass Bedenken gegen die Bestimmtheit des Tatbestandes bestehen und Wertungswidersprüche anzunehmen sind.

29 Ausnahmen bestehen lediglich für die Erfüllung staatlicher Aufgaben (§ 184b Abs. 5 Nr. 1 StGB) sowie für dienstliche Handlungen im Rahmen strafrechtlicher Ermittlungsverfahren (§ 184b Abs. 6 StGB).

30 Im Fall pornografischer Deepfakes, die nicht unter die Verbote der §§ 184a, b u. c StGB fallen, ist der Besitz und die Verbreitung im Grundsatz erlaubt. Für die Verbreitung gelten allerdings bestimmte Regeln (§ 184 StGB), die sich jedoch nicht auf die Zustimmung der abgebildeten Personen beziehen, das heißt, nicht einvernehmliche sexualisierte Deepfakes von Erwachsenen sind vom Pornografiestrafrecht nur bedingt erfasst.

weil der/die Urheber/in die Ministerin mit dem Deepfake in eine kompromittierende Situation bringt (Bundesgerichtshof, Beschluss vom 03.11.1983, Az. 1 StR 515/83). Die Rechtsprechung geht jedoch seit einiger Zeit bei (verbalen) sexuellen Übergriffen nur noch von einer für die Beleidigung konstitutiven Missachtung aus, wenn über die Tat hinaus eine Herabsetzung des Opfers erfolgt, sodass Ehrverletzungsdelikte nur bedingt auf Deepfakes anwendbar sein dürften (Härtlein 2025, S. 15). Für den Fall des Stalkings existiert seit 2021 in § 238 Abs. 1 Nr. 7 StGB ein Straftatbestand, der sich auf Deepfakes anwenden lässt. Demnach ist eine wiederholte Veröffentlichung von – scheinbar von den Betroffenen selbst stammenden – Äußerungen oder Darstellungen strafbar, die diese herabwürdigen oder verächtlich machen, sodass ihre Lebensgestaltung nicht unerheblich beeinträchtigt wird. Im Zusammenhang mit Ehrverletzungsdelikten ist insbesondere die „Wahrnehmung berechtigter Interessen" ein Rechtfertigungsgrund, der etwa im Rahmen der journalistischen Tätigkeit geltend gemacht werden kann. Angriffe auf die Menschenwürde (Art. 1 Abs. 1 GG) sind aber weder durch die Presse- noch durch die Kunstfreiheit gedeckt.

Das StGB enthält eine Reihe weiterer Vorschriften, die gegen Deepfakes angeführt werden können. So können Deepfakes als Betrug geahndet werden, wenn durch sie ein Irrtum hervorgerufen wird, der Menschen dazu bringt, Vermögen auf andere zu übertragen (§ 263 Abs. 1 StGB), ebenso kommen Bedrohung, Nötigung und Erpressung (zum Beispiel mittels Erzeugung kompromittierenden Materials) in Betracht. Nach § 23 des Gesetzes zum Schutz von Geschäftsgeheimnissen (GeschGehG)[31] macht sich zudem strafbar, wer Deepfakes dazu verwendet, um an fremde Geschäftsgeheimnisse zu gelangen (Busch/Oster 2023, S. 40). Allerdings kann es in Fällen, wie beispielsweise dem Identitätsdiebstahl mittels Deepfakes im Rahmen von Videoidentifizierungsverfahren (Video-Ident-Verfahren), der Fall sein, dass dieser für sich genommen „keine strafrechtlichen Konsequenzen nach sich zieht" (Salemi/Steffes 2022, S. 676), weil die gesetzlich vorgesehenen Tatbestandsmerkmale nicht erfüllt sind.

31 Gesetz zum Schutz von Geschäftsgeheimnissen vom 18. April 2019

5 Regelungslücken, Herausforderungen der Rechtsdurchsetzung und Handlungsoptionen

Mit Deepfakes geht eine ganze Reihe von Anwendungen einher, die Menschen, Unternehmen und Institutionen in demokratischen Gesellschaften schaden können. In mehreren Fällen ist es bereits zu materiellen Schäden durch Betrug mithilfe von Deepfakes gekommen. Der Bereich der pornografischen Deepfakes wird von Expert/innen als zahlenmäßig größter Anwendungsbereich angesehen, dabei besteht diesbezüglich eine große Ungewissheit und es ist mit einer hohen Dunkelziffer insbesondere nicht einvernehmlicher Fälle zu rechnen. Entgegen vieler Befürchtungen, Deepfakes würden politische Prozesse beeinflussen, kam es trotz vieler Fälle insbesondere im Kontext von Wahlkämpfen bisher nicht zu nennenswerten Auswirkungen, wobei Expert/innen weiterhin große Gefahren sehen.

Grundsätzlich beruhen Deepfakes auf Technologien, die sehr leistungsstark sind und auch Nutzen bringen können. Wie die Anwendungsbeispiele (Kapitel 3) zeigen, können Deepfakes in den Bereichen Kunst und Unterhaltung, Bildung, Marketing und auch der Politik und Strafverfolgung neue Angebotsformen ermöglichen und die Ansprache von bestimmten Zielgruppen verbessern. Die rasante technische Entwicklung im Bereich der Bild-, Video- und Texterzeugung und -bearbeitung durch KI-Technologien lässt dabei erwarten, dass Deepfakes in der absehbaren Zukunft noch leichter zu erstellen sein werden und ihre gesellschaftliche Bedeutung, sowohl im positiven wie auch im negativen Sinn, zunehmen wird.

5.1 Debatte zu den rechtlichen Instrumenten

Auch wenn es in Deutschland noch keine Gesetze mit explizitem Bezug zu Deepfakes gibt, werden im Zuge der Umsetzung bzw. Anwendung europäischer Rechtsakte wie der KI-Verordnung, dem DSA und der Richtlinie (EU) 2024/1385 Rahmenbedingungen für die Anwendung von Deepfakes vorgegeben. An diesen wird allerdings von einigen Seiten Kritik geäußert. Ferner lassen sich die Regelungen des deutschen Zivil- und Strafrechts zwar, wie in den Kapiteln 4.3 und 4.4 gezeigt, auf Deepfakes anwenden, allerdings lassen sie Regelungslücken offen.

5.1.1 Kritik an EU-rechtlichen Regelungen

Kritiker/innen geht die Vorgabe der KI-Verordnung, von einem KI-System generierte synthetische bzw. – als Deepfake – manipulierte Inhalte als künstlich erzeugt oder manipuliert zu kennzeichnen, nicht weit genug (Busch/Oster 2023, S. 42 f.). Eine böswillige Nutzung von Deepfakes könne gesellschaftliche Sprengkraft entfalten und beispielsweise im Kontext von Wahlmanipulation demokratische Grundwerte gefährden (Łabuz 2024; Vanberghen 2024). Die gegenwärtig bestehende Kennzeichnungspflicht wird als ineffektiv und nicht ausreichend erachtet. Zudem gilt sie nicht für natürliche Privatpersonen, *sofern* diese ausschließlich zu persönlichen Zwecken handeln (Härtlein 2025, S. 8), und es kann davon ausgegangen werden, dass Akteure, die eine Schädigung anderer Personen oder Unternehmen beabsichtigen, sich nicht an gesetzlich festgelegte (Kennzeichnungs)Regelungen halten werden (Block 2023; Karaboga 2023; Łabuz 2024). Am Beispiel von

Deepfakevideos von Papst Leo XIV. ist zudem zu beobachten, dass diese auf Social-Media-Plattformen eine enorme Reichweite erhalten, selbst wenn sie explizit als fiktiv gekennzeichnet sind (dpa factchecking 2025).[32]

In Bezug auf den DSA wird kritisiert, dass zum einen keine konkreten Fristen für Onlinediensteanbieter zur Löschung nach Meldung rechtswidriger Inhalte genannt werden. Die Verpflichtung zur unverzüglichen Reaktion seitens der Onlinediensteanbieter besteht demnach erst bei Kenntnisnahme (Busch/Oster 2023, S. 53; DJB 2023). Dem könnte durch behördliche Löschungsanordnungen begegnet werden, wobei fraglich ist, inwieweit dies bei der Vielzahl der gemeldeten Inhalte überhaupt durchsetzbar ist. Als exemplarisches Beispiel kann die Zentrale Meldestelle für strafbare Inhalte im Internet (ZMI) im Bundeskriminalamt angeführt werden, bei der im Zeitraum von Juni 2021 bis Februar 2024 nahezu 21.000 Meldungen eingingen, von denen 88 % als strafrechtlich relevant eingestuft wurden (in der Mehrzahl Fälle von Volksverhetzung) (Bundesregierung 2024).

Zum anderen ist die Implementierung eigener Maßnahmen zur Erkennung rechtswidriger Inhalte, beispielsweise zur Verhinderung des Uploads potenziell strafbarer Deepfakes, für Hostingdienste nicht obligatorisch. Im Rahmen der Haftungsprivilegierung, wie sie im DSA festgelegt ist, können Plattformen für den Upload und die Verbreitung von rechtswidrigen Inhalten per se nicht in Haftung genommen werden, wenn sie darüber keine Kenntnis haben (Busch/Oster 2023, S. 53; DJB 2023). Diese Konstellation bedingt die Verfügbarkeit eines externen Meldesystems, welches sich mutmaßlich auf aktive Nutzer/innen oder auch Organisationen verlassen muss. Zu diesem Zweck formuliert der DSA in Art. 22 das System der unabhängig agierenden „vertrauenswürdigen Hinweisgeber" (Trusted Flagger; die Statusvergabe erfolgt durch Prüfung nach Antragstellung durch den „Koordinator für digitale Dienste" des jeweiligen Mitgliedstaates), deren Meldungen bevorzugt zu behandeln sind. Die Funktionsfähigkeit eines solchen Systems lässt sich bislang jedoch noch nicht beurteilen. In diesem Zusammenhang ist zu überlegen, inwiefern die Erweiterung des Status „vertrauenswürdig" auch auf Contentproduzenten (zum Beispiel öffentlich-rechtliche Anstalten) zu befürworten wäre (Trusted Content), um der Verbreitung von Deepfakes im Zuge von Fake News entgegenzuwirken (Busch/Oster 2023, S. 54 f.).

Nicht zuletzt entspricht das im DSA verankerte Melde- und Prüfverfahren durch die Hostinganbieter eben nicht einer rechtlichen Prüfung im engeren Sinne. Gerade im Zusammenhang von Verstößen gegen Persönlichkeitsrechte, die immer eine rechtliche Einzelfallprüfung erfordern, ist dies gemäß Busch und Oster (2023, S. 54) durchaus problematisch, weil die Regelung den Rechtsschutz der Betroffenen einschränkt.[33] Zudem zeigt eine Analyse der französischen Parlamentswahlen 2024, dass die großen Onlineplattformen ihrer Verpflichtung aus dem DSA, Deepfakes zur Minderung der systemischen Risiken für Wahlen als solche zu kennzeichnen, nicht nachgekommen sind (auch die Parteien haben die von ihnen verwendeten Deepfakes und KI-generierten Inhalte nicht also solche gekennzeichnet) (Schueler et al. 2024, S. 16).

32 Ein weiteres Beispiel ist ein Video vom August 2025, in dem die indonesische Finanzministerin Sri Mulyani Indrawati verächtlich über Lehrer/innen spricht. Das Video war mit dem Wasserzeichen des KI-Systems Veo gekennzeichnet und somit als Deepfake erkennbar. Dennoch wurde es vielfach geteilt und soll zum Anheizen der gewaltsamen Proteste der Bevölkerung gegen die Regierung beigetragen haben (Saathoff 2025).

33 Auch in Fällen, in denen Deepfakes als rechtswidrig und daher zu löschen bewertet wurden, müssen Hostinganbieter dieser Pflicht nur für das gemeldete Video und sinngleiche Inhalte nachkommen. Bei darüber hinausgehenden Abweichungen muss das Meldeverfahren gegebenenfalls jeweils neu durchlaufen werden. In einem Beschluss vom 4.3.2025 (Az. 16 W 10/25) stellte das OLG Frankfurt allerdings klar, dass Hostinganbietern die Überprüfung anderer, „identischer und kerngleicher" Videos zumutbar ist. Hintergrund des Rechtsstreits waren mehrere sehr ähnliche Deepfakes, in denen scheinbar der Arzt und Wissenschaftsjournalist Eckart von Hirschhausen für Produkte wirbt. Er beklagt „jede Menge Verzögerungstaktik und Schutzbehauptungen" von Seiten der Plattformen (von Hirschhausen 2025).

5.1.2 Regelungslücken im deutschen Recht

Das deutsche Rechtssystem ist aufgrund seiner Technologieneutralität, also der primären Orientierung am Schutz von Rechtsgütern und nicht an bestimmten Technologien, prinzipiell gut für die Herausforderungen neuer Technologien wie Deepfakes gerüstet (Busch/Oster 2023, S. 67). Allerdings ist zu prüfen, inwiefern durch Deepfakes Lücken im Rechtsschutz entstehen, die zu schließen sind.

Im Gutachten und in der juristischen Diskussion wird in Bezug auf das Strafrecht herausgestellt, dass nicht geklärt ist, inwiefern § 201a StGB auf Deepfakes anwendbar ist (Busch/Oster 2023, S. 38; DJB 2023, S. 7; Härtlein 2025, S. 11 f.). Insbesondere bei nicht einvernehmlichen sexualisierten Deepfakes stellt sich die Frage, wie sie strafrechtlich zu erfassen sind (Härtlein 2025, S. 12 ff.). Eine Verletzung des Rechts am eigenen Bild ist zwar gemäß § 33 KunstUrhG strafbar, allerdings „handelt es sich nicht nur um ein absolutes Antragsdelikt, sondern auch um ein Privatklagedelikt" (Bundesverband Trans* e. V. et al. 2023, S. 17), daher wird befürchtet (und von vielen Betroffenen berichtet; Bröckling/Tanner 2024), dass entsprechende Vergehen kaum strafrechtlich verfolgt werden. Einen sehr weitreichenden, wenn auch noch nicht konkreten Vorschlag für ein Verbot von „Deepfakes von echten Menschen" hat eine Gruppe von Künstler/innen und anderen Prominenten im März 2025 in einer Petition vorgebracht.[34] Die Forderung umfasst Deepfakes und andere Darstellungen von scheinbar echten Menschen (zum Beispiel Social Bots), bei denen keine Einwilligung der Betroffenen vorliegt, auch in Bereichen wie Satire und politischer Kommunikation. Bis zum Juni 2025 wurde die Petition mehr als 300.000-mal auf der Plattform We Act! unterschrieben.

Eine „erhebliche Regelungslücke" (Busch/Oster 2023, S. 34) im Zivilrecht besteht in Bezug auf mögliche Diskriminierung durch Privatpersonen, da diese durch die geltenden Regelungen in vielen Fällen nur bei Vorsatz erfasst wird. Für den Bereich der behördlichen Nutzung von Deepfakes zur Rechtspflege und öffentlichen Sicherheit ergeben sich mögliche urheberrechtliche Probleme, da die entsprechende Schrankenregelung mit einem Änderungsverbot verknüpft ist (Busch/Oster 2023, S. 30 f.), für Einsatzszenarien im Rahmen verdeckter Ermittlungen besteht bisher keine Rechtsgrundlage (Härtlein 2025, S. 5).

5.2 Herausforderungen der Rechtsdurchsetzung

Während aufgrund von Regelungslücken die strafrechtliche Ahndung von Verletzungen der Persönlichkeitsrechte problematisch sein kann, fehlt es nicht grundsätzlich an materiellrechtlichen Regelungen, um Ansprüche aufgrund von Schädigungen durch Deepfakes zu erheben (Busch/Oster 2023, S. 49). Allerdings stellt die Durchsetzung zivilrechtlicher Ansprüche ein wesentliches Problem dar, auch Behörden und Staatsanwaltschaften stoßen bei Rechtsverstößen im Internet auf Ermittlungsprobleme. Deepfakes können global verbreitet werden, ihre Bekämpfung erfordert daher regelmäßig eine internationale Zusammenarbeit (siehe beispielsweise Meywirth et al. 2023 für eine Auswahl an Kollaborationen im Bereich Cybercrime). Ein Problem dabei ist, den für einen bestimmten Fall gültigen nationalen Rechtsrahmen sowie die Gerichtszuständigkeit festzustellen (Busch/Oster 2023, S. 58 f.; Karaboga et al. 2024, S. 157). Das deutsche Recht kann zwar in vielen Fällen auch dann angewendet werden, wenn Täter/innen vom Ausland aus agieren, bei

34 https://weact.campact.de/petitions/deepfakes-verbieten (13.6.2025)

einer Erweiterung um deepfakespezifische Straftatbestände ist allerdings gegebenenfalls das Strafanwendungsrecht zu bedenken.[35] Zudem bietet sich eine internationale Zusammenarbeit beim Umgang mit Deepfakes an, da viele Länder von den möglichen negativen Auswirkungen auf Individuen und Gesellschaft betroffen sind.

Ein weiteres Problem stellt die Ermittlung der für einen Deepfake verantwortlichen Person dar, da eine Verbreitung zum Beispiel über Onlineplattformen unter Pseudonym erfolgen kann. Die Plattformen sind zum Teil nicht im Besitz der für die Identifizierung nötigen Daten, in zivilrechtlichen Verfahren haben die Betroffenen zum Teil keine Handhabe oder müssen großen Aufwand betreiben, um sie zur Herausgabe der Daten zu bewegen (Busch/Oster 2023, S. 49 ff.; DJB 2023, S. 10 f.; Hogertz et al. 2024, S. 7; Rüdiger/Bayerl 2020). So verpflichtet der DSA zwar Onlinediensteanbieter zur Zusammenarbeit mit den EU-ansässigen Behörden, um die Strafverfolgung zu gewährleisten. Privatrechtliche Ansprüche von Geschädigten gegenüber Unternehmen mit Sitz außerhalb der EU sind davon jedoch nicht erfasst (Busch/Oster 2023, S. 59). Gerade Onlineplattformen spielen „eine zentrale Rolle bei der Verbreitung synthetischer Medien" (Busch/Oster 2023, S. 67), daher wird gefordert, sie entsprechend in Verantwortung zu nehmen.

Zwei laufende Gesetzgebungsprozesse zielen unter anderem darauf ab, Lücken bei der strafrechtlichen Verfolgung von Deepfakes zu schließen und könnten insofern auch die Rechtsdurchsetzung erleichtern. Zum einen regelt die Richtlinie (EU) 2024/1385 die Strafbarkeit pornografischer Deepfakes. Sie muss allerdings noch in nationales Recht umgesetzt werden. Zum anderen hat der Freistaat Bayern im Mai 2024 einen Gesetzesantrag „zum strafrechtlichen Schutz von Persönlichkeitsrechten vor Deepfakes" in den Bundesrat eingebracht (Freistaat Bayern 2024). Demnach soll das StGB durch den neuen Straftatbestand der „Verletzung von Persönlichkeitsrechten durch digitale Fälschung" ergänzt werden. Für einen Verstoß hiergegen ist entweder eine Geldstrafe oder eine Freiheitsstrafe von bis zu 2 Jahren vorgesehen, im Fall einer Veröffentlichung oder der Verletzung des höchstpersönlichen Lebensbereichs bis zu 5 Jahren (Freistaat Bayern 2024, S. 7). Der neu zu schaffende § 201b StGB soll, wie bereits § 201a, in den Kreis der Privatklagedelikte aufgenommen werden. Schon 2021 hatte die Konferenz der Justizministerinnen und Justizminister (Justizministerkonferenz 2021) in einem Beschluss die Prüfung gesetzgeberischen Handlungsbedarfs gefordert. 2023, 2024 und 2025 wurden die Appelle an das Bundesjustizministerium in Bezug auf den Schutz der demokratischen Willensbildung, den Schutz von Persönlichkeitsrechten, den Schutz vor digitaler Gewalt und bildbasierter sexualisierter Gewalt erneuert (Justizministerkonferenz 2023a, 2023b, 2024, 2025). Der Bundesrat hat im Juli 2024 und – aufgrund des Grundsatzes der Diskontinuität – erneut im Juli 2025 einen Gesetzentwurf auf Grundlage des bayerischen Vorschlags beschlossen, der seit August 2025 dem Bundestag zusammen mit einer Stellungnahme der Bundesregierung vorliegt (Bundesrat 2024, 2025).

35 Deutsches Strafrecht kann grundsätzlich angewendet werden, wenn eines der folgenden Kriterien erfüllt ist (Härtlein 2025, S. 9): Die Tat wurde in Deutschland begangen; der Taterfolg ist in Deutschland eingetreten; der/die Täter/in oder das Opfer ist deutsche/r Staatsangehörige/r und die Tat ist am Tatort strafbar; es gelten bestimmte Sonderregelungen. Das Kriterium des Taterfolgs betrifft besonders Gefährdungsdelikte, bei denen zur Vollendung kein messbarer Erfolg in der Außenwelt vorausgesetzt wird (zum Beispiel Volksverhetzung oder auch Stalking). Gemäß der aktuellen Rechtsprechung des Bundesgerichtshofs ist das deutsche Strafrecht in diesen Fällen nicht anwendbar.

5.3 Handlungsoptionen

5.3.1 Technologien zur Detektion und Verifizierung

Rechtliche Rahmenbedingungen spielen eine wichtige Rolle für die Vermeidung bzw. Bekämpfung negativer Auswirkungen von Deepfakes. Einen weiteren Baustein dazu stellt die technologiebasierte Erkennung von Deepfakes dar, die beispielsweise für Faktenprüfung, aber auch für die automatische Überprüfung des Contents auf großen Onlineplattformen besonders relevant ist. Zwar lassen sich eine Reihe von Merkmalen von Deepfakes angeben, wie beispielsweise auffällig verwaschene Konturen, eine monotone Sprachausgabe oder Unstimmigkeiten auf Ebene der Metadaten oder einzelner Pixel, anhand der die Erkennungsrate durch betroffene Personen bzw. medienforensische Expert/innen deutlich verbessert werden kann (BSI o. J.). Doch angesichts immer besser werdender Manipulationen sind zunehmend technologiebasierte Ansätze zur Erkennung von Deepfakes gefragt (Kapitel 2.3).

Mehrere Forschungsprojekte sind gegenwärtig damit befasst, Deepfakedetektoren zu entwickeln, welche eine automatisierte Erkennung von Deepfakes ermöglichen sollen. Als Beispiel sei hier das vom Bundesministerium für Bildung und Forschung (BMBF) geförderte Forschungsprojekt „Fake-ID“ angeführt, welches sich mit der automatisierten Detektion von Deepfakes, zum Beispiel im Zuge des weitverbreiteten VideoIdent-Verfahrens, befasst.[36] Seit Herbst 2024 werden zudem im Rahmen des Innovationswettbewerbs SPRIND Funke zwölf Projekte unterstützt, die innerhalb von 13 Monaten einen Prototypen zur Erkennung und Prävention von Deepfakes entwickeln sollen.[37] Für einen Überblick über die bisherige Technikentwicklung sei auf Masood et al. (2023) verwiesen, die eine Liste von Deepfakegeneratoren und Detektoren anführen. Die darin genannten Detektoren weisen zum Teil eine Detektionsquote von bis zu 99 % auf. Auch eine Übersicht des US-amerikanischen National Institute of Standards and Technology (NIST 2024, S. 27) berichtet für KI-generierte bzw. manipulierte Videos Detektionsraten zwischen 62 und 99 %. Allerdings ist zu berücksichtigen, dass sich die Detektoren auf bestimmte Deepfaketechnologien (zum Beispiel Face Reenactment und Lip Sync) konzentrieren und folglich mit einem Detektor nicht alle Arten von Deepfakes zuverlässig erkannt werden können.

Im Test von Detektoren, der im Rahmen der Deepfakestudie von TA-Swiss durchgeführt wurde (Karaboga et al. 2024, S. 107 ff.), wurden die positiven Ergebnisse nicht bestätigt. Vielmehr gaben die getesteten Detektoren „sowohl falsch-negative als auch falsch-positive Ergebnisse aus […], sodass durch die Nutzung von Detektoren […] auch die Vertrauenswürdigkeit echter Inhalte untergraben werden könnte“ (Karaboga et al. 2024, S. 113 f.). Zudem offenbarte der Test weitere Hürden bei der Anwendung der Detektoren, etwa einen aufwändigen Genehmigungsprozess sowie häufig unklare Ergebnisse, die auf Wahrscheinlichkeiten und nicht auf einer Ja-/Nein-Einstufung eines Videos als Deepfake beruhen.

Ein grundsätzliches Problem der automatisierten Detektion besteht darin, dass technologische Weiterentwicklungen im Bereich der Manipulationstechniken die Erkennungsmethoden rasch überholen können und somit den technologischen Wettbewerb fortschreiben dürften (siehe hierzu auch Beckmann et al. 2023; Busch/Oster 2023). Alternative technische Ansätze zur Kennzeichnung

36 www.hhi.fraunhofer.de/abteilungen/vit/projekte/fakeid.html; www.bioid.com/deepfake-detection/ (4.9.2025)
37 www.sprind.org/impulse/challenges/funke-deepfake (4.9.2025); im Mai 2025 wurden die sieben vielversprechendsten Projekte für die zweite Stufe der Förderung ausgewählt.

und Verifizierung authentischer Inhalte sind auf die Rückverfolgung des Ursprungs der Daten ausgerichtet (Busch/Oster 2023, S. 60): Digitale Wasserzeichen können sichtbar oder verborgen den eigentlichen Informationen hinzugefügt werden und so zum Beispiel erkennbar machen, durch wen ein gegebenes Artefakt erstellt wurde. Sie sind schwer zu entfernen bzw. verändern und können automatisiert ausgelesen werden (Lv 2021; NIST 2024, S. 5 ff.). Durch kryptografische Signierung können Wasserzeichen oder auch – zusätzlich zur eigentlichen Information gespeicherte – Metadaten geschützt werden, sodass spätere Veränderungen sichtbar gemacht werden (NIST 2024, S. 17). Digitale Signaturen können außerdem zur Nachverfolgung in eine Blockchain eingebunden werden (Fraga-Lamas/Fernandez-Carames 2020; Hasan/Salah 2019; Heidari et al. 2024). Diese Methoden sollen den Nachweis der Herkunft sowie der Unversehrtheit von Daten ermöglichen.

Allerdings muss berücksichtigt werden, dass malevolente Akteure Technologien etwa zur Manipulation von Wasserzeichen entwickeln können, zudem könnten durch die Verwendung von Wasserzeichen oder Metadaten ungewollt sensible Informationen über die Urheber/innen eingebettet werden (NIST 2024, S. 15, 19). Außerdem erfordern entsprechende Kennzeichnungsmaßnahmen ein einheitliches Vorgehen zumindest innerhalb einzelner Branchen wie etwa der Medienindustrie. Ansonsten existieren gekennzeichnete und ungekennzeichnete Inhalte nebeneinander und es wäre im Fall einer fehlenden Kennzeichnung unklar, ob dies auf einen Deepfake oder lediglich die fehlende Beteiligung an der Kennzeichnung hindeutet (Karaboga et al. 2024, S. 336).

5.3.2 Anpassungen im Rechtssystem und Sicherung demokratischer Prozesse

Neben der Beobachtung und gegebenenfalls Schließung von Rechtslücken im Zusammenhang mit Deepfakes und den dargestellten Herausforderungen im Bereich der Durchsetzung von Rechtsansprüchen besteht ein weiteres Risiko in der Bedrohung der Integrität gerichtlicher Verfahren. Die Möglichkeit der Manipulation von Beweismitteln durch Deepfakes stellt grundsätzlich die Authentizität von Beweismitteln in audiovisueller Form infrage. Zwar können diese durch Sachverständige überprüft werden, jedoch sind auch hier die fehlenden oder begrenzten Detektionsmöglichkeiten in Bezug auf qualitativ hochwertige Deepfakes zu bedenken.

Dies bezieht sich nicht nur auf die Prüfung der Authentizität audiovisueller Beweismittel, sondern auch auf die rechtlichen Konsequenzen für die am Verfahren Beteiligten. So wurde beispielsweise eine Frau in Pennsylvania (USA) angezeigt, weil sie verleumderische Deepfakes hergestellt und verbreitet haben sollte (Kleeman 2024). Im Rahmen der Überprüfung – allerdings nach Veröffentlichung der Anschuldigungen durch die Staatsanwaltschaft – stellte sich heraus, dass die ihr zugeordneten Videos authentisch waren und die Behauptung, es handle sich um Manipulation, eine Schutzbehauptung. In einem anderen Fall argumentierte der Tesla-CEO Elon Musk im Rahmen eines Gerichtsverfahrens, er sei Opfer eines Deepfakes geworden. Das zur Debatte stehende Video war von der Gegenseite eingebracht worden als Beleg dafür, dass Musk öffentlich die Überlegenheit des Tesla-Autopiloten über menschliche Fahrer/innen in puncto Sicherheit herausgestellt habe (Busch/Oster 2023, S. 62; Jin/Levine 2023). In der Folge wurde Musk dazu aufgefordert, unter Eid auszusagen. Es ist nicht bekannt, ob es dazu kam, das Verfahren wurde mittlerweile außergerichtlich beigelegt (Vigliarolo 2024).

Auch die fortschreitende Digitalisierung von Gerichtsverfahren kann durch die Deepfaketechnologie kompromittiert werden. In Bezug auf die Digitalisierung der Gerichtsbarkeit ist zunächst

festzustellen, dass gemäß § 128a Zivilprozessordnung (ZPO)[38] zivilrechtliche Videoverhandlungen zulässig sind (Busch/Oster 2023, S. 64 f.).[39] Für Verfahrensbeteiligte besteht die Möglichkeit, sich auf Antrag per Videotechnik von einem anderen Ort zuschalten zu lassen, auch eine Beweisaufnahme per Bild- und Tonübertragung kann gestattet oder angeordnet werden (§ 284 Abs. 2 ZPO). Diese Entwicklung ist grundsätzlich als positiv zu bewerten, jedoch ist zugleich ein erhöhtes Risiko zu verzeichnen, dass es sich bei der per Video zugeschalteten Person um einen Deepfake handelt.[40] Derartige Befürchtungen wurden auch von Busch und Oster (2023, S. 65) sowie Steffes und Zichler (2024) artikuliert, da inzwischen Open-Source-Programme zur Echtzeiterstellung von Deepfakes existieren.[41] Laut Einschätzung von Busch und Oster (2023, S. 65) bestehen im Prozessrecht ausreichende Möglichkeiten, den Verfahrensbeteiligten Mechanismen zur Deepfakeerkennung zur Verfügung zu stellen, wobei aufgrund des technologischen Wettbewerbs und möglicher Detektionsfehler im Zweifel auch auf Instrumente der Prozessorganisation wie die Übertragung aus Gerichtsräumen zurückzugreifen ist. Eine weitere Handlungsoption zielt darauf ab, die Schwierigkeiten der Rechtsdurchsetzung für Betroffene zu verringern. So könnten in Fällen, die nicht strafrechtlich verfolgt werden, spezialisierte Opferberatungsstellen den Betroffenen Unterstützung geben, wenn sie dafür personell und finanziell ausgestattet werden (Karaboga et al. 2024, S. 333).

Durch die Verbreitung unwahrer Informationen mittels Deepfakes ergibt sich eine potenzielle Gefährdung demokratischer Prozesse. Zwar ist die Beeinflussung der politischen Willensbildung nicht strafbar bzw. sogar erwünscht, jedoch wird es problematisch, wenn die Grenze zur Gefährdung demokratischer Prozesse überschritten wird (Busch/Oster 2023, S. 66). Dies kann beispielsweise durch die Verbreitung von Deepfakes erfolgen, welche Falschaussagen politischer Mandatsträger zu bestimmten Themen beinhalten. Hier existiert eine Überschneidung mit dem Phänomenbereich der Desinformation, der bei größerer Verbreitung das Potenzial zur Beeinflussung von Wahlen zugeschrieben wird (Busch/Oster 2023, S. 66; Gehringer et al. 2024). Desinformation kommt in Form von echten Videos, die aus dem Kontext gerissen werden, oder einfach zu produzierenden Memes vor und wird bisher als größeres Problem als Deepfakes angesehen (Łabuz 2025; Weikmann/Lecheler 2024). Auch die Beobachtungen von Wahlkämpfen haben bisher trotz vieler Fälle politischer Deepfakes keine Anhaltspunkte für eine deutliche Beeinflussung durch Deepfakes belegt, dennoch warnen Expert/innen vor der Gefahr einer schleichenden Verunsicherung, was den Wahrheitsgehalt öffentlicher Informationen angeht (Birrer/Just 2024, S. 8; Łabuz/Nehring 2024, S. 465; Weikmann et al. 2025).

Als Gegenmaßnahmen wird vorgeschlagen, transparent in der Öffentlichkeit über Vorfälle von schädigenden Deepfakes zu berichten, um die Aufmerksamkeit für das Problem zu erhöhen und Schutzmaßnahmen zu verbessern (Karaboga et al. 2024, S. 338). Zudem könnten Kommunikationsakteure wie Medienorganisationen, aber auch politische Parteien, sich selbst zur Einhaltung bestehender professioneller (etwa journalistischer) Standards bzw. noch zu schaffender Regeln im Umgang mit KI-generierten Medieninhalten verpflichten. Auf diese Weise kann die Grenze zwischen kreativen und gefahrlosen und schädigenden Anwendungen deutlich gemacht werden (Karaboga et al. 2024, S. 339). Entsprechende Vereinbarungen wurden etwa bereits von Parteien in den Niederlanden, Dänemark und der Schweiz getroffen und sind in weiteren Ländern in Vorbereitung (EPTA 2024).

38 Zivilprozessordnung in der Fassung der Bekanntmachung vom 5. Dezember 2005, zuletzt am 24. Oktober 2024 geändert
39 Durch die Neufassung der ZPO nach dem Gesetz zur Förderung des Einsatzes von Videokonferenztechnik in der Zivilgerichtsbarkeit und den Fachgerichtsbarkeiten vom 15. Juli 2024 wurden diese Möglichkeiten noch erweitert.
40 In diesem Fall würde sich auch die Frage stellen, wie bzw. auf wen die Aussagedelikte gem. §§ 153 ff. StGB anwendbar sind.
41 beispielsweise https://github.com/iperov/DeepFaceLive (4.9.2025)

5.3.3 Stärkung von Medienkompetenz und interdisziplinäre Zusammenarbeit

Im Rahmen präventiver Maßnahmen wird vielfach der Begriff der AI Literacy bzw. Cyberliteracy verwendet. Ziel ist es, potenziellen Anwender/innen aus allen gesellschaftlichen Bereichen Wissen über den Phänomenbereich Deepfakes sowie über Cybersecurity zu vermitteln und sie zu befähigen, mit diesen Themen adäquat umzugehen, um sie vor den Folgen malevolenter Anwendungen (zum Beispiel Betrugsversuche) zu schützen (Karaboga et al. 2024, S. 334 f.). Als Beispiel kann der „Aktionsplan Deepfakes" der österreichischen Regierung (BMI 2022) angeführt werden, welcher die Förderung von auf Deepfakes und Desinformation ausgerichteten Maßnahmen zur Steigerung der Medienkompetenz sowie zur Sensibilisierung der Bevölkerung für einen kritischen Umgang mit Informationen, insbesondere online, vorsieht. Auch der BMBF-Aktionsplan Künstliche Intelligenz (BMBF 2023) beinhaltet eine KI-Kompetenzoffensive, wobei jedoch keine explizite Bezugnahme auf Deepfakes und deren Konsequenzen erfolgt. Die Vermittlung solcher Kompetenzen wird auch als eine Form des menschlichen Deep Learnings bezeichnet, das Fähigkeiten zum kritischen, analytischen und kreativen Denken umfasst (Erduran 2024).

Ein weiteres wesentliches Element ist die interdisziplinäre und insbesondere internationale Zusammenarbeit. Diese kann einen wichtigen Beitrag leisten – und zwar nicht nur zu Bildungszwecken. Auch die weitere Erforschung des Phänomenbereichs sowie die zeitnahe Reaktion auf Deepfakes, die gesellschaftliche Prozesse zu stören suchen, durch Krisenreaktionsteams (Karaboga et al. 2024, S. 338) könnten dadurch erheblich verbessert werden. Zu wenig ist bisher noch über die Auswirkungen von Deepfakes und selbst über die bloße Ausbreitung des Phänomens bekannt (Birrer/Just 2024, S. 5). Den Bedenken, dass Deepfakes einen starken Einfluss auf gesellschaftliche Prozesse haben können, setzen Kritiker entgegen, dass entsprechende Effekte häufig überschätzt werden und es angesichts der großen Menge an Kommunikationsinhalten schwer ist, gezielt Einfluss zu nehmen (Budak et al. 2024). Ebenfalls wird darauf hingewiesen, dass neben täuschenden Deepfakes mittlerweile eine große Zahl von offen als Deepfakes erkennbare Medieninhalte Verbreitung finden, denen in der wissenschaftlichen und öffentlichen Debatte jedoch noch kaum Aufmerksamkeit geschenkt wird (Öhman 2024, S. 13 f.). Nicht zuletzt ist bei allen Maßnahmen zu bedenken, dass Regulierungen speziell für Deepfakes in Konflikt mit dem Gebot der Technologieneutralität geraten könnten und zudem im Bereich des öffentlichen Diskurses einen hochsensiblen Bereich der Demokratie betreffen (Busch/Oster 2023, S. 66).

6 Literatur

6.1 In Auftrag gegebenes Gutachten und Kommentargutachten

Busch, C.; Oster, J. (2023): Rechtliche Aspekte von Deepfakes. Osnabrück/Ober-Olm

Härtlein, J. (2025): Anmerkungen zur TA-Kurzstudie: Rechtliche und gesellschaftliche Herausforderungen sowie Innovationspotenziale von Deepfakes. Würzburg

6.2 Weitere Literatur

Acharya, B. et al. (2024): ScamChatBot: An end-to-end analysis of fake account recovery on social media via chatbots. arXiv, https://doi.org/10.48550/arXiv.2412.15072

Adam, D. (2024): Misinformation might sway elections – but not in the way that you think. In: Nature 630(8018), S. 807–809, https://doi.org/10.1038/d41586-024-01696-z

Ajder, H. et al. (2019): The state of deepfakes: Landscape, threats, and impact. Deeptrace, Amsterdam

Allen, K.; Nehring, C. (2025): AI-generated disinformation in Europe and Africa. Use cases, solutions and transnational learning. Konrad-Adenauer-Stiftung, Johannesburg

Angwin, J. et al. (2016): Machine bias. ProPublica, 23.5.2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (12.12.2024)

AP (2024): An illness took away her voice. AI created a replica she carries in her phone. Associated Press, WUSF.org, https://www.wusf.org/health-news-florida/2024-05-19/an-illness-took-away-her-voice-ai-created-a-replica-she-carries-in-her-phone (27.1.2025)

Beckmann, A. et al. (2023): Fooling state-of-the-art deepfake detection with high-quality deepfakes. In: IH&MMSec'23. Proceedings of the 2023 ACM Workshop on Information Hiding and Multimedia Security, 28.-30.6, New York, S. 175–180, https://doi.org/10.1145/3577163.3595106

Belghaus, N.; Cohrs, C. (2024): Model über KI in der Sexindustrie: „Ich wollte mich neu erschaffen". taz.de, 25.3.2024, https://taz.de/Model-ueber-KI-in-der-Sexindustrie/!5997391/ (29.11.2024)

bellingcat (2025): Unmasking MrDeepFakes: Canadian pharmacist linked to world's most notorious deepfake porn site. https://www.bellingcat.com/news/2025/05/07/canadian-pharmacist-linked-to-worlds-most-notorious-deepfake-porn-site/ (15.6.2025)

Beuth, P. et al. (2024): Wie das Weltbild einer künstlichen Intelligenz entsteht. Spiegel Online, https://www.spiegel.de/netzwelt/web/laion-5b-so-entsteht-das-weltbild-einer-kuenstlichen-intelligenz-a-dea0157b-d364-4622-8d77-3a1cb6c4afd1 (10.1.2025)

Bieber, C. et al. (2024): KI im Superwahljahr 2024. Generative KI im Umfeld demokratischer Prozesse. Lernende Systeme – Die Plattform für Künstliche Intelligenz, München, https://doi.org/10.48669/PLS_2024-5

Birrer, A.; Just, N. (2024): What we know and don't know about deepfakes: An investigation into the state of the research and regulatory landscape. In: New Media & Society, https://doi.org/10.1177/14614448241253138

Block, M. (2023): Deepfakes und Recht. Einführung in den deutschen Rechtsrahmen für synthetische Medien. Berlin, https://doi.org/10.1007/978-3-662-67427-7

BMBF (2023): BMBF-Aktionsplan Künstliche Intelligenz. Neue Herausforderungen chancenorientiert angehen. Bundesministerium für Bildung und Forschung, Berlin

BMI (2022): Aktionsplan Deepfake. Bundesministerium für Inneres, https://www.bmi.gv.at/bmi_documents/2779.pdf (12.8.2024)

Bohacek, M.; Farid, H. (2024): The making of an AI news anchor – and its implications. In: Proceedings of the National Academy of Sciences 121(1), Art. e2315678121, https://doi.org/10.1073/pnas.2315678121

Bröckling, M.; Tanner, B. (2024): Deepfake-Pornos: Die Jagd nach den Tätern. ZDF, Mainz

BSI (o. J.): Deepfakes – Gefahren und Gegenmaßnahmen. Bundesamt für Sicherheit in der Informationstechnik, https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Informationen-und-Empfehlungen/Kuenstliche-Intelligenz/Deepfakes/deepfakes.html?nn=1009560 (31.7.2024)

Budak, C. et al. (2024): Misunderstanding the harms of online misinformation. In: Nature 630(8015), S. 45–53, https://doi.org/10.1038/s41586-024-07417-w

Bundesrat (2024): Entwurf eines Gesetzes zum strafrechtlichen Schutz von Persönlichkeitsrechten vor Deepfakes. Deutscher Bundestag, Drucksache 20/12605, Berlin

Bundesrat (2025): Entwurf eines Gesetzes zum strafrechtlichen Schutz von Persönlichkeitsrechten vor Deepfakes. Deutscher Bundestag, Drucksache 21/1383, Berlin

Bundesregierung (2024): Arbeit und Kooperationspartner der Zentralen Meldestelle für strafbare Inhalte im Internet im Bundeskriminalamt. Antwort der Bundesregierung auf die Kleine Anfrage der Abgeordneten Eugen Schmidt, Barbara Benkstein, Edgar Naujok, weiterer Abgeordneter und der Fraktion der AfD – Drucksache 20/10559 –. Deutscher Bundestag, Drucksache 20/10786, Berlin

Bundesverband Trans* e.V. et al. (2023): Stellungnahme zu den Eckpunkten für ein Gesetz gegen digitale Gewalt des Bundesministeriums der Justiz. https://hateaid.org/wp-content/uploads/2023/05/Finale-Stellungnahme-Eckpunkte-BMJ-Gesetz-gegen-digitale-Gewalt.pdf (17.12.2024)

Burghard, M.; Hornung, A. (2024): Was ist schon normal? In: kritische berichte – Zeitschrift für Kunst- und Kulturwissenschaften 52(2), S. 64–72, https://doi.org/10.11588/KB.2024.2.103988

Cahlan, S. (2020): How misinformation helped spark an attempted coup in Gabon. Washington Post, 13.2.2020, https://www.washingtonpost.com/politics/2020/02/13/how-sick-president-suspect-video-helped-sparked-an-attempted-coup-gabon/ (12.8.2024)

CeMAS (2025): Autoritäre Strategien im Netz: Analyse und Monitoring digitaler Risiken rund um die Bundestagswahl 2025. Center für Monitoring, Analyse und Strategie, Berlin

Chowdhury, R. (2024): AI-fuelled election campaigns are here – where are the rules? In: Nature 628 (8007),S. 237–237, https://doi.org/10.1038/d41586-024-00995-9

Compton, S.; Hamlyn, R. (2023): Alptraum Deepfake-Pornos. Another Body/ARTE/ZDF, Straßburg

Croitoru, F.-A. et al. (2023): Diffusion models in vision: A survey. In: IEEE Transactions on Pattern Analysis and Machine Intelligence 45(9), S. 10850–10869, https://doi.org/10.1109/TPAMI.2023.3261988

Croitoru, F.-A. et al. (2024): Deepfake media generation and detection in the generative AI era: a survey and outlook. arXiv, https://doi.org/10.48550/arXiv.2411.19537

DJB (2023): Policy Paper „Bekämpfung bildbasierter sexualisierte Gewalt". Deutscher Juristinnenbund e. V., https://www.djb.de/presse/pressemitteilungen/detail/st23-17 (5.6.2024)

DLM (2003): Drittes Strukturpapier zur Unterscheidung von Rundfunk und Mediendiensten. Direktorenkonferenz der Landesmedienanstalten, https://www.medienanstalt-nrw.de/fileadmin/user_upload/lfm-nrw/Die_LfM/Rechtsgrundlagen/Materialien/drittes_strukturpapier.pdf (3.12.2024)

Dornis, T. W. (2024): The training of generative AI is not text and data mining. SSRN, https://doi.org/10.2139/ssrn.4993782

dpa factchecking (2025): Deepfake-Video – Vatikan warnt vor gefälschter Papst-Botschaft bei Youtube. https://dpa-factchecking.com/germany/250528-99-994852/ (13.6.2025)

Eberl, A. et al. (2022): Using deepfakes for experiments in the social sciences – A pilot study. In: Frontiers in Sociology 7, https://doi.org/10.3389/fsoc.2022.907199

EK (2022): The strengthened code of practice on disinformation. Europäische Kommission, https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation (4.12.2024)

EP (2023): KI-Gesetz: erste Regulierung der künstlichen Intelligenz. Europäisches Parlament, https://www.europarl.europa.eu/topics/de/article/20230601STO93804/ki-gesetz-erste-regulierung-der-kunstlichen-intelligenz (22.7.2024)

EPTA (2024): Artificial Intelligence and Democracy. EPTA Report 2024. European Parliamentary Technology Assessment, Oslo

Erduran, S. (2024): Deepfakes and students' deep learning: A harmonious pair in science? In: Science 385(6712), Art. eadr8354, https://doi.org/10.1126/science.adr8354

Farid, H.; Schindler, H.-J. (2020): Deepfakes: eine Bedrohung für Demokratie und Gesellschaft. Konrad-Adenauer-Stiftung, Berlin

Fraga-Lamas, P.; Fernandez-Carames, T. M. (2020): Fake news, disinformation, and deepfakes: leveraging distributed ledger technologies and blockchain to combat digital deception and counterfeit reality. In: IT Professional 22(2), S. 53–59, https://doi.org/10.1109/MITP.2020.2977589

Frances-Wright, I. et al. (2024): Disconnected from reality: American voters grapple with AI and flawed OSINT strategies. ISD digital dispatches, 7.11.2024, https://www.isdglobal.org/digital_dispatches/disconnected-from-reality-american-voters-grapple-with-ai-and-flawed-osint-strategies/ (20.12.2024)

Freistaat Bayern (2024): Entwurf eines Gesetzes zum strafrechtlichen Schutz von Persönlichkeitsrechten vor Deepfakes. Gesetzesantrag des Freistaates Bayern. Bundesrat, Drucksache 222/24, Berlin

Gehringer, F. et al. (2024): Der Einfluss von Deep Fakes auf Wahlen. Konrad-Adenauer-Stiftung, Monitor Sicherheit, Berlin

Graham, M. M. (2024): Deepfakes: Federal and state regulation aims to curb a growing threat. Thomson Reuters, 26.6.2024, https://www.thomsonreuters.com/en-us/posts/government/deepfakes-federal-state-regulation/ (3.12.2024)

Han, C. et al. (2025): Characterizing the MrDeepFakes sexual deepfake marketplace. arXiv, https://doi.org/10.48550/arxiv.2410.11100

Hasan, H. R.; Salah, K. (2019): Combating deepfake videos using blockchain and smart contracts. In: IEEE Access 7, S. 41596–41606, https://doi.org/10.1109/ACCESS. 2019.2905689

Hawkins, W. et al. (2025): Deepfakes on demand: the rise of accessible non-consensual deepfake image generators. arXiv, https://doi.org/10.48550/arXiv.2505.03859

Heidari, A. et al. (2024): A novel blockchain-based deepfake detection method using federated and deep learning models. In: Cognitive Computation 16(3), S. 1073–1091, https://doi.org/10.1007/s12559-024-10255-7

von Hirschhausen, E. (2025): Mein Gesicht gehört mir. Süddeutsche Zeitung, 25.3.2025, https://www.sueddeutsche.de/medien/eckart-von-hirschhausen-meta-urteil-deepfakes-li.3225072 (13.6.2025)

Hoek, S. et al. (2024): Promising for patients or deeply disturbing? The ethical and legal aspects of deepfake therapy. In: Journal of Medical Ethics, https://doi.org/10.1136/jme-2024-109985

Hogertz, A. et al. (2024): Deepfakes: Eine juristische Einordnung. Bundesverband Digitale Wirtschaft, Berlin

Hoppenstedt, M. et al. (2024): Wo KI zur Waffe gegen Frauen und Mädchen wird. Spiegel Online, 9.12.2024, https://www.spiegel.de/netzwelt/web/deepfake-pornos-das-perfide-geschaeft-mit-gefaelschten-sexvideos-a-ddf22e79-c2ef-47d0-8b5d-6934e4397b17 (16.1.2025)

Hoppenstedt, M.; Milatz, M. (2025): Die Männer hinter den Fake-Pornos. Spiegel Online, 10.6.2025, https://www.spiegel.de/netzwelt/apps/deepfakes-die-maenner-hinter-den-fake-pornos-a-451ce839-50ca-466d-aa32-e7ccd42de198 (15.6.2025)

Hsu, T.; Lu, Y. (2024): No, that's not taylor swift peddling le creuset cookware. New York Times, 9.1.2024, https://www.nytimes.com/2024/01/09/technology/taylor-swift-le-creuset-ai-deepfake.html (28.11.2024)

Jin, H.; Levine, D. (2023): Elon, or deepfake? Musk must face questions on autopilot statements. Reuters, 27.4.2023, https://www.reuters.com/legal/elon-or-deepfake-musk-must-face-questions-autopilot-statements-2023-04-26/ (7.8.2024)

Justizministerkonferenz (2021): TOP II. 8 Bekämpfung von Gefahren durch sog. Deepfakes. 92. Konferenz der Justizministerinnen und Justizminister, Düsseldorf

Justizministerkonferenz (2023a): TOP II.2 Schutz der demokratischen Willensbildung – Wirksame Bekämpfung von „Fake News". 94. Konferenz der Justizministerinnen und Justizminister, Berlin

Justizministerkonferenz (2023b): TOP I.3 Digitale Gewalt effektiver bekämpfen – Zugang zum Recht erleichtern. 94. Konferenz der Justizministerinnen und Justizminister, Berlin

Justizministerkonferenz (2024): TOP I.14 Rechtssicherheit bei Deepfakes schaffen und Rechtsschutz verbessern. 95. Konferenz der Justizministerinnen und Justizminister, Hannover

Justizministerkonferenz (2025): TOP II.18 Bildbasierte sexualisierte Gewalt – Verbesserung des strafrechtlichen Schutzes. 96. Konferenz der Justizministerinnen und Justizminister, Bad Schandau

Kain, E. (2020): ‚Star wars' deepfake ‚fixes' that crazy ‚Mandalorian' season 2 finale cameo. Forbes, 23.12.2020, https://www.forbes.com/sites/erikkain/2020/12/23/star-wars-fan-fixes-that-crazy-mandalorian-cameo-with-a-stellar-deepfake/ (28.11.2024)

Karaboga, M. (2023): Die Regulierung von Deepfakes auf EU-Ebene: Überblick eines Flickenteppichs und Einordnung des Digital Services Act- und KI-Regulierungsvorschlags. In: Jaki, S.; Steiger, S. (Hg.): Digitale Hate Speech. Berlin, S. 197–220, https://doi.org/10.1007/978-3-662-65964-9_10

Karaboga, M. et al. (2024): Deepfakes und manipulierte Realitäten – Technologiefolgenabschätzung und Handlungsempfehlungen für die Schweiz. TA-SWISS, https://doi.org/10.5281/zenodo.11643643

Kleeman, J. (2024): She was accused of faking an incriminating video of teenage cheerleaders. She was arrested, outcast and condemned. The problem? Nothing was fake after all. The Guardian, 11.5.2024, https://www.theguardian.com/technology/article/2024/may/11/she-was-accused-of-faking-an-incriminating-video-of-teenage-cheerleaders-she-was-arrested-outcast-and-condemned-the-problem-nothing-was-fake-after-all (11.6.2024)

Kleemann, A. (2023): Deepfakes – Wenn wir unseren Augen und Ohren nicht mehr trauen können. Stiftung Wissenschaft und Politik, SWP-Aktuell 2023/A43, Berlin, https://doi.org/10.18449/2023A43

Koltai, K. (2024): Behind a Secretive Global Network of Non-Consensual Deepfake Pornography. bellingcat, 23.2.2024, https://www.bellingcat.com/news/2024/02/23/behind-a-secretive-global-network-of-non-consensual-deepfake-pornography/ (15.6.2025)

Köver, C. (2023): Deepfakes in Spanien: Gefälschte Nacktbilder von Mädchen sorgen für Aufschrei. Netzpolitik.org, 22.9.2023, https://netzpolitik.org/2023/deepfakes-in-spanien-gefaelschte-nacktbilder-von-maedchen-sorgen-fuer-aufschrei/ (18.1.2024)

Kraetzig, V. (2024): Werk im Werk. Zu den Schranken bearbeitender Vervielfältigungen. In: GRUR – Gewerblicher Rechtsschutz und Urheberrecht 126(4), S. 170–175

Kramer, B. (2024): More and more states are enacting laws addressing AI deepfakes. MultiState Insider, 5.4.2024, https://www.multistate.us/insider/2024/4/5/more-and-more-states-are-enacting-laws-addressing-ai-deepfakes (11.6.2024)

Kreutzer, T. (2022): Der Pastiche im Urheberrecht. Gutachten über eine urheberrechtsspezifische Definition des Pastiche-Begriffs nach § 51a UrhG. Gesellschaft für Freiheitsrechte, Berlin

Laaff, M. (2019): Hello, Adele – bist du's wirklich? Golem.de, 21.11.2019, https://www.golem.de/news/deep-fakes-hello-adele-bist-du-s-wirklich-1911-145136.html (12.8.2024)

Łabuz, M. (2024): Die Deepfake-Regulierung im AI-Act ist nur der Anfang. Tagesspiegel Background Digitalisierung & KI, 30.1.2024, https://background.tagesspiegel.de/digitalisierung/die-deepfake-regulierung-im-ai-act-ist-nur-der-anfang (13.6.2024)

Łabuz, M. (2025): Memetic Warfare – sind wir bereit für dieses Schlachtfeld? Tagesspiegel Background Digitalisierung & KI, 29.1.2025, https://background.tagesspiegel.de/digitalisierung-und-ki/briefing/memetic-warfare-sind-wir-bereit-fuer-dieses-schlachtfeld (13.6.2025)

Łabuz, M.; Nehring, C. (2024): On the way to deep fake democracy? Deep fakes in election campaigns in 2023. In: European Political Science 23(4), S. 454–473, https://doi.org/10.1057/s41304-024-00482-9

Lakatos, S. (2023): A revealing picture. AI-generated „undressing" images move from niche pornography discussion forums to a scaled and monetized online business. Graphika, o. O.

Lang, E. (2025): GTA pharmacist allegedly behind deepfake porn site now on leave, says hospital network CEO. CBC News, 13.5.2025, https://www.cbc.ca/news/canada/toronto/pharmacist-mrdeepfakes-david-do-off-work-oak-valley-health-1.7533549 (15.6.2025)

Lantwin, T. (2020): Strafrechtliche Bekämpfung missbräuchlicher Deep Fakes. Geltendes Recht und möglicher Regelungsbedarf. In: MMR – Zeitschrift für IT-Recht und Recht der Digitalisierung 23(2), S. 78–81

Lauber-Rönsberg, A. (2024): UrhG § 51a Karikatur, Parodie und Pastiche. In: Götting, H.-P. et al. (Hg.): BeckOK Urheberrecht. München

Lees, D. (2024): Deepfakes in documentary film production: images of deception in the representation of the real. In: Studies in Documentary Film 18(2), S. 108–129, https://doi.org/10.1080/17503280.2023.2284680

LTO (2024): Gericht verbietet Scholz-Deepfake zu AfD-Verbot. Legal Tribune Online, 25.2.2024, https://www.lto.de/recht/nachrichten/n/lg-berlin-ii-15o579-23-olaf-scholz-bundeskanzler-deepfake-afd-verbot-zentrum-politische-schoenheit (26.11.2024)

Lv, L. (2021): Smart watermark to defend against deepfake image manipulation. In: 2021 IEEE 6th International Conference on Computer and Communication Systems (ICCCS), Chengdu, 23.-26. April. IEEE Xplore, S. 380–384, https://doi.org/10.1109/ICCCS52626.2021.9449287

Maamar, N. (2023): Urheberrechtliche Fragen beim Einsatz von generativen KI-Systemen. In: Zeitschrift für Urheber- und Medienrecht 67(7), S. 481–491

Mai, K. T. et al. (2023): Warning: Humans cannot reliably detect speech deepfakes. In: PLOS ONE 18(8), Art. e0285333, https://doi.org/10.1371/journal.pone.0285333

Margerie, M. A. (2024): Deepfakes, Voiceclones & Co. KI-basiertes Täuschen durch Strafverfolgungsbehörden. Tagung „Zwischen Blackbox und Deepfakes – KI in der Strafverfolgung“, 10.10.2024, Bayreuth

Margerie, M. A.; Hartmann, M. (2025): Strafverfolgung mit Hilfe von Deepfake-Technologien – Realität oder Wunschdenken? In: Zeitschrift für Europäisches Daten- und Informationsrecht 1(2), S. 84–89

Masood, M. et al. (2023): Deepfakes generation and detection: state-of-the-art, open challenges, countermeasures, and way forward. In: Applied Intelligence 53(4), S. 3974–4026, https://doi.org/10.1007/s10489-022-03766-z

Medler, F. (2024): Betrugsmasche mit Deepfake-Promis. ZDFheute, 13.8.2024, https://www.zdf.de/nachrichten/panorama/kriminalitaet/betrug-promi-deepfake-werbevideos-geldanlage-100.html (28.11.2024)

Meineck, S. (2024): KI-Nacktbilder: Wie Online-Shops mit sexualisierten Deepfakes abkassieren. Netzpolitik.org, 27.11.2024, https://netzpolitik.org/2024/ki-nacktbilder-wie-online-shops-mit-sexualisierten-deepfakes-abkassieren/ (30.11.2024)

Merkert, P. (2022): Tauchgang ins Innere. Die Technik hinter KI-Bildgeneratoren. In: c’t – Magazin für Computer-Technik 25/2022, S. 80–84

Meywirth, C. et al. (2023): Cybercrime und Cyber Security Intelligence – kollaborative Ansätze gegen Cyber- und Computerkriminalität. In: Rüdiger, T.-G.; Bayerl, P. S. (Hg.): Handbuch Cyberkriminologie. Wiesbaden, https://doi.org/10.1007/978-3-658-35450-3_54-1

Mihailova, M. (2021): To dally with Dalí: Deepfake (inter)faces in the art museum. In: Convergence: The International Journal of Research into New Media Technologies 27(4), S. 882–898, https://doi.org/10.1177/13548565211029401

Mukherjee, S. (2024): Few AI deepfakes identified in EU elections, Microsoft president says. Reuters, 3.6.2024, https://www.reuters.com/technology/few-ai-deepfakes-identified-eu-elections-microsoft-president-says-2024-06-03/ (12.8.2024)

Naik, R.; Nushi, B. (2023): Social biases through the text-to-image generation lens. In: AIES ’23: Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society, Montreal, 8.–10. August. New York, S. 786–808, https://doi.org/10.1145/3600211.3604711

Neff, C. (2025): AI of dead Arizona road rage victim addresses killer in court. The Guardian, 7.5.2025, https://www.theguardian.com/us-news/2025/may/06/arizona-road-rage-victim-ai-chris-pelkey (13.6.2025)

NIST (2024): Reducing risks posed by synthetic content: An overview of technical approaches to digital content transparency. National Institute of Standards and Technology, AI 100-4, Gaithersburg, https://doi.org/10.6028/NIST.AI.100-4

Nordling, L. (2024): Scientists are falling victim to deepfake AI video scams – here’s how to fight back. Nature, https://doi.org/10.1038/d41586-024-02521-3

Öhman, C. (2024): Who are the targets of deepfakes? Evidence from flagged videos on tiktok and youtube. SSRN, https://doi.org/10.2139/ssrn.4759677

Pawelec, M.; Bieß, C. (2021): Deepfakes. Technikfolgen und Regulierungsfragen aus ethischer und sozialwissenschaftlicher Perspektive. Baden-Baden

Rüdiger, T.-G.; Bayerl, P. S. (2020): Cyberkriminologie: Braucht die Kriminologie ein digitales Upgrade? In: Rüdiger, T.-G.; Bayerl, P. S. (Hg.): Cyberkriminologie. Wiesbaden, S. 3–12, https://doi.org/10.1007/978-3-658-28507-4_1

Rudl, T. (2025): Gegen Deepfakes: US-Kongress verabschiedet Take It Down Act. netzpolitik.org, 30.4.2025, https://netzpolitik.org/2025/gegen-deepfakes-us-kongress-verabschiedet-take-it-down-act/ (13.6.2025)

Saathoff, C. (2025): Ausschreitungen in Indonesien. Wie gefälschte Videos Proteste befeuern. Tagesschau.de, 2.9.2025, https://www.tagesschau.de/faktenfinder/proteste-indonesien-fakenews-100.html (10.9.2025)

Salemi, S.; Steffes, B. (2022): Deepfakes im VideoIdent-Verfahren: (fehlende) Straf- und zivilrechtliche Konsequenzen für Täter. In: Demmler, D. et al. (Hg.): INFORMATIK 2022. Hamburg, 26.–30. September. Proceedings. Bonn, S. 665–680, https://doi.org/10.18420/inf2022_53

Schueler, M. et al. (2024): Artificial elections: Exposing the use of generative AI imagery in the political campaigns of the 2024 French elections. AI forensics, 4.7.2024, https://aiforensics.org/work/french-elections-2024 (17.12.2024)

Schwieter, C.; Gandhi, M. (2024): Vilify, ridicule, disinform. Political communication and media trust in the age of generative AI. Friedrich-Naumann-Stiftung für die Freiheit, Berlin

Security Hero (2023): 2023 State of deepfakes: Realities, threats, and impact. https://www.securityhero.io/state-of-deepfakes/ (3.12.2024)

Sensity B. V. (2024): The state of deepfakes 2024. https://sensity.ai/reports/ (3.12.2024)

Seow, J. W. et al. (2022): A comprehensive overview of Deepfake: Generation, detection, datasets, and opportunities. In: Neurocomputing 513, S. 351–371, https://doi.org/10.1016/j.neucom.2022.09.135

Siggelkow, C.; Reveland, P. (2023): Künstliche Intelligenz: Falsche Tagesschau-Audiodateien im Umlauf. Tagesschau.de, 13.11.2023, https://www.tagesschau.de/faktenfinder/tagesschau-audio-fakes-100.html (3.12.2024)

Specht-Riemenschneider, L. (2018): KUG § 22 [Recht am eigenen Bilde]. In: Dreier, T.; Schulze, G. (Hg.): Urheberrechtsgesetz. Kommentar. München

Spiegel Online (2024): Stimme des Schulleiters gefälscht: Sportlehrer in Maryland festgenommen. 26.4.2024, https://www.spiegel.de/netzwelt/apps/ki-stimmengenerator-sportlehrer-soll-rassistische-bemerkungen-gefaelscht-haben-a-f267a6f9-a236-4c0a-b3e3-20f3d624235e (29.11.2024)

Steffes, B.; Zichler, A. (2024): Deepfakes in Videoverhandlungen vor Gericht: Herausforderungen der menschlichen und maschinellen Erkennung. In: Datenschutz und Datensicherheit 48(3), S. 158–163, https://doi.org/10.1007/s11623-023-1899-1

STOA (2021): Tackling deepfakes in European policy. European Parliamentary Research Service, Scientific Foresight Unit, Luxemburg

Stockwell, S. (2024): AI-enabled influence operations: Threat analysis of the 2024 UK and european elections. Centre for Emerging Technology and Security, CETaS briefing papers, London

Szeto, E. et al. (2025): This Canadian pharmacist is key figure behind world's most notorious deepfake porn site. CBC News, 7.5.2025, https://www.cbc.ca/news/canada/mrdeepfakes-porn-website-key-figure-1.7527626 (15.6.2025)

TAB (2020): Mögliche Diskriminierung durch algorithmische Entscheidungssysteme und maschinelles Lernen – ein Überblick. (Autor/innen: Kolleck, A.; Orwat, C.) Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag, TAB-Hintergrundpapier 24, Berlin, https://doi.org/10.5445/IR/1000127166

TAB (2022a): Algorithmen in digitalen Medien und ihr Einfluss auf die Meinungsbildung. (Autor/innen: Oertel, B. et al.) Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag, TAB-Arbeitsbericht 204, Berlin, https://doi.org/10.5445/IR/1000154065

TAB (2022b): Data-Mining – gesellschaftspolitische und rechtliche Herausforderungen. (Autorin: Gerlinger, K.) Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag, TAB-Arbeitsbericht 203, Berlin, https://doi.org/10.5445/IR/1000156297

TAB (2024): Künstliche Intelligenz in der Kreativwirtschaft. (Autorinnen: Kind, S.; Hille, M.) Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag, TAB-Themenkurzprofil 73, Berlin, https://doi.org/10.5445/IR/1000172202

Tagesschau.de (2024): Europäische Datenschützer reichen Beschwerde gegen OpenAI ein. 29.4.2024, https://www.tagesschau.de/wirtschaft/verbraucher/openai-chatgpt-datenschutz-beschwerde-100.html (22.7.2024)

Taipei Times (2023): Bill to curb deepfake pornography clears legislature. 8.1.2023, https://www.taipeitimes.com/News/front/archives/2023/01/08/2003792190 (13.6.2025)

Thiel, M. (2021): „Deepfakes" – Sehen heißt glauben? Gefahren, gesetzgeberischer Handlungsbedarf, Konsequenzen für die biometrische Gesichtserkennung. In: Zeitschrift für Rechtspolitik 54(7), S. 202–204

theally (2025): Policy Update: Removal of Real-Person Likeness Content. Civitai, https://civitai.com/articles/15022/policy-update-removal-of-real-person-likeness-content (15.6.2025)

Vaccari, C.; Chadwick, A. (2020): Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. In: Social Media + Society 6(1), https://doi.org/10.1177/2056305120903408

Van Minnen, A. et al. (2022): Initial development of perpetrator confrontation using deepfake technology in victims with sexual violence-related PTSD and moral injury. In: Frontiers in Psychiatry 13, Art. 882957, https://doi.org/10.3389/fpsyt.2022.882957

Vanberghen, C. (2024): The AI Act vs. deepfakes: A step forward, but is it enough? Euractiv, 26.2.2024, https://www.euractiv.com/section/artificial-intelligence/opinion/the-ai-act-vs-deepfakes-a-step-forward-but-is-it-enough/ (13.6.2024)

Vigliarolo, B. (2024): Tesla settles Apple engineer Autopilot fatality case. The Register, 9.4.2024, https://www.theregister.com/2024/04/09/tesla_settles_autopilot_case/ (20.12.2024)

Weikmann, T. et al. (2025): After deception: How falling for a deepfake affects the way we see, hear, and experience media. In: The International Journal of Press/Politics 30(1), S. 187–210, https://doi.org/10.1177/19401612241233539

Weikmann, T.; Lecheler, S. (2024): Cutting through the hype: Understanding the implications of deepfakes for the fact-checking actor-network. In: Digital Journalism 12(10), S. 1505–1522, https://doi.org/10.1080/21670811.2023.2194665

Westerlund, M. (2019): The Emergence of Deepfake Technology: A Review. In: Technology Innovation Management Review 9(11), S. 39–52, https://doi.org/10.22215/timreview/1282

Whittaker, L. et al. (2023): Mapping the deepfake landscape for innovation: A multidisciplinary systematic review and future research agenda. In: Technovation 125, Art. 102784, https://doi.org/10.1016/j.technovation.2023.102784

Wiederhold, B. K. (2021): Can Deepfakes Improve Therapy? In: Cyberpsychology, Behavior, and Social Networking 24(3), S. 147–148, https://doi.org/10.1089/cyber.2021.29209.editorial

Wilkens, A. (2022): Niederländische Polizei sucht mit Deepfake-Video eines Mordopfers nach Zeugen. Heise online, 24.5.2022, https://www.heise.de/news/Niederlaendische-Polizei-sucht-mit-Deepfake-Video-eines-Mordopfers-nach-Zeugen-7121459.html (29.11.2024)

Wolf, M. (2024): Als Sean Connery an der Redaktionskonferenz teilnahm. Golem.de, 22.8.2024, https://www.golem.de/news/deepfakes-als-sean-connery-an-der-redaktionskonferenz-teilnahm-2408-188172.html (26.11.2024)

Yang, J. et al. (2023): Model-agnostic method: Exposing deepfake using pixel-wise spatial and temporal fingerprints. In: IEEE Transactions on Big Data 9(6), S. 1496–1509, https://doi.org/10.1109/TBDATA.2023.3284272

7 Anhang

7.1 Abbildungen

7.2 Kästen

Herausgeber
Büro für Technikfolgen-Abschätzung
beim Deutschen Bundestag
Neue Schönhauser Straße 10
10178 Berlin

Telefon: +49 30 28491-0
E-Mail: buero@tab-beim-bundestag.de
www.tab-beim-bundestag.de

2025

Bildnachweis
terovesalainen/AdobeStock (S. 1)

ISSN: 2944-8077
DOI: 10.5445/IR/1000190774

Das Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag (TAB) berät das Parlament und seine Ausschüsse in Fragen des wissenschaftlich-technischen Wandels. Das TAB wird seit 1990 vom Institut für Technikfolgenabschätzung und Systemanalyse (ITAS) des Karlsruher Instituts für Technologie (KIT) betrieben. Grundlage ist ein Vertrag mit dem Deutschen Bundestag. Hierbei kooperiert es seit September 2013 mit dem IZT – Institut für Zukunftsstudien und Technologiebewertung gGmbH sowie dem Institut für Innovation und Technik (iit) in der VDI/VDE Innovation + Technik GmbH.