

Deepfakes als Beweismittel: Zur Notwendigkeit der Authentizitätsprüfung von Bild-, Audio- und Videodaten

Das digitale Rig als Authentifizierungsmethode zur Erkennung von synthetischen Bild- und Videodaten

Stefan Kellermann ¹, Dirk Labudde ²

Abstract: Mit der zunehmenden Leistungsfähigkeit von künstlicher Intelligenz werden die Anforderungen an eine Authentizitätsprüfung von Bild-, Audio-, und Videodaten unter anderem bei der Verwendung der Inhalte als Beweismittel für die Strafverfolgungsbehörden steigen. Dass der Gesetzgeber darüber hinaus exemplarisch bei der Protokollierung von Strafermittlungsverfahren auf Videodaten (Videovernehmungen) oder Audiodaten (Aufzeichnung von Gerichtsverhandlungen) auf die Integrität dieser setzt, zeigen die Gesetzesänderungen der letzten Jahre. Bei zunehmender Verbreitung von synthetischen (KI-generierten) Inhalten muss es für forensische Untersuchungen verlässlich möglich sein, synthetische Inhalte zu erkennen, um die Authentizität der Daten beweissicher bewerten zu können. In dieser Arbeit wird auf verschiedene Techniken der Herstellung von Deepfakes und grundsätzliche Ansatzpunkte der Detektion eingegangen. Das digitale Rig wird als Möglichkeit zur Authentizitätsprüfung bei body-puppetry-Imitationen vorgestellt und als KI-Anwendung selbst in den EU-gesetzlichen Kontext des EU AI Actes gesetzt.


Keywords: Authentizität, Deepfakes, Künstliche Intelligenz, Beweismittel

1 Synthetische Bild-, Audio- und Videodaten – eine Gefahr für den Rechtsstaat

Synthetische Daten sind im Kontext von Wahlen geeignet, die Integrität von Personen und Institutionen anzugreifen und damit den demokratischen Meinungsbildungsprozess negativ zu beeinflussen[Br24]. Dieser negative Einfluss lässt sich auch auf den Prozess der Beweisaufnahme zur Urteilsfindung übertragen, insbesondere wenn synthetisch generierte Inhalte als authentisch bewertet werden. Der Erkennung von synthetisch generierten Inhalten kommt demnach eine Schlüsselrolle zu, um Wahrheit von Manipulation bzw. Desinformation trennen zu können. In Bezug auf Bilddaten stellt Krupicka fest: „Die generelle Glaubwürdigkeit von Fotografien, ihr Anspruch, die Realität visuell exakt darzustellen, eventuell sogar ihre Eignung als Beweismittel in Gerichtsverfahren, wird durch die KI-basierten Bildgeneratoren nachhaltig in Frage

¹ Universität Greifswald, Ernst-Lohmeyer-Platz 1, Greifswald, 17489,

stefan.kellermann@stud.uni-greifswald.de,  <https://orcid.org/0009-0002-0536-0673>.

² Hochschule Mittweida, Technikumplatz 17, Mittweida, 09648, labudde@hs-mittweida.de,  <https://orcid.org/0000-0003-0466-0017>.

gestellt.“ [Kr23] Auch das Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag kommt bereits im Jahr 2019 bezüglich der Folgen von Deepfakes zu der Einschätzung, dass „es nur eine Frage der Zeit und des technischen Fortschritts [ist], bis kein Unterschied zur Realität mehr feststellbar ist.“ [Bo19] Digitale Inhalte sind aus verschiedenen Perspektiven zunehmend fester und gesetzlich forcierter Bestandteil im Strafverfahren.

Die Feststellung der Echtheit von digitalen Inhalten als Beweismittel ist von entscheidender Bedeutung, weil der kriminalistische Erkenntnisprozess auf die Wahrheitsfindung fokussiert, um kriminalistisch relevante Sachverhalte aufzuklären. [ACR22] Die technisch voranschreitenden Möglichkeiten zur Generierung synthetischer Bild-, Audio- und Videodaten stellen die Strafverfolgungsbehörden daher zunehmend vor neue Herausforderungen. Ist in der Zukunft die Echtheit zum Beispiel eines Videos nicht mehr von synthetisch erstellten Inhalten abzugrenzen, stellt dies die Verwendung des Datenmaterials als Beweismittel grundsätzlich in Frage.

Aber nicht nur die Verwendung als Beweis für die Tatbeteiligung kommen Bild-, Audio- und Videodaten in Betracht. Die vom Gesetzgeber in den letzten Jahren umgesetzten und forcierten Änderungen der Strafprozessordnung bezüglich der Aufzeichnung von Beschuldigtenvernehmungen oder dem aktuellen Vorhaben, die Protokollierung von erstinstanzlichen Hauptverhandlungen in Land- und Oberlandesgerichten mithilfe einer Tonaufnahme zeigen, dass selbiger die Digitalisierung des Prozesses zur Urteilsfindung transparenter gestalten und die Möglichkeiten zur retrograden Bewertung des Prozesses zur Urteilsfindung steigern will. [De23] Wie Steffes und Zichler bereits dargestellt haben, ist in der Zukunft in Videoverhandlung die Verwendung von in Echtzeit erstellten synthetischen Videoinhalten denkbar. [SZ24] Damit geht ein Vertrauen in die digitalen Aufzeichnungen des Prozesses einher, dass es durch geeignete technische und organisatorische Maßnahmen abzusichern gilt, um die Integrität der Aufzeichnungen und Übertragungen zu gewährleisten.

2 Erkennung von synthetisch generierten Bild-, Audio- und Videodaten

2.1 Wesentliche Erscheinungsformen von synthetisch generierten Bild-, Audio- und Videodaten

Um in der Folge die Detektionsansätze aufzuzeigen, werden zunächst die vielseitigen Erscheinungsformen synthetischer Bild-, Audio- und Videodaten überblicksartig nach Kietzmann et al. [Ki20] mit Ergänzungen nach Garrido et al. [Ga14] dargestellt.

Art des Deepfakes	Vorgehen	Beschreibung
Foto- Deepfake	Face and Body Swapping	Vornehmen von Änderungen an einem Gesicht, das Ersetzen oder Verschmelzen des Gesichts (oder des Körpers) durch das Gesicht (oder den Körper) einer anderen Person
Audio-Deepfake	Voice-Swapping	Ändern einer Stimme oder Nachahmung der Stimme einer anderen Person
	Text-to-Speech	Veränderung der Tonspur einer Aufnahme durch Eingabe eines neuen Textes
Video-Deepfakes	Face Swapping	Ersetzen des Gesichts einer Person in einem Video durch das Gesicht einer anderen Person
	Face-Morphing	Gesicht verwandelt sich in ein anderes Gesicht durch einen nahtlosen Übergang
	Facial Reenactment[Ga14]	Ersetzen eines Gesichtes oder der Bewegung eines Gesichtes in einem vorhandenen Zielbild oder Video durch ein anderes Gesicht oder eine andere Darstellung aus einem Quellbild oder -video
	Full body puppetry	Übertragung der Bewegungen vom Körper einer Person auf den einer anderen Person
Audio und Video-Deepfakes	Lip-syncing	Veränderung der Mundbewegungen und des gesprochenen Wortes in

einem Video in dem jemand spricht.

Tabelle 1: Überblick der Arten von Deepfake

2.2 Ansatzpunkte zur Erkennung von synthetischen Bild-, Audio- und Videodaten

Die Verfahren zur Erkennung von synthetischen Bild-, Audio- und Videodaten zielen darauf ab, die Fehler der Fälschungen oder fehlende Merkmale eines lebenden Menschen zu erkennen. Ein Grundproblem der automatisierten Detektion von manipulierten Medieninhalten ist die mangelnde Generalisierbarkeit der Detektionsmethoden. Diese Methoden basieren zum Teil auf trainierten Datenbeständen, die als Gesamtheit eine Spezifik abbilden. Diese Spezifik führt jedoch dazu, dass die Erkennungsmethoden nur eingeschränkte Wirkung entfalten. [Bu21] Die Erkennung von manipulierten Medieninhalten kann daher nur aus einem Vorgehen aus verschiedenen technischen Analyseverfahren sowie der kritischen Überprüfung durch die vorliegende Beweislage insgesamt bestehen.

So zählt das BSI u.a. beispielsweise bei Gesichtsmanipulationen folgende Fehler auf:

- Sichtbare Artefakte bei einem Face-Swapping-Verfahren an der Naht rund um das Gesicht sowie Wechsel der Hautfarbe und -textur an diesem Übergang.
- Scharfe Konturen an den Zähnen oder Augen wirken verwaschen.
- Begrenzte Mimik, unstimmige Belichtung durch fehlerhafte Darstellung von Gesichtsausdrücken oder Beleuchtungssituationen

Bei synthetischen Stimmen treten u.a. laut dem BSI noch folgende Fehler auf:

- Metallischer Klang
- Falsche Aussprache, insbesondere beim Wechsel der Sprache, z.B. vom Deutschen ins Englische
- Monotone Sprachausgabe bei der Betonung der Worte

Die KI-basierten Verfahren zur Erkennung von synthetischen Bild-, Audio- und Videodaten sind darauf ausgerichtet, unter anderem die o.g. Fehler zu erkennen. Zudem besteht ein Wettlauf zwischen sich qualitativ immer „besser?“ werdenden, also für die betrachtende Person weniger gut erkennbaren, synthetischen Inhalten und den Erkennungsmethoden. Aufgrund der bereits unter Punkt 1 dargestellten Bedeutung der Erkennung von synthetischen Inhalten stehen die Erkennungsmethoden, wie nachfolgend beispielhaft aufgeführt, derzeit im Forschungsfokus.

Das Ziel des Forschungsprojektes FAKE-ID des BMBF ist es unter anderem, für den Menschen unsichtbare Merkmale von Fake-Generatoren zu erkennen. [Di24] Zudem werden innerhalb des Projektes auch weitere Detektionsverfahren entwickeln, wie die Pulsschlagerkennung bei Bildaufnahmen. Einen ähnlichen Ansatz verfolgt auch die Software von Intel, die den „Blutfluss“ in Videopixeln zur Lebenderkennung nutzt. [In22] Das BMBF-Forschungsprojekt „SpeechTrust+“ ist darauf ausgerichtet, „mittels KI-basierter Technologien Sprachmanipulationen zu erkennen und Beweismittel auf Authentizität hin zu prüfen.“ [Bu22] Das emotionale Zusammenspiel zwischen Sprache und Video und Emotionen ist ebenfalls ein Ansatz, um synthetische Audio- und Videodaten zu erkennen. [Ho21]

Zu den Merkmalen, die zur Erkennung von synthetischen Inhalten genutzt werden könnten, gehören auch anthropologische Maße. Diese Merkmale werden durch das sogenannte digitale Rig abstrahiert. Das digitale Rig (anthropologischer Vergleich) könnte diese o.g. unterschiedlichen Ansätze, insbesondere bezüglich der Identifizierung von Personen anhand des Gangbildes, ergänzen. Damit geht inhärent auch die Erweiterung der Möglichkeiten zur Integritätsprüfung von Bild- und Videomaterial einher.

3 Das digitale Rig als Authentifizierungsmethode zur Erkennung von synthetischen Bild- und Videodaten

3.1 Anwendungsmöglichkeiten des digitalen Rig anhand von ausgewählten Szenarien

Bei dem digitalen Rig handelt es sich um ein passives biometrisches Merkmal, das bereits erfolgreich zur Identifikation von Personen innerhalb von Strafprozessen zur Anwendung gekommen ist. [La23] Die theoretischen und verfahrensbasierten Grundlagen wurden bereits in verschiedenen Werken umfangreich aufgezeigt. [PRL23], [Be22], [La23] Ziel des Verfahrens ist es, das Rig einer bekannten tatverdächtigen Person mit einem Rig einer im Videomaterial dargestellten Person abzugleichen und zu bewerten. Mit dem Vergleich kann ein Gewissheitsgrad angegeben werden.

Die weiteren Anwendungsmöglichkeiten des digitalen Rigs sollen in folgenden Szenarien kurz beschreiben werden.

Szenario 1:

Das Rig zur Authentifizierung von Videomaterial

Wenn in einem Strafverfahren Bild- und/oder Videomaterial als Beweismittel vorgelegt wird, könnten sowohl Zeugen als auch Beschuldigte geneigt sein, die Echtheit des Datenmaterials aus verschiedenen Gründen in Frage zu stellen. [PB21] Exemplarisch, weil eine oder mehrere Parteien vorgeben, dass es sich schlicht um eine andere Person in dem

Video handelt. Das digitale Rig könnte dabei helfen, die Echtheit zu bestätigen, indem die betreffenden Personen eindeutig identifiziert werden.

Szenario 2:

Aufklärung von Begehungsformen des Identitätsdiebstahls mithilfe des digitalen Rig

Wie Pawelec et al. ausführen, sind Politiker:innen bereits häufig von synthetischen Bildern und Videos, mit dem Ziel betroffen, sie öffentlich weitreichend zu diskreditieren.[PB21] Das Rig könnte dazu genutzt werden, Opfer von Identitätsdiebstählen als solche zu identifizieren, indem das synthetische Bild- und/oder Videomaterial eindeutig als solches erkannt wird. Insbesondere im Kontext von Wahlen könnten diese gefälschten Videos den demokratischen Meinungsbildungsprozess negativ beeinflussen. Eine schnelle Aufklärung dieser Sachverhalte ist daher für die Reputation der betroffenen Person und den Prozess der Wahl unabdingbar.

Der Vorteil der Methode liegt insbesondere darin, dass die Person auf dem Videomaterial nicht mit dem eigenen Gesicht zu erkennen sein muss, um den anthropologischen Vergleich durchzuführen. Gleichwohl bedarf es der Einzelfallprüfung, ob genügend Vergleichsmaterial, resultierend aus dem in Frage stehenden Bild-/Videomaterial, für einen Abgleich mit dem Rig der betreffenden Person zur Verfügung steht.

Erkennung von Deepfakes für die Szenarien 1 und 2 mithilfe des digitalen Rig

Neben den unter Punkt 2.2 dargestellten Ansatzpunkten zur Erkennung von Deepfakes könnte das digitale Rig ein zusätzliches Prüfmerkmal bilden, um die kriminalistischen Ermittlungen im Rahmen der Szenarien 1 und 2 zu unterstützen. Dies könnte in dergestalt erfolgen, als das von den beteiligten Personen (Szenario 1) oder dem Opfer (Szenario 2) das digitale Rig als Vergleichsmaterial erstellt und mit den im Videomaterial dargestellten Szenen abgeglichen wird. Voraussetzung zur Nutzung des anthropologischen Vergleichs ist, neben der Erstellung des digitalen Rig, die digitale dreidimensionale Vermessung des Bildausschnittes des zu überprüfenden Bild-/Videomaterials.

Jede im Strafverfahren angewandte Methodik bedarf der gesetzlichen Ermächtigungsgrundlage. Nachfolgend schließen sich daher Ausführungen zur rechtlichen Betrachtung der Rig-Methode an.

3.2 Rechtliche Betrachtung der Rig-Methode bei der Anwendung im Strafverfahren

Zur rechtlichen Einordnung der Methode in den strafprozessualen Kontext wird die Einordnung nach Engelhard et al. herangezogen. [En24] Demnach kann die Methode „in den derzeit geltenden Kanon strafprozessrechtlicher Normen eingebettet werden“. Die zunächst vorbereitend notwendige Rig-Ableitung des/der Tatverdächtigen könne als digitale Spurensicherungs- und Spurenauswertungsmaßnahme auf die

Ermittlungsgeneralklausel gem. §§ 161, 163 StPO gestützt werden, da es sich um Spuren einer Straftat handelt.

Hingegen kann die Maßnahme der Erfassung und Erstellung des Rigs eines/einer Beschuldigten, also die photogrammetrische Vermessung, auf die Eingriffsnormen der erkennungsdienstlichen Behandlung, hier § 81b Abs. 1 StPO, gestützt werden.

Der anschließend mögliche Abgleich zwischen dem Rig der Tatortaufnahme und dem erstellten Rig aus der erkennungsdienstlichen Behandlung kann auf Grundlage des maschinellen Abgleichs im Sinne des § 98c StPO erfolgen.

Ergänzend zu den Betrachtungen von Engelhardt et al. wird ausgeführt, dass sofern im Sinne des Szenarios 2 Zeugen bzw. Opfer von der Maßnahme betroffen sind, diese nicht auf den § 81b StPO gestützt werden können. In diesem Fällen darf die Maßnahme nicht gegen den Willen der Betroffenen durchgeführt werden. Die Einwilligung zur Maßnahme ist einzuholen.

4 Missbrauchspotenzial und Grenzen der Rig-Methode bei der Erkennung von synthetischen Bild- und Videodaten

4.1 Spezialfall: Das Rig als Full body puppetry im Deepfake

Ein Manipulationsgebiet sind die body-puppetry-Imitationen (siehe auch 2.1), bei denen Körperbewegungen analysiert, imitiert und transferiert werden. Gerade in einer Zeit, in welcher wir bewegten Bildern besonders vertrauen, ist diese Manipulationsform sehr gefährlich.

Sogenanntes Body-Puppetry überträgt bestimmte Bewegungen auf bestehendes Videomaterial einer anderen Person. Da die KI hierbei in der Lage sein muss, den Hintergrund authentisch zu rekonstruieren, gilt diese Form des Deepfakings heute noch als am kompliziertesten. Jedoch zeigen aktuelle KI-Frameworks, dass die Generierung solcher Body-Puppetry immer effektiver erfolgen können.

Im Prozess der Erstellung von Body-Puppetry werden die Bewegungen durch Posenerkennung übertragen. Für den menschlichen Betrachter ist es unmöglich die Besonderheiten eines individuellen Rigs einer Person vom Rig einer anderen Person zu unterscheiden. Das digitale Rig, welches automatisch detektiert werden kann, weist eine hohe Individualität auf. [Be22] Die sogenannte Duplikationswahrscheinlichkeit beträgt auf gemessenen anthropologischen Daten zwischen 10^{-8} und 10^{-15} . [HHL22], [HHL23] Wissenschaftlich muss jedoch festgestellt werden, was die Koordinatentransformation während des Body-Puppetry an Genauigkeit liefert.

Gerade im Zusammenhang mit Social Engineering besteht die besondere Gefahr, wenn diese Form als Angriffsvektoren verwendet werden.

4.2 Missbrauchspotenzial und Grenzen der Rig-Methode

Wie unter Punkt 4.1 bereits dargestellt, besteht theoretisch die Möglichkeit im Rahmen des Social Engineering das Rig einer Person abzuleiten, indem z.B. der Ort eines Presseauftritts des Opfers digital vermessen wird und anschließend mit dem durch das Opfer oder Medienvertreter:innen veröffentlichten Videomaterial auf dem selbigen Opfer abgebildet ist, verglichen wird. Diese Gefahr besteht insbesondere für Personen des öffentlichen Lebens. Dieser Umstand zeigt auch gleichzeitig die Grenzen der Rig-Methode auf. Wenn das Rig einer Person den Täter:innen bekannt ist, wäre es denkbar, dass dies in die Erstellung eines Deepfakes einfließt und somit als Ansatzpunkt zur Erkennung unwirksam wird. Hieraus ergibt sich auch, dass das digitale Rig einer Person als personenbezogener Datensatz besonders schützenswert ist, um einem Missbrauch der Daten entgegenzuwirken. Diesem Umstand geschuldet, erfolgt eine erste Einordnung des digitalen Rig in den Kontext der Verordnung über künstliche Intelligenz. [ER24]

5 Das digitale Rig im Kontext des Gesetzes über künstliche Intelligenz (EU AI Act)

5.1 Kernregelungen des EU AI Act

Am 12.07.2024 wurde die Verordnung (EU) 2024/1689 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz nach Bewilligung des europäischen Parlamentes und des Rates der Europäischen Union im Amtsblatt der Europäischen Union veröffentlicht. [ER24] Dies bedeutet, dass das weltweit erste Gesetz zur Regelung von künstlicher Intelligenz, das sog. KI-Gesetz, auch KI-Verordnung oder EU AI Act genannt, für den Geltungsbereich der EU verabschiedet wurde und nun sukzessive in Kraft treten wird. Zentral dabei ist, dass ein risikobasierter Ansatz zur Einordnung von KI-Systemen eingeführt wurde, der je nach Höhe des Risikos unterschiedliche Maßnahmen fordert, unter denen u.a. der Betrieb von KI-Systemen in der EU erlaubt ist. [Eu24] Zudem wurden verbotene KI-Praktiken, u.a ein Verbot des Einsatzes von KI, um soziales Verhalten zu bewerten, Menschen zu beeinflussen oder ihre Schwächen auszunutzen, eingeführt. [Eu24] Die Nutzung von biometrischen Fernidentifizierungssystemen durch Strafverfolgungsbehörden wurde beschränkt. [Eu24] Aus Gründen des Umfangs wird auf weitere Darstellungen zu allgemeinen Regelungen der Verordnung über künstliche Intelligenz, auch EU AI Act genannt, verzichtet. Um das digitale Rig in die Regelungen zu subsumieren, werden zuvor notwendige Begriffsbestimmungen des EU AI Acts aufgeführt.

5.2 Begriffsbestimmungen gem. Artikel 3 des EU AI Act [ER24]

1. KI-System

„Ein „KI-System“ ist ein maschinengestütztes System, das für einen in unterschiedlichem Grade autonomen Betrieb ausgelegt ist und das nach seiner Betriebsaufnahme anpassungsfähig sein kann und das aus den erhaltenen Eingaben für explizite oder implizite Ziele ableitet, wie Ausgaben wie etwa Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen erstellt werden, die physische oder virtuelle Umgebungen beeinflussen können.“ [ER24]Art. 3 Nr. 1.

2. Biometrische Daten

„Biometrische Daten [sind] mit speziellen technischen Verfahren gewonnene personenbezogene Daten zu den physischen, physiologischen oder verhaltenstypischen Merkmalen einer natürlichen Person, wie etwa Gesichtsbilder oder daktyloskopische Daten [gehören.]“ [ER24]Art. 3, Nr. 34.

3. Biometrische Identifizierung

Biometrische Identifizierung umfasst „die automatisierte Erkennung physischer, physiologischer, verhaltensbezogener oder psychologischer menschlicher Merkmale zum Zwecke der Feststellung der Identität einer natürlichen Person durch den Vergleich biometrischer Daten dieser Person mit biometrischen Daten von Personen, die in einer Datenbank gespeichert sind.“ [ER24]Art. 3, Nr. 35.

4. Biometrische Verifizierung

Biometrische Verifizierung ist definiert als eine „automatisierte Eins-zu-eins-Verifizierung, einschließlich Authentifizierung, der Identität natürlicher Personen durch den Vergleich ihrer biometrischen Daten mit zuvor bereitgestellten biometrischen Daten.“ [ER24]Art. 3, Nr. 36.

5. Deepfake

Ein „Deepfake“ bezeichnet „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.“ [ER24]Art. 3, Nr. 60.

5.3 Versuch einer Einordnung des digitalen Rigs in den Rechtsrahmen des EU AI Act

1. Grundsätzliche Verortung

Unstrittig ist, dass durch die Anwendung des digitalen Rig biometrische Daten im Sinne des EU AI Acts, siehe Punkt 5.2, Nr. 2, erhoben und verarbeitet werden. Bei dem Vorgehen gem. den Szenarien 1 und 2, siehe Punkt 3.1, handelt es sich um die Anwendung zur Identifizierung, siehe Punkt 5.2, Nr.3, und gleichzeitig auch der Verifizierung, siehe Punkt 5.2, Nr.4. Aus Sicht der Autoren ist die Verwendung des digitalen Rigs für sich im Rahmen der Szenarien 1 und 2 kein automatisierter Abgleich der biometrischen Daten der

betroffenen Personen. Dies wäre eher dann der Fall, wenn exemplarisch neben den in der INPOL-Datenbank bisher erfassten Daten, auch systematisch die biometrischen Daten der Rigs erfasst und automatisiert in einer Datenbank abgeglichen werden würden.

2. Einklassifizierung in Risikogruppen

Gemäß Anhang III Nr. 6 c gehören KI-Systeme, die bestimmungsgemäß von Strafverfolgungsbehörden oder in deren Namen oder von Organen, Einrichtungen und sonstigen Stellen der Union zur Unterstützung von Strafverfolgungsbehörden zur Bewertung der Verlässlichkeit von Beweismitteln im Zuge der Ermittlung oder Verfolgung von Straftaten verwendet werden sollen, zu den Hochrisiko-KI-Systemen gemäß Artikel 6 Absatz 2 des EU AI Act. Der Artikel 6 „Einstufungsvorschriften für Hochrisiko-KI-Systeme“ enthält unter Abs. 3 Ausnahmebedingungen, unter denen ein KI-System dann nicht als hochriskant einzustufen ist, wenn es kein erhebliches Risiko der Beeinträchtigung in Bezug auf die Gesundheit, Sicherheit oder Grundrechte natürlicher Personen in sich birgt. Dies gilt unter anderem insbesondere, wenn:

- das KI-System ist dazu bestimmt, eine eng gefasste Verfahrensaufgabe durchzuführen und/oder
- das KI-System, siehe Punkt 5.2, Nr. 1, ist dazu bestimmt, eine vorbereitende Aufgabe für eine Bewertung durchzuführen, die für die Zwecke der in Anhang III aufgeführten Anwendungsfälle relevant ist.

Sofern die Zuordnung zu diesen ausschließenden Regelungen einer juristischen Prüfung standhält, gehört die Rig-Methode, sofern sie Teil eines „KI-Systems“ ist, was ihrer zielführenden Nutzung immanent ist, nicht zur Hochrisiko-KI. Allerdings wurde unter Punkt 3.2 bereits aufgezeigt, dass die Methode, wenn sie zielführend zur Identifizierung eingesetzt werden soll, in die Grundrechte einer natürlichen Person eingreift. Die Ausnahmeregelungen des Artikel 6 EU AI Act kommen demnach nach Ansicht der Autoren nicht in Betracht. Für diese Auffassung sprechen auch die Ausführungen des EU-Parlamentes selbst, dass zur Bewertung der Zuverlässigkeit von Beweisen eingesetzte KI zur Hochrisikogruppe gehört. [Eu24]

3. Folgen der Einstufung als Hochrisiko-KI

Gem. Artikel 1 Abs. 2 lit. c des EU AI Act ergeben sich besondere Anforderungen an Hochrisiko-KI-Systeme und Pflichten für Akteure in Bezug auf solche Systeme. Aufgrund des Umfangs werden nur wesentliche Anforderungen an die KI-Systeme (Artikel 9 ff. EU AI Act) überblicksartig dargestellt:

- Einrichtung, Anwendung, Dokumentation und Aufrechterhaltung eines Risikomanagementsystems
- Daten und Daten-Governance – Verwendung von Trainings- Validierungs- und Testdaten bestimmter Güte
- Regelungen für den Umgang mit personenbezogenen Daten

- Technische Dokumentation, wie die gestellten Anforderungen erfüllt werden
- Aufzeichnungspflichten (Automatische Aufzeichnung von Ereignissen)
- Transparenz und Bereitstellung von Informationen für die Betreiber
- Wirksame menschliche Beaufsichtigung
- Erreichen einer angemessenen Genauigkeit, Robustheit und Cybersicherheit

Ferner ergeben sich folgende wesentliche Pflichten für Anbieter und Betreiber:

- Sicherstellen der o.g. Anforderungen
- Durchlaufen von Konformitätsbewertungsverfahren
- Einrichten eines Qualitätsmanagementsystems
- Aufbewahrungspflichten
- Informationspflichten
- Pflicht zur Zusammenarbeit mit Behörden

6 Fazit und Ausblick

Das digitale Rig ist eine innovative Methode zur Identifizierung von Personen anhand passiver biometrischer Merkmale. Die Anwendung kann die Erkennung von synthetischen Bild- und Videoinhalten ergänzen und damit zur Verifizierung von vorliegendem Beweismaterial beitragen sowie die Aufklärung von Identitätsdiebstählen unterstützen, sofern die entsprechenden Voraussetzungen zur Anfertigung eines Rigs erfüllt sind. Die Anwendung der Maßnahme stützt damit inhärent auch die Integrität der kriminalistischen Beweisführung mit dem Anspruch der Wahrheitsfindung. Dass die Methode erfolgreich funktionieren kann, zeigen die Fälle in denen die Rig-Methode zur Identifizierung von Personen und damit eindeutigen Zuordnung durch einen anthropologischen Vergleich verwendet wurde.

Die Methode kann derzeit unter Einhaltung bestehender Normen der StPO rechtssicher zur Anwendung kommen.

Es besteht jedoch auch die theoretische Gefahr, dass das digitale Rig, insbesondere von Personen des öffentlichen Lebens, aus veröffentlichtem Videomaterial mit zusätzlichem Aufwand abgeleitet und somit in die Erstellung von Deepfakes einfließen kann. In diesen Fällen kann das Rig nicht mehr uneingeschränkt zur Authentifizierungsbestimmung herangezogen werden. Der Datensatz eines personenbezogenen digitalen Rigs sollte demnach besonderen Schutzmaßnahmen unterliegen, um einem möglichen Missbrauch entgegenzuwirken.

Eine erste Einordnung der Maßnahme in den Regelungsrahmen des EU AI Actes zeigt, dass das digitale Rig als passives biometrisches Verfahren in Verbindung mit automatisierten Prozessen zur Bewertung der Zuverlässigkeit von Beweisen im Strafverfahren der Gruppe der Hochrisiko-KI gehört und damit den umfangreichen Regelungen des EU AI Actes unterliegt. Gleichwohl bedarf diese Einschätzung der weiteren juristischen Prüfung in der Zukunft. Eine hier derzeit noch nicht betrachtete Möglichkeit der Nutzung der Methode im Rahmen von Fernidentifizierungssystemen käme zukünftig in Betracht und wäre in jedem Falle der Hochrisiko-KI im Sinne des EU AI Actes zuzuordnen. Weiter zu betrachten wäre, welche rechtlichen Voraussetzungen zur generellen präventiven Erstellung des digitalen Rig im Rahmen von erkennungsdienstlichen Maßnahmen gem. § 81b Abs. 1 2. Alternative StPO zu erfüllen wären.

Literaturverzeichnis

- [ACR22] Ackermann, R.; Clages, H.; Roll, H.: Handbuch der Kriminalistik. Richard Boorberg Verlag GmbH & Co KG, 2022.
- [Be22] Becker, S. et al.: COMBI: Artificial Intelligence for Computer-Based Forensic Analysis of Persons. KI - Künstliche Intelligenz 2/36, S. 171–180, 2022.
- [Bo19] Bovenschulte, M.: Deepfakes – Manipulation von Filmsequenzen. Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag (TAB), 2019.
- [Br24] Brömmelhörster, L.: Potentielle Gefahren von Deepfakes. Inwieweit können Deepfakes einen Angriff auf die Demokratie darstellen? Kriminalistik -Unabhängige Zeitschrift für die kriminalistische Wissenschaft und Praxis 4, S. 248–256, 2024.
- [Bu21] Bundesamt für Sicherheit in der Informationstechnik: Deepfakes - Gefahren und Gegenmaßnahmen. https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Informationen-und-Empfehlungen/Kuenstliche-Intelligenz/Deepfakes/deepfakes_node.html, Stand: 09.05.2024.
- [Bu22] Bundesministerium für Bildung und Forschung: Erkennung KI-basierter Sprachsynthese und Stimmverfremdung (SpeechTrust+). https://www.sifo.de/sifo/shareddocs/Downloads/files/projektumriss_speechtrust.pdf?__blob=publicationFile&v=3, Stand: 12.05.2024.
- [De23] Deutscher Bundestag: Entwurf eines Gesetzes zur digitalen Dokumentation der strafgerichtlichen Hauptverhandlung. Drucksache 20/8096. <https://dserver.bundestag.de/btd/20/080/2008096.pdf>, Stand: 16.06.2024.
- [Di24] Die Bundesregierung: Deep Fakes erkennen mit KI. Interview zum Forschungsprojekt „Fake-ID“. <https://www.bundesregierung.de/breg-de/themen/digitalisierung/kuenstliche-intelligenz/ki-gegen-deep-fakes-2274090>, Stand: 12.05.2024.
- [En24] Engelhard, J. et al.: Digital-anthropometrischer Rigabgleich als forensisches Instrument zur bildgestützten, biometrischen Personenidentifizierung. Eine interdisziplinäre Betrachtung. Kriminalistik -Unabhängige Zeitschrift für die kriminalistische Wissenschaft und Praxis 5, 2024.
- [ER24]: VERORDNUNG (EU) 2024/1689 DES EUROPÄISCHEN PARLAMENTS UND DES RATES 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). VERORDNUNG (EU) 2024/1689: Amtsblatt der Europäischen Union, 2024.

- [Eu24] European Union: Shaping Europe's digital future. AI Act. <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>, Stand: 12.05.2024.
- [Ga14] Garrido, P. et al.: Automatic Face Reenactment, S. 1–8, 2014.
- [HHL22] Heinke, F.; Heuschkel, M.-L.; Labudde, D.: A frequentist estimation of duplicate probability as a baseline for person identification from image and video material using anthropometric measurements. Gesellschaft für Informatik, Bonn, 2022.
- [HHL23] Heinke, F.; Heuschkel, M.-L.; Labudde, D.: Analysing Distributions of Feature Similarities in the Context of Digital Anthropometric Pattern Matching Probability. Gesellschaft für Informatik e.V, 2023.
- [Ho21] Hosler, B. et al.: Do Deepfakes Feel Emotions? A Semantic Approach to Detecting Deepfakes Via Emotional Inconsistencies: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, S. 1013–1022, 2021.
- [In22] Intel Corporation: Intel Introduces Real-Time Deepfake Detector. Intel's deepfake detector analyzes 'blood flow' in video pixels to return results in milliseconds with 96% accuracy. <https://www.intel.com/content/www/us/en/newsroom/news/intel-introduces-real-time-deepfake-detector.html#gs.axw9j2>, Stand: 12.05.2024.
- [Ki20] Kietzmann, J. et al.: Deepfakes: Trick or treat? Business Horizons 2/63, S. 135–146, 2020.
- [Kr23] Krupicka, E.: Künstliche Bilder – eine Herausforderung für Polizei und Forensik?! Kriminalistik -Unabhängige Zeitschrift für die kriminalistische Wissenschaft und Praxis 11, S. 599–602, 2023.
- [La23] Labudde, D.: Das digitale Rig als intelligentes, bildbasiertes, forensisches Instrument. SIAKJournal – Zeitschrift für Polizeiwissenschaft und polizeiliche Praxis 4, S. 28–39, 2023.
- [PB21] Pawelec, M.; Bieß, C.: Deepfakes. Technikfolgen und Regulierungsfragen aus ethischer und sozialwissenschaftlicher Perspektive. Nomos Verlagsgesellschaft mbH & Co. KG, 2021.
- [PRL23] Pistorius, E.; Richter, S.; Labudde, D.: The digital skeleton in modern video analysis - inter- and intraspecific comparison of individual rigs. Gesellschaft für Informatik e.V, 2023.
- [SZ24] Steffes, B.; Zichler, A.: Deepfakes in Videoüberhandlungen vor Gericht. Datenschutz und Datensicherheit - DuD 3/48, S. 158–163, 2024.