Lea Uhlenbrock, Giulia Canova, Denise Moussa, Monika Simmler, Christian Riess

Synthetische Bilddaten vor Gericht

Technische und juristische Herausforderungen für die Beweiswürdigung im Strafverfahren

In Zeiten, in denen fotorealistische Falschbilder in wenigen Klicks erstellt werden können, steht der Beweiswert von Bildern in Frage. Für die entstehenden juristischen und technischen Herausforderungen bedarf es dringend Lösungen.

Das Aufkommen synthetischer Bilder, die mithilfe künstlicher Intelligenz (KI) erzeugt werden, wirft grundlegende Fragen zur Authentizität von Bilddaten auf. Die technologischen Fortschritte der künstlichen Generierung von Bildern erlauben mittlerweile die Herstellung und Verbreitung von realistisch anmutenden Bildern, bei denen rein visuell nichts auf deren künstliche Herkunft hinweist. Durch diese Fortschritte der generativen KI gepaart mit der einfachen Zugänglichkeit entsprechender Werkzeuge werden künstlich generierte Bilder bereits breitflächig zur Falschinformation, als Propagandamaterial oder zu Betrugszwecken eingesetzt. Früher oder später werden synthetische Bilder auch als potenzielle Beweismittel vor Gerichten landen.[1] Dies bringt Herausforderungen auf technischer und juristischer Seite und erfordert Lösungsideen.

1 Synthetische Bilder vor Gericht

Die Evolution synthetischer Bilder führt bei der Verwendung von Bilddaten als Beweismittel zu grundsätzlichen Problemstellungen IT-forensischer und rechtlicher Natur. Einerseits wird es mit fortschreitender Verbesserung der Bildgeneratoren schwieriger, die Echtheit oder eben Authentizität eines Bildes im Rahmen der forensischen Bildanalyse nachzuweisen. Andererseits wirkt sich

1 Vgl. Venema/Geradts, *Digital Forensics, Deepfakes and the Legal Process.* The Science Tech Lawyer 2020, 14.

Lea Uhlenbrock

ist wissenschaftliche Mitarbeiterin und Doktorandin in der Gruppe für Multimedia-Sicherheit am Lehrstuhl für IT-Sicherheitsinfrastrukturen an der Friedrich-Alexander-Universität Erlangen-Nürnberg. Ihr Forschungsschwerpunkt liegt in der automatisierten Erkennung KI-generierter Bilder.
E-Mail: lea.uhlenbrock@fau.de

Giulia Canova

ist wissenschaftliche Mitarbeiterin und Doktorandin am Kompetenzzentrum für Strafrecht und Kriminologie an der Universität St. Gallen. Sie forscht zur strafprozessrechtlichen Regulierung digitaler Ermittlungen.
E-Mail: giulia.canova@unisg.ch

Denise Moussa

ist wissenschaftliche Mitarbeiterin und Doktorandin in der Gruppe für Multimedia-Sicherheit am Lehrstuhl für IT-Sicherheitsinfrastrukturen an der Friedrich-Alexander-Universität Erlangen-Nürnberg. Ihre Forschungsschwerpunkte liegen im Bereich des Maschinellen Lernens in der Audio- und Bildforensik.
E-Mail: denise.moussa@fau.de

Prof. Dr. Monika Simmler

ist Assistenzprofessorin für Strafrecht, Strafprozessrecht und Kriminologie sowie Co-Direktorin des Kompetenzzentrums für Strafrecht und Kriminologie an der Universität St. Gallen. Ihre Forschungsschwerpunkte sind Grundlagenfragen der strafrechtlichen Verantwortlichkeit sowie das IT-Strafrecht.
E-Mail: monika.simmler@unisg.ch

PD Dr.-Ing. Christian Riess

leitet die Gruppe Multimedia-Sicherheit am Lehrstuhl für ITSicherheitsinfrastrukturen der Friedrich-Alexander-Universität Erlangen-Nürnberg. Seine Forschung beschäftigt sich mit Methoden der Signal- und Bildverarbeitung mit Anwendungen in der Cybersicherheit.
E-Mail: christian.riess@fau.de

bereits die Möglichkeit, dass es sich bei visuell echt anmutenden Bilddaten um generierte Bilder handelt, grundlegend auf deren Würdigung als Beweismittel aus. Das Vertrauen in die Echtheit eines Bildes ist Grundvoraussetzung für dessen Verwendung als Beweismittel. Das Gericht ist bei der Würdigung von Beweismitteln frei und an keine vorgegebenen Regeln gebunden. Es stellt sich dabei jedoch die Frage, wie ein Gericht den Beweiswert digitaler Bilddaten beurteilen soll, wenn sie künstlich erzeugt worden sein könnten, dies jedoch durch reinen Augenschein nicht feststellbar ist.

In Anbetracht dieser Ausgangslage widmet sich dieser Beitrag den technischen und strafprozessrechtlichen Herausforderungen, die entstehen, wenn potenziell synthetische Bilder ins Strafverfahren eingehen. Ziel ist es, die Vielschichtigkeit der Problematik aufzuzeigen und Handlungsvorschläge für die Evaluation von Bilddaten zu präsentieren.

2 Synthetische Bilddaten und Beweiswürdigung

2.1 Entstehung synthetischer Bilder

Herkömmliche digitale Bilddaten entstehen durch die Aufnahme einer realen Szene durch eine Kamera. Um eine nicht-reale Szene in ähnlicher Glaubhaftigkeit zu erzeugen, bedurfte es bisher entweder der aufwändigen Erstellung von künstlichen Szenen oder Techniken der Bildbearbeitung, um Elemente verschiedener Bilder zu einer neuen Darstellung zusammenzufügen. Mit der Veröffentlichung des KI-basierten Bildgenerators DALL-E[2] und den seither rapiden Entwicklungen im Bereich der Text-zu-Bild-Generatoren ändert sich dies jedoch. Nun ist es möglich, mit wenigen Klicks und einer kurzen Bildbeschreibung völlig neue Szenarien in fotorealistischer Qualität zu erzeugen. Diese Bilder werden als synthetisch bezeichnet. Ein Meilenstein in der Erzeugung realistischer Bildinhalte mithilfe von KI ist das Generative Adversarial Network (GAN)[3], das 2014 vorgestellt wurde. Seit 2022 erlangte eine Nachfolgetechnik, sog. Diffusionsmodelle[4], sprunghaft Berühmtheit. Zu den bekanntesten Modellen gehören DALL-E 3[5], Midjourney 5[6], Stable Diffusion XL[7] und das in Adobe Photoshop integrierte Firefly[8]. Während sich anfangs oft Fehler einschlichen, wie eine zu hohe Anzahl Finger an der Hand eines Menschen, sind neue Modelle in der Lage, fotorealistische Bilder meist ohne offensichtliche Makel herzustellen. Diese Entwicklung wirft Fragen auf, wovon einige die Untersuchung und Beweiswürdigung von Bilddaten betreffen. Es ist absehbar, dass synthetische Bilder in Ermittlungen und Gerichtsverfahren auftauchen.

2 Craiyon https://www.craiyon.com. Zuletzt abgerufen am 15.07.2024

3 Goodfellow et al. *Generative adversarial nets.* Advances in neural information processing systems, 2014.

4 Rombach et al., *High-resolution Image Synthesis with Latent Diffusion Models*, IEEE/CVF Conference on Computer Vision and Pattern Recognition 2022.

5 DALL-E 3 https://openai.com/dall-e-3. Zuletzt abgerufen am 15.07.2024

6 Midjourney https://www.midjourney.com/. Zuletzt abgerufen am 15.07.2024

7 Stable Diffusion https://stability.ai/stable-diffusion. Zuletzt abgerufen am 15.07.2024

8 Adobe Firefly https://www.adobe.com/de/products/firefly.html. Zuletzt abgerufen am 15.07.2024

2.2 Digitale Bilddaten als Beweismittel im Strafverfahren

Aufgrund ihrer Abbildungsfunktion dienen Bilddaten in Strafverfahren regelmäßig dem Nachweis von Sachverhalten, so z.B. in Form von Bildern aus Überwachungskameras oder Smartphones. Traditionell wird Bildern seit dem Einbezug von Analogfotografien ein hoher Beweiswert zugeordnet, da sie in (vermeintlich) objektiver Weise Begebenheiten visuell abbilden[9] und gut dazu geeignet sind, Falldetails intuitiv verständlich zu präsentieren.[10] In Zeiten der Digitalfotografie liegen sie nun allerdings oft nicht mehr in analogem Format vor, sondern als digitale Bilddaten.

Trotz der Relevanz digitaler Daten als Beweismittel existieren im deutschsprachigen Raum keine spezifischen Regeln für den Umgang mit ihnen im strafprozessualen Beweiskontext.[11] Das Deutsche Strafprozessrecht normiert eine Reihe von Beweismitteln: Den Augenscheinbeweis (§ 86 ff. DE-StPO) zur sinnlichen Wahrnehmung von Sachen oder Personen; den Zeugenbeweis (§ 48 ff. DE-StPO); sowie den Sachverständigenbeweis (§ 72 ff. DE-StPO), sobald es zur Ermittlung eines Sachverhaltes besonderer Sachkunde von Expertinnen oder Experten bedarf. Schließlich ist der Urkundenbeweis (§ 249 ff. DE-StPO) gesetzlich geregelt, der sich der Nutzung schriftlich verkörperter und verlesbarer Gedankenerklärungen widmet.

Digitale Bilddaten sind als Beweismittel hingegen nicht eigens gesetzlich vorgesehen. Sie können aber über den „Umweg" der normierten Beweismittel in Strafverfahren eingebracht werden. In der Praxis werden digitale Bilddaten meist über den Augenscheinbeweis als Beweismittel eingebracht.[12]

Um dem Augenscheinbeweis zugänglich zu werden, müssen maschinenlesbare, digitale Bilddaten zunächst in eine sinnlich wahrnehmbare Form gebracht werden (z.B. digital auf einem Bildschirm dargestellt oder auf Papier ausgedruckt).[13] Die ausgedruckten oder digital auf einem Bildschirm wiedergegebenen Bilddaten können dann im Rahmen des richterlichen Augenscheins beurteilt werden.[14] Auf diese Weise kann das für die Würdigung der Beweise zuständige Gericht das Bild wahrnehmen und beurteilen.

In einem zweiten Schritt geht es darum, im Rahmen der Beweiswürdigung den Beweiswert der eingebrachten Beweismittel festzustellen und basierend auf der gesamten Beweisgrundlage zu beurteilen, ob eine Tatsache als erwiesen gilt oder nicht. Diese Beurteilung folgt im deutschsprachigen Raum dem Grundsatz der freien richterlichen Beweiswürdigung.[15] Das Gericht fällt Urteile nach seiner freien Überzeugung, die es aus der Gerichtsverhandlung schöpft.[16] Bei der Würdigung der Beweise ist das Gericht an keine festen Regeln gebunden, die vorschreiben, wann eine Tat-

9 Knopp, *Rechtliche Perspektiven zur digitalen Beweisführung*, Informatik 2009, 156 f.

10 Hak, *Image-based Evidence in International Criminal Prosecutions: Charting a Path Forward*, Oxford 2024, xxiii ff.

11 Vgl. Sieber/Brodowski, Handbuch Multimedia-Recht, München 2023, Rz. 164; Knopp (Fn. 9), 156 ff.; Rückert, *Digitale Daten als Beweismittel im Strafverfahren*, Tübingen 2023, 651 ff.

12 Vgl. Sieber/Brodowski (Fn. 10), Rz. 163; Heinson, IT-Forensik, Tübingen 2014, 113.

13 Fährmann, *Digitale Beweismittel und Datenmengen im Strafprozess*, MMR 2020, 228; Heinson (Fn. 12), 113.

14 Heinson (Fn. 12), 113.

15 Sieber/Brodowski (Fn. 11), Rz. 164; Nay, *Freie Beweiswürdigung und in dubio pro reo*, ZstrR 1996, 87 ff.

16 Siehe § 261 DE-StPO.

sache als bewiesen gilt oder welchen Wert einem Beweismittel beizumessen ist.[17]

Das Gericht bestimmt den Wert eines Beweises folglich selbst. Anerkannt ist jedoch, dass sich das Gericht an gewissen (gesetzlich nicht vorgeschriebenen) forensischen, kriminalistischen und wissenschaftlichen Grundregeln orientieren muss.[18] Auch die Würdigung von digitalen Beweismitteln unterliegt der freien Beweiswürdigung, wobei forensische Grundregeln zu beachten sind. Liegt ein digitales Bild vor, hat das Gericht die Aufgabe, den Beweiswert des Bildes – und damit auch dessen Echtheit – zu beurteilen. Dabei sind die Grundregeln der IT-Forensik und forensischen Bildanalyse miteinzubeziehen. Diesen Regeln zufolge hängt der Beweiswert digitaler Daten maßgeblich davon ab, dass ihre Authentizität und Integrität gewährleistet ist. Die Daten müssen echt, unverändert und vollständig sein.[19]

Bei dem Beweis dienenden Bilddaten ist das Gericht deshalb gehalten, deren Authentizität und Integrität zu bewerten. In Anbetracht der aktuellen technologischen Entwicklung werden sich Gerichte zunehmend damit konfrontiert sehen, die Echtheit von potenziell KI-generierten Bilddaten beurteilen zu müssen, bei denen mit bloßem Auge keine Manipulation ersichtlich ist. Damit gerät der Beweiswert von Bilddaten grundsätzlich ins Wanken. Gleichzeitig wird es für Gerichte unmöglich, die Authentizität und Integrität von Bilddaten ohne technische Analyseverfahren festzustellen. Mangels Expertise zu eben diesem Nachweis der Authentizität und Integrität kann das Gericht allerdings über den Sachverständigenbeweis (§ 72 ff. DE-StPO) Sachverständige aus der IT-Forensik beziehen.

3 Stand der Technik: Forensische Bildanalyse

Die Prüfung der Authentizität ist ein Bestandteil der sog. forensischen Bildanalyse. Ziel der Authentizitätsprüfung ist es, die Echtheit eines Bildes zu bestätigen oder zu widerlegen. Vor dem Aufkommen von KI-Bildgeneratoren konnten überzeugend gefälschte Fotografien hauptsächlich in Handarbeit mit Bildbearbeitungswerkzeugen wie Adobe Photoshop oder Gimp hergestellt werden. Durch gezielte Bildmanipulationsoperationen kann dabei die Aussage eines Bildes für das menschliche Auge überzeugend verändert werden. Für die Detektion derartiger Manipulationen in digitalen Bildaufnahmen haben sich diverse forensische Analyseverfahren etabliert. Sie zielen im Regelfall auf eine Suche nach unerwarteten Anomalien ab, die nicht mit dem Bildaufnahmeprozess zusammenpassen.

Auf Basis der abgebildeten Szene kann etwa die Plausibilität geprüft werden. So können bspw. inkonsistente semantische oder physikalische Eigenschaften eine Fälschung entlarven. Unrealistische Größenverhältnisse, kopierte Bildbereiche oder auch inkonsistente Lichtverhältnisse innerhalb der Aufnahme weisen darauf hin, dass Bildinhalte nachträglich eingebracht wurden.[20]

Die Suche nach Manipulationen kann ebenfalls auf Spuren basieren, die während des Aufnahmeprozesses im Bild entstehen. Zum Beispiel enthält jeder Sensor fertigungsbedingte Unterschiede, welche ein charakteristisches Rauschmuster erzeugen.[21] Bildbearbeitungsoperationen können solche Rauschmuster verletzen und werden dadurch als Anomalie im extrahierten Rauschbild lokalisierbar.[22]

Am Ende des digitalen Bildaufnahmeprozesses steht die Speicherung der Datei auf einem Datenträger in einem bestimmten Dateiformat. Auch dies wird regelmäßig für eine Manipulationsanalyse ausgenutzt. So hinterlassen verlustbehaftete Dateiformate, wie z.B. das JPEG-Format, diverse und oft charakteristische Spuren in den Pixelwerten. Das JPEG-Format hinterlässt aufgrund seiner Kompressionsstrategie blockartige Artefakte in einer Bilddatei, die im Rahmen einer Analyse extrahiert werden können. Schon das Kopieren und Verschieben von Bereichen innerhalb desselben Bildes verletzt in fast allen Fällen die Konsistenz des regelmäßigen Blockmusters und kann manipulierte Regionen sichtbar machen.[23]

Gerade solche Spuren lassen sich mitunter auch noch bei teilsynthetischen Bildern, also solchen mit natürlichen und generierten Anteilen, nachweisen. Generell stellen KI-Bildgeneratoren die forensische Bildanalyse allerdings vor neue Herausforderungen. Klassische Bildaufnahmespuren sind in synthetischen Bildern i.d.R. nicht enthalten. Dennoch sind Plausibilitätsprüfungen durchführbar, die auf den gleichen Prinzipien beruhen. Zum Beispiel sind KI-Bildgeneratoren oft nur begrenzt in der Lage, komplexe physikalische Vorgänge wie konsistente Lichtbrechungen, Schattenwürfe oder Reflexionen korrekt zu simulieren.[24] Räumliche Zusammenhänge sind in synthetischen Bildern bei einer genauen Untersuchung oft unplausibel.[25] Zusätzlich enthalten synthetische Bilder oft charakteristische Muster, die mithilfe forensischer Werkzeuge identifiziert werden können.[26]

4 Herausforderungen für die technische Evaluation von Bilddaten

Technische Herausforderungen können für zwei unterschiedliche Szenarien unterschieden werden: Die Einzelfallanalyse und die Analyse von Massendaten. Bei Einzelfallanalysen spielt v.a. die Weiterentwicklung der Bildgeneratoren eine große Rolle, die das allmähliche Verschwinden von subtilen, sichtbaren Spuren bewirkt. Das kann dazu führen, dass Prüfungen von Bildern aufgrund augenscheinlicher, überzeugender Echtheit gar nicht erst in Auftrag gegeben werden. Umgekehrt kann durch die visuelle Ununterscheidbarkeit von synthetischen und Originalbildern die Echtheit jedes Bildes angezweifelt werden. Weitere Herausforderungen ergeben sich durch Fälle, bei denen nicht geklärt ist, ob sie als authentisch oder manipuliert betrachtet werden sollen. Da sich durch verschiedene, parallel laufende Entwicklungen die forensischen Eigenschaften von realen und synthetischen Bild-

17 KK StPO-Tiemann, § 261 Rz. 87.

18 BVerfG NJW 2003, 244; Möllers/Salemi/Schliwinski, *Digitale Beweise im Straf- und Zivilprozess*, Jusletter IT 2022, 173; vgl. Rückert (Fn. 11), 661.

19 Rückert (Fn. 11), 665.

20 Farid, *Photo Forensics*, Cambridge 2019, 5 ff., 73 ff.

21 Lukaš et al., *Digital Camera Identification from Sensor Pattern Noise*, IEEE Transactions on Information Forensics and Security 1.2 2006, 205 ff.

22 Chen et al., *Determining Image Origin and Integrity using Sensor Noise*, IEEE Transactions on information forensics and security 3.1, 2008, 74 ff.

23 Farid (Fn. 20), 217

24 Farid, *Lighting (In)consistency of Paint by Text*, arXiv preprint 2207.13744, 2022.

25 Farid, *Perspective (In)consistency of Paint by Text*, arXiv preprint 2206.14617, 2022.

26 Corvi et al., *On the Detection of Synthetic Images generated by Diffusion Models*, IEEE International Conference on Acoustics, Speech and Signal Processing 2023.

daten aneinander annähern, wird darüber hinaus die Unterscheidung erschwert. Für Massendatenanalysen ergeben sich darüber hinaus weitere Problemstellungen, da eine automatisierte Prüfung nicht trivial und mit komplexen Fragestellungen verknüpft ist. Im Folgenden werden die Herausforderungen für die Evaluation näher erörtert.

KI-Filter

In vielen Smartphone-Apps sind Filter enthalten, die das Bild leicht verändern, um eine bestimmte Ästhetik zu erreichen. Von der Weichzeichnung der Haut oder des Hintergrundes bis zur KI-basierten Anpassung der Farbeinstellungen ist vieles bereits automatisch aktiviert oder kann mit einem Klick hinzugefügt werden. Für die Gerichte stellt sich die Frage, inwieweit derartige Filter das Bild semantisch verändern und welche Änderungen man bei der Nutzung eines Bildes als Beweismittel zulassen will, wenn die Originalaufnahme gar nicht auf dem Gerät gespeichert wird. Abgesehen von den visuellen Veränderungen, die solche Filter verursachen, verändern sie ebenfalls die statistischen Eigenschaften eines Bildes, auf denen viele forensische Untersuchungen beruhen und erschweren damit auch die Analyse.

Veränderte Spuren

Eine kontinuierliche Verbesserung der Qualität der KI-Bildgeneratoren bedeutet, dass sich die Bilder stärker an Realbilder annähern, mithilfe derer sie trainiert wurden. Auch forensische Spuren von Realbildern wie zum Beispiel das Bildrauschen könnten so mit der Zeit übernommen werden.

Kompressionsverfahren, die in den meisten Apps enthalten sind, schwächen zudem Spuren des Aufnahmegeräts und erschweren damit eine Zuordnung.[27] Ein weiteres Problem wird auftreten, wenn der JPEG AI Standard[28] in Kraft tritt. Dieses neue Kompressionsverfahren basiert auf der Verwendung eines neuronalen Netzes, wodurch KI-komprimierte Realbilder ähnliche Spuren aufweisen wie KI-generierte Bilder.[29]

Zusammenfassend werden die bekannten Spuren von Realbildern also schwieriger nachzuweisen oder verschwinden gar gänzlich, während möglicherweise synthetische Bilder in Zukunft sehr ähnliche Spuren aufweisen werden.

Automatisierung und Erklärbarkeit

Bei dem Anfallen größerer Datenmengen ist es naheliegend, die Sichtung bzw. Analyse zu automatisieren. Dies trägt dazu bei, die Überlastung von Behörden und Dienstleistern zu reduzieren. Automatisierungsmethoden können Daten vorsortieren oder klassifizieren. In beiden Fällen sind verlässliche Analysemethoden sehr wichtig. Zusätzlich setzt der AI Act der Europäischen Union[30] einen rechtliche Rahmen für die Validierung und Dokumentation von automatisierten Entscheidungssystemen. Eine allgemeine Herausforderung automatisierter Verfahren liegt in der Vermeidung von systematischen Fehleinschätzungen aufgrund ungewollter Faktoren. Beispielsweise können neuronale Netze lernen, Entscheidungen anhand von Faktoren zu treffen, die nicht für die Klassifizierung vorgesehen waren. Aufgrund der Opazität solcher Methoden kann dies unbemerkt geschehen.[31] Diese Faktoren können dann unter anderem Aussehen, Geschlecht oder Hautfarbe sein anstatt der statistischen Merkmale eines Bildes. Solche Fehler können mit erklärbaren Methoden unter Umständen direkt gefunden und vermieden werden. Die Ausgaben moderner neuronaler Netze sind in der Regel jedoch wenig erklärbar, und müssen daher durch gründliches Testen validiert werden.

5 Herausforderungen für die Beweiswürdigung im Strafverfahren

Künstliche Bilddaten stellen sowohl die Nutzung von Bildern als Beweismittel als auch den Grundsatz der freien Beweiswürdigung vor Herausforderungen. Das Strafprozessrecht verfügt über kein spezifisches Instrument für die Einbringung von Daten als Beweismittel. Über den Augenscheinbeweis können digitale Bilder zwar als Beweismittel eingeführt werden; Voraussetzung dafür ist jedoch, dass die Bilddaten in eine visuell wahrnehmbare Form transformiert werden (durch einen Ausdruck oder eine Darstellung auf einem Bildschirm). Bei der Transformation in ein Augenscheinobjekt geht der Bezug zu den digitalen Daten, aus denen Bilder bestehen, verloren. Es werden nicht Daten gewürdigt, sondern die menschlich wahrnehmbaren visuellen Abbildungen. Das Einbringen von Bilddaten über sinnlich wahrnehmbare Beweisstücke macht es unmöglich zu erkennen, ob die Daten, in denen die Bildinformationen enthalten sind, manipuliert oder gar künstlich generiert wurden.[32]

Dass die Einbringung von Bilddaten an visuell wahrnehmbare Beweisstücke anknüpft, hat Konsequenzen für die Würdigung durch das Gericht. Zentrale Voraussetzungen für einen hohen Beweiswert von digitalen Bilddaten muss deren Echtheit sein, was die Authentizität und Integrität der Daten bedingt.[33] Diese Echtheit lässt sich durch das Gericht durch eine visuelle Betrachtung jedoch im digitalen Zeitalter nicht mehr beurteilen. Um Manipulationen oder Fälschungen zu erkennen (bzw. auszuschließen), bedarf es einer vertieften Datenanalyse. Ohne Kenntnisse der forensischen Bildanalyse ist dem Gericht eine solche Beurteilung nicht möglich. Das System der freien richterlichen Beweiswürdigung überlässt dem Gericht also die Verantwortung für die Würdigung von Bilddaten, die es ohne besondere Kenntnisse gar nicht beurteilen kann.

Fehlt es dem Gericht an Expertise zur Beurteilung der Bilddaten, besteht die Möglichkeit, über den Sachverständigenbeweis IT-Sachverständige zur forensischen Bildanalyse beizuziehen. Die Ergebnisse der forensischen Bildanalyse zur Echtheit von Bildern können dann in Form eines Sachverständigenberichts (zusätzlich zum Bild) als Beweismittel eingebracht werden. Wie in Kap. 3 ausgeführt, verfügt die IT-Forensik über verschiedene Techniken zur Evaluation von Bilddaten. Auf Seiten des für die Beweiswürdigung zuständigen Gerichts besteht jedoch weiterhin die

27 De Roos/Geradts, Factors That Influence PRNU-Based Camera-Identification via Videos, Journal of Imaging 7.1, 2021, 8.

28 Ascenso/Upenik, *JPEG AI Score and Framework [White Paper]*, https://ds.jpeg.org/whitepapers/jpeg-ai-white-paper.pdf. Zuletzt abgerufen am 15.07.2024

29 Bergmann et al., *Frequency-Domain Analysis of Traces for the Detection of AI-based Compression*, International Workshop on Biometrics and Forensics 2023.

30 Artificial Intelligence Act (verabschiedet vom Europäischen Parlament am 21.05.2024) https://artificialintelligenceact.eu/de/das-gesetz/. Zuletzt abgerufen am 15.07.2024

31 Geirhos, Robert, et al., *Shortcut Learning in Deep Neural Networks*. Nature Machine Intelligence 2020, 665-673.

32 Vgl. Rückert (Fn. 11), 655.

33 Vgl. Rückert (Fn. 11), 665.

Herausforderung, überhaupt erst zu erkennen, wann der Beizug von Sachverständigen zur technischen Evaluation der Bilder angezeigt wäre. Bei mittels moderner Generatoren erstellten synthetischen Bildern sind auf Manipulationen hinweisende Spuren zumeist gerade nicht mehr mit bloßem Auge zu erkennen.

Im Umgang mit digitalen Beweismitteln sollte ein Bewusstsein dafür geschaffen werden, dass auch bei realistisch anmutenden Bildern eine Wahrscheinlichkeit besteht, dass diese manipuliert bzw. künstlich generiert wurden. Um die in der IT-Forensik bestehende Expertise einbringen zu können, sollten grundlegende Möglichkeiten der ersten Überprüfung der Authentizität und Integrität von Bilddaten – wie auch digitaler Beweismittel im Allgemeinen – bei den für die Beweiswürdigung zuständigen Gerichtsbehörden vorhanden sein. Nur so lassen sich problematische Fälle, bei denen ein vertiefte Überprüfung angezeigt ist, erst erkennen. Ansonsten wären schlicht alle Bilder fachkundig zu überprüfen – mit den entsprechenden Konsequenzen bezüglich Ressourcenbedarf.

6 Handlungsvorschläge

Für den Umgang mit den beschriebenen Herausforderungen bestehen verschiedene Ansatzpunkte. Das Erlangen und die Aufbewahrung von Bilddaten ist in der digitalen Forensik bereits zufriedenstellend gelöst, da es großflächig akzeptierte Leitfäden[34] für die Beschaffung und Verwaltung von digitalen Daten gibt, die genauso für digitale Bilder anwendbar sind. Die Analyse sowie die Berichterstattung folgt jedoch keinen offiziellen Standards und ist methodisch stark fragmentiert.[35] Dies verhindert, dass Prozesse und Ergebnisse einheitlich nachvollzogen werden können, sowohl von Staatsanwaltschaft und Gericht, als auch von der Verteidigung, um Ergebnisse anfechten zu können. Die folgenden Vorschläge zielen darauf ab, Nachvollziehbarkeit und Überprüfbarkeit zu verbessern, ohne den Analyseaufwand zu erhöhen.

Erstanalyse durch Gericht

Die detaillierte Analyse eines Bildes, um genaue Schlussfolgerungen über seine Herkunft und Authentizität zu ziehen, muss gegenwärtig und auch zukünftig durch forensisches Fachpersonal erfolgen. Eine erste Analyse jedoch, die bereits Aussagen über die prinzipielle Echtheit eines Bildes treffen kann, ließe sich als

34 Bundesamt für Sicherheit in der Informationstechnik, Leitfaden IT-Forensik, https://www.bsi.bund.de/dok/6620610. Zuletzt abgerufen am 15.07.2024

35 Stoykova et al., *Machine Learning for Evidence in Criminal Proceedings: Techno-legal Challenges for Reliability Assurance*, Artificial Intelligence (AI) in Forensic Sciences 2023, 21 ff.

abgeschlossene Software umsetzen. Dies würde es Gerichtsbehörden erlauben, die Authentizität von Bilddaten in einem ersten „Screening" zu überprüfen und Anzeichen für Manipulationen oder künstliche Generierung selbst zu erkennen. So könnten problematische Bilddaten mit einem vertieften Abklärungsbedarf vorab identifiziert werden. Im Rahmen des gesetzlich normierten Augenscheinbeweises wäre es rechtlich durchaus möglich, dass sich das Gericht selbst unter Verwendung geeigneter Software Kenntnis von den für die Echtheit relevanten (Meta-) Daten verschafft und diese würdigt.[36]

Dabei könnte das Ergebnis entweder lauten, dass das vorliegende Bild zweifelsfreie Authentizität besitzt, oder dass die Echtheit nicht einwandfrei belegt werden kann. Im Zweifelsfall wäre eine weitere forensische Analyse angezeigt. Dieses Vorgehen würde es ermöglichen, einen ersten Test ohne Zuzug von technischen Dienstleistern durchzuführen, was die Analyselast verringern und Verfahren beschleunigen würde. Gleichzeitig wäre sichergestellt, dass Bilder aufgrund der niedrigen Hürde häufiger einem solchen ersten Test unterzogen werden und nicht ungeprüft in das Verfahren eingehen. Insgesamt könnte so verhindert werden, dass realistisch aussehenden, aber gleichwohl synthetischen Bilddaten aufgrund fehlender Entdeckung von Anzeichen ein hoher Beweiswert zugesprochen wird. Eine minimale Überprüfung durch das Gericht würde zum Standard.

Leitfäden für Analyseprozess und Report

Die starke methodische Fragmentierung der forensischen Analyse und der Berichterstattung beeinträchtigt die Nachvollziehbarkeit, Überprüfbarkeit, Effizienz und Entlastung von Gericht und Gutachtern. Gerade durch die rapide Weiterentwicklung von generativen Modellen ist es wichtig, großflächige Mindeststandards einzuführen, die die Qualität der Analyse und damit von Gerichtsentscheidungen sicherstellen. Da erfahrungsgemäß jeder Ermittlungsfall individuell ist, wäre es kaum möglich oder zielführend, den kompletten Prozess zu standardisieren. Für einzelne, klar abtrennbare Prozessschritte mit konkreter Fragestellung bietet sich jedoch die Entwicklung von Richtlinien an. Die Frage nach der Authentizität aller Teile eines Bildes ließe sich mit einem Analyseprozess auf der Basis von Richtlinien einerseits und mit einem standardisierten Berichtformat andererseits einheitlich beantworten. Die methodischen Vorgaben können dabei mit unterschiedlichen Werkzeugen umgesetzt werden. Ein prinzipielles Vorgehen zur Bildanalyse könnte wie folgt aussehen:

1. Zuordnung zu einem Quellgerät, falls möglich
2. Analyse der Metadaten auf Inkonsistenzen
3. Prüfung der Konsistenz von physikbasierten Bildeigenschaften wie Beleuchtung, Schatten und Spiegelungen, Geometrie und Perspektive
4. Analyse von statistischen Spuren des Bildes wie Farb- und Rauscheigenschaften, um manuelle oder KI-basierte Bearbeitung aufzudecken

Dabei können die Punkte, die bearbeitet wurden, in einer Checkliste abgehakt und dokumentiert werden. Dies zeigt auch, welche Punkte nicht bearbeitet wurden (bspw. aufgrund fehlender Metadaten). Für die technische Dokumentation sind Prozess- und Datenmodelle hilfreich, wie sie bereits im "Leitfaden IT-Forensik" des deutschen Bundesamts für Sicherheit in der Informationstechnik (BSI)[33] existieren. Durch klar bekannte Prozesse und Strukturen können alle angewandten Analysen validiert und reproduziert werden, was für einen fairen Prozess unabdingbar ist. Ein Vorschlag für ein angepasstes Datenmodell auf der Basis der Richtlinien des BSI, das zur KI-gestützten Analyse von Bildern genutzt werden kann, wurde 2022 veröffentlicht.[37]

Für die Validierung der genutzten Werkzeuge, besonders für KI-basierte Software, können standardisierte Evaluationsgerüste verwendet werden, die beispielsweise einen gut definierten Anwendungsrahmen oder ein gewisses Maß an Erklärbarkeit sicherstellen sollen. Ein Beispiel für ein solches Rahmenwerk ist das sich in der Entwicklung befindliche AP4AI Framework[38] von Europol.

Kategorisierung von Bilddaten im Hinblick auf Authentizität

Die zunehmende Unschärfe der Grenze zwischen realen und synthetischen Bildern fordert eine feinere Abstufung zwischen unterschiedlichen Arten von Bilddaten. Neben realen Fotoaufnahmen und vollständig synthetisch generierten Bildern ist es sinnvoll, auch die Bildarten zu unterscheiden, die auf dem Spektrum dazwischen liegen. Hier bietet sich eine Kategorisierung an, die Informationen zur Einstufung der Verlässlichkeit des Inhaltes wiedergibt. Dabei könnten die Dateien auf Spuren verschiedener Veränderungen untersucht werden, die nach ihrem Einfluss auf die Bildaussage klassifiziert werden können.

Für die Aufbereitung von Bilddaten als Beweismittel existiert bereits ein Vorschlag, der verschiedene Bildbearbeitungsschritte aufteilt in „inhalts- und sinnwahrende Bildbearbeitung" und „inhalts- und sinnändernde Bildbearbeitung".[39] Unter die erste Kategorie fallen Anpassung des Kontrasts, der Schärfe und der Helligkeit des Bildes, unter die zweite Kategorie fällt das Austauschen, Hinzufügen oder Entfernen von Bildinhalten, die die Aussage semantisch verändern. Diese Idee der Kategorisierung von Bildbearbeitungsoperationen könnte genutzt werden, um Bilddaten für ihre Gewichtung in der Beweiswürdigung einzuordnen. Eine Kategorisierung kann beispielsweise wie im Folgenden beschrieben vorgenommen werden:

1. **Originalbild**: Unveränderte Sensordaten. Keine Auswirkungen auf Authentizität, Bild kann als Beweismittel genutzt werden.
2. **Farben leicht bearbeitet, semantisch integer:** Leichte Anpassungen, die auf Ästhetik oder bessere Erkennbarkeit abzielen wie Helligkeit, Farbstimmung, leichte Kontrastverbesserung, Schärfe. Bis auf Einzelfälle, in denen bspw. der Kontrast die Bildaussage verändern kann, keine Auswirkungen auf Authentizität.
3. **Ästhetisch leicht bearbeitet:** Angewandte Filter zur leichten Verbesserung der Ästhetik wie weichzeichnende Filter auf der Haut. Kaum Auswirkungen auf Semantik, aber vermutlich Löschung von forensischen Spuren im vom Filter betroffenen Bereich und damit potenziell Auswirkungen auf Authentizität.
4. **Starke Bearbeitung:** Verzerrung, Verschiebung, Ersetzung von Bildteilen, stärkere Filter, die das Bild mehr als oberflächlich verbessern oder verändern sollen, wie Austauschen des Himmels, Änderung des Gesichtsausdrucks und der Gesichtsform. Einfluss auf Authentizität, da in großen Bildbereichen foren-

36 Dazu Rückert (Fn. 11), 657.

37 Siegel et al., *Forensic Data Model for Artificial Intelligence based Media Forensics-Illustrated on the Example of DeepFake Detection*, Electronic Imaging 2022, 1 ff.

38 https://ap4ai.eu/reports/2022/02/ap4ai-framework-blueprint (zuletzt abgerufen am 15.07.2024).

39 Rabe, *Polizeiliche Fotografie: Anforderungen an das Beweismittel „digitales Foto"*, Die Kriminalpolizei 2020.

sische Spuren gelöscht oder stark verändert werden und Bildaussage in relevantem Maß verändert sein kann.

5. **Teils erzeugte Anteile oder unklare Bildechtheit:** Bildanteile, die durch Nutzung von KI generiert oder durch manuelle Methoden wie 3D-Rendering oder Zeichnung erzeugt wurden. Starker Einfluss auf Authentizität, da Bildaussage des realen Bildanteils stark verändert worden sein kann durch Hinzufügen oder Entfernen von Bildinhalten. Von der Verwendung zu Beweiszwecken wird abgeraten. Gleichwertig zu behandeln sind Bilder, die so stark komprimiert oder anderweitig von ungeeigneter Qualität sind, dass ihre Echtheit nicht mehr forensisch nachweisbar ist.
6. **Vollständig synthetisches Bild**: Durch Nutzung von KI oder manuellen Werkzeugen erzeugtes Bild. Ungeeignet als Beweismittel, wenn die Bildaussage Gegenstand der Untersuchungen ist.

Eine derartige Kategorisierung erlaubt es einerseits, forensische Gutachten einheitlicher und eindeutiger zu gestalten und das Ergebnis verständlicher zu machen, andererseits bewahrt es die Freiheit des Gerichts, frei und im Einzelfall den Beweiswert eines Bildes festzustellen.

Hochqualitative Datensätze

Seit der Veröffentlichung von DALL-E 2[40] erschienen in etwas mehr als einem Jahr DALL-E 3[5], mehrere Versionen der alternativen Anwendung Midjourney[6] und der Open-Source-Variante Stable Diffusion[7], sowie viele weitere Variationen generativer KI-Systeme. Ein schnelles Reagieren auf derartige Entwicklungen sowie aktuelle Datensätze der jeweiligen Generatoren mit ebenfalls aktuellen Kompressionen sind unabdingbar dafür, Strafverfolgung und Judikative rechtzeitig auf technische Herausforderungen vorzubereiten. Hierbei müssen die im AIA spezifizierten Anforderungen an Datensätze eingehalten werden: Datensätze sollen vollständig und repräsentativ sein.[41] Das heißt in diesem Kontext, dass Datensätze aus unterschiedlichsten Bildgeneratoren bestehen sollten sowie diversifizierte Bildmotive enthalten und zusätzlich mit Kompressionen gängiger Social Media-Anwendungen abgelegt werden, um die Realität von Bilddaten möglichst gut abzubilden.

Aktive Forensik

Eine grundsätzlich andere Lösungsrichtung für die Vertrauenswürdigkeit und Integrität von Bildern wäre eine Signatur, die jedes Bild eindeutig identifiziert. Bei einer Speicherung dieser Signatur kann ein untersuchtes Bild auf seine Signatur zurückgeführt und alle seit der Aufnahme geschehenen Änderungen nachverfolgt werden. Allerdings wird diese nicht so schnell einheitlich verfügbar sein. Leica stellte 2023 ein Konzept vor, das Bilder bereits in der Kamera mit einem Label ausstattet und alle Bearbeitungen aufzeichnet.[42] Sony kündigte 2023 ebenfalls ein Kameramodell an, das ein ähnliches Feature beinhalten würde.[43]

Für den Privatgebrauch erzeugen solche Signaturen jedoch zusätzliche Herausforderungen, da unklar ist, wie die Privatsphäre des Fotografen geschützt werden kann. Für den Einsatz in der Pressefotografie oder bei der Dokumentation durch Beamte könnten solche Ansätze allerdings helfen, den hohen Beweiswert von Bildern sicherzustellen.

7 Fazit

Die Verbreitung synthetischer Bilder und deren signifikanter Einfluss auf die Frage der Authentizität von Beweismitteln in Strafverfahren stellen Strafverfolgungsbehörden und Gerichte vor Herausforderungen. Eine durch das Gericht durchführbare Erstanalyse von Bilddaten könnte verhindern, dass synthetische Bilder ungeprüft in Verfahren eingehen. Nachvollziehbare, überprüfbare und verständlich dokumentierte Prozesse zur Analyse von Bilddaten sowie eine Kategorisierung von Bildern in Bezug auf ihre Authentizität tragen dazu bei, dass individuell sinnvolle und wohlinformierte Entscheidungen getroffen werden können. Langfristig ermöglichen gut gepflegte und aktuelle Datensätze, mit der technischen Evolution Schritt zu halten und eine Einbindung von Signaturen in Bilder ermöglicht verlässliches Prüfen der Bilddaten, wenn sie als Beweismittel genutzt werden. Aktive forensische Forschung im Bereich der Bildanalyse bringt für jede neue Herausforderung wertvolle Lösungsansätze hervor, und es liegt an Strafverfolgungsbehörden und Gerichten, diese aktiv zu nutzen, um den hohen Beweiswert von Bildern zu sichern. Auf der anderen Seite sind rechtsanwendende Behörden gefordert, ein vertieftes Verständnis der Authentizität von Bildern im Zeitalter von KI-Generatoren zu entwickeln und sich zu befähigen, entsprechende Analysen würdigend einordnen zu können.

Danksagung

Die hier dokumentierten Arbeiten wurden gefördert mit Mitteln der Deutschen Forschungsgemeinschaft (DFG) als Teil des Graduiertenkollegs 2475 „Cyberkriminalität und Forensische Informatik“ (Projektnummer 393541319/GRK2475/2-2024).

40 DALL-E 2 https://openai.com/dall-e-2. Zuletzt abgerufen am 15.07.2024

41 Artikel 10 Absatz 3 AIA (Fn. 30).

42 Content Credentials https://leica-camera.com/de-DE/fotografie/content-credentials#. Zuletzt abgerufen am 15.07.2024

43 In-camera photo forgery detection technology for corporate business users https://pro.sony/ue_US/solutions/forgery-detection. Zuletzt abgerufen am 15.07.2024