

KI, Synthetische Medien und die Transformation des öffentlichen Raums

Transparenzpflichten für synthetische Medieninhalte: Rechtliche Anforderungen und technische Realität

Moritz Griesel und Hans Brorsen

Zusammenfassung

Der Beitrag untersucht die Regulierung synthetischer Medieninhalte im europäischen Recht mit besonderem Fokus auf Kennzeichnungs- und Transparenzpflichten. Ausgangspunkt ist die zunehmende Verbreitung täuschend echt wirkender, KI-generierter Medieninhalte, deren schnelle Produktion und Reichweite über digitale Plattformen erhebliche Risiken für individuelle Rechtsgüter sowie für die Funktionsfähigkeit des öffentlichen Diskurses bergen. Vor diesem Hintergrund analysiert der Beitrag die einschlägigen Regelungen der KI-Verordnung (KI-VO), des Digital Services Act (DSA) und des Europäischen Medienfreiheitsgesetzes (EMFA). Zentrale These ist, dass diese Rechtsakte einer wirkungsbezogenen Regulierungslogik folgen: Kennzeichnungspflichten knüpfen nicht allein an den Einsatz bestimmter Technologien an, sondern an deren potenzielle Täuschungs- und Manipulationswirkungen auf Rezipientinnen und Rezipienten. Der Beitrag zeigt, dass eine wirkungsorientierte Ausgestaltung von Kennzeichnungspflichten entscheidend ist, um Vertrauen zu stabilisieren und die Integrität öffentlicher Kommunikationsräume zu sichern. Anhand ausgewählter Kennzeichnungsmethoden wird die praktische Umsetzung der Kennzeichnungspflichten exemplarisch beleuchtet.

1. Einleitung

Die Nutzung generativer KI-Systeme und die damit verbundene Erstellung und Verbreitung synthetischer Medieninhalte haben in den vergangenen Jahren erheblich an Bedeutung gewonnen.¹ Zunehmend verdichtet sich

¹ Als weiterer relevanter Anwendungsfall sind KI-Systeme zur direkten Interaktion mit natürlichen Personen zu nennen. Art. 50 Abs. 1 KI-VO verpflichtet Anbieter, solche Systeme so auszugestalten, dass Nutzerinnen und Nutzer über die Interaktion mit einem

dabei die Erkenntnis, dass diese Systeme nicht nur Inhalte mit erheblichem Nutzwert, sondern ebenso solche mit missbräuchlichem oder schädlichem Charakter hervorbringen können.² Ihr inhaltliches Spektrum erstreckt sich über sämtliche Medienformate hinweg und scheint seine Grenzen vor allem in der Kreativität sowie der technischen Kompetenz seiner Nutzerinnen und Nutzer zu finden. Obwohl KI-generierte Inhalte erst gegen Ende der 2010er Jahre in das Bewusstsein einer breiteren Öffentlichkeit getreten sind³, ist eine Unterscheidung zwischen authentischen und synthetischen Medienausgaben für menschliche Rezipienten inzwischen vielfach kaum noch möglich.⁴ Die Verbreitung synthetischer Medien ist damit längst kein randständiges Phänomen mehr, sondern hat sich zu einem dynamisch wachsenden Bestandteil moderner Informationsökosysteme entwickelt.

Der Beitrag nimmt die Regulierungsansätze aus KI-VO⁵, DSA⁶ und EMFA⁷ zum Anlass, um zunächst die unterschiedlichen Erscheinungsformen

KI-System informiert werden; im nationalen Recht wird dies durch § 18 Abs. 3 MStV ergänzt. Der Beitrag beschränkt sich im Folgenden auf künstlich generierte Medieninhalte. Zur Diskussion um Chatbots s. *Kern*, Die Verwendung von Social Bots, KIR 2024, 94 (95 ff.); *Schultz*, Aufsichtsrechtliches Vorgehen gegen Fake-Profile, MMR 2025, 168 (169 ff.); *Löber/Roßnagel*, Kennzeichnung von Social Bots, MMR 2019, 493.

- 2 S. beispielhaft die Berichterstattung zu synthetisch erstellten pornografischen Darstellungen auf X/Grok: *Brokhues*/Deutschlandfunk, „EU-Kommission startet Verfahren gegen Chatbot Grok“, 2026; allgemeiner *Crone*/Verfassungsblock, „Bildbasiert aber unsichtbar“, 2026; Grundlegend *Karaboga u. a.*, Deepfakes und manipulierte Realitäten, 2024.
- 3 *Stosz*/Georgetown Security Studies Review, Policy Options for Fighting Deepfakes, 2019; *Klass*, Manipulierte Realität: Wie Deepfakes Recht und Gesellschaft herausfordern, ZUM 2025, 485; *Lantwin*, Deep Fakes – Düstere Zeiten für den Persönlichkeitsschutz? MMR 2019, 574.
- 4 So geben etwa 91% der Deutschen an, wegen des Einsatzes von generativer KI entsprechende Schwierigkeiten zu haben, vgl. <https://www.tuev-verband.de/studien/chatgpt-studie-2025-generative-ki-wird-zum-massenphaenomen-aber-sicherheit-und-orientierung-gehen-fehlen>.
- 5 Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates 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), ABl. 144/1.
- 6 Verordnung (EU) 2022/2065 des Europäischen Parlaments und des Rates vom 19. Oktober 2022 über einen Binnenmarkt für digitale Dienste und zur Änderung der Richtlinie 2000/31/EG (Gesetz über digitale Dienste), ABl. L 277/1.
- 7 Verordnung (EU) 2024/1083 des Europäischen Parlaments und des Rates vom 11. April 2024 zur Schaffung eines gemeinsamen Rahmens für Mediendienste im Binnenmarkt

digitaler Medienmanipulation zu skizzieren (2.). Im Anschluss werden technik- und wirkungsbezogene Kennzeichnungsverständnisse voneinander abgegrenzt (3.) und die einschlägigen Kennzeichnungspflichten des europäischen Rechtsrahmens eingeordnet (4.). Im letzten Abschnitt werden technische Kennzeichnungsstandards exemplarisch beleuchtet (5.).

2. Heterogene Formen digitaler Manipulation

Synthetische Medieninhalte bezeichnen medial vermittelte Darstellungen – Bild-, Video-, Audio- oder Textinhalte –, die mithilfe generativer künstlicher Intelligenz erzeugt oder so manipuliert werden, dass sie den Eindruck authentischer Wirklichkeit vermitteln, obwohl sie einen verfälschten oder vollständig fiktiven Sachverhalt abbilden.⁸ Technologisch wird zur Erstellung solcher Deepfakes⁹ nicht auf eine singuläre Methode zurückgegriffen, sondern auf eine Vielzahl unterschiedlicher Verfahren, die in diversen Anwendungen eingesetzt werden. Getragen wird diese Entwicklung insbesondere durch die breite Zugänglichkeit leistungsfähiger generativer KI-Anwendungen wie DALL-E 2, Midjourney oder Stable Diffusion. Diese ermöglichen es auch technisch wenig versierten Nutzerinnen und Nutzern innerhalb kurzer Zeit hyperrealistische Inhalte zu erzeugen¹⁰ – m. a. W. ist die Produktion synthetischer Medieninhalte damit demokratisiert worden.¹¹ Gleichzeitig verschieben fortlaufende technische Verbesserungen die Grenzen dessen, was als authentisch wahrgenommen wird. Das Auftreten und der Umfang wirklichkeitsverzerrender und irreführender Effekte lassen sich wohl nicht anhand von Parametern wie dem Alter, einem spezifischen Bildungsgrad oder der individuellen Technikaffinität der Rezipientinnen

und zur Änderung der Richtlinie 2010/13/EU (Europäisches Medienfreiheitsgesetz), ABL L 37/1.

8 Zum Begriff s. *Kumkar/Rapp*, Deepfakes, ZfDR 2022, 199 (200); s. die Legaldefinitionen in Art. 3 Nr. 60 KI-VO und Art. 35 Abs. 1 S. 2 lit. k DSA.

9 Überblick über verfügbare Technologien bei *van Huijstee u. a.*, Tackling deepfakes in European policy, 2021, S. 7 ff.; *Maras/Alexandrou*, Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos, International Journal of Evidence & Proof 2019, 255; *Hartmann*, Der persönlichkeitsrechtliche Schutz vor Deepfakes, K&R 2020, 350.

10 *Molavi Vasse'i*, Watermarking von KI-generierten Inhalten als regulatorisches Instrument, RD 2024, 406 (406, Rn. 2 ff.) mwNachw.

11 *Kumkar/Griesel*, Transparenzpflichten für Deepfakes und synthetische Medieninhalte in der KI-VO, KIR 2024, 117 (118).

und Rezipienten eingrenzen.¹² Anwendungen zum Klonen von Stimmen¹³ oder zur digitalen Reproduktion real existierender Personen sind mittlerweile in der Lage, menschliche Sprachmuster, Mimik und Silhouetten überzeugend nachzuahmen – z. T. mit erheblicher Reichweite über digitale Plattformen.¹⁴

Mögliche Inhalte KI-generierter Ausgaben sind außerordentlich vielfältig: In vergleichsweise unproblematischen Kontexten werden etwa audiovisuelle Inhalte übersetzt oder synchronisiert¹⁵ oder historische Persönlichkeiten¹⁶ und andere virtuelle Doppelgänger digital (re-) konstruiert. Hinzu treten kreative Nutzungen wie die Generierung neuer Musikstücke, satirischer Darstellungen, Memes oder Parodien sowie die synthetische Neuschöpfung von Kunstwerken im Stil bekannter Künstlerinnen und Künstler. Auch in Bildungs- und Lernkontexten, etwa bei der individualisierten Aufbereitung von Lernmaterialien oder bei Sprachübungen, haben sich entsprechende Technologien etabliert.¹⁷ Darüber hinaus spielen generative KI-Systeme zunehmend eine Rolle in der Medienproduktion, im Journalismus, in der Werbung sowie in politischen Kommunikationsräumen, etwa im Rahmen von Wahlkämpfen¹⁸ oder gesellschaftlichen Debatten.¹⁹

Gleichzeitig sind mit der Verbreitung synthetischer Medien erhebliche Gefährdungspotenziale verbunden. Aus der Perspektive der betroffenen Personen stehen vor allem Beeinträchtigungen des allgemeinen Persönlichkeitsrechts im Vordergrund – häufig in Form von Ehrverletzungen, nachhaltigen Reputationsschäden oder der falschen Zuschreibung von Hand-

12 Kleine, „Perception of Deepfake Technology“, 2022, S. 79.

13 Engel-Bunsas, Recht an der eigenen Stimme in Zeiten von Deepfakes, RDi 2025, 292; Ellenberg, Persönlichkeits- und Urheberrechte beim Voice Cloning, RDi 2024, 599 (600); Klass, Manipulierte Realität: Wie Deepfakes Recht und Gesellschaft herausfordern, ZUM 2025, 485 (486).

14 S. die aktuellen Nachweise in D'Anastasio/Bloomberg, „KI-Chatbot Grok sorgt auf X für massenhafte Nacktbilder ohne Zustimmung der Betroffenen“, 2026.

15 Engel-Bunsas, Recht an der eigenen Stimme in Zeiten von Deepfakes, RDi 2025, 292; Ellenberg, Persönlichkeits- und Urheberrechte beim Voice Cloning, RDi 2024, 599 (600).

16 Oswalt/Deutschlandfunk, „Wie Kennedy eine Rede nach seinem Tod hielt“.

17 Beispiele in Klass, Manipulierte Realität: Wie Deepfakes Recht und Gesellschaft herausfordern, ZUM 2025, 485 (486).

18 Bentzen/European Parliamentary Research Service, „Briefing: Information manipulation in the age of generative artificial intelligence“, 2025, S. 3 mwNachw.

19 Ausf. Karaboga, in: Jaki/Steiger, „Die Regulierung von Deepfakes auf EU-Ebene“, Digitale Hate Speech, 2023, S. 201 ff.; Pawelec/Bieß, Deepfakes, 2021, S. 43 ff.; vgl. auch Beitrag von Munoz/Laumann in diesem Band.

lungen oder Eigenschaften.²⁰ Besonders drastisch treten diese Gefahren in Erscheinung bei nicht einvernehmlichen pornografischen Deepfakes²¹, bei KI-generierten Darstellungen sexualisierter Gewalt gegenüber Kindern²², bei digitalem Identitätsmissbrauch oder bei der Imitation von Stimmen zur Durchführung betrügerischer Handlungen. Diese Entwicklung wird dadurch begünstigt, dass bereits minimale Eingangsdaten – etwa einzelne Bild- oder Tonsequenzen – ausreichen, um überzeugende synthetische Darstellungen zu erzeugen.²³

Auf gesellschaftlicher und politischer Ebene betreffen die Risiken synthetischer Medien vor allem die Stabilität deliberativer Prozesse sowie die Funktionsfähigkeit des öffentlichen, digitalen Informationsraums.²⁴ Die durch die einfache Erstellung und schnelle Zirkulation beförderte Dynamik stellt den öffentlichen Diskurs vor signifikante Herausforderungen, da die zunehmende Verbreitung authentisch wirkender Inhalte u. a. die gemeinsame Tatsachengrundlage untergräbt, auf die demokratische Meinungsbildungsprozesse angewiesen sind.²⁵ Als Konsequenz kann das Vertrauen in Medien, staatliche Institutionen und politische Akteure verloren gehen. Daneben sind auch wirtschaftliche Schutzgüter²⁶ betroffen – etwa, wenn

-
- 20 Kumkar/Rapp, Deepfakes, ZfDR 2022, 199 (203 ff.); Klass, Manipulierte Realität: Wie Deepfakes Recht und Gesellschaft herausfordern, ZUM 2025, 485 (487); Völzmann, Dein Körper, meine Bilder? Der Schutz vor bildbasierter sexualisierter Gewalt (auch) in der Gewaltschutzrichtlinie der EU, ZUM 2025, 1 (2 ff.).
 - 21 Nachweise in Kumkar/Griesel, Transparenzpflichten für Deepfakes und synthetische Medieninhalte in der KI-VO, KIR 2024, 117 (118); so veröffentlichte etwa Grok vom 5.-6.1.2026 pro Stunde 6.659 sexualisierte Deepfakes, vgl. D'Anastasio/Bloomberg, „KI-Chatbot Grok sorgt auf X für massenhafte Nacktbilder ohne Zustimmung der Betroffenen“.
 - 22 Ausführlich Steinebach, KI-generierte Abbildungen von Kindesmissbrauch, DuD 2024, 304.
 - 23 Home Security Heroes, State of Deepfakes, 2023; zum perspektivischen Nutzungsverhalten s. Ernst/Heise, „Jim Keller über KI: <In 10 Jahren ist alle Software weg>“, 2024.
 - 24 Karaboga, in: Jaki/Steiger, „Die Regulierung von Deepfakes auf EU-Ebene“, Digitale Hate Speech, 2023, S. 204 ff.; Pawelec/Bieß, Deepfakes, 2021, S. 44.
 - 25 Kumkar/Griesel, Transparenzpflichten für Deepfakes und synthetische Medieninhalte in der KI-VO, KIR 2024, 117 (118); Molavi Vasse'i, Watermarking von KI-generierten Inhalten als regulatorisches Instrument, RD 2024, 406 (406, Rn. 4); Kumkar/Rapp, Deepfakes, ZfDR 2022, 199 (201); Kumkar, Deepfakes – Risiken und Regulierung im europäischen Verordnungsentwurf für künstliche Intelligenz, K&R-Beil. zu Heft 10/2023, 32 (34).
 - 26 Zur Haftung für KI-generierte geschäftsschädigende Falschinformationen s. LG Kiel, Urteil vom 29.2.2024 – 6 O 151/23; zu Lizenzgebühren für Synchronsprecher s. LG Berlin II Urteil vom 20.8.2025 – 2 O 202/24; zu vermögenswerten Bestandteilen des

Unternehmen durch synthetische Inhalte geschädigt²⁷ oder Vermögenswerte durch täuschungsbasierte Kommunikationsformen gefährdet werden.

3. Technik- oder wirkungsbezogenes Verständnis von Transparenz

Als Reaktion auf die identifizierten Herausforderungen haben sowohl die nationalen als auch der europäische Gesetzgeber ihre regulatorischen Aktivitäten zuletzt deutlich ausgeweitet und einen umfassenden Regulierungsrahmen geschaffen.²⁸ Entsprechende Transparenzmechanismen versprechen zwar grundsätzlich einen von der individuellen technischen Kompetenz unabhängigen Einfluss auf die Glaubwürdigkeit und Genauigkeit von Inhalten.²⁹ Dabei ergeben sich jedoch Unterschiede je nach gewähltem Transparenzansatz.³⁰

Kennzeichnungen von Medieninhalten könnten zunächst an deren Entstehungsprozess anknüpfen und damit das Ziel verfolgen, sämtliche KI-generierten Ausgaben als solche – also unter Nutzung einer bestimmten Technologie erstellten Inhalte – zu deklarieren. Die Kennzeichnung steht damit unabhängig vom konkreten Inhalt oder den mit dessen Generierung und Verbreitung beabsichtigten Wirkungen. Dies gilt vor allem für die maßgeblichen Transparenzvorschriften der KI-VO, die dem (system- und produktbezogenen) Produktsicherheitsrecht zuzuordnen sind. Zwar ist keiner der dortigen Regelungen ein eindeutig technikbezogenes Transparenzverständnis zu entnehmen; Ansätze hierfür finden sich jedoch an einigen Stellen.³¹

Diesem technikbezogenen Verständnis von Transparenz steht indes die mit der Einführung von Kennzeichnungs- und Offenlegungspflichten häufig verbundene gesetzgeberische Motivation entgegen.³² Anlass für regulatorisches Tätigwerden war weniger die bloße Tatsache maschineller Generierung als vielmehr das Risiko, dass bestimmte Inhalte geeignet sind, Rezi-

Persönlichkeitsrechts s. Schertz, Postmortales Persönlichkeitsrecht in Zeiten von KI und Deepfakes, NJW 2025, 2801 (2803 f.).

27 LG Kiel, Urteil vom 29.2.2024 – 6 O 151/23.

28 Art. 50 Abs. 2, 4 KI-VO; Art. 35 Abs. 1 S. 2 lit. k DSA; Art. 18 Abs. 1 lit. e EMFA.

29 Kleine, „Perception of Deepfake Technology“, 2022.

30 Epstein u. a., „What label should be applied to content produced by generative AI?“, 2023, S. 2.

31 S. unten unter 4.I.

32 S. dazu im Einzelnen unten unter 4.

pientinnen und Rezipienten in die Irre zu führen und dadurch individuelle oder gesellschaftliche Schutzgüter zu beeinträchtigen. Diesen Schutz bieten Kennzeichnungs- und Offenlegungspflichten, die stärker wirkungsbezogen ausgestaltet sind. Hier tritt der Entstehungsprozess in den Hintergrund; entscheidend ist vielmehr Inhalte zu identifizieren, die ein besonderes Täuschungs- oder Manipulationspotenzial aufweisen. Das Problem künstlich generierter Inhalte ist damit besser im Kontext bekannter Erscheinungsformen missbräuchlicher Inhalte wie Fehl-, Mis- und Desinformation zu verorten, das durch den hohen Grad an Authentizität synthetischer Medieninhalte nur verstärkt wird.³³ Wirkungsbezogene Transparenzpflichten zielen damit *primär* auf den Schutz übergeordneter öffentlicher Interessen wie Authentizität, Richtigkeit und Verlässlichkeit von Medieninhalten. Die Vertrauenswürdigkeit einer Information soll dabei von ihrer nachvollziehbaren Erstellung bis zu ihrer Zirkulation, Nutzung und Bearbeitung über Plattform- und andere Systeminfrastrukturgrenzen hinweg gewährleistet sein. Methoden und Wirkmechanismen dieser Interventionsstrategie gegen den Verlust des Vertrauens in das Informationsökosystem wurden schließlich auch vielfach empirisch untersucht.³⁴

-
- 33 Zu den Schutzgütern der im DSA vorgegebenen Anti-Desinformationsmaßnahmen s. *Peukert*, Desinformationsregulierung in der EU - Überblick und offene Fragen, JZ 2023, 278 (281) mwNachw.; *Gerdemann/Spindler*, Das Gesetz über digitale Dienste (Digital Services Act) (Teil 2), GRUR 2023, 115 (123). Dementsprechend verfolgt der DSA grundsätzlich auch einen technikneutralen Ansatz, sodass auch ein ohne KI erzeugter, etwa durch irreführende Schnitte, falsche Kontextualisierung oder manipulative Beschriftung verfälschter Medieninhalt, der eine vergleichbare Gefährdungslage begründet wie ein vollständig synthetisch generiertes Medium, einer Kennzeichnungspflicht unterliegt; Umgekehrt die eine Kennzeichnungspflicht nicht exklusiv der Prävention und Kontextualisierung solcher Medieninhalte, sondern erfüllt daneben noch weitere Funktionen, vgl. *Becker*, Generative KI und Deepfakes in der KI-VO, CR 2024, 353 (360, Rn. 60); *Becker*, Das Urheberrecht als Trostpreis für den Menschen?, GRUR 2024, 505 (507).
- 34 Zur Terminologie von KI-Labels s. *Epstein u. a.*, „What label should be applied to content produced by generative AI?“, 2023; *Rauchfleisch u. a.*, „Deepfakes or Synthetic Media? The Effect of Euphemisms for Labeling Technology on Risk and Benefit Perceptions“, 2025; *Altay/Gilardi*, People are skeptical of headlines labeled as AI-generated, even if true or human-made, because they assume full AI automation, 2024; Zu KI-Labels in einzelnen Branchen s. *Toff/Simons*, „Or They Could Just Not Use It?\": The Dilemma of AI Disclosure for Audience Trust in News, The International Journal of Press/Politics 2025, 881-903; *Li/Yang*, Impact of Artificial Intelligence-Generated Content Labels On Perceived Accuracy, Message Credibility, and Sharing Intentions for Misinformation, 2024; *Wittenberg u. a.*, Labeling AI-generated media online, 2025; *Gamage u. a.*, Labeling Synthetic Content: User Perceptions of Warning Label

Der Anteil synthetisch generierter Inhalte im Internet ist bereits jetzt erheblich³⁵ und wird auch zukünftig weiter zunehmen. Bereits angesichts der enormen Menge an Medieninhalten stellt sich die Frage, wie eine grundsätzlich wünschenswerte wirkungsbezogene Transparenzkennzeichnung in der Praxis effektiv umgesetzt werden kann. Vor diesem Hintergrund ist auch in Betracht zu ziehen, authentische bzw. menschlich-erstellte Inhalte ebenfalls einer Kennzeichnungspflicht zu unterziehen³⁶ – zumindest soweit es sich um Medieninhalte handelt, an deren menschlicher Erstellung ein besonderes Interesse³⁷ besteht. Dies hätte den positiven Effekt, dass diese Inhalte leichter auffindbar und identifizierbar sind und eine höhere Wahrnehmung von Vertrauenswürdigkeit und Werthaltigkeit erfahren können.³⁸

4. Europäische Kennzeichnungs- und Offenlegungspflichten im Überblick

Die Regulierung der Nutzung generativer KI-Systeme wird in unterschiedlichen Rechtsgebieten³⁹ nachvollzogen: Während etwa in Dänemark die Einführung eines Urheberrechts an der individuellen Erscheinung einer

Designs for AI-generated Content on Social Media, 2025; Lockhart/Tessono, „Human or AI? Evaluating Labels on AI-Generated Social Media Content, 2025; zur Wirkung von ampelfarbenbasierten Warnhinweisen s. Dobber u. a., The effect of traffic light veracity labels on perceptions of political advertising source and message credibility on social media, 2025.

35 Spennemann, „Delving into: the quantification of Ai-generated content on the internet“, 2025.

36 Newmann/Reuters Institute, „Journalism and Technology Trends and Predictions 2026“, 2026.

37 Ähnlich Becker, Generative KI und Deepfakes in der KI-VO, CR 2024, 353 (360 Rn. 61, 364 ff. Rn. 96 ff.).

38 Campante u. a., GenAI Misinformation, Trust, And News Consumption: Evidence From A Field Experiment, 2025.

39 Zum Produktsicherheitsrecht: Kumkar/Griesel, Transparenzpflichten für Deepfakes und synthetische Medieninhalte in der KI-VO, KIR 2024, 117; Molavi Vasse'i, Watermarking von KI-generierten Inhalten als regulatorisches Instrument, RD i 2024, 406; Zum Strafrecht: Pfeiffer, Tatbestandsmerkmale, Rechtfertigungsgründe und Herausforderungen beim strafrechtlichen Schutz vor Deepfakes, MMR 2025, 934; Blocher, Strafbare Deepfakes als Täuschung über die Authentizität von Medieninhalten, KIR 2025, 225; Zum Deliktsrecht: Kumkar/Rapp, Deepfakes, ZfDR 2022, 199 (204 ff.); Lantwin, Deep Fakes – Düstere Zeiten für den Persönlichkeitsschutz, MMR 2019, 574.

natürlichen Person erwogen wird⁴⁰, hat Italien nach einer Reihe öffentlichkeitswirksamer pornografischer Deepfakes über Personen des öffentlichen Lebens sein Strafrecht verschärft. In Deutschland steht die Verschärfung strafrechtlicher Sanktionsmechanismen nach einer Bundesrats-Initiative des Freistaates Bayern⁴¹ aktuell genauso zur Debatte wie Verbesserungen in der Durchsetzung zivilrechtlicher Ansprüche auf Beseitigung, Unterlassung und Schadensersatz im Gesetzentwurf gegen digitale Gewalt.

Auf unionsrechtlicher Ebene stellen schließlich die im öffentlichen Produktsicherheits-, Medien- und Plattformrecht zu verortenden Transparenzpflichten für unterschiedliche Ausgaben von KI-Systemen einen nicht nur mitgliedsstaats-, sondern auch system- und inhaltsübergreifenden Ansatz dar. Die Transparenzpflichten knüpfen an unterschiedlichen Zeitpunkten im Lebenszyklus eines generativen KI-Systems bzw. eines synthetisch erzeugten Medieninhaltes an: Beginnend mit der Herstellung des KI-Systems über die Herstellung des synthetischen Medieninhalts bis zu dessen Zirkulation auf Online-Plattformen stellen die entsprechenden Vorschriften einen umfassenden Regulierungsrahmen dar.

4.1 Verordnung über künstliche Intelligenz (KI-VO)

Die Regelungen in Art. 50 KI-VO dienen dabei gewissermaßen als rechtsaktübergreifender Ausgangspunkt. In Art. 1 Abs. 1 und Erwägungsgrund 1 KI-VO werden mit dem Schutz vor schädlichen Auswirkungen von KI-Systemen in der EU und der Entwicklung von menschenzentrierten und vertrauenswürdigen KI-Systemen zentrale Regulierungsgründe benannt.

40 Der dänische Vorschlag scheint aus deutscher Perspektive besser i.R.d. nationalen Persönlichkeitsrecht dogmatik verortbar. Der dänische (Um-)Weg über das Urheberrecht wirft initial schon die Frage auf, wie der beabsichtigte Schutz „realistic digitally generated imitations of personal characteristics“ überzeugend unter den urheberrechtlichen Werkbegriff einer persönlichen geistigen Schöpfung eingeordnet werden kann. Vgl. dazu *Karttunen*, European Parliamentary Research Service, „The Danish approach to copyright and deepfakes: A model for the EU?“, 2026.

41 BT-DruckS. 222/24; dazu der Antrag der FDP-Fraktion im Landtag Nordrhein-Westfalen, LT NRW DruckS. 18/10528, wo zusätzlich detaillierte Vorschriften zur Kennzeichnung von Deepfakes verlangt werden.

Spezifischer sieht Art. 50 Abs. 2 KI-VO vor, dass die Ausgaben eines KI-Systems in einem maschinenlesbaren⁴² Format⁴³ „als künstlich erzeugt oder manipuliert“ gekennzeichnet werden müssen.⁴⁴ In ähnlicher Weise sehen auch die Transparenzpflichten für die Generierung von Deepfakes in Art. 50 Abs. 4 KI-VO und für die Zirkulierung synthetischer Medieninhalte in Art. 35 Abs. 1 S. 2 lit. k DSA vor, dass Inhalte als künstlich erzeugt oder manipuliert offengelegt werden müssen. Zumindest der KI-VO geht es dabei explizit um die Differenzierung zwischen von Menschen erzeugten und synthetisch generierten Inhalten.⁴⁵

Der Anwendungsbereich von Art. 50 Abs. 4 UAbs. 1 KI-VO wird nicht allein aufgrund der künstlichen Erzeugung eines Inhalts eröffnet, sondern setzt zusätzlich voraus, dass der Medieninhalt eine wirklichkeitsverzerrende oder wirklichkeitsimitierende Wirkung entfaltet – mithin einen Deepfake⁴⁶ darstellt. Art. 50 Abs. 4 UAbs. 2 KI-VO stellt darüber hinaus auf den Kontext der Veröffentlichung ab, indem er insbesondere Textinhalte erfasst, die zur Information der Öffentlichkeit über Angelegenheiten von öffentlichem Interesse verbreitet werden.

Weitergehende kontextvermittelnde Anforderungen an die Kennzeichnung sind den Vorschriften zunächst nicht zu entnehmen; insbesondere ist weder Art. 50 Abs. 4 KI-VO noch Art. 50 Abs. 2 KI-VO die Pflicht zur Angabe des Namens des Betreibers oder Anbieters eines KI-Systems zu entnehmen.⁴⁷ Die Entscheidung des Unionsgesetzgebers gegen eine entsprechend weite Verpflichtung wurde bereits in Zusammenhang mit der Generierung und Verbreitung persönlichkeitsrechtsverletzender Deepfakes kritisiert.⁴⁸

42 Zum Begriff s. *Brorsen/Falk*, Die „Maschinenlesbarkeit“ von Informationen, MMR 2025, 7ff.

43 Zur Frage ob auch ein menschenlesbares Format erforderlich ist, *Kumkar/Griesel*, Transparenzpflichten für Deepfakes und synthetische Medieninhalte in der KI-VO, KIR 2024, 117 (124).

44 *Kumkar/Griesel*, Transparenzpflichten für Deepfakes und synthetische Medieninhalte in der KI-VO, KIR 2024, 117 (122 ff.).

45 Erwägungsgrund 133 S. 1 KI-VO; Vergleichbare Erwägungen fehlen im DSA, da sich kein spezifischer Erwägungsgrund mit Art. 35 Abs. 1 S. 2 lit. k DSA beschäftigt.

46 Zum Begriff s. Art. 3 Nr. 60 KI-VO; *Meding/Sorge*, What constitutes a deepfake?, CSLAW '25: Symposium on Computer Science and Law, 2025.

47 *Kumkar/Griesel*, Transparenzpflichten für Deepfakes und synthetische Medieninhalte in der KI-VO, KIR 2024, 117 (122); Zur entsprechenden Kritik am Kommissionsentwurf, *Hinderks*, Die Kennzeichnungspflicht von Deepfakes, ZUM 2022, 110 (113).

48 *Kumkar/Griesel*, Transparenzpflichten für Deepfakes und synthetische Medieninhalte in der KI-VO, KIR 2024, 117 (122); *Hinderks*, Die Kennzeichnungspflicht von Deepfakes, ZUM 2022, 110 (113).

Doch auch jenseits dieser spezifischen Inhalte-Kategorie hätten entsprechende kontextvermittelnde Informationen belastbare Anknüpfungspunkte zur Einordnung und Kontextualisierung von Medieninhalten geliefert.

Das wirkungsbezogene Verständnis wird jedoch durch die Erwägungsgründe der KI-VO weiter untermauert. Die Kennzeichnungs- und Offenlegungspflichten stehen etwa in – wirkungsbezogenen – Erwägungen zentraler Risiken wie Identitätsbetrug, Täuschung und anderen schädlichen Auswirkungen auf Rezipienten⁴⁹ – nicht aber technikbezogenen Risiken und Gefahren auf Systemebene⁵⁰. Zugleich sind die Integrität des Informationsökosystems und das hierin gesetzte Vertrauen angesichts von Fehlinformation, Manipulation in großem Maßstab, Betrug und Täuschungshandlungen erheblichen Gefahren ausgesetzt.⁵¹ Auch unter Berücksichtigung der Anforderungen an die einzuführenden Abhilfemechanismen, wofür als Anknüpfungspunkte u. a. die Herkunft oder Echtheit von Inhalten benannt werden, scheint ein wirkungsbezogenes Verständnis naheliegender.⁵²

Die sekundärrechtlichen Vorgaben der KI-VO wurden zuletzt i. R. d. Entwurfs für einen Verhaltenskodex über die Transparenz KI-generierter Inhalte⁵³ (CoP-E) aufgegriffen. Die Vereinbarung konkreter Kennzeichnungs- und Offenlegungsmechanismen steht dabei erneut in Erwägung des engen Zusammenhangs zwischen synthetischer Generierung von Medieninhalten und den damit verbundenen Folgen für die Integrität und das Vertrauen in das Informationsökosystem.⁵⁴ Vorgesehen ist die Bereitstellung strukturierter, automatisiert auslesbarer Informationen über Herkunft, Entstehungs- und Bearbeitungsprozesse eines Inhalts. Transparenz wird dabei nicht allein als Information für Endnutzerinnen und -nutzer verstanden, sondern zugleich als infrastrukturelle Voraussetzung für die Verbreitung oder Nutzung von Medieninhalten auf Plattformen, Suchmaschinen und durch andere Intermediäre – etwa zur Bewertung der Vertrauenswürdigkeit von Inhalten oder zur Unterstützung weiterer Risikominderungsmaßnahmen. Ergänzend verweist der CoP-E auf den Einsatz standardisierter Pro-

49 ErwGr. 132 S. 1 KI-VO.

50 Im risikobasierten Ansatz der KI-VO sind entsprechende Compliance-Maßnahmen nur für Hochrisiko-KI-Systeme vorgesehen, vgl. Kapitel III Abschnitt 2 KI-VO.

51 ErwGr. 133 S. 2 KI-VO.

52 ErwGr. 133 S. 3 KI-VO; ErwGr. 135 KI-VO.

53 Second Draft Code of Practice on Transparency of AI-Generated Content, 03.03.2026.

54 Second Draft Code of Practice on Transparency of AI-Generated Content, 03.03.2026, S. 7.

venienzsysteme („Content Credentials“)⁵⁵, die solche Informationen in interoperablen Formaten bündeln, eine plattformübergreifende Nachvollziehbarkeit der Erstellung und Bearbeitung digitaler Inhalte ermöglichen und regelmäßig durch kryptografische Sicherungsmechanismen gegen Manipulation geschützt sind. Durch diese Kombination aus Maschinenlesbarkeit, Standardisierung und technischer Absicherung sollen Provenienzinformationen dauerhaft mit dem jeweiligen Inhalt verknüpft, überprüfbar gehalten und damit als funktionales Instrument zur Stärkung der Integrität des Informationsökosystems nutzbar gemacht werden.

4.2 Gesetz über digitale Dienste (DSA)

Darüber hinaus sind wirkungsbezogene Transparenzpflichten als Katalogmaßnahme zur Minderung systemischer Risiken auf „sehr großen Online-Plattformen und sehr großen Online-Suchmaschinen“ (VLOPSEs)⁵⁶ in Art. 35 Abs. 1 S. 2 lit. k DSA enthalten. Die Kennzeichnung kann als Risikominderungsmaßnahme ergriffen werden, nachdem i. R. e. plattformseitig durchzuführenden Risikobewertung ein systemisches Risiko i. S. v. Art. 34 DSA auf der Plattform festgestellt wurde. Zu den erfassten Risikokategorien⁵⁷ zählen die Verbreitung rechtswidriger Inhalte, eine nachteilige Auswirkung auf demokratische Prozesse, gesellschaftliche Debatten und Wahlprozesse sowie geschlechtsspezifische Gewalt.⁵⁸ Weitergehende besondere Gefährdungslagen können sich zudem aus den Funktionsweisen einer Plattform ergeben – etwa durch Nutzung von Reichweitenmechanismen, algorithmischer Verstärkung oder strukturellen Anreizen zur massenhaften Verbreitung rechtswidriger Inhalte.⁵⁹ Doch auch jenseits dieser Kategorien systemischer Risiken zielt der DSA auf eine wirkungsbezogene Regulierung, die in allgemeinen Schutzzweckbestimmungen wie einem sicheren,

55 S. im Einzelnen unten unter 5.

56 Very Large Online Platforms and Search Engines i.S.v. Art. 34 DSA.

57 Z. T. werden diese Bereiche auch als Regelbeispiele verstanden, vgl. v. *Lewinski/Kling*, Compliance-Pflichten nach dem DSA – ein umfassender Überblick, RD 2023, 183 (190).

58 Art. 34 Abs. 1 DSA; *Peukert*, Desinformationsregulierung in der EU - Überblick und offene Fragen, JZ 2023, 278 (281) erklärt daher, dass das vom DSA vorgegebene „Desinformationsregime nicht dem Schutz individueller Rechtsgüter bestimmter Einzelpersonen dient, sondern auf einer abstrakteren Ebene dem Schutz öffentlicher Güter/Interessen“.

59 ErwGr. 79 S. 1 DSA; ErwGr. 84 S. 3 DSA.

vorhersehbaren, berechenbaren und vertrauenswürdigen Online-Umfeld zum Ausdruck kommt.⁶⁰

Dies spiegelt sich auch in den DSA-spezifischen Leitlinien⁶¹ und einem Verhaltenskodex⁶² wider. Erneut ist der Anwendungsbereich der Regelungen nicht nur allgemein in Bezug auf KI-generierte Inhalte eröffnet; vielmehr bedarf es (zusätzlich) eines täuschenden, falschen oder irreführenden Charakters dieser Inhalte.⁶³ Maßnahmen zur Reduzierung damit verbundener Risiken bei Wahlen⁶⁴ und Desinformationen⁶⁵ beinhalten u. a. die Einführung von Faktenprüfhinweisen, Vertrauenssiegeln und Werkzeugen oder anderen Informationen zur Bewertung der Provenienz und Editierhistorie, um die Vertrauenswürdigkeit, Integrität und Echtheit digitaler Inhalte effektiv bewerten zu können.⁶⁶ Kennzeichnungspflichten dienen in diesem Rahmen weniger der Offenlegung des technischen Entstehungsprozesses als vielmehr der Befähigung von Nutzerinnen und Nutzern, die Vertrauenswürdigkeit und Aussagekraft von Medieninhalten eigenständig zu bewerten.⁶⁷

Für die Zirkulation synthetischer Medieninhalte bestehen zudem mit Art. 35 DSA auf der regulierungsintensivsten Ebene der abgestuften Pflicht-

60 Art. 1 Abs. 1 DSA; ErwGr. 3, 9 DSA.

61 Leitlinien der Kommission für Anbieter sehr großer Online-Plattformen und sehr großer Online-Suchmaschinen zur Minderung systemischer Risiken in Wahlprozessen gemäß Artikel 35 Absatz 3 der Verordnung (EU) 2022/2065, C/2024/3014.

62 Zur Vereinbarkeit mit den Anforderungen an Verhaltenskodizes nach Art. 45 DSA s. Commission Opinion of 13.2.2025 on the assessment of the Code of Practice on Disinformation within the meaning of Article 45 of Regulation 2022/2065, C(2025) 1008 final; *Kollmann*, Zur Bedeutung des CoP on Disinformation 2022 im Kontext von KI-generierten Inhalten, MMR-Aktuell 2023, 458365.

63 Leitlinien der Kommission für Anbieter sehr großer Online-Plattformen und sehr großer Online-Suchmaschinen zur Minderung systemischer Risiken in Wahlprozessen gemäß Artikel 35 Absatz 3 der Verordnung (EU) 2022/2065, C/2024/3014, S. 1, 6, 8, 9; Code of Conduct on Disinformation 2022, S. 4.

64 Leitlinien der Kommission für Anbieter sehr großer Online-Plattformen und sehr großer Online-Suchmaschinen zur Minderung systemischer Risiken in Wahlprozessen gemäß Artikel 35 Absatz 3 der Verordnung (EU) 2022/2065, C/2024/3014, S. 7.

65 Commitment 20 Code of Conduct on Disinformation 2022 benennt beispielhaft den Industriestandard C2PA als taugliche Umsetzungsmaßnahme; dazu *Brorsen/Falk*, Neue Compliance-Pflichten nach dem Digital Services Act, MMR 2024, 32 (35 f.).

66 S. Commitments 21, 22 Code of Conduct on Disinformation 2022.

67 S. dazu auch die vorgeschlagenen Initiativen zur Förderung der Medienkompetenz, vgl. Leitlinien der Kommission für Anbieter sehr großer Online-Plattformen und sehr großer Online-Suchmaschinen zur Minderung systemischer Risiken in Wahlprozessen gemäß Artikel 35 Absatz 3 der Verordnung (EU) 2022/2065, C/2024/3014, S. 6.

tenstruktur des DSA verschärfte Sorgfalts- und Transparenzpflichten für VLOPs und VLSEs zur Bekämpfung systemischer Risiken. Dazu gehört nach Art. 35 Abs. 1 S. 2 lit. k DSA auch eine plattformseitig durchzuführende Kennzeichnung synthetisch erzeugter Medieninhalte. Ergänzend ist den Nutzern der Plattform eine benutzerfreundliche Funktion zur Verfügung zu stellen, entsprechende Inhalte zu melden.

4.3 Europäisches Medienfreiheitsgesetz (EMFA)

Mittelbar entfalten die Vorgaben der KI-VO und des DSA auch Wirkung im EMFA, wo zugunsten von Mediendiensteanbietern im Sinn von Art. 2 Nr. 2 EMFA ein Medienprivileg für die Verbreitung ihrer Inhalte auf VLOPSEs vorgesehen ist: Nach Art. 18 Abs. 1 lit. e EMFA sollen sie eine Erklärung („Selbstdeklaration“⁶⁸) über die Art und Weise der eigenen Inhalteerstellung abgeben.⁶⁹ Bezugspunkt dieser Erklärung ist nach dem Wortlaut der Verordnung die Frage, ob der Mediendiensteanbieter Inhalte bereitstellt, die von Systemen künstlicher Intelligenz erzeugt wurden, ohne dass diese einer menschlichen Überprüfung und redaktionellen Kontrolle unterzogen worden sind.⁷⁰ An diese Erklärung knüpft der EMFA schließlich Erleichterungen für entsprechend notifizierte Mediendiensteanbieter im Melde- und Abhilfeverfahren nach Art. 16 DSA – insbesondere indem den Entscheidungen über die Löschung oder Beschränkung der Sichtbarkeit eines Medieninhalts zwingend ein Stellungnahmeverfahren vorzuschalten ist – sogenannte „Must-carry-Verpflichtung“.⁷¹ In Bezug auf die durchzuführende Notifizierung dürfen sich VLOPSEs zudem Zertifizierungen Dritter über die Einhaltung der Anforderungen aus Art. 18 EMFA bedienen. Mit der *Journalism Trust Initiative* wird dabei explizit eine Stelle benannt, die mit CWA 17493:2019 (E) einen maschinenlesbaren und neutralen Standard für die globale Zertifizierung vertrauenswürdiger Medien geschaffen hat. Unter „Section B: Professionalism and Accountability, Nr. 10.6 Automatically Generated Content“ des Standards werden die Adressaten der standardi-

68 Achenleiter, Das Medienprivileg im European Media Freedom Act, MMR 2025, 89 (91).

69 Art. 18 Abs. 1 lit. e EMFA.

70 Art. 18 Abs. 1 lit. e EMFA.

71 Achenleiter, Das Medienprivileg im European Media Freedom Act, MMR 2025, 89 (92 f.); Hennemann, Der European Media Freedom Act oder: Von Zielen und Mitteln des Medienrechts, ZEuP 2024, 498 (501).

sierten Norm verpflichtet, ihre Nachrichteninhalte, die ganz oder teilweise automatisch durch algorithmische Verfahren generiert werden, eindeutig als solche zu kennzeichnen. Eine allein am Wortlaut orientierte Auslegung würde auch insoweit für eine rein am Entstehungsprozess anknüpfende Kennzeichnung sprechen.

4.4 Persönlichkeitsrechtliche Erwägungen

Die genannten Transparenzpflichten sind schließlich für solche Medieninhalte, in denen individuelle Rechtspositionen – wie das allgemeine Persönlichkeitsrecht einschließlich seiner speziellen Ausprägungen im Recht am eigenen Bild oder dem Recht an der eigenen Stimme – betroffen sind, nur von begrenzter Bedeutung. Der durch Transparenz über die Inhalteerstellung vermittelbare Schutz scheint hier gegenüber den zuvor genannten Informationen mit meinungsbildendem Charakter oder von wirtschaftlicher Bedeutung deutlich beschränkter⁷² – immerhin verhindern Kennzeichnungen originär keinen digitalen Missbrauch von – vornehmlich – Frauen und Kindern jenseits des Körperlichen⁷³; dennoch ermöglichen sie die Einordnung eines ansonsten wirklichkeitsgetreu erscheinenden Inhalts als wirklichkeitsverzerrend und dienen damit auch in diesen Kontexten als Gradmesser zwischen Wirklichkeit und Fiktion. Solche Erwägungen könnten zumindest zivilrechtlich – etwa in Bezug auf die Bestimmung der Eingriffsintensität einer Rechtsgutverletzung – Berücksichtigung finden.

72 Ähnlich in Bezug auf Schadensersatzansprüche *Gerdemann/Spindler*, Das Gesetz über digitale Dienste (Digital Services Act) (Teil 2), GRUR 2023, 115 (123); s. die empirischen Erkenntnisse zum Interaktionsverhalten von Rezipienten mit unterschiedlichen Inhalten, die als KI-generiert gekennzeichnet sind, vgl. *Game u. a.*, Labeling Synthetic Content: User Perceptions of Warning Label Designs for AI-generated Content on Social Media, 2025.

73 Zum Begriff und verbreiteten Phänomenen s. *Burghardt/Schmidt/Steinl*, "Sexuelle Selbstbestimmung jenseits des Körperlichen. Zur Einführung", in *Burghardt/Schmidt/Steinl*, Sexuelle Selbstbestimmung jenseits des Körperlichen, 2024, S. 1 ff.; *Labuz*, Regulating Deep Fakes in the Artificial Intelligence Act, ACIG 2023, 252 (271); Centre for Digital Governance, The false promise of transparent deep fakes: How transparency obligations in the draft AI Act fail to deal with the threat of disinformation and image-based sexual abuse, 2022; so bereits zum Kommissionsentwurf *AccessNow*, Access Now's submission to the European Commission's adoption consultation on the AI Act, 2021, S. 6.

5. Praktische Erwägungen zur Kennzeichnung und Offenlegung

Offenlegungen von synthetischen Inhalten können grundsätzlich nur dann stattfinden, wenn bekannt ist, dass der Inhalt mithilfe einer generativen künstlichen Intelligenz hergestellt oder manipuliert wurde. Dieses Kenntnis hat vornehmlich der Betreiber eines KI-Systems und legt nach der Intention des Gesetzgebers seine synthetischen Inhalte im Einklang mit den Verpflichtungen aus Art. 50 Abs. 4 KI-VO offen. Mit großen praktischen Hürden ist allerdings die Offenlegung solcher Inhalte seitens Dritter (wie etwa sozialen Medien oder andere Plattformen) verbunden. Dies gilt insbesondere dann, wenn die Offenlegungen nicht existent (beispielsweise durch Entfernung von entsprechenden „Labels“) oder nicht mit dem entsprechenden Medium verbunden sind (bspw. fand die Offenlegung durch einen separaten textlichen Hinweis statt). In diesen Fällen sind zur Offenlegung verpflichtete Dritte entweder auf aufwendige und zugleich fehleranfällige Detektionsmechanismen angewiesen, die synthetische Inhalte mithilfe computerforensischer Verfahren identifizieren. Alternativ können sie auf die nach Art. 50 Abs. 2 KI-VO erforderliche maschinenlesbare Kennzeichnung zurückgreifen, die verlässlich, schnell und kostengünstig Auskunft über die synthetische Herstellung oder Manipulation eines Medieninhalts gibt.

In praktischer Hinsicht kann bei der Kennzeichnung von KI-Ausgaben zwischen zwei grundsätzlich unterschiedlichen Verfahren unterschieden werden: Der Kennzeichnung des Inhalts der Ausgabe und der Kennzeichnung des Datenträgers der Ausgabe.

Die *Kennzeichnung des Inhalts* der Ausgabe erfolgt beispielsweise durch die Einfügung von sogenannten Wasserzeichen⁷⁴. Dabei handelt es sich um für das menschliche Auge nicht erkennbare – aber dennoch für Maschinen auslesbare – technische Muster im jeweiligen Inhalt. Ein Beispiel für ein solches Verfahren ist Googles SynthID, bei dem digitale Spuren in synthetisch hergestellte Bilder eingefügt werden, die wiederum mit der passenden Software zwecks Verifikation ausgelesen werden können.⁷⁵ Eine andere Methode ist die Speicherung von sogenannten digitalen Fingerabdrücken. Bei diesem Verfahren wird eine Abstraktion des synthetischen Inhalts

74 Molavi Vasse'i, Watermarking von KI-generierten Inhalten als regulatorisches Instrument, RD 2024, 406 (408, Rn. 142 ff.) mwNachw.; s. auch Partnership on AI, Building a glossary for Synthetic Media Transparency Methods, Part 1: Indirect Disclosure.

75 <https://deepmind.google/models/synthid/>.

(bspw. ein Vektor oder ein Hash-Digest) hergestellt und in einer Datenbank samt entsprechenden Metadaten gespeichert. Bei zu verifizierenden Inhalten wird ebenfalls ein digitaler Fingerabdruck entnommen und mit dem Datenbestand an Fingerabdrücken verglichen. Ergibt der Vergleich einen Treffer, können die passenden Metadaten über den Inhalt offengelegt werden (Beispiel: „dieses Bild ist synthetisch“). Der Vorteil dieses Verfahrens ist die Schnelligkeit, die Robustheit gegenüber Veränderungen bei gekennzeichneten Inhalten sowie die Möglichkeit, sehr viele Informationen zum jeweiligen „fingerprinted“ Inhalt vorzuhalten.⁷⁶

Eine Methode zur *Kennzeichnung des Datenträgers* des Mediums ist das Verfahren nach dem Industriestandard C2PA.⁷⁷ Dabei werden Medien-Metadaten mittels eines kryptografischen Verfahrens mit den Mediendatenträgern verbunden.⁷⁸ Um das Verfahren der Kennzeichnung im C2PA-Standard möglichst sicher zu gestalten, können nur zertifizierte Unternehmen die kryptografischen Kennzeichnungen durchführen.⁷⁹ Auch diese dem Mediendatenträger eingefügten, standardisierten Metadaten können mittels Software ausgelesen werden. Vereinzelt soziale Medien lesen bereits die C2PA-Medien-Metadaten aus und legen diese ihren Nutzerinnen und Nutzern in Form von sog. Content Credentials offen.⁸⁰ Von dieser Offenlegung sind der Ursprung und die Authentizität des Medieninhalts und ggf. ein Hinweis auf die Herstellung oder Manipulation des Inhalts mit generativer künstlicher Intelligenz umfasst. Im gleichen Sinne könnten auch VLOSEs ihren Risikominderungspflichten gemäß Art. 35 Abs. 1 S. 2 lit. k DSA nachkommen, wenn sie die maschinenlesbaren C2PA-Kennzeichnungen auslesen und ihre Nutzerinnen und Nutzer auf den synthetischen Ursprung von ausgespielten Medien hinweisen.⁸¹ Dieses Ineinandergreifen von Art. 50 Abs. 2 KI-VO und Art. 35 Abs. 1 S. 2 lit. k DSA würde schließlich nach hier vertretener Auffassung auch die – eigentlich Betreiber-seitig zu

76 Beispielsweise nutzt die Applikation *Lumid* Fingerprinting-Technologien, um maschinen- und menschenlesbare Informationen zur Herkunft, zum Erstellungsdatum sowie zu urheberrechtlich relevanten Metadaten eines Medieninhalts bereitzustellen. (<https://lumid.tech>).

77 Coalition for Content Provenance and Authenticity, <https://c2pa.org/>.

78 S. https://spec.c2pa.org/specifications/specifications/2.2/specs/C2PA_Specification.html, Ziffer 9. Allerdings ist es auch möglich, C2PA-Metadaten mit Wasserzeichen- und Fingerprinting-Kennzeichnungen zu verbinden.

79 In Deutschland ist in etwa die Valid Technologies GmbH dazu befugt; s. <https://valid.tech>

80 So etwa LinkedIn, s. <https://www.linkedin.com/help/linkedin/answer/a6282984>.

81 S. o. 4.II.; Code of Conduct on Disinformation 2022, Commitment 20.

erfüllende – Offenlegungspflicht nach Art. 50 Abs. 4 KI-VO funktional ersetzen.⁸²

Zuletzt sei noch angemerkt, dass die Betreiber bzw. Anbieter von KI-Systemen – neben dem rechtlichen Zwang – auch eine technisch-wirtschaftliche Motivation zur Kennzeichnung von synthetischen Inhalten haben sollten. Denn das (versehentliche) Einspeisen von synthetischen Inhalten in die Trainingsdaten von KI-Systemen verringert die Leistungsfähigkeit des Systems bis hin zur Unbrauchbarkeit.⁸³

6. Fazit

Der Beitrag hat gezeigt, dass die zunehmende Verbreitung synthetischer Medien das Vertrauen in einzelne Medieninhalte und in digitale Informationsräume gleichermaßen unter Druck setzt und selbst medienkompetente Nutzerinnen und Nutzer vor erhebliche Verifikationsprobleme stellt. Transparenzpflichten stehen dabei im Zielkonflikt zwischen der schlichten Offenlegung der technischen Erstellung und der Minderung negativer Auswirkungen KI-generierter Ausgaben auf das Informationsökosystem. Entsprechend der identifizierten Schutzgüter ist ein wirkungsbezogener Ansatz zu präferieren. Mit Blick auf die Umsetzung der rechtlichen Vorgaben haben sich in der Praxis bereits einzelne Kennzeichnungsmechanismen etabliert.

Zugleich ist vor überhöhten Erwartungen an Transparenzpflichten zu warnen.⁸⁴ Transparenzmaßnahmen dienen primär der Bereitstellung von Kontextinformationen und fördern eine informierte Mediennutzung, stellen jedoch keinen eigenständigen Schutz vor der Erstellung oder dem (digitalen) Missbrauch mithilfe rechtswidriger Inhalte dar. Insbesondere bei gravierenden Rechtsgutverletzungen wie sexualisierter Gewalt stoßen Kennzeichnungen daher an ihre Grenzen.

82 So wohl auch Bomhard/Pieper/Wende/Merkle, 1. Aufl. 2025, KI-VO Art. 50 Rn. 150; allgemein zum Verhältnis der Transparenzpflichten aus KI-VO und DSA s. ErwGr. 136 KI-VO; Kumkar/Griesel, Transparenzpflichten für Deepfakes und synthetische Medieninhalte in der KI-VO, KIR 2024, 117 (125); Becker, Generative KI und Deepfakes in der KI-VO, CR 2024, 353 (363 Rn. 94).

83 Xing u. a., On the caveats of AI autophagy, Nature Machine Intelligence, 2025.

84 S. hierfür aktuell Clark/Lewandowsky, The continued influence of AI-generated deepfake videos despite transparency warnings, Communication Psychology, 2026.

Literatur

- Access Now (2021): Access Now's submission to the European Commission's adoption consultation on the Artificial Intelligence Act. URL: <https://www.accessnow.org/wp-content/uploads/2021/08/Submission-to-the-European-Commissions-Consultation-on-the-Artificial-Intelligence-Act.pdf> (besucht am 30.01.2026).
- Achenleiter, Ranja Andrea (2025): Das Medienprivileg im European Media Freedom Act. *Multimedia und Recht*, 02/2025, S. 89-93.
- Altay, Sacha und Gilardi, Fabrizio (2024): People are skeptical of headlines labeled as AI-generated, even if true or human-made, because they assume full AI automation. *PNAS Nexus*, 3 (10).
- Becker, Maximilian (2024): Generative KI und Deepfakes in der KI-VO. *Computer und Recht*, 06/2024, S. 353-366.
- Becker, Maximilian (2024): Das Urheberrecht als Trostpreis für den Menschen? *Gewerblicher Rechtsschutz und Urheberrecht*, 08/2024, S. 505-514.
- Bentzen, Naja (2025): Briefing: Information manipulation in the age of generative artificial intelligence. European Parliamentary Research Service. URL: [https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/779259/EPRS_BRI\(2025\)779259_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/779259/EPRS_BRI(2025)779259_EN.pdf) (besucht am 30.01.2026).
- Blocher, Janine (2025): Strafbare Deepfakes als Täuschung über die Authentizität von Medieninhalten. *Künstliche Intelligenz und Recht*, 06/2025, S. 225-230.
- Brokhues, Annabell (26. Januar 2026): EU-Kommission startet Verfahren gegen Chatbot Grok. *Deutschlandfunk*. URL: <https://www.deutschlandfunk.de/deepfakes-eu-kommission-startet-verfahren-gegen-chatbot-grok-100.html> (besucht am 30.01.2026).
- Brorsen, Hans und Falk, Richard (2024): Neue Compliance-Pflichten nach dem Digital Services Act. *Multimedia und Recht*, 01/2024, S. 32-36.
- Brorsen, Hans und Falk, Richard (2025): Die „Maschinenlesbarkeit“ von Informationen. *Multimedia und Recht*, 01/2025, S. 7-12.
- Burghardt, Boris; Schmidt, Anja und Steinl, Leonie (2024): Sexuelle Selbstbestimmung jenseits des Körperlichen. Zur Einführung. In: Boris Burghardt, Anja Schmidt und Leonie Steinl (Hrsg.): Sexuelle Selbstbestimmung jenseits des Körperlichen. Berlin: Mohr Siebeck. S. 1-14.
- Campante, Felipe R.; Durante, Ruben; Hagemeister, Felix und Sen, Ananya (2025): Genai Misinformation, Trust, And News Consumption: Evidence From A Field Experiment. National Bureau of Economic Research. Cambridge. URL: https://www.nber.org/system/files/working_papers/w34100/w34100.pdf (besucht am 30.01.2026).
- CWA17493 (2019): Journalism Trust Initiative. Cen Workshop Agreement. URL: <https://www.cencenelec.eu/media/CEN-CENELEC/CWAs/ICT/cwa17493.pdf> (besucht am 30.01.2026).
- Clark, Simon und Lewandowsky, Stephan (2026): The continued influence of AI-generated deepfake videos despite transparency warnings. *Communication Psychology* 4, article 13.

- Crone, Theresia (23. Januar 2026): Bildbasiert aber unsichtbar. *Verfassungsblog.de* URL: <https://verfassungsblog.de/pornografische-deepfakes/> (besucht am 30.01.2026).
- D'Anastasio, Cecilia (08. Januar 2026): KI-Chatbot Grok sorgt auf X für massenhafte Nacktbilder ohne Zustimmung der Betroffenen. *Bloomberg.com* URL: <https://www.bloomberg.com/news/articles/2026-01-08/x-gerat-wegen-grok-unter-druck-durch-flut-an-ki-nacktbildern> (besucht am 30.01.2026).
- Dobber, Tom; Kruikemeier, Sanne; Votta, Fabio und Helberger, Natali (2025): The effect of traffic light veracity labels on perceptions of political advertising source and message credibility on social media. *Journal of Information Technology & Politics*, Volume 22, Issue 1, S. 82-97.
- Ellenberg, Stefan (2024): Persönlichkeits- und Urheberrechte beim Voice Cloning. *Recht Digital* 12/2024, S. 599-607.
- Engel-Bunsas, Christoph (2025): Recht an der eigenen Stimme in Zeiten von Deepfakes. *Recht Digital*, 06/2025, S. 292-300.
- Epstein, Ziv; Fang, Cathy Mengying; Arechar, Antonio Alonso und Rand, David (2024): What label should be applied to content produced by generative AI?. An MIT Exploration of Generative AI.
- Ernst, Nico (21. April 2024): Jim Keller über KI: „In 10 Jahren ist alle Software weg“. *Heise.de*. URL: <https://www.heise.de/news/Jim-Keller-ueber-KI-In-10-Jahren-ist-alle-Software-weg-9692706.html> (besucht am 30.01.2026).
- Europäische Kommission (2024): Leitlinien der Kommission für Anbieter sehr großer Online-Plattformen und sehr großer Online-Suchmaschinen zur Minderung systemischer Risiken in Wahlprozessen gemäß Artikel 35 Absatz 3 der Verordnung (EU) 2022/2065. URL: https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=OJ:C_202403014 (besucht am 16.03.2026).
- Europäische Kommission (2025): Commission Opinion of 13.2.2025 on the assessment of the Code of Practice on Disinformation within the meaning of Article 45 of Regulation 2022/2065. URL: <https://digital-strategy.ec.europa.eu/en/library/code-conduct-disinformation>
- Europäische Kommission (2025): First Draft Code of Practice on Transparency of AI-Generated Content. URL: <https://digital-strategy.ec.europa.eu/en/library/first-draft-code-practice-transparency-ai-generated-content> (besucht am 16.03.2026).
- Europäische Kommission (2026): Second Draft Code of Practice on Transparency of AI-Generated Content. URL: <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-second-draft-code-practice-marking-and-labelling-ai-generated-content> (besucht am 16.03.2026).
- Europäische Kommission (2022): The Strengthened Code of Practice on Disinformation 2022. URL: <https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation>
- Europäisches Parlament und Rat (2022): Verordnung (EU) 2022/2065 über einen Binnenmarkt für digitale Dienste und zur Änderung der Richtlinie 2000/31/EG (Gesetz über digitale Dienste). URL: <https://eur-lex.europa.eu/eli/reg/2022/2065/oj> (besucht am 16.03.2026)

- Europäisches Parlament und Rat (2024): Verordnung (EU) 2024/1083 zur Schaffung eines gemeinsamen Rahmens für Mediendienste im Binnenmarkt und zur Änderung der Richtlinie 2010/13/EU (Europäisches Medienfreiheitsgesetz). URL: https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=OJ:L_202401083 (besucht am 16.03.2026).
- Europäisches Parlament und Rat (2024): Verordnung (EU) 2024/1689 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) (Text von Bedeutung für den EWR). URL: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj?locale=de> (besucht am 16.03.2026).
- Game, Dilrukshi; Sewwandi, Dilki; Bandara, Arosha und Zhang, Min (Mai 2025): Labeling Synthetic Content: User Perceptions of Warning Label Designs for AI-generated Content on Social Media. *Conference on Human Factors in Computing Systems CHI* 25. URL: <https://arxiv.org/pdf/2503.05711> (besucht am 30.01.2026).
- Gerdemann, Simon und Spindler, Gerald (2023): Das Gesetz über digitale Dienste (Digital Services Act) (Teil 2). *Gewerblicher Rechtsschutz und Urheberrecht*, 03/2023, S. 115-125.
- Hartmann, Frank (2020): Der persönlichkeitsrechtliche Schutz vor Deepfakes. *Kommunikation und Recht*, 05/2020, S. 350-356.
- Hennemann, Moritz (2024): Der European Media Freedom Act oder: Von Zielen und Mitteln des Medienrechts. *Zeitschrift für Europäisches Privatrecht*, 03/2024, S. 498-504.
- Hinderks, Tobias (2022): Die Kennzeichnungspflicht von Deepfakes. *Zeitschrift für Urheber- und Medienrecht*, 02/2022, S. 110-119.
- Home Security Heroes (2023): State of Deepfakes. URL: <https://www.securityhero.io/state-of-deepfakes/> (besucht am 30.01.2026).
- van Huijstee, Mariette; van Boheemen, Pieter; Das, Djurre; Nierling, Linda; Jahn, Jutta; Karaboga, Murat; Fatun, Martin; Kool, Linda und Gerritsen, Jost (2021): Tackling deepfakes in European policy. European Parliamentary Research Service. URL: [https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_STU\(2021\)690039_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf) (besucht am 30.01.2026).
- Karaboga, Murat (2023): Die Regulierung von Deepfakes auf EU-Ebene. In: Sylvia Jaki und Stefan Steiger (Hrsg.): *Digitale Hate Speech*. Berlin: J.B. Metzler, S. 197-220.
- Karttunen, Sofia (2026): The Danish approach to copyright and deepfakes: A model for the EU? *European Parliamentary Research Service*. URL: [https://www.europarl.europa.eu/RegData/etudes/ATAG/2026/782611/EPRS_ATA\(2026\)782611_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/ATAG/2026/782611/EPRS_ATA(2026)782611_EN.pdf) (besucht am 30.01.2026).
- Kern, Michael (2024): Die Verwendung von Social Bots: Transparenzpflichten gemäß Medienstaatsvertrag und KI-VO sowie deren Umsetzung in virtuellen Welten. *Künstliche Intelligenz und Recht*, 03/2024, S. 94-99.
- Klass, Nadine (2025): Manipulierte Realität: Wie Deepfakes Recht und Gesellschaft herausfordern. *Zeitschrift für Urheber- und Medienrecht*, 07/2025, S. 485-492.

- Kleine, Felix (2022): Perception of Deepfake Technology - The Influence of the Recipients' Affinity for Technology on the Perception of Deepfakes. Masterarbeit. Hochschule Darmstadt. URL: https://www.researchgate.net/publication/364254498_Perception_of_Deepfake_Technology_-_The_Influence_of_the_Recipients'_Affinity_for_Technology_on_the_Perception_of_Deepfakes (besucht am 30.01.2026).
- Kollmann, Katharina (2023): EU-Kommission: Unterzeichner des Verhaltenskodexes gegen Desinformation sollen KI-Nutzung klar kennzeichnen. *Multimedia und Recht-Aktuell*, 11/2023, 458365.
- Kumkar, Lea Katharina (2023): Deepfakes – Risiken und Regulierung im europäischen Verordnungsentwurf für künstliche Intelligenz. *Kommunikation und Recht*, Beilage zu 10/2023, S. 32-37.
- Kumkar, Lea Katharina und Griesel, Moritz (2024): Transparenzpflichten für Deepfakes und synthetische Medieninhalte in der KI-VO. *Künstliche Intelligenz und Recht*, 04/2024, S. 117-126.
- Kumkar, Lea Katharina und Rapp, Julian Philipp (2022): Deepfakes. *Zeitschrift für Digitalisierung und Recht*, 03/2022, S. 199-228.
- Łabuz, Mateusz (2023): Regulating Deep Fakes in the Artificial Intelligence Act. *Applied Cybersecurity & Internet Governance* 2. 1-42., S. 252-291.
- Lantwin, Tobias (2019): Deep Fakes – Düstere Zeiten für den Persönlichkeitsschutz? *Multimedia und Recht*, 09/2019, S. 574-578.
- v. Lewinski, Kai und Klink, Quentin (2023): Compliance-Pflichten nach dem DSA – ein umfassender Überblick. *Recht Digital*, 04/2023, S. 183-194.
- Li, Fan und Yang, Ya (2024): Impact of Artificial Intelligence-Generated Content Labels On Perceived Accuracy, Message Credibility, and Sharing Intentions for Misinformation. *JMIR formative research* 8.
- Löber, Lena und Roßnagel, Alexander (2019): Kennzeichnung von Social Bots: Transparenzpflichten zum Schutz integrier Kommunikation. *Multimedia und Recht*, 08/2019, S. 493-498.
- Lockhart, Angus und Tessono, Christelle (2025): Human or AI? Evaluating Labels on AI-Generated Social Media Content. Dais.ce. URL: <https://dais.ca/reports/human-or-ai/> (besucht am 30.01.2026).
- Maras, Marie-Helen und Alexandrou, Alex (2019): Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *International Journal of Evidence & Proof*, Volume 23, Issue 3, S. 255-262.
- Meding, Kristof und Sorge, Christoph (März 2025): What constitutes a deepfake? Symposium of Computer Science and Law, URL: <https://dl.acm.org/doi/10.1145/3709025.3712218> (besucht am 30.01.2026).
- Merkle, Marieke (2025): KI-VO Art. 50 Transparenzpflichten für Anbieter und Betreiber bestimmter KI-Systeme. In: David Bomhard, Fritz-Ulli Pieper und Susanne Wende (Hrsg.): *KI-VO*. Frankfurt am Main: Dfv Mediengruppe.
- Molavi Vasse'i, Ramak (2024): Watermarking von KI-generierten Inhalten als regulatorisches Instrument. *Recht Digital*, 09/2024, S. 406-414.

- Newman, Nic (Januar 2026): Journalism and Technology Trends and Predictions 2026. *The Reuters Institute For The Study Of Journalism*. URL: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2026-01/Trends_and_Predictions_2026.pdf (besucht am 30.01.2026).
- Oswalt, Stefanie (20. Juni 2018): Wie Kennedy eine Rede nach seinem Tod hielt. *Deutschlandfunk-Kultur* URL: <https://www.deutschlandfunkkultur.de/auditive-fake-news-wie-kennedy-eine-rede-nach-seinem-tod-100.html> (besucht am 30.01.2026).
- Partnership on AI (13. Dezember 2023): Building a glossary for Synthetic Media Transparency Methods, Part 1: Indirect Disclosure. *Partnershiponai.org*. URL: <https://partnershiponai.org/resource/glossary-for-synthetic-media-transparency-methods-part-1/> (besucht am 30.01.2026).
- Pawelec, Maria und Bieß, Cora (2021): Deepfakes. Baden-Baden: Nomos Verlagsgesellschaft mbH & Co. KG
- Peukert, Alexander (2023): Desinformationsregulierung in der EU - Überblick und offene Fragen. *Juristen Zeitung*, 07/2023, S. 278-286.
- Pfeiffer, Charlene (2025): Tatbestandsmerkmale, Rechtfertigungsgründe und Herausforderungen beim strafrechtlichen Schutz vor Deepfakes. *Multimedia und Recht*, 12/2025, S. 934-939.
- Rauchfleisch, Adrian; Vogler, Daniel und de Seta, Gabriele (2025): Deepfakes or Synthetic Media? The Effect of Euphemisms for Labeling Technology on Risk and Benefit Perceptions. *Social Media + Society*. Volume 11, Issue 3.
- Schertz, Christian (2025): Postmortales Persönlichkeitsrecht in Zeiten von KI und Deepfakes. *Neue Juristische Wochenschrift*, 39/2025, S. 2801-2806.
- Schultz, Annabel (2025): Aufsichtsrechtliches Vorgehen gegen Fake-Profile: Mögliche Maßnahmen im Kampf gegen die Verbreitung von Desinformationen. *Multimedia und Recht*, 03/2025, S. 168-173.
- Spennemann, Dirk H.R. (2025): ‚Delving into‘ the quantification of Ai-generated content on the internet (synthetic data). arXiv. URL: <https://arxiv.org/abs/2504.08755v1> (besucht am 10.02.2026).
- Steinebach, Martin (2024): KI-generierte Abbildungen von Kindesmissbrauch. *Datenschutz und Datensicherheit*, 05/2024, S. 304-309.
- Stosz, Conrad (03. Februar 2019): Policy Options for Fighting Deepfakes. *Georgetown Security Studies Review* URL: <https://georgetownsecuritystudiesreview.org/2019/02/03/policy-options-for-fighting-deepfakes/> (besucht am 30.01.2026).
- Toff, Benjamin und Simons, Felix M. (2025) “Or They Could Just Not Use It?": The Dilemma of AI Disclosure for Audience Trust in News. *The International Journal of Press/Politics*, S. 881-903.
- Völmann, Berit (2025): Dein Körper, meine Bilder? Der Schutz vor bildbasierter sexualisierter Gewalt (auch) in der Gewaltschutzrichtlinie der EU. *Zeitschrift für Urheber- und Medienrecht*, 01/2025, S. 1-8.
- Wittenberg, Chloe; Epstein, Ziv; Peloquin-Skulski, Gabrielle; Berinsky, Adam J. und Rand, David G. (2025): Labeling AI-generated media online. *PNAS Nexus*, Volume 4, Issue 6.

Xing, Xiaodan; Shi, Fadong; Huang, Jiahao; Wu, Yinzhe; Nan, Yang; Zhang, Sheng; Fang, Yingying; Roberts, Michael; Schönlieb, Carola-Bibiane; Del Ser, Javier und Yang, Guang (2025): On the caveats of AI autophagy, *Nature Machine Intelligence*, S. 172-180.