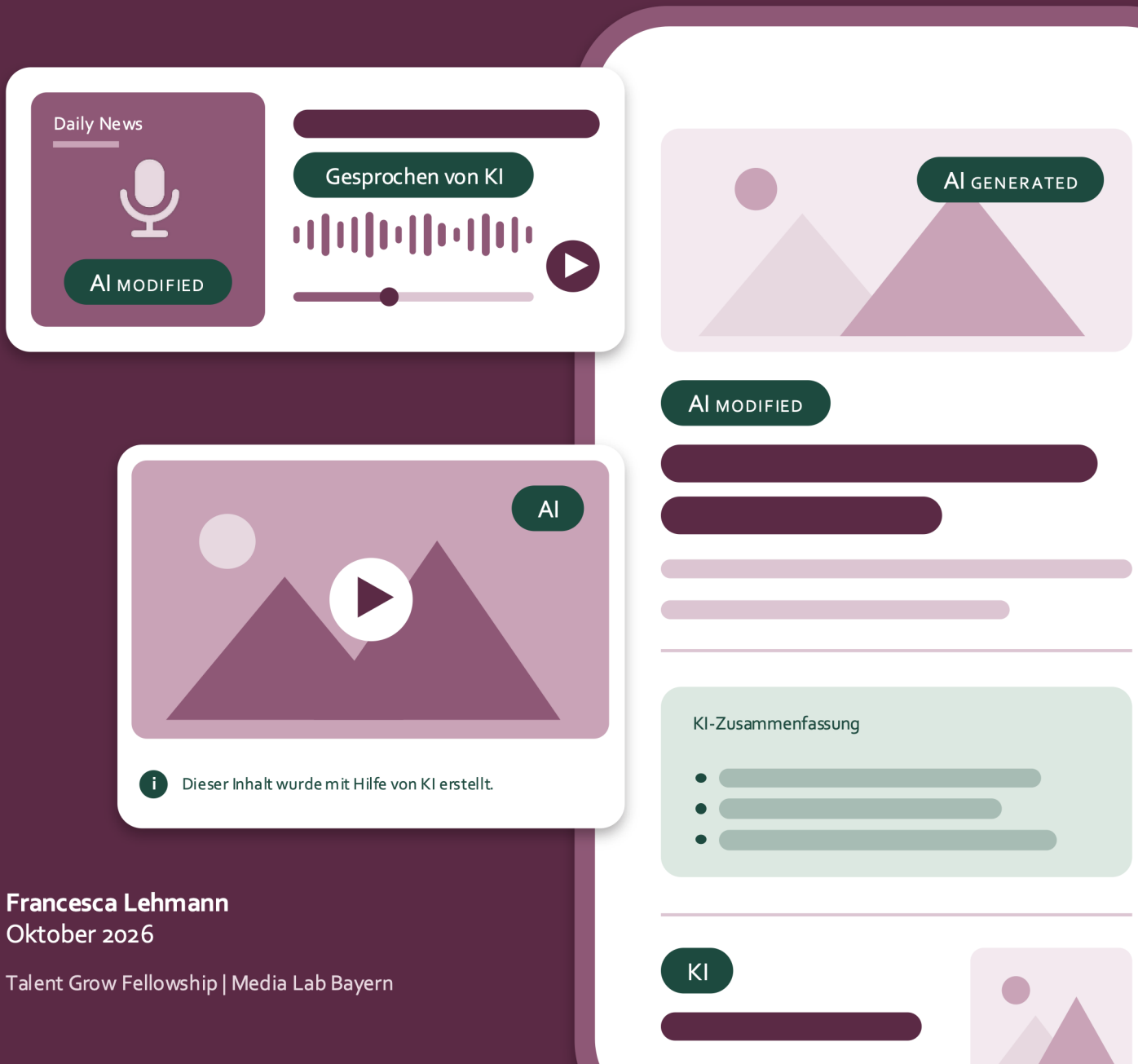
White Paper

Transparente Kennzeichnung von KI im Journalismus



Francesca Lehmann
Oktober 2026

Talent Grow Fellowship | Media Lab Bayern

Inhalt

Zusammenfassung 2

Über dieses White Paper 2

1. Einleitung 3

2. KI-Transparenz und die KI-Verordnung 3

2.1 Begrenztes Risiko = Transparenz: Artikel 50 der KI-VO 4

2.2 Leitlinien der EU-Kommission 5

2.3 Verhaltenskodex der EU-Kommission (Code of Practice) 6

2.4 Die KI-Verordnung auf nationaler Ebene 8

3. KI-Transparenz und Journalismus 9

4. KI-Kennzeichnung in der Praxis 10

4.1 Anforderungen an KI-Kennzeichnung 10

4.2 Wie wird KI-Kennzeichnung umgesetzt? 11

5. Stimmen aus der Praxis: Wie Medienhäuser mit KI-Kennzeichnung umgehen 15

6. Braucht es einen Standard für KI-Kennzeichnung? 16

7. Fazit & Ausblick 18

Literatur 21

Danksagung & Kontakt 25

Zusammenfassung

Seit dem 2. August 2026 gelten die Transparenzpflichten aus Artikel 50 der KI-Verordnung. Dieses White Paper untersucht, was sie für Medienhäuser bedeuten, welche selbstregulierenden Maßnahmen die Branche entwickelt und wie KI-Kennzeichnung in der Praxis aussieht.

Die KI-Verordnung betrifft Medienhäuser nur in Teilen. Kennzeichnungspflichtig sind vor allem Deepfakes sowie KI-generierte Texte zu Themen von öffentlichem Interesse, sofern diese nicht redaktionell geprüft und verantwortet werden. In der Rolle des „Betreibers" eines KI-Systems bewerten Häuser selbst, ob ein Inhalt täuschen könnte (und daher der Deepfake-Definition der KI-VO entspricht) und ob ein Text ausreichend geprüft ist und jemand die redaktionelle Verantwortung übernimmt (und daher die Ausnahme gilt, bei der keine Kennzeichnung erforderlich ist).

Die Branche reguliert sich vor allem über Leitlinien. 35 von 60 untersuchten Akteuren legen ihren KI-Einsatz öffentlich dar, am häufigsten in Leitlinien, daneben in Transparenzbeiträgen oder im Impressum. Wie Kennzeichnung konkret aussehen soll, bleibt darin meist offen.

In der Praxis entstehen eigene Formen der Kennzeichnung mit unterschiedlichen Bedeutungen. Hinweise stehen meist direkt am Produkt. Dort heißt die Kennzeichnung: Hier hat die Redaktion selbst KI eingesetzt, und sie legt den Prozess offen. Bei der Berichterstattung über Deepfakes ist die Kennzeichnung hingegen ein Warnsignal: Sie klärt darüber auf, dass ein Inhalt nicht echt ist.

Einen Standard gibt es bisher nicht und ob ein Standard sinnvoll wäre, ist offen. Die Labels des Verhaltenskodex haben sich bisher nicht flächendeckend durchgesetzt. Einheitliche Labels wären leichter erkennbar, eigene Formulierungen sind genauer.

Wie Kennzeichnung wirkt, ist offen. Einige argumentieren, dass Transparenz durch Kennzeichnung Vertrauen schafft. Andere befürchten, dass ein KI-Label Inhalte abwertet.

Über dieses White Paper

Dieses White Paper ist im Rahmen des Talent Grow Fellowship des Media Lab Bayern entstanden und richtet sich an alle, die sich mit KI-Transparenz im Journalismus, Kennzeichnungspraktiken und Herkunftsnachweisen digitaler Artefakte befassen. Es verbindet eine Analyse der Anforderungen der KI-Verordnung mit einer quantitativen Erhebung der Selbstverpflichtungen von 60 Akteuren der Medienbranche, Kennzeichnungsbeispielen aus der Praxis und Einblicken aus fünf Gesprächen mit Verantwortlichen aus Medienhäusern. Untersucht wurde der Zeitraum vom 1. Juli bis 30. September 2026. Das White Paper ordnet die rechtliche Lage ein, ersetzt aber keine Rechtsberatung.

1. Einleitung

Die sorgfältige Prüfung von Informationen auf ihren Wahrheitsgehalt, ihre wahrheitsgemäße Wiedergabe sowie die **transparente Kennzeichnung unbestätigter oder symbolischer Darstellungen** sind als zentrale **Normen im journalistischen Selbstverständnis** verankert (vgl. Pressekodex, Ziffern 1 und 2).

KI-Technologien stellen diese Normen und ihre Umsetzung in der Praxis in zweifacher Hinsicht vor neue Herausforderungen. Erstens wird es **schwieriger, Wahrheitsgehalt, Herkunft und Authentizität medialer Inhalte zu beurteilen**, da sich „echte" und „synthetische" Inhalte mit dem bloßen Auge nicht mehr unterscheiden lassen. Das betrifft in der Praxis vor allem extern bezogene Inhalte, die Redaktionen nicht selbst erstellt haben, zum Beispiel Material von Agenturen ebenso wie Material aus sozialen Netzwerken. Zweitens ist der **Einsatz von KI-Technologien** in journalistischen Prozessen, Produkten und Ausspielungsformen ein **zentraler Hebel für die digitale Transformation.**

Wenn Redaktionen KI zur Erstellung, Bearbeitung oder Aufbereitung eigener journalistischer Inhalte einsetzen, stellt sich die Frage, ob und in welcher Form dieser Einsatz im Sinne der journalistischen Normen **kenntlich gemacht werden sollte**.

Verbindliche Vorgaben, die diese Frage für den Journalismus beantworten, **existieren bislang nicht**. Das vorliegende White Paper untersucht deshalb zweierlei: welche Transparenzpflichten sich aus der KI-Verordnung ergeben und welche selbstregulierenden Maßnahmen die Branche entwickelt.

2. KI-Transparenz und die KI-Verordnung

Ob ein Inhalt mit KI erzeugt oder verändert wurde, lässt sich im Nachhinein technisch nur begrenzt nachweisen. Das gilt auch für die Erkennung von Deepfakes. Der Grund: **Erkennung und Erzeugung entwickeln sich in wechselseitiger Abhängigkeit.** Jede verbesserte Detektion setzt den Maßstab, an dem sich die nächste Generation generativer Modelle ausrichtet. Daraus ist ein grundlegend anderer Ansatz hervorgegangen: **Wenn sich die Herkunft eines Inhalts nachträglich nicht zuverlässig feststellen lässt, muss sie dort offengelegt werden, wo er entsteht.**

Dieser Logik folgt auch die KI-VO (EU AI Act). Transparenz bedeutet in Artikel 50 in erster Linie, offenzulegen, wenn Inhalte mit KI erzeugt oder verändert wurden. Die Pflichten richten sich deshalb **an diejenigen, die KI-Systeme entwickeln oder einsetzen**.

Die KI-VO ist eine horizontale Verordnung: Sie gilt für alle Branchen gleichermaßen und enthält keine direkten Regeln für die Medienbranche; die Pflichten werden erst relevant, wenn in einer

bestimmten Rolle (Anbieter oder Betreiber) bestimmte KI-Systeme eingesetzt werden oder die erzeugten Inhalte der KI-Systeme verbreitet werden.

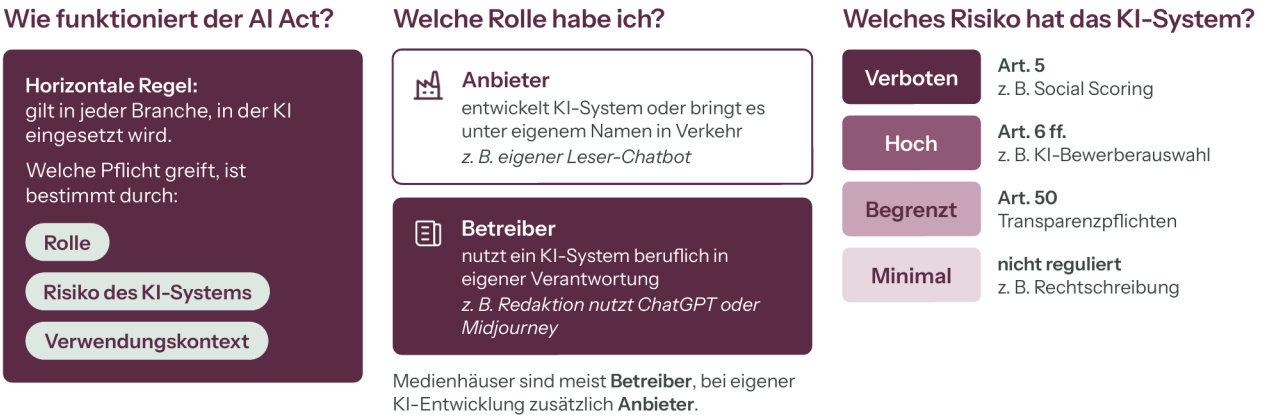


Abbildung 1: Die Logik der KI-Verordnung im Überblick (eigene Darstellung nach Verordnung (EU) 2024/1689)

2.1 Begrenztes Risiko = Transparenz: Artikel 50 der KI-VO

Seit dem 2. August 2026 gelten die Transparenzpflichten aus Artikel 50 der KI-Verordnung. Sie richten sich an zwei Rollen: **Anbieter**, die KI-Systeme entwickeln, und **Betreiber**, die KI-Systeme in eigener Verantwortung beruflich nutzen.

Anbieter müssen KI-Systeme, die direkt mit Menschen interagieren, so gestalten, dass die Nutzenden erfahren, dass sie es mit einer KI zu tun haben (Art. 50 Abs. 1). Außerdem müssen sie KI-erzeugte Audio-, Bild-, Video- und Textinhalte **maschinenlesbar markieren** und als künstlich erzeugt oder manipuliert erkennbar machen (Art. 50 Abs. 2).

Betreiber von Systemen zur Emotionserkennung oder biometrischen Kategorisierung müssen die betroffenen Personen darüber informieren (Art. 50 Abs. 3). Betreiber, die mit **KI Deepfakes** erzeugen oder manipulieren, müssen offenlegen, dass die Inhalte künstlich erzeugt oder verändert wurden (Art. 50 Abs. 4).

> Als **Deepfake** gelten Bild-, Audio- oder Videoinhalte, die real existierenden Personen, Gegenständen, Orten, Einrichtungen oder Ereignissen ähneln und fälschlich als echt oder wahrheitsgemäß erscheinen würden (Art. 3 Nr. 60).

Bei offensichtlich künstlerischen, kreativen, satirischen oder fiktionalen Werken beschränkt sich die Pflicht auf einen Hinweis, der Darstellung des Werks nicht beeinträchtigt (Art. 50 Abs. 4).

Betreiber müssen ebenfalls offenlegen, wenn sie **KI-generierte oder -manipulierte Texte** veröffentlichen, die die Öffentlichkeit über Angelegenheiten von öffentlichem Interesse informieren sollen. **Dies gilt nicht, wenn der Text von Menschen geprüft oder redaktionell kontrolliert wurde und eine natürliche oder juristische Person die redaktionelle Verantwortung trägt** (Art. 50 Abs. 4).

Für alle Pflichten gilt: Die Informationen müssen **klar und eindeutig erkennbar** sein, spätestens **bei der ersten Interaktion** oder dem ersten Kontakt mit dem Inhalt erfolgen und **barrierefrei** sein (Art. 50 Abs. 5).

Je Fall prüfen: **Rolle:** Anbieter oder Betreiber? | **Output:** Text, Bild, Audio, Video? | **Kontext:** öffentlich? Satire?

Abs.	Was entsteht?	Wer?	Pflicht	Ausnahme
1	**Chatbot / interaktive KI**	Anbieter	Nutzer informieren, dass sie mit KI interagieren	**entfällt**, wenn das offensichtlich ist
2	**Synthetischer Inhalt** Text, Bild, Audio, Video	Anbieter	maschinenlesbar markieren (z. B. Metadaten, Wasserzeichen)	**entfällt** bei reiner Assistenz, z. B. Standard-Bearbeitung
3	**Emotionserkennung / biometr. Kategorisierung**	Betreiber	betroffene Personen informieren	**entfällt** bei gesetzlich erlaubter Strafverfolgung
4	**Deepfake** Bild, Audio, Video realer Personen, Orte, Ereignisse	Betreiber	offenlegen, dass Inhalt künstlich erzeugt oder manipuliert ist	**abgeschwächt** bei Kunst, Satire, Fiktion: Hinweis, der das Werk nicht stört
4	**KI-Text zu Themen von öffentlichem Interesse**	Betreiber	offenlegen, dass Text KI-generiert oder -manipuliert ist	**entfällt** bei menschlicher Prüfung und redaktioneller Verantwortung

Abs. 5, für jeden Hinweis: klar, erkennbar, barrierefrei, spätestens beim ersten Kontakt

Abbildung 2: Begrenztes Risiko: Transparenzpflichten nach Art. 50 (eigene Darstellung nach Art. 50 KI-VO)

Wie die Offenlegung in der Praxis aussehen soll, lässt die Verordnung **weitgehend offen.**

Diese Fragen beantworten zwei ergänzende Instrumente: die **Leitlinien der EU-Kommission**, die die Pflichten auslegen, um das Rollenverständnis zu vereinfachen, und ein **freiwilliger Verhaltenskodex** (Code of Practice), der eine mögliche Umsetzung der Transparenzpflichten aus Artikel 50 beschreibt.

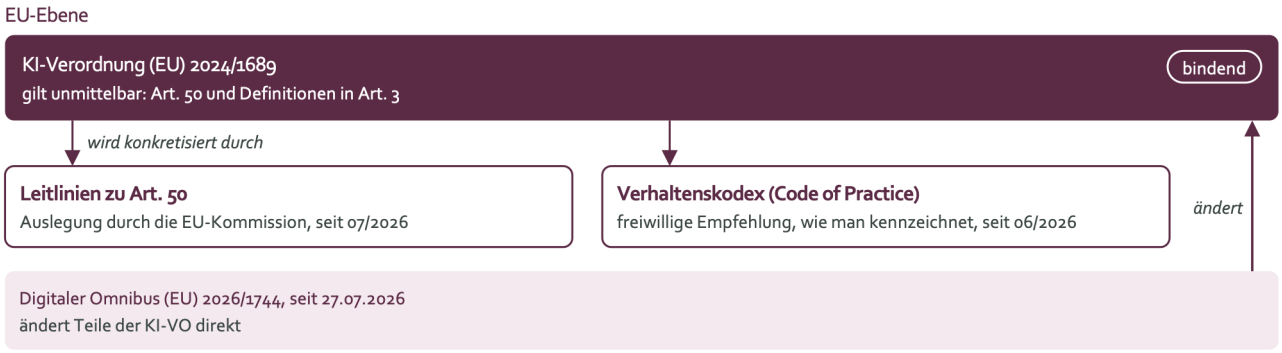


Abbildung 3: Was gehört zum Regelwerk? (eigene Darstellung nach KI-VO, Leitlinien der EU-Kommission, Verhaltenskodex und Verordnung (EU) 2026/1744)

2.2 Leitlinien der EU-Kommission

Wie die Pflichten auszulegen sind, hat die Kommission am 20. Juli 2026 in Leitlinien konkretisiert. Sie sollen Behörden, Anbietern und Betreibern helfen, die Pflichten aus Art. 50 einheitlich, wirksam und verhältnismäßig anzuwenden. Ein Entwurf war am 8. Mai 2026 erschienen und bis zum 3. Juni zur Konsultation gestellt. In die Endfassung flossen Rückmeldungen der Mitgliedstaaten, des KI-Gremiums und weiterer Stakeholder ein. Die **Leitlinien** grenzen ab, welche Pflichten für wen

entlang der Wertschöpfungskette gelten, **definieren zentrale Begriffe** und zeigen anhand von **Beispielen**, was unter die Pflichten fällt und was nicht.

Ein **Deepfake** liegt vor, wenn drei Kriterien zusammenkommen (FAQ):

1. hohe Ähnlichkeit mit dem dargestellten Gegenstand

2. ein Bezug zu etwas, das existiert, plausibel existieren könnte oder existiert haben könnte

3. die Eignung, über Echtheit oder Wahrheitsgehalt zu täuschen.

Für das dritte Kriterium darf der **Grad der Ähnlichkeit**, die **inhaltliche Aussage**, der vorgesehene **Einsatzkontext** sowie die **Zusammensetzung und die Erwartungen des Publikums** berücksichtigt werden.

Als **Betreiber** gilt, wer ein KI-System in eigener Verantwortung beruflich nutzt; rein private Nutzung ist ausgenommen (FAQ). Ist ein Unternehmen der Betreiber, sind die Beschäftigten, die das System in dessen Auftrag nutzen, keine eigenständigen Betreiber.

Die Kommission nennt als Beispiele Content Creator sowie Journalistinnen und Journalisten. Das Unternehmen bleibt auch dann Betreiber, wenn externe Dienstleister oder Freie das System in seinem Auftrag bedienen.

Kennzeichnungspflichtig sind Texte, die veröffentlicht werden und Angelegenheiten von öffentlichem Interesse betreffen (FAQ).

Die Ausnahme für geprüfte Texte definieren die Leitlinien genauer:

Menschliche Prüfung bedeutet eine **bewusste inhaltliche Prüfung durch fachkundige Personen**.

Redaktionelle Kontrolle bedeutet, dass eine **verantwortliche redaktionelle Instanz** den Text aus inhaltlichen Gründen freigeben, ändern oder ablehnen kann, **einschließlich Faktenprüfung.** Rein formale Prüfungen wie Rechtschreib- oder Grammatikkorrekturen genügen nicht.

Redaktionelle Verantwortung heißt, dass eine Person die letztliche rechtliche **Verantwortung für die Veröffentlichung** trägt.

Inhalte, die vor dem 2. August 2026 erzeugt wurden, müssen nicht rückwirkend gekennzeichnet werden; die Kommission empfiehlt es aber.

2.3 Verhaltenskodex der EU-Kommission (Code of Practice)

Wie die Kennzeichnungspflichten aus Art. 50 KI-VO in der Praxis umgesetzt werden können, beschreibt der Verhaltenskodex (Code of Practice), dessen finale Fassung die EU-Kommission am 10. Juni 2026 veröffentlicht hat.

Erarbeitet haben ihn unabhängige Expertinnen und Experten. Beteiligt waren laut Teilnehmerliste der Kommission zudem 136 Organisationen und 46 Forschende, aus der deutschen Medienbranche unter anderem BDZV, MVFP, VAUNET, der Deutsche Fotorat und die Medienanstalten.

Der Kodex besteht aus zwei Teilen: Der erste richtet sich an Anbieter von KI-Systemen und regelt die maschinenlesbare Markierung. Der zweite richtet sich an Betreiber und **regelt die sichtbare Kennzeichnung von Deepfakes und KI-generierten Texten zu Themen von öffentlichem Interesse**.

Wer die **Ausnahme** für redaktionell geprüfte Texte in Anspruch nimmt, muss die Prüfung **intern dokumentieren, etwa wer verantwortlich war und wann der Text freigegeben wurde**. Daraus ergibt sich eine einfache Logik: Entweder **nach innen dokumentieren** oder **nach außen kennzeichnen**.

Fall und Rolle	AI Act sagt	Verhaltenskodex: so geht's	Beispiele
2 **Synthetischer Inhalt** Text, Bild, Audio, Video — Anbieter	Art. 50 Abs. 2, Section 1 Output maschinenlesbar markieren und als KI erkennbar machen	**Mind. 2 Ebenen:** digital signierte Metadaten und unsichtbares Wasserzeichen (Freitext nur Wasserzeichen, ab 200 Tokens). **Erkennungstool** grundsätzlich kostenlos, für Medien und Faktenchecker ohne Mengenlimit.	**Erkennungstools** der KI-Anbieter müssen Redaktionen kostenlos zur Verfügung stehen.
4 **Deepfake: Bild, Video** realistisch, zeigt reale Personen, Orte, Ereignisse — Betreiber	Art. 50 Abs. 4, Section 2 offenlegen, dass Inhalt künstlich erzeugt oder manipuliert ist	**EU-Icon „AI"** oder gleichwertiges Label, direkt im Inhalt ohne Überlagerung, z. B. oben rechts. **Video:** zu Beginn, wenn möglich in Intervallen, mindestens nach jeder Unterbrechung.	Fotorealistische KI-Rekonstruktion im Erklärvideo: **Icon oben rechts**, nach der Werbepause erneut.
4 **Deepfake: Audio** z. B. nachgebildete Stimme einer realen Person — Betreiber	Art. 50 Abs. 4, Section 2 offenlegen, dass Inhalt künstlich erzeugt oder manipuliert ist	**Gesprochener Hinweis** zu Beginn, in der Sprache des Inhalts oder Englisch. Lange oder Live-Formate: **Erinnerung** in Intervallen, mindestens nach Unterbrechungen. Mit Display zusätzlich Icon.	Podcast mit KI-Stimme: **Ansage zu Beginn** und nach dem Werbeblock, in der App zusätzlich das Icon.
4 **Kunst, Satire, Fiktion** Werk mit Deepfake — Betreiber	Art. 50 Abs. 4, Section 2 Hinweis auf KI-Inhalt, der das Werk nicht stört	**Icon passend zum Werk:** im Begleittext, Vor- oder Abspann. Digital auch **neben** Bild oder Player. Offline am Eingang, im Flyer oder auf der Verpackung.	Satire-Clip mit KI-Stimme eines Politikers: Hinweis **im Abspann** oder neben dem Player.
4 **KI-Text zu Themen von öffentlichem Interesse** — Betreiber	Art. 50 Abs. 4, Section 2 offenlegen – entfällt bei menschlicher Prüfung und redaktioneller Verantwortung	**Icon über dem Text** oder an der Überschrift; auch nur der KI-Teil kann markiert werden. **Ausnahme nur bei redaktioneller Prüfung:** Medienanbieter (Medienfreiheitsgesetz, VO 2024/1083) belegen sie über bestehende Prozesse, andere benennen Verantwortliche öffentlich.	Automatischer Wahl-Ticker ohne Gegenlesen: **Icon an der Überschrift**. Geprüfter Text: keine Pflicht.
5 **Für jeden Hinweis**	Art. 50 Abs. 5 klar, erkennbar, barrierefrei – spätestens beim ersten Kontakt	sofort sichtbar, ohne Klick · lange genug sichtbar · vor jedem Hintergrund lesbar · Screenreader und Alt-Text · hoher Kontrast · Normen: ETSI EN 301 549, WCAG 2.1	

Abbildung 4: Wie soll so gekennzeichnet werden? Der Verhaltenskodex zu Art. 50 (eigene Darstellung)

Der Verhaltenskodex beschreibt deutlich differenzierter als die KI-VO, **wann gekennzeichnet werden soll und wann Ausnahmen gelten**.

Zusätzlich empfiehlt er, **wie** je nach Einfluss von KI **zu kennzeichnen ist**. Dafür schlägt die Kommission **drei unterschiedliche Kennzeichnungslabels** vor.



„AI" weist allgemein auf den Einsatz von KI hin und kann durch zusätzliche Angaben ergänzt werden. **„AI GENERATED"** steht für **vollständig KI-generierte Inhalte**, bei denen sich der menschliche Beitrag auf die Eingabe von Prompts beschränkt. **„AI MODIFIED"** kennzeichnet ursprünglich von Menschen erstellte Inhalte, die **teilweise mit KI verändert wurden.**

Die **Unterzeichnung** des Verhaltenskodex ist **freiwillig**. Wer unterschreibt, profitiert aber von einer vereinfachten Aufsicht, denn Kommission und KI-Gremium haben den Kodex als geeignetes Instrument eingestuft, um die Einhaltung der Pflichten nachzuweisen.

Bis Ende Juli 2026 haben rund 190 Organisationen unterzeichnet, darunter KI-Anbieter wie Google, Meta, Microsoft und OpenAI sowie einzelne Nachrichtenportale aus Litauen und Griechenland. **Große deutsche Verlage oder Sender stehen bislang nicht auf der Liste.**

2.4 Die KI-Verordnung auf nationaler Ebene

Welche Behörden die KI-VO in Deutschland durchsetzen, regelt das KI-Marktüberwachungs- und Innovationsförderungsgesetz (KI-MIG), das am 29. Juli 2026 in Kraft getreten ist. Grundsätzlich ist die Bundesnetzagentur als zentrale Anlauf- und Beschwerdestelle zuständig, sofern keine andere Fachbehörde benannt ist. Dort ist zudem ein Koordinierungs- und Kompetenzzentrum angesiedelt, das die Zusammenarbeit der Behörden bündelt.

Für die Medienbranche gilt eine Ausnahme. Nach § 2 Absatz 8 KI-MIG liegt die **Marktüberwachung bei den nach Landesrecht zuständigen Behörden,** wenn Mediendiensteanbieter KI-Systeme für journalistische Zwecke oder zu Werbezwecken einsetzen. Nach dem Willen von Bund und Ländern sollen das die **Landesmedienanstalten** sein. Festgeschrieben werden soll das im Neunten Medienänderungsstaatsvertrag, der voraussichtlich im August 2027 in Kraft tritt.

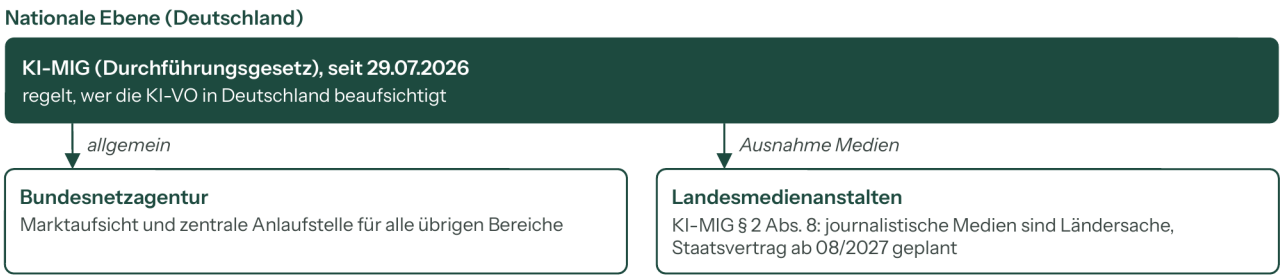


Abbildung 5: Wer ist für die Umsetzung der KI-Verordnung in Deutschland verantwortlich? (eigene Darstellung nach KI-MIG)

3. KI-Transparenz und Journalismus

Der Einsatz von KI-Technologien im Journalismus ist **keine neue Entwicklung** (Anderson, 2013; Dörr, 2016; Diakopoulos, 2019; Van Dalen, 2024). Seit vielen Jahren stützt sich die Automatisierung der Nachrichtenproduktion auf die natürliche Sprachgenerierung (Natural Language Generation, NLG). Dabei erzeugt Software aus **strukturierten Daten** automatisch **Texte in menschlicher Sprache** (Dörr, 2016).

KI-Technologien und Journalismus

Anfang der 2010er-Jahre zeigten Unternehmen zunehmend, dass sich auf diese Weise Nachrichtenbeiträge in großem Umfang erstellen lassen, vor allem **in Bereichen mit hochstrukturierten, regelmäßig aktualisierten Daten wie Sport, Finanzen, Wahlen und Wetter** (Dörr, 2016). In der wissenschaftlichen Debatte wurde diese Entwicklung als Teil einer **„quantitativen Wende" im Journalismus** beschrieben (Coddington, 2015) und unter Begriffen wie Datenjournalismus, computergestützter Journalismus oder „Roboterjournalismus" (Carlson, 2015) diskutiert. Algorithmen durchziehen seither die gesamte journalistische Wertschöpfungskette, von der Nachrichtenbeschaffung über Verifikation und Produktion bis zur Verbreitung und Personalisierung von Inhalten (Anderson, 2013; Cools & Diakopoulos, 2026).

Mit der Veröffentlichung von ChatGPT im November 2022 hat sich diese Entwicklung grundlegend verändert. Generative KI-Systeme erzeugen Texte, Bilder, Audio- und Videoinhalte mit minimaler menschlicher Eingabe. Sie entwerfen Artikel, erstellen Zusammenfassungen, übersetzen, formulieren Überschriften und produzieren visuelles Material (Cools & Diakopoulos, 2026; Altuncu et al., 2024). Anders als frühere Formen des automatisierten Journalismus, die auf strukturierten Daten basierten und formelhafte Ergebnisse lieferten, können generative Systeme **Inhalte in allen redaktionellen Formaten und Genres** erzeugen. Damit verschwimmt die Grenze zwischen werkzeuggestütztem und KI-generiertem Journalismus (Van Dalen, 2024).

Transparenz als journalistische Norm

Transparenz ist eine **etablierte berufliche Norm** im Journalismus. Sie soll der Öffentlichkeit Einblick in journalistische Produkte und interne Prozesse gewähren (Karlsson, 2010; Dubèl et al., 2025) und umfasst Angaben zu Autorinnen und Autoren, zu den Gründen der Berichterstattung sowie zu den verwendeten Quellen und Methoden (Craft & Heim, 2008; Ward, 2014; Dubèl et al., 2025).

Dubèl et al. (2025) nennen drei Gründe für die **wachsende Bedeutung von Transparenz im Journalismus.** Erstens hat das zunehmende Misstrauen gegenüber Medien diese dazu veranlasst, offener zu werden, um **Vertrauen zurückzugewinnen** (Karlsson, 2020). Zweitens hat sich Transparenz innerhalb des Berufsstands als Teil **professioneller Praxis** durchgesetzt (Loosen et al.,

2020). Drittens hat die Digitalisierung **neue Möglichkeiten** geschaffen, etwa durch Verlinkungen zu Quellen oder die laufende Aktualisierung von Beiträgen (Karlsson, 2010; Vos & Craft, 2017).

In einer Inhaltsanalyse von mehr als 27.000 Nachrichtenbeiträgen aus sechs niederländischen Medien unterscheiden Dubèl et al. (2025) **vier Transparenzroutinen**:

1. Autorentransparenz (Namen, Kontaktdaten, biografische Angaben)

2. Aktualisierungstransparenz (Zeitstempel, Hinweise auf Änderungen)

3. Quellentransparenz (Verlinkungen, Quellenangaben)

4. Produktionstransparenz (Einblicke in den Entstehungsprozess).

Alle untersuchten Medien setzen Transparenz in gewissem Maße um, doch Umfang und Art unterscheiden sich deutlich je nach Medientyp. Die **Umsetzung folgt also eher hausinternen Werten und Richtlinien als einem einheitlichen Standard.** Am seltensten praktizieren Redaktionen Produktionstransparenz, auch weil sie **befürchten**, die **Offenlegung ihrer internen Abläufe könne das Publikum eher verwirren als überzeugen**.

Genau in diese Kategorie fällt die Offenlegung von KI-Einsatz. Bemerkenswert ist, dass KI in der Analyse von Dubèl et al. (2025) nicht als eigener Gegenstand von Transparenz auftaucht. Technologie erscheint dort als Mittel, bestehende Transparenzpraktiken umzusetzen. Heute ist KI **ist selbst zum Gegenstand der Debatte um Transparenz geworden.**

4. KI-Kennzeichnung in der Praxis

4.1 Anforderungen an KI-Kennzeichnung

Neben den gesetzlichen Pflichten hat die Medienbranche **eigene Anforderungen an KI-Transparenz** formuliert.

Dazu zählt unter anderem der KI-Kodex der öffentlich-rechtlichen Rundfunkanstalten. Er wurde am 21. Januar 2026 veröffentlicht und ist als Teil des Reformstaatsvertrags verbindlich.

KI-Kodex der öffentlich-rechtlichen Sender

Der KI-Kodex wurde von ARD, ZDF, Deutschlandradio und Deutsche Welle gemeinsam vorgelegt. Die **sechs Grundsätze** reichen von Mehrwert und redaktioneller Kontrolle über KI-Kompetenz bis zu Nachhaltigkeit. Im Zentrum steht der **„Human in the Loop"-Ansatz**: Die publizistische Verantwortung bleibt beim Menschen, KI-generierte Inhalte werden klar gekennzeichnet. **Wie diese Kennzeichnung konkret aussieht, regelt der Kodex allerdings nicht.**

Der Deutsche Presserat hat den Pressekodex im September 2024 um Bestimmungen zu KI-generierten Inhalten ergänzt.

Anpassung des Pressekodex

„KI-generierte Bilder sind als Symbolbilder zu kennzeichnen".

Einen branchenweiten Standard zum Umgang mit KI-Transparenz oder KI-Kennzeichnung gibt es nicht. Solange branchenweite Standards fehlen, **entscheidet jedes Haus selbst, ob und wenn ja wie es seinen KI-Einsatz offenlegt.**

Selbstverpflichtende Maßnahmen der Medienhäuser

In den Interviews, die im Rahmen dieses Forschungsprojekts geführt wurden, zeichneten sich **drei Formate** ab, über die Medienhäuser ihren KI-Einsatz öffentlich erklären:

1. Leitlinien
2. Redaktionelle Beiträge in einem Transparenzblog / ein Artikel zu „In eigener Sache"
3. Hinweise im Impressum.

Ob das auch über die befragten Häuser hinaus gilt, habe ich in einer quantitativen Erhebung mit 60 Akteuren geprüft.

Leitlinien sind das verbreitetste Format der Selbstverpflichtung

35 der 60 Akteure nutzen mindestens eines der drei Formate. Leitlinien sind dabei das verbreitetste Format der Selbstverpflichtung.

Diese Erkenntnis deckt sich mit einer Empfehlung der Landesmedienanstalten: In ihrem Positionspapier „KI und Medien: Vielfalt stärken, Verantwortung regeln, Vertrauen wahren" vom März 2024 fordern sie Medienanbieter dazu auf, Leitlinien zu ihrem KI-Einsatz zu veröffentlichen.

Alle Dokumente zu KI-Transparenz thematisieren **Kennzeichnung als sinnvolle Maßnahme,** um auf KI-Einsatz aufmerksam zu machen. Allerdings wird nur **selten konkretisiert, wie die Kennzeichnung umgesetzt wird.**

4.2 Wie wird KI-Kennzeichnung umgesetzt?

Um diese Lücke zu schließen, habe ich recherchiert, wie Medienhäuser KI-Einsatz in ihren Angeboten tatsächlich kennzeichnen.

Aus den gefundenen Beispielen lassen sich **sechs Formen** ableiten: KI-Podcasts, KI-Zusammenfassungen, KI-Chatbots, KI-Hinweise unter Beiträgen und Bildern, KI-Personas sowie die Kennzeichnung von Deepfakes in der Berichterstattung.

Welche Form gewählt wird und welches Ziel die Kennzeichnung verfolgt, ergibt sich jeweils aus dem Kontext. **In den meisten Fällen legt sie den eigenen KI-Einsatz offen** und ist damit Produktionstransparenz im Sinne von Dubèl et al. (2025), also ein Einblick in den

Entstehungsprozess. Bei der Berichterstattung über **Deepfakes** ist die **Kennzeichnung hingegen ein Warnsignal**: Sie klärt darüber auf, dass ein Inhalt nicht echt ist und täuschen soll.

KI-Podcasts

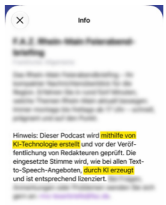
Bei **KI-generierten Podcasts** hört man den Hinweis auf den Einsatz von KI in der Regel direkt zu Beginn auf der Tonspur. Manche Häuser halten ihn zusätzlich schriftlich fest: im Titel der Folge, im Beschreibungstext oder auf dem Cover.



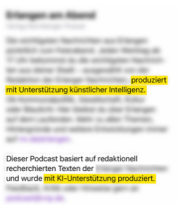
Kennzeichnung auf dem Podcast-Cover



Kennzeichnung im Episodentitel



Hinweis im Beschreibungstext des Podcasts

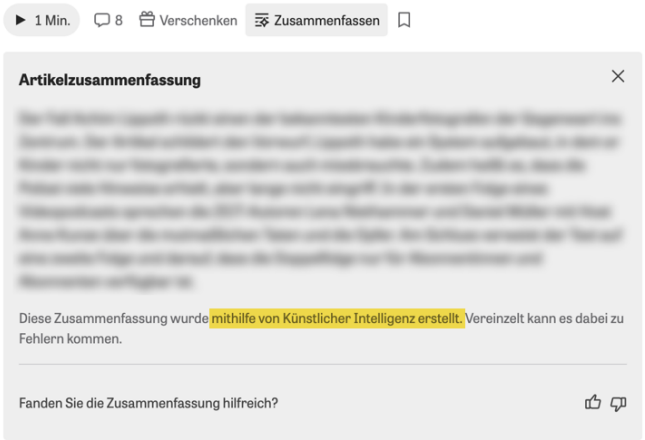


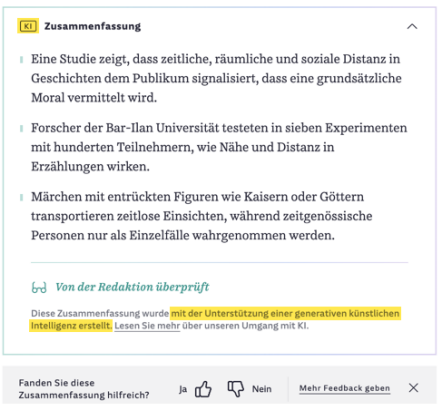
Hinweis im Beschreibungstext des Podcasts

Die Formulierungen reichen von „KI-produziert" über „mit KI-Unterstützung produziert" bis zu längeren Hinweisen, die erklären, wofür die KI eingesetzt wurde, etwa für die synthetische Stimme und den Recherche-Prozess.

KI-Zusammenfassungen

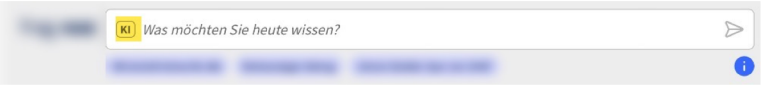
Bei **KI-Zusammenfassungen** von Artikeln ist die Kennzeichnung durch die Produktbeschreibung direkt verankert. Schon die Bezeichnung der Funktion, etwa „KI-Zusammenfassung", macht deutlich, dass der Text maschinell erstellt wurde. Ein separates Label ist deshalb oft nicht nötig. Teilweise ergänzen Redaktionen aber einen kurzen Zusatz, zum Beispiel, dass die Zusammenfassung automatisch erstellt wurde und Fehler enthalten kann.



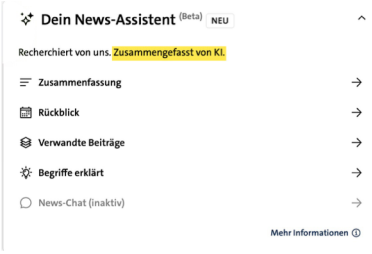


KI-Chatbots

Bei **Chatbots**, die Fragen auf Grundlage der eigenen Berichterstattung beantworten, oder **Assistenten**, die einen Artikel zusammenfassen, einordnen und auf verwandte Beiträge verweisen, ist die Kennzeichnung ebenfalls **durch die Produktbeschreibung** vorhanden.



KI-Kürzel im Eingabefeld des Chatbots



Hinweis „Zusammengefasst von KI" im News-Assistenten

KI-Hinweis unter Beiträgen / Bildern

Die einfachste Form der Kennzeichnung ist ein kurzer Texthinweis direkt am Inhalt. Bei Texten steht er meist als eigene Zeile am Ende des Beitrags. Bei Bildern nutzen Redaktionen den Bildnachweis, in dem ohnehin die Quelle steht. Dort taucht entweder das Kürzel „KI" auf oder der Name des KI-Werkzeugs, etwa „DALL-E". Beide Varianten nutzen bestehende Elemente des Beitrags und kommen ohne eigenes Symbol aus.



Hinweis unter dem Textbeitrag mit Verlinkung



Name des Bildgenerators im Bildnachweis



Hinweis unter dem Textbeitrag



Kürzel „KI" im Bildnachweis

KI-Persona

Ein Medienhaus hat sich dafür entschieden, **KI-gestützte Texte unter eigenen KI-Personas** zu veröffentlichen. Sie haben einen Namen, ein mit KI erzeugtes Profilbild und ein Autorenprofil.



Zusatz „(KI)" hinter dem Namen



Kürzel „(KIEC)" hinter dem Namen



Erklärung im Autorenprofil



Erklärung im Autorenprofil

Deepfakes

Die bisherigen Kennzeichnungspraktiken bezogen sich auf Inhalte, die Medienhäuser selbst mit KI erstellen. Ein anderer Fall von KI-Kennzeichnung lässt sich in der Berichterstattung beobachten: wenn Redaktionen **über KI-generierte Bilder oder Videos berichten, die Dritte in Umlauf gebracht haben**.

Entscheiden Redaktionen sich, diese Inhalte zu zeigen, um klarzustellen, dass es sich um Fälschungen handelt, kennzeichnen sie sie entsprechend.



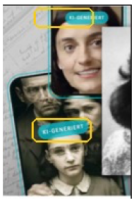
Stempel „KI-generiert" und „Manipuliert"



Wasserzeichen „KI-generiert" über dem ganzen Bild



Stempel „Fake" auf dem Bild


Label „KI-generiert" auf den Bildern

Diese Kennzeichnung ist keine Pflicht aus der KI-Verordnung, sondern eine freiwillige redaktionelle Entscheidung. Verpflichtet ist nach Artikel 50 Absatz 4 eigentlich **derjenige, der mit KI einen Deepfake erzeugt:** Er muss offenlegen, dass der Inhalt künstlich erzeugt oder verändert wurde. Kursieren Fälschungen ohne Kennzeichnung, hat sich der Urheber nicht an diese Pflicht gehalten, oder sie galt für ihn gar nicht, etwa weil er den Deepfake privat erstellt hat.

Hier zeigt sich ein **Widerspruch in der Logik der Deepfake-Definition:** Sie knüpft die Kennzeichnungspflicht an die Eignung zu täuschen. Wer aber täuschen will, hat wenig Anreiz, einen Deepfake zu kennzeichnen, denn mit dem Hinweis würde er nicht mehr täuschen.

Redaktionen springen hier freiwillig ein. Verpflichtet sind sie dazu nicht, doch Falschinformationen zu prüfen und richtigzustellen gehört ohnehin zu ihrem journalistischen Grundauftrag. Die Kennzeichnung am Bild oder Video ist dabei schlicht der Weg, auf dem sie ihr Prüfergebnis an das Publikum weitergeben.

Eine einheitliche Praxis für die Kennzeichnung von Deepfakes im Rahmen der Berichterstattung über Falschinformationen gibt es ebenfalls nicht. Manche Redaktionen bilden Fakes gar nicht ab, andere zeigen sie mit Stempeln, Wasserzeichen oder Labels. Auch die Wortwahl unterscheidet sich: Einige schreiben „Fake" quer über das Bild, andere unterscheiden zwischen „KI-generiert" und „manipuliert" oder legen „KI-generiert" als Wasserzeichen über das gesamte Bild.

5. Stimmen aus der Praxis: Wie Medienhäuser mit KI-Kennzeichnung umgehen

Kapitel 4 zeigt, wie Kennzeichnung aussieht. Warum ein Haus etwas kennzeichnet und etwas anderes nicht, ist von außen nicht zu erkennen. Dazu habe ich mit fünf Verantwortlichen aus Medienhäusern gesprochen: aus drei regionalen Mediengruppen, einem B2B-Fachmedium und einem Online-Nachrichtenportal.

Der 2. August war Anlass, sich mit Kennzeichnung auseinanderzusetzen, hat allerdings wenig verändert

Für die meisten meiner Gesprächspartner war die Auseinandersetzung mit der KI-VO und ihren Anforderungen eine Herausforderung, insbesondere wegen fehlender Konkretisierungen und unklarer Zuständigkeiten. Nach der Prüfung der Anforderungen wurde vielen klar: **So viel muss eigentlich gar nicht verändert werden.** Deepfakes, also Inhalte, die täuschen könnten, herzustellen, widerspricht ohnehin der journalistischen Praxis. Auch Prozesse für die Überprüfung von Inhalten und die Verantwortung für das, was veröffentlicht wird, gibt es bereits. Diskutiert wurden vor allem Grenzfälle wie KI-generierte Illustrationen. Dabei zeigte sich auch, dass die Kennzeichnungspflichten eher andere Abteilungen betreffen könnten als die Redaktion, etwa das Marketing. Die KI-VO war trotzdem **Anlass, diese Prozesse noch einmal zu überprüfen** und in den Redaktionsteams zu reflektieren. Bei vielen hat das außerdem dazu geführt, dass **interne Leitlinien** verfasst oder überarbeitet und teilweise auch veröffentlicht wurden.

Spannend war der Umgang mit Grenzfällen wie der **Artikelvertonung**, die in vielen Häusern bereits seit mehreren Jahren etabliert ist. Einige haben sich dafür entschieden, die Artikelvertonung zusätzlich visuell und auf der Tonspur zu kennzeichnen. Andere argumentieren, dass dies nicht notwendig sei, da es etablierte Praxis ist, dass dabei synthetische Stimmen zum Einsatz kommen.

Wo der Hinweis steht, hängt vom Ausspielweg ab

Entscheidend für den Ort der Kennzeichnung ist, **wo das Publikum einem Inhalt begegnet**. Die Artikelvertonung zum Beispiel läuft meist nur über die hauseigenen Ausspielwege, deshalb genügt dort ein sichtbarer Hinweis direkt am Audioplayer. Podcasts werden dagegen auch auf Drittplattformen ausgespielt und gehört. **Viele Häuser kennzeichnen sie deshalb an mehreren Stellen:** in der Beschreibung, am Anfang der Tonspur und in einem Fall zusätzlich als Text auf dem Cover.

Ein Medienhaus fasst seinen Grundsatz so zusammen: „**Gekennzeichnet wird direkt am Produkt, dort, wo KI tatsächlich drinsteckt**." Ein anderes betont: „**Wir haben nach dem Grundsatz gearbeitet, dass kein Nutzer an der Kennzeichnung vorbeikommen soll.**"

Transparenz hat eine Kehrseite und es ist unklar, wie Kennzeichnung wirkt

Ein Haus berichtet, dass ein KI-Hinweis die Reaktionen des Publikums verändert. Weist das Publikum auf Fehler hin, fällt der Ton bei gekennzeichneten Beiträgen deutlich schärfer aus als bei Beiträgen ohne Hinweis. **Viele lesen den Hinweis offenbar so, als habe die KI den Text erfunden, obwohl Menschen maßgeblich beteiligt waren.**

Ein Vergleich aus einem Gespräch bringt das Dilemma auf den Punkt: Der Dorfbäcker wirbt mit „handgemacht", nicht mit seiner Knetmaschine. Auch in anderen Gesprächen hieß es, dass **bislang unklar sei, wie Kennzeichnung beim Publikum wirkt**. Ob und wie weit freiwillig im Sinne der Produktionstransparenz (Dubèl et al. (2025) gekennzeichnet werden soll, ist daher umstritten.

Was sich daraus ableiten lässt

Gekennzeichnet wird vor allem dort, wo KI für das Publikum nicht mehr von menschlicher Arbeit zu unterscheiden ist, etwa bei synthetischen Stimmen. Wo die Grenze liegt, entscheidet jedes Haus selbst, meist in Form interner Leitlinien, die teilweise auch veröffentlicht werden.

Mehrere Gesprächspartnerinnen und -partner betonten zudem, dass **die größere Herausforderung** nicht in der Kennzeichnung eigener Inhalte liegt, sondern darin, **KI in externen Inhalten zu erkennen**, etwa in Leserfotos und Agenturbildern.

Aus den Gesprächen wird außerdem deutlich, dass die Auseinandersetzung mit Kennzeichnung nicht abgeschlossen ist. Nach der Umsetzung der ersten Hinweise beginnt nun eine Phase, in der die **Medienhäuser selbst prüfen wollen, wie Kennzeichnung beim Publikum ankommt** und wie unterschiedliche Umsetzungen verstanden werden. Dabei geht es auch um die Frage, wie sich **Kennzeichnung und Nutzerfreundlichkeit** verbinden lassen. Ähnlich wie bei Werbehinweisen ist davon auszugehen, dass wiederkehrende Hinweise das Publikum auf Dauer stören können. Einige Häuser denken deshalb über **Personalisierung** nach, zum Beispiel über eine **Opt-out-Möglichkeit** für Nutzerinnen und Nutzer, die regelmäßig denselben KI-Podcast hören und den Hinweis nicht jedes Mal brauchen. Wie ein solches Opt-out rechtlich einzuordnen ist, ist derzeit allerdings noch unklar.

6. Braucht es einen Standard für KI-Kennzeichnung?

Die Kapitel 4 und 5 zeigen: Im deutschsprachigen Raum kennzeichnet jedes Haus auf seine Weise. Damit stellt sich die Frage, **ob es einen gemeinsamen Standard braucht**.

Dafür spricht vor allem ein Argument: Das **digitale Informationsökosystem kennt keine Landesgrenzen**. Inhalte werden geteilt, eingebettet und über Plattformen weiterverbreitet, oft weit über das Angebot des Hauses hinaus, in dem sie entstanden sind. Damit Kennzeichnung überall verstanden wird, wäre eine **gemeinsame, wiedererkennbare Form** eigentlich notwendig.

Die Praxis sieht allerdings anders aus: Die Häuser kennzeichnen meist in ihrer Landessprache und mit eigenen Formulierungen. Das hat einen Vorteil, denn so lässt sich der **KI-Einsatz genau beschreiben und an Produkt und Publikum anpassen**.

Diese zwei Ansätze stehen sich also gegenüber: **Wiedererkennbarkeit über Sprach- und Ländergrenzen hinweg** und eine **möglichst genaue, auf Produkt und Publikum zugeschnittene Beschreibung des KI-Einsatzes**.

Dass Inhalte über Grenzen hinweg wirken, berücksichtigt auch die KI-VO mit dem Marktortprinzip (Art. 2 Abs. 1 KI-Verordnung): Sie gilt für alle, die KI-Systeme in der EU anbieten oder deren KI-Ergebnisse in der EU genutzt werden, unabhängig davon, wo sie ihren Sitz haben. Auf dieser Grundlage verfolgt die EU mit dem Verhaltenskodex ein Standardisierungsvorhaben, das für den ersten Ansatz steht: Er schlägt mit „AI", „AI GENERATED" und „AI MODIFIED" gemeinsame Kürzel vor, die **leicht wiederzuerkennen, aber zwangsläufig generisch** sind. Dazu kommt, dass viele die Option nutzen das allgemeine Label an ihre Landessprache anzupassen.

Ein weiteres Beispiel für ein Standardisierungsvorhaben ist das Forschungsprojekt **VITraDes**: Es soll den Stand von Regulierung und Selbstregulierung erfassen und daraus Leitlinien für die Kennzeichnung sowie Material für die journalistische Ausbildung entwickeln.

Forschungsprojekt VITraDes

VITraDes steht für „Vertrauenswürdige Information und Transparenz zu Algorithmen und lernenden Systemen im Journalismus". Das Projekt wird in Zusammenarbeit des Internationalen Zentrums für Ethik in den Wissenschaften (IZEW) der Universität Tübingen, des Instituts für Journalistik der TU Dortmund und des Bayreuth Center for Law and Technology der Universität Bayreuth durchgeführt und seit März 2026 drei Jahre lang mit rund einer Million Euro vom BMFTR gefördert. Assoziierte Partner sind unter anderem der BR, das Faktenforum von CORRECTIV, der KI-Hub des SWR, Studio 47 und der DJV.

Inhaltlich plant das Projekt eine Bestandsaufnahme von Regulierung und Selbstregulierung zur KI-Kennzeichnung im Journalismus, einschließlich der KI-Initiativen von Medienhäusern, Verlagen und Medienorganisationen. Daraus sollen praxisorientierte Leitlinien für die Kennzeichnung und Material für die journalistische Ausbildung entstehen. VITraDes gehört zur BMFTR-Kampagne „Desinformation – erkennen, verstehen, abwehren", die bis 2029 zunächst elf Forschungsprojekte fördert.

BMFTR. (2026, 24. März). *Auftakt in Berlin: Elf Forschungsprojekte gegen Desinformation gestartet*. https://www.forschung-it-sicherheit-kommunikationssysteme.de/service/aktuelles/auftakt-in-berlin-elf-forschungsprojekte-gegen-desinformation-gestartet. (Abgerufen am 25. September 2026).

TU Dortmund. (2026). *Transparente Kennzeichnung von KI im Journalismus*. https://www.tu-dortmund.de/detail/transparente-kennzeichnung-von-ki-im-journalismus-63003/. (Abgerufen am 25. September 2026).

In **Schweden** hat sich die Branche auf sieben gemeinsame, freiwillige Empfehlungen geeinigt. Führungskräfte aus 13 Medienhäusern erarbeiteten sie zwischen Dezember 2023 und Februar 2024 unter dem Dach des Branchenverbands Utgivarna, in dem öffentlich-rechtliche und private Sender sowie die Verlage organisiert sind. Laut dem **Branchenpapier** wird gekennzeichnet, wenn KI einen „erheblichen journalistischen Einfluss" hatte; was das heißt, entscheidet die Redaktion.

Routinewerkzeuge wie Transkription oder Rechtschreibprüfung brauchen keinen Hinweis. Als einheitliche Formulierung empfiehlt das Papier „med stöd av", also „mit Unterstützung von". **Varianten wie „durch" oder „zusammen mit" würden die Rolle der Menschen kleinreden. Von einem Symbol rät die Branche ab**: Es könne KI-gestützte Inhalte pauschal als weniger glaubwürdig erscheinen lassen, sei über Plattformen hinweg kaum einheitlich umzusetzen und veralte schnell.

Für die öffentlich-rechtlichen Medien sind auf europäischer Ebene die Aktivitäten der **Europäischen Rundfunkunion (EBU)** relevant. Einen gemeinsamen Kennzeichnungsstandard für ihre Mitglieder hat sie nicht, sie vertritt aber gemeinsame Positionen. In einer Analyse von 2024 hält sie fest, dass echte Transparenz mehr verlangt als Labels und dass bloße Labels Vertrauen sogar untergraben können. Im Juni 2026 forderte sie die EU-Kommission gemeinsam mit Verbänden privater Sender auf, Deepfakes in den Leitlinien zu Art. 50 enger zu fassen und **keine Dauerkennzeichnung zu verlangen, weil allgegenwärtige Hinweise niemand mehr beachte**. In einem Beitrag vom März 2026 verlangt sie von KI-Anbietern, ihre Quellen offenzulegen und auf Originalinhalte zu verlinken; Kennzeichnung und Herkunftsnachweise wie C2PA könnten Vertrauen zurückgewinnen. **Den Standard C2PA treibt die EBU mit einer eigenen Arbeitsgruppe voran.**

Herkunftsnachweise durch C2PA

C2PA ist ein offener technischer Standard, der Fotos, Videos und Audio **mit kryptografisch gesicherten Angaben zu Herkunft und Bearbeitung** versieht. Entwickelt wird C2PA von der „Coalition for Content Provenance and Authenticity", die 2021 unter anderem von Adobe, BBC und Microsoft gegründet wurde und heute zahlreiche Medien- und Technologieunternehmen vereint.

Die EBU treibt seine Einführung im Rundfunk mit einer eigenen Arbeitsgruppe voran, in der Mitglieder die technischen Anforderungen klären, Umsetzungswege bewerten und Leitlinien entwickeln, wie sich Herkunftsnachweise in bestehende Produktions- und Verbreitungsabläufe einbauen lassen.

Auf der EBU-Nachrichtenplattform News Pilot tragen neu veröffentlichte Bilder seit Dezember 2024 C2PA-Nachweise; Audio und Video sollen folgen. Gemeinsam mit CBC/Radio-Canada hat die EBU zudem einen quelloffenen Videoplayer entwickelt, der prüft, ob ein Inhalt unverändert ist und von einem verifizierten Herausgeber stammt; im April 2026 wurde er mit dem NAB Technology Innovation Award ausgezeichnet.

Am weitesten ist France Télévisions: Der Sender versieht seit 2025 seine beiden täglichen Hauptnachrichtenausgaben mit C2PA-Signaturen und bezeichnet sich damit als ersten Sender, der das Protokoll systematisch im Tagesgeschäft einsetzt.

7. Fazit & Ausblick

Die Debatte über die Transparenzpflichten der KI-Verordnung und der eigene Anspruch, gegenüber dem Publikum offenzulegen, wie man arbeitet, haben viele Medienhäuser dazu gebracht, sich mit KI-Kennzeichnung auseinanderzusetzen. Redaktionen haben Leitlinien geschrieben und entschieden, was sie kennzeichnen und wie.

Die KI-VO lässt ihnen dabei viel Spielraum. Ob ein Bild, Video oder Audio über seine Echtheit täuschen kann, müssen die Häuser **in der Logik der Deepfake-Definition**, also in einer Täuschungsfolgenabschätzung, selbst bewerten. Ob ein KI-gestützter Text ausreichend redaktionell verantwortet ist, beurteilen sie ebenfalls selbst. Dabei gilt das Prinzip, das der Verhaltenskodex vorgibt: **Entweder dokumentieren sie die Prüfung nach innen, oder sie kennzeichnen nach außen.**

Braucht es einen Standard?

Entsprechend vielfältig sind die Lösungen. Manche Häuser nennen das KI-Werkzeug im Bildnachweis, andere setzen den Hinweis „KI" direkt ans Produkt, weisen zu Beginn eines Podcast auf die synthetische Stimme hin oder schreiben den Hinweis auf das Cover. Ein Haus veröffentlicht KI-gestützte Texte sogar unter eigenen KI-Personas. **Einen branchenweiten Standard gibt es bislang nicht.** Auch die Labels, die der europäische Verhaltenskodex vorschlägt, haben sich in Deutschland noch nicht durchgesetzt: Kein großer deutscher Verlag oder Sender hat den Kodex bisher unterzeichnet, und in der Recherche fand sich kein Beispiel, in dem die vorgeschlagenen Labels verwendet werden.

Darin zeigt sich ein Spannungsfeld: Ein **einheitliches Kürzel** wie „AI" ist über Sprach- und Ländergrenzen hinweg leicht wiederzuerkennen, sagt aber wenig darüber, was die KI konkret getan hat. **Eigene Formulierungen** beschreiben den KI-Einsatz genauer und lassen sich an Produkt und Publikum anpassen, sind dafür aber von Haus zu Haus verschieden.

Vorerst entwickelt jedes Haus eigene Lösungen. Dieses Muster ist aus der Transparenzforschung bekannt: Dubèl et al. (2025) zeigen für etablierte Transparenzroutinen wie Autoren- oder Quellenangaben, dass Redaktionen zwar **ähnliche Ziele verfolgen, die Umsetzung sich aber deutlich unterscheidet** und eher hausinternen Werten folgt als einem gemeinsamen Standard. Bei der KI-Kennzeichnung zeichnet sich dieselbe Entwicklung ebenfalls ab.

Kennzeichnung, Herkunftsnachweis oder beides?

Neben Label und Beschreibung gibt es eine **dritte Ebene, in der die Medienbranche aktiv ist: Herkunftsnachweise** wie C2PA. Sie halten fest, wie ein digitales Artefakt entstanden ist und welche Schritte es durchlaufen hat, von der Aufnahme über die Bearbeitung bis zur Veröffentlichung. Diese Historie ist nicht nur maschinenlesbar, sondern lässt sich auch für das Publikum sichtbar machen. Herkunftsnachweise ermöglichen damit eine Transparenz, die über einen einzelnen Hinweis weit hinausgeht.

Label, redaktionelle Beschreibung und Herkunftsnachweis stehen deshalb nicht in Konkurrenz, sondern beantworten unterschiedliche Fragen des Publikums: ob KI beteiligt war, welche Rolle sie gespielt hat und wie ein Inhalt entstanden ist.

Wie wirkt Kennzeichnung?

Welcher Ansatz besser trägt, hängt davon ab, wie Kennzeichnung beim Publikum ankommt. **Genau das kann bislang niemand sicher beantworten.**

Es gibt Hinweise darauf, dass das Publikum Kennzeichnung grundsätzlich erwartet: In einer Befragung der Medienanstalten aus 2024 halten **77 Prozent** die Kennzeichnung KI-erstellter Beiträge für sehr wichtig. Dieselbe Studie zeigt allerdings, **dass ein Hinweis nur wirkt, wenn er auch wahrgenommen wird**. Bei einem Text erkannten mit Hinweis 60 Prozent den KI-Einsatz, ohne Hinweis nur 21 Prozent. Bei einem Audiobeitrag machte der Hinweis dagegen kaum einen Unterschied.

Ob Kennzeichnung Vertrauen stärkt, ist unklar. Mehrere Experimente deuten in die entgegengesetzte Richtung. Eine niederländische Studie aus 2025 fand für alle getesteten KI-Hinweise negative Effekte auf das Vertrauen in Nachrichten. Altay und Gilardi (2024) zeigen, dass Schlagzeilen mit KI-Label als weniger zutreffend gelten, selbst wenn sie wahr sind oder von Menschen stammen, weil Lesende eine vollständige Automatisierung vermuten.

Uneins ist die Forschung zudem darüber, wie ausführlich ein Hinweis sein sollte, und damit genau über die Frage zwischen knappem Label und genauer Beschreibung. In einer US-Studie (2025) stieg das Vertrauen, je mehr der Hinweis erklärte. Ein weiteres Experiment aus den Niederlanden (2026) kam zum gegenteiligen Ergebnis.

Solange die Wirkungsfrage offen ist, lässt sich auch nicht beantworten, ob ein einheitlicher Standard sinnvoll wäre und wie er aussehen müsste. Mehrere Häuser wollen die Wirkung ihrer Hinweise deshalb nun selbst testen und Kennzeichnung mit Nutzerfreundlichkeit in Einklang bringen. **Ein tragfähiger Ansatz für Transparenz im Journalismus wird dabei Label, Beschreibung und Herkunftsnachweis zusammendenken müssen.**

Literatur

Altay, S., & Gilardi, F. (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), pgae403. https://doi.org/10.1093/pnasnexus/pgae403

Altuncu, E., Franqueira, V. N. L., & Li, S. (2024). Deepfake: Definitions, performance metrics and standards, datasets, and a meta-review. *Frontiers in Big Data*, *7*, 1400024. https://doi.org/10.3389/fdata.2024.1400024

Anderson, C. W. (2013). Towards a sociology of computational and algorithmic journalism. *New Media & Society*, *15*(7), 1005–1021. https://doi.org/10.1177/1461444812465137

ARD, ZDF, Deutschlandradio & Deutsche Welle. (2026, 21. Januar). *KI-Kodex*. https://www.ard.de/die-ard/organisation-der-ard/rechtsgrundlagen/ki-kodex-104.html (Abgerufen am 30. September 2026)

BMFTR. (2026, 24. März). *Auftakt in Berlin: Elf Forschungsprojekte gegen Desinformation gestartet*. https://www.forschung-it-sicherheit-kommunikationssysteme.de/service/aktuelles/auftakt-in-berlin-elf-forschungsprojekte-gegen-desinformation-gestartet (Abgerufen am 30. September 2026)

Bundesministerium für Digitales und Staatsmodernisierung. (2026, 29. Juli). *Neues KI-Gesetz tritt in Kraft* [Pressemitteilung]. https://bmds.bund.de/aktuelles/pressemitteilungen/detail/neues-ki-gesetz-tritt-in-kraft (Abgerufen am 30. September 2026)

Carlson, M. (2015). The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital Journalism*, *3*(3), 416–431. https://doi.org/10.1080/21670811.2014.976412

Coalition for Content Provenance and Authenticity. (2021, 22. Februar). *C2PA founding press release*. https://c2pa.org/c2pa-founding-press-release/ (Abgerufen am 30. September 2026)

Coddington, M. (2015). Clarifying journalism's quantitative turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism*, *3*(3), 331–348. https://doi.org/10.1080/21670811.2014.976400

Cools, H., & Diakopoulos, N. (2026). Uses of generative AI in the newsroom: Mapping journalists' perceptions of perils and possibilities. *Journalism Practice*, *20*(3), 878–896. https://doi.org/10.1080/17512786.2024.2394558

Craft, S., & Heim, K. (2008). Transparency in journalism: Meanings, merits, and risks. In L. Wilkins & C. G. Christians (Hrsg.), *The handbook of mass media ethics*. Routledge.

Deutscher Bundestag. (2026, 23. März). *Zuspruch für Bundesnetzagentur als KI-Marktüberwachungsbehörde*. https://www.bundestag.de/dokumente/textarchiv/2026/kw13-pa-digitales-1155576 (Abgerufen am 30. September 2026)

Deutscher Presserat. (2024, 18. September). *Redaktionen auch für KI-generierte Inhalte ethisch verantwortlich* [Pressemitteilung]. https://www.presserat.de/presse-nachrichten-details/redaktionen-auch-fuer-ki-generierte-inhalte-ethisch-verantwortlich.html (Abgerufen am 30. September 2026)

Deutscher Presserat. (2025). *Publizistische Grundsätze (Pressekodex)* (Fassung vom 19. März 2025). https://www.presserat.de/pressekodex.html (Abgerufen am 30. September 2026)

Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press.

Die Medienanstalten. (2024). *Transparenz-Check: Wahrnehmung von KI-Journalismus*. https://lfk.de/fileadmin/PDFs/Publikationen/Studien/Transparenz-Check/Transparenz-Check_KI-Journalismus.pdf (Abgerufen am 30. September 2026)

Die Medienanstalten. (2024, 20. März). *KI und Medien: Vielfalt stärken, Verantwortung regeln, Vertrauen wahren* [Positionspapier]. https://www.die-medienanstalten.de/fileadmin/user_upload/die_medienanstalten/Service/Positionspapiere/20240320_KI_und_Medien.pdf (Abgerufen am 30. September 2026)

Dörr, K. N. (2016). Mapping the field of algorithmic journalism. *Digital Journalism*, *4*(6), 700–722. https://doi.org/10.1080/21670811.2015.1096748

Dubèl, R., Boukes, M., & Trilling, D. (2025). To disclose or not to disclose: A comprehensive analysis into the article transparency of news websites. *Journalism & Mass Communication Quarterly*. Vorabveröffentlichung online. https://doi.org/10.1177/10776990241306411

EBU Technology & Innovation. (o. J.). *Content provenance and authenticity* [Arbeitsgruppe]. https://tech.ebu.ch/groups/c2pa_workflows (Abgerufen am 30. September 2026)

EBU Technology & Innovation. (2024, 18. Dezember). *EBU joins early implementers of new standard to track media content provenance*. https://tech.ebu.ch/news/2024/12/ebu-joins-early-implementers-of-new-standard-to-track-content-provenance (Abgerufen am 30. September 2026)

EBU Technology & Innovation. (2025, 21. August). *France Télévisions leads the way with operational adoption of C2PA*. https://tech.ebu.ch/news/2025/08/france-televisions-leads-the-way-with-operational-adoption-of-c2pa (Abgerufen am 30. September 2026)

EBU Technology & Innovation. (2026, 21. April). *EBU and CBC/Radio-Canada win NAB Technology Innovation Award for C2PA-enabled video player*. https://tech.ebu.ch/news/2026/ebu-and-cbc-radio-canada-win-nab-technology-innovation-award-for-c2pa-enabled-video-player (Abgerufen am 30. September 2026)

epd medien. (2026, 13. Februar). Medienanstalten begrüßen Pläne der Bundesregierung zur KI-Aufsicht. https://medien.epd.de/article/4275 (Abgerufen am 30. September 2026)

Europäische Kommission. (o. J.). *Article 50: Transparency obligations for providers and deployers of certain AI systems*. AI Act Service Desk. https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-50 (Abgerufen am 30. September 2026)

Europäische Kommission. (2025, 8. Dezember). *Vollständige Liste der Teilnehmer der Arbeitsgruppen 1 und 2*. https://ec.europa.eu/newsroom/dae/redirection/document/122890 (Abgerufen am 30. September 2026)

Europäische Kommission. (2026). *Code of Practice on Transparency of AI-generated Content*. https://digital-strategy.ec.europa.eu/en/policies/code-practice-ai-generated-content (Abgerufen am 30. September 2026)

Europäische Kommission. (2026, 8. Mai a). *Draft of the guidelines on the implementation of the transparency obligations for certain AI systems under Article 50 of the AI Act*. https://digital-strategy.ec.europa.eu/en/library/draft-guidelines-implementation-transparency-obligations-certain-ai-systems-under-article-50-ai-act (Abgerufen am 30. September 2026)

Europäische Kommission. (2026, 8. Mai b). *Consultation on the draft guidelines on transparency obligations under the AI Act*. https://digital-strategy.ec.europa.eu/en/consultations/consultation-draft-guidelines-transparency-obligations-under-ai-act (Abgerufen am 30. September 2026)

Europäische Kommission. (2026, 10. Juni). *Commission publishes Code of Practice on marking and labelling AI-generated content*. https://digital-strategy.ec.europa.eu/en/news/commission-publishes-code-practice-marking-and-labelling-ai-generated-content (Abgerufen am 30. September 2026)

Europäische Kommission. (2026, 9. Juli). *Commission opinion on the assessment of the Code of Practice on Transparency of AI-generated Content*. https://digital-strategy.ec.europa.eu/en/library/commission-opinion-assessment-code-practice-transparency-ai-generated-content (Abgerufen am 30. September 2026)

Europäische Kommission. (2026, 20. Juli). *Guidelines on transparency obligations for providers and deployers of AI systems*. https://digital-strategy.ec.europa.eu/en/library/guidelines-transparency-obligations-providers-and-deployers-ai-systems (Abgerufen am 30. September 2026)

Europäische Kommission. (2026, 24. Juli). *Transparency obligations under Article 50 of the AI Act* [FAQ]. https://digital-strategy.ec.europa.eu/en/faqs/transparency-obligations-under-article-50-ai-act (Abgerufen am 30. September 2026)

Europäische Kommission. (2026, 29. Juli). *Guidelines on transparency obligations for providers and deployers of certain AI systems*. https://digital-strategy.ec.europa.eu/en/policies/guidelines-transparency-ai-generated-content (Abgerufen am 30. September 2026)

Europäische Kommission. (2026, 31. Juli). *Strong backing for the Code of Practice on Transparency of AI-generated Content*. https://digital-strategy.ec.europa.eu/en/news/strong-backing-code-practice-transparency-ai-generated-content (Abgerufen am 30. September 2026)

Europäische Kommission. (2026, 2. August). *Safer and more transparent AI*. https://commission.europa.eu/news-and-media/news/safer-and-more-transparent-ai-2026-08-02_en (Abgerufen am 30. September 2026)

European Broadcasting Union. (2024, 13. September). *AI regulation: Are public service media's needs being met?*. https://www.ebu.ch/guides/open/report/ai-regulation-public-service-media-analysis (Abgerufen am 30. September 2026)

European Broadcasting Union. (2026, 12. März). *The 4 burning policy questions for AI and the media sector*. https://www.ebu.ch/news/2026/03/the-4-burning-policy-questions-for-ai-and-the-media-sector (Abgerufen am 30. September 2026)

European Broadcasting Union. (2026, 16. Juni). *European media call on the Commission to refrain from unhelpful labelling obligations in the AI Act's transparency guidelines*. https://www.ebu.ch/news/2026/06/european-media-call-on-the-commission-to-refrain-from-unhelpful-labelling-obligations-in-the-ai-act-s-transparency-guidelines (Abgerufen am 30. September 2026)

Gesetz zur Marktüberwachung und Innovationsförderung von künstlicher Intelligenz (KI-Marktüberwachungs- und Innovationsförderungsgesetz – KI-MIG) vom 22. Juli 2026. *BGBl. 2026 I*, Nr. 223. https://www.gesetze-im-internet.de/ki-mig/BJNR0DF0B0026.html (Abgerufen am 30. September 2026)

Karlsson, M. (2010). Rituals of transparency: Evaluating online news outlets' uses of transparency rituals in the United States, United Kingdom and Sweden. *Journalism Studies, 11*(4), 535–545. https://doi.org/10.1080/14616701003638400

Karlsson, M. (2020). Dispersing the opacity of transparency in journalism on the appeal of different forms of transparency to the general public. *Journalism Studies, 21*(13), 1795–1814. https://doi.org/10.1080/1461670X.2020.1790028

Landesregierung Schleswig-Holstein. (2026, 3. Juli). Entwurf eines Neunten Staatsvertrags zur Änderung medienrechtlicher Staatsverträge (Digitale Medien-Staatsvertrag) [Unterrichtung 20/360]. Schleswig-Holsteinischer Landtag. https://www.landtag.ltsh.de/infothek/wahl20/unterrichtungen/00300/unterrichtung-20-00360.pdf (Abgerufen am 30. September 2026)

Loosen, W., Reimer, J., & Hölig, S. (2020). What journalists want and what they ought to do (in)congruences between journalists' role conceptions and audiences' expectations. *Journalism Studies, 21*(12), 1744–1774. https://doi.org/10.1080/1461670X.2020.1790026

Mattis, N. M., Kieslich, K., & de Vreese, C. (2025). *Feeling iffy about generative AI: When journalists disclose AI use, trust in news is lower* [Preprint]. OSF. https://doi.org/10.31219/osf.io/tmzq4_v1

Nordic AI Journalism & Utgivarna. (2024). *AI transparency in journalism*. https://www.nordicaijournalism.com/_files/ugd/efd447_06c5eb91a6084cf78da840af828ce845.pdf (Abgerufen am 30. September 2026)

Prajod, P., Cools, H., Röggla, T., Puttur Venkatraj, K., Kusters, A., ElKattan, A., Cesar, P., & El Ali, A. (2026). *Full disclosure, less trust? How the level of detail about AI use in news writing affects readers' trust* [Preprint]. arXiv. https://arxiv.org/abs/2601.09620

TU Dortmund. (2026). *Transparente Kennzeichnung von KI im Journalismus*. https://www.tu-dortmund.de/detail/transparente-kennzeichnung-von-ki-im-journalismus-63003/ (Abgerufen am 30. September 2026)

TV 2 Danmark & TV 2 Regionerne. (2024). *TV 2 Danmark og TV 2 Regionernes fælles retningslinjer for anvendelse af generativ AI i nyhedsjournalistik og dokumentarer*. https://obj3739.public-dk6.clu4.obj.storagefactory.io/syd-documents/tv_2_retningslinjer_for_ai_a4_lang_final.pdf (Abgerufen am 30. September 2026)

Utgivarna. (2024, 14. März). *Rekommendationer för AI-transparens inom journalistik* [Pressemitteilung]. https://utgivarna.se/pressmeddelande/rekommendationer-for-ai-transparens-inom-journalistik/ (Abgerufen am 30. September 2026)

Van Dalen, A. (2024). Revisiting the algorithms behind the headlines: How journalists respond to professional competition of generative AI. *Journalism Practice*. Vorabveröffentlichung online. https://doi.org/10.1080/17512786.2024.2389209

Verordnung (EU) 2024/1083 des Europäischen Parlaments und des Rates vom 11. April 2024 zur Schaffung eines gemeinsamen Rahmens für Mediendienste im Binnenmarkt (Europäisches Medienfreiheitsgesetz). (2024). *Amtsblatt der Europäischen Union*, L, 2024/1083. http://data.europa.eu/eli/reg/2024/1083/oj

Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (KI-Verordnung). (2024). *Amtsblatt der Europäischen Union*, L, 2024/1689. https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:32024R1689

Verordnung (EU) 2026/1744 des Europäischen Parlaments und des Rates vom 8. Juli 2026 zur Änderung der Verordnungen (EU) 2024/1689, (EU) 2018/1139 und (EU) 2023/1230 im Hinblick auf die Vereinfachung der Umsetzung harmonisierter Vorschriften für künstliche Intelligenz (Digital-Omnibus-Verordnung zur KI). (2026). *Amtsblatt der Europäischen Union*, L, 2026/1744. https://eur-lex.europa.eu/eli/reg/2026/1744/oj

Vos, T. P., & Craft, S. (2017). The discursive construction of journalistic transparency. *Journalism Studies, 18*(12), 1505–1522. https://doi.org/10.1080/1461670X.2015.1135754

Walsh, L., & Toff, B. (2025, 30. Juli). *How AI disclosures in news help — and also hurt — trust with audiences*. Trusting News. https://trustingnews.org/new-research-how-ai-disclosures-in-news-help-and-also-hurt-trust-with-audiences/ (Abgerufen am 30. September 2026)

Ward, S. J. A. (2014). The magical concept of transparency. In L. Zion & D. Craig (Hrsg.), *Ethics for digital journalists: Emerging best practices* (S. 45–58). Routledge.

White Paper

Transparente Kennzeichnung von KI im Journalismus

Oktober 2026

Danksagung & Kontakt

Mein Dank gilt dem Media Lab Bayern, das dieses White Paper durch seine Förderung möglich gemacht hat.

Ebenso herzlich danke ich meinen Gesprächspartnerinnen und Gesprächspartnern für ihre Zeit, die Einblicke in ihre Arbeit und ihre Expertise zum Thema.

Kontakt: LinkedIn