



Branchenerhebung Kennzeichnungsmöglichkeiten von Künstlicher Intelligenz in der Medienwirtschaft

Stand: 11. November 2024

Rundfunk und Telekom Regulierungs-GmbH (RTR-GmbH)

Mariahilfer Straße 77–79
1060 Wien, Österreich
www.rtr.at

E: rtr@rtr.at
T: +43 1 58058 - 0

FN 208312t, HG Wien
UID-Nr.: ATU43773001

Inhaltsverzeichnis

1.	Hintergrund und Zielsetzungen	3
2.	Grundprinzipien der KI-Nutzung und Kennzeichnung	5
2.1.	Gesetzmäßige Front-End Kennzeichnung	5
2.2.	Ergänzende Back-End Kennzeichnung in Metadaten	5
2.3.	Einheitliche Anwendung	6
2.4.	Eindeutiger Kennzeichnungsbegriff	6
2.5.	Iterativer Kennzeichnungs- und Nutzungsansatz	6
3.	Zeitpunkt der Kennzeichnungspflicht	6
4.	Ansätze für spezifische Medienformen	7
4.1.	Audio	7
4.2.	Video	7
4.3.	Bild	7
4.4.	Text	8
5.	Künstlerische, kreative, satirische, fiktionale und analoge Werke	8
6.	Ausblick	8

1. Hintergrund und Zielsetzungen

In einer zunehmend digitalisierten Welt spielt künstliche Intelligenz (KI) eine immer größere Rolle in der Produktion, Verbreitung und Konsumation von Medien. Die jüngsten Entwicklungen unterstreichen deutlich sowohl die Möglichkeiten als auch die Herausforderungen für den Mediensektor. In einer Ära, in der technologische Fortschritte die Medienlandschaft grundlegend verändern, ist Transparenz einer jener Grundsätze, die Integrität und Glaubwürdigkeit von Medieninhalten im Kern berühren. Folglich wird auch die Kennzeichnung von synthetischen Inhalten als adäquates Mittel zur Stärkung der Integrität des Informationsökosystems und des Vertrauens von Medienkonsumierenden international verfolgt. Erste Ansätze für technische und inhaltliche Kennzeichnungen sind bereits plattformübergreifend am Markt zu beobachten.

Auch der Artificial Intelligence Act (AI Act) setzt im Hinblick auf synthetische Inhalte bei den Grundprinzipien der Transparenz und Offenlegung an. Konkret definiert der AI Act erstmals rechtlich den Begriff 'Deep Fake' als einen durch KI erzeugten oder manipulierten Bild-, Ton- oder Videoinhalt, der wirklichen Personen, Gegenständen, Orten, Einrichtungen oder Ereignissen ähnelt und einer Person fälschlicherweise als echt oder wahrheitsgemäß erscheinen würde (Art 3 Z 60 AI Act). Indem der AI Act durch die Deep Fake Definition auf eine mögliche Irreführungseignung des synthetischen Contents abstellt, verfolgt er einen wirkungsorientierten Ansatz, der sich auf die Abwendung potenzieller schädlicher Inhalte konzentriert.

Zur Schadensabwehr von Deep Fakes wird durch den AI Act ein prozessorientierter Ansatz verfolgt, der die Offenlegung der KI-Nutzung zur Erstellung oder Bearbeitung erfordert. Zur Kennzeichnung verpflichtet sind die sogenannten 'Betreiber', folglich natürliche und juristische Personen, welche die KI-Systeme verwenden. Damit kombiniert der AI Act insgesamt betrachtet einen wirkungs- und prozessorientierten Ansatz, bei dem nur als Deep Fake definierte potenziell irreführende Inhalte, die mit KI erstellt oder bearbeitet wurden, verpflichtend gekennzeichnet werden, müssen.

Die Medienform Text wird aufgrund ihrer Besonderheiten separat behandelt. Mittels KI erzeugte oder bearbeitete Texte von Betreibern, welche veröffentlicht werden, um die Öffentlichkeit über Angelegenheiten von öffentlichem Interesse zu informieren, müssen als künstlich erzeugt oder bearbeitet gekennzeichnet werden. Die Pflicht zur Offenlegung besteht nicht, wenn diese Inhalte einem Verfahren der menschlichen Überprüfung oder redaktionellen Kontrolle unterzogen wurden und wenn eine natürliche oder juristische Person die redaktionelle Verantwortung für die Veröffentlichung der Inhalte trägt.

Trotz des international verbreiteten Ansatzes zur Einführung von Kennzeichnungspflichten KI-basierter Medien ist unklar, welche Begriffe für diese verwendet werden sollen und wie effektive und zielführende Labels aussehen können. KI-basierte Medien sind nicht monolithisch und unterscheiden sich

vielfältig in Art und Grad der algorithmischen Eingriffe sowie in ihren möglichen Konsequenzen für die Gesellschaft. Hervorzuheben ist auch, dass nicht alle KI-basierten Medien von Natur aus irreführend sind (zB digitale Kunst), und auch nicht alle Formen irreführender Inhalte mittels KI produziert werden (zB Verwendung von Informationen außerhalb des Originalkontext). Zudem ist das Niveau an Digital- und Medienbildung unterschiedlich ausgeprägt. Die Kenntnisse und die Vertrautheit von Medienkonsumierenden mit KI variiert stark und werfen weiter die Frage auf, wie und unter welchen Bedingungen spezifische Labels für die allgemeine Öffentlichkeit verständlich sein werden.

In ihrer Funktion als KI-Servicestelle und zentrale Plattform für die Vorbereitungen der Umsetzung des AI Acts in Österreich, hat die Rundfunk und Telekom Regulierungs-GmbH (RTR-GmbH), Fachbereich Medien proaktiv einen umfassenden Stakeholder-Prozess initiiert, der sich der praktischen Ausgestaltung der Kennzeichnungspflichten in den Medien widmete. Im Zeitraum vom 5. Mai 2024 bis zum 15. Mai 2024 wurden Stellungnahmen und Empfehlungen zur Frage, wie KI-basierte Medieninhalte zielführend gemäß den Vorgaben des AI Acts transparent gemacht werden können, gesammelt. Die rund 200 Rückmeldungen von mehr als 100 Teilnehmenden wurden als Grundlage für einen interaktiven Workshop aufbereitet, um Ansätze für eine gemeinsame, einheitliche Vorgehensweise im Mediensektor bei der Umsetzung der Pflichten zur Kennzeichnung zu entwickeln.

Diese Branchenerhebung, betreffend der Kennzeichnung KI-basierter Medien für den Mediensektor, ist das Ergebnis der beschriebenen umfassenden Vorarbeiten. Zielsetzung dieses Dokuments ist es, sowohl Transparenz als auch Vertrauen im Hinblick auf die Nutzung von KI bei der Medienproduktion zu fördern, und den österreichischen Mediensektor beim verantwortungsvollen und positiven Einsatz von KI zu unterstützen. Die KI-Servicestelle wird weiterhin in engem Austausch mit allen Stakeholdern bleiben, um die Strategien zur Transparenz kontinuierlich zu überprüfen und anzupassen, damit diese höchsten Standards entsprechen.

2. Grundprinzipien der KI-Nutzung und Kennzeichnung

Aus den Anforderungen des AI Acts im Zusammenwirken mit den Ergebnissen des Stakeholder-Prozesses haben sich die folgenden Grundprinzipien für eine effektive und verantwortungsvolle Kennzeichnung von KI-basierten Medieninhalten ergeben.

2.1. Gesetzmäßige Front-End Kennzeichnung

Kennzeichnungen wirken sich weit über den Kontext des jeweiligen Einzelinhalts aus. Belegt sind Effekte, dass Medienkonsumierende dazu neigen könnten, weniger Vertrauen in gekennzeichnete Inhalte zu haben, unabhängig von deren tatsächlicher Wahrhaftigkeit oder Herkunft. Ebenso zu beachten ist das Einsetzen eines 'impliziten Wahrheitseffekts', der die Glaubwürdigkeit nicht-gekennzeichneter Inhalte erhöhen könnte, selbst wenn diese unzuverlässig sein sollten.

In Abwägung und Bewertung der möglichen indirekten Effekte, müssen die gesetzlichen Vorgaben entsprechend der Kennzeichnungsvorgaben des AI Acts für Deep Fakes und KI-basierte Texte selbstverständlich eingehalten werden. Darüber hinaus sollen jedoch keine überschießenden zusätzlichen Anforderungen oder Kennzeichnungen geschaffen oder vorgenommen werden, um im Besonderen Verwirrung bei Medienkonsumierenden und daraus folgende nachteilige Effekte zu vermeiden.

KI-Nutzung im Medienbereich, die nicht unter das verpflichtende Kennzeichnungsregime des AI Acts fällt sollte laut den Ergebnissen der Erhebung nicht zusätzlich offengelegt oder gekennzeichnet werden. Essenziell ist es jedoch, diese (interne) KI-Nutzung entsprechend den rechtlichen Anforderungen und den etablierten Standards der Medienarbeit zu gestalten. Um dies sicherzustellen, sind interne Richtlinien, Prozesse und Konzepte empfehlenswert, welche die konkret verwendeten KI-Tools sowie die konkreten Use Cases abbilden.

2.2. Ergänzende Back-End Kennzeichnung in Metadaten

Der AI Act verpflichtet künftig Anbieter von KI-Systemen, die synthetische Audio-, Bild-, Video- oder Textinhalte erzeugen, sicherzustellen, dass der erzeugte synthetische Content in einem maschinenlesbaren Format (*Watermarking*) als KI-generiert oder KI-bearbeitet kenntlich gemacht wird. Ein Großteil der Anbieter verfolgt bereits aktuell eine Kennzeichnung durch entsprechende Angaben in den Metadaten (Exif) als auch durch unsichtbare Wasserzeichen.

Für Medienunternehmen, die KI-Systeme von Drittanbietern nutzen, ergibt sich daraus, Anbieter auszuwählen, welche die Anforderungen des AI Acts erfüllen. Darüber hinaus ist es empfehlenswert, gemäß AI Act kennzeichnungspflichtige KI-generierte oder KI-bearbeitete Inhalte zusätzlich zum sichtbaren Ausweis für Medienkonsumierende auch in den Metadaten zu kennzeichnen.

2.3. Einheitliche Anwendung

Die Pflichtkennzeichnung gemäß des AI Acts soll konsistent über alle Medienhäuser, Plattformen, Formate und Medientypen hinweg angewendet werden. Einheitliche Standards und Formulierungen sind notwendig, um Konfusion zu vermeiden und sicherzustellen, dass das Publikum immer die gleichen Informationen erhält, unabhängig davon, wo und wie der Inhalt konsumiert wird.

2.4. Eindeutiger Kennzeichnungsbegriff

Die Kennzeichnung von KI-basierten Inhalten soll eindeutig und verständlich sein. Die Kennzeichnungspflicht gemäß des AI Acts gibt einen wertorientierten Einstiegspunkt vor, indem nur Deep Fakes aufgrund ihrer Irreführungseignung und künstlich erzeugte oder bearbeitete Texte zur Information der Öffentlichkeit über Angelegenheiten von öffentlichem Interesse verpflichtend gekennzeichnet werden sollten. Die Beurteilung, ob ein Fall der verpflichtenden Kennzeichnung vorliegt, ist durch das Medienunternehmen vorzunehmen. Laut dem Wortlaut des AI Acts ist gegenüber Medienkonsumierenden in Bezug auf Deep Fakes bzw. kennzeichnungspflichtige Texte offenzulegen, dass diese Inhalte künstlich erzeugt oder bearbeitet wurden. Der dafür verwendete Kennzeichnungsbegriff sollte klar vermitteln, dass der Inhalt ganz oder teilweise durch KI erstellt oder bearbeitet wurde. (zB mit KI generiert oder bearbeitet).

2.5. Iterativer Kennzeichnungs- und Nutzungsansatz

Kennzeichnungsmittel sowie interne KI-Nutzungsrichtlinien sollten kontinuierlich weiterentwickelt werden, um den Fortschritten der Technologie, den Einsatzfeldern im Medienbereich sowie den Erwartungen der Medienkonsumierenden gerecht zu werden. Regelmäßige Überprüfungen und Aktualisierungen sowie Feedback-Schleifen innerhalb von Medienunternehmen sind hierbei essenziell, um stets auf dem neuesten Stand zu bleiben und auf die sich verändernden Umfeldbedingungen reagieren zu können.

3. Zeitpunkt der Kennzeichnungspflicht

Natürliche Personen sind gemäß dem AI Act spätestens in jenem Zeitpunkt, zu dem sie den ersten Interaktionspunkt mit kennzeichnungspflichtigem Content haben, in klarer und eindeutiger Weise darüber zu informieren, wobei Informationen auch den geltenden Anforderungen an Barrierefreiheit entsprechen müssen.

Daraus ergibt sich, dass eine Kennzeichnung gut sichtbar und leicht zugänglich sein sollte, damit sie von Medienkonsumierenden problemlos wahrgenommen werden kann. Dies bedeutet, dass Kennzeichnungen an prominenten Stellen platziert werden sollten, wie etwa am Anfang eines Artikels, am Anfang eines Audioinhalts, als Bildunterschrift oder als eingblendeter Text in Videos. Unsichtbare oder schwer auffindbare Kennzeichnungen untergraben das Ziel der Transparenz.

4. Ansätze für spezifische Medienformen

Die Kennzeichnung von Deepfakes nach dem AI Act ist ein zentrales Element der Transparenz zur Darstellung der Nutzung von generativer KI in den Medien. Medienrezipient:innen sollen auf einen Blick erkennen können, ob es sich um ein Deepfake gemäß dem AI Act handelt.

Aus diesem Grund hat der Fachbereich Medien der KI-Servicestelle in einem breiten Prozess alle relevanten Stakeholder dazu eingeladen, ihre Vorstellungen für eine effektive Kennzeichnung einzumelden. Aus über 100 Rückmeldungen wurden in einem anschließenden Workshop gemeinsam mit Vertreter:innen der Medienbranche die folgenden Ansätze für eine einheitliche, gemeinsame und transparente Kennzeichnung erarbeitet.

4.1. Audio

Wenn KI zur Erstellung oder Bearbeitung von Audioinhalten wie Podcasts oder Nachrichten verwendet wird, wurde erarbeitet, dass dies in der Beschreibung und akustischen Einleitung des Inhalts offengelegt werden kann. Beispielsweise könnte eine Podcast-Episode den Hinweis enthalten: "Diese Episode enthält Segmente, die mit KI-Technologie erstellt wurden."

Zusätzlich können Informationen über die KI-Nutzung in die Metadaten der Audiodateien eingefügt werden, sodass diese Informationen über verschiedene Plattformen und Geräte zugänglich sind.

4.2. Video

Für KI-generierte oder KI-bearbeitete Videoinhalte ist die Form einer Einblendung in das Video, die auf die Nutzung von KI hinweist, festgehalten worden. Ein Beispiel könnte sein: "Dieses Video enthält Elemente, die mit KI-Unterstützung erstellt wurden." Darüber hinaus sollten in den Videobeschreibungen auf Plattformen wie YouTube Details zur Nutzung von KI im Produktionsprozess angegeben werden.

Zusätzlich können Informationen über die KI-Nutzung in die Metadaten der Videodatei eingefügt werden, sodass diese Informationen über verschiedene Plattformen und Geräte zugänglich sind.

4.3. Bild

Für KI-generierte Bilder oder Illustrationen ist das Ergebnis des Prozesses ein direkt unter dem Bild eingefügte Bildunterschrift, die die Nutzung von KI offenlegt. Ein Beispiel könnte sein: "Dieses Bild wurde mit KI erstellt."

Bis zur Entwicklung eines weiter verbreiteten Standards sollte vermieden werden, visuelle Labels oder Symbole für KI-generierte oder KI-bearbeitete Inhalte zu verwenden, um Missverständnisse über Glaubwürdigkeitsunterschiede zwischen KI-generierten und menschlich erstellten Inhalten zu verhindern.

Zusätzlich können Informationen über die KI-Nutzung in die Metadaten der Bilddatei eingefügt werden, sodass diese Informationen über verschiedene Plattformen und Geräte zugänglich sind.

4.4. Text

Für KI-generierte Artikel oder Berichte sollte eine Erklärung am Beginn des Textes oder als Fußnote hinzugefügt werden, die den Einsatz von KI angibt. Ein Beispiel könnte lauten: „Dieser Artikel wurde mit KI erstellt.“

Hierbei ist zu beachten, dass Inhalte, welche vor der Veröffentlichung menschlich überprüft wurden, gemäß AI Act keiner Kennzeichnung bedürfen.

5. Künstlerische, kreative, satirische, fiktionale und analoge Werke

Sind KI-generierte oder KI-bearbeitete Inhalte Bestandteil eines offensichtlich künstlerischen, kreativen, satirischen, fiktionalen oder analogen Werks oder Programms, so beschränken sich die Transparenzpflichten des AI Acts darauf, das Vorhandensein solcher erzeugten oder bearbeiteten Inhalte in geeigneter Weise offenzulegen, die die Darstellung oder den Genuss des Werks nicht beeinträchtigt.

Denkbar sind hier beispielsweise

- Eindeutige, aber nicht störende Kennzeichnung im Werk selbst oder dezente Kennzeichnung im Vor- oder Abspann beispielsweise bei Filmen, Serien oder Videoinhalten ("Einige Inhalte in diesem Werk wurden durch KI generiert oder bearbeitet.").
- Fußnoten oder Anmerkungen am Ende des Dokuments oder relevantes Kapitales in schriftlichen oder gedruckten Werken ("Einige Passagen dieses Werkes wurden mithilfe von KI erstellt oder bearbeitet.").
- In interaktiven digitalen Medien, wie zB Spielen oder Webseiten, könnten Pop-up-Hinweise oder diskrete Informationssymbole verwendet werden. Ein Symbol könnte angeklickt werden, um weitere Informationen zu erhalten ("Dieses Segment wurde mit KI-Unterstützung erstellt.)
- Hinweise in Begleitmaterialien, wie Programmhefte, Broschüren oder Online-Beschreibungen können Informationen enthalten, wie etwa einen kurzen Abschnitt, welche Teile des Werks KI-generiert sind.

Durch diese Maßnahmen wird die Transparenz für das Publikum gewährleistet, ohne die künstlerische Integrität und den Genuss des Werkes zu beeinträchtigen.

6. Ausblick

Einheitliche europäische und globale Übereinkommen bei der Kennzeichnung von KI-basierten Medien sind von entscheidender Bedeutung, um Transparenz und



Vertrauen in digitale Inhalte zu gewährleisten. Durch einheitliche Regelungen wird zudem der internationale Austausch von Medieninhalten vereinfacht, da Medienproduzierende und Medien-konsumierende sich auf gemeinsame Prinzipien stützen können. Die KI-Servicestelle lädt daher auch weiterhin alle Akteure in der Medien- und Technologiebranche sowie Entscheidungsträger:innen und die Zivilgesellschaft ein, sich aktiv an der Entwicklung dieser gemeinsamen Standards zu beteiligen, um eine transparente und vertrauenswürdige Medienlandschaft für die Zukunft zu schaffen.