

Leitfaden: Transparenzpflichten für KI-Erzeugnisse

Kennzeichnungspflichten für Anbieter und Offenlegungspflichten für Betreiber nach Art. 50 KI-Verordnung

Stand: September 2026

I. Vorwort	2
II. Adressaten	3
III. Offenlegungspflicht der Betreiber	4
1. Allgemeines	4
a. Deepfakes	4
b. Texte zu Angelegenheiten von öffentlichem Interesse	4
c. Ausnahmen	5
2. Form und Platzierung der Offenlegung	5
a. Visuelle Inhalte (Bilder, Videos) und Texte	5
b. Audio-Inhalte	6
c. Barrierefreiheit	6
3. Musteroffenlegung	6
a. Formulierungsbeispiele	6
b. EU-Icons	7
4. Anwendungsfälle	8
IV. Kennzeichnungspflicht der Anbieter	9
1. Anwendungsbereich	9
2. Technische Lösungen	9
1. Ebene: Digital signierte Metadaten	9
2. Ebene: Wasserzeichen	10
3. Ergänzende Ebene: Fingerprinting und Logging	10
4. Detektion	10
V. Handlungsempfehlungen für Nutzende an Hochschulen	11
VI. Abgrenzung zu weiteren Transparenzpflichten	12

I. Vorwort

Geltungsbeginn: Die Pflichten nach Art. 50 Abs. 2, 4 und 5 KI-Verordnung gelten grundsätzlich ab dem **2. August 2026**. Diese Frist hat sich durch die Änderung der KI-Verordnung (Digital Omnibus on AI)¹ nur bedingt geändert! Der neue Art. 111 Abs. 4 KI-VO wurde ergänzt. Dieser gewährt den Anbietern von KI-System, die vor dem 2. August 2026 in Verkehr gebracht wurden, eine verlängerte Umsetzungsfrist bis zum 2. Dezember 2026.

Dieser Leitfaden wurde auf Grundlage des Art. 50 der KI-Verordnung (KI-VO)², den zugehörigen Erwägungsgründen, einschlägigen Kommentierungen sowie den „Guidelines on Transparency Obligations under Article 50“ (Guidelines)³ und des „Code of Practice on Transparency of AI-Generated Content“ (CoP)⁴ der EU-Kommission erstellt. Die KI-Verordnung ist verbindlich. Die Guidelines haben keinen verbindlichen Charakter und weisen ausdrücklich darauf hin, dass die Auslegung der KI-VO letztendlich beim Europäischen Gerichtshof liegt.⁵ Sie zeigen eine aktuell gut vertretbare und von der Kommission angenommene Betrachtungsweise und können daher eine taugliche Orientierung bieten. Auch der CoP ist nur von freiwilliger Natur, kann durch die Unterzeichnung für die Unterzeichnenden jedoch bindend werden. Er wurde von unabhängigen Experten mithilfe des AI-Office erstellt. Er bildet somit einen ersten anerkannten praktischen Rahmen zur Umsetzung von Art. 50 KI-VO, jedoch keinen abschließenden Nachweis für die Rechtmäßigkeit der Handlungen.⁶ Nichtunterzeichner können auch andere geeignete Maßnahmen wählen, müssen aber darlegen können, dass diese die gesetzlichen Anforderungen in gleichwertig angemessener Weise erfüllen.

Nicht alle KI-generierten Inhalte sind zwingend kennzeichnungspflichtig. Der hauptsächliche Schutzgedanke liegt darin, dass KI-Erzeugnisse nicht immer auf den ersten Blick als künstlich erzeugt erkennbar sind, was zu einer Verbreitung von Fehlinformationen führen kann. Die Transparenzpflichten dienen hier vor allem der Vermeidung von Täuschung und Manipulation und der Integrität von Informationen.

Für Betreiber erfasst Art. 50 Abs. 4 KI-VO daher insbesondere Deepfakes sowie bestimmte veröffentlichte KI-generierte oder KI-manipulierte Texte zu Angelegenheiten von öffentlichem Interesse. Daneben bestehen eigenständige Pflichten für Anbieter, insbesondere zur maschinenlesbaren Kennzeichnung bestimmter synthetischer Erzeugnisse nach Art. 50 Abs. 2 KI-VO.

Kurz gesagt: Anbieter markieren KI-Erzeugnisse technisch. Betreiber kennzeichnen bestimmte veröffentlichte Inhalte für Menschen sichtbar oder hörbar. Das sind zwei getrennte Pflichten.

Hinweis: Diese Handreichung dient der allgemeinen Orientierung und Aufklärung über die Transparenzpflichten, die an Hochschulen einzuhalten sind. Es ist wichtig zu beachten, dass es sich um eine reduzierte Darstellung der oben genannten Quellen handelt, um diese leichter

¹ [Digital-Omnibus-Verordnung zur KI \(Verordnung \(EU\) 2026/1744 vom 08. Juli 2026\)](#).

² [Verordnung \(EU\) 2024/1689 \(KI-Verordnung\)](#).

³ [EU-Kommission: Guidelines on transparency obligations under Article 50, Fassung vom 20. Juli 2026](#).

⁴ [EU-Kommission: Code of Practice on Transparency of AI-generated Content, Fassung vom 10. Juni 2026](#).

⁵ Guidelines, Rn. 5.

⁶ Vgl. Guidelines Rn. 146-149.

handhabbar zu machen. Im Zweifelsfall sollte die Tendenz stets zugunsten der Transparenzpflichten ausfallen, um einen möglichst sicheren Weg einzuschlagen.

II. Adressaten

Die Pflichten aus Art. 50 KI-VO lassen sich in Pflichten für „Anbieter“ (engl. Provider) und Pflichten für „Betreiber“ (engl. Deployer) unterteilen.

Anbieter trifft nach Art. 50 Abs. 2 KI-VO insbesondere die Pflicht, bestimmte synthetische Audio-, Bild-, Video- und Textausgaben maschinenlesbar zu kennzeichnen und ihre Erkennbarkeit als künstlich erzeugt oder manipuliert sicherzustellen (Kennzeichnungspflicht). Die Betreiber müssen gemäß Art. 50 Abs. 4 KI-VO für KI-Erzeugnisse, die Deepfakes oder Texte zu Angelegenheiten von öffentlichem Interesse sind, die KI-Nutzung offenlegen (Offenlegungspflicht). Insgesamt spricht die KI-Verordnung für die Pflichten aus Art. 50 KI-VO von Transparenzpflichten.⁷

Anbieter im Sinne der KI-VO ist *„eine natürliche oder juristische Person, Behörde, Einrichtung oder sonstige Stelle, die ein KI-System oder ein KI-Modell mit allgemeinem Verwendungszweck entwickelt oder entwickeln lässt und es unter ihrem eigenen Namen oder ihrer Handelsmarke in Verkehr bringt oder das KI-System unter ihrem eigenen Namen oder ihrer Handelsmarke in Betrieb nimmt, sei es entgeltlich oder unentgeltlich“* (Art. 3 Nr. 3 KI-VO).

Betreiber im Sinne der KI-VO ist *„eine natürliche oder juristische Person, Behörde, Einrichtung oder sonstige Stelle, die ein KI-System in eigener Verantwortung verwendet, es sei denn, das KI-System wird im Rahmen einer persönlichen und nicht beruflichen Tätigkeit verwendet“* (Art. 3 Nr. 4 KI-VO).

Eine Hochschule ist Betreiberin, wenn sie ein KI-System in eigener Verantwortung einsetzt. Entscheidend sind daher Verantwortung, Weisungs- und Kontrollzusammenhang beim Einsatz des KI-Systems; der Beschaffungsvorgang allein ist nicht ausschlaggebend, aber oftmals indikativ. Wird den Beschäftigten ein bestimmtes KI-System zur Verfügung gestellt, so kann regelmäßig von einer Betreibereigenschaft der Hochschule ausgegangen werden. Anbieterin kann eine Hochschule insbesondere dann sein, wenn sie ein KI-System selbst entwickelt oder entwickeln lässt und es unter eigenem Namen oder eigener Marke in Verkehr bringt oder in Betrieb nimmt.

Nutzen Beschäftigte ein KI-System unter Weisung, Kontrolle und Verantwortung der Hochschule, sind sie nach den Guidelines grundsätzlich nicht als eigenständige Betreiber anzusehen.⁸ Betreiberin bleibt die Hochschule (juristische Person). Bei eigenständiger beruflicher Nutzung außerhalb eines solchen Verantwortungs- und Kontrollzusammenhangs kann die Rollenverteilung anders zu beurteilen sein. Für die interne Praxis sollte deshalb festgelegt werden, welche KI-Systeme in wessen Verantwortung eingesetzt werden und wer die erforderlichen Transparenzmaßnahmen umsetzt.⁹

⁷ Zur Abgrenzung zu weiteren Transparenzpflichten siehe VI.

⁸ Guidelines Rn. 14.

⁹ Eine Festlegung interner Verantwortlichkeiten ist abhängig von vielen hochschulspezifischen Faktoren. Diese tiefergehenden Governance- und Compliance-Fragen können in dieser Handreichung nicht

III. Offenlegungspflicht der Betreiber

1. Allgemeines

Betreiber treffen insbesondere die Offenlegungspflichten nach Art. 50 Abs. 3 und 4 KI-VO. Art. 50 Abs. 3 KI-VO betrifft die Betreiber von Emotionserkennungssystemen oder Systemen zur biometrischen Kategorisierung.¹⁰ Diese spielen für die Hochschule eine nachrangige Rolle, sodass für diesen Leitfaden Art. 50 Abs. 4 im Mittelpunkt steht. Dieser regelt die für natürliche Personen wahrnehmbare Offenlegung von Deepfakes und bestimmten veröffentlichten Texten zu Angelegenheiten von öffentlichem Interesse. Es lassen sich zwei Fallgruppen sowie Ausnahmefälle bestimmen.

a. Deepfakes

Unter Deepfakes versteht die KI-VO „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 Nr. 60 KI-VO). Es geht also zum einen um vollständig neu erzeugte Inhalte als auch um die Verfälschung ursprünglich authentischer Inhalte.¹¹ Eine vollständig synthetische Darstellung kann daher ebenso ein Deepfake sein wie die KI-gestützte Veränderung eines authentischen Ausgangsinhalts. Umgekehrt ist nicht jede synthetische Stimme, jedes KI-Bild oder jedes fiktive Ereignis automatisch ein Deepfake; es kommt auch auf den Nutzungskontext, Aussagegehalt und die Erwartungen des vorgesehenen Publikums.

b. Texte zu Angelegenheiten von öffentlichem Interesse

Aus der KI-Verordnung ergibt sich nicht, wann ein Text als veröffentlicht anzusehen ist. Texte, die ausschließlich einen zahlenmäßig überschaubaren, dem Betreiber bekannten Empfängerkreis erreichen können und sollen sind von der Pflicht aber ausgeschlossen.¹² Hilfsweise kann der Veröffentlichungsbegriff des Urheberrechts herangezogen werden. Hiernach ist die Öffentlichkeit betroffen, wenn der Text bei Veröffentlichung *einem unbestimmten Personenkreis zugänglich gemacht* wird.¹³ Auch die Guidelines gehen davon aus, dass der Text einer unbestimmten, hinreichend große Zahl nicht miteinander verbundener potenzieller Empfänger, gleichzeitig oder nacheinander zugänglich gemacht werden muss.¹⁴

Angelegenheiten von öffentlichem Interesse sollen jedenfalls politische, gesellschaftliche, wirtschaftliche und kulturelle Angelegenheiten betreffen.¹⁵ Beispiele sind Zeitungsartikel/ Social Media-/ Blog-/ Forenbeiträge mit Berichten zu Wahlen, Kommentierungen von Alltagsgeschehen, Börsenberichte und ähnliches.

Schwierig ist die Bewertung von Infografiken, Diagrammen, etc., die keinen Text beinhalten, also eigentlich ein Bild darstellen (z.B. Klimagrafiken). Diese dienen dazu die Öffentlichkeit über eine Angelegenheit von öffentlichem Interesse zu informieren, sind aber streng genommen kein Text.

weitergehend behandelt werden. Oftmals wird aber die Regelung im Rahmen einer Dienstvereinbarung oder Hochschulordnung (z.B. in Form einer KI-Policy) sinnvoll sein.

¹⁰ Für solche KI-Systeme ist ggf. zusätzlich das Verbot aus Art. 5 KI-VO zu berücksichtigen.

¹¹ Vgl. *Martini* in Martini/Wendehorst, Art. 50, Rn. 101.

¹² Vgl. *Martini* in Martini/Wendehorst, Art. 50, Rn. 113.

¹³ *Merkle* in Bomhard/Pieper/Wende KI-VO, Art. 50 Rn. 159.

¹⁴ Guidelines, Rn. 133.

¹⁵ Vgl. *Martini* in Martini/Wendehorst, Art. 50, Rn. 113.

Sie sind vermutlich aber auch kein Deepfake, da die Abbildung keinen wirklichen Personen, Gegenständen, Orten, Einrichtungen oder Ereignissen ähnelt. Die Abbildung ist in der Regel als erstellte Grafik erkennbar und es besteht keine Verwechslungsgefahr. Der Normzweck des Schutzes der öffentlichen Meinungsbildung gebietet ggf. aber dennoch eine Kennzeichnung, wenn eine Täuschung bzw. eine Fehlinformation durch einen KI-generierten Inhalt naheliegt. Eine eindeutige Rechtspflicht ist derzeit aus dem Wortlaut jedoch nicht herleitbar. Zum Zwecke der Rechtssicherheit sollte daher auch hier eine -zumindest freiwillige- Offenlegung erfolgen.

c. Ausnahmen

Art. 50 Abs. 4 S. 5 2. HS KI-VO¹⁶ nimmt veröffentlichte Texte aus der Offenlegungspflicht aus, wenn eine substantielle menschliche Prüfung oder redaktionelle Kontrolle stattgefunden hat und eine natürliche oder juristische Person die redaktionelle Verantwortung für die Veröffentlichung trägt. Kontrolliert also eine Person den Text und macht sich diesen nach umfassender Prüfung zu eigen, so ist die oben beschriebene Gefahr ungeprüfter Fehlinformationen eingedämmt, sodass keine Offenlegung erforderlich ist.¹⁷

Art. 50 Abs. 4 S. 3 KI-VO kennt zusätzlich eine Teilausnahme für Deepfakes, die Teil eines „*offensichtlich künstlerischen, kreativen, satirischen, fiktionalen oder analogen Werks oder Programms*“ sind. Hier hat die Offenlegung auf eine angemessene Art und Weise zu erfolgen, die die Darstellung oder den Genuss des Werks nicht beeinträchtigt.

2. Form und Platzierung der Offenlegung

Grundsatz: Die Kennzeichnung muss nach Art. 50 Abs. 5 KI-VO *klar und unterscheidbar, spätestens bei der ersten Wahrnehmung des Inhalts und unter Beachtung der anwendbaren Anforderungen an die Barrierefreiheit* erfolgen.

a. Visuelle Inhalte (Bilder, Videos) und Texte

- **Klar und unterscheidbar:** sofort verständlich, gut lesbar, ausreichender Kontrast; nicht im Kleingedruckten/AGB/Metadaten „verstecken“.
- **Spätestens bei der ersten Wahrnehmung:** Die Person muss den Hinweis sehen, bevor oder sobald sie den Inhalt wahrnimmt.
- **Ohne zusätzliche Handlung:** Kein erforderlicher Klick, kein Hover und kein Scrollen, um die Grundinformation an den Wahrnehmenden zu übermitteln.
- **Direkt im Inhalt oder gleichwertig als stabiler Overlay-Hinweis.** Beim Weiterteilen oder Herunterladen sollte die Kennzeichnung möglichst erhalten bleiben.
- Das EU-Icon (s.u.) ist optional. Ein gleichwertiges Label ist zulässig, wenn es die Design- und Platzierungsanforderungen erfüllt. Beim CoP steht „AI“ als Hauptbestandteil; ergänzend sind „vollständig KI-generiert“ (AI generated) oder „teilweise KI-modifiziert“ (AI modified) sinnvoll.

¹⁶ Eine weitere Ausnahme besteht in Art. 50 Abs. 4 S. 5 1. HS KI-VO, soweit die Nutzung gesetzlich zur Aufdeckung, Verhütung, Ermittlung oder Verfolgung von Straftaten zugelassen ist. Diese Konstellation dürfte im normalen Hochschulbetrieb keine Rolle spielen.

¹⁷ Vgl. *Martini* in *Martini/Wendehorst*, Art. 50 Rn. 114.

b. Audio-Inhalte

- Bei Audio muss die Offenlegung spätestens bei der ersten Wahrnehmung klar hörbar sein; ein kurzer, verständlicher Hinweis in der Sprache des Inhalts oder auf Englisch ist regelmäßig geeignet.
- Bei langen, live übertragenen oder leicht in Ausschnitte zerlegbaren Inhalten sollte sichergestellt werden, dass auch Personen, die die Wiedergabe zu einem späteren Zeitpunkt beginnen oder nur einen Ausschnitt wahrnehmen, die Information angemessen erhalten. Art und Häufigkeit des Hinweises richten sich nach Inhalt und Nutzungskontext. (Beispiel: Ansage zu Beginn jedes Kapitels bei einem Podcast)
- Ist zugleich eine visuelle Benutzeroberfläche vorhanden, kann ein zusätzlicher visueller Hinweis die Wahrnehmbarkeit und Barrierefreiheit verbessern. (Beispiel: Nutzeroberfläche, über die der Play-Button bedient wird)

c. Barrierefreiheit¹⁸

- Visuelle Hinweise sollten ausreichend groß und kontrastreich sein. Soweit möglich, sollten Icons oder begleitende Hinweise durch Alternativtext bzw. ARIA-Kennzeichnungen für assistive Technologien erfassbar sein.
- Zeitlich begrenzte Hinweise müssen lange genug wahrnehmbar sein, damit auch Personen mit kognitiven oder Verarbeitungsschwierigkeiten sie erfassen können.

3. Musteroffenlegung

Es gibt kein allgemeingültiges Muster, das zwingend einzuhalten ist. Wichtig ist nur, dass die Voraussetzungen des Art. 50 Abs. 5 KI-VO eingehalten werden (s.o.).¹⁹

a. Formulierungsbeispiele

- Bild/Video: „Dieses Bild/Video wurde (vollständig) mit KI erzeugt.“ / „This image is (fully) AI generated.“
- Teilmanipulation: „Dieses Bild wurde mit KI verändert.“ / „This image is AI modified.“ Wenn möglich sollte die konkrete Manipulation benannt werden.
- Audio: „Hinweis: Die folgende Stimme wurde künstlich mit KI erzeugt.“ / „Attention: The following voices are artificially generated by AI.“
- Text: „Dieser Beitrag wurde mit KI erzeugt.“ / „This text was created using AI.“

Hinweis: Eine Offenlegung in AGB, Metadaten, einem allgemeinen Website-Hinweis oder hinter einem Info-Button/Link ist nicht ausreichend, wenn beim konkreten kennzeichnungspflichtigen Inhalt die klare Kennzeichnung bei der ersten Wahrnehmung fehlt.

¹⁸ Die jeweils anwendbaren Barrierefreiheitsvorschriften sind zu beachten, insbesondere das Barrierefreiheitsstärkungsgesetz (BFSG) und das Behindertengleichstellungsgesetz NRW (BGG NRW).

¹⁹ Eine Hochschule könnte sich aber natürlich dazu entscheiden, den Mitarbeitenden eine einheitliche Offenlegungsweise vorzugeben.

b. EU-Icons

Der CoP und die Kommission stellen drei EU-Icon-Typen für die Betreiberoffenlegung bereit: ein Basis-AI-Icon „AI“, „AI-Generated“ und „AI-Modified“. Die Icons sind optional und sollen die klare, unterscheidbare Offenlegung unterstützen.²⁰ Die EU-Icons dürfen von allen frei und ohne Quellenangabe verwendet werden, also auch von Nichtunterzeichnern des CoP. Bei Nichtunterzeichnern darf ihre Verwendung lediglich nicht als Erklärung eines Beitritts zum CoP verstanden werden. Unterzeichner müssen die im CoP festgelegten Darstellungs- und Platzierungsvorgaben beachten

Icon	Kennzeichnungszweck
	Allgemeiner Hinweis darauf, dass KI an der Erzeugung oder Veränderung eines offlegungspflichtigen Deepfakes oder veröffentlichten Textes beteiligt war. Besonders geeignet in Verbindung mit einem erläuternden Text oder einer interaktiven zweiten Informationsebene.
	Offenlegung eines vollständig KI-generierten Deepfakes oder eines vollständig KI-generierten Textes zu einer Angelegenheit von öffentlichem Interesse. Abgesehen von der Eingabe der Prompts enthält der Inhalt keine menschlich erstellten Bestandteile und unterlag keiner menschlichen redaktionellen Kontrolle.
	Offenlegung eines ursprünglich menschlich erstellten Inhalts, der durch KI teilweise so verändert wurde, dass ein Deepfake oder ein kennzeichnungspflichtiger Text zu einer Angelegenheit von öffentlichem Interesse entstanden ist.

Wichtig: Die Beispielformulierungen oder EU-Icons allein genügen den Transparenzpflichten nur dann, wenn auch die Platzierung, der Zeitpunkt, die Wahrnehmbarkeit und die Barrierefreiheit eingehalten werden, Art. 50 Abs. 5 KI-VO (s.o.).

²⁰ Weitere Formen der Icons: [EU-Kommission: EU Icons for labelling AI-generated content](#).

4. Anwendungsfälle

KI-Erzeugnis	Kennzeichnung	Art der Kennzeichnung
Realistisch wirkendes KI-Porträt einer echten Person	Ja, Deepfake	Sichtbares AI-Icon oder gleichwertige visuelle Kennzeichnung spätestens bei erster Wahrnehmung.
Menschlich klingende KI-Stimme (z.B. in Podcast oder Telefonansage)	Ja/Nein, möglicher Audio-Deepfake, Entscheidend ist, ob die Stimme einem realen oder realistisch vorstellbaren menschlichen Subjekt merklich ähnelt und im konkreten, vorhersehbaren Nutzungskontext fälschlich als authentische menschliche Aufnahme erscheinen kann. Umgekehrt muss nicht zwingend die Stimme einer bestimmten identifizierbaren Person imitiert werden.	Kurzer verständlicher Audiohinweis am Anfang; bei langen Inhalten angemessene Wiederholungen. Ggfls. zusätzlich visuell.
KI-Video eines nie geschehenen real wirkenden Ereignisses z.B. Medienbericht/Dokumentation	Ja, Deepfake	Label am Anfang, möglichst regelmäßig und mindestens nach Unterbrechungen; so platzieren, dass es nicht verdeckt wird.
KI-generierter Bericht zu z.B. Wahl, Gesundheit, Klima oder Wirtschaft	Ja, Text zu Angelegenheit von öffentlichem Interesse <u>Ausnahme:</u> Findet eine substanzielle menschliche Prüfung (redaktionelle Kontrolle) statt und übernimmt eine juristische oder natürliche Person erkennbar die redaktionelle Verantwortung, so kann gem. Art. 50 Abs. 4 KI-VO auf die Kennzeichnung verzichtet werden.	Label z. B. über dem Text, nahe der Überschrift oder zu Beginn im Impressums-/Kolophonbereich.
Offensichtlich fiktionale Darstellung mit Deepfake-Element	Ja, aber erleichtert	Angemessener Hinweis z. B. in Begleittext, Vorspann/Abspann oder direkt angrenzend, ohne den Werkgenuss zu stören.
KI-Hintergrundbild ohne Anschein eines realen Vorbilds	Regelmäßig nein (Art. 50 Abs. 4 KI-VO)	
Rechtschreib-, Format- oder Standardbearbeitung ohne wesentliche Inhaltsänderung	Regelmäßig kein relevantes Erzeugnis/ keine Manipulation	Einzelfall prüfen, wenn Bedeutung oder Authentizität verändert wird.

IV. Kennzeichnungspflicht der Anbieter

1. Anwendungsbereich

Kennzeichnungspflichtig: Art. 50 Abs. 2 KI-VO verlangt von Anbietern von KI-Systeme, die *synthetische Audio-, Bild-, Video- oder Textinhalte erzeugen*, dass die Ausgaben in einem maschinenlesbaren Format markiert und als künstlich erzeugt oder manipuliert erkennbar sind.²¹ Ebenso erfasst sind KI-Systeme mit allgemeinem Verwendungszweck, die entsprechende synthetische Inhalte erzeugen. Als synthetisch erstellt gelten sowohl vollständig durch KI-Systeme erzeugte Inhalte als auch bestehende Inhalte, die durch KI manipuliert bzw. verändert wurden. Unklar ist, wie weit der Begriff der Manipulation reicht. Grenzfälle sind beispielsweise KI-basierte Bildfilter, das Entfernen von Hintergrundgeräuschen („Noise Reduction“) oder das KI-gestützte Hochskalieren von Videos („Upscaling“).²² Im Zweifelsfall sollte die Wahl stets zugunsten der Kennzeichnung gehen.

Ausnahmen: Nicht kennzeichnungspflichtig sind KI-Systeme, die lediglich eine unterstützende Funktion für die Standardbearbeitung darstellen. Werden die Eingabedaten oder deren Semantik nicht wesentlich verändert, so muss ebenfalls keine Kennzeichnung stattfinden. Hierunter zählen zum Beispiel automatische Rechtschreib- oder Grammatikkorrekturen, automatische Formatierungen oder Bildbearbeitungen.

2. Technische Lösungen

Die technischen Lösungen sollen die Kennzeichnung fest mit dem synthetischen Inhalt verbinden. Die Anbieter müssen sicherstellen, dass die Lösungen *wirksam, interoperabel, belastbar und zuverlässig* sind. Hierbei sind der Inhaltstyp, die technischen Möglichkeiten/Grenzen, Umsetzungskosten und der allgemein anerkannte Stand der Technik zu berücksichtigen. Wie diese Vorgaben konkret zu erfüllen sind lässt der Gesetzgeber offen.²³

Section 1 des CoP befasst sich mit der Kennzeichnung durch den Anbieter. Er geht davon aus, dass es aktuell keine einzelne Kennzeichnungstechnik gibt, welche die Voraussetzungen *wirksam, interoperabel, belastbar und zuverlässig* erfüllt.²⁴ Dies mag sich zwar in der Zukunft ändern, derzeit wird aber deshalb von einem mehrschichtigen Ansatz ausgegangen.

1. Ebene: Digital signierte Metadaten

In geeigneten Dateiformaten sollen maschinenlesbare Metadaten hinterlegt werden, aus denen die künstliche Erzeugung oder Manipulation hervorgeht.

Die Provenienzinformationen (engl. „provenance information“) sollen möglichst:

- digital signiert,
- manipulationssicher bzw. manipulationsnachweisbar und
- gegebenenfalls mit Zeitinformationen versehen

²¹ Wenn eine Anbieterstellung sowie ein betroffenes KI-System vorliegen, dann sollte der Code of Practice für das entsprechende KI-System sorgfältig angewandt werden.

²² Merkle in Bomhard/Pieper/Wende KI-VO, Art. 50 Rn. 66f.

²³ Martini in Martini/Wendehorst, Art. 50 Rn. 77.

²⁴ CoP S. 5, Recital b).

sein. Die Herkunft und Bearbeitungshistorie eines Inhalts sollen technisch überprüfbar sein.

2. Ebene: Wasserzeichen

Zusätzlich können unsichtbare bzw. nicht wahrnehmbare Wasserzeichen unmittelbar in den Inhalt eingebettet werden. Das Wasserzeichen soll möglichst robust²⁵ gegenüber typischen Weiterverarbeitungsschritten sein, also auch nach Dateiänderungen fortbestehen, z.B. durch:

- Kompression,
- Größenänderung,
- Formatkonvertierung oder
- vergleichbaren technischen Veränderungen.

Vorteil gegenüber Metadaten: Metadaten können beim Kopieren oder Konvertieren einer Datei entfernt werden. Ein in den Inhalt eingebettetes Wasserzeichen kann demgegenüber auch nach solchen Verarbeitungsschritten erhalten bleiben.

3. Ergänzende Ebene: Fingerprinting und Logging

Der CoP berücksichtigt außerdem Fingerprinting- und Logging-Verfahren. Dabei kann beispielsweise ein charakteristischer digitaler Fingerabdruck eines Outputs registriert werden. Ein später vorgelegter Inhalt kann dann mit den registrierten Informationen abgeglichen werden. Diese Verfahren kommen insbesondere als Ergänzung in Betracht, wenn Metadaten oder Wasserzeichen für den jeweiligen Inhalt technisch nur eingeschränkt geeignet sind.

4. Detektion

Die technische Kennzeichnung soll nicht nur vorhanden, sondern auch erkennbar und überprüfbar sein. Der CoP verbindet deshalb „**Marking**“ und „**Detection**“: Der Anbieter soll also nicht lediglich eine technische Kennzeichnung in das KI-Erzeugnis integrieren. Es muss auch eine Möglichkeit bestehen, dieses Signal zu erkennen bzw. die KI-Herkunft des Inhalts technisch zu verifizieren.²⁶

Dabei spielt Interoperabilität eine zentrale Rolle. Die Erkennung soll möglichst nicht davon abhängen, dass für jeden KI-Anbieter ein vollständig separates und proprietäres Prüfverfahren verwendet werden muss. **Kennzeichnung und Detektion müssen also zusammenpassen.** Ein eingebettetes Wasserzeichen oder Herkunftsinformationen nützen nur, wenn sie mit einem geeigneten Verfahren zuverlässig erkannt werden können.

Wichtig: Eine sichtbare Offenlegung durch den Betreiber ersetzt die technische, maschinenlesbare Kennzeichnung des Anbieters nicht. Ebenso wenig ersetzt die Anbieterkennzeichnung die Offenlegung durch den Betreiber.

²⁵ Das Merkmal „belastbar“ aus Art. 50 Abs. 2 KI-VO heißt in der englischen Fassung „robust“, weswegen dieser Begriff z.T. auch in deutschen Quellen zu finden ist.

²⁶ Vgl. Guidelines, Rn. 69–78; CoP, Section 1, Commitment 2.

V. Handlungsempfehlungen für Nutzende an Hochschulen

1. Rolle und Anwendungsfall feststellen: Wird als Betreiber gehandelt? Werden Deepfakes erzeugt oder manipuliert oder sollen Texte zu öffentlichem Interesse veröffentlicht werden?

a. Wird ein Bild, eine Audio-Datei oder ein Video durch ein KI-System erzeugt oder manipuliert und wirkt es wie eine echte Person, ein echtes Objekt, einen echten Ort, eine echte Einrichtung oder ein echtes Ereignis?

- **Ja:** Es liegt grundsätzlich ein Deepfake nahe. Offenlegung ist erforderlich, soweit keine gesetzliche Ausnahme greift. Weiter mit c.
- **Nein:** Art. 50 Abs. 4 greift grundsätzlich nicht. Weiter mit b.

Wichtig: Eine neben der Betreiberpflicht bestehende technische Anbieterpflicht nach Art. 50 Abs. 2 ist separat anhand ihrer eigenen Voraussetzungen zu prüfen (siehe **IV.**).

b. Wird ein KI-generierter oder KI-manipulierter Text veröffentlicht, um die Öffentlichkeit über eine Angelegenheit von öffentlichem Interesse zu informieren?

- **Ja:** Offenlegung ist erforderlich, soweit keine gesetzliche Ausnahme greift. Weiter mit c.
- **Nein:** keine Offenlegung erforderlich. (z.B. Werbetexte, interne Notizen, private Nachrichten oder rein persönliche Inhalte)

c. Greift eine Ausnahme?

- Handelt es sich bei dem erzeugten oder manipulierten Inhalt offensichtlich um ein künstlerisches, kreatives, satirisches, fiktionales oder analoges Werk?
 - **Ja:** Keine Vollaussnahme. Offenlegung bleibt notwendig, darf aber so gestaltet werden, dass Werkgenuss und normale Nutzung nicht beeinträchtigt werden.
 - **Nein:** Vollständige Offenlegung, die den Anforderungen von Art. 50 Abs. 5 KI-VO genügt, d.h. klar und unterscheidbar, spätestens bei der ersten Wahrnehmung des Inhalts und unter Beachtung der anwendbaren Anforderungen an die Barrierefreiheit (siehe **III.2.**).
- Wurde der künstlich erzeugte oder manipulierte Text von öffentlichem Interesse menschlich überprüft oder einer redaktionellen Kontrolle unterzogen und übernimmt eine natürliche oder juristische Person die redaktionelle Verantwortung für die Veröffentlichung?
 - **Ja:** Keine Offenlegungspflicht nach Art. 50 Abs. 4 KI-VO
 - **Nein:** Vollständige Offenlegung, die den Anforderungen von Art. 50 Abs. 5 KI-VO genügt, d.h. klar und unterscheidbar, spätestens bei der ersten Wahrnehmung des Inhalts und unter Beachtung der anwendbaren Anforderungen an die Barrierefreiheit (siehe **III.2.**).

2. Ggf. Ausnahme dokumentieren: Bei Texten menschliche Prüfung, redaktionelle Kontrolle sollte die redaktionelle Verantwortung geklärt und für die Empfänger ersichtlich gemacht werden. Der Text sollte hierzu den Ersteller (also die redaktionell verantwortliche Person) erkennen lassen (z.B. Rektorat, Organisationseinheit der Hochschule, Lehrende oder andere natürliche Person).

3. Inhalt klassifizieren: Vollständig KI-generiert oder teilweise KI-modifiziert?

4. Offenlegung anwenden: Offenlegung entsprechend der inhaltlichen Klassifizierung (3.) gem. Art. 50 Abs. 5 KI-VO wählen und anwenden, d.h. klar und unterscheidbar, spätestens bei der ersten Wahrnehmung des Inhalts und unter Beachtung der anwendbaren Anforderungen an die Barrierefreiheit (siehe III.2.). Ggfls. Beispielformulierungen und EU-Icons verwenden.

5. Weitergabe absichern: Markierung und Kennzeichnung bei Export, Upload, Download und Weiterverwendung möglichst erhalten.

VI. Abgrenzung zu weiteren Transparenzpflichten

Diese Handreichung konzentriert sich auf die Kennzeichnungs- und Offenlegungspflichten nach Art. 50 Abs. 2 und Abs. 4 KI-VO. Art. 50 KI-VO enthält daneben weitere Transparenzpflichten: Anbieter interaktiver KI-Systeme (z.B. ChatBots) müssen natürliche Personen grundsätzlich darüber informieren, dass sie mit einem KI-System interagieren, sofern dies nicht offensichtlich ist. Auch diese Hinweise müssen klar, unterscheidbar, barrierefrei und spätestens bei der ersten Interaktion oder Exposition erfolgen.

Auch das Datenschutzrecht kennt Transparenzpflichten. Art. 13 DSGVO schreibt vor, dass betroffene Personen zum Zeitpunkt der Datenerhebung mindestens über den Verantwortlichen und den Zweck der Datenverarbeitung informiert werden müssen. Im Kontext mit KI-Systemen kann eine Verarbeitung personenbezogener Daten –etwa durch Prompting– durchaus möglich sein, z.B. bei der Eingabe von Personal- oder Studierendendaten, Auswertungen von Fragebögen, Erhebung und Auswertung von Lernstatistiken. Hier sind viele Konstellationen denkbar. Insgesamt sollten die datenschutzrechtlichen Anforderungen immer geprüft werden, wenn personenbezogene Daten –also Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen (Art. 4 Nr. 1 DSGVO) – verarbeitet werden sollen. Die Einhaltung der Transparenzpflichten nach Art. 50 KI-VO ersetzt weder eine datenschutzrechtliche Rechtsgrundlage noch die erforderlichen Datenschutzinformationen.