Community of Practice KI*PerWeb*

Austausch zur Nutzung und Entwicklung KI-gestützter Webanwendungen

KIPERWEB

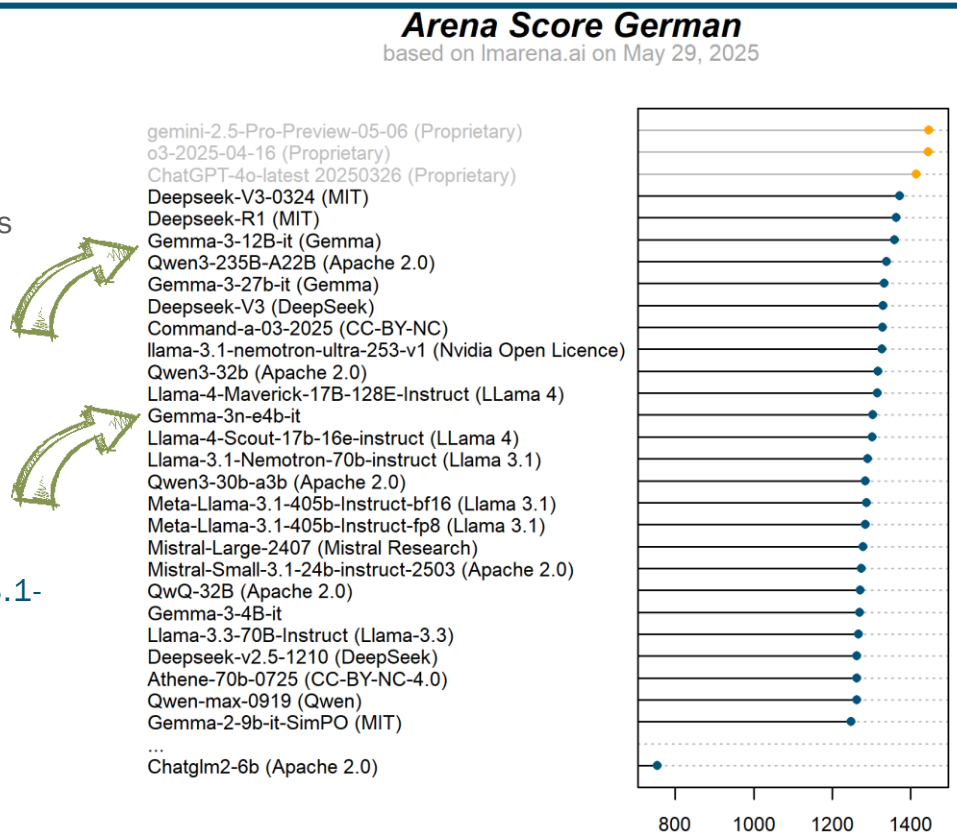
f-bb Forschungsinstitut Betriebliche Bildung

Agenda

- Update
  - News & Leaderboard-Update
- Input
  - „ Transparenzpflichten zu KI nach Art. 50 KI-VO“
- Diskussion

Leaderboard-Update (28.05.2025)

- Mit Blick auf die Performanz in der Kategorie „German“ liegt nach wie vor Gemini-2.5-Pro-Preview-05-06 ganz vorne auf dem Leaderboard.
- Arena-Scores von *nicht*-proprietären Modelle sind rechts ausgewiesen sofern sie mindestens das Niveau von Gemma-2-9b-it-SimPo erreichen:
  - Gemma-3-12B-it bleibt das beste Modell für den Hausgebrauch (noch vor Gemma-3-27B-it)
  - Gemma-3n-e4b-it liegt vor Gemma-3-4B-it bei geringerem Ressourcenverbrauch (https://ai.google.dev/gemma/docs/gemma-3n) und noch *vor* Schwergewichten wie Meta-Llama-3.1-405b-Instruct!
- Schlusslicht auf dem Leaderboard bleibt Chatglm2-6b



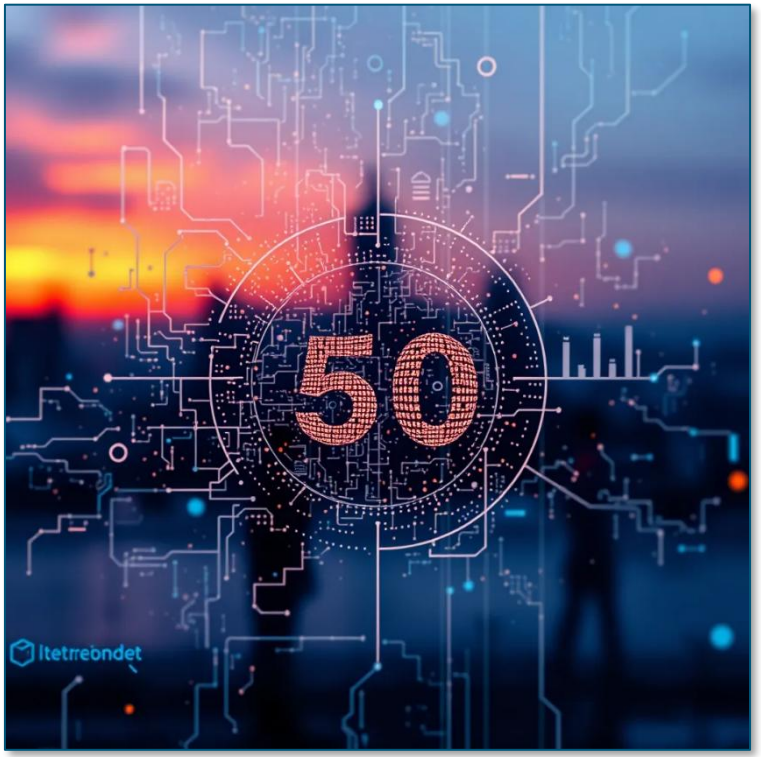
Datengrundlage: https://lmarena.ai/leaderboard/text/german

Fokusthema: Transparenzpflichten

- Prompt:
  „Transparenzpflichten zu KI nach Art. 50 der KI-Verordnung“

  (rechts visualisiert von FLUX.1-schnell, seed 480535002)

...



Quelle: https://huggingface.co/spaces/black-forest-labs/FLUX.1-schnell

Inkrafttreten und Geltungsbeginn

Artikel 113

Inkrafttreten und Geltungsbeginn

Diese Verordnung tritt am zwanzigsten Tag nach ihrer Veröffentlichung im *Amtsblatt der Europäischen Union* in Kraft.

Sie gilt ab dem 2. August 2026.

Jedoch:

a) Die Kapitel I und II gelten ab dem 2. Februar 2025;

b) Kapitel III Abschnitt 4, Kapitel V, Kapitel VII und Kapitel XII sowie Artikel 78 gelten ab dem 2. August 2025, mit Ausnahme des Artikels 101;

c) Artikel 6 Absatz 1 und die entsprechenden Pflichten gemäß dieser Verordnung gelten ab dem 2. August 2027.

Diese Verordnung ist in allen ihren Teilen verbindlich und gilt unmittelbar in jedem Mitgliedstaat.

Geschehen zu Brüssel am 13. Juni 2024.

Im Namen des Europäischen Parlaments	*Im Namen des Rates*
Die Präsidentin	*Der Präsident*
R. METSOLA	M. MICHEL

Kap. I&II umfassen Artikel 1-5, fordern insb. KI-Kompetenz und verbieten gewisse Praktiken und gelten bereits. Kapitel IV mit Art.50 gilt ab dem 02.08.2026, aber Berücksichtigung von Rechten und Pflichten ist Teil des Begriffs KI-Kompetenz *(vgl. Art. 3, Begriffbestimmung 56)*

Transparenz von KI-Systemen

Whitepaper des BSI legt den Fokus auf die Anbieter-Pflichten:
„Transparenz von KI-Systemen ist die Bereitstellung von Informationen über den gesamten Lebenszyklus eines KI-Systems sowie über dessen Ökosystem. Transparenz forciert die Zugänglichkeit zu Informationen, die eine Einschätzung des Systems hinsichtlich unterschiedlicher Bedarfe und Ziele ermöglichen, für alle Interessenträge“

Vgl. insb. Art. 13 KI-VO
für Hochrisiko-KI-Systeme

Art. 50 KI-VO umfasst
jedoch auch Pflichten für
Nutzende bzw. Betreiber.
(vgl. für Überblick
auch CoP_KIPerWeb Session 16)

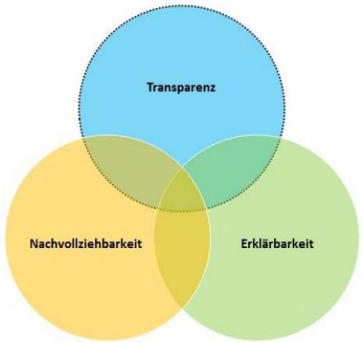
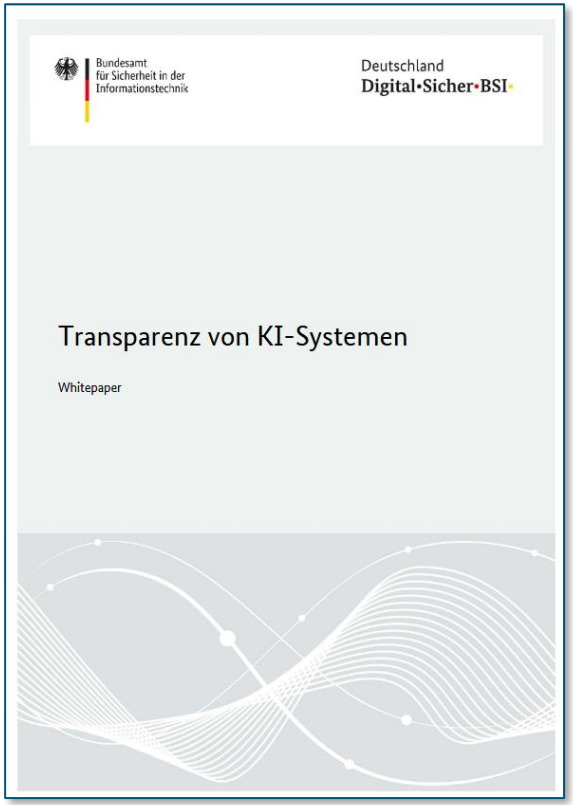


Abbildung 1: Venn-Diagramm zur Verdeutlichung des Zusammenhangs von Transparenz, Erklärbarkeit und Nachvollziehbarkeit im Kontext der Vertrauenswürdigkeit von KI-Systemen. Die verschiedenen Bereiche überschneiden sich, beleuchten aber jeweils einen eigenen Schwerpunkt.



Quelle: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/KI/Whitepaper-Transparenz-KI-Systeme.pdf?__blob=publicationFile&v=3

Exkurs: Art. 13 KI-VO

- An Hochrisiko-KI-Systeme werden besondere Transparenz-Anforderungen gestellt (vgl. Art. 13)

Artikel 13

Transparenz und Bereitstellung von Informationen für die Betreiber

(1) Hochrisiko-KI-Systeme werden so konzipiert und entwickelt, dass ihr Betrieb hinreichend transparent ist, damit die Betreiber die Ausgaben eines Systems angemessen interpretieren und verwenden können. Die Transparenz wird auf eine geeignete Art und in einem angemessenen Maß gewährleistet, damit die Anbieter und Betreiber ihre in Abschnitt 3 festgelegten einschlägigen Pflichten erfüllen können.

(2) Hochrisiko-KI-Systeme werden mit Betriebsanleitungen in einem geeigneten digitalen Format bereitgestellt oder auf andere Weise mit Betriebsanleitungen versehen, die präzise, vollständige, korrekte und eindeutige Informationen in einer für die Betreiber relevanten, barrierefrei zugänglichen und verständlichen Form enthalten.

(3) Die Betriebsanleitungen enthalten mindestens folgende Informationen:

a) den Namen und die Kontaktangaben des Anbieters sowie gegebenenfalls seines Bevollmächtigten;

b) die Merkmale, Fähigkeiten und Leistungsgrenzen des Hochrisiko-KI-Systems, einschließlich

  i) seiner Zweckbestimmung,

  ii) des Maßes an Genauigkeit — einschließlich diesbezüglicher Metriken —, Robustheit und Cybersicherheit gemäß Artikel 15, für das das Hochrisiko-KI-System getestet und validiert wurde und das zu erwarten ist, sowie aller bekannten und vorhersehbaren Umstände, die sich auf das erwartete Maß an Genauigkeit, Robustheit und Cybersicherheit auswirken können;

  iii) aller bekannten oder vorhersehbaren Umstände bezüglich der Verwendung des Hochrisiko-KI-Systems im Einklang mit seiner Zweckbestimmung oder einer vernünftigerweise vorhersehbaren Fehlanwendung, die zu den in Artikel 9 Absatz 2 genannten Risiken für die Gesundheit und Sicherheit oder die Grundrechte führen können,

  iv) gegebenenfalls der technischen Fähigkeiten und Merkmale des Hochrisiko-KI-Systems, um Informationen bereitzustellen, die zur Erläuterung seiner Ausgaben relevant sind;

  v) gegebenenfalls seiner Leistung in Bezug auf bestimmte Personen oder Personengruppen, auf die das System bestimmungsgemäß angewandt werden soll;

  vi) gegebenenfalls der Spezifikationen für die Eingabedaten oder sonstiger relevanter Informationen über die verwendeten Trainings-, Validierungs- und Testdatensätze, unter Berücksichtigung der Zweckbestimmung des Hochrisiko-KI-Systems;

  vii) gegebenenfalls Informationen, die es den Betreibern ermöglichen, die Ausgabe des Hochrisiko-KI-Systems zu interpretieren und es angemessen zu nutzen;

c) etwaige Änderungen des Hochrisiko-KI-Systems und seiner Leistung, die der Anbieter zum Zeitpunkt der ersten Konformitätsbewertung vorab bestimmt hat;

d) die in Artikel 14 genannten Maßnahmen zur Gewährleistung der menschlichen Aufsicht, einschließlich der technischen Maßnahmen, die getroffen wurden, um den Betreibern die Interpretation der Ausgaben von Hochrisiko-KI-Systemen zu erleichtern;

e) die erforderlichen Rechen- und Hardware-Ressourcen, die erwartete Lebensdauer des Hochrisiko-KI-Systems und alle erforderlichen Wartungs- und Pflegemaßnahmen einschließlich deren Häufigkeit zur Gewährleistung des ordnungsgemäßen Funktionierens dieses KI-Systems, auch in Bezug auf Software-Updates;

f) gegebenenfalls eine Beschreibung der in das Hochrisiko-KI-System integrierten Mechanismen, die es den Betreibern ermöglicht, die Protokolle im Einklang mit Artikel 12 ordnungsgemäß zu erfassen, zu speichern und auszuwerten.

Art. 50(1) KI-VO

- (1) Die **Anbieter** stellen sicher, dass KI-Systeme, die für die direkte Interaktion mit natürlichen Personen bestimmt sind, so konzipiert und entwickelt werden, **dass die betreffenden natürlichen Personen informiert werden, dass sie mit einem KI-System interagieren, es sei denn, dies ist aus Sicht einer angemessen informierten, aufmerksamen und verständigen natürlichen Person aufgrund der Umstände und des Kontexts der Nutzung offensichtlich**. Diese Pflicht gilt nicht für gesetzlich zur Aufdeckung, Verhütung, Ermittlung oder Verfolgung von Straftaten zugelassene KI-Systeme, wenn geeignete Schutzvorkehrungen für die Rechte und Freiheiten Dritter bestehen, es sei denn, diese Systeme stehen der Öffentlichkeit zur Anzeige einer Straftat zur Verfügung.

Quelle: https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=OJ:L_202401689#cpt_IV

Art. 50(2) KI-VO

- (2) **Anbieter** von KI-Systemen, einschließlich KI-Systemen mit allgemeinem Verwendungszweck, die synthetische Audio-, Bild-, Video- oder Textinhalte erzeugen, stellen sicher, **dass die Ausgaben des KI-Systems in einem maschinenlesbaren Format gekennzeichnet und als künstlich erzeugt oder manipuliert erkennbar sind**. Die Anbieter sorgen dafür, dass — soweit technisch möglich — ihre technischen Lösungen wirksam, interoperabel, belastbar und zuverlässig sind und berücksichtigen dabei die Besonderheiten und Beschränkungen der verschiedenen Arten von Inhalten, die Umsetzungskosten und den allgemein anerkannten Stand der Technik, wie er in den einschlägigen technischen Normen zum Ausdruck kommen kann. Diese Pflicht gilt nicht, soweit die KI-Systeme eine unterstützende Funktion für die Standardbearbeitung ausführen oder die vom Betreiber bereitgestellten Eingabedaten oder deren Semantik nicht wesentlich verändern oder wenn sie zur Aufdeckung, Verhütung, Ermittlung oder Verfolgung von Straftaten gesetzlich zugelassen sind.

Vgl. zu Wasserzeichen in textgenerierender KI auch https://www.golem.de/news/wasserzeichen-chatgpt-hinterlaesst-unsichtbare-zeichen-im-text-2504-195509.html

Quelle: https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=OJ:L_202401689#cpt_IV

Art. 50(3) KI-VO

- (3) Die **Betreiber eines Emotionserkennungssystems oder eines Systems zur biometrischen Kategorisierung informieren** die davon betroffenen natürlichen Personen über den Betrieb des Systems und verarbeiten personenbezogene Daten gemäß den Verordnungen (EU) 2016/679 und (EU) 2018/1725 und der Richtlinie (EU) 2016/680. Diese Pflicht gilt nicht für gesetzlich zur Aufdeckung, Verhütung oder Ermittlung von Straftaten zugelassene KI-Systeme, die zur biometrischen Kategorisierung und Emotionserkennung im Einklang mit dem Unionsrecht verwendet werden, sofern geeignete Schutzvorkehrungen für die Rechte und Freiheiten Dritter bestehen.

Quelle: https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=OJ:L_202401689#cpt_IV

Art. 50(4) KI-VO

- (4) **Betreiber eines KI-Systems, das Bild-, Ton- oder Videoinhalte erzeugt oder manipuliert, die ein Deepfake sind, müssen offenlegen, dass die Inhalte künstlich erzeugt oder manipuliert wurden.** Diese Pflicht gilt nicht, wenn die Verwendung zur Aufdeckung, Verhütung, Ermittlung oder Verfolgung von Straftaten gesetzlich zugelassen ist. Ist der Inhalt Teil eines offensichtlich künstlerischen, kreativen, satirischen, fiktionalen oder analogen Werks oder Programms, so beschränken sich die in diesem Absatz festgelegten Transparenzpflichten darauf, **das Vorhandensein solcher erzeugten oder manipulierten Inhalte in geeigneter Weise offenzulegen, die die Darstellung oder den Genuss des Werks nicht beeinträchtigt**.

vgl. Art 3, Begriff 60: Für die Zwecke dieser Verordnung bezeichnet der Ausdruck (…) „Deepfake“ 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;

Quelle: https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=OJ:L_202401689#cpt_IV

Art. 50(4) KI-VO (Forts.)

- **Betreiber eines KI-Systems, das Text erzeugt oder manipuliert, der veröffentlicht wird, um die Öffentlichkeit über Angelegenheiten von öffentlichem Interesse zu informieren, müssen offenlegen, dass der Text künstlich erzeugt oder manipuliert wurde.** Diese Pflicht gilt nicht, wenn die Verwendung zur Aufdeckung, Verhütung, Ermittlung oder Verfolgung von Straftaten gesetzlich zugelassen ist oder **[nicht] wenn die durch KI erzeugten 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.**

Quelle: https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=OJ:L_202401689#cpt_IV

Art. 50(5) KI-VO

- (5) Die in den Absätzen 1 bis 4 genannten Informationen werden den betreffenden natürlichen Personen **spätestens zum Zeitpunkt der ersten Interaktion oder Aussetzung in klarer und eindeutiger Weise bereitgestellt. Die Informationen müssen den geltenden Barrierefreiheitsanforderungen entsprechen.**

Quelle: https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=OJ:L_202401689#cpt_IV

Art. 50(6) KI-VO

- (6) Die **Absätze 1 bis 4 lassen die in Kapitel III festgelegten Anforderungen und Pflichten unberührt und berühren nicht andere Transparenzpflichten**, die im Unionsrecht oder dem nationalen Recht für Betreiber von KI-Systemen festgelegt sind.

Quelle: https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=OJ:L_202401689#cpt_IV

Art. 50(7) KI-VO

- (7) Das Büro für Künstliche Intelligenz fördert und erleichtert die Ausarbeitung von Praxisleitfäden auf Unionsebene, um die wirksame Umsetzung der Pflichten in Bezug auf die Feststellung und Kennzeichnung künstlich erzeugter oder manipulierter Inhalte zu erleichtern. Die Kommission kann Durchführungsrechtsakte zur Genehmigung dieser Praxisleitfäden nach dem in Artikel 56 Absatz 6 festgelegten Verfahren erlassen. Hält sie einen Kodex für nicht angemessen, so kann die Kommission einen Durchführungsrechtsakt gemäß dem in Artikel 98 Absatz 2 genannten Prüfverfahren erlassen, in dem gemeinsame Vorschriften für die Umsetzung dieser Pflichten festgelegt werden.

Quelle: https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=OJ:L_202401689#cpt_IV

Diskussion

Wie schätzt ihr die Anforderungen ein und welche Umsetzungsmöglichkeiten haltet ihr für verbreitet oder empfehlenswert, welche für problematisch?

- **Anbieter** müssen KI-Interaktion offenlegen und Ausgaben generativer KI maschinenlesbar kennzeichnen (auch jenseits von Hochrisiko-KI-Systemen)
- **Betreiber** müssen ggf. über Emotionserkennung & biometrische Kategorisierung informieren, Deepfakes offenlegen (ohne Genuss zu beeinträchtigen) und die KI-basierte Erzeugung oder Manipulation von informativen Texten bei der Veröffentlichung offenlegen (oder überprüfen/kontrollieren und verantworten)
- **Information** hat in den genannten Fällen unmittelbar und barrierefrei zu erfolgen und weitere Pflichten sind davon unberührt