Community of Practice KI*PerWeb*

Austausch zur Nutzung und Entwicklung KI-gestützter Webanwendungen

KIPERWEB

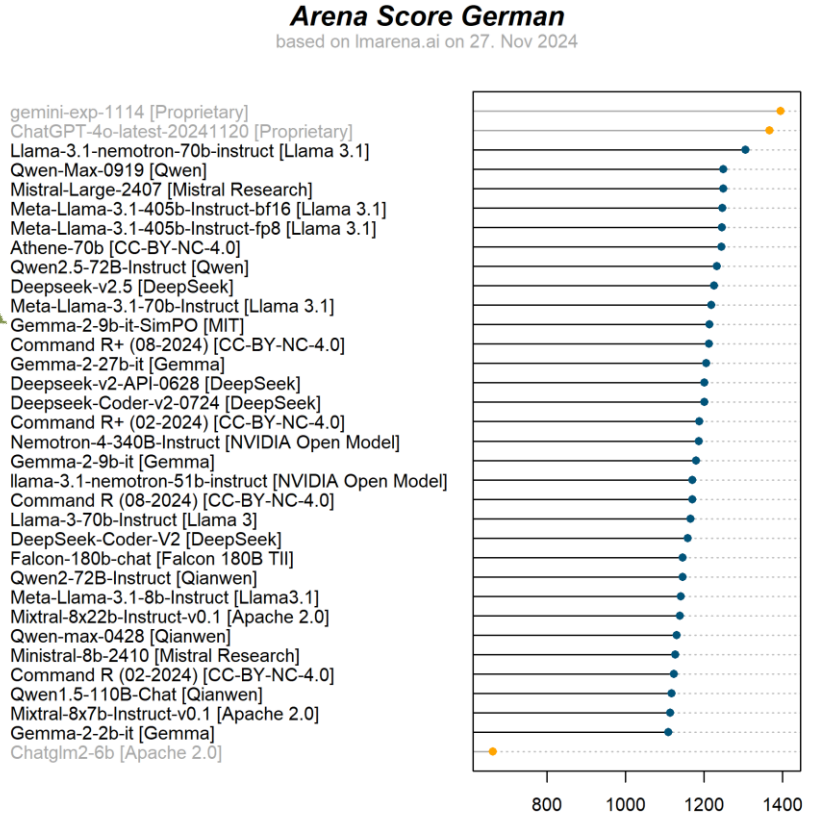
f-bb Forschungsinstitut Betriebliche Bildung

Agenda

- Update
  - News & Leaderboard-Update
- Input
  - „EU AI Act“
- Diskussion

News & Update (27.11.2024)

- Gemini-exp-1114 löst in der Kategorie „German“ ChatGPT-4o-latest-20241120 (2024-09-03) als Klassenprimus ab
- Chatglm2-6b aktuell ausgewiesenes Schlusslicht
- Llama-3.1-nemotron-70b-instruct steigt als bestes nicht-proprietäres Modell neu ein.
- Top-Kandidat für den Hausgebrauch on-premises ist m.E. Gemma-2-9b-it-SimPO mit 1215 (unter MIT Lizenz) Arena-Scores weiterer nicht-proprietärer Modelle sind rechts ausgewiesen sofern sie mindestens das Niveau von Gemma-2-2b-it erreichen



Datengrundlage: https://chat.lmsys.org/?leaderboard

EU AI Act - Verordnung (EU) 2024/1689

- 21.04.21: Vorschlag der Kommission
- 21.05.24: Verabschiedung durch den EU-Rat
- 13.06.24: Datum des Rechtsakts
- 12.07.24: Veröffentlichungsdatum
- 01.08.24: Inkrafttreten
  - Geltung erster Vorschriften ab 02.02.25

Vorschriften u.a. für **Inverkehrbringen**, **Inbetriebnahme** und **Verwendung** von KI-Systemen (Art.1), insb. für **Hochrisiko-KI-Systeme** (im Bildungsbereich u.a. Systeme zu Schülerbewertung, Bewerberauswahl, Zugangsprüfung) mit Sonderregeln für **„General Purpose AI“ (GPAI)**

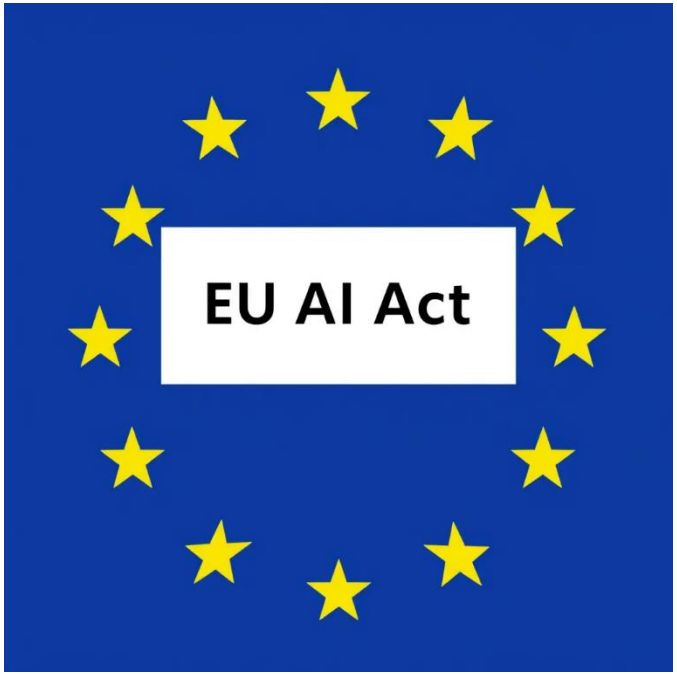


Bild von KI „Flux.1[schnell]“ zum Prompt „Die Flagge der EU als Hintergrundbild mit einem weißen Rechteck in der Mitte. Innerhalb des weißen Rechtecks steht der Titel 'EU AI Act' in schwarzer Schrift. Das Design ist minimalistisch und professionell, mit klaren Linien und einer modernen Schriftart.“

Quellen: https://artificialintelligenceact.eu/de/entwicklungen/ https://www.bundesregierung.de/breg-de/aktuelles/ai-act-2285944 https://eur-lex.europa.eu/eli/reg/2024/1689/oj

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

Quelle: https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Ausgewählte Grenzen des Anwendungsbereichs

Artikel 2

Anwendungsbereich

(1) Diese Verordnung gilt für

a) Anbieter, die in der Union KI-Systeme in Verkehr bringen oder in Betrieb nehmen oder KI-Modelle mit allgemeinem

(6) Diese Verordnung gilt nicht für KI-Systeme oder KI-Modelle, einschließlich ihrer Ausgabe, die eigens für den alleinigen Zweck der wissenschaftlichen Forschung und Entwicklung entwickelt und in Betrieb genommen werden.

(12) Diese Verordnung gilt nicht für KI-Systeme, die unter freien und quelloffenen Lizenzen bereitgestellt werden, es sei denn, sie werden als Hochrisiko-KI-Systeme oder als ein KI-System, das unter Artikel 5 oder 50 fällt, in Verkehr gebracht oder in Betrieb genommen.

e) Produkthersteller, die KI-Systeme zusammen mit ihrem Produkt unter ihrem eigenen Namen oder ihrer Handelsmarke in Verkehr bringen oder in Betrieb nehmen;

f) Bevollmächtigte von Anbietern, die nicht in der Union niedergelassen sind;

g) betroffene Personen, die sich in der Union befinden.

Quelle: https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Auswahl zentraler Begriffsbestimmungen zu KI

Artikel 3

Begriffsbestimmungen

1. „KI-System“ ein maschinengestütztes System, das für einen in unterschiedlichem Grade autonomen Betrieb ausgelegt ist und das nach seiner Betriebsaufnahme anpassungsfähig sein kann und das aus den erhaltenen Eingaben für explizite oder implizite Ziele ableitet, wie Ausgaben wie etwa Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen erstellt werden, die physische oder virtuelle Umgebungen beeinflussen können;

63. „KI-Modell mit allgemeinem Verwendungszweck“ ein KI-Modell — einschließlich der Fälle, in denen ein solches KI-Modell mit einer großen Datenmenge unter umfassender Selbstüberwachung trainiert wird —, das eine erhebliche allgemeine Verwendbarkeit aufweist und in der Lage ist, unabhängig von der Art und Weise seines Inverkehrbringens ein breites Spektrum unterschiedlicher Aufgaben kompetent zu erfüllen, und das in eine Vielzahl nachgelagerter Systeme oder Anwendungen integriert werden kann, ausgenommen KI-Modelle, die vor ihrem Inverkehrbringen für Forschungs- und Entwicklungstätigkeiten oder die Konzipierung von Prototypen eingesetzt werden;

66. „KI-System mit allgemeinem Verwendungszweck“ ein KI-System, das auf einem KI-Modell mit allgemeinem Verwendungszweck beruht und in der Lage ist, einer Vielzahl von Zwecken sowohl für die direkte Verwendung als auch für die Integration in andere KI-Systeme zu dienen;

Quelle: https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Auswahl zentraler Begriffsbestimmungen zu Akteuren

3. „Anbieter“ 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;

4. „Betreiber“ 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;

8. „Akteur“ einen Anbieter, Produkthersteller, Betreiber, Bevollmächtigten, Einführer oder Händler;

20. „Konformitätsbewertung“ ein Verfahren mit dem bewertet wird, ob die in Titel III Abschnitt 2 festgelegten Anforderungen an ein Hochrisiko-KI-System erfüllt wurden;

(vgl. diesbezüglich Ausführungen zu notifizierten Stellen & notifizierender Behörde)

Quelle: https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Auswahl zentraler Passagen zu KI-Kompetenz

Artikel 4

KI-Kompetenz

Die Anbieter und Betreiber von KI-Systemen ergreifen Maßnahmen, um nach besten Kräften sicherzustellen, dass ihr Personal und andere Personen, die in ihrem Auftrag mit dem Betrieb und der Nutzung von KI-Systemen befasst sind, über ein ausreichendes Maß an KI-Kompetenz verfügen, wobei ihre technischen Kenntnisse, ihre Erfahrung, ihre Ausbildung und Schulung und der Kontext, in dem die KI-Systeme eingesetzt werden sollen, sowie die Personen oder Personengruppen, bei denen die KI-Systeme eingesetzt werden sollen, zu berücksichtigen sind.

Artikel 3

Begriffsbestimmungen

56. „KI-Kompetenz“ die Fähigkeiten, die Kenntnisse und das Verständnis, die es Anbietern, Betreibern und Betroffenen unter Berücksichtigung ihrer jeweiligen Rechte und Pflichten im Rahmen dieser Verordnung ermöglichen, KI-Systeme sachkundig einzusetzen sowie sich der Chancen und Risiken von KI und möglicher Schäden, die sie verursachen kann, bewusst zu werden.

Quelle: https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Zentrale Anforderungen an Hochrisiko-KI-Systeme

Artikel 8

Einhaltung der Anforderungen

(1) Hochrisiko-KI-Systeme müssen die in diesem Abschnitt festgelegten Anforderungen erfüllen, wobei ihrer Zweckbestimmung sowie dem allgemein anerkannten Stand der Technik in Bezug auf KI und KI-bezogene Technologien Rechnung zu tragen ist. Bei der Gewährleistung der Einhaltung dieser Anforderungen wird dem in Artikel 9 genannten Risikomanagementsystem Rechnung getragen.

- Risikomanagementsystem (Art. 9)
- Daten und Daten-Governance (Art. 10)
- Technische Dokumentation (Art. 11)
- Aufzeichnungspflichten (Art. 12) iSv. Logging
- Transparenz und Bereitstellung von Informationen für die Betreiber (Art. 13)
- Menschliche Aufsicht (Art. 14)
- Genauigkeit, Robustheit und Cybersicherheit (Art. 15)

Quelle: https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Zentrale Anforderungen an Anbieter und Betreiber

Allg. spezifiziert der EU AI Act „Verbotene Praktiken im KI-Bereich“ (Art. 5)

Zentrale Anforderungen an Anbieter von Hochrisiko-KI-Systemen:

- Pflichten der Anbieter von Hochrisiko-KI-Systemen (Art. 16)
- Qualitätsmanagementsystem (Art. 17)
- Aufbewahrung der Dokumentation (Art. 18)
- Automatisch erzeugte Protokolle (Art. 19)
- Korrekturmaßnahmen und Informationspflicht (Art. 20)
- Zusammenarbeit mit den zuständigen Behörden (Art. 21)
- EU-Konformitätserklärung (Art. 47)
- Transparenzpflichten für Anbieter und Betreiber bestimmter KI-Systeme (Art. 50)
- Beobachtung nach dem Inverkehrbringen durch die Anbieter und Plan für die Beobachtung nach dem Inverkehrbringen für Hochrisiko-KI-Systeme (Art. 72)
- Meldung schwerwiegender Vorfälle (Art. 73)
    - Z.B. Hinweise auf KI-Interaktion, Kennzeichnung von generierten Inhalten etc.

Zentrale Anforderungen an Betreiber von Hochrisiko-KI-Systemen:

- Pflichten der Betreiber von Hochrisiko-KI-Systemen (Art. 26)
- Grundrechte-Folgenabschätzung für Hochrisiko-KI-Systeme (Art. 27)
- Art. 50 (s.o.)

Quelle: https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Zentrale Passagen zu KI mit systemischem Risiko

Artikel 51

Einstufung von KI-Modellen mit allgemeinem Verwendungszweck als KI-Modelle mit allgemeinem Verwendungszweck mit systemischem Risiko

(1) Ein KI-Modell mit allgemeinem Verwendungszweck wird als KI-Modell mit allgemeinem Verwendungszweck mit systemischem Risiko eingestuft, wenn eine der folgenden Bedingungen erfüllt ist:

a) Es verfügt über Fähigkeiten mit hohem Wirkungsgrad, die mithilfe geeigneter technischer Instrumente und Methoden, einschließlich Indikatoren und Benchmarks, bewertet werden;

b) einem unter Berücksichtigung der in Anhang XIII festgelegten Kriterien von der Kommission von Amts wegen oder aufgrund einer qualifizierten Warnung des wissenschaftlichen Gremiums getroffenen Entscheidung zufolge verfügt es über Fähigkeiten oder eine Wirkung, die denen gemäß Buchstabe a entsprechen.

(2) Bei einem KI-Modell mit allgemeinem Verwendungszweck wird angenommen, dass es über Fähigkeiten mit hohem Wirkungsgrad gemäß Absatz 1 Buchstabe a verfügt, wenn die kumulierte Menge der für sein Training verwendeten Berechnungen, gemessen in Gleitkommaoperationen, mehr als 10^{25} beträgt.

Quelle: https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Zentrale Anforderungen an Anbieter allg. KI

Artikel 53

Pflichter

(1) Anbieter von KI-Mod

a) erstellen und aktualisi
Testverfahrens und der
enthält, damit sie dem
Verfügung gestellt werde

b) erstellen und aktualisier
Verfügung, die beabsich
Unbeschadet der Notwe
Geschäftsgeheimnisse im
die Informationen und

i) die Anbieter von KI-
Verwendungszweck g

ii) zumindest die in Anl

c) bringen eine Strategie
insbesondere zur Ermitt
gemachten Rechtsvorbel

d) erstellen und veröffentli
allgemeinem Verwendur
Vorlage.

Artikel 55

Pflichten der Anbieter von KI-Modellen mit allgemeinem Verwendungszweck mit systemischem Risiko

(1) Zusätzlich zu den in den Artikeln 53 und 54 aufgeführten Pflichten müssen Anbieter von KI-Modellen mit allgemeinem Verwendungszweck mit systemischem Risiko

a) eine Modellbewertung mit standardisierten Protokollen und Instrumenten, die dem Stand der Technik entsprechen, durchführen, wozu auch die Durchführung und Dokumentation von Angriffstests beim Modell gehören, um systemische Risiken zu ermitteln und zu mindern,

b) mögliche systemische Risiken auf Unionsebene — einschließlich ihrer Ursachen —, die sich aus der Entwicklung, dem Inverkehrbringen oder der Verwendung von KI-Modellen mit allgemeinem Verwendungszweck mit systemischem Risiko ergeben können, bewerten und mindern,

c) einschlägige Informationen über schwerwiegende Vorfälle und mögliche Abhilfemaßnahmen erfassen und dokumentieren und das Büro für Künstliche Intelligenz und gegebenenfalls die zuständigen nationalen Behörden unverzüglich darüber unterrichten,

d) ein angemessenes Maß an Cybersicherheit für die KI-Modelle mit allgemeinem Verwendungszweck mit systemischem Risiko und die physische Infrastruktur des Modells gewährleisten.

Quelle: https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Ausgewählte Passagen zu Innovationsförderung

KAPITEL VI

MASSNAHMEN ZUR INNOVATIONSFÖRDERUNG

Artikel 57

KI-Reallabore

(1) Die Mitgliedstaaten sorgen dafür, dass ihre zuständigen Behörden mindestens ein KI-Reallabor auf nationaler Ebene einrichten, das bis zum 2. August 2026 einsatzbereit sein muss. Dieses Reallabor kann auch gemeinsam mit den zuständigen Behörden anderer Mitgliedstaaten eingerichtet werden. Die Kommission kann technische Unterstützung, Beratung und I

Die Verpflichtu
eine solche Bet

(5) Die nach Absatz 1 eingerichteten KI-Reallabore bieten eine kontrollierte Umgebung, um Innovation zu fördern und die Entwicklung, das Training, das Testen und die Validierung innovativer KI-Systeme für einen begrenzten Zeitraum vor ihrem Inverkehrbringen oder ihrer Inbetriebnahme nach einem bestimmten zwischen den Anbietern oder zukünftigen Anbietern und der zuständigen Behörde vereinbarten Reallabor-Plan zu erleichtern. In diesen Reallaboren können auch darin beaufsichtigte Tests unter Realbedingungen durchgeführt werden.

(6) Die zuständigen B
bereit, um Risiken, insb
maßnahmen sowie dere
anderem Unionsrecht u

Artikel 60

Tests von Hochrisiko-KI-Systemen unter Realbedingungen außerhalb von KI-Reallaboren

(1) Tests von Hochrisiko-KI-Systemen unter Realbedingungen können von Anbietern oder zukünftigen Anbietern von in Anhang III aufgeführten Hochrisiko-KI-Systemen außerhalb von KI-Reallaboren gemäß diesem Artikel und — unbeschadet der Bestimmungen unter Artikel 5 — dem in diesem Artikel genannten Plan für einen Test unter Realbedingungen durchgeführt werden.

Quelle: https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Hochrisiko-KI-Systeme laut EU AI Act Annex III

1. Biometrics, in so far as their use is permitted under relevant Union or national law
2. Critical infrastructure
3. **Education and vocational training:**
   a) AI systems intended to be used to determine access or admission or to assign natural persons to **educational and vocational training institutions** at all levels;
   b) AI systems intended to be used to evaluate **learning outcomes**, including when those outcomes are used **to steer the learning process** of natural persons in educational and vocational training institutions at all levels;
   c) AI systems intended to be used for the purpose of **assessing the appropriate level** of education that an individual will receive or will be able to access, in the context of or within educational and vocational training institutions;
   d) AI systems intended to be used for **monitoring and detecting prohibited behaviour** of students during tests in the context of or within educational and vocational training institutions.
4. Employment, workers management and access to self-employment
5. Access to and enjoyment of essential private services and essential public services and benefits
6. Law enforcement, in so far as their use is permitted under relevant Union or national law