UNIVERSITÄT
ZU KÖLN

DIE REGULIERUNG VON CHATGPT & CO. DURCH DIE EUROPÄISCHE KI-VERORDNUNG

–

ZU DEN ANFORDERUNGEN AN SOG. KI-MODELLE MIT ALLGEMEINEM VERWENDUNGSZWECK

Institut für Strafrecht und Strafprozessrecht / Mathias Schuh

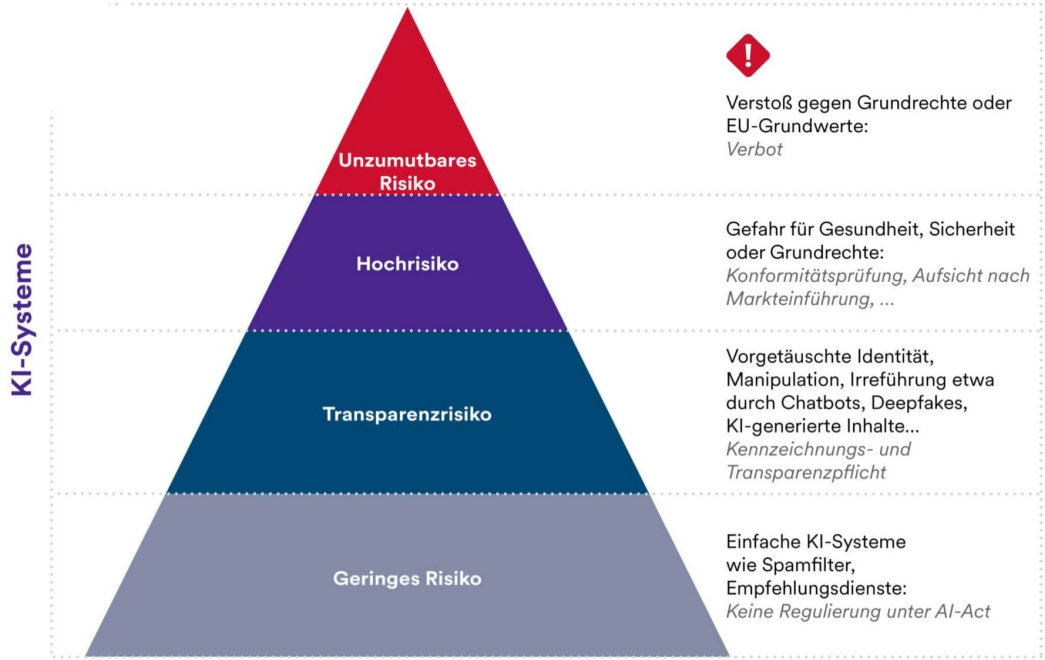
11.09.24

Agenda

- Entstehung, Ziele und Konzepte der KI-VO: Umbruch durch ChatGPT & Co.
- Systematik und Begriffe der KI-Verordnung
- Verantwortung in der Wertschöpfungskette:
  - Anbieter von KI-Modellen
  - Anbieter von KI-Systemen, die KI-Modelle integrieren
  - Betreiber von KI-Systemen

Entstehung, Ziele und Konzepte der KI-VO

- <u>Ziele:</u> „Europa soll das globale Zentrum für vertrauenswürdige KI werden“
  - Wahrung europäischer Grundrechte und –werte
  - Steigerung der Wettbewerbsfähigkeit im KI-Sektor
- <u>Entstehungsgeschichte:</u>
  - April 2021: Entwurf der Europäischen Kommission Risikobasierter Ansatz nach Verwendungszweck
  - November 2022: ChatGPT-Boom
    - **(P)** wie umgehen mit Allzweck-KI?
  - Verabschiedung nach Trilog-Verhandlungen (Mai 2024)



European Parlimanet: Artificial intelligence act, europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf

Verwendungszweck als Anknüpfungspunkt der Regulierung (1)

- „Zweckbestimmung“, Art. 3 Nr. 12 KI-VO
  „die **Verwendung**, für die ein KI-System **laut Anbieter bestimmt ist**, einschließlich der besonderen Umstände und Bedingungen für die Verwendung, entsprechend den **vom Anbieter bereitgestellten Informationen** in den Betriebsanleitungen, im Werbe- oder Verkaufsmaterial und in diesbezüglichen Erklärungen sowie in der technischen Dokumentation;“

- „KI-***Modell*** mit allgemeinem Verwendungszweck“, Art. 3 Nr. 63 KI-VO
  „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 […]“

- „KI-***System*** mit allgemeinem Verwendungszweck“, Art. 3 Nr. 66 KI-VO
  „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;“

Begriffe: KI-Modell und KI-System (1)

- Unterscheidung erheblich:
  - Unterschiedliche Anforderungen an KI-Modelle und KI-Systeme
  - Verbote und Hochrisiko-Anforderungen nur an KI-Systeme
  - KI-Modelle nur dann ausdrücklich adressiert, wenn einen allgemeinen Verwendungszweck haben

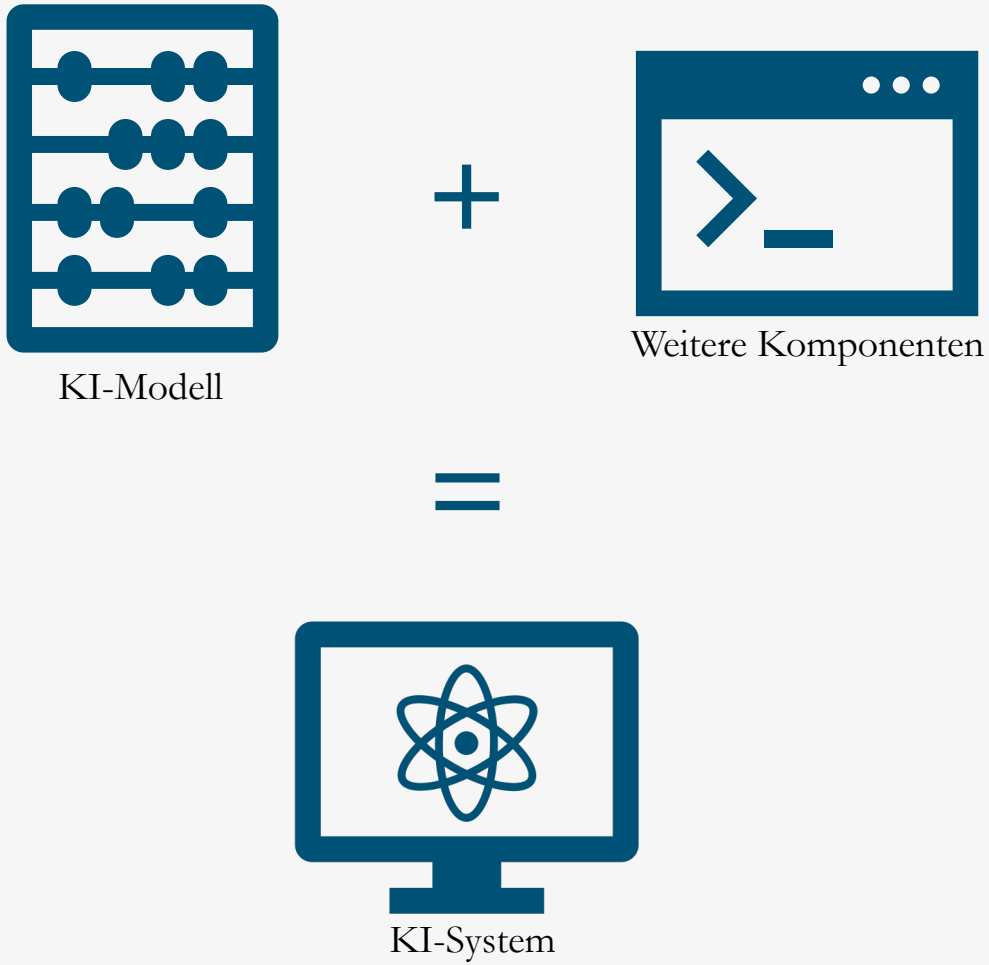
Begriffe: KI-Modell und KI-System (2)

- Keine explizite Abgrenzung des „Modells“ zum „System“ in der Begriffsbestimmung nach Art. 3 Nr. 63 KI-VO:

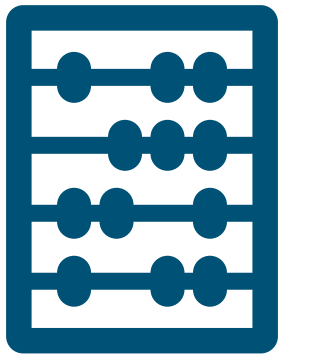
  „KI-Modell mit allgemeinem Verwendungszweck“ (ist) ein KI-Modell […]“

- Abgrenzung in Erwägungsgrund 97 KI-VO:

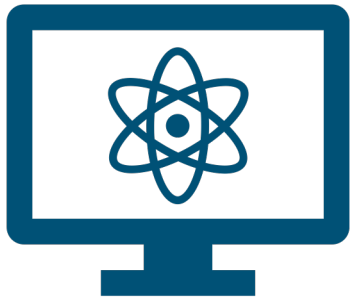
  „Obwohl KI-Modelle wesentliche Komponenten von KI-Systemen sind, stellen sie für sich genommen keine KI-Systeme dar. Damit KI-Modelle zu KI-Systemen werden, ist die **Hinzufügung weiterer Komponenten, zum Beispiel einer Nutzerschnittstelle, erforderlich**.“



Verwendungszweck als Anknüpfungspunkt der Regulierung (2)



KI-Modelle



KI-Systeme

	Allgemeiner Verwendungszweck	Spezifischer Verwendungszweck
KI-Modelle	**Besondere (vorgelagerte) Pflichten nach Art. 50-55 KI-VO**	Keine *ausdrücklichen* Pflichten
KI-Systeme	**Transparenzpflichten für generative KI-Systeme, Art. 50 Abs. 2 KI-VO**	**Verbote und Hochrisiko-Anforderungen (Kap. 2 und 3)**

Zweckbestimmung
(Art. 3 Nr. 12 KI-VO)

Adressieren Modelle als Bestandteil von Systemen

Pflichten entlang der Wertschöpfungskette von Hochrisiko-KI-Systemen (vereinfacht)

Anbieter gem. Art. 3 Nr. 3 KI-VO
= „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**“

Betreiber gem. Art. 3 Nr. 4 KI-VO
= 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

Rollen können auch in einer Person gebündelt sein!

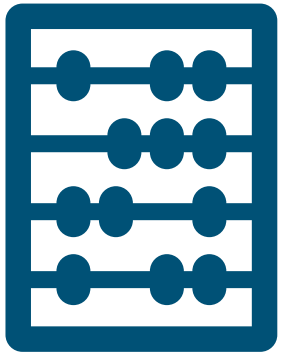
Pflichten entlang der Wertschöpfungskette von Hochrisiko-KI-Systemen (vereinfacht)

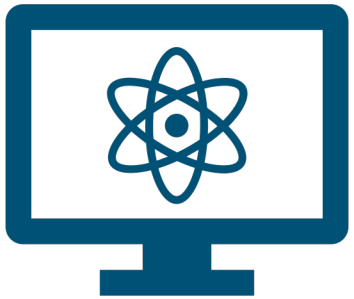
„Modell-Anbieter“ | „System-Anbieter“ | „System-Betreiber“

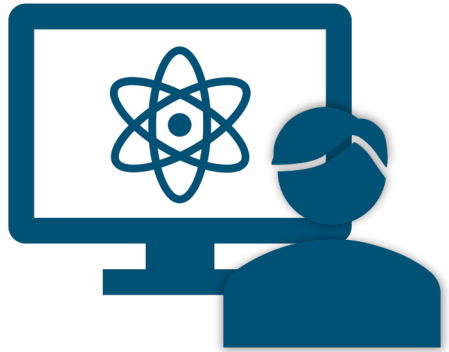
- **Entwickelt KI-Modell** mit allgemeinem Verwendungszweck und bringt es auf den Markt

- **Integriert KI-Modell** mit allgemeinem Verwendungszweck
- Bringt KI-System (*optional: mit Zweckbestimmung*) auf den Markt

- **Verwendet KI-System** in eigener Verantwortung







Pflichten entlang der Wertschöpfungskette von Hochrisiko-KI-Systemen (vereinfacht)

„Modell-Anbieter“ | „System-Anbieter“ | „System-Betreiber“

Risikominderung und Information für risikogeminderte Nutzung

Risikomanagementsystem, Art. 9:
Risikoermittlung und Vornahme geeigneter Minderungsmaßnahmen

Daten-Governance, Art. 10

Technische Dokumentation, Art. 11:
insb. Systemarchitektur und eingebundene Komponenten

Informationen an Betreiber, Art. 13:
Über Daten, Fähigkeiten, Integrationen, Interpretation

Genauigkeit, Robustheit & Cybersicherheit, Art. 15

Ggf. Transparenzpflichten, Art. 50 Abs. 2, 3

Pflichten entlang der Wertschöpfungskette von Hochrisiko-KI-Systemen (vereinfacht)

„Modell-Anbieter“ | „System-Anbieter“ | „System-Betreiber“

Risikominderung und Information für risikogeminderte Nutzung

Überwachung, risikogem. Nutzung

Risikomanagementsystem, Art. 9:
Risikoermittlung und Vornahme geeigneter Minderungsmaßnahmen

Daten-Governance, Art. 10

Technische Dokumentation, Art. 11:
insb. Systemarchitektur und eingebundene Komponenten

Informationen an Betreiber, Art. 13:
Über Daten, Fähigkeiten, Integrationen, Interpretation

Genauigkeit, Robustheit & Cybersicherheit, Art. 15

Überwachung des Betriebs, Art. 26 Abs. 5, insb. Pflicht zur Beobachtung im Hinblick auf Risiken nach Art. 79 Abs. 1 **Überwachungs-, Melde- und ggf. Abschaltpflichten**, insbesondere wenn gesteigertes Risiko identifiziert wird (Abs. 3)

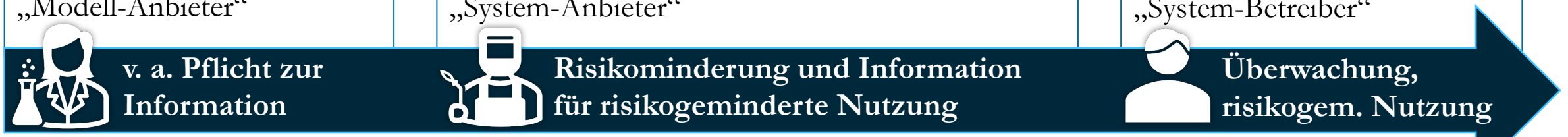
Verwendung nach Betriebsanleitung, Art. 26 Abs. 2

Ggf. Transparenzpflichten, Art. 50 Abs. 2, 3

UNIVERSITÄT ZU KÖLN

⇐ = Pflichterfüllung aufbauend auf

Pflichten entlang der Wertschöpfungskette von Hochrisiko-KI-Systemen (vereinfacht)



Technische Dokumentation und Informationen an Anbieter, Art. 53:
Architektur, Entwurfsspezifikationen, Trainingsmethoden, Trainingsdaten und Trainingstechniken (erweitert bei syst. Risiko)

Strategie zur Einhaltung des Urheberrechts und weiterer Schutzrechte

Risikomanagementsystem, Art. 9:
Risikoermittlung und Vornahme geeigneter Minderungsmaßnahmen

Daten-Governance, Art. 10

Technische Dokumentation, Art. 11:
insb. Systemarchitektur und eingebundene Komponenten

Informationen an Betreiber, Art. 13:
Über Daten, Fähigkeiten, Integrationen, Interpretation

Genauigkeit, Robustheit & Cybersicherheit, Art. 15

Überwachung des Betriebs, Art. 26 Abs. 5, insb. Pflicht zur Beobachtung im Hinblick auf Risiken nach Art. 79 Abs. 1 **Überwachungs-, Melde- und ggf. Abschaltpflichten,** insbesondere wenn gesteigertes Risiko identifiziert wird (Abs. 3)

Verwendung nach Betriebsanleitung, Art. 26 Abs. 2

Ggf. Transparenzpflichten, Art. 50 Abs. 2, 3

⇐ = Pflichterfüllung aufbauend auf

Regulierung von KI-Modellen durch KI-Verordnung

- Anbieter von KI-Modellen mit allgemeinem Verwendungszweck ***mit*** systemischem Risiko
  - Definiert in Art. 51 Abs. 1 KI-VO:
    - a) Es verfügt über **Fähigkeiten mit hohem Wirkungsgrad**, die mithilfe geeigneter technischer Instrumente und Methoden, einschließlich Indikatoren und Benchmarks, bewertet werden; oder
    - 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
  - Konkretisierung in Art. 51 Abs. 2 KI-VO:
    - 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

Regulierung von KI-Modellen durch KI-Verordnung

- Anbieter von KI-Modellen mit allgemeinem Verwendungszweck ***mit*** systemischem Risiko
  - Wesentliche ergänzende Pflichten sind in Art. 55 Abs. 1 KI-VO kodifiziert:
    - a) eine **Modellbewertung** mit standardisierten Protokollen und Instrumenten **durchzuführen**, die dem Stand der Technik entsprechen, wozu auch die Durchführung und Dokumentation von Angriffstests beim Modell gehören, um systemische Risiken zu ermitteln;
    - b) **mögliche systemische Risiken** auf Unionsebene – einschließlich ihrer Ursachen – zu **bewerten und** zu **mindern**;
    - c) **einschlägige Informationen über schwerwiegende Vorfälle** und mögliche Abhilfemaßnahmen zu erfassen, zu dokumentieren und die zuständigen Behörden hierüber zu informieren; sowie
    - d) ein **angemessenes Maß an Cybersicherheit** zu gewährleisten.

Pflichten entlang der Wertschöpfungskette von Hochrisiko-KI-Systemen (vereinfacht)

„Modell-Anbieter“ | „System-Anbieter“ | „System-Betreiber“

v. a. Pflicht zur Information

Risikominderung und Information für risikogeminderte Nutzung

Überwachung, risikogem. Nutzung

Technische Dokumentation und Informationen an Anbieter, Art. 53:
Architektur, Entwurfsspezifikationen, Trainingsmethoden, Trainingsdaten und Trainingstechniken (erweitert bei syst. Risiko)

Strategie zur Einhaltung des Urheberrechts und weiterer Schutzrechte

Risikomanagementsystem, Art. 9:
Risikoermittlung und Vornahme geeigneter Minderungsmaßnahmen

Daten-Governance, Art. 10

Technische Dokumentation, Art. 11:
insb. Systemarchitektur und eingebundene Komponenten

Informationen an Betreiber, Art. 13:
Über Daten, Fähigkeiten, Integrationen, Interpretation

Genauigkeit, Robustheit & Cybersicherheit, Art. 15

Überwachung des Betriebs, Art. 26 Abs. 5, insb. Pflicht zur Beobachtung im Hinblick auf Risiken nach Art. 79 Abs. 1 **Überwachungs-, Melde- und ggf. Abschaltpflichten,** insbesondere wenn gesteigertes Risiko identifiziert wird (Abs. 3)

Verwendung nach Betriebsanleitung, Art. 26 Abs. 2

Ggf. Transparenzpflichten, Art. 50 Abs. 2, 3

Risikominderungsmaßnahmen bei syst. Risiko, Art. 55
Angriffstests
Risikobewertung/Minderung systemischer Risiken

⇐ = Pflichterfüllung aufbauend auf

UNIVERSITÄT
ZU KÖLN

VIELEN DANK FÜR IHRE AUFMERKSAMKEIT!

Kontakt: mathias.schuh@uni-koeln.de