Die Klassifizierung von KI-Systemen nach der KI-VO

Vorschlag zur matrixbasierten Risikoklassifizierung

Ulrich Spiegel/Maximilian Höving

Ziel des Beitrages ist, die Risikoklassifizierung von KI-Systemen nach der KI-Verordnung (KI-VO) greifbarer zu machen. Dazu gibt der Beitrag zunächst eine Übersicht über die de lege lata eingeführten Klassen sowie deren Beziehung untereinander und analysiert sodann die bestehenden grafischen Aufarbeitungen dieser Klassen. Ausgehend von der Analyse gibt der Beitrag abschließend den Vorschlag, grafische Darstellungen auf Grundlage einer matrixbasierten Risikoklassifizierung vorzunehmen.

I. Notwendige Rechtssetzung: Regeln für KI-Systeme?

KI-Systeme zu regulieren war, wie es die Europäische Union (EU) nunmehr mit der KI-VO[1] umgesetzt hat, überfällig. KI- und andere automatisierte Systeme können zwar zur Lösung gesellschaftlicher Herausforderungen beitragen, schaffen aber gleichzeitig zu bewältigende Risiken. Zwei Praxisbeispiele sollen typische Risiken solcher Systeme verbildlichen:

Automatisierte Systeme schaffen schon seit geraumer Zeit eine Gefahr für die Benutzer und die Allgemeinheit. Drastisch zeigt das ein Fall, der sich bereits im Jahr 2012 bei Alzenau[2] ereignete: Ein mit einem Spurhalteassistent ausgerüstetes Fahrzeug fuhr mit hoher Geschwindigkeit in eine Ortschaft und tötete eine Frau und ihr Kind. Obwohl der Fahrer infolge eines Schlaganfalls das Lenkrad verrissen hatte, führte der Spurhalteassistent das Fahrzeug wieder zurück auf die Straße und in die Ortschaft.

Ein weiteres Risiko betrifft die Anwendung von KI zur Begehung von Straftaten, insbesondere durch Deepfakes[3]. So haben sich Betrugsmaschen mit täuschend echten Imitationen von Stimmen entwickelt. Es versuchten etwa Unbekannte, mit der gefälschten Stimme von Ferrari-Chef Benedetto Vigna einen Manager dazu zu bringen, Geld zu überweisen.[4] Das Risiko von Deepfakes steigt mit dem technischen Fortschritt. Bereits 2023 glaubten 81 % der Teilnehmer einer Umfrage des Bitkom e. V., Deepfakes nicht erkennen zu können.[5]

Die Ausgangslage ist de facto und de jure herausfordernd: Anbieter von KI-Systemen sind mitunter zwangsläufig außerstande vorherzusagen, welche Entscheidung ein KI-System warum trifft. Die KI-VO soll das bisherige Produktsicherheitsrecht[6], das den Digitalisierungstrend teilweise adressiert, durch umfassende Anforderungen ergänzen, um den Herausforderungen von KI-Systemen zu begegnen.[7]

Aufgabe der KI-VO als Teil der Harmonisierungsvorschriften der EU[8] ist es, mögliche regulatorische Lücken zu schließen. Die KI-VO hat es sich ausweislich ihrer Erwägungsgründe zur Aufgabe gemacht, den spezifischen Herausforderungen von KI-Systemen – insbesondere im Hinblick auf ein einheitliches Schutzniveau – mit verschiedenen Pflichten und Verantwortlichkeiten der maßgeblichen Akteure zu begegnen. Ihren risikobasierten Ansatz nimmt die KI-VO aus internationalen Rechtsquellen: Sie schafft zB einen Gleichklang mit der aktualisierten Definition von KI-Systemen der OECD.[9]

II. Rechtsquellen: Klassifizierung von KI-Systemen als weltweite Idee des risikobasierten Ansatzes gemäß der OECD

Der „OECD Framework for the Classification of AI systems“[10] („KI-Grundsätze der OECD“) stellt fest, dass Gesetzgeber bei der Regulierung von KI einen risikobasierten Ansatz bevorzugen, um Aufsicht und Eingriffe dort zu konzentrieren, wo sie am nötigsten sind. Gleichzeitig versuchten die Gesetzgeber so, Innovationshemmnisse zu vermeiden.[11]

In den KI-Grundsätzen der OECD heißt es sinngemäß: KI-Akteure sollten auf der Grundlage ihrer Rolle, ihres Kontexts und ihrer Handlungsfähigkeit in jeder Phase des Lebenszyklus von KI-Systemen kontinuierlich einen systematischen Risikomanagementansatz anwenden, um Risiken im Zusammenhang mit KI-Systemen, einschließlich des Schutzes der

Privatsphäre, der digitalen Sicherheit, der Sicherheit und der Bias-Problematik, zu begegnen.[12]

Die KI-Grundsätze der OECD halten weiter fest, dass die mit dem Einsatz von KI-Systemen verbundenen Risiken stark von der jeweiligen Anwendung abhängen. Der Spagat zwischen abstrakter und spezifischer Regulierung soll gemeistert

werden, indem KI-Systeme in wenige Risikogruppen eingeteilt werden. Verschiedene Gremien (zB die deutsche Datenethikkommission und die IEC SEG10) haben vier bis fünf Risikostufen vorgeschlagen.[13]

Die Europäische Kommission (EU-Kommission) hat eine Grafik des risikobasierten Ansatzes auf ihrer Webseite zu dem KI-Regulierungsvorhaben veröffentlicht und vier Risikoklassen vorgeschlagen: Unannehmbares Risiko, hohes Risiko, begrenztes Risiko und minimales Risiko.[14] Diese Klassifizierung von KI-Systemen bildet die zentrale Weichenstellung für die gesetzgeberische Umsetzung und praktische Anwendung der KI-VO.

III. Umsetzung: Die Systematik der Klassifizierung von KI-Systemen nach der KI-VO konkretisiert die Idee des risikobasierten Ansatzes

Die aus der KI-VO herauszuarbeitende Klassifizierungssystematik von KI-Systemen ist von höchster Relevanz, vor allem für den Rechtsanwender. Denn auf der Klassifizierung beruhen die Verpflichtungen und Verantwortlichkeiten der beteiligten „Akteure“[15], die sich nach Risikoklasse unterscheiden.

1. Trotz horizontaler Regelungswirkung sind gewisse KI-Systeme von der KI-VO vollständig ausgenommen

Die KI-VO enthält hinsichtlich ihres sachlichen Anwendungsbereichs Ausnahmen für gewisse KI-Systeme. Die Ausnahmen sind unabhängig von der Frage des Risikos, das von den jeweiligen KI-Systemen ausgeht, und stellen die Frage nach dem Zweck oder der Beschaffenheit des KI-Systems. Ausgenommen sind KI-Systeme

- ausschließlich für militärische Zwecke, Verteidigungszwecke oder Zwecke der nationalen Sicherheit (Art. 2 Abs. 3, UAbs. 2 und 3 KI-VO),
- für den alleinigen Zweck der wissenschaftlichen Forschung und Entwicklung entwickelt und in Betrieb genommen werden (Art. 2 Abs. 6 KI-VO) sowie
- die unter freien und quelloffenen Lizenzen bereitgestellt werden, es sei denn, sie werden als Hochrisiko-KI-Systeme oder als ein KI-System, das unter Art. 5 KI-VO oder Art. 50 KI-VO fällt, in Verkehr gebracht oder in Betrieb genommen (Art. 2 Abs. 12 KI-VO).

2. Praktiken mit inakzeptablem Risiko sind verboten (Verbotene KI-Praktiken)

Kapitel II KI-VO und der einzige darin enthaltene Artikel, Art. 5 KI-VO, verbietet bestimmte „Praktiken". Verboten ist als Praktik das „Inverkehrbringen, die Inbetriebnahme oder die Verwendung“ von KI-Systemen, und zwar dann, wenn KI-Systeme bestimmte, in Art. 5 KI-VO näher beschriebene, Eigenschaften oder Funktionalitäten aufweisen[16]. Damit verfolgt Art. 5 KI-VO prima facie einen anderen Regulierungsansatz als die Kapitel III-V, in denen der Gesetzgeber explizit eine risikobasierte Klassifizierung von KI-Systemen vorgesehen hat. Trotz des ersten Eindrucks bietet es sich an, die in Art. 5 KI-VO näher beschriebenen KI-Systeme als eigene Risikoklasse einzuordnen. Denn bei näherer Betrachtung liegt auch den KI-Systemen nach Art. 5 KI-VO der risikobasierte Ansatz der KI-VO zugrunde. Risiken der in Art. 5 KI-VO genannten Praktiken sind nach Ansicht des Gesetzgebers inakzeptabel hoch – insbesondere im Hinblick auf die Grundrechte der betroffenen Personen.[17]

„Verbote“ ergeben sich indes nicht nur aus der KI-VO selbst. Jedes der Verbote in Art. 5 Abs. 1 KI-VO hat inhaltliche Bezüge zu anderen Rechtsakten, die ähnliche Verbote oder zusätzliche Beschränkungen enthalten, wie zB die Richtlinie 2005/29/EG oder die Verordnung (EU) 2016/679.[18] Darüber hinaus kann sich ein Verbot auch aus anderen Vorschriften des Unionsrechts ergeben, Art. 5 Abs. 8 KI-VO. Damit ist gemeint, dass jedes Verhalten, das generell verboten ist, auch unter Zuhilfenahme von KI-Systemen verboten bleibt.[19]

3. Die Klassifizierung von KI-Systemen mit hohem Risiko ist an ihrer Zweckbestimmung ausgerichtet (Hochrisiko-KI-Systeme)

Die Regulierung der Hochrisiko-KI-Systeme (Kapitel III KI-VO) ist das Herzstück der KI-VO.[20] Akteure, die ein solches KI-System entwickeln und in Betrieb nehmen oder in Verkehr bringen möchten, unterliegen den weitreichendsten Pflichten nach der KI-VO. Maßgeblich für die Einordnung ist die Zweckbestimmung des Anbieters eines KI-Systems.

Die KI-VO definiert den Begriff Zweckbestimmung in Art. 3 Nr. 12 KI-VO. Maßgeblich sind allein die vom Anbieter[21] bereitgestellten Informationen (zB Betriebsanleitung, Werbe- oder Verkaufsmaterial und diesbezügliche Erklärungen sowie technischen Dokumentation).[22] Für die klassifizierungsrele-

vante Zweckbestimmung bleibt die übliche Verwendung,[23] damit grundsätzlich außer Betracht. Die KI-VO folgt damit dem üblichen, wenn auch kaum ausdrücklich definierten,[24] produktrechtlichen Zweckbestimmungsansatz.

a) Die Systematik des Art. 6 Abs. 1-3 KI-VO

Die KI-VO enthält einen zweigeteilten Ansatz der Risikozuweisung. Sie unterscheidet zwischen KI-Systemen, die als (Teil bestimmter) Produkte unter die in Anhang I Abschnitt A KI-VO aufgeführten EU-Rechtsvorschriften fallen, weswegen ein

hohes Risiko von ihnen ausgehen soll, und sog. eigenständigen KI-Systemen, die intrinsisch ein hohes Risiko tragen und in Anhang III KI-VO genannt sind.[25]

Die KI-VO greift zur Bestimmung der Risikoqualität, also einerseits auf andere Normen, etwa des sog. „New Legislative Framework" (NLF)[26], aber auch weiterer Harmonisierungsvorschriften zurück,[27] Art. 6 Abs. 1 KI-VO. Das KI-System muss kumulativ erstens ein Sicherheitsbauteil eines der in Anhang I KI-VO genannten Produkte sein oder selbst unter den Anwendungsbereich dieser Normen fallen, Art. 6 Abs. 1 lit. a KI-VO und zweitens einem Konformitätsbewertungsverfahren unterliegen, das vorsieht, dass Dritte (etwa benannte Stellen) zu involvieren sind, Art. 6 Abs. 1 lit. b KI-VO.

Andererseits hängt die Klassifikation von dem Einsatzkontext des KI-Systems ab, Art. 6 Abs. 2 und 3 KI-VO. Der EU-Gesetzgeber definiert diese kritischen Einsatzgebiete in Anhang III KI-VO selbst. Maßgeblich ist die Zweckbestimmung für das Einsatzgebiet.[28]

Im Hinblick auf das Verhältnis von Art. 6 Abs. 1 und Abs. 2 KI-VO stellt sich die Frage nach einer Sperrwirkung, dh ob für ein KI-System, das den Tatbestand des Art. 6 Abs. 1 KI-VO erfüllt, die Anwendung von Art. 6 Abs. 2 KI-VO aufgrund von Spezialität ausgeschlossen wäre (und vice versa). In der Literatur gibt es hierzu ein offenes Meinungsbild. Autoren greifen das Verhältnis zwischen den beiden Absätzen auf und verweisen auf Erwägungsgrund 52 KI-VO und Art. 43 KI-VO, argumentieren mit den Schutzgütern der KI-VO und der Vermeidung von Schutzlücken.[29] Um sicherzustellen, alle formalen Anforderungen der KI-VO einzuhalten, empfiehlt es sich – bis zur obergerichtlichen Klärung – eine fehlende Sperrwirkung zugrunde zu legen.

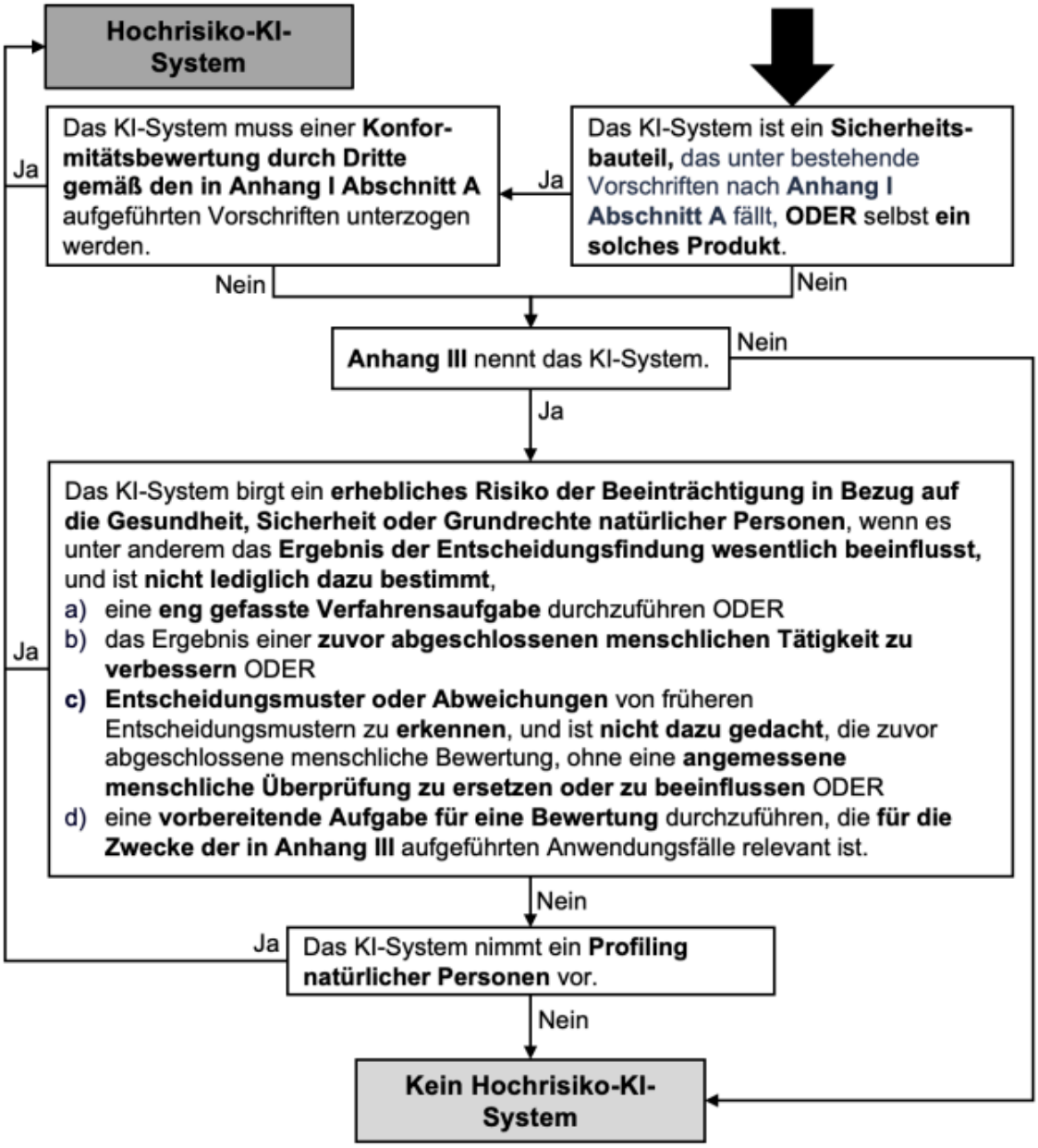


Abb. 1: Entscheidungsbaum zum schrittweisen Vorgehen bei Beurteilung von KI-Systemen nach Art. 6 Abs. 1-3 KI-VO

Die Grafik (Abb. 1) berücksichtigt, dass unter der KI-VO Hochrisiko-KI-System nicht gleich Hochrisiko-KI-System ist, und hält sich an die Reihenfolge der Absätze in Art. 6 KI-VO[30]. So bleiben diejenigen Hochrisiko-KI-Systeme in der Grafik außen vor, deren Klassifizierung aus Anhang I Abschnitt B KI-VO herrührt. Denn die KI-VO schließt diese Hochrisiko-KI-Systeme auf der Rechtsfolgenseite de facto aus ihrem Anwendungsbereich aus[31] und damit auch die unmittelbare Anwendung der Pflichten unter der KI-VO. Folge ist, dass die Umsetzung der Anforderungen an diese Hochrisiko-KI-Systeme von weiteren Rechtsakten abhängig ist (s. Abschnitt III. 3. b).

b) Das Verhältnis zwischen Hochrisiko-KI-Systemen und der Einschränkung des Anwendungsbereichs der KI-VO ist zu klären

Es ist zwischen Hochrisiko-KI-Systemen, die die KI-VO selbst regelt, und solchen Hochrisiko-KI-Systemen zu unterscheiden, deren Anforderungen mit weiteren Rechtsakten festzulegen sind.

Die KI-VO enthält für Hochrisiko-KI-Systeme nach Art. 6 Abs. 1 KI-VO iVm Anhang I KI-VO Anforderungen. Aber: Nur für Hochrisiko-KI-Systeme nach Art. 6 Abs. 1 KI-VO iVm Anhang I Abschnitt A KI-VO sind solche Anforderungen auch unmittelbar anwendbar. Für die in Art. 6 Abs. 1 KI-VO iVm Anhang I Abschnitt B KI-VO genannten (Hochrisiko-)KI-Systeme enthält Art. 2 Abs. 2 KI-VO auf Rechtsfolgenseite eine Einschränkung des sachlichen Anwendungsbereichs.[32] Demnach gelten nur die Art. 6 Abs. 1, Art. 102-109 KI-VO, Art. 112 KI-VO sowie, mit Einschränkungen, Art. 57 KI-VO.

Das lässt sich wie folgt begründen: Es handelt sich bei den in Anhang I Abschnitt B KI-VO zitierten Rechtsakten überwiegend um solche, die nicht oder nur teilweise dem NLF folgen.[33] Aus diesen ergibt sich regelmäßig eine eigenständige Detailregulierung auf Gesetzesebene – ohne Rückgriff auf harmonisierte Normen.[34] (Hochrisiko-) KI-Systeme, die unter Vorschriften nach Anhang I Abschnitt B KI-VO fallen, reguliert die KI-VO nur mittelbar und nur teilweise. Denn die Anforderungen an KI-Systeme nach Kapitel III Abschnitt 2 KI-VO sind noch in den jeweiligen Rechtsakten bzw. durch die auf diese Rechtsakte gestützten delegierten Rechtsakte, Durchführungsrechtsakte, technischen Spezifikationen, Prüfnormen und ähnlichen Maßnahmen der Kommission zu integrieren.[35] Dies wirkt sich vor allem auf den Umfang von Anbieterpflichten aus, denn die Verweise in den Art. 102 bis 109 KI-VO auf Kapitel III Abschnitt 2 KI-VO entsprechen lediglich Art. 16 lit. a KI-VO. Es fehlen indes Verweise auf Art. 16 lit. b bis lit. l KI-VO.

Schließlich stellt sich wieder die Frage nach einer Sperrwirkung, dh ob für ein KI-System, das in den Tatbestand des Art. 6 Abs. 1 KI-VO iVm Anhang I Abschnitt B KI-VO fällt, die Anwendung von Art. 6 Abs. 2 KI-VO aufgrund von Spezialität ausgeschlossen ist (und vice versa) (s. o. unter III.3.a) Nach einer Literaturansicht soll der Rechtsetzungsprozess darauf hindeuten, dass ein KI-System, obwohl es gem. Art. 2 Abs. 2 KI-VO vom Anwendungsbereich der KI-VO ausgenommen ist, als ein Hochrisiko-KI-System zu klassifizieren ist, wenn es die Voraussetzungen des Art. 6 Abs. 2 KI-VO erfüllt.[36] Dagegen spricht der Wortlaut von Art. 2 Abs. 2 Satz 1: „gelten nur Artikel 6 Absatz 1 …“. Ebenso die fehlende Vereinbarkeit der in Anhang I Abschnitt B KI-VO zitierten Normen mit dem NLF spricht für eine Ablehnung einer Sperrwirkung.

c) Unternehmen müssen in der Praxis Entscheidungen über Umsetzungsmaßnahmen treffen

In der Praxis stellen sich vielfältige Konstellationen, die sich mithilfe des aktuellen Gesetzestextes keiner eindeutigen Lösung zuführen lassen. Als Beispiel mag man an eine Emotionserkennungsfunktion (nach Art. 6 Abs. 2, 3 UAbs. 3 KI-VO) in einem Kraftfahrzeug zur Personenbeförderung, wie etwa einem Stadtbus (gem. Art. 2 Abs. 2 iVm Anhang I Abschnitt B KI-VO), denken. In diesem Fall fällt das Emotionserkennungssystem unter die KI-VO, das Kraftfahrzeug aber unter Anhang I Abschnitt B KI-VO. Unklar ist auch der Umgang mit einer Maschine mit einem KI-System (nach Art. 6 Abs. 1 KI-VO) zum Einsatz in kritischer Infrastruktur (nach Art. 6 Abs. 2), welche nicht die Ausnahmevorschriften (nach Art. 6 Abs. 3 KI-VO) erfüllt. In diesem Fall stehen die Anbieter vor einer Herausforderung im Hinblick auf das anzuwendende Konformitätsbewertungsverfahren: Gem. Art. 6 Abs. 2 iVm 43 Abs. 2 KI-VO ist das Konformitätsbewertungsverfahren auf der Grundlage einer internen Kontrolle gem. Anhang VI KI-VO durchzuführen, das von einer Beteiligung einer notifizierten Stelle absieht. Gem. Art. 6 Abs. 1 und Art. 43 Abs. 3 KI-VO ist allerdings das einschlägige, durch die KI-VO modifizierte Konformitätsbewertungsverfahren der entsprechenden Harmonisierungsrechtsakte der Union (hier zB Maschinenrichtlinie) durchzuführen.

4. Der Gesetzgeber begegnet spezifischen Risiken von KI-Systemen mit Transparenzpflichten (Bestimmte KI-Systeme)

Neben den Hochrisiko-KI-Systemen enthält die KI-VO in Kapitel IV (Art. 50 KI-VO) eine eigenständige Regulierungsklasse, die sog. „bestimmten KI-Systeme“. Der europäische Gesetzgeber führt besondere Schutzbestimmungen zur Bewältigung der Gefahren ein, die von „bestimmten KI-Systemen“ ausgehen, weil deren Leistungen menschlichen Fähigkeiten nahe kommen und die betroffenen KI-Systeme deshalb Manipulations- bzw.

Verwechslungsrisiken beim Betroffenen hervorrufen. Nach dem risikobasierten Ansatz der EU-Kommission geht von diesen „bestimmten" KI-Systemen ein besonderes Risiko in Bezug auf Identitätsbetrug oder Täuschung aus.[37]

Aus diesen Erwägungen heraus stellen die bestimmten KI-Systeme zunächst eine eigenständig regulierte Gruppe von KI-Systemen dar. Diese Regulierungsklasse umfasst insbeson-

dere die in Art. 3 Abs. 66 KI-VO definierten „KI-Systeme mit allgemeinem Verwendungszweck". Um das Risiko von bestimmten KI-Systemen abzumildern, enthält Art. 50 KI-VO Transparenzpflichten und ist zugleich bei der Abgrenzung zwischen Hochrisiko-KI-Systemen und bestimmten KI-Systemen behilflich.

Ist ein KI-System sowohl gem. Art. 6 KI-VO als ein Hochrisiko-KI-System zu klassifizieren als auch ein bestimmtes KI-System gem. Art. 50 Abs. 1-4 KI-VO („bestimmtes KI-System"), sieht Art. 50 Abs. 6 KI-VO vor, dass die Anforderungen und Pflichten beider Klassen anzuwenden sind (s. unten Abb. 2).

Das bedeutet, dass ein Anbieter eines Hochrisiko-KI-Systems, das auch ein bestimmtes KI-System ist, sowohl die Transparenzpflichten der bestimmten KI-Systeme als auch die Anforderungen eines Hochrisiko-KI-Systems und die korrespondierenden Anbieterpflichten zu erfüllen hat.

Die normativen Vorgaben des Art. 13 KI-VO (Transparenzpflichten im Hinblick auf Hochrisiko-KI-Systeme) und des Art. 50 KI-VO (Transparenzpflichten im Hinblick auf bestimmte KI-Systeme) unterscheiden sich in ihrem Regelungsgehalt. Art. 13 KI-VO zielt bereits auf die Entwicklungsphase des Hochrisiko-KI-Systems, während Art. 50 KI-VO neben Anbietern auch Betreiber, also diejenigen, die ein KI-System in eigener Verantwortung verwenden (Art. 3 Nr. 4 KI-VO), adressiert.

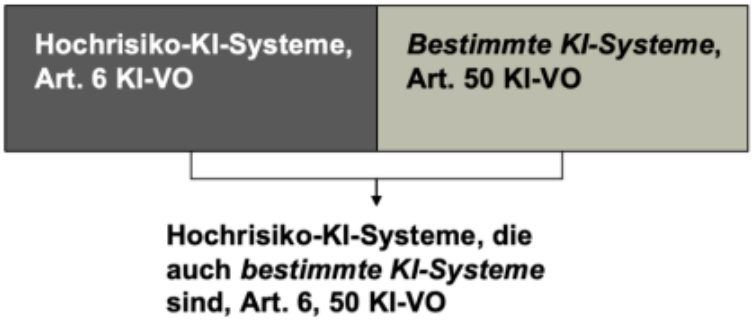


Abb. 2: Übersicht zum Verhältnis von Hochrisiko-KI-Systemen und bestimmten KI-Systemen sowie ihrem Verhältnis

5. Für die übrigen KI-Systeme sieht die KI-VO freiwillige Selbstregulierung vor (Andere als Hochrisiko-KI-Systeme)

Alle KI-Systeme, die entweder gemäß Art. 6 Abs. 1, 2 KI-VO oder nach Art. 6 Abs. 3 KI-VO als Hochrisiko-KI-Systeme ausscheiden („Andere als Hochrisiko-KI-Systeme"), sollen freiwilligen Verhaltenskodizes unterliegen, Art. 95 KI-VO. Diese Verhaltenskodizes können einige oder alle Anforderungen für Hochrisiko-KI-Systeme aus Kapitel III Abschnitt 2 KI-VO einschließen. Zudem sollen die Anforderungen auf die Zweckbestimmung der Systeme und das niedrigere Risiko angepasst werden.[38] Sprachlich und systematisch schließen „Andere als Hochrisiko-KI-Systeme" die bestimmten KI-Systeme grundsätzlich ein. Etwas anderes gilt, wenn ein bestimmtes KI-System zugleich ein Hochrisiko-KI-System ist.

Der Anschein, „Andere als Hochrisiko-KI-Systeme" und bestimmte KI-Systeme bewegten sich in regulierungsfreiem Raum, trügt. So gilt gem. Art. 74 Abs. 1 KI-VO die Marktüberwachungsverordnung, VO (EU) 2019/1020 für alle KI-Systeme unter der KI-VO. Im Hinblick auf Erwägungsgrund 166 KI-VO sind „Andere als Hochrisiko-KI-Systeme" auch insofern einer gewissen Regulierungswirkung unterworfen, als sie sicher sein müssen. Denn die Produktsicherheitsverordnung (VO (EU) 2023/988) soll als Sicherheitsnetz dienen und soll – so zumindest Wille der EU-Kommission – auch auf Software Anwendung finden[39].

6. Zusammenfassende Übersicht zu KI-Systemen

Aus den vorstehenden Erwägungen ergibt sich die folgende Übersicht (Abb. 3) mit den unterschiedlichen Arten von KI-Systemen. Die Verhältnisse, die als Text dargestellt sind, werden mit Pfeilen verdeutlicht, wobei noch zu klären ist, wie Hochrisiko-KI-Systeme zu von der KI-VO ausgenommenen KI-Systemen stehen.

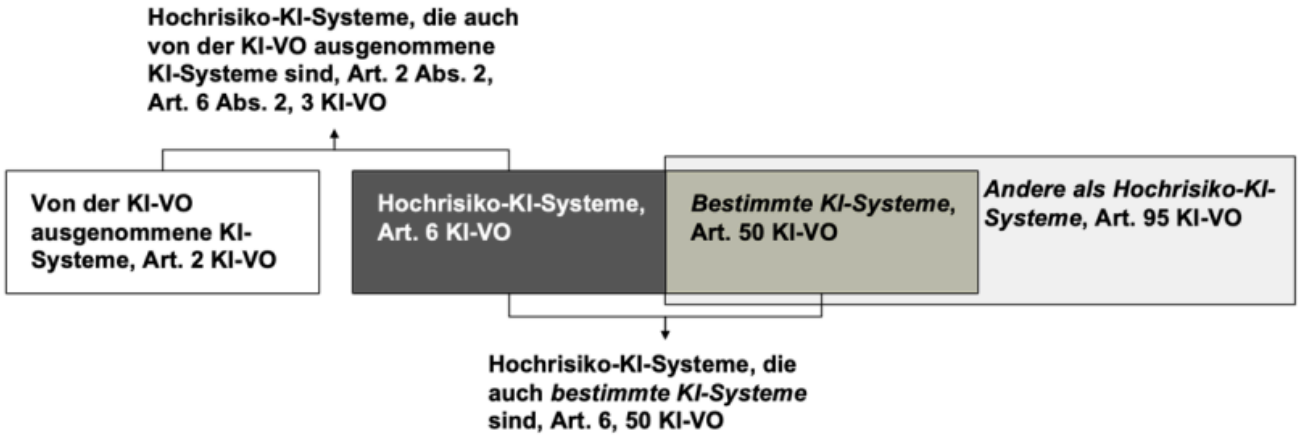


Abb. 3: Übersicht zu den Klassen von KI-Systemen mit Fragen zum Verhältnis untereinander

Spiegel/Höving: Die Klassifizierung von KI-Systemen nach der KI-VO (KIR 2025, 231)

IV. Rezeption in der Praxis: Untersuchung der bisherigen grafischen Systematisierung

Die zuvor dargestellten Klassen sind in der Praxis bisher anders visualisiert worden. Insbesondere der Internetauftritt der EU-Kommission zeigt, dass die grafische Visualisierung noch angepasst werden könnte.

1. Die Visualisierung auf der Webseite der EU-Kommission

Die EU-Kommission bietet auf ihrer Webseite[40] eine Grafik (Abb. 4) an, die KI-Systeme und Risikoklassen wie folgt illustriert:

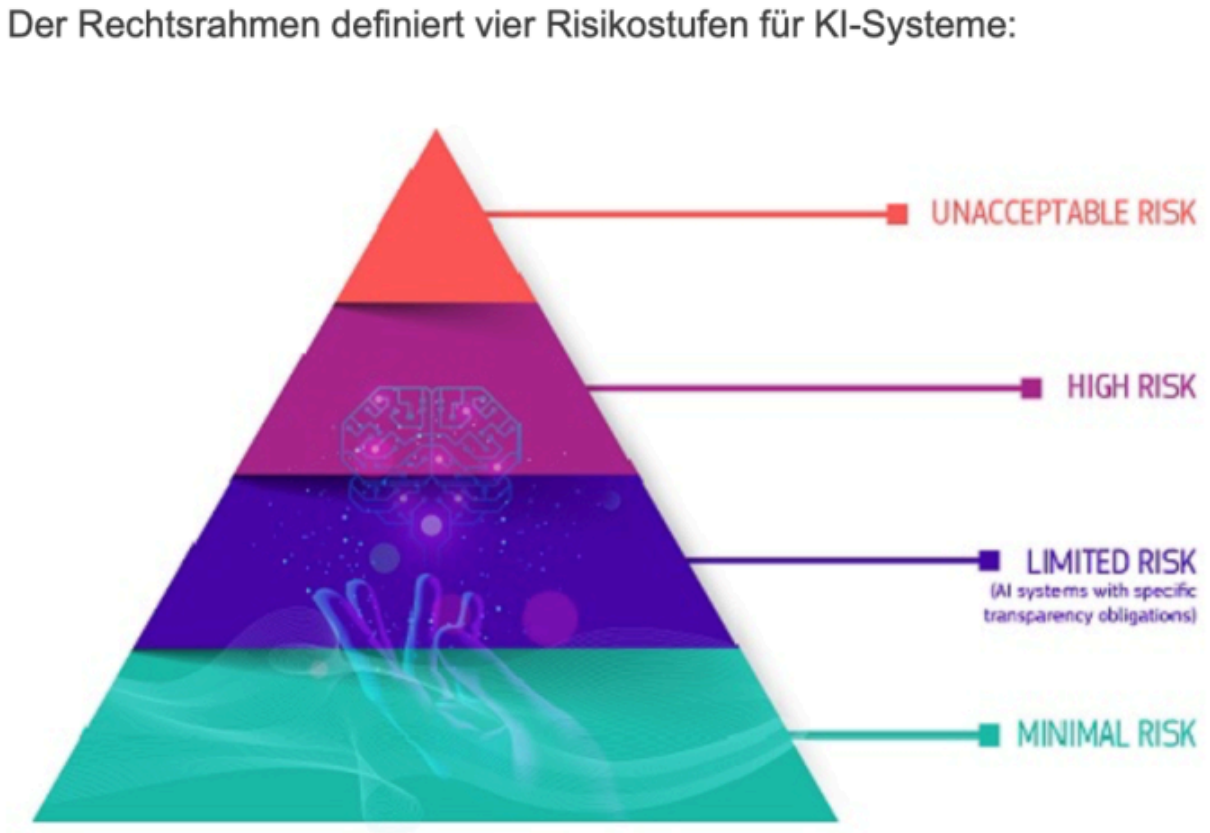


Abb. 4: Vier Risikostufen für KI-Systeme

Zu der Grafik ist anzumerken:

- Zu kritisieren ist, dass ein Hinweis auf das übergreifende Verhältnis zwischen „limited risk“ und „high risk“ fehlt. So sind (vgl. unter III.1.b) die Anforderungen beider Klassen einzuhalten – sofern ein KI-System unter den Tatbestand beider Klassen fällt. Der Umstand, ein KI-System in eine der Klassen einzuordnen, verleitet zu der falschen Annahme, dass entweder (1) alle Voraussetzungen der unteren „Stufen“ einzuhalten sind oder (2) die „Stufen“ unabhängig voneinander sind.

• Die Visualisierung ist unvollständig: Es fällt auf, dass nur bei dem sog. „Limited Risk“, dem begrenzten Risiko, gleichzeitig auch Verpflichtungen genannt sind.

• Die Visualisierung ist möglicherweise irreführend:

(a) Im Hinblick auf Systeme mit „limited risk“: Soweit in einem bestimmten KI-System ein KI-Modell mit allgemeinem Verwendungszweck (Art. 3 Nr. 63 KI-VO) und systemischem Risiko (Art. 3 Nr. 65 KI-VO) enthalten ist, ist die Begriffswahl begrenztes Risiko („limited risk“) missverständlich, weil sich die Begriffe auf den ersten Blick widersprechen.

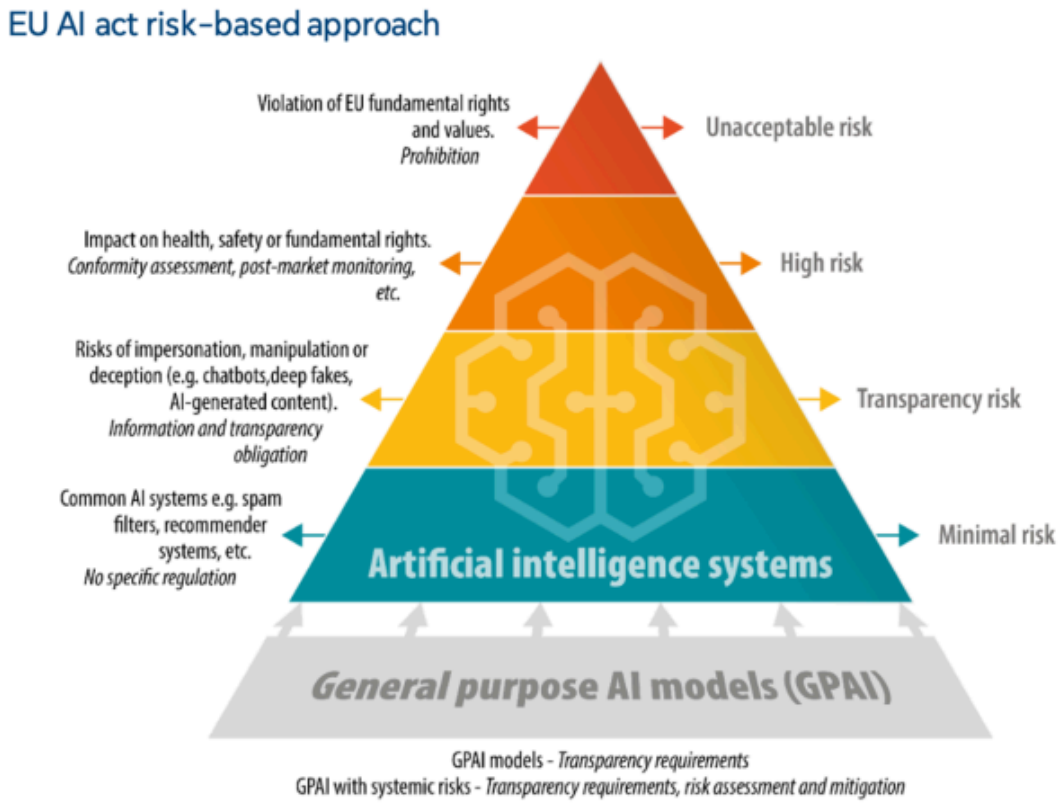
(b) Im Hinblick auf die Risikoklasse „unacceptable risk“: Für Verwirrung sorgt möglicherweise die Aussage, dass die Risikoklasse „unacceptable risk“ für KI-Systeme gelten soll. Denn Art. 5 KI-VO knüpft schon im Titel an Praktiken an.

(c) Unklar ist, welche Aussage die Pyramidenform enthalten soll, denn das Risiko steigt doch in der Übersicht von unten nach oben an.

• Die Visualisierung widerspricht begründungslos ihrer ursprünglichen Vorlage: Der Kommissionsentwurf sah folgenden dreistufigen Ansatz vor: „Die Verordnung verfolgt einen risikobasierten Ansatz, bei dem zwischen Anwendungen von KI unterschieden wird, die ein i) unannehmbares Risiko, ii) ein hohes Risiko und iii) ein geringes oder minimales Risiko darstellen.“[41]

2. Die nachfolgende Veröffentlichung der EU-Kommission vom September 2024 vertieft die Schwächen

Die EU-Kommission hat die vorherige Übersicht inzwischen weiterentwickelt und wie folgt (Abb. 5) ergänzt:[42]



Spiegel/Höving: Die Klassifizierung von KI-Systemen nach der KI-VO (KIR 2025, 231)

Abb. 5: Detailliertere Identifikation

Hierzu ist anzumerken:

• Auf der Stufe „unacceptable risk“ vermischt die Grafik Verletzung („violation“) und Risiko, das gemäß Art. 3 Nr. 2 KI-VO als „die Kombination aus der Wahrscheinlichkeit des Auftretens eines Schadens und der Schwere dieses Schadens“ definiert ist.

• Auf der Stufe des „minimal risk“ fehlt eine Erklärung des Risikos für oder der entsprechenden Auswirkungen auf die Rechte der betroffenen Person.

• Indem KI-Modelle mit allgemeinem Verwendungszweck („GPAI“) breiter als die KI-Systeme dargestellt sind, ist die Beziehung zu den KI-Systemen unklar. Erwägungsgrund 97 KI-VO fordert eine eindeutige Einordnung: „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 … erforderlich. KI-Modelle sind in der Regel in KI-Systeme integriert und Teil davon. …“

• Das Verhältnis von KI-Modellen mit allgemeinem Verwendungszweck und solchen mit systemischem Risiko bleibt offen.

3. Ein Visualisierungsvorschlag der Literatur setzt umfangreiches Hintergrundwissen voraus

Die Literatur hat die Risikopyramide der EU-Kommission aufgegriffen und versuchte schon vor dem Inkrafttreten der KI-VO – basierend auf dem Entwurf der KI-VO –, die Schwächen der Visualisierung durch die EU-Kommission auszugleichen (vgl. Abb. 6).[43]

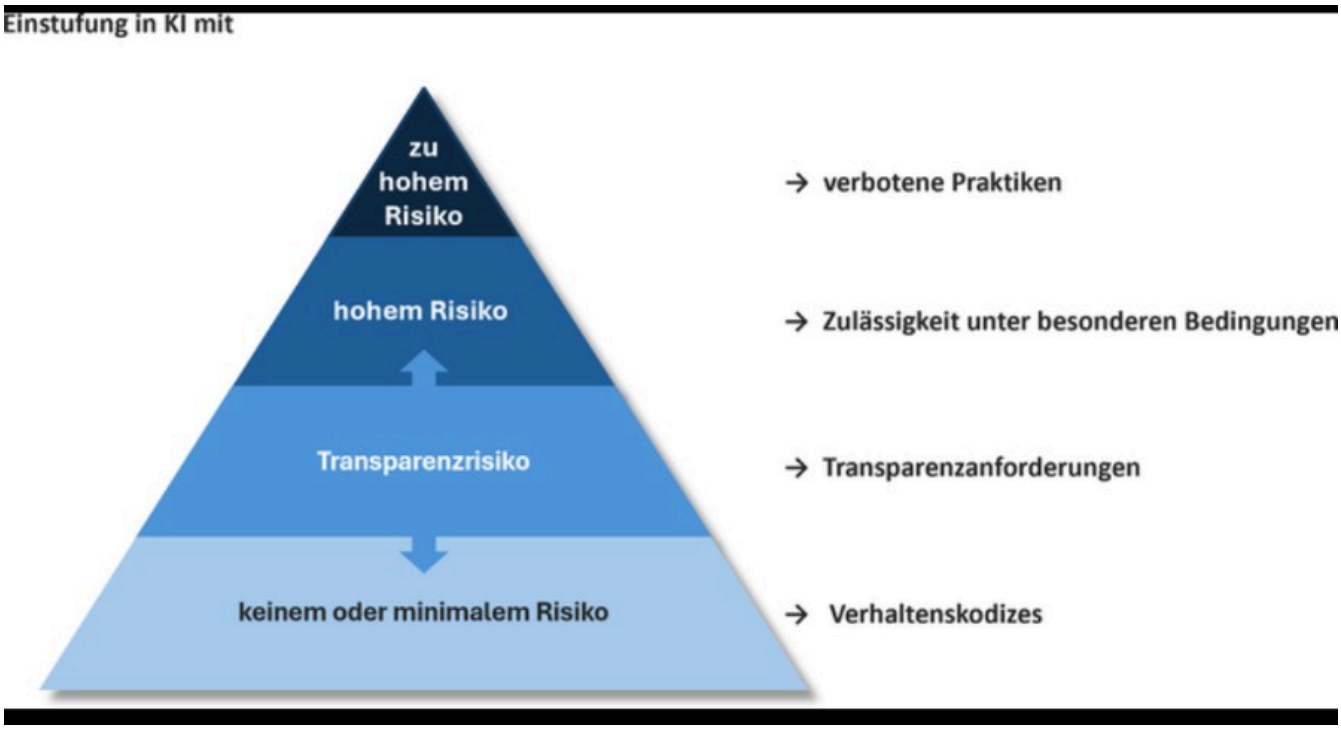


Abb. 6: Aufarbeitung in der Literatur

Spiegel/Höving: Die Klassifizierung von KI-Systemen nach der KI-VO (KIR 2025, 231)

Aus sich heraus deutlich verständlicher ist der Vorschlag einer Risikopyramide von Prof. Dr. Janine Wendt und Prof. Dr. Domenik H. Wendt.[44]. Visuell hilfreich ist, dass sich in der Pyramide die Risikoklassen befinden, während außerhalb die entsprechende Regulierungsmechanismen zugeordnet sind. Zudem deuten die beiden Pfeile im Bereich Transparenzrisiko eine „Durchlässigkeit“ in den Risikoklassen an. Leicht unscharf ist noch die Einleitung, die eine „Einstufung in KI“ vornimmt. Das ist dem Umstand geschuldet, dass das Gesetz uneinheitlich an KI-Systeme bzw. Praktiken anknüpft.

V. Beitrag zur Systematisierung: zwei Vorschläge zur grafischen Darstellung

Um die Darstellung der Sach- und Rechtslage zu vereinfachen, bieten sich zwei präzisere Darstellungsformen an, wobei sie denselben Ansatz – eine Risikomatrix statt einer Risikopyramide – verfolgen und sich möglichst nah am Gesetz orientieren sollen.

1. Präzise Darstellung des risikobasierten Ansatzes anhand der KI-VO

Die nachfolgende Risikomatrix (Abb. 7) soll auf verständliche Art den risikobasierten Ansatz mit den regulatorischen Maßnahmen des EU-Gesetzgebers verbinden und den Gesetzeswortlaut – den risikobasierten Ansatz und die regulatorischen Maßnahmen – exakt umsetzen. Daher sind die unterschiedlichen Regelungsgegenstände (KI-Systeme und KI-Praktiken) in eine Matrix eingeordnet.

Der im Gesetz verankerte risikobasierte Ansatz findet sich auf der „Y-Achse“. Als Vorlage dienen die Beschreibung aus Erwägungsgrund 26 KI-VO[45] und dessen Umsetzung in Art. 1 Abs. 2 KI-VO: „In dieser Verordnung wird Folgendes festgelegt: … b) Verbote bestimmter Praktiken im KI-Bereich; c) besondere Anforderungen an Hochrisiko-KI-Systeme

und Pflichten für Akteure in Bezug auf solche Systeme; d) harmonisierte Transparenzvorschriften für bestimmte KI-Systeme … .“ Festzuhalten ist, dass die KI-VO diese Klassifizierung von Risiken an einer Stelle aufbricht, vgl. Erwägungsgrund 51 KI-VO: „für Produkte mit mittlerem und hohem Risiko“. Da an keiner Stelle der KI-VO ein weiterer Bezug zu mittlerem Risiko auftritt, ist ein redaktionelles Versehen wahrscheinlich.

Auf der „X-Achse“ ist die Intensität der regulatorischen Maßnahmen aufgereiht. Die Übersicht verzichtet bewusst darauf, Verhaltenskodizes zu nennen: Der risikobasierte Ansatz schweigt sich über eine weitere (vierte) Stufe aus[46], außerdem sind die Verhaltenskodizes freiwillig.

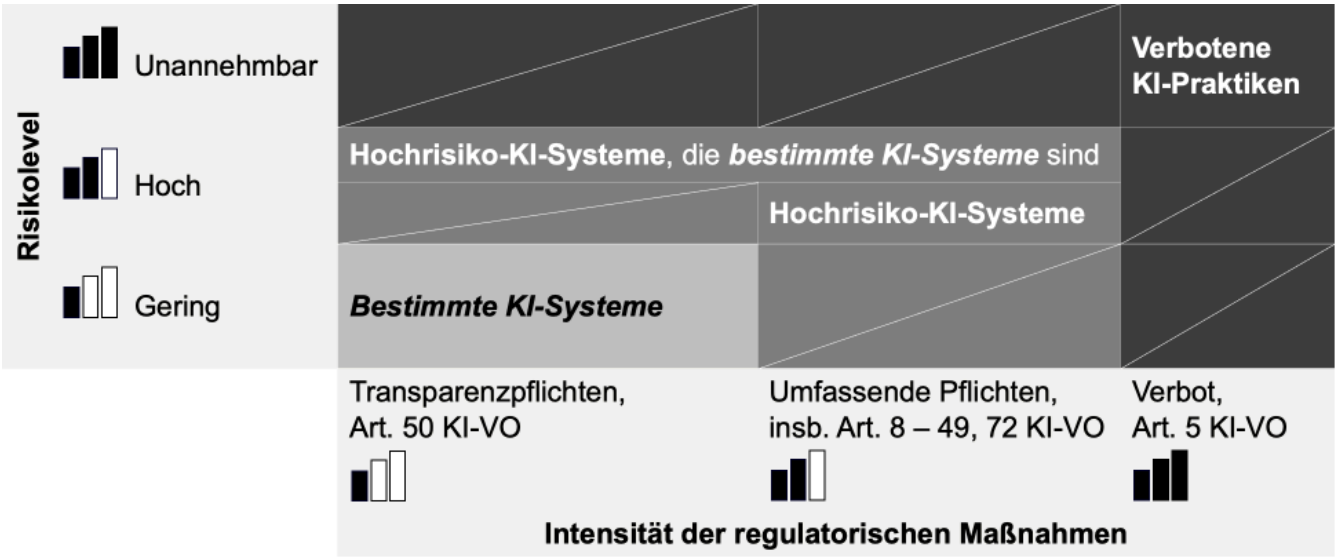


Abb. 7: Vorschlag einer Risikomatrix

In der Matrix sind die Regelungsgegenstände anhand ihres „Risikolevels“ und anhand der korrespondierenden regulatorisch angeordneten Maßnahme eingetragen. Das Inverkehrbringen von Hochrisiko-KI-Systemen, die auch bestimmte KI-Systeme sind, setzt gem. Art. 50 Abs. 6 KI-VO sowohl die Einhaltung der Transparenzpflichten als auch der Pflichten in Bezug auf Hochrisiko-KI-Systeme voraus. Aufgrund der Häufung der Risiken erfolgt eine Einordnung über der Einordnung von bloßen Hochrisiko-KI-Systemen – dies ist allerdings nicht zwingend.[47]

2. Praxisorientierte, vollständige Darstellung der Regelungsgegenstände in der Risikomatrix

In der Praxis dürften „Andere als Hochrisiko-KI-Systeme“ durchaus relevant sein, obwohl sie regulatorisch nur mit freiwilligen Maßnahmen bedacht sind. Sowohl die EU-Kommission als auch die juristische Literatur[48] greifen daher vier Risikoklassen auf. Die folgende Übersicht (Abb. 8) ordnet daher noch die „Anderen als Hochrisiko-KI-Systeme“ in die Risikomatrix ein:

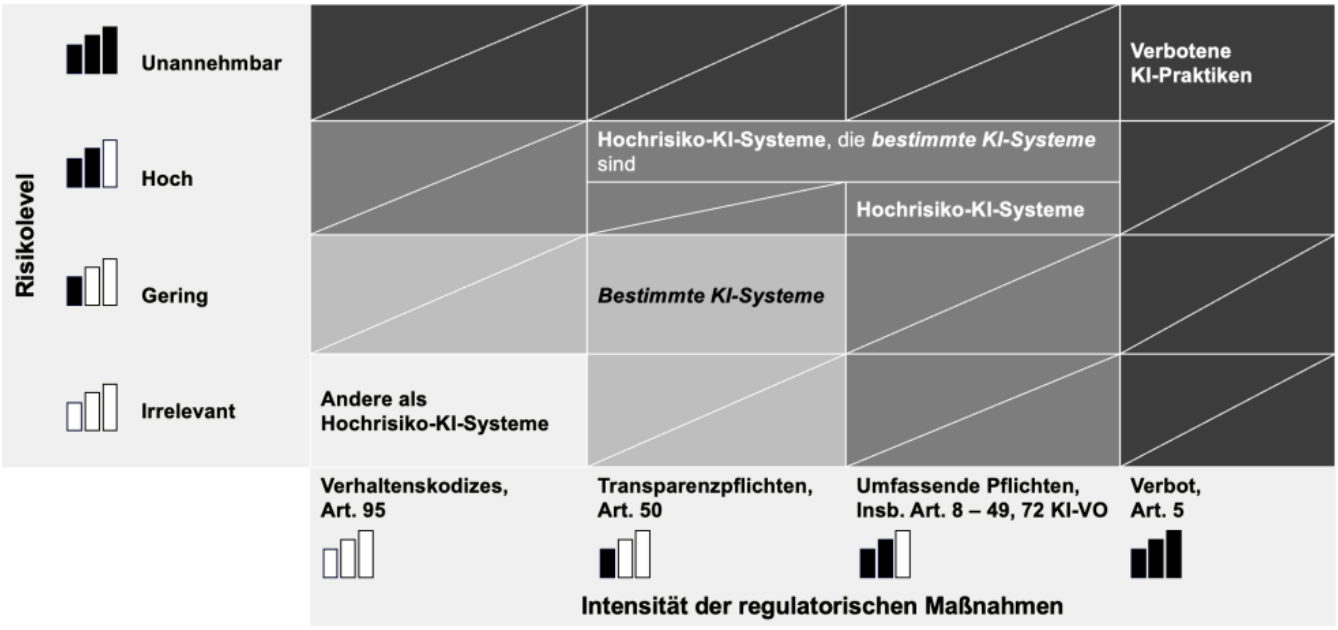


Abb. 8: Praxisorientierte Risikomatrix

Umgesetzt in der Risikomatrix sind die Verhaltenskodizes im Bereich „Intensität der regulatorischen Maßnahmen“ hinzuzufügen. Im Bereich „Risikolevel“ ist mit „irrelevant“ ein weiterer Begriff einzuführen, den das Gesetz bisher frei lässt und der ein niedrigeres Risiko als gering festlegt. Die Formulierung „kein Risiko“ ist abzulehnen, denn ohne Risiko wäre keine Maßnahme erforderlich.

VI. Ausblick

Der in diesem Beitrag vorgenommene grafische Systematisierungsvorschlag ist ein Angebot auf Grundlage der KI-VO, wie sie und ihre Rezeption in der Praxis aktuell „steht und liegt“. Das Ende der interpretativen Fahnenstange ist bei weitem noch nicht erreicht: Allein während der Entstehung dieses Beitrags hat die EU-Kommission[49] hunderte neue Seiten an Interpretationsmaterial veröffentlicht. Es bleibt zu hoffen, dass der an sich lobenswerte Ansatz, für KI-Systeme eine klassenspezifische – und dadurch angemessene, weil nach Risiken abgestufte – Regulierung zu finden, für die Praxis handhabbar gemacht wird.

Dr. Ulrich Spiegel

ist Senior Associate in der Rechtsanwaltskanzlei TaylorWessing.

Maximilian Höving

ist Legal Counsel Special Law – Technical Regulation & Standardization bei der Siemens AG.

Schnell gelesen…

Die Klassifizierung von KI-Systemen ist eine entscheidende Weiche in der KI-VO, um die gesetzlich angeordneten Anforderungen an KI-Systeme zu bestimmen.

Bisherige Grafiken der Risikoklassen der KI-VO können ergänzt werden. Zur präziseren Darstellung empfiehlt sich – getreu des Gesetzestextes der KI-VO – eine matrixbasierte Darstellung in drei Risikoklassen.

Sollen auch KI-Systeme abgebildet werden, denen die KI-VO keine unmittelbaren Anforderungen oder gar Verbote auferlegt, lässt sich die Matrix um eine weitere Klasse erweitern.

Neben der weiteren Präzisierung bietet die matrixbasierte Darstellung „vertrautes Terrain“ für den Produktsicherheitsrechtler durch die Verknüpfung von Risko, Klasse des KI-Systems und darauf basierenden regulatorischen Anforderungen oder gar Verbote.

---

1 Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates v. 13.6.2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG) Nr. 300/2008, (EU) Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828.

2 Staub, Wenn Technik tötet, LTO v. 4.9.2019, abrufbar unter: https://www.lto.de/recht/hintergruende/h/automatisiertes-fahren-selbstfahrend-auto-strafrecht-fahrzeugfuehrer-chain-of-supply; eine mögliche zivilrechtliche Betrachtung eines ähnlichen Sachverhalts findet sich in Thöne/Kellner JA 2020, 253.

3 Vgl. Art. 3 Abs. 60 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“.

4

Martin-Jung, Hier spricht der Ferrari-Chef – nicht, Süddeutsche Zeitung v. 29.7.2024, abrufbar unter: https://www.sueddeutsche.de/wirtschaft/deepfakes-ceo-fraud-betrug-ferrari-lux.D6i4WqmRwWhL9iT8spKGrG.

5 Bitkom, Presseinformation v. 24.7.2023, abrufbar unter: https://www.bitkom.org/Presse/Presseinformation/63-Prozent-haben-Angst-vor-Deepfakes.

6 Vgl. Art. 6 Abs.1 lit. g, h VO (EU) 988/2023 und VO (EU) 1230/2023, eingehend zur erstmals umfassenden produktsicherheitsrechtlichen Regulierung von Online-Marktplätzen etwa Spiegel ZVertriebsR 2023, 71.

7 Zur KI-VO als Teil des Produktsicherheitsrecht zB Rohrßen ZfPC 2024, 111.

8 Vgl. Erwägungsgrund 9 KI-VO.

9 Vgl. https://www.oecd-ilibrary.org/docserver/623da898-en.pdf?expires=1728741969&id=id&accname=guest&checksum=BB9EB880889487F5D10BC5BF59DB40E2 und die Leitlinien zur Definition von KI-Systemen der EU-Kommission https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application.

10 S. OECD Publishing, OECD Framework for the Classification of AI systems, OECD digital economy papers, February 2022 No. 323, S. 67, abrufbar unter: https://www.oecd-ilibrary.org/docserver/cb6d9eca-en.pdf?expires=1728756080&id=id&accname=guest&checksum=6C361269B9B0971942C8ED1BD701A242.

11 S. OECD Publishing, OECD Framework for the Classification of AI systems, OECD digital economy papers, February 2022 No. 323, S. 67, abrufbar unter: https://www.oecd-ilibrary.org/docserver/cb6d9eca-en.pdf?expires=1728756080&id=id&accname=guest&checksum=6C361269B9B0971942C8ED1BD701A242.

12 S. OECD Publishing, OECD Framework for the Classification of AI systems, OECD digital economy papers, February 2022 No. 323, S. 67, abrufbar unter: https://www.oecd-ilibrary.org/docserver/cb6d9eca-en.pdf?expires=1728756080&id=id&accname=guest&checksum=6C361269B9B0971942C8ED1BD701A242.

13 S. OECD Publishing, OECD Framework for the Classification of AI systems, OECD digital economy papers, February 2022 No. 323, S. 67, abrufbar unter: https://www.oecd-ilibrary.org/docserver/cb6d9eca-en.pdf?expires=1728756080&id=id&accname=guest&checksum=6C361269B9B0971942C8ED1BD701A242.

14 EU-Kommission, KI-Gesetz, abrufbar unter: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai sowie die Fragen und Antworten der EU-Kommission.

15 Anders als das althergebrachte Produktrecht spricht die KI-VO grundsätzlich von Akteuren, nicht von Wirtschaftsakteuren, vgl. Art. 3 Nr. 8 KI-VO.

16 Roth-Isigkeit MMR 2024, 621 (622).

17 Martini/Wendehorst, KI-VO/Wendehorst, 2024, Art. 5 Rn. 1 f.

18 Martini/Wendehorst, KI-VO/Wendehorst, 2024, Art. 5 Rn. 14.

19 Martini/Wendehorst, KI-VO/Wendehorst, 2024, Art. 5 Rn. 192.

20 Roth-Isigkeit MMR 2024, 621 (623) und Ebers/Streitbörger RDi 2024, 393.

21

Vgl. aber Art. 25 Abs. 1 lit. c KI-VO.

22 Inwieweit der EuGH die Wertungen der Rechtsprechung des BGH zum mittlerweile außer Kraft getretenen § 3 Abs. 1 MPG übernimmt – wie von Ebers/Streitbörger in RDi 2024, 393 (395) vorgeschlagen – bleibt abzuwarten. Grenze der Zweckbestimmungsfreiheit des Anbieters wäre dann die Willkür.

23 Vgl. § 2 Nr. 5 lit. a ProdSG.

24 Der Begriff Zweckbestimmung wird nur selten iRd NLF definiert. Ausnahmen etwa im Medizinproduktrecht, zB Art. 2 Nr. 12 VO (EU) 2017/745.

25 Ebers/Streitbörger RDi 2024, 393 (394).

26 „NLF", bestehend aus VO (EG) Nr. 765/2008, Beschluss Nr. 768/2008/EG und der VO (EU) 2019/1020.

27 Anhang I Teil A und B KI-VO.

28 Martini/Wendehorst, KI-VO/Ruschemeier, 2024, Art. 6 Rn. 85 ff.

29 Fragend: Ebers/Streitbörger RDi 2024, 393 (395), dagegen: Hilgendorf/Roth-Isigkeit/Martini, Die neue Verordnung der EU zur Künstlichen Intelligenz, HdB, 2023, § 4 Rn. 50, offen: Martini/Wendehorst, KI-VO/Ruschemeier, 2024, Art. 6 Rn. 84.

30 Es wäre auch möglich, mit Art. 6 Abs. 2 KI-VO zu beginnen. Im Hinblick auf die leichtere Darstellungsweise der Bereichsausnahmen nach Art. 2 Abs. 2 KI-VO iVm Anhang I Abschnitt B KI-VO bietet sich hier die vorliegende Reihenfolge an.

31 Vgl. Art. 2 Abs. 2 KI-VO.

32 Erwägungsgrund 49 KI-VO.

33 Erwägungsgrund 9 KI-VO.

34 Gerdemann NJW 2024, 2209 (2212).

35 S. Art. 102-109 KI-VO.

36 Martini/Wendehorst, KI-VO/Wendehorst, 2024, Art. 2 Rn. 46 f.

37 Vgl. Erwägungsgrund 132 KI-VO.

38 Vgl. Martini/Wendehorst, KI-VO/Hartmann, 2024, Art. 95 Rn. 7.

39 Q&A der EU-Kommission zur Produktsicherheitsverordnung, abrufbar unter: https://webgate.ec.europa.eu/safety/consumers/consumers_safety_gate/obligationsForBusinesses/documents/Q&A.pdf Versuche, „stand alone" Software in die Definition „Produkt" aufzunehmen, sind gescheitert, vgl. dazu Schucht/Wiebe, HK-GPSR/Wilrich, 2025; EU-Produktsicherheits-VO, Art. 3 Rn. 10 f.

40 Abrufbar unter https://digital-strategy.ec.europa.eu/de/policies/regulatory-framework-ai.

41

Abschnitt 5.2.2 des Vorschlags für eine VO des europäischen Parlaments und des Rates v. 21.4.2021 zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz (Gesetz über Künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union, COM/2021/206 final.

42 Vgl. EU-Parlament, Artificial intelligence act, Briefing v. 2.9.2024, abrufbar unter : https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf.

43 Aus Bomhard/Siglmüller RDi 2024, 45.

44 Wendt/Wendt, Das neue KI-Recht, 2024, § 4 Rn. 1.

45 S. Erwägungsgrund 26 KI-VO: „Es ist daher notwendig, bestimmte inakzeptable Praktiken im Bereich der KI zu verbieten und Anforderungen an Hochrisiko-KI-Systeme und Pflichten für die betreffenden Akteure sowie Transparenzpflichten für bestimmte KI-Systeme festzulegen."

46 Rohrßen ZfPC 2025, 6 (15).

47 Vgl. die Hinweise zur Risikobewertung im Hinblick auf eine Mehrheit von Risiken im Durchführungsbeschluss (EU) 2019/417 der Kommission v. 8.11.2018 zur Festlegung von Leitlinien für die Verwaltung des gemeinschaftlichen Systems zum raschen Informationsaustausch „RAPEX" gem. Artikel 12 der Richtlinie 2001/95/EG über die allgemeine Produktsicherheit und für das dazugehörige Meldesystem (Bekannt gegeben unter Aktenzeichen C(2018) 7334), Anhang I Teil 3 Abschnitt 2.3 „Ist eine Kumulierung von Risiken möglich?": „Risiken werden also nicht einfach kumuliert. Wenn jedoch mehrere relevante Risiken bestehen, muss gegebenenfalls schneller und mit entschiedeneren Maßnahmen gegen die Risiken vorgegangen werden."

48 Vgl. zB Martini/Wendehorst, KI-VO/Martini, 2024, Art. 50 Rn. 9.

49 Vgl. etwa EU-Kommission, Commission publishes the Guidelines on prohibited artificial intelligence (AI) practices, as defined by the AI Act, Pressemitteilung v. 4.2.2025, abrufbar unter: https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act; EU-Kommission, The Commission publishes guidelines on AI system definition to facilitate the first AI Act's rules application, Pressemitteilung v. 6.2.2025, abrufbar unter: https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application; EU-Kommission, Third Draft of the General-Purpose AI Code of Practice published, written by independent experts, Pressemitteilung v. 11.3.2025, abrufbar unter: https://digital-strategy.ec.europa.eu/en/library/third-draft-general-purpose-ai-code-practice-published-written-independent-experts.