Die neue Verordnung der EU zur Künstlichen Intelligenz

Hilgendorf / Roth-Isigkeit

2. Auflage 2025
ISBN 978-3-406-82760-0
C.H.BECK

umfassendes Spezialsortiment im Bereich Recht, Steuern und Wirtschaft mit rund 700.000 lieferbaren Fachbuchtiteln.

technischen Beschränkungen einer Weiterverwendung sowie Sicherheits- und Datenschutzmaßnahmen, einschließlich Pseudonymisierung, unterliegen müssen und sichergestellt werden muss, dass die Daten gesichert, geschützt und Gegenstand angemessener Sicherheitsvorkehrungen sind. Zu solchen Vorkehrungen dürfte der Einsatz von starker Kryptographie ebenso wie andere technische Maßnahmen gehören. Organisatorische Maßnahmen wie Zugangsbeschränkungen werden damit nicht ausgeschlossen, sondern sind implizit mit umfasst.

Der Rechtsfertigungstatbestand des Art. 10 Abs. 5 KI-VO beschränkt sich auf die gem. Art. 16 Buchst. a KI-VO zur Sicherstellung der Anforderungskonformität verpflichteten Anbieter. **Für Betreiber** von KI-Systemen bleibt es hingegen bei den engen Voraussetzungen des Art. 9 DS-GVO, insbes. der ausdrücklichen Einwilligung, wenn besonders sensible Daten verarbeitet werden sollen. 39

VII. Anforderungen für KI-Systeme ohne Trainingsdaten

Da der Begriff des KI-Systems, von dem der KI-VO ausgeht, außerordentlich breit gefasst ist – wenngleich auch im Gesetzgebungsprozess sehr umstritten[60] – und nicht nur Systeme des **maschinellen Lernens** umfasst, ist Art. 10 Abs. 6 KI-VO eine konsequente Ausdehnung der Anforderungen an die Data-Governance auch auf solche Systeme, die kein Training mit Datensätzen voraussetzen. Letztlich bleibt es bei den Anforderungen des Art. 10 KI-VO, mit dem Unterschied, dass Trainingsdatensätze nicht erfasst werden – alle anderen Kriterien bzw. Anforderungen bleiben erhalten, insbes. hinsichtlich der Erkennung und Vermeidung von Verzerrungen von Daten. Im Einzelfall kann es schwierig sein, die Modifizierungen zu bestimmen, da dies stark von dem jeweiligen KI-System, und wieweit es auf Datensätze zurückgreift, abhängt. Die grundsätzlichen Kriterien aber, allen voran die Eignung, Repräsentativität, Fehlerfreiheit und Vollständigkeit, bleiben auch den nicht-maschinell lernenden Systeme erhalten. 40

C. Technische Dokumentation (Art. 11 KI-VO)

Voraussetzung für externe Konformitätsbewertungsverfahren sind aussagekräftige Dokumentationen, wie sie von Art. 11 Abs. 2 KI-VO verlangt werden. Aber auch solche Systeme, die gem. Art. 43 Abs. 1–3 KI-VO bislang nur interne Kontrollen durchführen müssen und damit extern *„nur"* der Überwachung durch die (Markt-)Aufsichtsbehörden unterliegen, müssen Dokumentationen mit dem **Mindestinhalt nach Anhang IV** KI-VO aufweisen. Nur mit diesen Informationen lassen sich Zertifizierungsverfahren durchführen[61] ebenso wie spätere Kontrollen durch Marktüberwachungsbehörden. Auf Basis des Kompromissvorschlags der tschechischen Ratspräsidentschaft[62] wurde in Art. 11 Abs. 1 UAbs. 2 S. 3–6 KI-VO allerdings eine **Erleichterung für KMU, inklusive Start-Ups,** eingefügt. Diese können die Dokumentationspflichten auch auf Basis eines von der Kommission bereitzustellenden, vereinfachten Formulars erfüllen, das auf die Bedürfnisse dieser Unternehmen zugeschnitten sein soll. 41

[60] So wird etwa in Art. 3 Abs. 1 Ratsdok. 11124/22 ein engeres Verständnis des KI-Systems vorgeschlagen: *„'artificial intelligence system' (AI system) means a system that is designed to operate with a certain level of autonomy and that, based on machine and/or human-provided data and inputs, infers how to achieve a given set of human-defined objectives using machine learning and/or logic- and knowledge based approaches, and produces system-generated outputs such as content (generative AI systems), predictions, recommendations or decisions, influencing the environments with which the AI system interacts".*

[61] Zu deren hohen Stellenwert iRd KI-VO Rostalski/Weiss ZdDR 2021, 329 (354 ff.); Martini/Wendehorst/Braun Binder/Egli KI-VO Art. 11 Rn. 4.

[62] Art. 11 Abs. 1 S. 3 Ratsdok. 11124/22.

I. Zweck der Dokumentation

42 Die technische Dokumentation dient als **Nachweis** dafür, wie das Hochrisiko-KI-System die Anforderungen der Art. 8 ff. KI-VO erfüllt, insbes. damit die jeweiligen Überwachungsstellen und notifizierten Stellen ihre Aufgaben wahrnehmen können, Art. 11 Abs. 1 UAbs. 2 S. 1 KI-VO. Die Dokumentation muss erstellt werden, **bevor** die jeweiligen Systeme in den Markt gebracht oder in Betrieb genommen werden. Ferner müssen sie stets aktualisiert werden.

II. Integration in produktsicherheitsrechtliche Dokumentationsvorgaben

43 Art. 11 Abs. 2 KI-VO ist quasi die „Scharnier"-Vorschrift zu den von dem KI-VO ebenfalls indirekt geregelten Rechtsakten, die den Konformitätsbewertungsverfahren unterfallen. Demnach wird nur eine einzige technische Dokumentation erstellt, die aber alle in Anhang IV KI-VO verlangten Informationen enthalten muss, so dass dem beabsichtigten Gleichlauf der horizontalen Regulierung Rechnung getragen wird (→ Rn. 4).

III. Detaillierte Beschreibungen der Funktionsweisen des KI-Systems und seiner Komponenten

44 Die technische Dokumentation reflektiert die **Anforderungen** des zweiten Kapitels der KI-VO. Nach Anhang IV Nr. 1 KI-VO muss das KI-System umfassend beschrieben werden, etwa die Zweckbestimmung, die Interaktion von KI-System mit Hardware oder anderer Software, die jeweiligen Versionen, die Hardware, auf der das KI-System betrieben werden soll, im Falle von embedded KI-Systemen Fotografien oder Abbildungen der Produkte einschließlich ihres inneren Aufbaus bis hin zu den Gebrauchs- und Installationsanweisungen für die Betreiber. Hierzu zählt auch die nach Anhang IV Nr. 5 KI-VO erforderliche *„[D]etaillierte Beschreibung des **Risikomanagementsystems** gemäß Artikel 9"*. Denn das Risikomanagementsystem nach Art. 9 KI-VO stellt mit seiner Klammerwirkung und Auffangfunktion den Grundbaustein der an hochriskanten KI-Systeme gestellten Anforderungen dar (→ Rn. 12).

45 Darüber hinaus verlangt Anhang IV Nr. 2 KI-VO eine detaillierte Beschreibung aller **Bestandteile** des KI-Systems *„und seines Entwicklungsprozesses",* was weiter aufgefächert wird – allerdings wieder unter etlichen Redundanzen –, etwa hinsichtlich des *„Einsatzes von Dritten bereitgestellter vortrainierter Systeme oder Werkzeuge, und wie diese vom Anbieter benutzt, integriert oder verändert wurden"* (Anhang IV Nr. 2 Buchst. a KI-VO); ferner den **Spezifikationen** des Systems, insbes. der allgemeinen Logik des KI-Systems und der eingesetzten Algorithmen, dazu die *„wichtigste[n] Entwurfsentscheidungen mit den Gründen und Annahmen, auch in Bezug auf Personen oder Personengruppen, auf die das System angewandt werden soll"*. Auch soll detailliert beschrieben werden, *„was das System optimieren soll und welche Bedeutung den verschiedenen Parametern dabei zukommt"* (Anhang IV Nr. 2 Buchst. b KI-VO). Wiederum überlappend mit anderen Anforderungen schreibt Anhang IV Nr. 2 Buchst. c KI-VO eine Beschreibung der **Systemarchitektur** vor, wie die Softwarekomponenten miteinander agieren. Neu dagegen ist das Verlangen der Angabe von ***„Rechenressourcen",*** gem. Anhang IV Nr. 2 Buchst. c KI-VO, die *„zum Entwickeln, Trainieren, Testen und Validieren des KI-Systems"* verwandt wurden – daraus lassen sich mittelbar Rückschlüsse auf den Verwendungsumfang von Datensätzen für das KI-System herleiten.

46 Das besondere Augenmerk, dass die KI-VO auf die **Qualität der Datensätze** legt (Art. 10 KI-VO), schlägt sich auch in den Anforderungen an die technische Dokumentation in Anhang IV Nr. 2 Buchst. d KI-VO nieder, der eine detaillierte Beschreibung von „Datenanforderungen in Form von Datenblättern, in denen die Trainingsmethoden und -techniken und die verwendeten Trainingsdatensätze beschrieben werden, mit Angaben zu Herkunft, Umfang und Hauptmerkmalen dieser Datensätze" verlangt. Ferner müssen

„Angaben zur Beschaffung und Auswahl der Daten“, ebenso wie zu *„Kennzeichnungsverfahren (zB für überwachtes Lernen)“* und *„Datenbereinigungsmethoden (zB Erkennung von Ausreißern)“* gemacht werden. Alle diese Anforderungen dienen letztlich der **Unterstützung** der materiell-rechtlichen Anforderungen von Art. 10 Abs. 2, Abs. 3 KI-VO hinsichtlich der **Vollständigkeit und Fehlerfreiheit der Datensätze.**

Aber auch die **Testverfahren,** die gerade für KI-Systeme besonders bedeutsam sind, werden in der technischen Dokumentation nach Anhang IV Nr. 2 Buchst. g KI-VO ausdrücklich adressiert. Demnach müssen die *„verwendete[n] Validierungs- und Testverfahren [angegeben werden], mit Angaben zu den verwendeten Validierungs- und Testdaten und deren Hauptmerkmalen“.* Ferner gehören zur Beschreibung *„Parameter, die zur Messung der Genauigkeit, Robustheit, Cybersicherheit und der Erfüllung anderer einschlägiger Anforderungen nach Titel III Kapitel 2 sowie potenziell diskriminierender Auswirkungen verwendet werden“.* Schließlich müssen auch die *„Testprotokolle und alle von den verantwortlichen Personen datierten und unterzeichneten Testberichte“* dokumentiert werden. **47**

Die technische Dokumentation muss zudem detailliert die Bewertung der Verfahren der menschlichen Aufsicht nach Art. 14 KI-VO umfassen, Anhang IV Nr. 2 Buchst. e KI-VO, ebenso die technischen Maßnahmen nach Art. 13 Abs. 3 Buchst. d KI-VO, um den Betreibern *„die Interpretation der Ergebnisse von KI-Systemen zu erleichtern“* (Einzelheiten zur menschlichen Aufsicht: → § 6 Rn. 23). **48**

Schließlich adressiert Anhang IV Nr. 2 Buchst. f KI-VO die „***vorab bestimmten Änderungen** an dem KI-System*“, indem verlangt wird, dass diese detailliert beschrieben werden, einschließlich der *„einschlägigen Angaben zu den technischen Lösungen, mit denen sichergestellt wird, dass das KI-System die einschlägigen Anforderungen nach Titel III Kapitel 2 KI-VO weiterhin dauerhaft erfüllt“.* Anhang IV Nr. 2 Buchst. g KI-VO bezieht auch die insoweit einschlägigen Testprotokolle und -berichte in die Dokumentationspflicht ein. Die darin enthaltenen Informationen können nicht nur für notifizierte Stellen und Überwachungsbehörden von Interesse sein, sondern auch für die Betreiber des KI-Systems, soweit diese einschätzen möchte, ob von ihnen geplante Verwendungsarten als wesentliche Veränderung zu qualifizieren sind und sie daher ggf. Gefahr laufen, in die Pflichtenposition des Anbieters einzutreten (→ Rn. 26). **49**

IV. Detaillierte Informationen über die Überwachung, Funktionsweise und Kontrolle des KI-Systems

Während Anhang IV Nr. 1, 2 KI-VO den Schwerpunkt auf die Beschreibung der Entwicklung und des Aufbaus des KI-Systems legen, wendet Anhang IV Nr. 3 KI-VO sich mehr der fortlaufenden Überwachung und Kontrolle bzw. Funktionsweise des KI-Systems zu. Demnach müssen die „***Fähigkeiten und Leistungsgrenzen,** mit dem Genauigkeitsgrad für bestimmte Personen oder Personengruppen, auf die das System angewandt werden soll*“ detailliert beschrieben werden, ebenso der „*insgesamt erwartete Genauigkeitsgrad in Bezug auf seine **Zweckbestimmung***“. Zu den Angaben zählen ferner vorhersehbare unbeabsichtigte **Ergebnisse und Quellen von Risiken** für die Gesundheit und Sicherheit sowie Grundrechte und Diskriminierung, zudem Beschreibungen der menschlichen Aufsicht und der Maßnahmen, um Betreibern die Interpretation der Ergebnisse von KI-Systemen zu erleichtern – offenbar aber auf die laufende Überwachung bezogen, da schon Anhang IV Nr. 2 Buchst. e KI-VO Angaben hierzu verlangt. **50**

Flankierend tritt die nach Anhang IV Nr. 9 KI-VO erforderliche „*[D]etaillierte **Beschreibung des Systems** zur Bewertung der Leistung des KI-Systems in der Phase nach dem Inverkehrbringen gemäß Artikel 72, mit dem in Artikel 72 Absatz 3 genannten Plan für die Beobachtung nach dem Inverkehrbringen*“ hinzu. Denn hier handelt es sich um die Produktbeobachtung des KI-Systems nach dessen Inverkehrbringung, mit deren Hilfe das Phänomen der unvorhersehbaren Entwicklung und daraus resultierender Risiken bewältigt werden soll. **51**

V. Dokumentation der angewandten harmonisierten Normen oder Lösungen

52 Der technischen Dokumentation müssen ferner nach Anhang IV Nr. 7 KI-VO alle angewandten harmonisierten Normen einschließlich deren Fundstellen im Amtsblatt der EU angehören (→ Rn. 5) und falls diese nicht einschlägig sind, die *„detaillierte Beschreibung der Lösungen, mit denen die Anforderungen in Titel III Kapitel 2 erfüllt werden sollen, mit einer Aufstellung anderer einschlägiger Normen und technischer Spezifikationen"*. Letzteres reflektiert zum einen den bisherigen Mangel an einschlägigen technischen Normungen und die vorerst erforderlichen eigenständigen technischen Lösungen, zum anderen aber auch die erheblichen praktischen Anreize zur freiwilligen Befolgung harmonisierter Normen, für die gerade keine *„detaillierte Beschreibung"* erforderlich ist.[63]

53 Schließlich muss nach Anhang IV Nr. 8 KI-VO auch eine Kopie der EU-Konformitätserklärung beigefügt werden, was aber ersichtlich nur für diejenigen KI-Systeme gelten kann, für die ein solches Konformitätsbewertungsverfahren durchgeführt wurde, mithin produktsicherheitsbezogene KI-Systeme (→ Rn. 1).

VI. Änderungen des Anhangs IV KI-VO

54 Nach Art. 11 Abs. 3 KI-VO wird der Kommission die Möglichkeit eingeräumt, im Wege eines delegierten Rechtsaktes den Anhang IV KI-VO zu verändern, wenn der technische Fortschritt dies erfordert im Hinblick auf die Einhaltung der Anforderungen des zweiten Kapitels durch das KI-System. Hierbei handelt es sich um eines von vielen Beispielen einer inhaltlichen Flexibilisierung der KI-VO mittels Kompetenzübertragung an die Exekutive, womit der Gesetzgeber auf die rasante technologische Entwicklung zu reagieren versucht.

D. Robustheit, Genauigkeit und IT-Sicherheitsanforderungen (Art. 15 KI-VO)

I. Überblick

55 Die KI-VO verlangt hinsichtlich der Konzeption und Entwicklung eine ausreichende Sicherheit und Genauigkeit der KI-Systeme, **einschließlich der Cybersicherheit,** auch während der **gesamten Lebensdauer** des KI-Systems, Art. 15 Abs. 1 KI-VO. Aber auch ein angemessenes Maß an **Robustheit** und **Genauigkeit** im Hinblick auf ihre Zweckbestimmung wird von Art. 15 Abs. 1 KI-VO eingefordert – wobei die Frage, was als *„angemessen"* gelten kann, von der KI-VO offengelassen wird und daher nur im Einzelfall im Hinblick auf Kosten-Nutzen und -Risiko-Verhältnis bestimmt werden kann.[64]

II. Genauigkeitsgrade

56 Art. 15 Abs. 3 KI-VO setzt sich zwar mit den Genauigkeitsgraden der KI-Systeme auseinander, verlangt aber nur, dass die *„Genauigkeitsgrade und die relevanten Genauigkeitskennzahlen von Hochrisiko-KI-Systemen […] in der ihnen beigefügten Gebrauchsanweisung angegeben"* werden. Damit gibt der KI-VO letztlich **keinen bestimmten Standard** hinsichtlich der Genauigkeit der Empfehlungen oder Handlungsweisen von KI-Systemen vor, sondern überlässt dies dem Anwender.[65] Implizit verlangt der KI-VO aber auch einen angemessenen Grad an Genauigkeit, was in Art. 9 Abs. 2 KI-VO und Art. 10 Abs. 2 KI-VO verdeutlicht

[63] Vgl. zur hierin anklingenden Praxis der negativen Konformitätsvermutung Martini/Wendehorst/Gerdemann/Wöbbeking KI-VO Art. 40 Rn. 17.
[64] Zur Begrifflichkeit Ehtechami/Gilch ZfPC 2024, 206.
[65] Martini/Wendehorst/Martini KI-VO Art. 15 Rn. 39 f. mit Vergleich zu Art. 5 Abs. 1 Buchst. d DS-GVO.

wird, ebenso in der erforderlichen technischen Dokumentation nach Anhang IV Nr. 2, 4, 9 KI-VO. Art. 15 Abs. 2 KI-VO stellt insofern in Aussicht, dass die Kommission in Zusammenarbeit mit einschlägigen Interessenträgern und Organisationen wie Metrologie- und Benchmarking-Behörden *„gegebenenfalls"* die Entwicklung von Benchmarks und Messmethoden fördert. Faktische Konkretisierungen des erforderlichen Grads der Genauigkeit können auch in den von der Kommission bereits bin Auftrag gegebenen harmonisierten KI-Normen erfolgen (→ Rn. 5).

III. Robustheit der Systeme

Hinsichtlich der Robustheit der Systeme lässt die KI-VO die genauen Anforderungen weitgehend offen, weist aber darauf hin, dass diese durch technische Maßnahmen wie Back-Up-Systeme oder *„fail-safe-plans"* erreicht werden können, Art. 15 Abs. 4 KI-VO. Jedenfalls bis entsprechende harmonisierte KI-Normen oder die in Art. 15 Abs. 3 KI-VO in Aussicht gestellten Benchmarks und Messmethoden vorliegen, kann insoweit auf thematisch einschlägige Normungen wie den **ISO/IEC TR 24029-1:2021** zurückgegriffen werden. **57**

Jedenfalls erfasst Art. 15 Abs. 4 UAbs. 1 S. 1 KI-VO **jegliche Fehler,** Störungen etc, die nicht nur innerhalb des Systems, sondern expressis verbis auch durch die Umgebung und die Interaktion des KI-Systems mit natürlichen Personen oder anderen Systemen entstehen können. Damit können auch Netzstörungen oder systemische Effekte aufgrund von Störungen von Intermediären erfasst werden, etwa wenn es um den Austausch von Daten zwischen KI-Systemen geht (zB beim autonomen Fahren[66]). Aber auch menschliche *„Fehler"* werden durch den Begriff der Interaktion mit natürlichen Personen erfasst – scheinbar kommt es in diesem Rahmen nicht darauf an, ob es sich um *„vernünftigerweise vorhersehbare"* Ereignisse handelt, da Art. 15 Abs. 3 KI-VO anders als Art. 9 Abs. 2 Buchst. b KI-VO nicht darauf Bezug nimmt. Über den allgemeinen Grundsatz der Orientierung an der Zweckbestimmung eines KI-Systems und den allgemeinen Stand der Technik gem. Art. 8 Abs. 1 KI-VO dürften aber jedenfalls völlig fernliegende Ereignisse unberücksichtigt bleiben dürfen. **58**

Bemerkenswert ist in diesem Zusammenhang, dass Art. 15 Abs. 4 S. 3 KI-VO auch die sog. *„feedback loops"* erfasst, in denen das selbst-lernende System anhand seiner eigenen Ergebnisse praktisch auf eine Pfadabhängigkeit seiner Bewertungen gelangt; diese sollen durch geeignete Maßnahmen abgemildert (aber nicht unbedingt vollständig verhindert) werden. Eine Erfolgspflicht, dass es nicht zu derartigen Verzerrungen kommt, ist Art. 15 Abs. 4 S. 3 KI-VO damit nicht zu entnehmen, was im Einklang mit Art. 10 Abs. 2 Buchst. g KI-VO steht, der geeignete Maßnahmen zur Erkennung, Verhinderung und Abschwächung von Verzerrungen fordert. **59**

IV. Cybersicherheit

Bezüglich der Cybersicherheit verlangt Art. 15 Abs. 5 KI-VO, dass KI-Systeme gegen Attacken unautorisierter Dritter gesichert sind, wobei auch die Manipulationen von Trainingsdaten (sog. *„Datenvergiftung"*) oder Inputs (sog. *„feindliche Beispiele"*) erfasst werden;[67] iRd Konformitätsbewertungen bezieht die KI-VO dabei auch Zertifizierungen nach dem **Cybersecurity Act** ein, für die dann nach Art. 42 Abs. 2 KI-VO ebenfalls eine **Ver-** **60**

[66] Bericht Ethik Kommission Automatisiertes und Vernetztes Fahren, 2017, 15, abrufbar unter: https://www.bmvi.de/SharedDocs/DE/Publikationen/DG/bericht-der-ethik-kommission.pdf?__blob=publicationFile (zuletzt aufgerufen am 4.10.2022); Albrecht SVR 10/2005, 373 (375); Jänich/Schrader/Reck NZV 2015, 313 (317); Haupt NZV 2021, 172 (174); kritisch bzgl. der Relevanz des Problems beim autonomen Fahren Lutz/Tang/Lienkamp NZV 2013, 57 (61).

[67] Zu Risiken der Cybersicherheit für KI-Systeme: EHKS KI/Ebers § 3 Rn. 55 ff. mwN; Ehtechami/Gilch ZfPC 2024, 206 (207 ff.).

mutungswirkung eingreift.[68] Allerdings verlangt Art. 15 Abs. 5 UAbs. 2 KI-VO nur, dass die technischen Lösungen zur Gewährleistung der Cybersicherheit den jeweiligen Umständen und Risiken nach *„angemessen"* sein müssen,[69] ohne dass dies mit Referenzen auf den Cybersecurity Act oder einen anderen Rechtsakt oder Standard näher spezifiziert würde.[70]

61 Dass Art. 15 Abs. 5 UAbs. 2 KI-VO den Maßstab der Angemessenheit nur auf technische Lösungen, nicht aber auf organisatorische Sicherheitsmaßnahmen bezieht, wird man wohl nicht so verstehen müssen, dass diese von Art. 15 Abs. 5 KI-VO nicht ggf. gefordert werden bzw. von Anbietern nicht als Beleg insgesamt angemessener Cybersicherheitsmaßnahmen herangezogen werden dürfen; soweit organisatorische Schutzvorkehrungen und Kontrollen dazu beitragen können, die Ausnutzung von Systemschwachstellen durch Dritte gem. Art. 15 Abs. 5 UAbs. 1 KI-VO zu verhindern.[71] Der Fokus der von Art. 15 Abs. 5 KI-VO geforderten Maßnahmen liegt indes eindeutig auf technischer Ebene, konkret auf der Minimierung auf KI-spezifischen Cybersicherheitsrisiken, wie sie in Art. 15 Abs. 5 UAbs. 3 KI-VO exemplarisch aufgeführt werden.

[68] Verordnung (EU) 2019/881 des Europäischen Parlaments und des Rates vom 17. April 2019 über die ENISA (Agentur der Europäischen Union für Cybersicherheit) und über die Zertifizierung der Cybersicherheit von Informations- und Kommunikationstechnik und zur Aufhebung der Verordnung (EU) Nr. 526/2013 (Rechtsakt zur Cybersicherheit) (ABl. 2019 L 151, 15); s. auch Roos/Weitz MMR 2021, 844 (847); Martini/Wendehorst/Martini KI-VO Art. 15 Rn. 51.

[69] In diese Richtung auch ZVEI (Zentralverband Elektrotechnik- und Elektronikindustrie e.V.), Stellungnahme zum Vorschlag der EU-Kommission für eine Verordnung zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz („AI Act"), 2021, 7, abrufbar unter: https://www.zvei.org/fileadmin/user_upload/Presse_und_Medien/Publikationen/2021/September/EU-KI-Gesetz/ZVEI-Comments-on-AI-Proposal_2021-08_DE.pdf (zuletzt aufgerufen am 4.10.2022).

[70] Krit. dazu zu Recht: Schallbruch DuD 2021, 438 (443).

[71] Vgl. auch → Rn. 38; insoweit wohl aA Martini/Wendehorst/Martini KI-VO Art. 15 Rn. 57.

§ 6. Transparenzanforderungen an Hochrisiko- und andere KI–Systeme

Übersicht

Wie ein KI-System zu seinen Arbeitsergebnissen gelangt, lässt sich für Außenstehende nur mit hohem technischem Aufwand oder – bei den zunehmend komplexeren Formen des maschinellen Lernens – auch überhaupt nicht mehr nachvollziehen (Stichwort: black box-Problematik). Gleichzeitig ist ein gewisses Maß an Systemverständnis unabdingbar, um das KI-System ordnungsgemäß bedienen und seine Arbeitsergebnisse verwerten zu können. Vor diesem Hintergrund verwundert es wenig, dass die **Transparenz von KI-Systemen** aus regulatorischer Perspektive eine zentrale Rolle einnimmt. Die KI-Verordnung greift Transparenzgesichtspunkte an zwei zentralen Stellen auf: als ein Teilaspekt der Anforderungen an Hochrisiko-KI-Systeme in Kapitel III KI-VO (insbes. Art. 12–14 KI-VO, zu den sonstigen Anforderungen für Hochrisiko-KI-Systeme → § 5 Rn. 1 ff.) sowie iRd *„Transparenzpflichten für bestimmte KI-Systeme"* in Kapitel IV KI-VO (Art. 50 KI-VO). **1**

A. Einleitung

Bevor die Transparenzanforderungen der KI-Verordnung im Detail dargestellt und analysiert werden (→ Rn. 10 ff. und → Rn. 37 ff.) soll zunächst die klassische Transparenz- **2**

Problematik im Kontext von KI-Systemen (KI als black box, → Rn. 3 ff.) näher beleuchtet werden. Die sich anschließende Untersuchung des Begriffs der Transparenz sowie seine spezifischen Verwendungskontexte in der Verordnung legen nahe, dass der Verordnungsgeber kein einheitliches Verständnis für *„Transparenz"* kennt, sondern den Begriff vielmehr in sehr unterschiedlichen Zusammenhängen verwendet. Insgesamt lassen sich dabei vier zentrale Wirkungsdimensionen von Transparenz beschreiben, die sogar noch über das klassische black box-Problem hinausreichen (→ Rn. 4 ff.).

I. Die klassische Transparenz-Problematik: KI als black box

3 Mit Blick auf das Transparenzziel verhalten sich Algorithmen und KI-Systeme dialektisch. Einerseits **lösen** KI-Systeme ein drängendes Transparenzproblem unserer modernen Informationsgesellschaft, indem sie den Umgang mit großen Datenbeständen bewältigen. Andererseits **entsteht** durch den Einsatz von KI-Systemen ein neues Transparenzproblem:[1] Denn wie das KI-System zu seinen Ergebnissen gelangt ist, lässt sich für Außenstehende nur noch eingeschränkt nachvollziehen; dies gilt in besonderer Weise für die zunehmend komplexer werdenden Formen des **maschinellen Lernens** (deep learning). Man spricht deshalb auch vom KI-System als **black box.** Die black box-Eigenschaft stellt sich vor allem deshalb als Problem dar, weil sich die Funktionsweise des Systems aufgrund der eingeschränkten Nachvollziehbarkeit von außen nicht mehr vollumfänglich kontrollieren lässt. Dies macht das KI-System nicht nur anfällig für Fehlfunktionen, wie zB Diskriminierungen,[2] die bei lernenden KI-Systemen etwa entstehen können, wenn sich Verzerrungen oder Benachteiligungen zulasten bestimmter Personengruppen aus den verwendeten Trainingsdaten in den Ausgabeentscheidungen der *„angelernten"* KI fortsetzen.[3] Auch gehen eine hohe Systemkomplexität und fehlende Transparenz häufig mit einer gering(er)en Vertrauenswürdigkeit des Systems einher, was sich dann auch in der Akzeptanz entsprechender KI-Systeme auf Nutzerseite niederschlagen kann.[4] Dabei steht zu erwarten, dass sich dieses Problem mit der fortschreitenden technischen Entwicklung in der Tendenz eher noch weiter verschärfen wird.[5] Mit Blick auf diese Konfliktlage hat sich zuletzt ein ganzes Forschungsfeld der Suche nach Ansätzen und Methoden für algorithmische Transparenz verschrieben:[6] Die Forschung zu **eXplainable AI** (XAI) zielt darauf ab, bessere Einblicke in algorithmische Entscheidungsprozesse zu erlangen. Zentral geht es dabei um die Frage, wie der komplexe Arbeitsprozess der Maschine überhaupt in einer für den Menschen verständlichen Form dargestellt werden kann.

II. Begriff und Wirkungsdimensionen

4 Obwohl der Begriff „Transparenz" im Normtext der KI-VO mehrfach aufgegriffen wird,[7] findet sich dort keine spezifische Begriffsdefinition. Ein Blick auf die jeweiligen Verwendungszusammenhänge legt nahe, dass **kein einheitliches Begriffsverständnis** für Transparenz existiert. So verlangt Art. 13 Abs. 1 S. 1 KI-VO (vgl. dazu → Rn. 16 ff.), dass Hochrisiko-KI-Systeme so konzipiert und entwickelt werden, dass ihr Betrieb *„hinreichend transparent"* ist. Auf diese Weise sollen die Betreiber des Systems in die Lage versetzt werden, die Ergebnisse des Systems angemessen interpretieren und verwenden zu können.

[1] Vgl. Roth-Isigkeit GesR 2022, 278 (280) („Transparenzproblem zweiter Ordnung").
[2] Dazu Dzida/Groh NJW 2018, 1917; Sesing/Tschech MMR 2022, 24; Steege MMR 2019, 715.
[3] Dazu Barocas/Selbst Cal. L. Rev. 2016, 671 (680 ff.).
[4] Vgl. mit Beispielen Käde/von Maltzan CR 2020, 66 (66 f.). Zur Bedeutung der Vertrauenswürdigkeit für die Nutzung von KI-Systemen vgl. von Eschenbach Philosophy&Technology 2021, 1607 (1609 ff.).
[5] Dazu bereits Kumkar/Roth-Isigkeit JZ 2020, 277 (285).
[6] Zu den verschiedenen Ansätzen Käde/von Maltzan CR 2020, 66 (68 ff.).
[7] So etwa in Art. 1 Abs. 1 Buchst. d KI-VO, Art. 13 Abs. 1 KI-VO und Art. 50 KI-VO sowie in Erwgr. 59 f., 72, 131 f. KI-VO.