

Leibniz und der AI Act

– Was ist KI und wer ist für sie verantwortlich?

Malte Grützmacher

Eine Festschrift zum 50-jährigen Bestehen der Juristischen Fakultät der Leibniz Universität Hannover kommt im Jahr 2024 nicht umhin, sich mit der Frage zu beschäftigen, was künstliche Intelligenz (KI) ist und wann sie zu regulieren bzw. durch die EU reguliert ist.

Schon *Leibniz* philosophierte 1714 über denkende Maschinen.¹ Auch entwickelte er für die KI mitentscheidend die alte aristotelische Logik erheblich weiter² und ersann ab 1671 seine *machina arithmetica*.³ Wäre ihm, dem Universalgelehrten – Mathematiker, Philosophen und auch Juristen – bewusst gewesen, dass seine Ideen und Erfindungen zum vor uns liegenden Zeitalter der Künstlichen Intelligenz beitragen würden, so hätte er sich im Rahmen seiner berühmt-berüchtigten Zettelwirtschaft gewiss auch Gedanken darüber gemacht, ab wann Rechenmaschinen – sprich Computer und deren Software – zu „riskanter“ KI mutieren und wer für

-
- 1 *Leibniz*, Monadologie, § 17 (in französischer Originalsprache wiedergegeben in *Nolen* (Hrsg.), *Leibniz - La Monadologie*, 1881, S.195 f. [abrufbar unter https://fr.wikisource.org/wiki/Monadologie/%C3%89dition_Nolen_1881]; in deutscher Sprache wiedergegeben in *Guth* (Hrsg.), *Gottfried Wilhelm Leibniz: Monadologie*. Deutsch mit einer Abhandlung über Leibniz' und Herbarts Theorien des wirklichen Geschehens von Dr. Robert Zimmermann, Erstdruck Wien 1847, Berlin 2017, S. 16 f.; vgl. dazu auch *Gabriel*, Erfinder der besten aller Welten, *Süddeutsche Zeitung* v. 14.11.2016; *Lossau*, Leibniz – der Urvater des Digitalzeitalters, *Welt* v. 13.11.2016.
 - 2 *Manhart*, KI und Machine Learning: Eine kleine Geschichte der Künstlichen Intelligenz, *Computerwoche* v. 2.2.2024 (<https://www.computerwoche.de/a/eine-kleine-geschichte-der-kuenstlichen-intelligenz>).
 - 3 *Leupold*, *Theatrum Arithmetico-Geometricum*, Leipzig 1727, S. 35 ff., 244 (Nachdruck Hannover 1982; abrufbar auch unter https://www.db-thueringen.de/rsc/viewer/dbt_derivate_00019342/590528750_090202_02_590528750_leupold.pdf?page=1); siehe dazu die Wiedergabe der Maschine unter <https://www.hannover.de/Wirtschaft-Wissenschaft/Wissenschaft/Initiative-Wissenschaft-Hannover/Leibniz-in-Hannover/Leibniz%27-Nachlass/Leibniz-Rechenmaschine>; https://de.wikipedia.org/wiki/Gottfried_Wilhelm_Leibniz#cite_ref-33).

diese verantwortlich zeichnen sollte. Aber hätte er die Linien dort gezogen, wo die EU es nun mit dem sog. AI Act⁴ getan hat?

Auch an der Juristischen Fakultät der Universität Hannover nahm sich der zu deren Gründungszeiten frisch zum Thema „*Juristische Entscheidung und Elektronische Datenverarbeitung: Methodenorientierte Vorstudie*“ habilitierte und 1974 berufene Prof. Wolfgang Kilian früh nicht nur der rechtsinformatischen,⁵ sondern in der Folge auch der juristischen Aspekte der KI an.⁶ Zahlreiche aktuelle und ehemalige Ordinarien der Fakultät haben seitdem mit ihrem wertvollen Input zur rechtsgebietsübergreifenden Durchleuchtung der die KI betreffenden Rechtsfragen beigetragen.⁷

-
- 4 Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates 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 (Verordnung über künstliche Intelligenz), ABl. Nr. L vom 12.7.2024, S. 1 ff.
 - 5 Kilian, *Juristische Entscheidung und Elektronische Datenverarbeitung – Methodenorientierte Vorstudie*, Frankfurt a.M. 1974; ders., *Zur Methode juristischer Entscheidungen aus der Sicht der elektronischen Datenverarbeitung*, DVR 5 (1976), 69 ff.; ders., *Juristische Expertensysteme mit Entscheidungshilfescharakter und die Evolution des Rechts*, in: Nickel/Roßnagel/Schlink (Hrsg.), *Die Freiheit und die Macht – Wissenschaft im Ernstfall*, Festschrift für Adalbert Podlech, Baden-Baden 1994, S. 201 ff.
 - 6 Kilian, *Rechtliche Auswirkungen moderner Informationssysteme*, in: Rechtsanwaltskammer Hannover e.V. (Hrsg.), *Festschrift zur 150-Jahr-Feier des Rechtsanwaltsvereins Hannover*, Hannover 1981, S. 152 ff.; ders., *Liability for Deficient Medical Expert Systems*, in: Savory (Hrsg.), *Expertensysteme*, 2. Aufl., München/Wien 1989, S. 393 ff.; ders., *EEC Products Liability Directive and Medical Expert Systems*, in: Kaspersen/Oskamp (Hrsg.), *Amongst Friends in Computers and Law – A Collection of Essays in Remembrance of Guy Vandenberghe*, Deventer/Boston 1990, S. 93 ff.; siehe auch die Einladung der von Kilian mit geleiteten Gesellschaft für Rechts- und Verwaltungsinformatik e.V. (GRVI) und der Deutschen Gesellschaft für Informationstechnik und Recht e.V. (DGIR) zu einem Workshop zum Thema „Künstliche Intelligenz und Produkthaftung“, CR 1990, 431 f.
 - 7 Siehe bspw. die Monografien und Sammelwerke (mit zahlreichen Beiträgen von Beck, Buck-Heeb, Eichelberger, Heinze, Krügel, Oppermann, Stender-Vorwachs, Wolf) von bzw. unter Mitwirkung von Beck, *Jenseits von Mensch und Maschine – Ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs*, Baden-Baden 2012; Beck/Kusche/Valerius (Hrsg.), *Digitalisierung, Automatisierung, KI und Recht – Festgabe zum 10-jährigen Bestehen der Forschungsstelle RobotRecht*, Baden-Baden 2020; Buck-Heeb/Oppermann (Hrsg.), *Automatisierte Systeme*, München 2022; Oppermann/Stender-Vorwachs (Hrsg.), *Autonomes Fahren – Rechtsprobleme, Rechtsfolgen, technische Grundlagen*, 2. Aufl., München 2020; Ebers/Heinze/Krügel/Steinrötter (Hrsg.), *Rechtshandbuch Künstliche Intelligenz und Robotik*, München 2020; Chibanguza/Kuß/Steege (Hrsg.), *Künstliche Intelligenz – Recht und Praxis automati-*

Nunmehr hat das Vordenken teils ein Ende gefunden. Denn der EU-Gesetzgeber hat sich am 13. Juni 2024 final zur Regulierung von sog. „riskanter“ KI geäußert und den AI Act – zu Deutsch KI-Gesetz oder auch KI-VO – verabschiedet.⁸ Vieles, was in der KI-VO niedergeschrieben ist, ist sicher überdenkenswert, mitunter aber auch für die tradierten Disziplinen des Rechts bedenkenswert.⁹

Zwei fundamentale Fragen sollen hingegen im Mittelpunkt dieses Beitrags stehen: Was ist KI nach dem AI Act? Und wer ist für sie verantwortlich? Zu verstehen, was KI i.S.d. KI-VO ist, ist dabei wichtig, um erkennen zu können, wer im Rahmen einer arbeitsteiligen Entwicklung sowie bei der Nutzung von KI-Systemen für diese verantwortlich sein kann oder sollte. Denn verantwortlich sein kann nur, wer das Risiko dieser Technologie reduzieren kann.

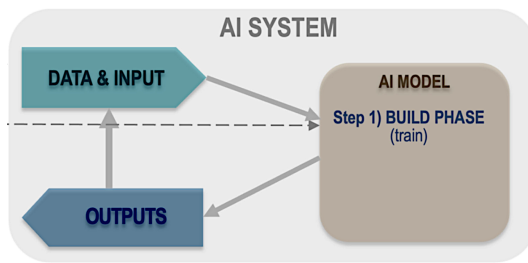
I. Entstehung und Nutzung von KI-Systemen (Lifecycle)

KI-Systeme entstehen in mehreren Entwicklungsschritten und Iterationen. So wird im ersten Schritt Grundlagentechnologie (etwa neuronale Netze oder Genetic Algorithms) genutzt, um ein KI-Modell zu schaffen.

sierter und autonomer Systeme, Baden-Baden 2022; *Wischmeyer/Rademacher* (Hrsg.), *Regulating Artificial Intelligence*, Cham 2020; sowie bspw. auch die Einzelbeiträge von *Dornis*, *Der Schutz künstlicher Kreativität im Immaterialgüterrecht*, GRUR 2019, 1252; *ders.*, *Künstliche Intelligenz und Patentrecht – Klarstellungen zur „Erfindung ohne Erfinder“*, GRUR Patent 2023, 14; *ders.*, *Künstliche Intelligenz und Vertragsschluss*, AcP 2023, 717 oder *Kramme*, *Regulierung von ChatGPT & Co.: Neue Regeln für eine neue Technik?*, *Future Law Working Papers* 2024 (2), 3.

⁸ Oben, Fn. 4.

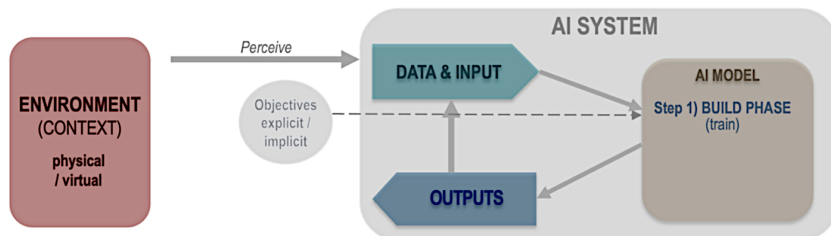
⁹ Zur Ausstrahlwirkung *Grützmaker*, *Die zivilrechtliche Haftung für KI nach dem Entwurf der geplanten KI-VO – Potentielle zivilrechtliche Auswirkungen des geplanten KI-Sicherheitsrechts: ein neues Schutzgesetz i.S.v. § 823 Abs. 2 BGB am Horizont*, CR 2021, 433 (442).



Quelle: Grafik OECD (2024), Explanatory Memorandum on the Updated OECD Definition of an AI System (Fn. 44)

Naturgegeben kann der Aufbau zu Qualitätsunterschieden führen und damit Mängel der KI begründen. Soweit ein erstes generelles Training stattfindet, kann es bei einer zu engen Orientierung an bestimmten Trainingsdaten zu einem sog. *overfitting* kommen, bei dem das Modell nicht abstrakt genug für neue Daten vortrainiert ist. Oder es entsteht durch *underfitting* ein zu schlichtes Modell, das die Struktur der Daten nicht ausreichend präzise erfasst.

Das KI-Modell wird sodann in ein KI-System mittels Software (API, etc.) integriert und mithilfe von Datensätzen (weiter) trainiert:



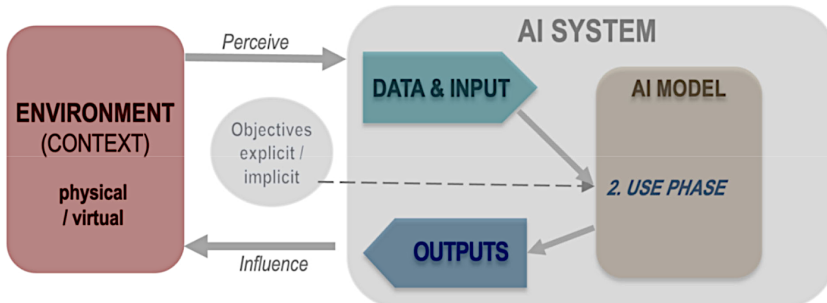
Quelle: Grafik OECD (2024), Explanatory Memorandum on the Updated OECD Definition of an AI System (Fn. 44)

Die Qualität der Datensätze wie des Trainings bestimmt in erheblichem Maß die Qualität, Verhaltensweise und Fehleranfälligkeit bzw. -toleranz des KI-Systems.¹⁰ Das Training kann dabei beaufsichtigt oder unbeaufsichtigt

¹⁰ Dazu Huth, Federated Learning, in: Kaulartz/Braegelmann (Hrsg.), Rechtshandbuch Artificial Intelligence und Machine Learning, München 2020, S. 37.

erfolgen.¹¹ Ziel des Trainings kann die Schaffung eines spezifischen KI-Systems oder auch „nur“ die Schaffung einer General Purpose AI sein, also eines KI-Modells mit allgemeinem Verwendungszweck i.S.v. Art. 3 Nr. 63 KI-VO. Letzteres wird sodann nochmals für spezifische Zwecke trainiert.

In der Folge wird das System genutzt. Das System erhält als Input Daten, aus denen Output in Form etwa von Empfehlungen oder Handlungen generiert werden kann:



Quelle: Grafik OECD (2024), Explanatory Memorandum on the Updated OECD Definition of an AI System (Fn. 44)

Mitunter setzt das KI-System bei der Nutzung aber das Lernen fort und entwickelt sich weiter. Es findet also gewissermaßen ein „Training on the Job“ statt, welches die Qualität des Systems aber auch wieder reduzieren kann.¹²

II. Was ist KI im Sinne des AI Acts?

Die KI-Verordnung betrachtet das Objekt ihrer Regulierung – die KI – zweistufig sowie sodann differenziert. Genau genommen reguliert die KI-VO „nur“ bestimmte KI-Systeme, beschäftigt sich also selektiv und nicht generalistisch¹³ mit der KI. Hinzuge treten sind auf den letzten Metern Ende 2023 noch die KI-Modelle mit allgemeinem Verwendungszweck.

11 Kaulartz, Trainieren von Machine-Learning-Modellen, in: ders./Braegelman (Hrsg.), Rechtshandbuch AI (Fn. 10), S. 32 (35 f.).

12 Dazu Huth, Federated Learning (Fn. 10), S. 37.

13 Zu den, wie gesagt, gleichwohl bestehenden Ausstrahlwirkungen Grützmaier (Fn. 9), CR 2021, 442.

1. KI-Systeme

Zunächst einmal definiert die KI-VO in Art. 3 Nr. 1 KI-VO das regelungsgegenständliche „KI-System“ als

„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“.

[illegible]

Quelle: *Leibniz*, Memoires de l'Academie Royale, 1703, S. 86

Es fragt sich, was hier reguliert wird – klassische Software, gewissermaßen also der auf der Leibniz'schen Algebra¹⁴ beruhende Stoff für den Computer als Nachfolger von Leibniz' *machina arithmetica*, oder aber Machine oder gar Deep Learning (bzw. starke KI), wie sie nach Vorarbeiten von Minsky 1951 erst die von Rosenblatt und Wightman 1957¹⁵ „erfundenen“ neuronalen Netze ermöglichten?

Die Definition eines Systems, das anpassungsfähig sein kann, scheint auf den ersten Blick auch Software zu beschreiben. Software arbeitet auch in unterschiedlichem Grad autonom. Und kann normale Software nicht auch aus den erhaltenen „Eingaben für explizite oder implizite Ziele ableite[n]“, wie die Umgebungen beeinflussende „Ausgaben“ hervorbringen?

Der AI Act würde, so verstanden, auch auf ganz normale, klassische Software zielen. Eine genauere Analyse zeigt, dass dies nicht der Fall ist.

a) Gesetzgebungsgeschichte und Hintergründe

aa) Der Kommissions-Entwurf und die an ihm geübte Kritik

Der Kommissions-Entwurf der KI-VO sah in Art. 3 Nr. 1 zunächst noch folgende Definition vor:

„System der künstlichen Intelligenz‘ (KI-System) eine Software, die mit einer oder mehreren der in Anhang I aufgeführten Techniken und Konzepte entwickelt worden ist und im Hinblick auf eine Reihe von Zielen, die vom Menschen festgelegt werden, Ergebnisse wie Inhalte, Vorhersagen, Empfehlungen oder Entscheidungen hervorbringen kann, die das Umfeld beeinflussen, mit dem sie interagieren“

wobei Anhang I folgende Beschreibung enthielt:

*„ANHANG I
TECHNIKEN UND KONZEPTE DER KÜNSTLICHEN INTELLIGENZ
gemäß Art. 3 Absatz 1*

14 Wikipedia, „Leibniz algebra“ (abrufbar unter https://en.wikipedia.org/wiki/Leibniz_algebra); siehe auch Lenzen, Leibniz und die Boolesche Algebra, *Studia Leibnitiana* 1984, 187 ff.

15 Wikipedia, „Künstliches neuronales Netz“ (abrufbar unter https://de.wikipedia.org/wiki/K%C3%BCnstliches_neuronales_Netz).

- a) *Konzepte des maschinellen Lernens, mit beaufsichtigtem, unbeaufsichtigtem und bestärkendem Lernen unter Verwendung einer breiten Palette von Methoden, einschließlich des tiefen Lernens (Deep Learning);*
- b) *Logik- und wissensgestützte Konzepte, einschließlich Wissensrepräsentation, induktiver (logischer) Programmierung, Wissensgrundlagen, Inferenz- und Deduktionsmaschinen, (symbolischer) Schlussfolgerungs- und Expertensysteme;*
- c) *Statistische Ansätze, Bayessche Schätz-, Such- und Optimierungsmethoden.“*

Ziel der so festgelegten Begriffsbestimmung für KI-Systeme war es, „so *technologieneutral und zukunftstauglich wie möglich zu sein und den rasanten Entwicklungen in der KI-Technologie und auf dem KI-Markt Rechnung zu tragen*“; Anhang I sollte durch die konkrete Aufführung von Konzepten und Techniken für Rechtssicherheit sorgen und an neue technologische Entwicklungen anpassbar sein.¹⁶

Die Definition wurde aber zu Recht gerade wegen der Gefahr des Einbezugs einfacher Software als zu weit empfunden und entsprechend kritisiert.¹⁷ Vorgeschlagen wurde u.a. vom Verfasser, nur auf solche Systeme abzustellen, „*deren Entscheidungen bzw. deren Verhalten (als mit zumutbaren Mitteln) praktisch unvorhersehbar anzusehen sind*“.¹⁸ Adressiert worden wäre somit das sog. Blackbox-Problem der künstlichen Intelligenz,¹⁹ das ge-

16 Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz (Gesetz über Künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union vom 21.4.2021, COM(2021) 206 final, 2021/0106 (COD), S. 14.

17 Etwa *Bomhard/Merkle*, Europäische KI-Verordnung – Der aktuelle Kommissionsentwurf und praktische Auswirkungen, RD 2021, 276 (277); *Grützmacher/Füllsack*, Der Entwurf einer EU-KI-Verordnung, ITRB 2021, 159 (160); *Spindler*, Der Vorschlag der EU-Kommission für eine Verordnung zur Regulierung der Künstlichen Intelligenz (KI-VO-E) – Ansatz, Instrumente, Qualität und Kontext, CR 2021, 361 (373); *Steege*, Definition von Künstlicher Intelligenz in Art. 3 Abs. 1 KI-VO-E – Ein Meilenstein auf dem Weg zu einem harmonisierten Rechtsrahmen?, MMR 2022, 926 (928); *Deutscher Anwaltverein*, Stellungnahme Nr. 57/2021 zum Vorschlag der EU-Kommission für eine Verordnung zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Gesetz über Künstliche Intelligenz) vom 21.04.2021 (COM (2021) 206 final), Brüssel, November 2021, S. 7 (abrufbar unter <https://anwaltverein.de/de/newsroom/s-n-57-21-ki-verordnungsvorschlag-der-eu-kommission>).

18 *Deutscher Anwaltverein*, Stellungnahme Nr. 57/2021 (Fn. 17), S. 7.

19 Dazu ausführlich *Käde/von Maltzan*, Die Erklärbarkeit von Künstlicher Intelligenz (KI) – Entmystifizierung der Black Box und Chancen für das Recht, CR 2020, 66 (67 ff.).

rade zu den von ihm ausgehenden und spezifischen, regelungsbedürftigen Gefahren wie auch dessen Schwierigkeiten bei der Zurechnung führt.²⁰

bb) OECD-Definition und Parlaments-Entwurf

Statt einen solchen Ansatz zu wählen, hat sich das Parlament, auch um den kritisierten Einbezug einfacher Software zu vermeiden, sodann für den heutigen Entwurf der KI-VO an einer OECD-Definition für KI-Systeme orientiert und diese in den Entwurf eingebracht.²¹ Diese OECD-Definition lautete in der Fassung vom 24.5.2019:

“AI system: An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.”²²

Leider nur in Teilen²³ übernommen wurden die Änderungen der aktualisierten Fassung der OECD-Definition vom 8.11.2023:

“AI system: An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.”²⁴

20 Dazu Grützmacher (Fn. 9), CR 2021, 435; ders., Der Entwurf der EU-Verordnung für Künstliche Intelligenz, in: Buck-Heeb/Oppermann (Hrsg.), Automatisierte Systeme (Fn. 7), S. 459 (477 f.).

21 Vgl. Dok. 2021/0106(COD) vom 26.1.2024, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, II.2., S. 3: *“The definition [...] has been modified to align it more closely with the work of international organisations working on artificial intelligence, notably the OECD. Moreover, the corresponding Recital 6 [...] clarifies that the definition is not intended to cover simpler traditional software systems or programming approaches, which are based on the rules defined solely by natural persons to automatically execute operations.”*

22 OECD, Recommendation of the Council on Artificial Intelligence, OECD/C/MIN(2019)3/FINAL, S. 3 (abrufbar unter [https://one.oecd.org/document/C/MIN\(2019\)3/FINAL/en/pdf](https://one.oecd.org/document/C/MIN(2019)3/FINAL/en/pdf)).

23 Dazu näher unter b) bb) (1) und (3) und dd).

24 OECD, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, S. 7 (abrufbar unter <https://legalinstruments.oecd.org/en/instrument/s/oecd-legal-0449>).

Legt man den englischen Text des AI Act daneben, zeigen sich zu beiden Versionen eine Vielzahl wörtlicher Übereinstimmungen, aber auch Unterschiede:

“AI system’ means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments”.

Dazu heißt es heute im Erwägungsgrund 12 der KI-VO:

„Der Begriff ‚KI-System‘ in dieser Verordnung sollte klar definiert und eng mit der Tätigkeit internationaler Organisationen abgestimmt werden, die sich mit KI befassen, um Rechtssicherheit, mehr internationale Konvergenz und hohe Akzeptanz sicherzustellen und gleichzeitig Flexibilität zu bieten, um den raschen technologischen Entwicklungen in diesem Bereich Rechnung zu tragen. Darüber hinaus sollte die Begriffsbestimmung auf den wesentlichen Merkmalen der KI beruhen, die sie von einfacheren herkömmlichen Softwaresystemen und Programmierungsansätzen abgrenzen, und sollte sich nicht auf Systeme beziehen, die auf ausschließlich von natürlichen Personen definierten Regeln für das automatische Ausführen von Operationen beruhen.“

b) Auslegung des Art. 3 Nr. 1 KI-VO

Wann von KI-Systemen auszugehen ist, kann der Gesetzgeber nicht am in der Informatik lange gängigen sog. *Turing-Test* festmachen; denn dieser kann als eine Art manuelle „Blindverkostung“ in Form der parallelen Befragung von Mensch und Maschine²⁵ für eine gesetzliche Definition nicht fruchtbar gemacht werden. Und so versucht auch die KI-VO, wie gesagt angelehnt an die OECD-Definition (dazu oben unter a) bb)), KI allgemeiner zu definieren.

25 Vgl. dazu Wikipedia, „Turing-Test“ (abrufbar unter <https://de.wikipedia.org/wiki/Turing-Test>).

aa) „maschinengestütztes“ System

Betrachtet man nun die Definition des Art. 3 Nr. 1 KI-VO, fällt zunächst auf, dass das Merkmal „maschinengestütztes“ System wenig bis nichts besagt.²⁶ In Erwägungsgrund 12 der KI-VO heißt es dazu lapidar: *„Die Bezeichnung ‚maschinenbasiert‘ bezieht sich auf die Tatsache, dass KI-Systeme von Maschinen betrieben werden“*. Das ist zumindest aus heutiger Sicht noch²⁷ eine Selbstverständlichkeit. Das Merkmal, das der OECD-Definition entstammt, dürfte mithin zumindest in den nächsten Jahren unbedeutend bleiben.

bb) Auslegung auf in unterschiedlichem Grade autonomen Betrieb

Die Abgrenzung zwischen Entscheidungen bzw. Handlungen auf Basis deterministischer Algorithmen und autonomen Entscheidungen bzw. Handlungen steht seit jeher im Mittelpunkt der Diskussion um die Grenze klassischer Software zur KI.²⁸ Und so überrascht auf den ersten Blick nicht, dass auch laut der Legaldefinition in Art. 3 Nr. 1 KI-VO ein System vorliegen muss, *„das für einen in unterschiedlichem Grade autonomen Betrieb ausgelegt“* ist.

Auf den zweiten Blick ist dieses Tatbestandsmerkmal aber eher problematisch bzw. nichtssagend. Denn es stellt sich zum einen die Frage, ab wann von einem autonomen Betrieb gesprochen werden kann, wann also autonomes Handeln des KI-Systems vorliegt, zum anderen, was es bedeutet, wenn das System so „ausgelegt“ sein soll, dass Autonomie beim Betrieb *„in unterschiedlichem Grade“* möglich ist.

Phänotypisch lassen sich die Autonomie und die unterschiedlichen Grade (Level) an Autonomie mit dem von *Prasuraman, Sheridan* und *Wi-*

26 Ähnlich *Bomhard/Siglmüller*, AI Act – das Trilogergebnis, RD 2024, 45.

27 Siehe aber zum Brain-Computer-Interface (BCI) von Neuralink den Eintrag bei Wikipedia, „Neuralink“ (abrufbar unter <https://de.wikipedia.org/wiki/Neuralink>).

28 Dazu insb. *Boden*, Autonomy and Artificiality, in: dies. (Hrsg.), *The Philosophy of Artificial Life*, Oxford 1996, S. 95 ff.; *Ball/Callaghan*, *Explorations of Autonomy – An Investigation of Adjustable Autonomy in Intelligent Environments*, Essex 2012, passim; *Wheeler*, Autonomy, in: *Dubber/Pasquale/Das* (Hrsg.), *The Oxford Handbook of Ethics of AI*, New York 2020, S. 343 ff.

ckens²⁹ in 10 Stufen definierten Automatisierungsgrad von Entscheidungen und Handlungen von Computern beschreiben:

Hoch	10	Der Computer entscheidet alles, handelt autonom, ignoriert den Menschen
	9	Informiert den Menschen nur, wenn er, der Computer, dies beschließt
	8	Informiert den Menschen nur, wenn er gefragt wird
	7	Führt automatisch aus und informiert dann notwendigerweise den Menschen
	6	Erlaubt dem Menschen eine begrenzte Zeit, um ein Veto einzulegen, bevor die Aktion ausgeführt wird
	5	Führt einen gegebenen Vorschlag aus, wenn der Mensch zustimmt
	4	Schlägt eine Alternative vor
	3	Schränkt die Auswahl auf einige wenige ein
	2	Der Computer bietet einen vollständigen Satz von Entscheidungs-/Aktionsalternativen
Niedrig	1	Der Computer bietet keine Unterstützung; Der Mensch muss alle Entscheidungen und Handlungen treffen

(1) Autonomer Betrieb als Handeln ohne menschliche Entscheidung?

Mit dieser Abstufung wird aber weder beschrieben, ab wann von künstlicher Intelligenz auszugehen ist, noch, was ein autonomes Handeln bzw. einen autonomen Betrieb ausmacht. Allenfalls ließe sich argumentieren, dass erst ab Level 6 ein autonomer Betrieb vorliegt, weil bei den Stufen 1-5 letztlich immer noch der Mensch entscheidet.³⁰ Dafür kann auch Erwägungsgrund 12 KI-VO angeführt werden, der feststellt, dass

„KI-Systeme [...] mit verschiedenen Graden der Autonomie ausgestattet [sind], was bedeutet, dass sie bis zu einem gewissen Grad unabhängig von menschlichem Zutun agieren und in der Lage sind, ohne menschliches Eingreifen zu arbeiten.“

29 Prasuraman/Sheridan/Wickens, A model for types and levels of human interaction with automation, IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans 2000, 286 (287); aufgegriffen auch von Ball/Callaghan, Autonomy (Fn. 28), S. 2; die 10 Stufen basieren auf den Vorbereitungen von Sheridan/Verplank, Human and Computer Control of Undersea Teleoperator, MIT Man-Machine Systems Laboratory, Technical rept., Cambridge, MA 1978, S. 8-17 ff. (abrufbar unter <https://apps.dtic.mil/sti/pdfs/ADA057655.pdf>).

30 Vgl. Reichwald/Pfisterer, Autonomie und Intelligenz im Internet der Dinge, CR 2016, 208 (210); in diesem Sinne auch Wheeler, Autonomy (Fn. 28), S. 343, 350.

Aber dies wiederum hieße, dass ein System, das nicht ohne menschliches Zutun entscheidet, nicht als KI-System im Sinne der KI-VO angesehen würde. Es bestehen erhebliche Zweifel, dass das so wörtlich zu nehmen ist. – Soll etwa ein in der Radiologie eingesetztes mustererkennendes, auf neuronalen Netzen beruhendes System zur Tumorerkennung nur deshalb kein KI-System im Sinne des KI-VO sein, weil medizinisches Personal sich auch noch einmal mit der Aufnahme beschäftigen kann und (einstweilen hoffentlich noch) wird? – Diese Zweifel werden noch dadurch verstärkt, dass die OECD für ihre aktualisierte Definition aus 2023 (dazu oben unter a) bb)), die leider so nicht übernommen wurde, lediglich ergänzend ausgeführt hat, dass verschiedene KI-Systeme einen unterschiedlichen Grad an Autonomie aufweisen können. Der OECD kam es evident gerade nicht darauf an, wie autonom das KI-System am Ende des Tages agiert. Es wird in der OECD-Definition vielmehr gerade klargestellt, dass ein KI-System unabhängig vom genauen Grad der Autonomie vorliegt.

Auch führen Vorhersagen und Empfehlungen, wie sie laut Art. 3 Nr. 1 und Erwägungsgrund 12 KI-VO der Output („Ausgabe“) der KI-Systeme i.S.d. KI-VO sein können, oft nicht zu Handlungen ohne weitere menschliche Interaktion. Viel spricht vor diesem Hintergrund dafür, den autonomen Betrieb nur auf die Datenverarbeitung vom Input zum Output zu beziehen. KI liegt auch ohne direkte Einflussnahme auf die Umwelt vor.³¹

Andersherum wird kritisch angemerkt, dass auch „nicht-intelligente“ Computerprogramme zu einem gewissen Grad unabhängig und in der Lage seien, ohne menschliches Eingreifen zu arbeiten, und daher bezweifelt, dass durch die Einführung des Begriffes der Autonomie der Kritik am zu weiten Begriff des KI-Systems gem. dem früheren Anhang I ausreichend Rechnung getragen wurde.³² Bei einem solchen – allerdings zu hinterfragenden – Verständnis (dazu sogleich unter (2)) würde das Merkmal klassische Software in der Tat nicht ohne Weiteres aus dem Anwendungsbereich der KI-VO ausnehmen.

Fest steht nach dem Gesagten insofern zunächst einmal nur, dass eine gewisse Autonomie vorliegen muss, dass aber die Definition gerade nicht festlegt, welches Level an Autonomie bestehen muss. Man wird das Merkmal

31 Vgl. *Stiemerling*, Technische Grundlagen, in: Kaulartz/Braegelmann (Hrsg.), *Rechtshandbuch AI* (Fn. 10), S. 15, (23, 25); *Borges*, Der Begriff des KI-Systems – Tatbestandsmerkmale und Auslegungsansätze, CR 2023, 706 (715).

32 *Becker/Feuerstack*, Der neue Entwurf des EU-Parlaments für eine KI-Verordnung, MMR 2024, 22.

mithin dahingehend verstehen müssen, dass nur ein gewisses Mindestmaß an Autonomie erforderlich ist.

(2) Autonomie als Ausdruck mangelnder Vorhersehbarkeit und fehlender Determiniertheit

Damit fragt sich, wann das besagte Mindestmaß an Autonomie erreicht ist. Mit der Frage, wann Autonomie vorliegt, hat sich grundlegend *Boden* befasst.³³ Nach *Boden* ist für die Autonomie relevant das Ausmaß, in dem

- die *Reaktion* auf die Umwelt direkt, d.h. determiniert, oder indirekt, d.h. *vermittelt durch innere Mechanismen*, die teilweise von der *individuellen Vergangenheit* abhängen, ist,
- die Steuerungsmechanismen *selbst generiert* und nicht von außen aufgezungen wurden und
- die inneren Steuerungsmechanismen reflektiert und/oder selektiv im Lichte allgemeiner Interessen oder der Besonderheiten des aktuellen Problems in seinem umweltbezogenen Kontext *modifiziert werden können*.³⁴

Reichwald und *Pfisterer* fassen dieses als Bestimmung der Autonomität durch die Anteile des gelernten Wissens, der Kontrolle über Eingangsparameter sowie der ausgeübten Selbstkontrolle zusammen.³⁵ Auch *Zech* stellt die Autonomie insofern in den Kontext der eigenen Lernfähigkeit der KI, damit aber auch verminderter Kontrollmöglichkeiten.³⁶ Deutlich wird in all diesen Äußerungen, dass die Autonomie korrespondierend zu einem

33 *Boden*, Autonomy (Fn. 28), S. 95 ff.

34 *Boden*, Autonomy (Fn. 28), S. 102 ff. (“the extent to which response to the environment is direct (determined only by the present state in ten external word) or indirect (mediated by inner mechanisms partly dependent on the creature’s previous history) ..., the extent to which the controlling mechanisms were self-generated rather than externally imposed ... the extent to which the inner directing mechanisms can be reflected upon, and/or selectively modified in the light of general interests or the particularities of the current problem in its environmental context ...”).

35 Vgl. *Reichwald/Pfisterer* (Fn. 30), CR 2016, 210.

36 *Zech*, Entscheidungen digitaler autonomer Systeme: Empfehlen sich Regelungen zu Verantwortung und Haftung?, Gutachten zum 73. Deutschen Juristentag, München 2020, S. A31 ff., A 39.

mangelnden Determinismus und damit zu einer gewissen – als solche gar nicht als negativ zu qualifizierenden – Unvorhersehbarkeit führt.³⁷

In der juristischen Literatur finden sich weitere Definitionsversuche in diese Richtung. So fasst *Linardatos*³⁸ den Begriff der Autonomie treffend wie folgt zusammen:

- „(1) *Das System trifft Entscheidungen in Situationen der Ungewissheit. Mit Entscheidung ist hierbei im funktionalen Sinne eine Reaktion des Systems auf adressierte Erwartungen gemeint.*
- (2) *Das System kann zwischen Alternativen wählen, die Auswahl ist mithin nicht prädestiniert.*
- (3) *Die getroffene Auswahl beruht auf angelernten Erfahrungswerten des Systems, es handelt somit nach eigenen Gesetzmäßigkeiten.*
- (4) *Das Verhalten des Systems kann infolge der technischen Veränderlichkeit für die jeweils konkrete Situation nicht vorhergesagt werden.*
- (5) *Nach getroffener Entscheidung und entsprechender Aktion des Systems bestehen für alle Beteiligten nachhaltige Schwierigkeiten, die Entscheidungsnarrativen des technischen Aktanten mit angemessenem Ressourcenaufwand nachzuvollziehen.“*

Ähnlich äußert sich *Steege*, der für Autonomie ebenfalls das *Fehlen eines Determinismus* im Sinne einer herkömmlichen Wenn-Dann-Software ausmacht, wobei so verstandene KI ohne jegliche bzw. mit geringer menschlicher Steuerung oder Aufsicht auskäme.³⁹ Auch er betont in diesem Kontext als Unterschied zu konventionellen IT-Systemen die *Unvorhersehbarkeit* der Handlungen aufgrund selbständigen Lernens und der Anpassung des KI-Systems sowie eben die Unsicherheit der Reproduzierbarkeit der Handlungen selbst bei gleichen Voraussetzungen.⁴⁰

Borges sieht als Autonomie zwar einen erheblichen Spielraum bei einem nicht menschlich prädestinierten Entscheidungsprozess an und die Nichtnachvollziehbarkeit desselben eigentlich als ein von der Autonomie abzugrenzendes Phänomen; er führt insoweit zu Recht an, dass natürlich ein Zufallsgenerator kein KI-System sei, geht aber letztlich ebenfalls davon

37 Vgl. *Grützmacher* (Fn. 9), CR 2021, 435; s. auch *Boden*, Autonomy (Fn. 28), S. 105.

38 *Linardatos*, Autonome und vernetzte Aktanten im Zivilrecht, in: Baumgärtel/Kiparski (Hrsg.), DGRI Jahrbuch 2021/2022, Köln 2023, S. 443 (449).

39 *Steege*, Produkt- und Produzentenhaftung im Kontext künstlicher Intelligenz, in: Buck-Heeb/Oppermann (Hrsg.), Automatisierte Systeme (Fn. 7), S. 367 (371).

40 *Steege*, Produkt- und Produzentenhaftung (Fn. 39), S. 371.

aus, dass es entscheidend sei, dass Ausgabewerte zwar nachvollziehbar auf Eingabewerten beruhten, die *Autonomie des Entscheidungsprozesses* aber aufgrund der Komplexität nicht nachvollzogen oder vorausgesagt werden kann.⁴¹ Auch *Borges* stellt also auf die mangelnde Vorhersehbarkeit ab. Weniger überzeugend ist es hingegen, auch das Kriterium der Komplexität zu bemühen. Denn dieses ist nicht wirklich entscheidend und übersieht den Einfluss des Trainings bzw. der Lernfähigkeit von KI.

Nach alledem liegt aber Autonomie jedenfalls bzw. wohl eher „nur“ bei einer mit zumutbaren Mitteln bestehenden Unvorhersehbarkeit vor.

(3) In unterschiedlichem Grade?

Schließlich fragt sich, ob, wenn die KI-VO in Anlehnung an die OECD-Definition von 2019 fordert, dass das KI-System für einen in „unterschiedlichem Grade“ autonomen Betrieb ausgelegt sein muss („*designed to operate with varying levels of autonomy*“), kein KI-System vorliegen soll, wenn das System nicht in unterschiedlichem Grad agieren kann.

Leider ist dazu anzumerken, dass hier die alte OECD-Definition offenbar unreflektiert und daher mit ihren Unzulänglichkeiten übernommen wurde. Zumindest hat die OECD zwischenzeitlich erkannt, dass ihre Definition insofern mindestens missverständlich war⁴² und daher den nachgestellten Satz geändert in: „*Different AI systems vary in their levels of autonomy ...*“. Klar wird hier, dass nicht mehr gefordert wird, dass ein und dasselbe KI-System in unterschiedlichem Grad autonom agieren können muss. In Einklang mit dem Wortlaut des Erwägungsgrunds 12 KI-VO bringend, sollte sich das Merkmal in Art. 3 Nr. 1 KI-VO aber auch derart auslegen lassen, dass KI-Systeme „*mit verschiedenen Graden der Autonomie ausgestattet*“ sein können, solange sie ein Stück weit unabhängig agieren und ohne menschliches Eingreifen arbeiten.

Im Grundsatz wird damit also durch die Definition der KI-VO geregelt, dass nicht ein erhöhter Grad von Autonomie erforderlich ist, sondern allen-

41 *Borges* (Fn. 31), CR 2023, 713 f.

42 Siehe dazu die Erläuterung von *Russell/Perset/Grobelnik*, Updates to the OECD's definition of an AI system explained, The OECD Artificial Intelligence Policy Observatory v. 29.11.2023 (<https://oecd.ai/en/wonk/ai-system-definition-update>): „Also, the previous wording, ‘operate with varying levels’, might be read as describing a single system whose level of autonomy and adaptiveness might change over time, which was not the intended reading.“

falls eine gewisse Selbständigkeit bestehen muss, die mitunter zu unvorhersehbaren Ausgaben (Output) samt Black-Box-Phänomen führt. Klassische Software ist insofern von der Definition ausgenommen.⁴³

cc) Anpassungsfähigkeit als Kriterium?

Nachdem Art. 3 Nr. 1 KI-VO weiter auf ein System zielt, „*das nach seiner Betriebsaufnahme anpassungsfähig sein kann*“, kommt man als Rechtsanwender neuerlich ins Grübeln. Denn unklar ist nach dem Wortlaut, ob die Anpassungsfähigkeit ein zwingendes Kriterium ist, ob die Eigenschaft, anpassungsfähig zu sein, also strikt verlangt wird, oder ob ein System optional so gestaltet sein kann. Schon die englische Sprachfassung („*that may exhibit adaptiveness after deployment*“) spricht eher für die letztgenannte Auslegung.

Zurückzugreifen ist weiter auf die OECD-Definition aus 2023/24, die wohl eigentlich übernommen werden sollte, wie Erwägungsgrund 12 S. 1 KI-VO zeigt, aber wörtlich nur teilweise übernommen wurde (dazu oben a) bb) a.E.). Im Explanatory Memorandum der OECD heißt es zu dieser nämlich:

*“Adaptiveness (contained in the revised definition of an AI system) is usually related to AI systems based on machine learning that can continue to evolve after initial development. The system modifies its behaviour through direct interaction with input and data before or after deployment. Examples include a speech recognition system that adapts to an individual’s voice or a personalised music recommender system. AI systems can be trained once, periodically, or continually and operate by inferring patterns and relationships in data. Through such training, some AI systems may develop the ability to perform new forms of inference not initially envisioned by their programmers.”*⁴⁴

Es ging der OECD also offenbar darum, klarzustellen, dass es KI-Systeme gibt, die sich auch nach ihrer Inbetriebnahme bzw. ihrem Inverkehrbringen durch fortlaufendes Training bzw. Lernen weiter verändern („*Different AI*

43 Siehe auch Bomhard/Siglmüller (Fn. 26), RD i 2024, 45.

44 OECD, Explanatory Memorandum on the Updated OECD Definition of an AI System, OECD Artificial Intelligence Papers, Nr. 8, März 2024, S. 6 (abrufbar unter <https://doi.org/10.1787/623da898-en>).

systems *vary* in their levels of [...] adaptiveness after deployment.“), ohne dass dieses ein zwingendes Merkmal von KI-Systemen ist.⁴⁵

In diesem Sinne ist auch Art. 3 Nr. 1 KI-VO zu lesen, der den Sinngehalt dieser OECD-Definition aus 2023/24 offenbar übernehmen wollte. Die Anpassungsfähigkeit ist also ebenfalls kein ausschlaggebendes Kriterium. Im Erwägungsgrund 12 KI-VO heißt es dazu ebenfalls im Konjunktiv:

„Die Anpassungsfähigkeit, die ein KI-System nach Inbetriebnahme aufweisen könnte, bezieht sich auf seine Lernfähigkeit, durch sie [sic] es sich während seiner Verwendung verändern kann.“

Dass ein konkretes System sich also auch noch nach der „Auslieferung“ ändert, ist keine Voraussetzung für die Klassifikation als KI-System. Allein dieses wird durch die Definition der KI-VO klargestellt.⁴⁶

dd) Fähigkeit zur Ableitung von Ausgaben aus Eingaben, insbesondere durch Anlernen und Training

Wirklich entscheidend ist nach alledem, dass das System „aus den erhaltenen Eingaben für explizite oder implizite Ziele ableitet [bzw. ableiten kann], wie Ausgaben wie etwa Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen hervorgebracht werden, die physische oder virtuelle Umgebungen“⁴⁷ beeinflussen können“.

Auch diese Formulierung ist nicht sofort eingängig. Dafür heißt es in Erwägungsgrund 12 KI-VO dazu aber relativ klar:

„Ein wesentliches Merkmal von KI-Systemen ist ihre Fähigkeit, abzuleiten. Diese Fähigkeit bezieht sich auf den Prozess der Erzeugung von Ausgaben, wie Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen, die physische und digitale Umgebungen beeinflussen können, sowie auf die Fähigkeit von KI-Systemen, Modelle oder Algorithmen oder beides aus Eingaben oder Daten abzuleiten. Zu den Techniken, die während der Gestaltung ei-

45 Noch deutlicher die Erläuterung von Russell et al., Updates (Fn. 42): „Adaptiveness: This reflects that some AI systems can continue to evolve after their design and deployment (for example, recommender systems that adapt to individual preferences or voice recognition systems that adapt to user’s voice) and is an additional characterisation of an important group of AI systems.“

46 Ebenso Bomhard/Siglmüller (Fn. 26), RD i 2024, 45: „nicht statisch sein muss“.

47 „Umgebungen“ sollen nach ErwGr 12 KI-VO „als Kontexte verstanden werden, in denen KI-Systeme betrieben werden“.

nes KI-Systems das Ableiten ermöglichen, gehören Ansätze für maschinelles Lernen, wobei aus Daten gelernt wird, wie bestimmte Ziele erreicht werden können, sowie logik- und wissensgestützte Konzepte, wobei aus kodierten Informationen oder symbolischen Darstellungen der zu lösenden Aufgabe abgeleitet wird. Die Fähigkeit eines KI-Systems, abzuleiten, geht über die einfache Datenverarbeitung hinaus, indem Lern-, Schlussfolgerungs- und Modellierungsprozesse ermöglicht werden.“

Aus diesem Merkmal und der Beschreibung ergibt sich mithin, dass das System auf eine bestimmte Aufgabenstellung, Use Cases oder dergleichen trainiert sein muss. Denn ohne ein solches Training trägt das KI-System keine Ableitung in sich, und zwar schon gar nicht mit Blick auf explizite oder implizite Ziele.

Für die Notwendigkeit eines solchen Trainings spricht weiter, dass auch die Experten der OECD von diesem auszugehen scheinen, wie u.a. auch die Abbildungen im Explanatory Memorandum zur KI-Definition der OECD zeigen.⁴⁸ Vor allem aber wären die Anbieterpflichten für Hochrisiko-KI-Systeme andernfalls fast durchgängig sinnentleert. Denn sie zielen zum einen auf das Training und die Data Governance, zum anderen auf das Risikomanagement sowie Beobachtungs- und andere betriebsbezogene Pflichten.⁴⁹ Im Kontrast dazu sind KI-Modelle mitunter untrainiert oder zumindest für die spezifische Fragenstellung nicht final trainiert, wie auch Art. 1 Nr. 63 KI-VO indiziert.⁵⁰

ee) Beeinflussung der Umgebung durch die Ausgaben

Schließlich muss das KI-System nach Art. 3 Abs. 1 KI-VO die Umgebung beeinflussen können. Das Kriterium ist aber nicht nur vage, sondern nicht wirklich als Hürde zu verstehen; ausreichend muss es mit *Borges* sein, wenn das System das Potential hat, eine Person oder ein System in einer Entscheidung irgendwie zu beeinflussen.⁵¹

48 OECD, Explanatory Memorandum (Fn. 44), S. 7 (abgedruckt oben unter I.).

49 S. insb. Art. 9 ff. KI-VO; dazu näher unten unter III. 2.

50 Vgl. auch Kaulartz, Trainieren (Fn. 11), S. 32: „*Machine Learning ... fußt im Kern auf Modellen, die auf der Grundlage von Trainingsdaten trainiert und weitertrainiert werden und die Vorhersagen zur Klassifizierung von unklassifizierten Daten erlauben*“.

51 Dazu ausführlich *Borges* (Fn. 31), CR 2023, 714 ff.

ff) Eingebettete und nicht integrierte Systeme

Schließlich findet sich in Erwägungsrund 12 KI-VO noch der Hinweis, dass KI-Systeme

„eigenständig oder als Bestandteil eines Produkts verwendet werden [können], unabhängig davon, ob das System physisch in das Produkt integriert (eingebettet) ist oder der Funktion des Produkts dient, ohne darin integriert zu sein (nicht eingebettet)“.

Letztlich ist auch das eine Selbstverständlichkeit.

c) Erfasste Systeme

Erfasst werden nach der Definition daher Systeme, die der sich bis dato allerdings allenfalls ansatzweise abzeichnenden starken KI zugehörig sind oder auch Systeme, die selbstlernend sind bzw. sich im Laufe der Nutzung selbstständig weiter anpassen. Überdies fallen unter die Definition auf trainierten neuronalen Netzen beruhende Systeme. Auch KI-Systeme, die auf einem Genetic Algorithm-Ansatz basieren, werden erfasst. Für all diese Systeme gilt in mehr oder minder starkem Maße, dass ihr ganz konkreter Output regelmäßig nicht vorhersehbar ist; sie alle zeigen ein Black-Box-Phänomen.

Einfache Software ist deterministisch, wird nicht trainiert und agiert, wenn auch ex ante oft nur schwer zu analysieren, voraussehbar. Zumindest lassen sich Fehler im Nachhinein auffinden und ihr Fehlverhalten damit erklären.

Schwierigkeiten bereitet die Einordnung wissensbasierter Expertensysteme, die in der Informatik klassisch zum Bereich der KI gezählt werden.⁵² Hier kommt es auf die Ausgestaltung im Einzelfall an. Oftmals dürften sie aber keine KI-Systeme im Sinne der KI-VO darstellen.

52 Dazu *Bund*, Einführung in die Rechtsinformatik, Freiburg 1990, S. 291 ff.; *Stiemerling*, Kritische Eigenschaften (Fn. 10), S. 15 (22).

2. KI-Modelle mit allgemeinem Verwendungszweck

a) KI-Modelle

KI-Systeme sind aber nicht der einzige Gegenstand, den der AI Act reguliert. Überdies definiert Art. 3 Nr. 63 KI-VO von der KI-VO insbesondere in Art. 51 ff. KI-VO regulierte KI-Modelle mit allgemeinem Verwendungszweck (General Purpose AI), und zwar wie folgt:

„KI-Modell mit allgemeinem Verwendungszweck‘ ein 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“.

Dass ein KI-Modell dabei nur als Modell und damit nicht weiter definiert ist, ist wenig hilfreich und erschwert die Abgrenzung zum KI-System.⁵³ Eine Idee davon, was die EU unter einem KI-Modell versteht, vermitteln nur die Sätze 6 ff. des Erwägungsgrunds 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. KI-Modelle sind in der Regel in KI-Systeme integriert und Teil davon. Diese Verordnung enthält spezifische Vorschriften für KI-Modelle mit allgemeinem Verwendungszweck und für KI-Modelle mit allgemeinem Verwendungszweck, die systemische Risiken bergen [...]“.

⁵³ Fast schon zynisch wirkt insofern Erwägungsgrund 97 KI-VO: „Der Begriff ‚KI-Modelle mit allgemeinem Verwendungszweck‘ sollte klar bestimmt und vom Begriff der KI-Systeme abgegrenzt werden, um Rechtssicherheit zu schaffen. Die Begriffsbestimmung sollte auf den wesentlichen funktionalen Merkmalen eines KI-Modells mit allgemeinem Verwendungszweck beruhen, insbesondere auf der allgemeinen Verwendbarkeit und der Fähigkeit, ein breites Spektrum unterschiedlicher Aufgaben kompetent zu erfüllen. ...“; kritisch und das Abgrenzungsproblem bemängelnd auch Bomhard/Siglmüller (Fn. 26), RdI 2024, 50.

Offenbar hielt der Gesetzgeber den Begriff des KI-Modells im Übrigen für technisch hinreichend klar. Die OECD versteht das KI-Modell als die Kernkomponente des KI-Systems; leicht abweichend davon sehen Bomhard und Siglmüller im KI-Modell „die KI“ und im „KI-System“ das vom Endnutzer nutzbare Produkt.⁵⁴ Das korreliert ein Stück weit mit dem oben unter I. dargestellten Entstehungsprozess und der oben unter dd) dargelegten Abgrenzung anhand eines nicht finalen Trainings.

b) Allgemeiner Verwendungszweck

Relevant ist die Definition in Art. 3 Nr. 63 KI-VO jedenfalls insoweit, als nicht jedes KI-Modell der KI-VO unterfällt, sondern nur ein solches Modell, das – ggf. gerade auch mit Blick auf allgemeine Bedienungselemente substantiell trainiert – „eine erhebliche allgemeine Verwendbarkeit aufweist“ und „ein breites Spektrum unterschiedlicher Aufgaben kompetent [...] erfüllen“ kann. Wann in Ansehung dieser vagen Rechtsbegriffe ein solches allgemein verwendbares KI-Modell vorliegt, ist schwer zu prognostizieren und im Zweifel der Gerichtspraxis überlassen. Umgrenzte KI-Modelle für branchenspezifische Aufgaben hingegen scheint der Gesetzgeber nicht in den Blick genommen zu haben.⁵⁵ Letztlich ging es ihm wohl vielmehr um Modelle wie Gemini von Google,⁵⁶ GPT von OpenAI,⁵⁷ Amazon Titan, AI21 Jurassic von AI21 Labs,⁵⁸ LLaMA von Meta, Claude von Anthropic,⁵⁹ Cohere,⁶⁰ Bloom⁶¹ und Hugging Face.⁶²

54 OECD (Fn. 44), S. 8: „AI model is a core component of an AI system used to make inferences from inputs to produce outputs“; Bomhard/Siglmüller (Fn. 26), RD i 2024, 51.

55 So wohl auch Bomhard/Siglmüller (Fn. 26), RD i 2024, 50.

56 https://en.wikipedia.org/wiki/Google_Gemini.

57 https://en.wikipedia.org/wiki/Generative_pre-trained_transformer.

58 https://en.wikipedia.org/wiki/AI21_Labs.

59 <https://en.wikipedia.org/wiki/Anthropic>.

60 <https://en.wikipedia.org/wiki/Cohere>.

61 [https://en.wikipedia.org/wiki/BLOOM_\(language_model\)](https://en.wikipedia.org/wiki/BLOOM_(language_model)).

62 https://en.wikipedia.org/wiki/Hugging_Face.

III. Wer ist nach der KI-VO verantwortlich für die KI?

Kausale Beiträge zu einem Fehlverhalten der KI kann jeder liefern, der zu deren Schaffung und Betrieb beiträgt. Aber verantwortlich gemacht werden kann er dafür mangels Vorhersehbarkeit und Zurechenbarkeit oftmals nicht.⁶³

Verantwortlich im Sinne der KI-VO⁶⁴ können letztlich (nur) alle Akteure im Sinne von Art. 3 Nr. 8 KI-VO sein; denn diese treffen unter der KI-Verordnung zahlreiche, wenn auch recht unterschiedliche Pflichten. Diese kompensieren letztlich ein Stück weit, dass KI-Systeme, wie dargelegt und in der Definition von KI der KI-VO angelegt, autonom und zunehmend selbstlernend agieren, insofern also auch unvorhersehbar, was eine Zurechnung ihres Verhaltens ein Stück weit verhindert.

1. Akteure der KI-VO

Die KI-VO in Art. 3 Nr. 8 KI-VO definiert ihre Akteure als solche und damit die potentiell Verantwortlichen. Zu diesen zählt sie danach „den Anbieter, den Produkthersteller, den Betreiber, den Nutzer, den Bevollmächtigten, den Einführer oder den Händler“. Die zwei Hauptprotagonisten sind dabei der Anbieter und der Betreiber. Die sonstigen Akteure – der Bevollmächtigte,⁶⁵ der Einführer (Importeur)⁶⁶ und der Händler⁶⁷ – stehen eher in zweiter Reihe bzw. neben dem Anbieter; sie werden ergänzend zum Anbieter in die Pflicht genommen.⁶⁸

Anbieter ist nach Art. 3 Nr. 3 KI-VO insofern, wer „ein KI-System oder ein KI-Modell mit allgemeinem Verwendungszweck entwickelt oder entwickeln lässt und es [...] in Verkehr bringt oder das KI-System [...] in Betrieb nimmt“. Betreiber hingegen ist laut Art. 3 Nr. 4 KI-VO, wer „ein KI-System in eigener

63 Für die deliktische Haftung etwa Grützmacher, Die deliktische Haftung für autonome Systeme – Industrie 4.0 als Herausforderung für das bestehende Recht?, CR 2016, 695 (697); ders., Verantwortung und Haftung bei (digitalen) autonomen KI-Systemen – Sollte der Gesetzgeber neue Regelungen schaffen?, AnwBl Online 2022, 498 (499 f.); Zech, Entscheidungen (Fn. 36), S. A41 ff.

64 Allgemeine Pflichten auch der nicht regulierten Zulieferer zum KI-System (etwa eines Softwareanbieters von Teilkomponenten) lässt das Regime der KI-VO aber unberührt.

65 Art. 3 Nr. 5 KI-VO.

66 Art. 3 Nr. 6 KI-VO.

67 Art. 3 Nr. 7 KI-VO.

68 Vgl. dazu Art. 22 – 24 KI-VO.

Verantwortung verwendet, es sei denn, das KI-System wird im Rahmen einer persönlichen und nicht beruflichen Tätigkeit verwendet“.

2. Die Pflichten der Akteure und ihre daraus resultierende jeweilige Verantwortlichkeit

a) Pflichten der einzelnen Akteure

Neben den verbotenen Praktiken gem Art. 5 KI-VO einerseits und bestimmten, nicht hochriskante KI-Systeme betreffenden Transparenzpflichten des Art. 50 KI-VO andererseits regelt die KI-VO in den Abschnitten 2 und 3 vor allem Anforderungen an (Art. 8 ff. KI-VO) und Pflichten für sog. Hochrisiko-KI-Systeme (Art. 16 ff. KI-VO), also besonders sicherheits- bzw. grundrechtsrelevante⁶⁹ KI-Systeme. Es ergeben sich für diese im Wesentlichen Pflichten in folgenden Bereichen:

- Risiko- und Qualitätsmanagement
- Information und Dokumentation
- Training, Datenqualität und Data Governance
- Cybersicherheit
- Menschliche Aufsicht und Beobachtung
- Protokollierung
- Registrierung, Konformitätsprüfung und Zusammenarbeit sowie ggf., soweit erforderlich, Bestellung von Bevollmächtigten
- Meldung von Vorfällen.⁷⁰

Auffällig ist dabei, dass die Pflichten zwar zunächst insbesondere Anbieter (Pflichten mit Blick auf das KI-System in Art. 16 – 21 KI-VO sowie die Pflicht für Anbieter aus Drittländern zur Bestellung eines Bevollmächtigten gem. Art. 22 KI-VO) treffen, dann aber auch die Betreiber vom Gesetz in ganz erheblichem Maße verpflichtet werden (Art. 25 – 27 KI-VO). Hier spiegelt sich die Verschiebung des zurechenbaren Risikos weg von der Entwicklung hin zum Training und Lernen sowie entsprechend⁷¹ der Risikobeherrschung hin zur Produktbeobachtung; anzumerken ist dabei kritisch, dass sich die Pflichten zur Aufsicht und Beobachtung vor allem auch

⁶⁹ Vgl. Art. 7 Abs. 3 KI-VO.

⁷⁰ Zum Pflichtenprogramm näher Grützmacher, Entwurf (Fn. 20), S. 466 ff.

⁷¹ Grützmacher (Fn. 9), CR 2021, 437; Zech, Entscheidungen (Fn. 36), S. A51.

durch den Betreiber möglicherweise als überzogen⁷² erweisen könnten. Die Protokollierungspflicht unterstützt eine evtl. nachträgliche Aufklärung beim Fehlverhalten der Systeme.⁷³ Gesonderte Pflichten treffen Anbieter von KI-Modellen mit allgemeinem Verwendungszweck (Art. 53 ff. KI-VO).

b) Verantwortungsübergang, Beschränkung und Wegfall der Verantwortung beim Durchlaufen der KI- und sonstigen Wertschöpfungskette

Nach Art. 25 Abs. 1 KI-VO, der die Verantwortlichkeiten entlang der KI-Wertschöpfungskette regeln soll, werden nicht nur Händler und Einführer (Importeure), sondern auch *Betreiber oder sonstige Dritte zu Anbietern* von Hochrisiko-KI-Systemen und unterliegen damit den Anbieterpflichten gemäß Art. 16 KI-VO,

- wenn sie das Hochrisiko-KI-System mit ihrem *Namen* oder ihrer *Handelsmarke* versehen (lit. a);
- wenn sie eine *wesentliche Änderung* an dem Hochrisiko-KI-System vornehmen, soweit dieses dann noch weiterhin ein Hochrisiko-KI-System ist (lit. b);
- wenn sie die *Zweckbestimmung* eines KI-Systems so *verändern*, dass das betreffende KI-System zu einem *Hochrisiko-KI-System* wird (lit. c).

Auch hier kommt es zu einer Verschiebung der Verantwortlichkeit, und zwar in Richtung des nachgelagerten Akteurs, nämlich des Betreibers. Denn gleichzeitig soll nach Art. 25 Abs. 2 KI-VO der ursprüngliche *Anbieter grds.*⁷⁴ *seine Rolle als Anbieter des spezifischen KI-Systems* und damit seine Verantwortlichkeit als Anbieter *verlieren*. Das entlastet ihn stark, auch wenn er verpflichtet bleibt, eng mit neuen Anbietern zusammenzuarbeiten, die erforderlichen Informationen zur Verfügung zu stellen und mit Blick auf

72 Kritisch schon Grützmacher (Fn. 9), CR 2021, 441, 443; ders. (Fn. 63), AnwBl Online 2022, 502.

73 Zu den Beweisproblemen Grützmacher (Fn. 9), CR 2021, 437 ff.; Zech, Entscheidungen (Fn. 36), S. A84 f.

74 Anders ist die Rechtslage nach Art. 25 Abs. 3 KI-VO im Übrigen auch bei Hochrisiko-KI-Systemen bestimmter *Sicherheitskomponenten von Produkten*, die unter die in Anhang I Abschnitt A aufgeführten Harmonisierungsrechtsvorschriften der Union fallen; hier soll es die Entlassung des Originalherstellers aus der Haftung offenbar nicht ohne Weiteres geben.

die Anbieterpflichten nach der KI-VO grds. für den vernünftigerweise zu erwartenden technischen Zugang und sonstige Unterstützung zu sorgen.⁷⁵

Dass die bloße Nutzung des eigenen Namens oder der eigenen Marke eines Betreibers diesen zum *alleinigen* Anbieter eines KI-Systems werden lassen soll, ist rechtspolitisch zweifelhaft. Richtig und sinnvoll erscheint diese Erweiterung der Verantwortlichkeit hingegen schon eher in den weiteren unter lit. b) und c) geregelten Fällen, wenn der Betreiber das System im Betrieb *weitertrainiert* und/oder für *andere Zwecke* einsetzt. Denn wie angedeutet ist das Training essentiell für die Qualität des KI-Systems und die von ihm ausgehenden Risiken falscher Handlungen, Vorschläge oder Entscheidungen. Andererseits muss sich der Betreiber auf die Datenqualität des originären KI-Anbieters verlassen können und kann den auch ihn dann treffenden Pflichten zum Risiko- und Qualitätsmanagement, Pflichten mit Blick auf Datenqualität und die Data Governance sowie die Cybersicherheit nur mit dessen Hilfe gerecht werden. Ob insofern aber – zumal bei entsprechenden Lock-In-Effekten – bei den oftmals marktmächtigen außereuropäischen Anbietern die Verpflichtung zur Zusammenarbeit und zur Überlassung von Informationen nach Art. 25 Abs. 2 KI-VO für den „neuen Anbieter“ ausreichend sein wird, wird wohl erst die Praxis zeigen. Es ist jedenfalls zu befürchten, dass sich diese Anbieter zumindest ihrer vertraglichen Haftung entziehen werden. Insofern könnte sich die komplette Haftungsverschiebung des Art. 25 Abs. 2 KI-VO – sprich Entlastung des Ersherstellers – als problematisch erweisen, zumal dessen Pflichtverstöße im Rahmen von Art. 25 Abs. 2 KI-VO nach Art. 99 Abs. 4 KI-VO nicht sanktioniert sind.

Schwierigkeiten bringen könnte – je nach Lesart des Merkmals „... *beabsichtigen ... zu integrieren*“ – weiter, dass die Verpflichtungen des Anbieters von KI-Modellen mit allgemeinem Verwendungszweck zur Information und Dokumentation nicht zugunsten des Betreibers gelten, dem im Rahmen von Art. 25 Abs. 1 KI-VO Anbieterpflichten auferlegt werden; allerdings ist insofern entweder über eine analoge Anwendung des Art. 53 Abs. 1 lit. b KI-VO nachzudenken oder die Verpflichtung zur Zusammenarbeit und Information durch den Erstanbieter entsprechend Art. 25 Abs. 2 KI-VO weit auszulegen.

75 Das gilt nach Abs. 2 S. 3 nur dann nicht, wenn der Erstanbieter eines KI-Systems eindeutig vertraglich festgelegt hat, dass sein KI-System nicht in ein Hochrisiko-KI-System umgewandelt werden darf und er daher nicht der Pflicht zur Übergabe der Dokumentation unterliegt.

Auch wie weit Art. 25 Abs. 4 KI-VO im stark konzentrierten IT-Markt hilft, wird sich zeigen. Denn auch dieser unterfällt der Sanktion des Art. 99 Abs. 4 KI-VO nicht. Immerhin gibt es etwa mit den Art. 53 Abs. 1 lit. d KI-VO sowie Art. 16 lit. a KI-VO i.V.m. Art. 11 KI-VO und Art. 13 KI-VO auch (nach Art. 99 bzw. 101 KI-VO sanktionierte) gesetzliche Pflichten, die für einen gewissen Informationsfluss entlang der Wertschöpfungskette sorgen sollten. Der unterstützenden Vertragsgestaltung wird es aber gleichwohl bedürfen.

c) Die Bestimmung der Verantwortlichkeit des Betreibers durch den Anbieter

Von besonderem Interesse ist schließlich die Regelung in Art. 26 Abs. 1 KI-VO, nach der der Betreiber eines Hochrisiko-KI-Systems dazu verpflichtet ist, dieses *entsprechend der Betriebsanleitung* zu verwenden. Der Anbieter legt mithin die Pflichten bzw. den Sorgfaltsmaßstab für die Nutzung fest, und zwar gerade auch mit Blick auf die Dritte treffenden Risiken. Ein solche Regelung ist rechtspolitisch, wenn nicht gar verfassungsrechtlich, problematisch.⁷⁶ Das gilt erst recht vor dem Hintergrund, dass Verstöße nach Art. 99 Abs. 4 lit. e KI-VO mit einem Bußgeld von bis zu 3% des weltweiten Jahresumsatzes sanktioniert werden können und die KI-VO auch Schutzgesetz i.S.v. § 823 Abs. 2 BGB sein dürfte.⁷⁷

IV. Fazit

Der EU ist es mit der KI-VO gelungen, eine tragfähige, wenn auch optimierungsfähige Definition für die zu regulierenden KI-Systeme zu finden.

76 Kritisch zum privaten Setzen von Standards und der Selbstregulierung durch private Unternehmen zur insofern vergleichbaren Regelung des § 1a Abs. 1 StVG vor dem Hintergrund des Rechtsstaatsprinzips *Schmidt/Vöneky*, *Fostering the Common Good: An Adaptive Approach Regulating High-Risk AI-Driven Products and Services*, in: Vöneky/Kellmeyer/Müller/Burgard (Hrsg.), *The Cambridge Handbook of Responsible Artificial Intelligence – Interdisciplinary Perspectives*, Cambridge 2022, S. 123, 133; siehe kritisch zur Rolle des Herstellers als „Ersatzgesetzgeber“ im Rahmen des § 1a StVG schon *Grützmacher*, *Gesetzentwurf zum autonomen Fahren: Drum prüfe, was der Hersteller findet*, LTO v. 13.1.2017 (<https://www.lto.de/recht/hintergruende/h/autonomes-fahren-gesetzentwurf-haftung-fahrer-hersteller/>); *ders.* (Fn. 63), *AnwBl Online* 2022, 501.

77 *Grützmacher* (Fn. 9), CR 2021, 438 ff.

Die Definition setzt mit ihren Schwerpunkten bei der Unvorhersehbarkeit (Autonomie) und der Frage des Trainings und Lernens (Fähigkeit zur Ableitung und ggf. Anpassungsfähigkeit) an, also dort, wo die herkömmliche Haftung in ihre Grenzen gewiesen wird. Für Hochrisiko-KI werden kompensierende Sorgfaltspflichten bei der Datenauswahl sowie beim Training statuiert und die Produktbeobachtungspflichten verschärft.

Bedenklich aber ist, dass die Anbieter von KI-Systemen (wie auch von KI-Modellen), die im Markt absehbar ohnehin in einer starken, wenn nicht gar dominanten Position sein werden, mitunter allzu früh aus ihren Pflichten entlassen werden. Ob sich die Betreiber hier ausreichend absichern können, bleibt abzuwarten, ist aber in Ansehung der Marktverhältnisse zumindest offen. Problematisch ist insofern auch, dass Verstöße gegen Art. 25 Abs. 2 und 4 KI-VO nicht nach Art. 99 Abs. 4 KI-VO sanktioniert sind.

Die KI-VO zeigt lehrbuchartig, wie schwer es heute sowohl der Gesetzgeber wie auch der Rechtsanwender haben, wenn es darum geht, hochkomplexe Sachverhalte der modernen Informationsverarbeitung zu regeln oder gar zu regulieren bzw. derartige Gesetzeswerke in der Praxis zu implementieren. Die KI-VO zeigt damit auch auf, dass eine moderne juristische Fakultät ganz im Sinne des Universalgelehrten *Leibniz* nicht nur im Bereich des IT-Rechts ausbilden, sondern dass auch das Unterrichtsfach der Rechtsinformatik – neudeutsch „Legal Tech“ – seine Berechtigung im Lehrplan einer juristischen Fakultät haben sollte. Die Juristische Fakultät der Universität Hannover ist insofern für die nächsten Jahrzehnte bestens gerüstet.