Daniel Becker und Daniel Feuerstack

Die EU-KI-Verordnung – Überblick und Bewertung mit Fokus auf Entwicklung und Einsatz von KI-Systemen an Hochschulen[1]

Nachdem knapp drei Jahre zuvor die EU-Kommission den ersten Entwurf für eine Verordnung über künstliche Intelligenz (KI-VO) vorgelegt hat und der Rat sowie das EU-Parlament diese Entwürfe angepasst und überarbeitet haben, ist am 13. März 2024 nun die KI-VO final beschlossen worden. Der risikobasierte Ansatz, wonach die Verordnung zwischen inakzeptablen, hochriskanten und minimal riskanten KI-Systemen unterscheidet, wird grundsätzlich beibehalten, jedoch um eine vierte Kategorie, KI-Modelle mit generellem Verwendungszweck, ergänzt. Kernanliegen dieses Beitrages ist es, die wichtigsten Regelungen und Vorgaben der KI-VO darzustellen und zu bewerten sowie deren Implikationen für den Einsatz und die Entwicklung von KI-Systemen an Hochschulen herauszuarbeiten.

I. Einleitung

Mit Beschluss vom 13. März 2024 hat das EU-Parlament nach vorheriger Zustimmung des EU-Ministerrates die Verordnung für Künstliche Intelligenz (KI-VO)[2] angenommen. Aufbauend auf dem Ursprungsentwurf der Kommission[3] setzt die KI-VO teilweise auch zentrale Änderungsvorschläge des Rats[4] und des Parlaments[5] um, wozu insbesondere die Regulierung von KI-Systemen mit generellem Verwendungszweck sowie umfassende Ausnahmen für die Forschung an KI-Systemen gehören. Ziel der KI-VO ist gem. Art. 1 die Förderung vertrauenswürdiger und menschenzentrierter KI bei gleichzeitiger Innovationsförderung. Als zentrale Schutzgüter werden in Art. 1 KI-VO Gesundheit, Sicherheit und Grundrechte sowie kollektive Rechtsgüter wie Demokratie, Rechtsstaatlichkeit und Umweltschutz benannt. Zur Erreichung dieser Ziele werden KI-Systeme in unterschiedliche Risikokategorien unterteilt: KI-Systeme mit inakzeptablem Risiko, die in der KI-VO grds. verboten werden, KI-Systeme mit hohem Risiko, die in der Verordnung detaillierten Regeln unterworfen werden und KI-Modelle mit generellem Verwendungszweck, hinsichtlich derer insbesondere Risikobewertungs-, Transparenz- und Informationspflichten bestehen. Beim Einsatz von KI-Systemen mit minimalem Risiko können bestimmte Transparenz- und Informationspflichten bestehen, im Regelfall sind diese jedoch von der KI-VO nicht erfasst. In diesem Beitrag soll die KI-VO inhaltlich dargestellt werden, insbesondere mit Blick auf ihre Relevanz für den Einsatz und die Entwicklung von KI in der Wissenschaft und Forschung sowie im Rahmen der Lehre an Hochschulen.

II. Anwendungsbereich

1. Sachlicher Anwendungsbereich: KI-Systeme

Der Begriff des KI-Systems wird nach Vorbild der KI-Leitlinien der OECD in Art. 3 Nr. 1 KI-VO definiert 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

1 Die Autoren danken Prof. Dr. Silja Vöneky für ihre hilfreichen Anmerkungen sowie Anna-Lena Lieder für die wertvolle Korrekturarbeit.

2 Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG)Nr. 300/2008, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2020/1828 (Verordnung über künstliche Intelligenz), 13.06.2024, PE-CONS 24/1/24 REV 1, 2021/0106(COD).

3 Europäische Kommission, 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, 21.04.2021, COM(2021) 206 final.

4 Rat der Europäischen Union, 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 - General approach, 14954/22, 25.11.2022.

5 Europäisches Parlament, Gesetz über künstliche Intelligenz, 14.06.2023, P9_TA(2023)0236.

etwa Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen erstellt werden, die physische oder virtuelle Umgebungen beeinflussen können".[6]

Entscheidend für die Abgrenzung zu nicht-intelligenter Software ist nach dieser Definition das in Art. 3 Nr. 1 KI-VO eingeführte Kriterium der Autonomie.[7] Autonom agieren KI-Systeme gem. ErwG 12, wenn sie "bis zu einem gewissen Grad unabhängig von menschlichem Zutun agieren und in der Lage sind, ohne menschliches Eingreifen zu arbeiten."

Das Kriterium der Autonomie ermöglicht jedoch keine trennscharfe Abgrenzung zu herkömmlicher, nicht-intelligenter Software.[8] So könnten auch automatisierte Prozesse in einer Excel-Tabelle unter den Begriff des KI-Systems subsumiert werden,[9] da solche Softwareprogramme zu einem gewissen Grad unabhängig agieren und in der Lage sind, ohne menschliches Eingreifen zu arbeiten.[10] Die Definition ist jedoch insgesamt als hinreichend zu bewerten. Im Übrigen ist ein weites Verständnis des Begriffes KI-System auch nicht zu beanstanden, da sich auch bei regelbasierten, nichtlernenden Algorithmen, sofern sie in riskanten Anwendungsbereichen eingesetzt werden, die mit KI-Systemen assoziierten Probleme stellen können.[11] Eine Eingrenzung des zwar weit definierten sachlichen Anwendungsbereiches erfolgt in einem zweiten Schritt über den in der KI-VO verfolgten risikobasierten Ansatz.

2. Personeller und räumlicher Anwendungsbereich

Der personelle Anwendungsbereich der KI-Verordnung erfasst nach Art. 2 I KI-VO Anbieter und Betreiber von KI-Systemen, soweit diese KI-Systeme in der EU auf den Markt bringen oder nutzen wollen, sowie weitere Akteure in der KI-Wertschöpfungskette. Daneben richtet sich die KI-VO gem. Art. 2 Abs. 1 lit. g KI-VO auch an Betroffene, für die sie jedoch keine Pflichten, sondern lediglich Rechte vorsieht. Hochschulen können dabei sowohl als Anbieter als auch als Betreiber zu qualifizieren sein. Ausweislich der Legaldefinition in Art. 3 Nr. 3 KI-VO ist Anbieter "jede natürliche oder juristische Person, die ein KI-System oder ein GPAI-Modell entwickelt oder entwickeln lässt, um es auf dem Markt anzubieten oder in Betrieb zu nehmen". Dies gilt auch für Hochschulen, soweit diese entsprechend tätig werden. Als Betreiber wird nach Art. 3 Nr. 4 KI-VO "jede natürliche oder juristische Person verstanden, die ein KI-System in eigener Verantwortung und nicht im Rahmen einer rein-persönlichen oder nicht-beruflichen Tätigkeit verwendet". Bei Hochschulen werden KI-Systeme stets im Rahmen einer beruflichen Tätigkeit angewendet, so dass sie unter den Betreiberbegriff subsumiert werden können.

Gemäß Art. 2 Abs. 1 lit. a KI-VO ist diese auf Anbieter anwendbar, unabhängig davon, ob diese selbst in der EU ansässig sind. Des Weiteren findet die KI-VO gemäß Art. 2 Abs. 1 lit. b KI-VO auf alle Betreiber von KI-Systemen Anwendung, die ihren Sitz oder eine Niederlassung in der EU haben. Schließlich gilt die KI-VO auch für alle Anbieter und Betreiber, die ihren Sitz in einem Drittland haben, sofern der Output des KI-Systems sich in der Union auswirkt.[12]

3. Ausnahmen

Von besonderer Bedeutung für Hochschulen ist zunächst die in der KI-VO implementierte Privilegierung der Forschung durch die Bereichsausnahme nach Art. 2 Abs. 6 KI-VO[13] sowie die Privilegierung nach Art. 2 Abs. 8 KI-VO.

Ausweislich Art. 2 Abs. 6 KI-VO, gilt die KI-VO nicht für „KI-Systeme oder KI-Modelle, einschließlich ihrer Ausgabe, die eigens für den alleinigen Zweck der wissenschaftlichen Forschung und Entwicklung entwickelt und in Betrieb genommen werden".

6 *Bomhard/Siglmüller*, RDi 2024, 45. Siehe OECD, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, 22.5.2019, abrufbar unter: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

7 *Bomhard/Siglmüller*, RDi 2024, 45, ebenso zum Parlamentsentwurf: *Becker/Feuerstack*, MMR 2024, 22, 23.

8 Hierzu bereits zum Ursprungsentwurf der Kommission *Ebers* et al., RDi 2021, 528, 529; *Grützmacher/Füllsack*, ITRB 2021, 159, 160; *Ebert/Spiecker* gen. *Döhmann,* NVwZ 2021, 1188, 1189; *Smuha* et al., How the EU Can Achieve Legally Trustworthy AI, 2021, S. 14 f.; Spindler, CR 2021, 361, 373; *Bomhard/Merkle*, RDi 2021, 276, 278; Deutscher Bundesrat, 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 COM(2021) 206 final, Drucksache 488/21 (Beschluss), 2021, S. 5, 7 f.; *Steege*, MMR 2022, 926f. Zum Entwurf des Parlaments siehe *Becker/Feuerstack*, MMR 2024, 22, 23; *Hacker/Berz*, ZRP 2023, 226, 227.

9 Vgl. zu diesem konkreten Beispiel *Zenner* et al., Ein KI-Gesetz für Europa, FAZ, 12.06.2023, https://www.faz.net/aktuell/wirtschaft/digitec/ki-eu-parlament-einigt-sich-auf-rahmen-fuer-gesetz-18955926-p2.html.

10 Hierzu bereits *Becker/Feuerstack*, MMR 2024, 22, 23 sowie *Hacker/Berz*, ZRP 2023, 226, 227.

11 *Smuha et al.*, How the EU Can Achieve Legally Trustworthy AI, 2021, S. 14 f, wobei je nach verwendeter Technik das Risiko einer Menschenrechtsverletzung variieren kann, vgl. ebd.

12 *Bomhard/Siglmüller*, RDi 2024, 45, 46.

13 Zur Privilegierung der Wissenschaft im Kommissionsentwurf siehe *Becker*, ZfDR 2023, 164 ff.

Die begriffliche Einbeziehung von KI-Modellen legt nahe, dass diese Ausnahme vom Anwendungsbereich auch für die durch die KI-VO umfangreich regulierten GPAI-Modelle gelten soll. Entscheidend ist dabei aber, dass das KI-System oder das KI-Modell „eigens für den alleinigen Zweck der wissenschaftlichen Forschung und Entwicklung entwickelt und in Betrieb genommen wird". Wenn also ein KI-System oder ein KI-Modell von Hochschulen neben der Forschung auch für andere Zwecke verwendet oder entwickelt wird, kann die Privilegierung bereits entfallen. Zu Abgrenzungsproblemen kann dies bei anwendungsbezogenen Drittmittelprojekten führen. Bei diesen wird die Anwendbarkeit der Bereichsausnahme des Art. 2 Abs. 6 KI-VO im Einzelfall zu bestimmen sein. Sobald das KI-System auch für einen weiteren Zweck entwickelt oder verwendet wird, also beispielsweise einer möglichen Nutzung in einem aus dem Projekt ausgegliederten Start-Up mit auch wirtschaftlicher Zielsetzung, kommt eine Privilegierung nach Art. 2 Abs. 6 KI-VO nicht mehr in Betracht. Auch bei Dual-Use-Anwendungen greift die Bereichsausnahme für dieses KI-System insgesamt nicht mehr und die Anwendbarkeit der KI-VO bleibt bestehen. Der Umstand, dass die Bereichsausnahme des Art. 2 Abs. 6 KI-VO die Anwendbarkeit derselben in ihrer Gänze ausschließt, spricht dabei für eine enge Auslegung des alleinigen Zwecks. Hierfür kann auch angebracht werden, dass unabhängig von der Anwendbarkeit des Art. 2 Abs. 6 KI-VO die Vorfeldprivilegierung wissenschaftlicher Forschung nach Art. 2 Abs. 8 KI-VO bestehen bleibt.

Die Einbeziehung des Entwicklungsbegriffes deutet darauf hin, dass in Übereinstimmung mit der Auslegung des Grundrechts auf Forschungsfreiheit aus Art. 13 GRCh auch die private sowie die anwendungsorientierte Forschung erfasst sein sollen.[14] Zu beachten ist dabei, dass die Nichteröffnung des Anwendungsbereichs der KI-VO die Bindung an das Unionsrecht im Übrigen unberührt lässt.

Fraglich ist, ob die Bereichsausnahme des Art. 2 Abs. 6 KI-VO die Lehre erfasst. Dagegen spricht zunächst, dass Art. 2 Abs. 6 KI-VO allein von Forschung und Entwicklung spricht, nicht aber von Lehre. Demgegenüber wird in ErwG 25 S. 1 davon gesprochen, dass die KI-VO die Freiheit der Wissenschaft nicht untergraben soll, worunter auch die Freiheit der Lehre verstanden werden kann. Im Übrigen geht aber auch ErwG 25 allein auf die Privilegierung wissenschaftlicher Forschung ein, nicht aber auf die Lehre. Daneben kann auch auf Art. 13 GRCh rekurriert werden, der Forschungsfreiheit und akademische Freiheit zwar in eine enge Verbindung stellt, aber als zwei verschiedene Grundrechte benennt. Des Weiteren spricht auch die Einstufung von Bewertungssystemen für Schüler und Studenten als Hochrisiko-KI-Systemen nach Anhang III Nr. 3 b der KI-VO für ein Verständnis der Bereichsausnahme des Art. 2 Abs. 6 KI-VO, das die Lehre nicht erfasst, da diese ansonsten systemwidrig wäre. Im Ergebnis erfasst daher Art. 2 Abs. 6 KI-VO den Schutz der Lehre nicht. Neben die Privilegierung durch die Bereichsausnahme des Art. 2 Abs. 6 KI-VO tritt die Privilegierung von Forschungs-, Test- und Entwicklungsaktivitäten im Vorfeld des Inverkehrbringens von KI-Systemen. Diese sind nach Art. 2 Abs. 8 S. 1 KI-VO unabhängig vom Akteur, nicht vom Anwendungsbereich der KI-VO erfasst und sind nach Art. 2 Abs. 8 S. 2 KI-VO im Einklang mit dem geltenden Unionsrechts durchzuführen. Gemäß Art 2 Abs. 8 S. 3 KI-VO findet die Regelung keine Anwendung auf Tests unter Realbedingungen. Die Regelung ist dabei primär als Klarstellung zu verstehen, da bereits der Anbieter- und der Betreiberbegriff eine Anwendbarkeit der KI-VO im Vorfeld des Inverkehrbringens ausschließt.

Insgesamt kann die Privilegierung, insbesondere gegenüber dem ursprünglichen Kommissionsentwurf[15], als weitgehend und insgesamt als hinreichend bewertet werden. Insbesondere die weite Privilegierung rein wissenschaftlicher Nutzung von KI-Systemen und KI-Modellen reduziert die Eingriffsdichte der KI-VO hinsichtlich der Wissenschaftsfreiheit nach Art. 13 GRCh auf ein Mindestmaß. Allein dort wo Forschende tatsächlich wie Unternehmer handeln, werden diese weiterhin durch die KI-VO reguliert, was in Anbetracht der Ziele der KI-VO, relevante Risiken einzuhegen, auch gerechtfertigt werden kann.

III. Risikobasierter Ansatz

Die KI-VO verfolgt einen risikobasierten Ansatz. Dabei wird zwischen KI-Systemen mit inakzeptablem Risiko, KI-Systemen mit hohem Risiko, KI-Systemen mit minimalem Risiko und KI-Modellen mit generellem Verwendungszweck unterschieden.

14 Vgl. *Becker*, ZfDR 2023, 164.

15 Hierzu umfassend *Becker*, ZfDR 2023, 164.

1. Verbotene KI-Praktiken

Sogenannte „KI-Praktiken" mit einem inakzeptablen Risiko werden in Art. 5 KI-VO aufgezählt und sind grundsätzlich verboten. Darunter fallen unter anderem manipulative Techniken, die das Verhalten von Personen beeinflussen und zu erheblichen Schäden für diese führen können (Art. 5 Abs. 1 lit. a KI-VO), die KI-basierte Bewertung des sozialen Verhaltens (Social Scoring, Art. 5 Abs. 1 lit. c KI-VO), die Beurteilung des Risikos der (erneuten) Straftatbegehung (Art. 5 Abs. 1 lit. d KI-VO) sowie die KI-basierte biometrische Echtzeit-Fernidentifikation (Art. 5 Abs. 1 lit. h, Abs. 2-7). Für letzteres Verbot bestehen dabei allerdings Ausnahmen zur Aufklärung und Verhütung schwerer Straftaten. Auch bei den anderen Verboten existieren bestimmte Ausnahmen, die die Verbotsnorm etwas verwässern und aufgrund ihrer teilweisen Unbestimmtheit für Rechtsunsicherheit sorgen.

2. Hochrisiko-KI-Systeme

KI-Systeme mit hohem Risiko werden in den Art. 6 ff. KI-VO geregelt und stellen den zentralen Regelungsgegenstand der KI-VO dar. Die Einordnung als Hochrisiko-KI-System hängt dabei von dem Bereich ab, in dem das KI-System eingesetzt werden soll. KI-Systeme, die unter die im Anhang I enthaltenen Unionsrechtsakte fallen, werden gem. Art. 6 Abs. 1 KI-VO als Hochrisiko-KI-Systeme eingeordnet. Der Anhang III enthält zudem eine Liste von Anwendungsbereichen mit hohem Risiko. Für die Wissenschaft und Forschung ist dabei zum einen relevant, dass der Einsatz in einem dieser Bereiche die Anwendbarkeit der Wissenschaftsprivilegierung ausschließt, da das KI-System dann nicht mehr allein Forschungszwecken dient.

Zum anderen enthält Anhang III, wie bereits genannt, einige Anwendungsbereiche, die unmittelbar für Hochschulen relevant sind. Dies sind unter anderem der Einsatz von KI-Systemen zur Zulassung zu Bildungseinrichtungen (Annex III Nr. 3 a)), zur Bewertung von Prüfungsleistungen (Annex III Nr. 3 b)), zur Bewertung des Bildungsniveaus (Annex III Nr. 3 c)) sowie zur Überwachung von Schülern oder Studierenden bei Prüfungen (Annex III Nr. 3 d). Auch der Einsatz von KI bei der Auswahl von (Lehr-)Personal sowie die Leistungsbewertung von Angestellten (Annex III Nr. 4 a) und b)) stellt einen Anwendungsbereich mit hohem Risiko dar, der für Hochschulen im Allgemeinen relevant sein dürfte.

Gem. Art. 9 KI-VO müssen für Hochrisiko-KI-Systeme Risikomanagementsysteme eingerichtet werden, wodurch absehbare Grundrechtsrisiken identifiziert und abgemildert werden sollen. Art. 10 KI-VO schreibt bestimmte Qualitätskriterien für Trainings-, Validierungs- und Testdatensätze vor. Demnach müssen diese Datensätze unter anderem hinreichend relevant, repräsentativ und nach Möglichkeit frei von Fehlern und im Hinblick auf die beabsichtigte Nutzung vollständig und möglichst kontextsensitiv sein und sind auf potenzielle Biases zu untersuchen, die verbotene Diskriminierungen zur Folge haben. Die Vorschrift adressiert damit teilweise das Problem unfairer KI, ohne dabei jedoch konkrete Fairnessmetriken vorzuschreiben.[16] Dies ist allerdings nicht zu beanstanden, da sich verschiedene Fairnessmetriken häufig ausschließen und widersprechen.[17] Ebenfalls kein Nachteil ist, dass keine weitergehenden Vorgaben an die Nichtdiskriminierung von KI-Systemen enthalten sind. Fragen wie etwa die Rechtfertigung algorithmischer Diskriminierungen, sind Gegenstand des europäischen und nationalen Antidiskriminierungsrechts. Die Beantwortung ob ein Verstoß gegen diese nicht-KI-spezfischen Normen vorliegt, hängt vom Einzelfall ab und die Feststellung obliegt letztlich den Gerichten. Auf Ebene einer präventiven Regulierung wie der KI-VO kann der Frage der Verhinderung von Bias dabei nur mit bestimmten Qualitätsanforderungen an die Datensätze begegnet werden. Zu kritisieren ist hier, dass die Anforderungen der KI-VO sehr unbestimmt sind.[18] Wann Datensätze *hinreichend* repräsentativ, relevant und fehlerfrei sind, bleibt weitestgehend unklar und liegt damit im Ermessen der Anbieter.

In Art. 11 KI-VO wird zudem eine Pflicht zur technischen Dokumentation bestimmter Informationen aufgestellt. Anhand dieser soll den Marktüberwachungsbehörden nach Art. 3 Nr. 26 KI-VO gezeigt werden können, dass die Anforderungen an das KI-System eingehal-

16 Vgl. *Feldkamp et al.*, ZfDR 2024, 60, 93. Als Fairnessmetriken bezeichnet man im allgemeinen verschiedene Versuche in der Informatik, KI-Systemen im Trainingsprozess Fairness, das heißt Gleichbehandlung, anhand quantitativer Gleichbehandlungsbegriffe einzuprogrammieren, vgl. *Gesellschaft für Informatik*, Technische und rechtliche Betrachtungen algorithmischer Entscheidungsverfahren. Studien und Gutachten im Auftrag des Sachverständigenrats für Verbraucherfragen, 2018, S. 39; *Verma/Rubin*, Fairness Definitions Explained, 2018 ACM/IEEE International Workshop on Software Fairness, 2018, S. 1; *Mehrabi et al.*, A Survey on Bias and Fairness in Machine Learning, ACM Computing Surveys, Vol. 54(6), 2021, Article 115, S. 11.

17 Vgl. etwa *Berk et al.*, Fairness in Criminal Justice Risk Assessments: The State of the Art, Sociological Methods & Research 50(I) (2021), 3 ff.

18 Zu dieser Kritik im Allgemeinen auch *Bomhard/Siglmüller*, RDi 2024, 45, 54.

ten werden. Nach Anhang IV müssen unter anderem der Zweck des Systems, dessen Funktionsweise, die Einsatzmöglichkeiten, die Hardwareanforderungen, das User-Interface, dessen Entscheidungslogik und zentrale Annahmen, die zentralen Klassifikationsentscheidungen, die zu optimierenden Parameter, mögliche Kompromisse in Bezug auf die technischen Lösungen, für die jeweiligen Anforderungen der KI-VO, die Qualität der Inputdaten, verwendete Leistungskennzahlen d.h. Metriken zur Beurteilung der Leistung des KI-Systems[19], das erwartete Maß an Genauigkeit und etwaige Diskriminierungsrisiken dokumentiert werden.

Zudem müssen bestimmte Ereignisse gem. Art. 12 KI-VO automatisch aufgezeichnet werden. Dies soll die Identifikation von Risiken nach Art. 79 Abs. 1 KI-VO sowie die Überwachung nach Markteinführung ermöglichen und ein hinreichendes Niveau an Rückverfolgbarkeit der Funktionsweise gewährleisten.

Nach Art. 13 Abs.1 KI-VO müssen Hochrisiko-KI-Systeme zudem hinreichend transparent sein, um Betreiber in die Lage zu versetzen, den Output des Systems zu interpretieren und das System richtig zu verwenden. Betreibern ist zu diesen Zwecken eine Gebrauchsanweisung zu überlassen. Unklar bleibt jedoch, wie die Begriffe Transparenz und Interpretation zu verstehen sind.

Ergänzt wird die Pflicht in Art. 13 KI-VO durch das in Art. 86 KI-VO enthaltenen Recht auf Erläuterung individueller Entscheidungsfindung für betroffene Personen. Demnach sind Personen, die von einer Entscheidung betroffen sind, die mithilfe eines Hochrisiko-KI-Systems getroffen wurde und auf sie rechtliche oder ähnlich beeinträchtigende Auswirkungen hat, berechtigt, vom Betreiber eine klare und aussagekräftige Erklärung über die wichtigsten Elemente der Entscheidung und die Rolle des KI-Systems zu verlangen. Dies könnte für Hochschulen insofern relevant werden, dass diese in den oben genannten Hochrisiko-Anwendungen, etwa bei der Bewertung von Prüfungsergebnissen, wenn dabei KI-Systeme zum Einsatz kommen, die wichtigsten Elemente erklären müssen. Art. 86 KI-VO ist jedoch ebenfalls inhaltlich zu unbestimmt. So wird nicht klar, worauf sich die "Elemente" einer Entscheidung beziehen. ErwG 171 stellt lediglich klar, dass die Erklärung klar und aussagekräftig sein soll und dadurch eine Grundlage geschaffen werden soll, die betroffene Personen in die Lage versetzt, ihre Rechte wahrzunehmen.

Des Weiteren müssen KI-Systeme mit hohem Risiko gem. Art. 14 KI-VO unter menschlicher Aufsicht stehen. Eine zuständige natürliche Person muss in der Lage sein, Fehlfunktionen zu erkennen, den Output korrekt zu interpretieren und einen „Stopp"-Knopf zu drücken. Übermäßiges Vertrauen in den Output soll verhindert werden.

Hochrisiko-KI-Systeme müssen schließlich gem. Art. 15 KI-VO über ein gewisses Maß an Genauigkeit, Robustheit und Cybersicherheit verfügen. Die Beeinflussung des Systems durch vergangene verzerrte Outputs (sog. Feedback-Loops) soll nach Möglichkeit verhindert werden.

Zudem wird in Art. 27 KI-VO eine Grundrechtefolgenabschätzung vorgeschrieben. Diese Pflicht gilt grundsätzlich nur für öffentlich-rechtliche Betreiber, worunter jedenfalls öffentlich-rechtliche Bildungseinrichtungen, wie etwa staatlichen Hochschulen, zu subsumieren sind. Diese Akteure sind verpflichtet, eine Abschätzung der Auswirkungen durchzuführen, die die Verwendung des jeweiligen Hochrisiko-KI-Systems für die Grundrechte haben kann. Diese Grundrechtefolgenabschätzung umfasst unter anderem die Kategorien der betroffenen Personen, die spezifischen Schadensrisiken und eine Beschreibung der Umsetzung von Maßnahmen der menschlichen Aufsicht. Unklar ist allerdings der konkrete Mehrwert der Grundrechtefolgenabschätzung gegenüber der inhaltlich weitgehend gleichlaufenden Verpflichtung zur Errichtung eines Risikomanagementsystems in Art. 9 KI-VO.[20] Sofern Hochschulen außerhalb der Forschungsprivilegierung nach Art. 2 Abs. 6 und 8 KI-VO - wie sie zuvor beschrieben wurde - KI-Systeme als Betreiber in einem Hochrisikobereich einsetzen oder Hochrisiko-KI-Systeme auf den Markt bringen, sind sie umfassend an die Vorgaben der KI-VO gebunden. Dies ist vor dem Hintergrund, dass sie in diesem Fall auch vergleichbar zu klassischen Anbietern und Betreibern – in der Regel private Unternehmen – tätig würden, auch gerechtfertigt.

19 *Gesellschaft für Informatik*, Technische und rechtliche Betrachtungen algorithmischer Entscheidungsverfahren. Studien und Gutachten im Auftrag des Sachverständigenrats für Verbraucherfragen, 2018, S. 38.

20 Hierzu *Hacker/Berz*, ZRP 2023, 226, 228.

3. KI-Systeme mit minimalem Risiko und Transparenzrisiko

KI-Systeme mit minimalem Risiko werden von der KI-VO grundsätzlich nicht reguliert. Eine Ausnahme hiervon stellen KI-Systeme dar, die direkt mit natürlichen Personen interagieren. Diese müssen gem. Art. 50 KI-VO als solche gekennzeichnet werden, sofern dies nicht offensichtlich ist. Automatisch generierte Inhalte wie Videos, Bilder oder Texte sowie sogenannte Deep Fakes müssen ebenfalls gekennzeichnet werden, auch wenn sie grds. nicht als Hochrisiko-KI-System eingeordnet werden.

Sonstige KI-Systeme, die weder ein hohes Risiko noch ein Transparenzrisiko aufweisen, bleiben jedoch unreguliert. Während der Parlamentsentwurf der VO noch im operativen Teil einen Katalog allgemeiner Prinzipien wie z.B. Fairness und Accountability enthielt, der für alle KI-Systeme gelten sollte,[21] wurde dieser nun in ErwG 27 und damit in den nicht rechtsverbindlichen Teil der Verordnung verschoben. Diesen abstrakten und unbestimmten Vorgaben wird in der Rechtspraxis wohl kein normativer Wert zukommen.

4. KI-Modelle mit generellem Verwendungszweck

Für viel Aufsehen und Diskussion hat letztlich auch die seit dem Ratsentwurf eingefügte Regulierung von KI-Systemen mit generellem Verwendungszweck gesorgt (kurz GPAI-Modelle, Abkürzung für den im Englischen verwendeten Begriff "general purpose AI Model").

Der Begriff wird in Art. 3 Nr. 63 KI-VO definiert als "KI-Modell […] 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". Unklar bleibt dabei, was unter dem Begriff des Modells zu verstehen ist und inwiefern sich dieser von dem Begriff des KI-Systems unterscheidet.[22] Darunter dürften jedenfalls große Sprachmodelle (engl. Large Language Models, kurz LLMs) wie ChatGPT oder Google Gemini sein.

Dieser "gestufte" Ansatz ist das Ergebnis eines politischen Kompromisses, dessen Zustandekommen das Scheitern der gesamten KI-VO verhindert hat, wobei das Ziel verfolgt wurde, kleinere und mittlere Unternehmen sowie Startups zu privilegieren.[23] Dieses Ziel mag zwar nachvollziehbar sein, ändert jedoch nichts daran, dass auch von kleineren GPAI-Modellen Grundrechtsrisiken ausgehen können.[24] Überzeugender wäre es daher gewesen, auf den gestuften Ansatz zu verzichten und stattdessen alle GPAI-Modelle auf bestimmte mit diesen einhergehenden Risiken zu bewerten.

Die komplexen Fragen, die den Einsatz von GPAI-Modellen an Hochschulen, etwa im Rahmen der Lehre oder durch Studierende bei der Erstellung von Seminar- und Hausarbeiten betreffen, werden durch die KI-VO nicht beantwortet. GPAI-Modelle werden vielmehr nur hinsichtlich des Zeitpunktes reguliert, an dem ihre konkrete Verwendung noch gar nicht feststeht. Inwiefern GPAI-Modelle etwa zulässige Hilfsmittel für Studierende darstellen, wird daher nicht geregelt. Sofern ein GPAI-Modell in den oben genannten Risikobereichen, etwa zur Bewertung und Benotung von Studierenden eingesetzt wird, ist es als Risiko-KI-System einzustufen. Sodann gelten jedoch die Regeln zu Hochrisiko-KI-Systemen. Wenn universitäre Einrichtungen GPAI-Modelle zu Forschungszwecken entwickeln und einsetzen, greift die Forschungsprivilegierung in Art. 2 Abs. 6 KI-VO. Gem. Art. 2 Abs. 8 KI-VO gilt die KI-VO auch nicht für Forschungs-, Test- und Entwicklungstätigkeiten zu GPAI-Modellen vor Inbetriebnahme oder Inverkehrbringen. Dies entbindet freilich Forschende und Entwickelnde nicht von der Pflicht, in diesem Stadium Grundrechte und einschlägiges Unionsrecht zu beachten.
Wollen Forschende GPAI-Modelle auf dem Markt anbieten oder in Betrieb nehmen, ohne ausschließliche Forschungszwecke zu verfolgen, so gelten sie als Anbieter i.S.v. Art. 2 Abs. 1 KI-VO. Sie müssen dann auch die speziell GPAI-Modelle betreffenden Regeln beachten. Diese finden sich vor allem in den Art. 51 ff. KI-VO. Dabei wird zwischen Modellen mit und ohne systemisches Risiko unterschieden. Um ein GPAI-Modell mit systemischem Risiko handelt es sich gem. Art. 51 Abs. 1 KI-VO, wenn dieses Fähigkeiten mit hoher Wirkkraft („high impact capabilities") besitzt. Diese wird gem. Art. 51 Abs. 2 KI-VO vermutet bei einer Leistungsfähigkeit von 10^{25} FLOPs. Diese Maßeinheit bezieht sich gem. Art. 3 Nr. 67 KI-VO auf die Zahl der Gleitkommazahl-Operationen,

21 Hierzu *Becker/Feuerstack*, MMR 2024, 22, 22.
22 *Bomhard/Siglmüller*, RDi 2024, 45, 50.
23 Hierzu *Frisch/Kohpeis*, ZD-Aktuell 2023, 01318.
24 *Martini/Wiesehöfer*, NVwZ – Online-Aufsatz 1/2024, 14.
25 *Hacker/Berz*, ZRP 2023, 226, 228.

die pro Sekunde ausgeführt werden können. Für alle GPAI-Modelle ungeachtet ihrer Einordnung gelten zunächst die Anforderungen des Art. 53 KI-VO. Demnach sind Anbieter zur technischen Dokumentation verpflichtet, die unter anderem die Aufgaben, die das Modell bewältigen soll, die Architektur und die Zahl der Parameter, die Lizenz, Trainingsmethoden und -techniken, Informationen über die verarbeiteten Daten sowie die Zahl der FLOPs enthalten muss. Nachgelagerten Anbietern müssen gem. Art. 53 Abs. 1 lit. b KI-VO Informationen über das Modell mitgeteilt werden, die diese in die Lage versetzen sollen, die Fähigkeiten und Limitationen des GPAI-Modells zu verstehen. Dies gilt gem. Art. 53 Abs. 2 KI-VO nicht für Modelle, die unter einer freien Lizenz verfügbar sind, sofern diesen kein systemisches Risiko innewohnt. Anbieter müssen zudem eine Policy zur Einhaltung des EU-Urheberrechts einführen und hinreichend detaillierte Zusammenfassungen der genutzten Trainingsdaten öffentlich zur Verfügung stellen.

Für GPAI-Modelle mit systemischem Risiko gilt darüber hinaus Art. 55 KI-VO. Dieser schreibt etwa eine Modellbewertung nach anerkannten technischen Standards vor, um systemische Risiken zu bewerten und abzumildern. Auch sind ernste Vorfälle, einschließlich möglicher Korrekturmaßnahmen zu dokumentieren und dem AI Office zu melden, und ein hinreichendes Maß an Cybersicherheit muss gewährleistet werden.

Synthetische Text-, Bild-, Video- und Audioinhalte, die von GPAI-Modellen generiert wurden, müssen zudem nach Art. 50 Abs. 2 KI-VO in computerlesbarem Format als künstlich generiert gekennzeichnet sein. Die Pflicht zur Modellbewertung nach Art. 55 Abs. 1 lit. a KI-VO erscheint angemessen, da nur nach erfolgter Modellbewertung eine tatsächliche Risikoeinschätzung durch den Anbieter des GPAI-Modells tatsächlich erfolgen kann.

Zu kritisieren ist jedoch die unterschiedliche und widersprüchliche Verwendung des Begriffes des systemischen Risikos: einerseits beschreibt der Begriff GPAI-Modelle mit Fähigkeiten mit hoher Wirkkraft, wodurch die strengen Anforderungen des Art. 55 KI-VO zur Anwendung gelangen. Gleichzeitig wird im Zusammenhang mit der Bewertungspflicht aus Art. 55 KI-VO das Begriffsverständnis aus den Art. 34 f. DSA zugrunde gelegt, wonach es insbesondere auf nachteilige Auswirkungen für bestimmte Rechtsgüter ankommt. In der Folge sind Anbieter von GPAI-Modellen, die aufgrund ihrer hohen FLOP-Zahl ein systemisches Risiko aufweisen, verpflichtet, das systemische Risiko nochmals zu identifizieren und zu bewerten und sodann abzumildern.

Es stellt sich dann die Frage, ob dies bedeutet, dass Anbieter die Rechenleistung reduzieren müssen. Unklar ist auch, inwiefern sich systemische Risiken auf abstrakter Ebene, ohne dass sie das konkrete künftige Anwendungsfeld kennen, überhaupt bewerten und abmildern lassen.[25] Da sich systemische Risiken für Grund- und Menschenrechte bei allen GPAI-Modellen, ungeachtet der FLOP-Zahl, stellen können,[26] da auch kleine GPAI-Modelle grundrechtsrelevante Bereiche betreffen können, ist der gestufte Ansatz zudem wenig überzeugend.

IV. Fazit

Die KI-VO enthält an verschiedenen Stellen Vorgaben für den Einsatz von KI in Wissenschaft und Forschung sowie in der Lehre an Hochschulen. Zwar gelten für „reine" Forschungs- und Entwicklungstätigkeiten zunächst die inzwischen umfassenden und großzügigen Ausnahmen vom Anwendungsbereich der KI-VO. Der Einsatz von KI-Systemen, bspw. bei der Einstellung von Lehrkräften und Angestellten sowie bei der Zulassung und Bewertung von Studierenden führt jedoch zur Anwendung der Vorschriften über Hochrisiko-KI-Systeme. In diesen Bereichen dürfen Hochschulen nur KI-Systeme einsetzen, die den Anforderungen der KI-VO entsprechen.

Die komplexen Fragen, die sich seit der Popularität von ChatGPT im Zusammenhang mit dem Einsatz von GPAI-Modellen an Hochschulen stellen, werden von der KI-VO nicht beantwortet, weshalb sie weiterhin Anlass zur Debatte geben. Die umfassenden, gestuften Regeln zu GPAI-Modellen betreffen vielmehr ein Stadium, in dem der konkrete Einsatz der Modelle noch nicht unbedingt feststeht. Sofern die Forschungsprivilegierung nicht greift und sie als Anbieter einzustufen sind, müssen auch Hochschulen die Vorschriften zu GPAI-Modellen beachten.

Aufgrund der doppeldeutigen Definition des systemischen Risikos und dem wenig überzeugenden gestuf-

26 So auch *Martini/Wiesehöfer*, NVwZ – Online-Aufsatz 1/2024, 14.

ten Ansatz muss die Regulierung von GPAI-Modellen insgesamt als misslungen bewertet werden. Auch die generell inflationäre Verwendung von unbestimmten Rechtsbegriffen, mit der erhebliche Rechtsunsicherheit einhergehen kann, ist negativ zu bewerten.[27] Die Kommission ist diesbezüglich gehalten, ihrer Aufgabe zur Erarbeitung von Leitlinien für die praktische Umsetzung gem. Art. 96 KI-VO möglichst noch vor Geltung der jeweiligen Vorschriften[28] nachzukommen.

Positiv zu bewerten ist jedoch zunächst der seit Beginn des Gesetzgebungsprozesses verfolgte risikobasierte Ansatz der KI-Regulierung. Daneben überzeugt auch die Angleichung der Definition von KI-Systemen in der KI-VO mit der Definition der OECD. Schließlich ist auch die im Vergleich zum Kommisionsentwurf deutlich erweiterte und praxistaugliche Privilegierung der Forschung in der finalen Fassung der KI-VO zu begrüßen.

Daniel Becker ist Rechtsreferendar am Landgericht Freiburg.

Daniel Feuerstack ist akademischer Mitarbeiter am Institut für öffentliches Recht (Abt II: Völkerrecht, Rechtsvergleichung) der Universität Freiburg. Er ist dort tätig in den interdisziplinären Forschungsnetzwerken ReScale (gefördert von der Carl-Zeiss-Stiftung) und Adaptive Governance of Emerging Technologies (AdGovEm) der Universität Freiburg.

27 *Bomhard/Siglmüller*, RDi, 45, 54.

28 Vgl. hierzu Art. 113 KI-VO, der grds. eine Umsetzungsfrist von 24 Monaten ab Inkrafttreten nennt, jedoch für verschiedene Vorschriften auch kürzere oder längere Umsetzungsfristen enthält.