Neue Zeitschrift für Verwaltungsrecht – Online-Aufsatz

In Zusammenarbeit mit der Neuen Juristischen Wochenschrift

Gegründet von Rechtsanwalt **Prof. Dr. Hermann Weber**, Schriftleiter von 1982–2001

1 2024
Seite 1–19
43. Jahrgang
1. Februar 2024

Professor Dr. Mario Martini und Wiss. Mitarbeiterin Christine Wiesehöfer*

Auf dem Weg zur Regulierung von General-Purpose-AI – eine erste Bestandsaufnahme und Kritik der Regelungsentwürfe

Der Jahreswechsel 2022/2023 wird womöglich als Zeitenwende Künstlicher Intelligenz in die Annalen eingehen. So wie einst die Dampfmaschine oder die Eisenbahn als Basisinnovationen die Produktion von Gütern bzw. das Verkehrswesen grundlegend umwälzten, machen nunmehr generative Anwendungen wie ChatGPT oder der Bildgenerator Midjourney KI-Technologien einer breiten Öffentlichkeit zugänglich. Ähnlich wie frühere technologische Umbrüche bergen aber auch sie gravierende Risiken. Deshalb ist weltweit eine intensive Debatte darüber entbrannt, wie die Rechtsordnung den Schattenseiten der sog. Basismodelle (Foundation Models, General-Purpose-AI-Modelle) begegnen sollte, auf denen viele generative KI-Systeme aufbauen. Die Europäische Union wird als Pionierin auf dem weitgehend unerschlossenen Terrain der KI-Regulierung das erste umfassende Regelwerk vorlegen. Die Schemen der normativen Vorgaben für Foundation Models zeichnen sich bereits politisch ab. Vieles ist aber noch unklar, insbes. liegt eine finale Textfassung noch nicht vor. Der Beitrag bewertet die Regulierungskonzepte der drei KI-VO-Entwürfe und entwickelt eigene Regelungsvorschläge für die Zukunft generativer KI.

I. Foundation Models und die Zukunft der KI

1. Bedeutung

In gleichem Maße, in dem Künstliche Intelligenz die Phantasie der Menschen beflügelt, polarisiert sie auch. Optimisten erkennen in ihr den größten Segen der Menschheitsgeschichte, Skeptiker hingegen ein Auslöschungsrisiko für die Menschheit insgesamt.[1] So warnte bspw. der KI-Pionier Geoffrey Hinton: „Diese Dinger sind völlig anders als wir. Manchmal denke ich, es ist, als wären Außerirdische gelandet und die Menschen hätten es nicht bemerkt, weil sie sehr gut Englisch sprechen".[2]

Der KI-Sektor erlebt schon seit einigen Jahren einen wahren Hype. Sog. *Foundation Models*[3] verleihen ihm wie ein Turbolader zusätzliche Schubkraft. Obgleich sie erst in jüngerer Zeit in den Fokus einer breiten Öffentlichkeit gerückt sind,[4] gelten sie bereits heute als Meilenstein der Technikgeschichte. Manche erkennen in ihnen gar einen Vorläufer Allgemeiner Künstlicher Intelligenz, die eigenständig strategische Entscheidungen trifft.

Besondere mediale Aufmerksamkeit erlangte zuletzt ChatGPT. Der Chatbot des Unternehmens *OpenAI* ist aber nur ein prominentes Beispiel für das dynamisch wachsende Feld unterschiedlicher generativer KI-Systeme, die derzeit im Mittelpunkt der Debatte über die künftige Rolle Künstlicher Intelligenz im gesellschaftlichen und individuellen Leben stehen.

2. Funktionsweise

Generative Chatbots basieren in der Regel auf sog. großen Sprachmodellen (*Large Language Models*).[5] Diese Modelle zeichnen sich dadurch aus, gewaltige Datenmengen anhand

* Mario Martini ist Lehrstuhlinhaber an der DUV Speyer und Leiter des Programmbereichs „Digitale Transformation im Rechtsstaat" am Deutschen Forschungsinstitut für öffentliche Verwaltung. Christine Wiesehöfer ist dort Forschungsreferentin. Die Autoren danken ganz besonders dem Verbundkoordinator des Programmbereichs Martin Feldhaus sowie Leonard Nalbantis, Katja Neumann und Michael Reichenthaler für die sehr gute Mitwirkung. Der Beitrag ist auf dem Stand vom 18.12.2023.

1 Vgl. etwa Barten/Meindertsma, An AI Pause Is Humanity's Best Bet For Preventing Extinction, Time.com v. 20.7.2023; Gebru et al., Statement from the listed authors of Stochastic Parrots on the "AI pause" letter, 31.3.2023.

2 Heaven, Geoffrey Hinton tells us why he's now scared of the tech he helped build, MIT Technology Review v. 2.5.2023.

3 Viele der Modelle sind auch unter den Begriffen „(große) Sprachmodelle" (Engl.: „Large Language Models") oder „KI-Systeme mit allgemeinem Verwendungszweck" (Engl.: general purpose artificial intelligence systems („GPAI-Systeme")) bekannt, wobei sich die Termini trotz vieler Überschneidungen in ihrer Bedeutung nicht entsprechen.

4 Bommasani et al., On the Opportunities and Risks of Foundation Models, Center for Research on Foundation Models (CRFM), Stanford Institute for Human-Centered Artificial Intelligence (HAI) at Stanford University, 2022, S. 4.

5 ChatGPT baut ebenfalls auf einem Sprachmodell auf (derzeit GPT-3.5 bzw. GPT-4 bei zahlenden Kunden), vgl. https://openai.com/blog/chatgpt.

sehr vieler Parameter auswerten zu können. Indem sie Wort- bzw. Satzmuster erkennen, sind sie in der Lage, die Wahrscheinlichkeit für die jeweils passenden Folgewörter oder Textbausteine (genauer die Tokens) auf der Basis erlernter Muster anhand statistischer Hochrechnungen zu bestimmen – und die entsprechenden Tokens sodann auszugeben. Obgleich es sich bei Basismodellen – vergröbert ausgedrückt – um „stochastische Papageien"[6] handelt, die kein Verständnis von Sprache haben, vermitteln sie Nutzern den Eindruck, menschenähnlich kommunizieren zu können. Die auf ihnen aufsetzenden Systeme können nicht nur Texte erkennen, erstellen oder übersetzen und selbst komplexe (Fach-)Fragen (obgleich teilw. noch mit hoher Fehlerquote) beantworten sowie bspw. Essays oder Zeitungsartikel generieren.[7] Sie lassen sich auch nutzen, um etwa komplexe ökonomische Zusammenhänge zu modellieren oder sogar Musikstücke zu komponieren. Aufgrund ihrer Lern- und Anpassungsfähigkeit lassen sich die Modelle zudem in Bereichen einsetzen, für die sie ursprünglich gar nicht konzipiert waren oder in denen sie kein Training durchlaufen haben – und können mitunter selbst unvorhergesehene Aufgaben bewältigen (daher auch die Bezeichnung „general purpose"). Bereits das Sprachmodell GPT-3 war zu „in-context learning" imstande: Schon ein sog. Prompt (eine Beschreibung der Aufgabe in natürlicher Sprache) reichte aus, um es an eine nachgelagerte Aufgabe anzupassen. KI-Anwendungen, die Texte erkennen und generieren sollten, lernten zB zu programmieren (obgleich bisweilen mit einer nicht unerheblichen Fehlerquote). Zugleich zeichnet sich ab, dass die Systeme nicht zwingend auf die eingespeiste Datenbasis beschränkt sein müssen. So ist GPT-4 mittlerweile (mithilfe von Drittanbieter-Plug-Ins) in der Lage, bspw. Einkäufe oder Urlaubsbuchungen vorzunehmen sowie mittels Websuche Informationen zu tagesaktuellen Ereignissen zu liefern.

Kein Wunder also, dass unterdessen eine Goldgräberstimmung um sich greift. Unternehmen weltweit werben gegenwärtig Milliardensummen ein, um sich im globalen Rennen um die leistungsstärksten Foundation Models und die KI-Vorherrschaft einen günstigen Startplatz zu verschaffen. Das französische Start-up *Mistral AI* sammelte bspw. in einer ersten Finanzierungsrunde bereits 105 Mio. Euro ein,[8] zuletzt gar 385 Mio. Euro.[9] US-amerikanische Unternehmen streichen noch viel größere Summen ein.[10]

3. Technische Annäherung

Foundation Models bauen auf lernenden Systemen[11] und neuronalen Netzen[12] auf, die sich kontinuierlich weiterentwickeln. Sie formen zwar keine homogene, klar umrissene Gruppe. Einige grundlegende technische Gemeinsamkeiten lassen sich aber ausmachen: Ebenso wie bei vielen anderen KI-Anwendungen – aber in signifikant höherem Ausmaß – gründen ihre Stärken auf Transferlernen und Skalierung. Sie können Lernergebnisse nicht nur auf andere Aufgaben übertragen (*emergence*), sondern diese darüber hinaus (angepasst) anwenden (sowohl als *few-shot* als auch als *zero-shot learner*). Ihre Fähigkeiten sind mithin so breit gefächert, dass eine Vielzahl nachgelagerter Anwender dasselbe Modell zu spezifischen neuen Zwecken als Werkzeug nutzen können (*homogenization*[13]).

Neben Chatbots entwickeln sich Systeme, wie zB DALL·E 2 des Unternehmens *OpenAI*, die im Bereich der „Computer Vision", also Bilderkennung und Bildgenerierung, operieren. Sie setzen ebenfalls auf Prompts auf.

Andere Systeme, wie zB DeepMinds *MuZero*, erlernen während des Spielvorgangs selbstständig die Regeln vielfältiger Spiele[14] oder assistieren in unterschiedlichen Szenarien, wie bspw. DeepMinds *Gato*, ein KI-Agent, der über 600 verschiedene Aufgaben (unterschiedlich gut) beherrscht; er soll neben Texten und Bildern u.a. auch Anweisungen für die Bewegungen eines Roboters ausgeben können.[15]

II. Nutzen und Risiken von Foundation Models

1. Potenziale

Dass GPAI-Modelle in den Fokus der öffentlichen Aufmerksamkeit gerückt sind, gründet insbes. darauf, dass sie eine breite Automatisierung verschiedener Sektoren entscheidend beflügeln können. Fast alle Bereiche der Wirtschafts- und Arbeitswelt werden in Zukunft von ihren Potenzialen nachhaltig profitieren.

So verheißen leistungsfähige Basismodelle etwa im Medizinsektor durchschlagende Effizienz- und Qualitätssteigerungen. Sie können zwar einen ausgebildeten Arzt nicht ersetzen, diesen aber immerhin spürbar entlasten. Auf ihnen aufsetzende Anwendungen sind in der Lage, aufwändige Routineaufgaben zu übernehmen, zB die Patientendokumentation vorzubereiten, aber auch im Vorfeld eines Arztbesuches mit den Patienten zu kommunizieren, um ihre Symptome zu erfragen und auf dieser Grundlage mögliche Krankheitsbilder einzugrenzen. *Google* arbeitet bspw. bereits daran, ein Sprachmodell zu optimieren, das auf den medizinischen Bereich spezialisiert ist (Med-PaLM bzw. Med-PaLM-2).[16] Es soll „qualitativ hochwertige Antworten auf medizinische Fragen geben" können.[17] Manche prognostizieren gar, dass die Modelle künftig über „fortgeschrittene Fähigkeiten zu medizinischem Denken" verfügen werden, und etwa „frei formulierte Erklärungen, mündliche Empfehlungen oder Bildanmerkungen" ausgeben können.[18]

Auch bspw. das Bildungssystem und die öffentliche Verwaltung werden Foundation Models voraussichtlich erheblich umwälzen. Auf ihnen basierende Anwendungen werden ganz verschiedene, individuelle Anliegen unterstützen können – indem sie zB Bewerbungen, Kurzzusammenfassungen oder Lösungsentwürfe für verschiedene Aufgabenstellungen formulieren.

6 Bender et al., On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?, 2021, S. 610 (616 f.).
7 Vgl. zB The Guardian, Sample, ChatGPT: what can the extraordinary artificial intelligence chatbot do?, The Guardian v. 13.1.2023.
8 Bspw. Holzki, Mistral AI – 105 Mio. Euro für kaum mehr als eine Idee, Handelsblatt online v. 16.6.2023.
9 Bomke/Holzki, Pariser KI-Start-up sammelt 385 Mio. Euro ein, Handelsblatt online v. 11.12.2023.
10 Allein Microsoft unterstütze OpenAI 2023 mit einer Fördersumme von wohl mehreren Mrd. US-Dollar, vgl. bspw. Tagesschau, Milliarden für die Künstliche Intelligenz, tagesschau online v. 23.1.2023.
11 Maschinelles Lernen zeichnet sich v.a. dadurch aus, „Vorhersagemodelle auf der Grundlage historischer Daten (zu) trainieren und zur Erstellung von Zukunftsprognosen (zu) verwende(n)." Bommasani et al., On the Opportunities and Risks of Foundation Models, S. 3.
12 Dazu bspw. Martini, Blackbox Algorithmus, 2019, S. 24.
13 „Homogenisierung bedeutet die Konsolidierung von Methoden zur Entwicklung von Systemen, die maschinell lernen, in einem breiten Spektrum von Anwendungen; (…)", Bommasani et al., On the Opportunities and Risks of Foundation Models, S. 3.
14 Vgl. bspw. Schrittwieser et al., Mastering Atari, Go, chess and shogi by planning with a learned model, Nature 2020, 604.
15 Vgl. Reed et al., A Generalist Agent, Transactions on Machine Learning Research (11/2022), 2022, 7.
16 Matias/Corrado, Our latest health AI research updates, Google The Keyword Blog v. 14.3.2023.
17 Vgl. Introduction, Med-Palm, Google, https://sites.research.google/med-palm/.
18 Moor et al., Foundation models for generalist medical artificial intelligence, Nature 2023, 259.

2. Risiken

Ihres herausragenden Potenzials zum Trotz bergen Foundation Models substanzielle Risiken. Deren Bandbreite reicht von unvorhersehbaren Fehlern bis hin zu bislang nicht identifizierten Sicherheitsrisiken.

a) Unberechenbare Fähigkeiten

Große Sprachmodelle werfen zwar auf Knopfdruck Texte zu verschiedenen (Sach-)Themen in ansprechender sprachlicher Qualität aus. Komplexe logische Fragen in die Tiefe zu denken, gelingt ihnen indes weniger gut. GPT-3 war zB nicht in der Lage, Rechenaufgaben mit hohen Zahlenwerten zu lösen – sehr wohl aber dazu, einen Programmcode zu erstellen, der eben jene Kalkulation ausführt.[19]

So dynamisch, wie sich die Modelle derzeit entwickeln, lässt sich kaum prognostizieren, über welche Fähigkeiten sie alsbald verfügen werden, wie sie also arbeiten/lernen/entscheiden und sich selbst fortentwickeln. Sie hinreichend zu regulieren, heißt daher zwangsläufig, nicht nur den heutigen Status quo in Rechnung zu stellen, sondern auch zukünftige technologische Fortschritte zu antizipieren. Setzt der Unionsgesetzgeber nicht frühzeitig rechtliche Leitplanken, läuft er Gefahr, die Entwicklung der Modelle später nicht mehr ohne Weiteres einfangen zu können – nicht zuletzt, da die ihnen anhaftenden Risiken auf alle Anwendungsfelder und -bereiche ausstrahlen, in denen sie als Grundlage anderer KI-Systeme dienen. Dass die zukünftigen Einsatzbereiche sich ebenfalls noch nicht konkret herauskristallisieren, reißt naturgemäß Wissenslücken, die es wesentlich erschweren, einen passgenauen regulatorischen Rahmen zu zeichnen.

b) Automatisierte Desinformation und Halluzinationen

Spätestens seit den US-Präsidentschaftswahlen im Jahre 2016 steht der Welt deutlich vor Augen, wie wirkmächtig „Alternative Facts“ und „Fake News“ als Waffen in der politischen Auseinandersetzung sein können. Foundation Models lassen sich hervorragend einsetzen, um gezielt Desinformationen großflächig zu verbreiten – besonders wenn Nutzer Chatbots als Suchmaschine verwenden.

Sprachmodelle neigen überdies zu sog. „Halluzinationen“: Sie erfinden mitunter Informationen und geben diese als objektive Fakten aus. Für Schlagzeilen sorgte bspw. der auf einem Sprachmodell basierende *Google*-Chatbot „Bard“. Er lieferte bereits in seiner ersten Werbepräsentation eine inkorrekte Antwort auf eine Frage zum James-Webb-Weltraumteleskop.[20]

c) Gefährdung der Rechte der Schöpfer kreativer Inhalte

Generative KI hat das Wissen, das sie darbietet, nicht selbst erschaffen. Sie bedient sich vielmehr bei der schöpferischen Kraft der besten kreativen Köpfe, indem sie deren Erkenntnisse neu arrangiert und zusammenführt.

Unbemerkt bekannte Persönlichkeiten täuschend echt zu imitieren, überfordert die Systeme keineswegs. Nicht nur Musiker, Schauspieler, Schriftsteller und Drehbuchautoren, sondern auch Wissenschaftler und Verlage befürchten, dass generative KI-Anwendungen ihre wissenschaftliche oder künstlerische Leistung – ohne Quellenhinweis und ohne Vergütung – abschöpfen und damit Raubbau an ihren kreativen Verdiensten betreiben. Die Hoffnungen, dies zu verhindern, ruhen derzeit auf digitalen Wasserzeichen[21] und Transparenzpflichten. Einige Plattformbetreiber haben bereits reagiert: *YouTube* hat bspw. eigene Regeln für KI-generierte Videos aus der Taufe gehoben.[22] Diese unterliegen nunmehr einer gesonderten Kennzeichnungspflicht.

d) Diskriminierungen

Sprachmodelle operieren mit Datensätzen, die ihrem Wesen nach die soziale Realität abbilden. Sie nehmen in ihren Informationsstaubsauger deshalb zB auch homophobe, rassistische, antisemitische oder islamophobe Wertungen auf, die mitunter in ihren Trainingsdaten stecken. Diese benachteiligenden Muster geben die Modelle dann typischerweise ungefiltert an die auf ihnen basierenden nachgelagerten KI-Systeme weiter. Auf diese Weise spiegeln sie tradierte Muster sozialer Anerkennung und gesellschaftlich verankerte diskriminierende Stereotype – und können so einzelne Gruppen systematisch benachteiligen.

Damit kann sich zugleich ein sog. „value lock-in“ verbinden: Sofern Modelle Trainingsdaten dauerhaft verwenden, verfestigen sich die (u.a. diskriminierenden) Wertungen, welche die eingespeisten Daten widerspiegeln. Schließlich entstammen die Trainingsdaten denklogisch immer der Vergangenheit.

e) Risiken für den Schutz personenbezogener Daten

(Trainings-)Datensätze enthalten mitunter vertrauliche Informationen, welche die Systeme auf gezielte Nutzernachfrage hin ungefiltert preisgeben. Sofern zB politische oder militärische Daten in die Hände Unbefugter gelangen, gehen davon substanzielle Gefahren aus. Im schlimmsten Fall geben die Systeme auch solche Informationen aus, die den Kernbereich privater Lebensgestaltung betreffen.[23]

Schon mit Blick auf die unübersehbare Menge an Daten, welche in die Modelle einfließen, ist eine individuelle Einwilligung jedes Betroffenen zur Datenverarbeitung keine realistische Option. Der große Trainingsdatenumfang – gepaart mit der Datenbeschaffung aus frei zugänglichen Internetquellen – wird es zudem mit sich bringen, dass die Modelle mitunter auch besonders sensible personenbezogene Daten (zB solche, die Rückschlüsse auf die politische Meinung zulassen) verarbeiten. Für diese statuiert die DS-GVO jedoch grds. ein Verarbeitungsverbot (Art. 9 I DS-GVO).

Häufig sind die Datensätze, auf denen Foundation Models aufbauen, nicht angemessen (technisch) dokumentiert; soweit dies aber der Fall ist, ist die (wegen der großen Datenmenge oftmals nicht vollständige) Dokumentation typischerweise nicht ausreichend einsehbar.[24] Datenschutzrechtsver-

19 Woolridge, Exploring Foundation Models – Session 1, The Alan Turing Institute, 22.2.2023, ab Minute 20:00, https://www.youtube.com/watch?v = n9OkJBluOa4.

20 Vgl. Tagesschau.de, Google-Textroboter gibt falsche Antwort v. 9.2.2023.

21 Zu den Problemen generativer KI und Wasserzeichen bspw. Henderson, Should the United States or the European Union Follow China's Lead and Require Watermarks for Generative AI?, Georgetown Journal of International Affairs, 2023.

22 Flannery O'Connor/Moxley, Unser Ansatz für KI-Innovationen mit Verantwortung, YouTube Official Blog v. 14.11.2023.

23 Vgl. zu dieser Problematik bspw. Carlini et al., Extracting Training Data from Large Language Models, 2020 (letzte Version 2021); Weidinger et al., Ethical and social risks of harm from Language Models, S. 18 ff.

24 Vgl. zur Problematik bspw. Bender et al., On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?, 2021.

stöße lassen sich folglich kaum nachvollziehen. Betroffene können sich nicht sicher sein, dass sie ihre Rechte, etwa das Recht auf Löschung (Art. 17 DS-GVO), bei großen Datenmodellen überhaupt noch effektiv durchzusetzen vermögen. Denn der Befehl „Löschen" überschreibt in Trainingsmodellen die Daten grds. nicht.[25] Bei Systemen, die maschinell lernen, wäre es an sich unumgänglich, sie auf Grundlage der verbleibenden Daten von Grund auf neu zu trainieren – ein Weg, der regelmäßig die Wirtschaftlichkeit und Fähigkeiten des Modells infrage stellen kann.[26]

Zudem wird vielen Betroffenen gar nicht bewusst sein, ob bzw. welche persönlichen Daten die Systeme verarbeiten. Sekundärrechtliche Transparenzvorgaben, wie diejenigen des Art. 12 DS-GVO, laufen dann gegebenenfalls ins Leere. Foundation Models konfrontieren daher das Datenschutzregime der DS-GVO mit neuen Herausforderungen.

f) Sicherheitsrisiken

Foundation Models vereinfachen es Cyberkriminellen substanziell, Straftaten im Internet zu begehen, zB Betrugsdelikte in Gestalt von Phishing oder Scamming. Terroristische Organisationen könnten sie als Mittel missbrauchen, um Propaganda breiter zu streuen oder sonstige Aktivitäten vorzubereiten. Denn die Modelle ermöglichen es, gefährdende Inhalte in noch höherer Qualität und deutlich schneller zu generieren.

Künstliche Intelligenz lässt sich auch einsetzen, um Trainingsdaten zu manipulieren oder große Sprachmodelle zu korrumpieren. Unter Einsatz von Malware lassen sich programmierfähige große Sprachmodelle nicht zuletzt missbrauchen, um mithilfe des gewonnenen Wissens noch einfacher Anschlussstraftaten zu begehen. Die Befürchtungen reichen bis dahin, dass Foundation Models besondere Gefahren für die „Biosicherheit" auslösen, da Dritte an kritische Informationen gelangen könnten, die sie benötigen, um biologische Waffen herzustellen.[27] Bereits jetzt versuchen Nutzer, das Sprachmodell (durch sog. Jailbreaking) dazu zu bringen, vorgegebene Anweisungen und Sicherheitsvorkehrungen zu ignorieren oder Nutzerdaten abzugreifen, indem sie Prompts auf Webseiten verbergen.[28]

g) Abhängigkeit von großen Tech-Unternehmen

Wer riesige Datenmengen verarbeiten und nutzen will, ist auf große Rechenleistungen und -kapazitäten angewiesen. Derzeit sind aber nur wenige Tech-Magnaten überhaupt imstande, leistungsstarke Foundation Models zu entwickeln. Daraus kann eine noch stärkere Monopolisierung und – damit verbunden – eine noch höhere gesellschaftliche sowie staatliche Abhängigkeit von einigen wenigen privaten Akteuren erwachsen.

Nutzer der Foundation Models bzw. der auf ihnen basierenden Anwendungen werden im Zweifel in die Verlegenheit geraten, (eigene) Daten mit den (dahinterstehenden) Modellen und Cloud-Systemen der großen amerikanischen bzw. chinesischen Unternehmen teilen zu müssen. Dies löst Spannungslagen zu dem Ziel der Datensparsamkeit (vgl. Art. 5 I Buchst. c DS-GVO) sowie datenschutzrechtlichen Sicherheitsanforderungen (vgl. bspw. Art. 32 DS-GVO) aus. Eine solche „Zwangslage" gilt es insbes. im Verhältnis Bürger-Staat, zB im Rahmen der öffentlichen Verwaltung, zu vermeiden. Vor allem quelloffene und lokale Foundation Models könnten insoweit eine Chance eröffnen, entsprechenden Abhängigkeiten zu entgehen.[29]

h) Umweltschutz

Foundation Models verschlingen gigantische Energiemengen – sowohl für das initiale Training als auch während ihres gesamten Lebenszyklus. Damit fordern sie die nationalen und supra- sowie internationalen Umwelt- und Klimaschutzziele in besonderer Weise heraus.

III. Foundation Models als regulatorische Herkulesaufgabe

Dass generative KI zusehends im Alltag der Menschen ankommt, schärft auch das Risikobewusstsein der Politik und der Bevölkerung. Rufe nach einer effektiven Regulierung ertönen – auf nationaler, unionaler und internationaler Ebene – immer lauter. Bereits im Jahr 2021 hatte der Europarat ein dauerhaftes Committee on Artificial Intelligence (CAI) eingerichtet. Weder dessen Entwurf für eine völkerrechtlich bindende „Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law"[30] noch die unverbindlichen Empfehlungen der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD)[31] oder der Organisation der Vereinten Nationen für Erziehung, Wissenschaft und Kultur (UNESCO)[32] geben jedoch verbindliche Antworten auf die zahlreichen Herausforderungen und Fragen, die GPAI-Modelle aufrufen.[33]

Umso heller fällt das Schlaglicht derzeit auf die EU: Ihr Vorschlag für ein Gesetz über Künstliche Intelligenz (KI-VO-E)[34] soll KI umfassend regulieren. Im April 2021 stellte die Kommission ihren ersten Entwurf[35] vor. Eineinhalb Jahre später veröffentlichte der Rat seine Allgemeine Ausrichtung (KI-VO-E (Rat)).[36] Im Juni 2023 verabschiedete schließlich das Europäische Parlament seinen Entwurf (KI-VO-E (EP)).[37] Auf Grundlage dieser drei Entwürfe haben die Unionsorgane in den Trilog-Verhandlungen Anfang Dezember 2023 einen vorläufigen politischen Abschluss erzielt.[38]

Der Versuch, Foundation Models zu regulieren, ist ein Balanceakt: Einerseits gilt es, die Gefahren der Modelle im

25 Das ruft die Frage auf den Plan, ob das Recht „das Löschen aus dem Suchindex, das Überschreiben im Dateisystem, das Tilgen aus den Logdateien und Backups oder sogar das Entfernen aus allen internen Mechanismen" umfasst, Fosch-Villaronga et al., Computer Security & Law, Review 2017, 12.

26 Ginart et al., Making AI Forget You: Data Machine/Deletion Learning, Abstract, 2019.

27 Ausf. dazu Maham/Küspert, Governing General Purpose AI, A Comprehensive Map of Unreliability, Misuse and Systematic Risks, Stiftung Neue Verantwortung, 2023, S. 30.

28 Heikkilä, Three ways AI chatbots are a security disaster, MIT Technology Review v. 3.4.2023.

29 Mit Blick auf „quelloffene" Foundation Models krit. Widder et al., Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI.

30 Council of Europe, CAI – Committee on Artificial Intelligence, Revised Zero Draft (Framework) Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law, 6.1.2023.

31 OECD Legal Instruments, Recommendation of the Council on Administrative and Technical Regulations which Hamper the Expansion of Trade (aufgehoben am 12.7.2017).

32 UNESCO, Recommendation on the Ethics of Artificial Intelligence, verabschiedet am 23.11.2021.

33 Die UNESCO veröffentlichte 2023 immerhin ein Policy Paper, das sich explizit mit Foundation Models auseinandersetzt, vgl. UNESCO, Foundation models such as ChatGPT through the prism of the UNESCO Recommendation on the Ethics of Artificial Intelligence, 2023.

34 Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung bestimmter Rechtsakte der Union.

35 Europäische Kommission, COM (2021) 206 final v. 21.4.2021.

36 Allgemeine Ausrichtung des Rates der Europäischen Union, Ratsdok. 15698/22v. 6.12.2022.

37 Verhandlungsposition des Europäischen Parlaments zur KI-VO, P9_TA (2023)0236 v. 14.6.2023.

38 Deren Details sind zurzeit noch nicht bekannt bzw. noch offen.

gebotenen Umfang einzuhegen und die europäischen Werte sowie Grundrechte der Union hinreichend zu schützen. Andererseits ist die Union aufgerufen, dem technologischen Fortschritt und der daraus erwachsenden wirtschaftlichen Wertschöpfung ausreichenden Spielraum zu belassen.

1. Regulierung Künstlicher Intelligenz in der EU

So sehr sich die drei Entwürfe unterscheiden, so sehr teilen sie doch eine gemeinsame Grundphilosophie: Das Pflichtenprogramm, welches die Verantwortlichen von KI-Systemen erfüllen müssen, hängt davon ab, welche Risiken von dem konkreten System für die Grundrechte und die Sicherheit ausgehen (sog. risikobasierter Ansatz).[39]

Der Kommissionsentwurf konzipiert eine Risikopyramide mit vier Stufen: (1) KI-Systeme ohne oder mit nur minimalen Risiken, für die er lediglich Verhaltenskodizes empfiehlt (Art. 67, 69 KI-VO-E); (2) KI-Systeme mit geringen Risiken, für die spezifische Transparenzpflichten greifen (Art. 52 KI-VO-E); (3) Hochrisiko-KI-Systeme (Art. 6 f. KI-VO-E), die der KI-VO-E zwar grds. erlaubt, aber in ein engmaschiges Regulierungskorsett presst (Art. 8 ff. KI-VO-E);[40] (4) generell verbotene Praktiken (Art. 5 KI-VO-E), mit denen unannehmbare Risiken einhergehen, so dass (unter engen Ausnahmen) niemand entsprechende KI-Systeme in den Verkehr oder auf den Markt bringen darf.

In diese Regelungsarchitektur fügen sich GPAI-Modelle nicht bruchfrei ein: Ihrem Wesen nach kennen sie keine konkrete Zweckbestimmung, sondern können in einer Vielzahl unterschiedlicher Kontexte und Einsatzszenarien zur Anwendung kommen. Daher liegen sie gleichsam quer zu einem an konkrete Risiken anknüpfenden Regulierungsmodell:[41] Da dieses sich im Kern dadurch auszeichnet, nicht einzelne *Technologien*, sondern riskante *Anwendungen* zu regulieren, vermag es die Besonderheiten der General-Purpose-AI nicht recht abzubilden. Schon bald nachdem ChatGPT das disruptive Potenzial generativer KI-Anwendungen erkennbar zutage treten ließ, begann daher das Ringen um geeignete legislatorische Antworten.

2. Foundation Models in den KI-VO-Entwürfen

a) Allgemeine Ausrichtung des Rates

Unter der Bezeichnung „KI-Systeme mit allgemeinem Verwendungszweck“ („general purpose AI systems“ („GPAI-Systeme“)) brachte die slowenische Ratspräsidentschaft im November 2021 erstmals die neue Systemkategorie in den Gesetzgebungsprozess der KI-VO ein.[42] Der Ratsentwurf versteht darunter Systeme, die keiner konkreten, feststehenden Zweckbestimmung verschrieben sind und daher – auch auf nachgelagerter Ebene – einer Vielzahl unterschiedlicher Verwendungen offenstehen. Sie können zB „generell anwendbare Funktionen wie Bild- und Spracherkennung, Audio- und Videogenerierung, Mustererkennung, Fragenbeantwortung, Übersetzung etc. ausführen“ (Erwgr. 70 a Ratsdok. 14278/21). Ihre Nutzer können sie mithin sowohl in hoch risikobehafteten Einsatzszenarien als auch in risikoarmen Kontexten verwenden. Der Vorschlag sah deshalb zunächst vor, den Anwendungsbereich der KI-VO *nicht* auf polyvalent einsetzbare KI-Anwendungen zu erstrecken.[43]

Dieser Ansatz stieß jedoch auf harsche Kritik. Der Rat gab ihn unter französischer Präsidentschaft auf.[44] Diese legte stattdessen einen Kompromissvorschlag vor, dessen Begriffsbestimmung die vielseitigen Nutzungsmöglichkeiten der Modelle aufgriff und die Verwendungsabsicht des Anbieters in den Vordergrund rückte. An dieser Definition hält der Rat in seiner allgemeinen Ausrichtung fest und nimmt GPAI-Systeme explizit in das regulatorische Fadenkreuz:[45] Er möchte für sie regelmäßig die allg. Vorschriften der Art. 8 ff. KI-VO-E (Rat), also das strenge Pflichtenregime für Hochrisiko-KI-Systeme, gelten lassen. Immer dann, wenn GPAI-Systeme „als Hochrisiko-KI-Systeme oder als Komponenten von Hochrisiko-KI-Systemen (…) verwendet werden können”, sollen die Normen auf sie anwendbar sein (Art. 4b I 1 KI-VO-E (Rat)).[46] Um ihren Besonderheiten Rechnung zu tragen, sollen aber Durchführungsrechtsakte der Kommission die Anforderungen aus Titel III Kapitel 2 KI-VO-E „präzisier(en)” und an KI-Systeme mit allgemeinem Verwendungszweck „an (…)pass(en)“ (Art. 4b I 2 KI-VO-E (Rat)). In diesem Regulierungsmodell gelten die Regeln der Art. 8 ff. KI-VO-E (Rat) für GPAI-Systeme erst, wenn die Durchführungsrechtsakte Anwendung finden, spätestens aber 18 Monate, nachdem die VO in Kraft getreten ist (Art. 4b I 1 KI-VO-E (Rat)).

Auch die Vorschriften zur Konformitätsbewertung (Art. 40 ff., Erwgr. 64 ff. KI-VO-E (Rat)) will der Rat auf die GPAI-Systeme anwenden, so dass die Anbieter das Verfahren „auf der Grundlage einer internen Kontrolle“ durchführen müssten (Art. 4b II, III iVm 16 Buchst. e KI-VO-E (Rat)). Eine weitere Vorgabe zielt darauf, die Transparenz der Systeme zu erhöhen: Ihre Anbieter sollen sie in einer Datenbank registrieren müssen (Art. 4b II, 16 Buchst. f, 51, 60 KI-VO-E (Rat)) – ebenso wie bereits der Anbieter eines Hochrisiko-KI-Systems nach dem Entwurf der Kommission (Art. 16 Buchst. f, 51, 60 KI-VO-E (KOM)).

b) Vorschlag des Parlaments

Anders als der Rat hat das Parlament für alle KI-Systeme, die in den Anwendungsbereich der KI-VO fallen – explizit auch für Foundation Models (vgl. Art. 4a KI-VO-E (EP)) – ein Standardregelungsprogramm vor Augen. Dieses schließt einige Grundsätze bzw. Vorgaben ein, welche deren Betreiber erfüllen sollen (etwa mit Blick auf die Menschenwürde, technische Sicherheit, Transparenz uva., vgl. Art. 4a I KI-VO-E (EP)). Als Spezifikum für Foundation Models bestimmt Art. 4a II 3 KI-VO-E (EP) zusätzlich, dass ihre Anbieter „die allgemeinen Grundsätze durch die in den Art. 28–28b fest-

39 Erwgr. 14 KI-VO-E (KOM). Dazu ausf. Hilgendorf/Roth-Isigkeit/Martini, Die neue Verordnung der EU zur Künstlichen Intelligenz, 2023, § 4 Rn. 5 ff. mwN.

40 Die Entscheidung, ob ein Hochrisiko-KI-System vorliegt, richtet sich u.a. nach der Zweckbestimmung des KI-Systems.

41 Die Kommission erwähnte die Systeme nicht explizit. Ihr Entwurf hätte sie nur umfasst, sofern das einzelne Modell die Voraussetzungen von bspw. Hochrisiko-KI-Systemen (mit entsprechender Zweckbestimmung) erfüllt hätte.

42 Ratsdok. 14278/21 v. 29.11.2021.

43 Vgl. Ratsdok. 14278/21 II. Main Changs, Nr. 6 General Purpose AI Systems.

44 Ratsdok. 10069/22 v. 15.6.2022.

45 Er definiert ein GPAI-System als „ein KI-System, das – unabhängig davon, wie es in Verkehr gebracht oder in Betrieb genommen wird, auch in Form quelloffener Software – vom Anbieter dazu vorgesehen ist, allgemein anwendbare Funktionen wie Bild- oder Spracherkennung, Audio- und Videogenerierung, Mustererkennung, Beantwortung von Fragen, Übersetzung und Sonstiges auszuführen“ (Art. 3 Nr. 1b KI-VO-E (Rat)). Die Begriffsbestimmung rekurriert zudem ausdrücklich darauf, dass GPAI-Systeme „in einer Vielzahl von Kontexten eingesetzt und in eine Vielzahl anderer KI-Systeme integriert werden“ können.

46 Eine Ausnahme von der Anwendbarkeit der Hochrisiko-KI-Regeln soll dann gelten, wenn der Anbieter eines GPAI-Systems „ausdrücklich jegliche Verwendung mit hohem Risiko ausgeschlossen hat“ (Art. 4c I KI-VO-E (Rat)), es sei denn, der Anbieter hat „hinreichende Gründe für die Annahme (…), dass es zu einer Fehlanwendung des Systems kommen könnte“ (Art. 4c II KI-VO-E (Rat)).

gelegten Anforderungen (…) (umsetzen) und (…) (einhalten)" sollen. Art. 28b KI-VO-E (EP) verpflichtet Anbieter eines Basismodells u.a. dazu, eine Risikoprüfung nachzuweisen, eine angemessene Daten-Governance durchzuführen, seine Sicherheit zu gewährleisten und seinen Energieverbrauch zu senken. Die allgemeinen Gebote sollen eine Ergänzung durch „Normen" erfahren (Erwgr. 60 g S. 6 KI-VO-E (EP)). Auch das Parlament möchte, dass die Anbieter von Foundation Models diese in einer Datenbank registrieren müssen (Erwgr. 69, Art. 28b II Buchst. g, Art. 60, Anhang VIII (Abschn. C) KI-VO-E (EP)).

c) Selbstregulierungsinitiative Deutschlands, Frankreichs und Italiens

Nicht alle am Gesetzgebungsverfahren beteiligten Akteure wollten ein umfassendes Anforderungs- und Pflichtenregime für Foundation Models in die KI-VO integrieren: Die Befürchtung, durch eine strikte Regulierung die Innovationsfreiheit zu stark zu beschränken, mündete im Rahmen der Trilog-Verhandlungen in Rufe nach *freiwilligen Selbstverpflichtungen* der Anbieter von Foundation Models. Frankreich, Italien und Deutschland brachten über den Rat den Vorschlag ein, grds. auf eine gesetzliche Regelung und Sanktionen zugunsten eines Selbstregulierungsansatzes zu verzichten.[47] Die Regierungen beschlich nicht zuletzt die Sorge, dass strenge regulatorische Anforderungen den Aufstieg ihrer heimischen dynamischen KI-Unternehmen, insbes. *Mistral AI* und *Aleph Alpha*, zu stark ausbremsen könnten.

Das Instrument der Selbstverpflichtung hat grds. den Charme, – stärker als starre Regulierungsmodelle – zuständige Aufsichtsbehörden zu entlasten und der Dynamik schnelllebiger Entwicklungen gegebenenfalls besser zu entsprechen. Verhaltenskodizes sind den KI-VO-E auch nicht gänzlich unbekannt: Bisher sind sie allerdings nur auf der Stufe der KI-Systeme vorgesehen, die keine oder nur minimale Risiken bergen (vgl. Art. 69 KI-VO-E).

Eine denkbare Ausgestaltungsoption wäre bspw. ein Kodex, der ethische Grundwerte definiert und dem sich die Anbieter grds. unterwerfen.[48] Für einen solchen Regulierungsansatz kann das Konzept „comply or explain", wie es etwa der Corporate Governance Kodex entwickelt hat, in Gestalt eines KI-Responsibility-Kodex eine über die bisherigen Regelungsentwürfe hinausgehende taugliche Blaupause liefern.[49] Die Anbieter müssten dann erklären, ob und inwiefern sie den Verhaltensanforderungen entsprechen und aus welchen Gründen sie von einzelnen Vorgaben gegebenenfalls abweichen.

Selbstregulierungskonzepte stehen jedoch im Verdacht, nicht die scharfen Zähne zu zeigen, die es bräuchte, um die regulatorischen Zielsetzungen in praxi tatsächlich zu erreichen, insbes. einen adäquaten Schutz Betroffener zu verbürgen. Die durchwachsenen Erfahrungen mit Selbstregulierungsmodellen bspw. bei sozialen Netzwerken unterfüttern bisher diese Befürchtung. Der Ansatz, heimische KI-Hoffnungsträger für Foundation Models gleichsam zu „protegieren", indem die KI-VO ihnen möglichst wenige Pflichten auferlegt, ist womöglich insgesamt auch etwas kurzsichtig. Denn von „Regulierungsferien" profitieren zum einen häufig gerade bereits etablierte Anbieter in besonderer Weise. Zum anderen sind für die wirtschaftliche Wertschöpfung im Binnenmarkt, die von KI ausgeht, im Zweifel weniger die Anbieter der Foundation Models allein maßgeblich, sondern v.a. die zahlreichen Unternehmen, die auf der Grundlage dieser Modelle Anwendungen anbieten.

IV. Bewertung der KI-VO-Entwürfe

1. Definition und Regulierungstiefe

Wie grundlegend sich die Vorschläge für eine KI-VO in ihrer konzeptionellen Ausrichtung unterscheiden, verdeutlichen bereits die abweichenden Begriffsbestimmungen für Systeme, die technisch als „Foundation Models" gelten.

Besonders die im Entwurf des *Rates* vorgesehene Definition des „KI-System(s) mit allgemeinem Verwendungszweck" rief schnell Kritiker auf den Plan. Denn sie zählt zwar beispielhaft einige „allgemein anwendbare Funktionen" auf, setzt jedoch keine konkreten Benchmarks,[50] um die Eigenschaften von GPAI-Systemen umfassend oder etwa mit Blick auf bestimmte Parameter qualitativ bewerten zu können. Art. 3 Nr. 1b Hs. 2 KI-VO-E (Rat) stellt zudem lediglich eine „Kann"-Bedingung auf („kann (…) eingesetzt und (…) integriert werden"), und öffnet den potenziellen Anwendungsbereich der Regelungen für GPAI-Systeme damit sehr weit. Indem die Definition nicht nur große Sprachmodelle erfasst, sondern auch andere Systeme, die in vielfältigen nachgelagerten Anwendungen und unterschiedlichen Szenarien (zB auch in Hochrisiko-Bereichen) zum Einsatz kommen können, verhindert sie einerseits empfindliche Schutzlücken.[51] Andererseits läuft ihre Weite jedoch Gefahr, auch solche KI-Systeme einer engmaschigen Regulierung zu unterstellen, die nicht im selben Maße risikogeneigt sind. Um die Begriffsbestimmung sachgerecht einzugrenzen, kamen zB „Checklisten" oder „Beispiele" als ergänzende Instrumente zur Sprache.[52]

Quelloffene Systeme bezieht die Definition ebenfalls in den Anwendungsbereich der KI-VO ein.[53] Die Besonderheiten bspw. (kleinerer) Foundation Models, die auf lokalen Rechnern laufen können und nicht Cloud-basiert sind (etwa das quelloffene GPT4ALL), greift der Rat demgegenüber nicht ausdrücklich auf. Da Art. 3 Nr. 1b KI-VO-E (Rat) keinen Rekurs auf die Menge verwendeter (Trainings-)Datensätze oder sonstige (Größen-)Parameter nimmt, erfasst der Ratsentwurf auch kleinere Modelle. Gerade ihre Anbieter könnte es jedoch vor enorme Schwierigkeiten stellen, den strengen Anforderungskanon des KI-VO-E für Hochrisiko-KI-Systeme zu erfüllen. Die damit verbundene Markteintrittshürde träfe dann insbes. kleinere Unternehmen, die im Binnenmarkt einen Gegenpol zu den großen marktbeherrschenden Tech-Firmen (zumeist US-amerikanisch) bilden könnten.

Das *Parlament* hat diese Kritik an der weiten Begriffsbestimmung des Rates teilw. aufgegriffen. Sein Entwurf unterscheidet zwischen Foundation Models (Art. 3 I Nr. 1c KI-VO-E (EP)) und GPAI-Systemen (Art. 3 I Nr. 1d KI-VO-E (EP)). Ein GPAI-System definiert es als „KI-System, das in einem breiten Spektrum von Anwendungen eingesetzt und an diese angepasst werden kann, für die es nicht absichtlich und

47 Bertuzzi, France, Germany, Italy push for 'mandatory self-regulation' for foundation models in EU's AI law, Euractiv, 19.11.2023.
48 Ausf. dazu bspw. Martini, Blackbox Algorithmus, 2019, S. 320 ff.
49 Dazu Martini, Blackbox Algorithmus, 2019, S. 328 ff.
50 Diese wären allerdings zum jetzigen Zeitpunkt angesichts der Vielseitigkeit der GPAI-Systeme und geringer öffentlich verfügbarer Informationen über sie wohl auch nur schwer bestimmbar.
51 Dazu aufschlussreich auch der offene Brief von Gebru et al., Five considerations to guide the regulation of "General Purpose AI" in the EU's AI Act, 2023.
52 Bspw. Solaiman in Center for Data Innovation, Panel: Should the EU Regulate General-Purpose AI Systems?, 13.9.2022, (ab Minute 14:30), https://datainnovation.org/2022/09/should-the-eu-regulate-general-purpose-ai-systems/.
53 Vgl. IV. 3.

speziell entwickelt wurde“ (Art. 3 I Nr. 1d KI-VO-E (EP)). Unter einem Basismodell versteht es hingegen ein „KI-Systemmodell, das auf einer breiten Datenbasis trainiert wurde, auf eine allgemeine Ausgabe ausgelegt ist und an eine breite Palette unterschiedlicher Aufgaben angepasst werden kann“ (Art. 3 I Nr. 1c KI-VO-E (EP)). Die Begriffsbestimmung eines Foundation Models ist folglich enger als diejenige eines GPAI-Systems, das sich auch eines/mehrerer Foundation Models bedienen kann.[54]

Wo die Grenze zwischen den beiden Systemkategorien exakt verläuft, bleibt aber unklar. GPAI-Systeme sollen zwar auf den Basismodellen aufbauen und diese implementieren können. Warum das Parlament insoweit aber überhaupt differenziert und inwieweit sich etwaige Überschneidungen auswirken, lässt sein Entwurf unbeantwortet.

Die Erwägungsgründe ermöglichen ebenfalls keine eindeutige Unterscheidung. Erwgr. 60e S. 1 KI-VO-E (EP) erkennt ein Merkmal von Basismodellen darin, „im Hinblick auf Allgemeinheit und Vielseitigkeit der Ergebnisse optimiert“ worden zu sein. Was genau diese Allgemeinheit und Vielseitigkeit – in Art. 3 I Nr. 1c KI-VO-E (EP) als „allgemeine Ausgabe” benannt – ausmacht, bleibt jedoch unklar. Sollte diese Wendung auf die möglichen unterschiedlichen Fähigkeiten und Aufgaben zielen, stellt sich die Frage, wie viele das Modell konkret beherrschen muss, bevor beide Attribute als erfüllt gelten.

Erwgr. 60e S. 2 KI-VO-E (EP) hilft bei dem Unterfangen, Foundation Models möglichst trennscharf von anderen Systemkategorien abzugrenzen, ebenfalls nicht entscheidend weiter. Er hebt als eines ihrer Kennzeichen hervor, dass sie „häufig auf der Grundlage eines breiten Spektrums von Datenquellen und großer Datenmengen trainiert (werden), um eine Fülle nachgelagerter Aufgaben zu erfüllen, (…)“. Zwar nimmt das Parlament insoweit – anders als der Rat – Bezug auf die Größe des Foundation Models, um den Anwendungsbereich deutlicher abzustecken. Sowohl das wertungsoffene Adverb „häufig” als auch der nicht weiter konkretisierte Bezug auf breite Datenquellen und große Datenmengen helfen allerdings nur begrenzt, wenn es zu entscheiden gilt, ob ein konkretes System als Foundation Model einzustufen ist.

Die Abgrenzung zu vortrainierten Modellen, bspw. Mehrzweck-KI, ist ebenfalls nur bedingt zweckdienlich: Diese sollen für eine „enger gefasste, weniger allgemeine und begrenztere Reihe von Anwendungen entwickelt (worden sein) und nicht an ein breites Spektrum von Aufgaben angepasst werden können“ (Erwgr. 60g S. 10 KI-VO-E (EP)). Auch diese Merkmale bleiben zu vage und unbestimmt, um eine klare Grenze zu ziehen. Gerade die Modellgröße kann erheblich divergieren; zudem ist ungewiss, wie diese sich zukünftig entwickeln werden. Es dürfte daher vermutlich künftig nur bedingt sinnstiftend sein, allein auf die beiden Parameter Ausgabe und Größe zu rekurrieren.

Zwischen KI-Systemen mit spezifischer Zweckbestimmung und solchen mit allgemeinem Verwendungszweck zu differenzieren, ist auch nur dann erforderlich, wenn für Letztere besondere Regelungen gelten. Solche speziellen Vorgaben fehlen im Parlamentsentwurf jedoch. Es entsteht zumindest der Eindruck, dass das Parlament die KI-Systeme mit allgemeinem Verwendungszweck v.a. deshalb in seine Liste der Begriffsbestimmungen in Art. 3 KI-VO-E (EP) aufgenommen hat, um diese neue Systemkategorie aus dem Ratsentwurf zumindest begrifflich aufzugreifen – der Terminus aber praktisch eine leere Worthülse bleibt.

Neben den Definitionen erweisen sich auch die einzelnen normativen Vorgaben, die die Entwürfe des Rates und des EP für die „Systeme“ – ob als Foundation Model oder als GPAI-System betitelt – vorsehen, als kritikwürdig. Besonders problematisch erscheint der Vorschlag des Rates, es der Kommission zu überlassen, in Durchführungsrechtsakten die Regelungen bzw. Hochrisiko-KI-Normen zu präzisieren bzw. anzupassen, die auf die GPAI-Systeme anwendbar sein sollen (Art. 4b I 2 KI-VO-E (Rat)). Damit verlagert er zum einen die erforderliche Konkretisierung bzw. Modifizierung der Regelungen noch weiter in die Zukunft. Reguliert der Unionsgesetzgeber die Risiken der Systeme erst nach einem Markteintritt, dürfte deren Entwicklung schon so weit fortgeschritten sein, dass es aufwendiger nachgelagerter technischer Anpassungen oder gar ganzer Umstrukturierungen bedürfte, um die Architektur der Systeme noch an einen strengeren Pflichtenkatalog anzupassen. Zum anderen vertraut der Rat der Kommission mit der Befugnis, solche Durchführungsrechtsakte zu erlassen, zentrale politische und grundrechtswesentliche Wertentscheidungen für den KI-Sektor an, die nicht ohne klare Steuerungsvorgaben verantwortbar sind.

Diese Kritik gilt ebenso für den Ansatz des Parlaments, der Kommission Befugnisse für Durchführungsrechtsakte zu übertragen, welche die Anforderungen des Art. 28b KI-VO-E (EP) näher spezifizieren sollen (vgl. bspw. Art. 41 I a KI-VO-E (EP)). Normungsorganisationen Konkretisierungsmacht zu übertragen, wie es das Parlament ebenfalls vorschlägt (vgl. Erwgr. 61a, Art. 40 I a f. KI-VO-E (EP))[55], überzeugt v.a. aus demokratischen Gesichtspunkten erst recht nicht. Wenn das Gesetz unbestimmt formuliert ist, können Normungsorganisationen fehlende demokratische Legitimation nicht ersetzen.

2. KI-Wertschöpfungskette

Zu den größten Herausforderungen bei dem Versuch, Foundation Models regulatorisch einzuhegen, gehört es, den gesamten Lebenszyklus der KI-Wertschöpfungskette sachgenau normativ zu erfassen. Denn in ihrem Funktionsgefüge übernehmen in der Regel mehrere Akteure verschiedene Rollen und Aufgaben – darunter auch solche Beteiligte, die sich bspw. nicht passgenau in die Kategorien der regulatorischen Blaupause des allgemeinen Produkthaftungsrechts einordnen lassen. Während dieses Rechtsregime zwischen (End-)Hersteller, Zulieferer bzw. Teilhersteller, Händler und Einführer usw. als verantwortlichen Akteuren unterscheidet, lassen sich diese Funktionen auf die Akteure einer KI-Wertschöpfungskette nicht bruchlos übertragen.

a) Vorschlag der Kommission

Der Kommissionsentwurf untergliedert die Personen, die das Pflichtenprogramm der KI-VO adressiert, in Anbieter, Nutzer sowie andere Akteure (vgl. Art. 16 ff. KI-VO-E (KOM)), wie zB Einführer (Art. 26 KI-VO-E (KOM)) oder Händler (Art. 27 KI-VO-E (KOM)). Die Hauptverantwortung für Hochrisiko-KI-Systeme sieht die Kommission beim Anbieter,

54 Vgl. Erwgr. 60 e S. 4 KI-VO-E (EP). Das Parlament differenziert zudem zwischen Basismodellen und „(v)ortrainierte(n) Modelle(n)” wie „einfache(n) Mehrzweck-KI-Systeme(n)”, die weniger breit gestreut Anwendung finden können (Erwgr. 60 g S. 10 KI-VO-E (EP)). Diese gelten im Anwendungsbereich der KI-VO nicht als Foundation Models, da „ihr Verhalten weniger unvorhersehbar” erscheint (Erwgr. 60 g KI-VO-E (EP)). Das Training eines Basismodells entspricht teilw. den Anforderungen des neu eingebrachten Terminus „große Trainingsläufe” (Art. 3 I Nr. 1e KI-VO-E (EP)).

55 Dazu auch vgl. IV. 4. a) aa).

also demjenigen, der das System auf den Markt oder in den Verkehr bringt – und das unabhängig davon, ob er „das System konzipiert oder entwickelt hat".[56]

Da Foundation Models im Kommissionsentwurf noch keine explizite Erwähnung finden, kennt dieser auch keine gesonderten Vorgaben für ihre Entwickler und Anbieter. Konsequenterweise unterfielen dann grds. allein die Anbieter nachgelagerter Anwendungen, die Foundation Models bspw. in ihre Hochrisiko-KI-Software integrieren und diese zu einem konkreten Zweck in einem Hochrisiko-Gebiet einsetzen, dem umfassenden Pflichtenkatalog für Anbieter von Hochrisiko-KI-Systemen (Art. 8 ff. KI-VO-E (KOM)).[57]

Viele nachgelagerte Hochrisiko-KI-Anbieter, die Foundation Models einsetzen, sähen sich in dieser Situation mit Pflichten konfrontiert, die sie mangels entsprechender Informationen nicht erfüllen können. Die Anbieter von Foundation Models hingegen könnten ihre Modelle so entwickeln, dass sie – trotz der besonderen Risiken ihrer Systeme – dem strengen Anforderungs- und Pflichtenregime der KI-VO nicht unterfallen. Somit wären sie nicht für mögliche Haftungsfälle und Gefahren verantwortlich – eine Art Freibrief mit vielschichtigen Konsequenzen.

b) Vorschlag des Rates

Die allgemeine Ausrichtung des Rates erwähnt den Anbieter eines GPAI-Systems ausdrücklich (Art. 4b II KI-VO-E (Rat)). Ihn treffen bereits dann entsprechende Pflichten, wenn sich das System als Hochrisiko-KI-System (oder als Komponente eines solchen) verwenden lässt (Art. 4b I 1, II KI-VO-E (Rat)). Für KMU strebt der Rat jedoch eine Ausnahmeregelung an: Für sie sollen die Hochrisiko-Vorgaben (Art. 4b KI-VO-E (Rat)) grds. nicht gelten (vgl. Art. 55a III KI-VO-E (Rat)). Die Vorschrift des Art. 28 KI-VO-E (KOM), der unter bestimmten Voraussetzungen Händler, Einführer, Nutzer oder sonstige Dritte als Anbieter einstuft, möchte der Rat ersatzlos streichen. Ferner verlangt sein Entwurf den Anbietern der GPAI-Systeme ab, mit solchen Anbietern zusammenzuarbeiten, die derartige Systeme für einen konkreten Hochrisiko-Zweck einsetzen wollen (Art. 4b V KI-VO-E (Rat)). Das ist auch insofern sinnstiftend, als Anbieter nachgelagerter KI-Systeme mit konkretem Verwendungszweck die Anforderungen der KI-VO denklogisch nur dann erfüllen können, wenn sie auf entsprechende Informationen Zugriff haben.

Für besonders komplexe Konstellationen hält der Rat allerdings keine Lösungen bereit. So schweigt sein Vorschlag bspw. zu den keineswegs fernliegenden Sachverhalten, in denen mehrere Akteure gemeinsam dasselbe GPAI-System anbieten. Wie sich die Pflichten unter den Anbietern dann aufteilen (vgl. etwa das Regelungsmodell des Art. 26 DS-GVO), bleibt ungeklärt.

Erwgr. 52a KI-VO-E (Rat) stellt zwar immerhin klar, dass Akteure in der KI-Wertschöpfungskette „(i)n bestimmten Situationen (…) mehr als eine Rolle gleichzeitig wahrnehmen (können) und (…) daher alle einschlägigen Pflichten, die mit diesen Rollen verbunden sind, kumulativ erfüllen (sollten)". Das betrifft allerdings nur einen singulären Anbieter – und hilft nicht, die Verantwortungsbereiche bei komplexen Beziehungen mehrerer Akteure gegeneinander abzuschichten. Der KI-VO-E (Rat) differenziert bspw. nicht danach, ob ein Akteur bei dem Entstehungsprozess des Systems eine besondere Rolle übernommen hat. Die Pflichten sollen vielmehr weiterhin den Anbieter treffen, der die Systeme auf den Markt bringt. Am Entstehungsprozess beteiligte Akteure, die nicht (zugleich) als Anbieter (Art. 3 Nr. 2 KI-VO-E (Rat)) einzustufen sind, träfen viele Regelungen folglich nicht.

Es empfiehlt sich jedoch, GPAI-Systeme grds. bereits ab dem Beginn ihrer Entwicklung regulatorisch einzuhegen. Anderenfalls wird sich kaum gewährleisten lassen, dass sie während ihres gesamten Lebenszyklus den notwendigen Anforderungen – insbes. mit Blick auf Transparenz, Dokumentation, (Trainings-)Daten oder Filter-Mechanismen – entsprechen. Ob es vor diesem Hintergrund sinnvoll erscheint, stets dem Anbieter die alleinige Verantwortung zuzuweisen, lässt sich durchaus in Zweifel ziehen.

Der Rat möchte dem Nutzer die Verpflichtung auferlegen, den Anbieter zu informieren, falls ein Hochrisiko-KI-System, das entsprechend der Gebrauchsanweisung zum Einsatz kommt, besondere Risiken iSd Art. 65 I KI-VO-E iVm Art. 3 Nr. 19 VO (EU) 2019/1020 offenbart (vgl. Art. 29 IV KI-VO-E (Rat)). Das überzeugt. Der Unionsgesetzgeber sollte den Umfang dieser Pflicht in der finalen Fassung der KI-VO weiter ausdehnen und explizit auf Foundation Models erstrecken, um einen umfassenderen, reibungslosen Austausch zwischen Nutzer und Anbieter sicherzustellen.

Zwar stellt der Rat besonders heraus, dass KI-Systeme mit allgemeinem Verwendungszweck „(a)ufgrund ihrer besonderen Merkmale und zur Gewährleistung einer gerechten Verteilung der Verantwortung entlang der KI-Wertschöpfungskette (…) verhältnismäßigen und spezifischeren Anforderungen und Pflichten unterliegen" sollten (Erwgr. 12c S. 4 KI-VO-E (Rat)). Diese grundlegende Erkenntnis münzt er jedoch im verfügenden Teil des Entwurfs nicht (hinreichend) in konkrete normative Vorgaben um. Durchführungsrechtsakte, welche die Kommission erlassen soll (vgl. auch Erwgr. 6c, 12c S. 6, Art. 4b I KI-VO-E (Rat)), können angesichts ihrer schwachen demokratischen Unterfütterung nur eingeschränkt für die angestrebte faire und flexible Verteilung der Verantwortlichkeitslasten bürgen. Es bräuchte vielmehr bereits konkrete sekundärrechtliche Vorgaben, die die unterschiedlichen Akteure während der jeweiligen Entwicklungsstufen des Systems – auch im Verhältnis untereinander – in die Pflicht nehmen. Entsprechende Durchsetzungsrechte, effektive Formen der Kontrolle und Sanktionen, die zB dann greifen, wenn Anbieter eines Foundation Models Informationen nicht vollständig oder wahrheitsgetreu übermitteln, sollten diese Anforderungen flankieren. Anderenfalls laufen die materiellen Vorgaben in praxi leer.

c) Vorschlag des Parlaments

Anders als der Rat unterscheidet das Parlament nicht nur ausdrücklich zwischen Foundation Models und GPAI-Systemen.[58] Sein Entwurf stellt zudem explizit fest, dass Foundation Models nicht automatisch als Hochrisiko-KI-Systeme einzustufen sind, aber besonderen Verpflichtungen unterfallen sollen (Erwgr. 60g S. 4, 9 KI-VO-E (EP)): Es hebt neue

56 Vgl. Erwgr. 53 KI-VO-E (KOM).

57 Nach Art. 28 I Buchst. a KI-VO-E (KOM) sollen auch Nutzer (bzw. Händler, Einführer oder Dritte) die Anbieterpflichten treffen, wenn sie ein Hochrisiko-KI-System unter ihrem Namen in Verkehr bringen oder in Betrieb nehmen. Ebenso wenn sie die Zweckbestimmung eines bereits auf dem Markt erhältlichen Hochrisiko-KI-Systems verändern oder eine wesentliche Änderung an dem besagten System vornehmen (Art. 28 I Buchst. b, c KI-VO-E (KOM)). Für diese Konstellationen hätte ein Foundation Model aber bereits als Hochrisiko-KI-System mit einer Zweckbestimmung auf dem Markt sein müssen. Zudem soll der ursprüngliche Anbieter in diesem Fall nicht mehr als Anbieter gelten, so dass die Verantwortung alleinig der neue Anbieter (der neuen Anwendung) trägt (Art. 28 II KI-VO-E (KOM)).

58 Vgl. III. 2. b) und IV. 1.

Regelungen aus der Taufe, die es auf den Anbieter eines Foundation Models zuschneidet (vgl. Erwgr. 60g, Art. 28b KI-VO-E (EP)) und etabliert so – neben den Pflichten, die alle KI-Anbieter treffen (vgl. Art. 4a KI-VO-E (EP)) – konkrete Anforderungen an das Datenmanagement, Transparenz, Sicherheit etc (Erwgr. 60g S. 5 KI-VO-E (EP)). Korrespondierende Normen sollen diese Verpflichtungen ergänzen (Erwgr. 60g S. 6 KI-VO-E (EP)). Bei Ausgaben generativer KI-Systeme, die Foundation Models anwenden, soll zudem u.a. hinreichend erkennbar sein, dass sie KI-basiert sind (Erwgr. 60g S. 8 KI-VO-E (EP)).

Die Pflichtenzuordnungsregelung des Art. 28 KI-VO-E möchte das Parlament – im Unterschied zum Rat – nicht streichen, sondern um weitere Vorgaben ergänzen.[59] Der Parlamentsentwurf spricht nicht mehr vom Nutzer („user"), sondern vom sog. Betreiber („deployer"; vgl. bspw. Art. 28 I KI-VO-E (EP)). Besondere Anforderungen an ihn stellt der KI-VO-E (EP) nur, sofern dieser das System im Hochrisiko-KI-Segment einsetzt (zB ist er dann verpflichtet, eine Grundrechte-Folgenabschätzung vorzunehmen, vgl. Art. 29a KI-VO-E (EP), Erwgr. 58 a KI-VO-E (EP)).

Anbieter, die ihr Foundation Model etwa „nur" durch einen API-Zugang[60] veröffentlichen, sollen zukünftig über den „gesamten Zeitraum (…), in dem dieser Dienst bereitgestellt und unterstützt wird" mit den nachgeschalteten Anbietern zusammenarbeiten (Erwgr. 60f Hs. 1 KI-VO-E (EP)). Dieser dauerhaften Kooperationsverpflichtung können sie nur entgehen, indem sie das „Trainingsmodell sowie umfassende und angemessene Informationen über die Datensätze und den Entwicklungsprozess des Systems" zur Verfügung stellen oder den „Dienst, zB den API-Zugang" so einschränken, dass die nachgelagerten Anbieter die normativen Vorgaben fortan allein erfüllen können (Erwgr. 60f Hs. 2 KI-VO-E (EP)).

Der Vorschlag, die Zusammenarbeit zwischen dem Anbieter eines Foundation Models und dem einer auf dem Modell basierenden Anwendung gesondert zu regeln, überzeugt im Grundsatz. Zu vage bleibt der Entwurf indes hinsichtlich der Handlungsoption, das „Trainingsmodell" und weitere Informationen an nachgelagerte Anbieter zu übertragen. Welche Anforderungen insoweit gelten sollen, bleibt unklar.

Die weitere Alternative, den Dienst so einzuschränken, dass nachgeschaltete Anbieter „ohne weitere Unterstützung" alle sie betreffenden Pflichten der KI-VO erfüllen können, ruft die Frage auf den Plan, wie sich das in praxi umsetzen lassen soll. Sie birgt erhebliches Konfliktpotenzial zwischen den betreffenden, häufig nicht auf Augenhöhe agierenden Akteuren. Nachgelagerte Anbieter müssten gegebenenfalls gegenüber den Anbietern der Basismodelle behaupten, dass es entgegen deren Verlautbarung in einem konkreten Fall nicht möglich ist, ohne weitergehende Unterstützung alle Pflichten zu erfüllen. Ob die Regelung zum Schutz der schwächeren (Vertrags-)Partei ausreicht und ob das Parlament dieses mögliche Ungleichgewicht ausreichend beachtet hat, ist zu bezweifeln.

Ohnedies ist die Gefahr, dass KMU als nachfolgende Anbieter sich mit missbräuchlichen Vertragsklauseln konfrontiert sehen, die ihnen die Anbieter der Foundation Models (zumeist große Tech-Unternehmen) einseitig stellen, mit Händen zu greifen. Mit Blick auf die allgemeine KI-Wertschöpfungskette will das Parlament dritten Parteien, die bestimmte „Tools und Dienstleistungen, aber auch Komponenten oder Prozesse" zu einem KI-System beisteuern, auferlegen, den Anbietern die entsprechenden Informationen zur Verfügung zu stellen (Erwgr. 60 S. 4 KI-VO-E (EP)). Zum Schutz von KMU sollen in diesem Rahmen einseitig auferlegte Vertragsbestimmungen nicht gelten, wenn diese Klauseln „die Lieferung von Werkzeugen, Dienstleistungen, Bauteilen oder Verfahren, die in einem Hochrisiko-KI-System verwendet oder integriert werden (…) regeln" (Erwgr. 60 a S. 3 KI-VO-E (EP)). Foundation Models unterfallen wohl nicht der Formulierung „Lieferung von Werkzeugen, Dienstleistungen, Bauteilen oder Verfahren". Allerdings sprechen Erwgr. 60 a S. 1 sowie Art. 28 II a und Art. 28a KI-VO-E (EP), die die entsprechenden Pflichten und den Schutz vor missbräuchlichen Vertragsklauseln näher ausgestalten, von Komponenten eines Hochrisiko-KI-Systems – hierunter ließen sich auch Basismodelle fassen. Ob ihre Anbieter „Dritte" sind (vgl. Art. 28 II a KI-VO-E (EP)), bleibt allerdings unklar. Zudem hat das Parlament – anders als in Art. 28 II UAbs. 2 KI-VO-E (EP) – (wohl bewusst) nicht klargestellt, dass Abs. 2a auf Foundation Models anwendbar sein soll.[61]

d) Zwischenergebnis

Eine gute Regulierung von GPAI-Modellen hat die Verantwortlichkeitssphären sachgerecht abzuschichten. Denn verschiedene Akteure können Basismodelle jeweils zu unterschiedlichen Zeitpunkten während ihrer Entwicklungs- und Nutzungsdauer prägen, heranbilden, hosten und verwenden. Bis zu einem gewissen Grad muss der Unionsgesetzgeber daher die Pflichten und Lasten der KI-VO flexibel aufteilen. Das Pflichtenprogramm muss einen oder mehrere Akteure gemeinsam oder getrennt voneinander treffen können und – in Grenzen – übertragbar sein.

Bei diesen Regulierungsbemühungen ist es aber zugleich essenziell, die Innovationskraft europäischer Unternehmen nicht abzuwürgen. Sofern diese (insbes. KMU) sich mit gesetzlichen Pflichten oder privatrechtlichen Vertragsklauseln konfrontiert sehen, denen sie nicht entsprechen können, drohen außerunionale Tech-Unternehmen sie vom Markt zu verdrängen. Ein Ausnahmevorbehalt für KMU, wie ihn Art. 55a III KI-VO-E (Rat) vorsieht, kann somit sinnvoll sein. Von Pflichten, die hohe Risiken für wichtige Schutzgüter (etwa bei der Verwendung von Foundation Models in nachgelagerten Anwendungen) abwenden sollen, sollten sie jedoch nicht in Gänze ausgenommen sein.

59 So sollen nach Art. 28 I Buchst. c KI-VO-E (EP) nunmehr auch Akteure als Anbieter eines Hochrisiko-KI-Systems gelten, die ein KI-System (einschließlich eines KI-Systems für allgemeine Zwecke), das kein Hochrisiko-KI-System darstellt „und bereits in Verkehr gebracht oder in Betrieb genommen wurde", so wesentlich verändern, dass es danach ein Hochrisiko-KI-System darstellt. Dies führt dann dazu, dass der ursprüngliche Anbieter nicht mehr als Anbieter dieses konkreten modifizierten KI-Systems gilt (Art. 28 II KI-VO (EP)). Er soll dafür aber den neuen Anbieter mit den entsprechenden Informationen versorgen, die er benötigt, um die Vorgaben der KI-VO zu befolgen, etwa mit der „technische(n) Dokumentation und alle(n) anderen relevanten und vernünftigerweise zu erwartenden Informationen und Fähigkeiten des KI-Systems, de(m) technischen Zugang oder sonstige(r) Unterstützung auf der Grundlage des allgemein anerkannten Stands der Technik". Diese Regelung soll auch für Anbieter von Foundation Models gelten, wenn „das Basismodell direkt in ein Hochrisiko-KI-System integriert ist" (Art. 28 II UAbs. 2 KI-VO-E (EP)).

60 Die Abkürzung steht für application programming interface, bezeichnet also eine Anwendungsprogrammierschnittstelle.

61 Art. 28 II UAbs. 1 KI-VO-E (EP) statuiert zwar ebenfalls konkrete Informationspflichten des Anbieters eines Foundation Models gegenüber dem Anbieter eines Hochrisiko-KI-Systems, der ein Foundation Model in sein System integriert. Dieser Absatz trifft aber anders als Abs. 2a keine weitergehenden Feststellungen mit Blick auf Mustervereinbarungen etc. Auch Art. 28a KI-VO-E (EP), der festlegt, wann eine Vertragsklausel als missbräuchlich gilt, bezieht sich nicht ausdrücklich auf Foundation Models. Ein Grund dafür, wieso KMU in den Fällen, in denen sie Foundation Models in ihre Anwendungen integrieren, um marktfähig zu sein, kein gesonderter Schutz zukommen sollte, ist nicht erkennbar.

Es empfiehlt sich, die Zuteilung der Verantwortlichkeit v.a. daran zu knüpfen, in welcher Gestalt der Anbieter das Basismodell veröffentlicht und welche Eingriffe er etwa bei Nutzung des Modells gewährt und zulässt, wie also das Verhältnis zwischen Anbieter, Nutzer bzw. Betreiber und anderen Akteuren (auch solchen, die die KI-VO-Entwürfe bisher noch nicht explizit umfassen, bspw. Hosts, Cloud-Anbietern etc) ausgestaltet ist. Die KI-VO muss die notwendigen Kooperationen und Koordinationen zwischen unterschiedlichen Akteuren forcieren, damit diese ihre jeweiligen Pflichten erfüllen (können).[62] Sowohl in dem Entwurf des Rates als auch im Parlamentsentwurf fehlen jedoch konkrete Vorschriften, die besondere Konstellationen (mehrere Anbieter, komplexe Sachverhalte, gemeinsame Verantwortlichkeiten usw.) rechtlich einrahmen.

Der Unionsgesetzgeber ist auch gut beraten, den Informationsaustausch zwischen den Entwicklern und Anbietern von Foundation Models und den Nutzern oder Anbietern nachgelagerter Anwendungen sowie weiteren beteiligten Akteuren und Dritten in unterschiedliche Richtungen zu etablieren[63] – und nicht lediglich eine Informationskette „top-down" aufzubauen. Es ist zwar sehr wichtig, dass die Anbieter der Foundation Models die Informationen teilen, die für die weiteren Akteure notwendig sind, um die Vorgaben der KI-VO einzuhalten. Jedoch ist es umgekehrt ebenso von Bedeutung, dass Nutzer bzw. Betreiber (also gerade auch nachgelagerte Anbieter) den Entwicklern und Anbietern der Foundation Models mitteilen (müssen), wenn ihnen Fehler, Bias oder andere Probleme am Basismodell oder bei dessen Integration in ein (Hochrisiko-)KI-System auffallen – folglich immer dann, wenn diese der „Sphäre" des Basismodells entspringen und unabhängig davon, ob die nachfolgenden Anbieter diese selbst beseitigen oder beeinflussen können.[64] Die derzeit in den Entwürfen vorgesehenen Bestimmungen reichen nicht aus, um diesen Informationsfluss zu gewährleisten. Betroffene sollten Probleme bis zu ihrem Ursprung zurückverfolgen können. Solche Erkenntnisse für andere Nutzer zu veröffentlichen, stärkte Transparenz und Nachprüfbarkeit.[65]

Die Verantwortlichen sollten zudem – sanktionsbewehrt – gewährleisten müssen, dass sie die ihnen übermittelten Informationen nachweislich untersuchen und ihre Systeme gegebenenfalls im notwendigen Maß anpassen. Das darf die Anbieter zugleich aber nicht dazu nötigen, ihre Geschäftsgeheimnisse (insbes. untereinander) preiszugeben.[66] Bereits die Prompts, die ein nachgelagerter Anbieter zur Steuerung nutzt, könnten allerdings als eben solche anzusehen sein.

Passgenaue Informationspflichten aus der Taufe zu heben, die der komplexen Wertschöpfungskette bei Basismodellen gerecht werden, gleicht im Ergebnis einem Drahtseilakt: Der Unionsgesetzgeber muss einerseits dafür Sorge tragen, dass die Akteure solche Belange, die unter den Schutz des geistigen Eigentums fallen oder die bspw. Geschäftsgeheimnisse darstellen (vgl. auch Erwgr. 60, Art. 28 II b KI-VO-E (EP)), nicht preisgeben müssen. Andererseits bleibt es ohne eine umfassende Verpflichtung zur Informationsweitergabe ungewiss, ob die Entwickler und / oder Anbieter der Foundation Models überhaupt selbst über alle Informationen verfügen, die etwa für umfangreiche Risikobewertungen[67] und Korrekturmaßnahmen erforderlich sind.

3. Quelloffene Foundation Models

Open-Source-Software (OSS) normativ einzuhegen, wirft bereits bei herkömmlichen Systemen zahlreiche Schwierigkeiten auf – etwa in puncto Lizenzierung, Urheberrechtsschutz oder Haftungsverteilung bei Schadensfällen, die auf der Nutzung der Software beruhen. Umso mehr gilt das bei KI-Modellen, die multiplen Verwendungsmöglichkeiten offenstehen.

a) Spezifika und Vorzüge quelloffener Systeme

Quelloffene Software zeichnet sich v.a. durch zwei Spezifika aus: Ihre Quellcodes sind (zumeist kostenfrei) öffentlich einsehbar und unterschiedliche Entwickler können diese regelmäßig überprüfen. Lizenznehmer der Software[68] dürfen bspw. die zugrundeliegenden Codes weiterverwenden und auch entsprechend modifizieren (gegebenenfalls unter der Verpflichtung, diese erneut zu veröffentlichen). Quelloffene Foundation Models basieren dadurch zumindest in Teilen auf Eingaben und Programmierungen mehrerer Entwickler und dienen als Grundlage neuer Entwicklungsprojekte.[69]

Den Quellcode (für jedermann) zugänglich zu machen, ermöglicht sowohl der Forschung als auch der Öffentlichkeit, die Funktionsweise der Systeme besser zu verstehen und sie weiterzuentwickeln – im Idealfall sogar unabhängig von großen Tech-Unternehmen. Frei verfügbare quelloffene Systeme bergen ferner das Potenzial, die Konzentration der Large Language Models auf die mit ihnen naturgemäß verbundenen kulturellen Maßstäbe ein Stück weit aufzubrechen, da Entwickler aus verschiedenen Staaten das System (inkl. bspw. seiner Filtertechniken) weiterentwickeln. Bias und Diskriminierungsgefahren lassen sich so im Idealfall abmildern und Monopolstellungen einzelner Tech-Giganten auf diesem zukunftsträchtigen KI-Feld aufbrechen.

b) Besondere Risiken

Der besondere Entstehungs- bzw. Entwicklungsprozess quelloffener Systeme bringt es mit sich, dass einige Risiken, die Basismodellen im Allgemeinen anhaften,[70] bei ihnen besonders ausgeprägt sind: Large Language Models müssen etwa gezielt und gesondert lernen, dass bestimmte in Datensätzen enthaltene Informationen verwerflich, gefährlich, falsch oder vertraulich sind, und dass sie diese daher nicht ausgeben dürfen. Entsprechende Kontrollen lassen sich im Rahmen quelloffener Systeme allerdings weitaus schwieriger durchführen.

Dass jedermann auf den Quellcode zugreifen kann, verstärkt auch die allgemeinen Gefahren von Foundation Models. Unterschiedliche Akteure können die auf ihnen aufsetzenden Anwendungen insbes. unbemerkt missbrauchen, zB indem sie diese einsetzen, um gezielte (politische) Desinformationskampagnen zu lancieren. Zudem besteht die Gefahr, dass Nutzer, die quelloffene Foundation Models weiterverwen-

62 Wie zB in Erwgr. 12 c, Art. 4b V, Art. 4c III KI-VO-E (Rat) sowie in Erwgr. 60 g, Art. 28 II UAbs. 2 KI-VO-E (EP).
63 So wohl auch Brown, Expert explainer: Allocating accountability in AI supply chains, Ada Lovelace Institute Blog, 29.6.2023.
64 Vgl. dazu IV. 5. b).
65 Vgl. auch Bommasani et al., On the Opportunities and Risks of Foundation Models, S. 156.
66 Ausf. dazu Hacker et al., Regulating ChatGPT and other Large Generative AI Models, S. 1116.
67 Vgl. auch IV. 4. b).
68 Vgl. etwa die Lizenzen der Open Source Initiative (OSI).
69 Ausf. und instruktiv dazu mit Blick auf generative KI-Systeme Solaiman, The Gradient of Generative AI Release: Methods and Considerations, 2023.
70 Vgl. II. 2.

den bzw. modifizieren, Sicherheitsmaßnahmen, die bereits Bestandteil eines Modells waren, nachträglich entfernen. Quelloffene Systeme können es aufgrund ihrer eingeschränkten Beherrschbarkeit im Extremfall vereinfachen, Cyberangriffe zu unternehmen oder terroristische Akte in der analogen Welt zu planen und ins Werk zu setzen.

c) Regulatorische Herausforderungen

aa) Anwendbare Vorschriften?

Die *allgemeine Ausrichtung des Rates* differenziert in ihrer Begriffsbestimmung für KI-Systeme mit allgemeinem Verwendungszweck nicht nach der Art und Weise, in der Anbieter GPAI-Systeme in Verkehr bringen bzw. in Betrieb nehmen (vgl. Art. 3 Nr. 1b KI-VO-E (Rat)). Quelloffene Modelle müssten folglich die Anforderungen des Titels III Kapitel 2 der VO (also die Vorgaben für Hochrisiko-KI-Systeme) ebenfalls erfüllen, soweit sie „als Hochrisiko-KI-Systeme oder als Komponenten von Hochrisiko-KI-Systemen (…) verwendet werden können" (Art. 4b I 1 KI-VO-E (Rat)). Ihre Anbieter wären den nach Art. 4b II KI-VO-E (Rat) anwendbaren Pflichten – u.a. Teilen des umfassenden Pflichtenkatalogs des Art. 16 KI-VO-E (Rat) – unterworfen und hätten somit bspw. ein Konformitätsbewertungsverfahren durchzuführen (Buchst. e) und Registrierungspflichten (Buchst. f) zu erfüllen.

Allerdings will der Ratsentwurf die KI-VO ausdrücklich nicht für solche KI-Systeme und deren Ergebnisse gelten lassen, die „eigens für den alleinigen Zweck der wissenschaftlichen Forschung und Entwicklung entwickelt und in Betrieb genommen werden" (Art. 2 VI KI-VO-E (Rat)). Eine weitere Ausnahme gesteht er „Forschungs- und Entwicklungsvorhaben zu KI-Systemen" zu (Art. 2 VII KI-VO-E (Rat)). Zumindest originäre Forschungs- und Entwicklungstätigkeiten sub specie quelloffener Systeme dürften daher nicht dem strengen Regelungsprogramm für Hochrisiko-KI-Systeme unterfallen.

Das *Parlament* erkennt die Bedeutung quelloffener Systeme für den europäischen KI-Markt und die KI-Entwicklung in seinem Entwurf im Grundsatz ausdrücklich an (vgl. Erwgr. 12 a S. 1 KI-VO-E (EP)). Bereits in den Erwägungsgründen[71] nehmen sie einen prominenten Platz ein. Zugleich möchte das Parlament quelloffene Foundation Models gleichwohl dem allgemeinen Anforderungs- und Pflichtenregime für Foundation Models unterwerfen.[72]

bb) KI-Wertschöpfungskette

Normative Vorgaben für Foundation Models (oder gar Regelungen für Hochrisiko-KI-Systeme) auf quelloffene Systeme anzuwenden, birgt neben der Frage, welche Vorschriften auf sie Anwendung finden sollen, in besonderer Weise die Schwierigkeit, die bereits proprietäre Foundation Models in puncto KI-Wertschöpfungskette aufwerfen.[73] Denn bei ihnen wirken noch mehr Akteure auf den (Weiter-)Entwicklungsprozess ein – bspw. in Fällen, in denen viele unterschiedliche Entwickler und potenzielle Anbieter gemeinsam ein System erstellen, oder wenn Dritte, etwa externe Model-Hub-Anbieter[74] oder Cloud-Anbieter, hinzukommen.

Um eine faire und gerechte Lastenverteilung sicherzustellen, bräuchte es flexible regulatorische Rahmenbedingungen. Die Vorschriften, die die Entwürfe derzeit vorhalten, und der generelle Ausschluss, auf den sich die Gesetzgebungsorgane im Trilog einigten,[75] tragen der hohen Komplexität und Vielfalt quelloffener Entwicklungsprozesse, die sich in kein einheitliches Raster pressen lassen, jedoch noch nicht hinreichend Rechnung. Ein Anbieter hätte zB nach Art. 4c I KI-VO-E (Rat) die Möglichkeit, die Anwendbarkeit der KI-VO zu umgehen, indem er jegliche Nutzung des Systems in Hochrisiko-Bereichen ausschließt; dies ist so allerdings bei quelloffenen Systemen in praxi kaum denkbar.

(1) Regulatorische Herausforderungen

Bei quelloffenen Foundation Models, die ein Unternehmen hauptverantwortlich entwickelt, sind die regulatorischen Herausforderungen insgesamt weniger komplex: Es ist als Anbieter im Sinne des Wertschöpfungskette einzustufen und daher Pflichtenadressat der KI-VO. Sofern das Unternehmen mit anderen Beteiligten zusammenarbeitet, könnte es – wie regelmäßig, sobald quelloffene Software zum Einsatz kommt – die Modalitäten durch gesonderte (Lizenz-)Vereinbarungen regeln.

Schwierigkeiten werfen jedoch Fallkonstellationen auf, in denen kein einzelner übergeordneter Akteur die Hauptverantwortung trägt. Wer dann tatsächlich für das System „verantwortlich" ist, und welcher Akteur für welchen Abschnitt des Lebenszyklus des KI-Systems das Pflichtenprogramm der KI-VO erfüllen soll, ist unklar.[76] Die Entwürfe regeln die Lastenverteilung zwischen mehreren Anbietern (nach jetzigem Stand) zumindest nicht ausdrücklich.[77]

(2) Vor- und Nachteile privilegierender Ausnahmeregelungen

So herausfordernd es auch ist, die Verantwortlichkeit bei quelloffenen Systemen zuzurechnen, so sehr ist ihnen im Rahmen der KI-Wertschöpfungskette[78] jedenfalls ein besonderer Vorzug eigen: Je nach Veröffentlichungsumfang verfügen die Anbieter nachgelagerter Anwendungen eher über die Informationen, die sie gegebenenfalls benötigen, um ihrerseits den Anforderungen der KI-VO nachzukommen. Zudem können bspw. Hub-Anbieter in der Lage sein, be-

71 Nr. 12a ff. KI-VO-E (EP).

72 Art. 2 V e 1 KI-VO-E (EP) iVm Erwgr. 12a S. 4 KI-VO-E (EP) nimmt „freie (…) und quelloffene (…) KI-Komponenten" zwar grds. aus dem Anwendungsbereich der VO aus. Wenn „sie (…) von einem Anbieter als Teil eines Hochrisiko-KI-Systems oder eines KI-Systems, das unter Titel II oder IV dieser Verordnung fällt, in Verkehr gebracht oder in Betrieb genommen (werden)" sind sie den Regeln jedoch grds. unterworfen (Erwgr. 12a S. 4, Art. 2 V e 1 Hs. 2 KI-VO-E (EP)). Für Foundation Models iSd Art. 3 soll dies zwar nicht gelten (Art. 2 V e 2 KI-VO-E (EP)). Art. 28b I KI-VO-E (EP) verpflichtet die Anbieter eines Foundation Models jedoch dazu, die besonderen Anforderungen des Art. 28b zu erfüllen, und zwar „unabhängig davon, ob es als eigenständiges Modell oder eingebettet in ein KI-System oder ein Produkt oder unter freien und Open-Source-Lizenzen als Dienstleistung sowie über andere Vertriebskanäle bereitgestellt wird". Um eine Anwendbarkeit der KI-VO auf quelloffene Foundation Models auszuschließen, bliebe folglich nur noch der Weg, bei Entwicklern, die quelloffene und kostenlose Systemkomponenten eines Basismodells veröffentlichen (Erwgr. 12 c KI-VO-E (EP)), die Anbietereigenschaft abzulehnen. Dies scheint das Parlament allerdings nicht im Sinn gehabt zu haben. Auch Inbetriebnahmen zu Forschungszwecken scheinen in den Anwendungsbereich des KI-VO-E (EP) zu fallen (Art. 2 V d und Ve KI-VO-E (EP)).

73 Vgl. IV. 2.

74 Ausf. Härlin et al., Exploring opportunities in the generative AI value chain, 26.4.2023.

75 Dazu Stierle, So will die EU Künstliche Intelligenz regulieren, Tagesspiegel Background, 11.12.2023.

76 Bei diesen Punkten ist jedoch auch entscheidend, inwieweit das System tatsächlich öffentlich ist. Wegen der Öffentlichkeit der Informationen automatisch die gesamte Verantwortung an den jeweils „nächsten" Akteur in der Lieferkette zu übertragen, wird der Komplexität der Sachverhalte zumindest nicht gerecht.

77 Vgl. IV. 2.

78 Ausf. dazu auch Küspert et al., The value chain of general-purpose AI, Ada Lovelace Institute Blog v. 10.2.2023.

stimmte Systeme zumindest auf den gängigsten Plattformen zu sperren.[79]

Zugleich geht von einem generellen Ausnahmevorbehalt für „quelloffene" Foundation Models die Gefahr aus, dass Unternehmen ihre Basismodelle „vorschnell" als quelloffen deklarieren. Das verlagerte die Verantwortung gegebenenfalls ausschließlich auf die Anbieter, die das Modell bspw. finetunen. Das Problem verschärft sich noch dadurch, dass es derzeit weder eine allgemeingültige Definition quelloffener KI gibt noch die Regelungen, die auf „traditionelle" OSS Anwendung finden, ohne Weiteres auf den Bereich der Künstlichen Intelligenz übertragbar sind.[80] Eine potenzielle Ausnahme, die ebenfalls quelloffene Basismodelle umfassen soll, müsste der Unionsgesetzgeber an eine klare Begriffsbestimmung knüpfen und die Regelung entsprechend konkret eingrenzen.

Allerdings läuft auch der Vorschlag, alle quelloffenen Systeme ohne Differenzierung der KI-VO zu unterwerfen, Gefahr, die Entwicklung (und den Einsatz) quelloffener Foundation Models in Europa nachhaltig zu schwächen. Unterschiedlicher Open-Source-Entwickler überhaupt habhaft zu werden, wenn sie selbst nicht in der EU ansässig sind und gerade kein rein kommerzielles Interesse verfolgen,[81] wird die Vollzugspraxis vor erhebliche Herausforderungen stellen.[82] Im Falle europäischer Entwickler besteht die Gefahr darin, dass diese im Gefolge strenger normativer Vorgaben der KI-VO eher geneigt sein werden, von einer Mitarbeit an quelloffenen Systemen gänzlich abzusehen. Schließlich drohen möglicherweise unmittelbar (persönliche) Haftungsrisiken und Lasten. Gegebenenfalls müssten sie zudem Anforderungen erfüllen, die sie finanziell und organisatorisch (zB in Fällen, in denen Unternehmen mit Forschenden kollaborieren und entsprechende Modelle oder Komponenten erstellen) nicht stemmen können oder wollen.[83] Beschränkt der Unionsgesetzgeber den Markt für Systeme, die ein Nutzer rechtmäßig verwenden kann, nachhaltig und sehen die europäischen Akteure deshalb von der Zusammenarbeit an quelloffenen Projekten oder der Nutzung von quelloffenen Systemen ab, könnte daraus ein entscheidender Nachteil für die europäische KI-Entwicklung und damit den Binnenmarkt insgesamt erwachsen. Dies widerspräche nicht zuletzt dem in sämtlichen Entwürfen zum Ausdruck kommenden allgemeinen Leitgedanken, die Entwicklung, Verwendung und Verbreitung von KI im Binnenmarkt zu fördern (Erwgr. 5).

Um die Risiken, die von quelloffenen Systemen ausgehen, hinreichend einzufangen, sollte der Verordnungsgeber in die finale Fassung zumindest umfassende Registrierungspflichten aufnehmen, die sicherstellen, dass nur derjenige an einem besonders leistungsfähigen System mitwirken kann, dessen Identität rückverfolgbar ist. An die missbräuchliche Nutzung des Modells sollten sich zudem scharfe Sanktionen knüpfen.

4. Transparentes Daten- und Risikomanagement sowie Evaluierung

Um die Risiken, welche Foundation Models anhaften, wirksam einzudämmen, braucht es nicht erst ab Markteintritt, sondern bereits für das Entwicklungsstadium ein passgenaues normatives Anforderungsregime der Risikosteuerung und Daten-Governance sowie der Evaluierung.

a) Daten und Daten-Governance

Soll das Wirken der Daten, die in Foundation Models eingehen, – auch in nachgelagerten Anwendungen – nachvollziehbar bleiben, sind transparente (Trainings-)Datensätze ab dem ersten Entwicklungsschritt unabdingbar.[84] Sofern in den Trainingsdatensätzen (die größtenteils aus archivierten Daten von Webseiten bestehen) bereits Bias angelegt sind, können sich diese in nachgelagerten Anwendungen fortschreiben und somit besondere Risiken oder Schäden für Betroffene, bspw. infolge von Diskriminierungen, nach sich ziehen. Dies steht etwa bereits dann in besonderer Weise zu befürchten, wenn keine hinreichenden (Trainings-)Datensätze aus unterschiedlichen Ländern, Sprachen und Kulturen verfügbar sind.[85]

aa) Ansätze der Regelungsentwürfe

Im ursprünglichen Entwurf der Kommission formuliert Art. 10 KI-VO-E (KOM) – ohne Bezug zu GPAI-Systemen – Vorgaben für Daten-Governance- und Datenverwaltungsverfahren (Art. 10 II KI-VO-E (KOM)).[86]

Der *Rat* möchte diese Vorgaben grds. auch auf GPAI-Systeme anwenden, sofern diese sich als Hochrisiko-KI-Systeme bzw. als deren Komponenten verwenden lassen (Art. 4b I KI-VO-E (Rat)). Er schwächt für sie jedoch die hohen Anforderungen an die Datenqualität aus Art. 10 III 1 KI-VO-E (KOM) deutlich ab: Die Datensätze müssen bspw. nur noch „so weit wie möglich fehlerfrei und vollständig sein" (vgl. Art. 10 III 1 KI-VO-E (Rat)). Der Rat wollte damit vermutlich der Kritik den Wind aus den Segeln nehmen, die technischen Möglichkeiten der Anbieter seien zu beschränkt, um das Niveau der Datenqualität, das die Kommission verlangt, in praxi gewährleisten zu können.

Das *Parlament* schlägt hingegen vor, Art. 10 KI-VO-E (KOM) nicht auf Foundation Models zu übertragen. Es sieht stattdessen für sie in Art. 28b KI-VO-E (EP) eine spezielle Regelung vor (vgl. insbes. II Buchst. b). Dieser Ansatz überrascht insofern, als die normativen Vorgaben des Art. 10 KI-VO-E (KOM) gerade bei Foundation Models von besonderer Bedeutung sind. Denjenigen, die Anwendungen auf ihnen aufsetzen, ist es regelmäßig nicht möglich, das Modell und die dahinterstehenden (Trainings-)Datensätze zu überprüfen. Dokumentationspflichten sind jedoch für jegliche Art externer Kontrolle unabdingbar.

Das erkennt das Parlament im Grundsatz auch an: Es stellt in Art. 4a iVm 28 b KI-VO-E (EP) Anforderungen auf, die ebenfalls Dokumentations- und Transparenzvorgaben einschließen. Allerdings bleiben diese Vorschriften recht vage und überlassen den Anbietern von Foundation Models somit weite Ermessens- und Entscheidungsspielräume: Art. 28b II Buchst. b KI-VO-E (EP) trägt Anbietern zwar auf, „nur Datensätze (zu) verarbeiten und ein (zu)beziehen, die angemes-

79 Bsp. von Hugging Face; vgl. Interview mit Solaiman, Responsible Release and Accountability for Generative AI Systems, Podcast Tech Policy Press v. 28.5.2023.

80 Instruktiv Widder et al., Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI, bspw. S. 1, 3; die Autoren argumentieren zudem u. a., dass entsprechende Ausnahmen für quelloffene Systeme unter Umständen tatsächlich die Marktpositionen der großen Tech-Unternehmen stärken könnten (vgl. S. 4, 18).

81 Letzteres ist im Falle der großen Tech-Unternehmen wohl ausgeschlossen, so dass das Argument wegen bspw. der großen Rechenleistungen, die die Entwicklung der Modelle benötigt, zumindest derzeit noch leerläuft.

82 Ein Ausschluss nach Art. 2 VII KI-VO-E (Rat) griffe wegen seiner absoluten Ausschließlichkeit dennoch selten.

83 Vgl. dazu Creative Commons et al., Supporting Open Source and Open Science in the EU AI Act v. 16.7.2023, S. 12.

84 So auch European Parliamentary Research Service, Auditing the quality of datasets used in algorithmic decision-making systems, Juli 2022, III.

85 Vgl. II. 2. d).

86 Dazu etwa ausf. Hilgendorf/Roth-Isigkeit/Spindler, Die neue Verordnung der EU zur Künstlichen Intelligenz, 2023, § 5 Rn. 21 ff. mwN.

senen Data-Governance-Maßnahmen für Basismodelle unterliegen". Es obliegt aber letztlich v.a. dem Anbieter selbst, festzulegen, welche konkreten Data-Governance-Maßnahmen er als angemessen einstuft. Insoweit mangelt es an entwickelten Standards, an denen sich alle Anbieter zu orientieren haben oder zumindest orientieren können – auch um eine entsprechende Vergleichbarkeit herzustellen.

Nach den Vorstellungen des Parlaments soll Normungsorganisationen die Aufgabe zufallen, die maßgeblichen Anforderungen nach Inkrafttreten der KI-VO zu präzisieren (vgl. Art. 40 ff. KI-VO-E (EP)). Zudem soll die Kommission für das Pflichtenprogramm des Art. 28b KI-VO-E (EP) unter bestimmten Voraussetzungen im Wege von Durchführungsrechtsakten „gemeinsame Spezifikationen" festlegen können (Art. 41 I a KI-VO-E (EP)). Zentrale grundrechtliche Anforderungen an die Datenmodelle festzulegen, muss jedoch dem Unionsgesetzgeber als Herzkammer des politischen Systems vorbehalten bleiben.

bb) Konkretisierungsvorschläge

Ein möglicher Ansatz, die offenen normativen Anforderungen an Datensätze in der KI-VO zu präzisieren, kann darin bestehen, die Dokumentation der Daten, die bei der Modellbildung Verwendung finden, auf die Beweggründe zu erstrecken, die die Datenauswahl und Datensammlung anleiten[87] – sowie die dokumentierten Datensätze zu kennzeichnen, um so für mehr Transparenz zu sorgen. Die denkbaren Ansätze reichen von Datasheets[88] bis hin zu sog. Dataset Nutrition Labels[89].

Eine solche Herangehensweise kann dazu beitragen, den Daten innewohnende Bias besser zu detektieren. Für die nachgelagerten Anwender von KI-Systemen mit konkretem Verwendungszweck wären solche Informationen ebenfalls relevant, insbes. wenn die entsprechenden Daten auf dem aktuellen Stand bleiben müssten. Ergänzend könnten sog. Modellkarten („Model Cards") zum Einsatz kommen, die Eigenschaften des trainierten Modells (wie zB seinen Typus, die beabsichtigten Anwendungsfälle, Leistungsverfahren und Messungen der Leistungen) erfassen, um die Transparenz zu erhöhen.[90] Zwar dürfte es nahezu unmöglich sein, die Datengrundlage des Trainingsprozesses im Detail nachzuvollziehen.[91] Denn die Trainingsdatensätze stammen aus verschiedenen Quellen und sind unübersehbar groß. Transparente Informationsblätter setzen aber zumindest ab dem Zeitpunkt der Auswahl der Datensätze eine konkrete Auseinandersetzung mit ihnen voraus.

Der Parlamentsentwurf verpflichtet Anbieter eines Foundation Models explizit, dieses in der öffentlichen EU-Datenbank für Hochrisiko-KI-Systeme (vgl. Art. 60 KI-VO-E (EP)) zu registrieren (Art. 28b II Buchst. g KI-VO-E (EP));[92] sie müssen zudem eine „Beschreibung der Datenquellen, die bei der Entwicklung des Basismodells verwendet wurden" hinzufügen[93] – eine technische Dokumentation muss der Anbieter demgegenüber nicht in die Datenbank einfügen.[94] Die vagen Vorgaben des Art. 28b II Buchst. b KI-VO-E (EP) schließen insbes. nicht die Pflicht ein, entsprechende Motivationen etc aufzuzeichnen. Der Vorschlag gesteht den Anbietern damit zu weite Entscheidungsspielräume in der wichtigen Frage zu, wie sie die Anforderungen erfüllen möchten. Der Gesetzgeber sollte Daten- und Modellkarten zudem explizit in sein Regulierungskorsett für Anbieter von Foundation Models integrieren.

Denkbar ist es auch, die Qualität von Datensätzen, die in algorithmischen Systemen zum Einsatz kommen, durch eine „database certification" abzusichern.[95] Wenn die Anbieter von Foundation Models die Datenbanken zertifizierten, könnte dies gerade den Anbietern nachgelagerter Anwendungen eine Orientierungshilfe an die Hand geben, um ein konkretes Modell auszuwählen, welches sie (zB für Hochrisiko-KI-Systeme) nutzen wollen. Den Anbietern der Foundation Models könnte eine Zertifizierung als Qualitätssignal Anreize setzen, sich (freiwillig) ihren strengen Anforderungen zu unterwerfen, um einen Wettbewerbsvorteil zu erlangen. Auf diese Weise ließe sich die Transparenz von Foundation Models ab dem ersten Entwicklungsschritt signifikant erhöhen. Das Parlament hat jedoch entsprechende Vorgaben nicht in seinen Verordnungsentwurf einfließen lassen.

b) Risikomanagement

So vielfältig und dynamisch die Risiken sind, die Foundation Models bergen, so sehr sind sie auf ein adäquates Risikomanagement angewiesen. Nur dann gelingt es hinreichend sicher, Gefahren zu identifizieren und – auch im Einzelfall – Methoden zu entwickeln, die diese wirksam eingrenzen. Zudem ermöglicht es, Risiken – sofern notwendig – rechtzeitig zu kommunizieren.

aa) Vorschlag des Rates und des Parlamentes

Verpflichtungen, ein Risikomanagementsystem[96] einzurichten, anzuwenden und aufrechtzuerhalten, kannte bereits der Entwurf der Kommission (Art. 9 I KI-VO-E (KOM)). Er nimmt aber nur die Anbieter eines Hochrisiko-KI-Systems in die Pflicht.

Der *Rat* möchte die Risikomanagementvorgaben auch auf GPAI-Systeme anwenden, soweit sie als Hochrisiko-KI-Systeme Verwendung finden können (Art. 4b I 1 KI-VO-E (Rat)) – mit zwei gravierenden Einschränkungen: zum einen nur für Risiken, „die mit Blick auf die Zweckbestimmung des Hochrisiko-KI-Systems höchstwahrscheinlich die Gesundheit, Sicherheit und Grundrechte beeinträchtigen",[97] zum anderen nur bei solchen Risiken, „die durch die Entwicklung oder Konzeption des hochriskanten KI-Systems oder durch die Bereitstellung ausreichender technischer Informationen angemessen gemindert oder behoben werden können" (Art. 9 II 2 Buchst. a, UAbs. 2 KI-VO-E (Rat)).

Das *Parlament* will bei dieser Frage einen anderen regelungstechnischen Weg einschlagen; seine Marschrichtung bleibt im Grundsatz jedoch ähnlich: Art. 9 KI-VO-E (EP) soll zwar keine Anwendung auf Foundation Models finden. Ein neuer

87 Bender et al., On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?, 2021, S. 618.
88 Aus ihnen sollen insbes. die Entstehung, Zusammensetzung, der vorgesehene Verwendungszweck, die Pflege und andere Eigenschaften der verwendeten Daten hervorgehen, Gebru et al., Datasheets for Datasets, 2018 (letzte Version 2021), S. 2 f.
89 Holland et al., The Dataset Nutrition Label: A Framework to Drive Higher Data Quality Standards, 2018.
90 Mitchell et al., Model Cards for Model Reporting, 2019, S. 1 f.
91 Liao/Wortman Vaughan (Microsoft Research), AI Transparency in the Age of LLMs: A Human-Centered Research Roadmap, S. 4.
92 Auch der Rat sieht die Registrierung vor, vgl. III. 2. a).
93 Anhang VIII – Abschnitt C (neu) Nr. 5 KI-VO-E (EP).
94 Mylly, IIC 2023, 1013 (1023).
95 European Parliamentary Research Service, Auditing the quality of datasets used in algorithmic decision-making systems, 2022, S. III.
96 Die Kommission versteht dieses als „kontinuierliche(n) iterative(n) Prozess", welcher sich auf den „gesamten Lebenszyklus" des Systems erstreckt (Art. 9 II 1 KI-VO-E (KOM)). Der Anbieter muss Einzelrisiken sowie das Gesamtrisiko identifizieren sowie auf ein vertretbares Maß reduzieren.
97 Der Terminus „Zweckbestimmung" wäre für die GPAI-Systeme anders zu lesen (Art. 4b VI KI-VO-E (Rat)).

Art. 28b II Buchst. a KI-VO-E (EP) soll den Anbietern von Foundation Models aber abverlangen, „vernünftigerweise vorhersehbare (…) Risiken für Gesundheit, Sicherheit, Grundrechte, Umwelt sowie Demokratie und Rechtsstaatlichkeit" vor und während des Entwicklungsprozesses zu identifizieren und zu minimieren. Konkrete „geeignete" Methoden bzw. Maßnahmen auszuwählen, obliegt hingegen dem Anbieter selbst. Die Verpflichtung bleibt daher unbestimmt.

Eine Grundrechte-Folgenabschätzung iSd Art. 29a KI-VO-E (EP) verlangt das Parlament den Anbietern der Foundation Models nicht ab – dabei könnte gerade diese dazu beitragen, besondere Gefahren für die Grundrechte vorab zu identifizieren. Immerhin müssen die Anbieter im Rahmen der Registrierung Informationen zu den „Leistungsgrenzen des Basismodells, einschließlich der vernünftigerweise vorhersehbaren Risiken und der ergriffenen Maßnahmen zu ihrer Minderung sowie der nicht geminderten Restrisiken mit einer Erklärung, warum sie nicht gemindert werden können" angeben.[98] Vor allem der letzte Punkt ermöglicht es im Idealfall, ein Bild der Risikominimierungsstrategie des Anbieters zu zeichnen.

Den Anbietern grds. Risikomanagementsysteme aufzuerlegen, ist bereits deshalb sinnvoll, weil diese Personen u.a. als Entwickler der Modelle über wertvolle Kenntnisse hinsichtlich potenzieller Risiken verfügen. Nur auf dem Weg eines frühzeitig ansetzenden, dokumentierten Risikomanagements lässt sich insbes. sicherstellen, dass Unternehmen zumindest die Basismodelle und ihre (möglichen) Risiken analysieren, bevor sie diese gegebenenfalls für eine breite (Weiter-)Verwendung veröffentlichen. Jegliche Verwendungen und Nutzungen ihrer Foundation Models samt der damit verbundenen Risiken vorherzusehen, dürfte aber selbst den Anbietern kaum möglich sein. Um dem Risikomanagementsystem nachhaltige Wirkung zu verleihen, ist der Gesetzgeber gehalten, flankierend eine externe Kontrolle vorzusehen. Dabei müssen die Schritte, die die Unternehmen wegen der von ihnen identifizierten Risiken unternehmen, überprüfbar sein.[99]

bb) Differenzierungsmodell

Um den unterschiedlichen Risikograden sowie den mit Beobachtungspflichten verbundenen Belastungen Rechnung zu tragen, kann ein regelungstechnischer Weg darin bestehen, bei den Anforderungen an das Risikomanagement zwischen verschiedenen Modellen, namentlich besonders systemisch risikoträchtigen und sonstigen Systemen, zu differenzieren.[100] In Art. 34 Digital Services Act (DSA) hat dieser Gedanke bereits Niederschlag gefunden:[101] Die Vorschrift verpflichtet sehr große Online-Plattformen sowie Suchmaschinen dazu, regelmäßig die ihren Systemen innewohnenden und daraus resultierenden systemischen Risiken (zB für die Verbreitung rechtswidriger Inhalte oder auf „negative Auswirkungen" auf Gesellschaft und Grundrechte[102]) zu ermitteln, zu analysieren und zu bewerten.

Diesen Grundgedanken auf das Risikomanagement im Rahmen der KI-Regulierung zu übertragen,[103] kann der Gefahr vorbeugen, dass die Risiken sehr großer generativer KI-Modelle in praxi voll durchschlagen. Für Foundation Models hätte dieser Ansatz ebenfalls Charme: Er trüge einerseits ihren Besonderheiten Rechnung, indem er (lediglich) verlangte, ihre konkreten systemischen Risiken zu ermitteln. Andererseits befreite er die Anbieter aber nicht von der Pflicht, mögliche Risiken und Gefährdungen ihrer Systeme präventiv in den Blick zu nehmen und – u.a. auf einer grundlegenden sowie systemischen Ebene – zu analysieren. Allerdings sollten sub specie der möglichen vielfältigen Implementierungen der Basismodelle in nachgelagerten Anwendungen auch für KMU (und OS-Kollaborationen) Vorgaben zum Risikomanagement greifen.

Auch der Trilog-Kompromiss der Unionsorgane unterscheidet nunmehr zwischen besonders leistungsfähigen Modellen und sonstigen Foundation Models.

Die Gesetzgebungsorgane haben schnell erkannt, dass es nicht zielführend ist, allein anhand von Nutzerzahlen zu differenzieren. Sie knüpfen vielmehr an Schwellenwerte von Rechenoperationen (eine bestimmte Anzahl Floating Point Operations Per Second)[104] an. Die Kommission ist befugt, diesen Schwellenwert zu verändern oder weitere Faktoren auszuwählen.[105]

Der derzeit festgelegte Schwellenwert greift jedoch zum einen sehr hoch. Er erfasst nur die wenigsten Modelle.[106] Es droht daher eine Schutzlücke.[107] Rechenoperations-Schwellenwerte sind zum anderen eindimensional. Stattdessen an ein Potpourri unterschiedlicher Schwellenwerte (bspw. mit Blick auf die verwendeten Parameter, Tokens, die Nutzerzahlen, die Rechenleistungen) anzuknüpfen, dürfte zwar in praxi sowohl schlecht umsetzbar als auch schwer überprüfbar sein (insbes. wegen der Notwendigkeit, rechtzeitigen Zugang zu allen dafür erforderlichen Informationen zu erhalten) und weitere Fragen aufwerfen (zB wie viele dieser Aspekte erfüllt sein müssen oder ob diese untereinander kompensationsfähig sind, in welchem Turnus alle Modelle zu überprüfen sind und von wem).[108]

Wegen der gravierenden Risiken, die Foundation Models und ihr Einsatz bergen, wäre es jedoch die bessere Lösung gewesen, alle Anbieter auf dem Markt verfügbarer Foundation Models im Grundsatz der Verpflichtung zu unterwerfen, ein Risikomanagementsystem zu betreiben. Die modellspezifischen Gefahren treten genauso bei „kleineren" Modellen auf und können – nicht zuletzt auf nachgelagerter Ebene – erhebliche Schäden bei Betroffenen hervorrufen.[109] Gerade bei Foundation Models ist Risikomanagement strukturell eine Daueraufgabe. Wenn ein Anbieter etwa die Trainingsdaten verändert, um ein ihm angezeigtes Diskriminierungsmuster zu unterbinden, können Ergebnisse die Folge sein, die wegen des neuen Trainings nunmehr andere „neue" Be-

98 Anhang VIII – Abschnitt C (neu) Nr. 6 KI-VO-E (EP).
99 Zur Kritik an dem Vorschlag, in den Risikoanalysen auch potenzielle zukünftige Gefahren prognostizieren zu müssen, Hacker et al., Regulating ChatGPT and other Large Generative AI Models, S. 1115 f.
100 Vgl. zu dieser Idee insbes. Anderljung et al., Frontier AI regulation: Managing Emerging Risks to Public Safety, 7/2023.
101 Helberger/Diakopolous, ChatGPT and the AI Act, Internet Policy Review, 2023, S. 4.
102 Gerdemann/Spindler RDi 2023, 183 Rn. 54, Raue/Heesen NJW 2022, 3537 Rn. 37.
103 So etwa die Forderung von Helberger/Diakopolous, ChatGPT and the AI Act, S. 4.
104 Vgl. bspw. Stierle, So will die EU Künstliche Intelligenz regulieren, Tagesspiegel Background v. 11.12.2023.
105 Stierle, So will die EU Künstliche Intelligenz regulieren, Tagesspiegel Background v. 11.12.2023.
106 The Future Society, EU AI Act Comliance Analysis, 2023, 4; Hacker, What's Missing from the EU AI Act, Verfassungsblog, 13.12.2023.
107 Instruktiv zu den Problemen der Abgrenzung vgl. bspw. Strait, Emerging processes for frontier AI safety, The Ada Lovelace Institute v. 27.10.2023.
108 Diese Entscheidungen an die Kommission auszulagern, ist zudem unter demokratischen Gesichtspunkten problematisch.
109 Zur Kritik, dass sich das UK AI Safety Summit 2023 auf Frontier AI konzentrierte, bspw. Morrison, UK government urged to widen scope of AI safety summit beyond frontier models, Tech Monitor v. 28.9. 2023.

troffene diskriminieren und zB Stereotype bedienen. Aus diesem Grund jeweils alle vor der Änderung bereits vorgenommenen Risikobewertungen bei jedem neuen Trainingslauf erneut zu überprüfen, dürfte die Zumutbarkeitsgrenze überschreiten. Der Unionsgesetzgeber sollte deshalb einerseits klarstellend festschreiben, dass das Risikomanagementsystem während des gesamten KI-Lebenszyklus aufrechtzuerhalten ist. Der Pflichtenradius sollte sich andererseits aber nur auf „vernünftigerweise vorhersehbare Risiken" erstrecken, die zB auch ein sachverständiger Dritter prognostiziert hätte.

Für KMU bräuchte es ein Sonderregime, das weniger weitgehende Verpflichtungen statuiert, da eine umfassende dauerhafte Überprüfung und ein dementsprechend engmaschiges Risikomanagementsystem ihre (finanziellen) Ressourcen in vielfältiger Weise überfordern dürfte. Aus diesen Gründen ganz auf es zu verzichten, darf angesichts der hohen Risiken für die Grundrechte allerdings keine Lösung sein.

c) Nachträgliche Evaluierungen

Im Verlauf des Lebenszyklus eines Foundation Models ist es unumgänglich, Risiken und Problemfelder der Modelle und ihres Einsatzes in verschiedenen Kontexten zu evaluieren.

Auf internationaler Ebene existieren bereits Vorstöße, Soft-Law-Instrumente zu implementieren, welche dies vereinfachen sollen. So schlagen etwa US-amerikanische Wissenschaftler ein „Foundation Model Review Board" vor.[110] Das HAI der Stanford University entwickelte die sog. „HELM"-Methode („Holistic Evaluation of Language Models"),[111] um die Transparenz großer Sprachmodelle zu erhöhen. Anders als herkömmliche Verfahren soll HELM besonders in Rechnung stellen, dass sich Sprachmodelle „für viele unterschiedliche Zwecke" nutzen lassen.[112] Um sie sinnvoll evaluieren zu können, sieht es u.a. einen „Zugriff auf das Modell" und „einen einheitlichen Standard" vor, zudem will es „alle Faktoren (…) in umfassender Weise betrachte(n)".[113] Auch HELM zielt jedoch lediglich auf Sprachmodelle ab[114] – die EU benötigt hingegen Methoden, die auch andere Modelle abdecken.[115]

aa) Vorschlag des Parlaments: ein Amt für Künstliche Intelligenz

Um Foundation Models ganzheitlich zu erfassen, hat das Parlament vorgeschlagen, auf unionaler Ebene eine neue Aufsichtsbehörde einzurichten – das sog. Amt für Künstliche Intelligenz.[116] Für dieses etabliert Art. 56b KI-VO-E (EP) einen breit gefächerten Aufgabenkatalog. Das Amt soll zukünftig u.a. den Dialog mit den Entwicklern von Foundation Models institutionalisieren sowie die Anforderungen des Art. 28b KI-VO-E (EP) und die „bewährte(n) Verfahren der Branche für die Selbstverwaltung" beaufsichtigen sowie überwachen (vgl. Art. 56b Buchst. q S. 1 KI-VO-E (EP)). Ferner soll es jährlich Leitlinien zu den Schwellenwerten veröffentlichen, „ab denen das Trainieren eines Basismodells als großer Trainingslauf gilt"; zudem möchte das Parlament „bekannte Fälle von großen Trainingsläufen" aufgezeichnet sowie überwacht, ferner „den Rechts- und Verwaltungsrahmen für solche Modelle (…) regelmäßig bewerte(t)" sehen und einen jährlichen Bericht „über den Stand der Entwicklung, Verbreitung und Nutzung von Basismodellen" veröffentlichen (Art. 56b Buchst. r KI-VO-E (EP); Erwgr. 60 h S. 4 KI-VO-E (EP)). Des Weiteren sollen die europäischen Benchmarking-Behörden (vgl. auch Art. 15 I a KI-VO-E (EP)) im Tandem mit dem Amt für Künstliche Intelligenz „kosteneffiziente Leitlinien und Kapazitäten zur Messung und zum Vergleich von Aspekten" für u.a. Foundation Models setzen, die von Bedeutung sind, um die KI-VO befolgen zu können (Art. 58a KI-VO-E (EP)).

Allerdings setzt der KI-VO-E (EP) vorrangig auf interne Prüfungen. So sehr das normative Wissen um Konformitätsbewertungen noch bruchstückhaft und die Methodik „zur Prüfung durch Dritte (…) noch in der Entwicklung" (Erwgr. 60 h S. 1 KI-VO-E (EP)), ist, so wenig ist es sachgerecht, der „Branche" selbst zu überlassen, „neue Methoden zur Bewertung von Basismodellen" zu entwickeln (vgl. Erwgr. 60 h S. 2 KI-VO-E (EP)).[117] Ein solcher Weg machte diese im schlimmsten Fall zum Richter in eigener Sache und wird damit einem angemessenen Grundrechtsschutz nicht gerecht.

Die Idee eines Benchmarkings als solche (vgl. Art. 58a KI-VO-E (EP)) überzeugt prinzipiell. In ihrer aktuellen Ausgestaltung bleibt sie aber zu unkonkret und vage. Ein Dialog zwischen den Anbietern und dem Amt für Künstliche Intelligenz (Art. 56b Buchst. o, q KI-VO-E (EP)) ersetzt keine Transparenzanforderungen oder externe Bewertungen, die sich an klaren Benchmarks orientieren. Ohne eine valide, umfassende Evaluierungsmethode geht ein Benchmarking im Ergebnis nicht über einen Papiertiger hinaus. Der Erfolg einer Evaluierungsmethode ist zudem maßgeblich von verfügbaren Informationen über die Modelle abhängig. Schließlich bräuchte es bereits umfassenden Zugang zu den relevanten Datensätzen, um entsprechende Standards auszuarbeiten. Viele der großen Tech-Unternehmen und Anbieter der Foundation Models sind allerdings nicht gewillt, diese in ausreichendem Maße preiszugeben.

bb) System eines Informationsaustauschs

Eine erste Idee, wie es gelingen könnte, mehr Informationen für bspw. die Entwicklung von „Methoden zur Prüfung durch Dritte" (vgl. Erwgr. 60h S. 1 KI-VO-E (EP)) zu gewinnen, kann darin bestehen, ein (zunächst freiwilliges) System des Informationsaustausches (zB via Schnittstelle) zwischen Regierung und Anbietern zu etablieren. Es könnte bspw. Bewertungen der Modellkompetenzen vornehmen und Risiken, die von Foundation Models ausgehen, frühzeitig detektieren.[118] Ein solcher tiefer Einblick ist die unerlässliche Basis dafür, hinreichende Informationen zu sammeln, um die Modelle in geeigneter Form nachträglich zu evaluieren sowie Standards und Benchmarks zu entwickeln. Ferner erleichtert er einen fortlaufenden Austausch mit der KI-Wirtschaft.

110 Liang et al., The Time Is Now to Develop Community Norms for the Release of Foundation Models, Center for Research on Foundation Models (CRFM), Stanford Institute for Human-Centered Artificial Intelligence (HAI) at Stanford University, 2022.
111 Vgl. Center for Research on Foundation Models (CRFM), Stanford Institute for Human-Centered Artificial Intelligence (HAI) at Stanford University, HELM; ausf. dazu Liang et al., Holistic Evaluation of Language Models.
112 Bommasani et al., Improving Transparency in AI Language Models: A Holistic Evaluation, Issue Brief HAI Policy & Society, 02/2023, S. 1.
113 Bommasani et al., Issue Brief HAI Policy & Society, 2/2023, S. 2.
114 Ebenso bspw. BIG-Bench, vgl. Srivastava et al., Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models, 2023; instruktiv auch Mökander et al., Auditing large language models: a three-layered approach, AI and Ethics, 2023.
115 Instruktiv zu Sozialverträglichkeitsprüfungen verschiedener generativer KI-Systeme (Text, Audio, Bild und Video) Solaiman et al., Evaluating the Social Impact of Generative AI Systems in Systems and Society, under review, 2023.
116 Vgl. auch Erwgr. 76 KI-VO-E (EP).
117 Erwgr. 60h S. 2 KI-VO-E (EP) führt insoweit exemplarisch „Modellevaluierung, Red-Teaming oder Verifizierungs- und Validierungstechniken des maschinellen Lernens" auf.
118 Ausf. dazu Martini, Blackbox Algorithmus, 2019, S. 130 f., 256, 351.

Eine denkbare Blaupause für ein solches System liefert womöglich das Vereinigte Königreich: Der britische Premierminister verkündete im letzten Jahr, dass drei Unternehmen, die Foundation Models entwickeln und betreiben, den staatlichen Behörden nunmehr einen „frühen Zugang“ zu den Modellen zugesichert haben.[119] Dieser soll dabei helfen, „mögliche Risiken zu verstehen“.[120] Wie dieser Zugang bzw. diese Vereinbarung konkret ausgestaltet ist oder sein wird, ist derzeit noch nicht bekannt – ebenso wenig, ob dieser frühe Zugang genug bewirken kann, um „der (…) Regierung ausreichend gründliche und langfristige Vorausschau zu gewähren.“[121] Auf ähnliche Vorstöße einigten sich bspw. die Beteiligten des AI Safety Summit 2023 im Vereinigten Königreich.[122] Dem britischen Vorbild[123] folgend könnten solche Verständigungen auch für die EU ein erster sinnvoller Schritt sein – dürfen aber zugleich keine Pflichten der KI-VO ersetzen.

cc) Schlussfolgerungen

In summa ergänzt eine nachträgliche Evaluation von Foundation Models den Regelungsrahmen, den die KI-VO zieht, in sinnvoller Weise – allerdings sollte bereits die Veröffentlichung des Modells voraussetzen, dass es die notwendigen Sicherheitsaspekte und Anforderungen der europäischen Gesetze erfüllt. Zumindest die in den KI-VO-Entwürfen für Hochrisiko-KI-Systeme vorgesehene Konformitätsbewertung (vgl. etwa Art. 40 ff. KI-VO-E (KOM)) sollte ebenso auf Foundation Models und generative KI-Systeme Anwendung finden.[124] Wegen der beträchtlichen Risiken, die sich zudem in (allen) nachgelagerten Anwendungen niederschlagen können, empfiehlt sich jedoch ein externes Verfahren. So lässt sich auch für nachfolgende Anbieter ein Mindestmaß an Rechtssicherheit herstellen: Sie können davon ausgehen, dass sie die Modelle auf dem europäischen Markt für ihre Entwicklungen nutzen können, da diese zumindest die erforderlichen (Mindest-)Standards erfüllen.

d) Zwischenergebnis

Die Unionsorgane bemühen sich intensiv darum, Foundation Models in das regulatorische Korsett der KI-VO zu spannen. Ihre zweckoffene Struktur liegt jedoch quer zu der Logik der nach konkreten Einsatzformen fragenden Risikoklassen des regulatorischen Gesamtkonzepts. Es macht eben einen Unterschied, ob ein Foundation Model zum Einsatz kommt, um Gesichter zu erkennen oder Text sprachlich zu optimieren.

Obgleich die KI-VO nicht die Technik als solche regulieren soll, sondern einzelne gefahrenträchtige Anwendungen, die von der Technik ausgehen, ist es sachgerecht, den grundrechtlichen, rechtsstaatlichen und demokratietheoretischen Risiken, die von Foundations Models ausgehen, bereits auf der Modellebene zu begegnen. Nicht zuletzt sind diejenigen, die auf den Modellen Anwendungen aufsetzen, technisch regelmäßig gar nicht in der Lage, allen Diskriminierungsrisiken oder Transparenzmängeln der Systeme zu begegnen. So sind etwa Vorkehrungen vonnöten, die sicherstellen, dass bereits die Systementwickler – im Rahmen des Möglichen – nur Trainingsdatensätze auswählen, die ethisch vertretbar sind und sich im Einklang mit den geltenden Datenschutzvorgaben verarbeiten lassen. Der Gesetzgeber muss insoweit zumindest für die notwendige Transparenz sorgen, indem er den Anbietern (sowohl von Foundation Models als auch von darauf basierenden Hochrisiko-KI-Systemen) aussagekräftige Dokumentationen abverlangt. Nur auf dieser Basis können externe Stellen bzw. Behörden Datensätze und etwaige Filtermechanismen zuverlässig überprüfen.

Unterdessen setzt sich der Regulierungsansatz durch, Foundation Models einzelnen Anforderungen für Hochrisiko-Systeme zu unterwerfen, aber zwischen verschiedenen Systemtypen zu unterscheiden: Nur solche Modelle, von denen ein systemisches Risiko ausgeht, sollen strengen Vorgaben unterliegen. Dahinter steckt nicht zuletzt die regulierungspolitische Strategie, den Aufstieg insbes. kleinerer unionaler Anbieter, wie *Aleph Alpha* und *Mistral AI*, nicht zu stark auszubremsen. Die Trilog-Fassung rekurriert als Trennlinie v.a. auf die Rechenstärke. Ab einer Rechenleistung von 10~25 FLOPS vermutet der Unionsgesetzgeber in Zukunft grds. ein systemisches Risiko. Die Kommission soll künftig entsprechende KI-Modelle auf dieser Grundlage sowie qualitativer Kriterien, welche ein systemisches Risiko begründen, in einer Liste erfassen.[125]

Ob die Rechenleistung jedoch ein tauglicher Gradmesser für das Risiko von Modellen ist, ist zu bezweifeln. An die Rechenleistung anzuknüpfen, gleicht dem Versuch, den jugendgefährdenden Charakter eines Filmes an seiner Länge festzumachen. Es wird erforderlich sein, vorrangig qualitative Kriterien, bspw. die Art und Weise, wie Trainingsdaten erhoben und eingespeist werden, welche Schutzmechanismen bestehen, für den Gefährdungsgrad auszumachen, die von einem Foundation Model ausgehen.

Kleine Foundation Models von Regulierungsanforderungen vollständig auszunehmen, wäre ebenfalls nicht angebracht. Sie sollten Mindestanforderungen an die (Daten-/Cyber-) Sicherheit, die Auswahl der Trainingsdaten und den Schutz vor Missbrauch etc einhalten müssen. Nur so lassen sich die weitreichenden Risiken von Foundation Models bestmöglich einhegen und nur so entspricht das Regulierungsniveau dem besonderen Gefährdungsgrad, den die Modelle – unabhängig von der Größe bzw. Umsatz ihrer Anbieter – bergen.

5. Lehren aus anderen Regulierungsfeldern für die KI-VO

Die Rechtsordnung steht nicht zum ersten Mal vor der Aufgabe, normative Vorgaben für Angebote zu schaffen, die sich für sehr vielfältige Zwecke nutzen lassen. Vergleichbare Fragen stellen sich insbes. bei sog. Dual-Use-Gütern, also solchen Produkten, die sowohl für zivile als auch für militärische Zwecke verwendbar sind (a) – ebenso bei Chemikalien, die sich in sehr unterschiedlichen Szenarien einsetzen lassen und bei Arzneimitteln, die unvorhergesehene Nebenwirkungen hervorbringen (b). Die Antworten, die der Gesetzgeber in diesen Konstellationen gefunden hat, können wertvolle Anhaltspunkte für die Regulierung von Basismodellen liefern. Dies gilt bspw. in Hinblick auf die Frage, wer für die Systeme und die potenziellen Risiken verantwortlich zeichnet

119 Sunak, Tweet v. 12.6.2023.
120 Sunak, Tweet v. 12.6.2023.
121 Mulani/Whittlestone, Proposing a Foundation Model Information-Sharing Regime for the UK, Centre for the Government of AI, 16.6.2023.
122 Vgl. bspw. Press Release der britischen Regierung v. 2.11.2023. Krit. mit Blick auf das AI Summit bspw. Riekeles/v. Thun, Rishi Sunak's AI plan has no teeth – and once again, big tech is ready to exploit that, Opinion, The Guardian v. 16.11.2023; Open Letter unterschiedlicher Akteure, AI Safety Summit Open Letter to the UK Prime Minister v. 30.10.2023.
123 Zu dem konkreten Vorschlag s. Mulani/Whittlestone, Proposing a Foundation Model Information-Sharing Regime for the UK, Centre for the Government of AI, 16.6.2023.
124 So wie es bereits der Rat für GPAI-Systeme, vgl. Art. 4b KI-VO-E (Rat), und das Parlament für die Foundation Models, vgl. Art. 40 ff. KI-VO-E (EP), vorgeschlagen haben.
125 FAZ, Strikte EU-Auflagen für ChatGPT-Basismodell v. 9.12.2023, 21.

und inwieweit eine ausgereifte Kommunikation aller Beteiligten notwendig ist, um Risiken zu minimieren.

a) Dual-Use-VO

Wie den Gefahren zu begegnen ist, die aus dem Export von Gütern mit denkbarem doppeltem (sc. zivilem und militärischem) Verwendungszweck erwachsen,[126] regelt die Union in der Dual-Use-VO (VO (EU) 2021/821)[127]. Um Risiken für die europäische Sicherheit sowie für die Menschenrechte und das humanitäre Völkerrecht einzuhegen, die von multipel einsetzbaren Produkten ausgehen, entschied sie sich dafür, ein *Lizenzierungssystem* einzuführen: Besonders risikobehaftete Dual-Use-Güter zu exportieren, bedarf der vorherigen Genehmigung der nationalen Exportkontrollbehörden.[128] Welche konkreten Güter dies betrifft, ergibt sich grds. aus einer Kontrollliste (Art. 3 I iVm Anhang I Dual-Use-VO). Der Genehmigungspflicht unterliegen unter bestimmten Voraussetzungen aber ebenfalls weitere, nicht gelistete Güter (vgl. Art. 3 II Dual-Use-VO) – insbes. sofern diese Güter den sog. Catch-all-Klauseln der Dual-Use-VO unterfallen, wenn also die Kenntnis oder der begründete Verdacht besteht, dass sie militärische Verwendung finden.

Bei der Exportkontrolle von Dual-Use-Gütern nimmt folglich der Staat eine Schlüsselrolle ein. Gleichwohl müssen Unternehmen zusätzlich besondere Sorgfalt bei ihren Exporten walten lassen: Sie können dazu verpflichtet sein, sog. Internal Compliance Programme (ICP) einzuführen, also Strategien und Verfahren, die ein rechtskonformes Verhalten fördern sollen[129] (Erwgr. 7, Art. 12 IV UAbs. 3 Dual-Use-VO). Das gilt insbes. bei Globalgenehmigungen (Erwgr. 18 S. 3 Dual-Use-VO). Die verantwortlichen Unternehmen müssen dann Risiko- und Betroffenheitsanalysen vornehmen.[130] Eine „Risikobewertung" kann und muss „nicht alle potenziellen künftigen Schwachpunkte und Risiken (…) aufzeigen".[131] Es geht vielmehr, nur darum, diejenigen zu finden, „die dann im Rahmen des ICP verringert werden können."[132] Obgleich die unterschiedlichen Verwendungsmöglichkeiten der Güter also offen sind, obliegt es dem Unternehmen, im Rahmen dieser Analysen zu prüfen, welche Aspekte das ICP enthalten sollte, um die normativen Vorgaben der Dual-Use-VO möglichst effektiv einzuhalten.

Ebenso wie den Herstellern von Dual-Use-Produkten ist es auch Anbietern von Foundation Models nicht möglich, jegliches Risiko, dass Akteure ihre Modelle gegebenenfalls missbräuchlich verwenden könnten, präventiv auszumerzen. Der Gesetzgeber tut dennoch gut daran, die Last, sich – präventiv – mit potenziellen Risiken auseinandersetzen zu müssen, nicht gänzlich von den Schultern der verantwortlichen Unternehmen zu nehmen. Ohne deren Unterstützung können die nachgelagerten Anbieter bzw. Nutzer in vielen Fällen die Verantwortung für ihre Modelle mit konkretem Verwendungszweck nicht vollständig übernehmen. Auch der Staat ist nicht in der Lage, diese Rolle vollständig auszufüllen. Deshalb sollten Unternehmen – nach dem Vorbild der Dual-Use-VO – „Foundation Model ICP" etablieren, und auf deren Grundlage (je nach ermitteltem Risiko) die notwendigen Schritte einleiten, um die Compliance mit den übrigen Anforderungen der KI-VO und sonstigem Unionsrecht zu gewährleisten. Die kontinuierliche Zusammenarbeit zwischen staatlichen Behörden und den betreffenden Unternehmen ist auch für Basismodelle im Rahmen der KI-VO geboten.[133]

Ein staatliches Genehmigungssystem, wie es die Dual-Use-VO etabliert, schösse für Foundation Models demgegenüber über das Ziel hinaus.[134] Es schwächte die Innovationskraft in diesem potenzialträchtigen KI-Sektor substanziell, ohne als Ex-ante-Prüfung einen hinreichenden Mehrwert sicher versprechen zu können. Die Verwendungsmöglichkeiten der Modelle sind so mannigfaltig und dynamisch, dass gesetzliche Genehmigungspflichten, die sich u.a. an Listen orientieren, schnell an ihre Grenzen stoßen. Eine Catch-all-Klausel ist ebenfalls nicht generell sinnvoll. Sie kann aber in Gestalt einer gesteigerten Prüf- und Beobachtungspflicht des Betreibers in solchen Bereichen eine sinnvolle Funktion entfalten, für die der begründete Verdacht besteht, dass Foundation Models missbräuchlich Einsatz finden.

Ein weiterer regulatorischer Ansatz der Dual-Use-VO lässt sich übertragen: Die Anbieter der Foundation Models sollten – vergleichbar einer Endverbleibsklausel im Exportkontrollrecht (vgl. Art. 12 IV UAbs. 2 Dual-Use-VO) – die Möglichkeit erhalten, spezifische Einsatzszenarien in nachgelagerten Anwendungen vertraglich auszuschließen (bspw. wenn es sich hierbei um verbotene Praktiken iSd Art. 5 KI-VO handelt) und sich dadurch von der Verantwortung für Missbrauch grds. regulatorisch freizeichnen zu können.[135] Diese Vertragsklausel müsste aber hinreichend konkret und durch flankierende technische Sicherungsmaßnahmen gegen missbräuchliche Nutzung abgesichert sein. Sonst öffnet sich ein Schlupfwinkel, über dessen Hintertreppe die Entwickler der Foundation Models ihre Verantwortlichkeit pauschal via Haftungsausschluss auf nachgelagerte Anwender auslagern können, die im Zweifel nicht über das Wissen und die Mittel verfügen, alle notwendigen Maßnahmen einzuleiten oder jegliche Pflichten zu erfüllen.[136]

b) REACH-VO und Pharmakovigilanz im AMG

Ähnlich wie bei Foundation Models steht der Unionsgesetzgeber bei der Chemikalienregulierung vor der Herausforderung, ein normatives Korsett für komplexe Sachverhalte mit geschichteten Verantwortlichkeitssphären bei Produkten zu schnüren, deren Wirkungen sich ex ante nicht sicher voraussagen lassen. Wie Foundation Models kommen auch chemische Stoffe in ganz unterschiedlichen Einsatzszenarien zur

126 In den letzten Jahren ist insoweit noch ein zusätzliches Problem entstanden: Nach dem Verwendungszweck der Güter (zivil oder militärisch) eindeutig abzugrenzen, ist wegen der technologischen Fortschritte häufig nur noch schwer möglich. Vormals klare Zuordnungen verschwimmen zunehmend. Zudem schließt die Dual-Use-VO nicht nur gegenständliche Güter ein, sondern ebenso (immaterielle) Technologien und Software (Art. 2 Nr. 1 Dual-Use-VO). Somit kommt auch ein KI-System grds. als Dual-Use-Gut in Betracht.

127 ABlEU 2021 L 206, 1.

128 In Deutschland ist dies im Wesentlichen das Bundesamt für Wirtschaft und Ausfuhrkontrolle (BAFA).

129 Zur Definition des Begriffs s. Art. 2 Nr. 21 Dual-Use-VO sowie die Empfehlung 2021/1700 der Kommission, ABlEU 2021, L 338, 1 (7).

130 Für den konkreten Inhalt der ICP hält der Verordnungsgeber keine Schablone vor. Vielmehr variieren die erforderlichen Maßnahmen je nach Unternehmensgröße, Geschäftsbereich und Kundenstamm. Empfehlung 2019/1318/ (EU) der Kommission v. 30.7.2019 zu internen Compliance-Programmen für die Kontrolle des Handels mit Gütern mit doppeltem Verwendungszweck (Dual-Use-Gütern) nach Maßgabe der VO (EG) 428/2009 des Rates, L 205/17.

131 Fn. 130, L 205/19.

132 Fn. 130, L 205/19.

133 Alternativ ließen sich Erwgr. 78 KI-VO-E (bzw. Art. 61 ff. KI-VO-E oder sogar Art. 21 KI-VO-E (EP)) in zweifacher Weise ausweiten: Zum einen sollten sie auch die Anbieter von Foundation Models umfassen, zum anderen sollte eine Meldung nicht lediglich dann erfolgen müssen, wenn es zu schwerwiegenden Vorfällen kam.

134 Vgl. dazu auch Martini, Blackbox Algorithmus, 2019, 229.

135 Sofern die Anbieter die Foundation Models nicht vollständig quelloffen bereitstellen.

136 Ausf. dazu auch der offene Brief von Gebru et al., Five considerations to guide the regulation of „General Purpose AI" in the EU's AI Act, Apr. 2023.

Anwendung; stoffbezogene Risikobestimmungen können sich zudem noch während des Einsatzes oder „Lebenszyklus“ des Stoffes verändern.

Um den besonderen Gefahren von Chemikalien für Gesundheit und Umwelt wirksam zu begegnen, etabliert die REACH-VO (VO (EG) 1907/2006)[137] ein umfassendes Informationsmanagementsystem: Da das Wissen über die Sicherheitseigenschaften der Stoffe über viele Akteure verteilt ist, gibt sie der Industrie auf, über die gesamte Wertschöpfungskette eines chemischen Stoffes hinweg alle relevanten Informationen zu sammeln und zu teilen, um dessen gefährliche Eigenschaften sowie Empfehlungen für Risikomanagementmaßnahmen zu ermitteln und weiterzugeben (vgl. Erwgr. 17 REACH-VO). Dies soll nicht nur einen kontinuierlichen Informationsfluss zwischen allen Beteiligten – Unternehmen in der Lieferkette und Behörden, aber auch Verbrauchern und der interessierten Öffentlichkeit – sicherstellen, sondern auch den Akteuren in der Wertschöpfungskette einen Anreiz setzen, Stoffe gegebenenfalls durch sicherere Alternativen auszutauschen, und einen Wettbewerb initiieren, der Akteure belohnt, die möglichst sichere Stoffe verwenden.[138]

Möchte ein Unternehmen einen neuen chemischen Stoff auf dem europäischen Markt lancieren, muss es den Stoff zunächst registrieren (Art. 6 I REACH-VO).[139] Der Hersteller des Stoffes muss dem Abnehmer grds. ein sog. Sicherheitsdatenblatt zur Verfügung stellen (vgl. Art. 31 I, III und IV REACH-VO). Die Informationspflicht greift aber genauso in umgekehrter Richtung, dh alle Akteure in der Lieferkette müssen nicht nur nachgelagerten Akteuren, sondern auch dem vorgeschalteten Akteur oder Händler Erkenntnisse mitteilen, bspw. wenn ihnen „neue Informationen über gefährliche Eigenschaften“ zur Verfügung stehen (vgl. Art. 34 UAbs. 1 Buchst. a REACH-VO). Die Händler reichen diese Informationen wiederum an den vorgeschalteten Akteur bzw. Händler weiter (vgl. Art. 34 UAbs. 2 REACH-VO). Das ermöglicht es allen anderen Akteuren, auf die neuen Gegebenheiten zu reagieren. Eine weitere Besonderheit: Sofern ein nachgeschalteter Anwender einen Stoff anders beurteilt als sein Lieferant, setzt er zudem die Europäische Chemikalienagentur davon in Kenntnis (Art. 38 IV, Art. 75 ff. REACH-VO).

Die REACH-VO etabliert so ein umfassendes Informationsökosystem, um die Gefahren der Stoffe bestmöglich erfassen und kontrollieren zu können. Alle beteiligten Akteure sind zu jedem Zeitpunkt gefordert, jegliche neue Informationen und vom bisherigen Wissensstand abweichende Erkenntnisse aufzunehmen und mit den weiteren Beteiligten zu teilen. In Situationen, in denen verschiedene Anwender chemische Stoffe in unterschiedlichen Einsatzszenarien verwenden oder wenn bspw. ein Gemisch[140] entsteht, für welches wiederum andere Sicherheitsstandards gelten können, hilft dieser weitreichende Informationsaustausch dabei, Wissenslücken und die darauf basierenden Gefahren einzugrenzen. Im Idealfall ermöglicht der eingeforderte Informationsfluss, allen beteiligten Akteuren auf neue stoffbezogene Erkenntnisse schnellstmöglich und effektiv zu reagieren.

Ein sehr ähnlicher Regelungsansatz liegt dem Pharmakovigilanz-System (§ 63b AMG) des Arzneimittelrechts zugrunde, das Zulassungsinhaber einrichten und betreiben müssen (§ 63 I AMG). Sie sind verpflichtet, insbes. nach der Zulassung Erfahrungen bei der Anwendung eines Medikaments fortlaufend und systematisch zu sammeln und auszuwerten (§ 63b II AMG). Die Norm zielt darauf, dessen unerwünschte Wirkungen kontinuierlich zu beurteilen und zu überwachen, um zeitnah geeignete Risikominderungsmaßnahmen ergreifen zu können. Denn zu dem Zeitpunkt, zu dem Arzneimittel auf den Markt gelangen, sind ihre Wirkungen an einer kleinen Personenzahl getestet und deshalb in ihren Dimensionen noch nicht vollständig voraussehbar.

Für Arzneimittel sowie chemische Stoffe sind unterschiedliche Einsatzszenarien und -modalitäten sowie disparat verteilte Informationsquellen und ex ante nicht in jeder Hinsicht voraussehbare Wirkungen ebenso wie bei Foundation Models gleichermaßen charakteristisch. In allen drei Fällen sollten die verantwortlichen Akteure diejenigen Informationen erhalten, die sie benötigen, um Risiken zu minimieren. Der Informationsaustausch sollte sowohl „stromaufwärts“ als auch „stromabwärts“ zwischen vor- und nachgelagerten Akteuren reibungsarm funktionieren – und die zuständigen Behörden einschließen.

Insoweit besteht bei den Entwürfen der KI-VO noch Nachbesserungsbedarf:[141] Dem Regulierungsansatz der REACH-VO und der Pharmakovigilanz des AMG folgend sollte die KI-VO auch für Foundation Models ein umfangreiches Informationssystem etablieren. Die derzeit vorgesehene technische Dokumentation (Art. 11 KI-VO-E) und die Aufzeichnungspflichten (Art. 12 KI-VO-E) erweisen sich insoweit als unzureichende Grundlage. Denn sie sind nicht darauf ausgerichtet, den Informationsfluss zwischen den Beteiligten in der KI-Wertschöpfungskette zu verbürgen. Zwar machen die Anbieter den Nutzern entsprechende technische Dokumentationen zugänglich (vgl. Erwgr. 47 KI-VO-E (KOM)). Diese Freigabe kann indes umfassende Informations- und Mitteilungspflichten für Fälle, in denen ein Akteur bspw. neue Risiken (auf Ebene der Foundation Models oder auf nachgelagerter Ebene) entdeckt, nicht ersetzen. Gleiches gilt für die Beobachtungspflichten iSd Art. 61 ff. KI-VO-E (KOM) und das Risikomanagementsystem (Art. 9 KI-VO-E). Um die Gefahren von „Schlupflöchern“ für Risiken oder missbräuchliche Verwendungen zu minimieren, bedarf es eines über diese Verpflichtungen hinausreichenden aktiven Austauschs zwischen allen Beteiligten der KI-Wertschöpfungskette, die auf einem Foundation Model aufbaut.

Gerade für Anbieter von Foundation Models ist es nur mit den notwendigen Informationen möglich, auf Risiken ihrer Systeme, die bspw. in nachgelagerten Anwendungen auftreten, angemessen zu reagieren und diese in ihre eigenen Risikoanalysen einfließen zu lassen. Sie müssen daher nicht nur ihre nachgelagerten Anbieter gleichsam in einer „Einbahnstraße“ mit Informationen versorgen. Ebenso wichtig ist es, dass auch sie stets die neuesten Erkenntnisse in ihre Arbeit an den Modellen einfließen lassen können. Eine Zusammenarbeit personell und finanziell gut aufgestellter sowie umfassend informierter Behörden und sämtlicher Akteure der Lie-

137 ABlEU 2006 L 396, 1-851.

138 Eifert/Hoffmann-Riem, Innovationsfördernde Regulierung/Bizer/Führ, Innovationen entlang der Wertschöpfungskette: Impulse aus der REACH-Verordnung, 2009, S. 293; Schulze-Rickmann, Das Recht auf Zugang zu Informationen und auf ihre Verwertung nach der europäischen REACH-Verordnung, 2010, S. 33; Seulen, Strategien zur Substitution umweltgefährdender Stoffe im europäischen und deutschen Gefahrstoffrecht, 2015, S. 413.

139 Hierzu teilt es der Europäischen Chemikalienagentur (ECHA) u.a. Leitlinien für die sichere Verwendung eines Stoffes (Art. 10 Buchst. a UBuchst. v REACH-VO) sowie die Ergebnisse umfassender Untersuchungen zu gefährlichen Eigenschaften des betroffenen Stoffes (insbes. Art. 10 Buchst. a UBuchst. vi REACH-VO) mit.

140 Art. 3 Nr. 2 VO (EG) 1907/2006.

141 Vgl. IV. 2. d).

ferkette sowie – soweit möglich – Nutzern eröffnete allen Beteiligten die Chance, die bekannten Risiken der Modelle im Blick zu behalten, und neue Risiken schnellstmöglich aufzudecken.

V. Schlussfolgerungen und Empfehlungen

So sehr Basismodelle eine Zeitenwende der KI einläuten, so sehr stellen sie die Gesellschaft vor Herausforderungen. Einerseits gilt es, die Risiken einzugrenzen, welche die Modelle in vielfältigen Kontexten – v.a. in Hinblick auf die Grundrechte – auslösen (können). Andererseits sollte der Gesetzgeber die vielfältigen Innovationsmöglichkeiten und damit einhergehenden wirtschaftlichen Wertschöpfungsquellen, die Foundation Models eröffnen, nicht ausbremsen. Ein (zu) strenger oder mit heißer Nadel gestrickter Regelungsrahmen könnte insbes. in der EU ansässige Unternehmen im globalen Wettbewerb auf dem KI-Markt benachteiligen und so die Wettbewerbsfähigkeit des Binnenmarktes auf internationalem Parkett signifikant schwächen. Die Union tut gut daran, faktische Monopolstellungen außereuropäischer Tech-Unternehmen, welche die europäische Wirtschaft in eine Abhängigkeitsposition drängen, nicht zu stärken und zugleich die Entwicklung der innovativen Modelle im Interesse des gesamten gesellschaftlichen und wirtschaftlichen Lebens in Bahnen zu lenken, die mit den europäischen Werten und Grundrechten in Einklang stehen.

Mit der Einigung über die finale Fassung der KI-VO möchte die Union die Chance ergreifen, sich global als einer der Vorreiter auf dem Feld der KI-Regulierung zu positionieren. Dass der Rat und das Parlament in ihren Entwürfen Konzepte einer Regulierung von Basismodellen entworfen haben, ist prinzipiell zu begrüßen. Allerdings ist das normative Grundgerüst (soweit es bereits öffentlich bekannt ist) bislang nicht vollständig tragfähig. Es hätte insbes. davon profitiert, Anleihen bei Blaupausen anderer Regulierungskontexte mit strukturell ähnlichen Herausforderungen zu nehmen, wie zB der Dual-Use-VO sowie der REACH-VO und den arzneimittelrechtlichen Regelungen zur Pharmakovigilanz. Ungenügend und optimierungsfähig sind in diesem Lichte etwa die angestrebten Vorgaben für das Informations- und Datenmanagement. Zum Schutz der Grundrechte sowie europäischer KMU muss die KI-VO insbes. der hochgradigen Komplexität der KI-Wertschöpfungsketten stärkere Beachtung schenken. Es gilt v.a., die Informationskette zu stärken, damit die Anbieter der Foundation Models neue Erkenntnisse und / oder nachträglich aufgetretene Risiken bewerten können. Auch nachgelagerte Anbieter bzw. Betreiber können nur auf der Grundlage entsprechender Informationen gezielt auswählen, ob und inwieweit sie ein Modell nutzen wollen bzw. abschätzen, welche Gefahren, zB Diskriminierungsrisiken, von ihm ausgehen. Zudem können sie nur dann ihren eigenen Verpflichtungen in ausreichendem Maße nachkommen. Die KI-VO sollte den Anbietern auch sanktionsbewehrt auferlegen, auf die neuen Erkenntnisse zu reagieren, und zB – sofern nötig – ihre eigenen Risikomanagementsysteme anzupassen und fortzuentwickeln. Entsprechende Bestimmungen – in diesem Fall ebenso für die nachgelagerten Anwender – fehlen in den derzeitigen Entwürfen jedoch in der notwendigen Tiefe.

Erhöhte Anforderungen an die Dokumentation und Nutzung von (Trainings-)Datensätzen sind gerade auch im Rahmen hoheitlicher Einsatzszenarien unerlässlich: Staatliche Verantwortungsträger müssen die Foundation Models und ihre Trainingsgrundlagen – zumindest soweit überhaupt möglich – verstehen und prüfen können.

Dass unionale KI-Unternehmen das Zeug haben, sich im globalen Wettbewerb erfolgreich zu behaupten, illustrieren die Unternehmen *Aleph Alpha, Mistral AI* bzw. Bloom (ein auf einem französischen Supercomputer trainiertes Sprachmodell) eindrucksvoll. Um deren Innovationsrahmen zu stärken, sollte der Gesetzgeber die Regelungen zu KI-Reallaboren (Art. 53 ff. KI-VO-E) um konkrete Vorgaben zur Erforschung von Foundation Models ergänzen. Auch weitere Ausnahmetatbestände für KMU mit partiellen Befreiungen von Anforderungen sind denkbar. Auf diese Weise sollte die EU gezielt die potenziellen Markteintrittshürden senken, die von strengen Anforderungen an Datensätze, Training, Transparenz uvam für die europäische KI-Wirtschaft ausgehen. Zusätzlich empfiehlt sich ein Strauß an Fördermaßnahmen – auch finanzielle Unterstützungen sowie der Zugriff auf Hochleistungsrechnerressourcen –, die es insbes. KMU oder gemeinwohlorientierten Zusammenschlüssen im Binnenmarkt ermöglichen, von den gegenwärtigen technologischen Sprüngen zu profitieren und diese einflussreich mitzugestalten.

Allein auf die Selbstregulierungskräfte des Marktes zu setzen, wäre demgegenüber kein adäquater Weg. General-Purpose-AI ist zu sensitiv, um sie nicht zu regulieren – sie ist aber auch zu innovativ, um sie regulatorisch zu strangulieren.[142]

142 In Anlehnung an Sundar Pichai in einem FAZ-Interview v. 31.5.2023.