Deutscher Bundestag
Ausschuss für Digitales

Ausschussdrucksache
zu 20(23)151

23.05.2023

Stellungnahme

Prof. Dr. Philipp Hacker, LL.M. (Yale)*

22. Mai, 2023

für die:

Öffentliche Anhörung „Generative Künstliche Intelligenz“ am Mittwoch, 24. Mai 2023, 14:30 – 16:30 Uhr, Sitzungssaal Reichstagsgebäude (RTG) 3 N 001

Inhalt:

* Chair for Law and Ethics of the Digital Society, European New School of Digital Studies, European University Viadrina.

I. Allgemeine Überlegungen

Die Regulierung der KI steht am Scheideweg, sowohl in der EU als auch darüber hinaus.

Wir erleben derzeit in Echtzeit die Geburt einer neuen Generation von KI-Systemen, insbesondere im Bereich der generativen KI. Diese Modelle bieten enorme Chancen und werden die Art und Weise, wie wir arbeiten, kommunizieren und leben, erheblich verändern - ja, sie verändern sie bereits. Gleichzeitig birgt diese neue Generation von KI-Systemen spezifische Risiken, auf die die Regulierung eingehen muss. Meiner Meinung nach sind die folgenden sechs Themen kurz- und mittelfristig am dringlichsten: Datenschutz, Nichtdiskriminierung, Qualität (von Daten und Output), Inhaltsmoderation, ökologische Nachhaltigkeit und zivilrechtliche Haftung. Längerfristig müssen wir uns auch auf eine mögliche Umstrukturierung des Arbeitsmarktes mit entsprechenden Auswirkungen auf das Steueraufkommen sowie auf die Nutzung von KI durch böswillige Akteure einstellen.

Wie ich bereits in früheren Veröffentlichungen dargelegt habe,[1] muss der Rechtsrahmen für KI und insbesondere für generative KI ein empfindliches Gleichgewicht zwischen der angemessenen Bewältigung dieser Risiken und der Möglichkeit für KI-Entwickler, Modelle zu erstellen und sie in der Praxis auf gesellschaftlich nützliche Weise einzusetzen, herstellen. Wichtig ist, dass die Regulierung ein Ökosystem fördert, das eine weitere Marktkonzentration in den Händen einiger weniger Unternehmen mit Sitz außerhalb der EU verhindert; vielmehr müssen große und kleine Unternehmen sowohl außerhalb als auch innerhalb der EU in der Lage sein, KI-Systeme innerhalb der gesetzlichen Leitplanken rasch zu entwickeln und einzusetzen. Daher ist es von größter Bedeutung, die Regulierung so zu gestalten, dass sie sowohl wirksam als auch umsetzbar ist, d. h. von allen Akteuren in der KI-Wertschöpfungskette möglichst reibungslos umgesetzt und befolgt werden kann.

[1] Siehe z.B., Philipp Hacker, Andreas Engel und Marco Mauer, 'Regulating ChatGPT and other Large Generative AI Models' (2023) ACM Conference on Fairness, Accountability, and Transparency (FAccT '23, in Vorbereitung) https://arxiv.org/abs/2302.02337; Philipp Hacker, 'The European AI Liability Directives - Critique of a Half-Hearted Approach and Lessons for the Future' (2022) Working Paper, https://arxivorg/abs/221113960 ; siehe auch Philipp Hacker, Andreas Engel und Amelie Berz, The EU AI Act is improving - but still contains fundamental flaws, TechMonitor (Mai 19, 2023), https://techmonitor.ai/comment-2/eu-ai-act-improving.

II. Antworten auf die vom Digitalausschuss vorbereiteten spezifischen Fragen

1. Die Regulierung der generativen KI ist derzeit Gegenstand der Verhandlungen zum Europäischen KI-Gesetz (AIA). Wie kann generative KI Ihrer Meinung nach effektiv in den AIA einbezogen und geregelt werden und wie beurteilen Sie die vorgeschlagenen Differenzierungen innerhalb der generativen KI zwischen "General Purpose AI" und "Foundation Models"?

Am 11. Mai hat das Europäische Parlament mit zwei entscheidenden Abstimmungen im Ausschuss den Weg für seinen Standpunkt zum KI-Gesetz freigemacht. Der Entwurf geht wohl in die richtige Richtung, aber es gibt noch erhebliche Mängel, die die Entwicklung und den Einsatz von KI in der EU zum Scheitern bringen könnten.

a) Terminologie

Was die Terminologie anbelangt, so enthält die EP-Fassung des KI-Gesetzes eine neue Bestimmung, die einen spezifischen Rahmen für so genannte "Basismodelle" vorgibt. Dieser Begriff umfasst hoch entwickelte KI-Modelle, einschließlich verschiedener generativer KI-Systeme wie ChatGPT, GPT-4, Bard oder Stable Diffusion. Das Konzept eines "Basismodells" hat in der Informatikgemeinschaft große Anerkennung gefunden,[2] unterstreicht in angemessener Weise das breite Spektrum von Aufgaben und Ergebnissen, die sie bewältigen können.[3] So würde beispielsweise ein linearer Klassifikator, der lediglich in der Lage ist, zwischen Menschen und Hunden in Bildern zu unterscheiden, die Kriterien nicht erfüllen, während ein Textgenerator wie GPT-4 oder Luminous, der in der Lage ist, Texte zusammenzufassen, zu vervollständigen und eigenständig zu generieren, als Basismodell gelten würde. Meines Erachtens ist es sinnvoll, eine Terminologie zu verwenden, die auch in der Informatik weit verbreitet ist, und sich speziell auf Basismodelle zu beziehen.

Im Gegensatz dazu ist der Begriff "Allzweck-KI-System" von Natur aus vage, wird mit unterschiedlichen Konnotationen verwendet - wenn überhaupt - in der technischen Gemeinschaft,[4] und sollte aufgegeben werden.

b) Regulierungsarchitektur für Stiftungsmodelle: drei Ebenen

Ich beziehe mich auf unser Papier, in dem wir den Begriff "großes generatives KI-Modell" (LGAIM) verwenden, um uns auf die Gründungsmodelle zu beziehen.

"Eine Regulierung von [großen generativen KI-Modellen] LGAIMs ist notwendig, muss aber besser auf die konkreten Risiken zugeschnitten sein, die sie mit sich bringen. Daher schlagen wir eine Verlagerung weg von der umfassenden KI-Gesetzesregulierung, wie sie im allgemeinen Ansatz des Rates der EU vorgesehen ist, hin zu spezifischen Regulierungspflichten und inhaltlicher Moderation vor. Wichtig ist, dass die Einhaltung der Vorschriften für große

[2] Siehe z.B., Rishi Bommasani und andere, "Über die Chancen und Risiken von Gründungsmodellen" (2021) arXiv preprint arXiv:210807258 .
[3] Hacker, Engel und Mauer, 'Regulierung von ChatGPT und anderen großen generativen KI-Modellen'.
[4] Carlos Ignacio Gutierrez und andere, 'A Proposal for a Definition of General Purpose Artificial Intelligence Systems' (2022) Working Paper, https://ssrncom/abstract=4238951 ; Hacker, Engel und Mauer, 'Regulierung von ChatGPT und anderen großen generativen KI-Modellen'.

und kleine LGAIM-Entwickler möglich sein muss, um ein "Winner-takes-all"-Szenario und eine weitere Marktkonzentration zu vermeiden [82]. Dies ist nicht nur für die Innovation und das Wohlergehen der Verbraucher entscheidend [33, 159, 160], sondern auch für die ökologische Nachhaltigkeit. Während der Kohlenstoff-Fußabdruck von IT und KI erheblich ist und stetig zunimmt [54-58] und das Training von LGAIMs besonders ressourcenintensiv ist [161], könnten große Modelle letztlich weniger Treibhausgasemissionen verursachen als ihre kleineren Brüder, wenn sie an mehrere Verwendungszwecke angepasst werden können.

Vor diesem Hintergrund sehen wir **drei Ebenen** von Verpflichtungen für LGAIMs vor: eine erste Gruppe von **Mindeststandards** für alle LGAIMs; eine zweite Gruppe von **spezifischen Hochrisiko-Regeln**, die nur für LGAIMs gelten, die in **konkreten Hochrisiko-Anwendungsfällen** eingesetzt werden; und die dritte Gruppe von Regeln für die **Zusammenarbeit entlang der KI-Wertschöpfungskette** (siehe Abschnitt 3.2.2), um die effektive Einhaltung der ersten beiden Gruppen von Regeln zu ermöglichen."[5]

i. Der Vorschlag des Europäischen Parlaments vom 9. Mai

(1) Erste Ebene: Mindeststandards für alle Stiftungsmodelle

"Die erste Ebene wird für die Anbieter (=Entwickler) einer Teilmenge von GPAIS gelten, die als "Basismodelle" (Artikel 28b(1)-(3) KI-Gesetz EP-Version) und generative KI (Artikel 28b(4) KI-Gesetz EP-Version) bezeichnet werden. In Anlehnung an einen bekannten Begriff aus der Computerwissenschaftdefiniert die EP-Fassung Basismodelle als ein KI-System, "das auf einer breiten Datenbasis in großem Umfang trainiert wird, auf die Allgemeinheit der Ergebnisse ausgelegt ist und an ein breites Spektrum spezieller Aufgaben angepasst werden kann" (Artikel 3 Absatz 1c KI-Gesetz EP-Fassung) . Der Fokus auf die Allgemeinheit des Outputs und der Aufgaben ist in der Tat besser geeignet, die Besonderheiten großer generativer KI-Modelle zu erfassen, als die vage Definition von GPAIS. Im Einklang mit den Vorschlägen dieses Papiers umfassen die allgemeinen Verpflichtungen für alle Basismodelle Maßnahmen zur Datenverwaltung, insbesondere im Hinblick auf die Verringerung von Verzerrungen (Artikel 28b Absatz 2 Buchstabe b KI-Gesetz EP-Fassung). Darüber hinaus muss während des gesamten Lebenszyklus des Modells ein angemessenes Niveau an Leistung, Interpretierbarkeit, Korrigierbarkeit, Sicherheit und Cybersicherheit aufrechterhalten werden. Diese Anforderungen müssen von unabhängigen Sachverständigen getestet, dokumentiert und überprüft werden (Artikel 28b Absatz 2 Buchstabe c KI-Gesetz EP-Fassung). Entscheidend ist jedoch, dass alle Stiftungsmodelle auch Risikobewertungen, Risikominderungsmaßnahmen und Risikomanagementstrategien im Hinblick auf vernünftigerweise vorhersehbare Risiken für die Gesundheit, die Sicherheit, die Grundrechte, die Umwelt, die Demokratie und die Rechtsstaatlichkeit umsetzen müssen, ebenfalls unter Einbeziehung unabhängiger Sachverständiger, Artikel 28b(2)(a) AI Act EP Version. Diese Forderung kommt faktisch einer Einstufung von Stiftungsmodellen als risikoreich per se gleich.

Ein wesentliches Element der Mindeststandards für generative KI ist in der "ChatGPT-Regel" enthalten, Art. 28b(4) KI-Gesetz EP Version. Sie enthält drei Hauptelemente. (i) Die Transparenzpflicht bezüglich des Einsatzes von KI (Art. 28b(4) KI-Gesetz EP-Fassung, Art. 52(1) AI Act) ist ein Schritt in die richtige Richtung. Sie adressiert Pflichten der Anbieter

[5] Hacker, Engel und Mauer, 'Regulierung von ChatGPT und anderen großen generativen KI-Modellen', 17.

gegenüber den Nutzern von KI-Systemen. Aus unserer Sicht sind darüber hinaus in einigen Fällen auch Pflichten der Nutzer gegenüber den Empfängern gerechtfertigt, um die Verbreitung von Fake News und Fehlinformationen zu bekämpfen. (ii) Die Vorschrift zur Verhinderung von Verstößen gegen das EU-Recht könnte ebenfalls detaillierter gefasst werden. Hier sollten die Compliance-Mechanismen des DSA viel konkreter übertragen werden, zum Beispiel durch klare, verbindliche Melde- und Aktionsverfahren und vertrauenswürdige Flagger. Es versteht sich von selbst, dass die Modelle dem geltenden Recht entsprechen müssen. (iii) Die Offenlegung von urheberrechtlich geschütztem Material, das in Trainingsdaten enthalten ist, kann in der Tat den Autoren und Urhebern helfen, ihre Rechte durchzusetzen. Allerdings streiten selbst Experten oft darüber, ob bestimmte Werke überhaupt urheberrechtsfähig sind oder nicht. Es muss vermieden werden, dass Entwickler, die z. B. 20 Millionen Bilder verarbeitet haben, nun eine umfassende rechtliche Prüfung dieser 20 Millionen Bilder durchführen müssen, um selbst zu entscheiden, ob sie urheberrechtsfähig sind oder nicht. Es muss daher ausreichen, auch übergreifend die Werke offen zu legen, die urheberrechtsfähig sein können, auch solche, bei denen nicht klar ist, ob sie letztlich urheberrechtsfähig sind oder nicht. Andernfalls entstehen wiederum praktisch unerschwingliche Due-Diligence-Kosten. Der einzelne Urheber muss dann, wenn er sein Werk entdeckt, entscheiden, ob er es für urheberrechtlich geschützt hält oder nicht. [...]

Unserer Ansicht nach enthält dieser Vorschlag zwar Schritte in die richtige Richtung, wäre aber letztlich nicht überzeugend, da er Stiftungsmodelle effektiv als Hochrisikoanwendungen behandelt. Wie bereits erwähnt und weiter unten ausführlich erörtert, kann der Output von KI für schädliche Äußerungen und Handlungen missbraucht werden (wie fast jede Technologie). Aber dies scheint nicht nur eher die Ausnahme als die Regel zu sein. Das Argument der nachteiligen Folgen für den Wettbewerb gilt auch hier. In der EP-Fassung konzentrieren sich Risikobewertung, -minderung und -management nach wie vor auf das Modell selbst und nicht auf die anwendungsspezifische Anwendung (Art. 28b(2)(a) und (f) AI Act EP Version), auch wenn in Erwägungsgrund 58a anerkannt wird, dass Risiken im Zusammenhang mit KI-Systemen aus ihrer spezifischen Anwendung resultieren können. Auch dies führt zu einer aufwändigen Bewertung und Abmilderung hypothetischer Risiken, die möglicherweise nie eintreten werden, anstatt Risiken auf der Anwendungsebene zu managen, wo der konkrete Einsatz berücksichtigt werden kann."[6]

(2) Zweite Ebene: spezifische Anwendungen mit hohem Risiko

"Die zweite Ebene bezieht sich auf "neue Anbieter", die das AI-System erheblich verändern, Art. 28(1)(b) und (ba) AI Act EP Version. Dieser neue Anbieter, der in unserem Papier als "deployer" bezeichnet wird, übernimmt bei einer wesentlichen Änderung die Pflichten des früheren Anbieters; der neue Anbieter übernimmt diese Rolle (Art. 28(1) und (2)(1) KI-Gesetz EP-Fassung).

Die neue Vorschrift über eine Grundrechtsfolgenabschätzung (Art. 29a AI-Gesetz EP-Fassung) gilt auch auf dieser Ebene der konkreten Anwendung. Auch diese Regelung scheint kaum operationalisierbar. Die Grundrechte sind eine unscharfe Kategorie und auf technischer Ebene schwer umzusetzen, wo spezifische sekundäre Regelungen sinnvoller sein könnten (DSGVO, Richtlinien zum Antidiskriminierungsrecht usw.). Vor allem aber ist sie auch doktrinär fehlgeleitet, da private Unternehmen im Allgemeinen nicht die Grundrechte anderer

[6] Ibid, 7 ff..

Privatpersonen verletzen können (Grundrechte binden den Staat, nicht die Bürger; es gibt nur einige Ausnahmen von dieser Regel[7])."[8] Darüber hinaus ist das Verhältnis der Grundrechtsfolgenabschätzung zur allgemeinen Risikobewertung (Artikel 9 und 28b Absatz 1 Buchstabe a) unklar und kann zu unnötiger Doppelarbeit führen.

(3) Die dritte Ebene: Zusammenarbeit entlang der KI-Wertschöpfungskette

"Eine dritte Anforderungsebene bezieht sich auf die KI-Wertschöpfungskette (Art. 28 Abs. 2 Ziff. 2 KI-Gesetz EP-Fassung), in Übereinstimmung mit den unten in diesem Papier gemachten Vorschlägen."[9]

ii. Eigener Vorschlag

Meiner Meinung nach brauchen wir in der ersten Schicht bestimmte Regeln, die für alle Stiftungsmodelle gelten. Nicht Risikobewertung und -management, sondern andere Mechanismen (siehe unten). In der zweiten Schicht sollten Risikobewertung und -management an bestimmte Anwendungsfälle gebunden sein. Die dritte Ebene muss Zugangs- und Informationsrechte vorsehen, damit die Akteure der KI-Wertschöpfungskette die für die Einhaltung des KI-Gesetzes erforderlichen Informationen sammeln können.[10]

"Was die Mindeststandards betrifft, so gilt der EU-Besitzstand in erster Linie auch für die Entwickler von LGAIMs, wobei die Datenschutzgrundverordnung (siehe Abschnitt 5), das Antidiskriminierungsgesetz (Abschnitt 4),[11] sowie die Produkthaftung [24] im Mittelpunkt stehen. Darüber hinaus müssen Transparenzvorschriften gelten, die nun auch vom Europäischen Parlament [70] vorgeschlagen wurden (siehe unten, Abschnitt 7.1). Darüber hinaus müssen spezifische Risiken, die von so herausragender Relevanz sind, dass sie auf der vorgelagerten Ebene adressiert werden sollten, anstatt sie in bestimmten Anwendungsfällen an die Hersteller zu delegieren, den Entwicklern als Teil der Mindeststandards zugewiesen werden. Dies betrifft unseres Erachtens ausgewählte Data-Governance-Pflichten (Art. 10 KI-Gesetz, siehe Abschnitt 4) und Regeln zum immer wichtiger werdenden Thema Cybersecurity (Art. 15 KI-Gesetz). Schliesslich sollten auch Nachhaltigkeitsregeln [24] sowie die Inhaltsmoderation (siehe unten, Abschnitt 7.4) zu den Mindeststandards gehören, die für alle LGAIMs gelten."[12]

(1) Risikobewertung und -management: Anwendungsfallbezogen

"Wichtig ist, dass der volle Umfang des Abschnitts über hohe Risiken des AI-Gesetzes, einschließlich des formellen Risikomanagements, nur dann gelten sollte, wenn eine bestimmte LGAIM (oder GPAIS) tatsächlich für Zwecke mit hohem Risiko verwendet wird. Diese Strategie steht im Einklang mit einem allgemeinen Grundsatz des Produktsicherheitsrechts :

---

[7] Siehe z. B. EuGH, Rechtssache C-414/16 (Egenberger) und folgende Urteile in dieser Richtung.
[8] Hacker, Engel und Mauer, "Regulierung von ChatGPT und anderen großen generativen KI-Modellen", 7 ff..
[9] Ebd., 7 ff.
[10] Zu diesem Punkt im Einzelnen siehe ebd., 10 ff.
[11] Siehe z. B., Sandra Wachter, "Die Theorie der künstlichen Unveränderlichkeit: Protecting Algorithmic Groups Under Anti-Discrimination Law' (2022) arXiv preprint arXiv:220501166 ; Frederik J Zuiderveen Borgesius, 'Strengthening legal protection against discrimination by algorithms and artificial intelligence' (2020) 24 The International Journal of Human Rights 1572; Philipp Hacker, 'Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic discrimination under EU law' (2018) 55 Common Market Law Review 1143.
[12] Hacker, Engel und Mauer, 'Regulierung von ChatGPT und anderen großen generativen KI-Modellen', 17.

Nicht jede Schraube und jeder Bolzen muss nach den höchsten Standards hergestellt werden. Nur wenn sie zum Beispiel für Raumschiffe verwendet werden, gelten die strengen Produktsicherheitsvorschriften für die Herstellung von Luftfahrtmaterial[13] - nicht aber, wenn sie im örtlichen Baumarkt für den allgemeinen Gebrauch verkauft werden. Dasselbe Prinzip sollte auch für LGAIMs gelten".[14]

(2) Moderation von Inhalten

"Eine der größten Herausforderungen für LGAIMs ist wohl ihr potenzieller Missbrauch für Desinformation, Manipulation und schädliche Äußerungen. Unseres Erachtens müssen die für traditionelle soziale Netzwerke konzipierten DSA-Regeln entsprechend erweitert und angepasst werden. Das Europäische Parlament hat sich dieser Herausforderung teilweise angenommen, indem es festlegt, dass Gründungsmodelle nicht gegen EU-Recht verstoßen dürfen [76]. Unserer Ansicht nach sollte die Regulierung jedoch einen Schritt weiter gehen und die DSA-Regeln selektiv auf LGAIM-Entwickler und -Einrichter ausweiten. LGAIMs und die Gesellschaft würden von verpflichtenden Melde- und Aktionsmechanismen, vertrauenswürdigen Flaggern und umfassenden Audits für Modelle mit besonders vielen Nutzern profitieren. Die Regelungslücke ist besonders virulent für LGAIMs, die als eigenständige Software angeboten werden, wie es derzeit der Fall ist. Für die Zukunft ist mit einer zunehmenden Integration in Plattformen verschiedener Art, wie Suchmaschinen oder soziale Netzwerke, zu rechnen, wie die Entwicklung oder Übernahme von LGAIM durch Microsoft, Meta oder Google zeigt. Obwohl der DSA dann technisch anwendbar wäre, müsste er dennoch aktualisiert werden, um sicherzustellen, dass LGAIM-generierte Inhalte genauso abgedeckt sind wie nutzergenerierte Inhalte. Da der LGAIM-Output derzeit besonders anfällig für die Verbreitung von Fehlinformationen ist, erscheint es ratsam, LGAIM-generierte Inhalte als solche zu kennzeichnen, sofern dies technisch möglich ist. Doktrinell könnte dies durch eine Änderung des DSA oder des Art. 29 AI-Gesetzes erreicht werden, das bereits in seinem Abs. 4 Meldepflichten enthält (siehe Teil 4). 4 enthält (siehe Teil 4). Angesichts des derzeitigen politischen Prozesses in der EU scheint die letztere Option realistischer zu sein.[15]

(3) Nachhaltigkeit

"Das KI-Gesetz sollte eine "Nachhaltigkeitsfolgenabschätzung" für KI-Systeme vorschreiben. Eine solche Prüfung könnte sich an den vielfältigen Vorschlägen zu Folgenabschätzungen für KI-Systeme im Allgemeinen orientieren.[16] Zu diesem Zweck sollte eine Bestimmung in das KI-Gesetz aufgenommen werden, die sich strukturell an Artikel 9 KI-Gesetz (Risikomanagementsystem) orientiert. Sie würde für Entwickler von KI-Systemen mit hohem und ohne hohem Risiko gleichermaßen gelten, da der Kohlenstoff-Fußabdruck von KI-Systemen nicht mit dem Grad ihres Risikos für Gesundheit, Sicherheit oder Grundrechte zusammenhängt. Entscheidend ist, dass die Entwickler während der Modellierungsphase verschiedene Modelltypen (z. B. lineare Regression versus Deep Learning) nicht nur hinsichtlich ihrer Leistung, sondern auch ihres geschätzten Klima-Fußabdrucks vergleichen

[13] Siehe z.B. Produktnormen, Reihe Luft- und Raumfahrt, DIN EN 4845-4851 (Dezember 2022) zu Schrauben.
[14] Hacker, Engel und Mauer, 'Regulierung von ChatGPT und anderen großen generativen KI-Modellen', 19.
[15] Ibid, 19 ff.
[16] Andrew D. Selbst, "An Institutional View of Algorithmic Impact Assessments" (2021) 35 Harvard Journal of Law & Technology; Kaminski und Malgieri, Multi-layered explanations from algorithmic impact assessments in the GDPR; Margot E. Kaminski und Gianclaudio Malgieri, "Algorithmic impact assessments under the GDPR: producing multi-layered explanations" (2020) International Data Privacy Law 19.

sollten.[17] Es gibt bereits Tools zur Messung der Kohlenstoffauswirkungen von Modellen.[18] Einfach ausgedrückt: Wenn zwei Modelltypen eine ähnliche Leistung aufweisen, wären die Entwickler nach der neuen Bestimmung verpflichtet, das nachhaltigere Modell für die weitere Entwicklung und den Einsatz zu wählen. Auf diese Weise kann die derzeitige Fixierung auf Leistungsmessungen durch ein noch größeres Umweltbewusstsein und konkrete, wartungsarme Schritte zur Einbeziehung der Nachhaltigkeit in die breitere Zielfunktion der ML-Entwicklung ergänzt werden.

In vielen Szenarien können Nachhaltigkeit und Leistung sogar Hand in Hand gehen. Ein aktueller Trend beim maschinellen Lernen ist die Verwendung von vortrainierten Modellen, die auf einigen allgemeineren Daten für eine bestimmte Aufgabenklasse trainiert wurden (z. B. Bild[19] oder Spracherkennung[20]).[21] Sie werden dann von Entwicklern, die an einem konkreten Problem mit domänenspezifischen Daten arbeiten, vollständig trainiert. Solche vortrainierten Modelle sind nicht nur oft leistungsfähiger und haben sich bei zahlreichen Aufgaben als State-of-the-Art-Architektur durchgesetzt,[22] , sondern sie verbrauchen auch insgesamt weniger Energie, da das Vortraining für viele verschiedene Modelleinsätze nur einmal durchgeführt werden muss.[23] Ironischerweise könnte die Regulierung diese Bemühungen jedoch zunichte machen. Die leistungsfähigsten vortrainierten Modelle sind [Basismodelle] - genau diejenigen, deren Entwicklung durch die aktuelle Version des KI-Gesetzes und die KI-Haftungsrichtlinien erheblich erschwert wird. Hier schließt sich der Kreis, allerdings nicht auf effiziente oder nachhaltige Weise. Dies zeigt erneut, wie wichtig es ist, die Regeln für"[24] Gründungsmodelle zu ändern.

2. Generative KI bietet zahlreiche Anwendungsmöglichkeiten in unterschiedlichsten Berufen und kann den Arbeitsmarkt entlasten. Wie schätzen Sie die Potenziale und Risiken der generativen KI für die Arbeitswelt ein und wo sehen Sie Regulierungsbedarf?

Zu Beginn möchte ich anmerken, dass ich Spezialist für Arbeitsökonomie bin.

Ich bin jedoch der festen Überzeugung, dass generative KI den Arbeitsmarkt erheblich verändern wird, und zwar sowohl zum Vorteil als auch zum Nachteil, je nachdem, welche

[17] Die genauen Auswirkungen sind nicht leicht zu messen. Ein Index, der die Scope-1-, -2- und -3-Emissionen für die notwendigen Rechenressourcen (z. B. Energie, Kohlendioxidemissionen) für Schulungen und Umschulungen umfasst, könnte als Näherungswert verwendet werden. Für eine umfassendere Messung der Auswirkungen (einschließlich Produktion, Transport und End-of-Life sowie Wasserverbrauch) siehe OECD, Measuring the Environmental Impacts of AI Compute and Applications: The AI Footprint, Anhang A; zu Scope 1, 2 und 3 Emissionen, siehe IPCC, Working Group III Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (2014), 122.

[18] Überblick in OECD, Measuring the Environmental Impacts of AI Compute and Applications: The AI Footprint, 28.

[19] Siehe z. B. Gustavo Carneiro, Jacinto Nascimento und Andrew P Bradley, Unregistered multiview mammogram analysis with pre-trained deep learning models (Springer 2015).

[20] Juliette Millet und andere, 'Toward a realistic model of speech processing in the brain with self-supervised learning' (2022) NeurIPS https://arxiv.org/abs/2206.01685.

[21] Xu Han und andere, 'Pre-trained models: Vergangenheit, Gegenwart und Zukunft" (2021) 2 AI Open 225.

[22] Ebd.

[23] David Patterson und andere, 'Carbon emissions and large neural network training' (2021) arXiv preprint arXiv:210410350, 15.

[24] Hacker, The European AI Liability Directives - Critique of a Half-Hearted Approach and Lessons for the Future", 63 ff.

Arbeitsplätze betroffen sind. Diese Veränderungen lassen sich nicht vollständig verhindern, sondern nur in bestimmte Richtungen kanalisieren.

Viele Aufgaben im Wissens- und Kreativbereich werden durch generative KI-Anwendungen unterstützt, die wahrscheinlich die Produktivität steigern werden. Wichtig ist, dass generative KI in vielen Sektoren auch die wachsende Lücke schließen kann, die durch den allgemeinen Mangel an qualifizierten Arbeitskräften entsteht.

Andererseits ist es kein Geheimnis, dass generative KI auch einige Arbeitsplätze ersetzen wird. Zum jetzigen Zeitpunkt erscheint es verfrüht, mit Sicherheit abzuschätzen, ob alle diese Arbeitsplätze mittelfristig durch andere Arbeitsplätze ersetzt werden, die durch den Aufstieg der generativen KI ermöglicht werden, oder ob einige von ihnen einfach verloren gehen werden. Meines Erachtens sollte eine umsichtige Regulierung zumindest Notfallpläne für den erheblichen Verlust von Arbeitsplätzen in bestimmten, vom Aufkommen der generativen KI besonders betroffenen Sektoren erstellen. Wichtig ist, dass sich der Verlust einer erheblichen Anzahl von Arbeitsplätzen auf die Steuereinnahmen auswirken wird. Daher müssen im Rahmen eines solchen Notfallplans neue Modelle für potenziell steuerpflichtige generative KI-Anwendungen ins Auge gefasst werden, die tatsächlich Arbeitsplätze ersetzen, die formal Steuern generieren.

Es versteht sich von selbst, dass auch erhebliche Investitionen in die Ausbildung und das Upscaling getätigt werden müssen. Unsere wissensbasierte Gesellschaft wird nur dann wohlhabend und produktiv bleiben, wenn wir uns die generative KI zu eigen machen und bei ihrer Entwicklung und Anwendung führend werden.

3. Inwieweit können Anwendungen aus Staats- oder Wirtschaftssystemen, die nicht immer demokratische und liberale Werte teilen, die europäische Gesellschaft beeinflussen, und wie sollten die EU und Deutschland damit umgehen?

Die größte Gefahr, die von der KI ausgeht, ist wohl nicht das Modell, sondern der Mensch, der es für böswillige Zwecke einsetzt. Ein besonderes Problem ist die automatische Massengenerierung von Fake News und Hassreden.

"Jüngste Experimente haben gezeigt, dass ChatGPT trotz angeborener Schutzmechanismen [32] genutzt werden kann, um Hassrede-Kampagnen in großem Umfang zu produzieren, einschließlich des für eine maximale Verbreitung erforderlichen Codes [8]. Darüber hinaus machen die Geschwindigkeit und die syntaktische Genauigkeit von LGAIMs sie zum perfekten Werkzeug für die massenhafte Erstellung von hochglanzpolierten, scheinbar faktengeladenen, aber zutiefst verdrehten Fake News [7, 17]. In Kombination mit der faktischen Abschaffung der Inhaltsmoderation auf Plattformen wie Twitter bahnt sich ein perfekter Sturm für den nächsten globalen Wahlzyklus an."[25]

Die EU verfügt natürlich über ein Instrument, um die Flut illegaler Online-Inhalte einzudämmen, insbesondere im Hinblick auf Hassreden und bestimmte Arten von Fake News:

[25] Hacker, Engel und Mauer, 'Regulierung von ChatGPT und anderen großen generativen KI-Modellen', 2.

den Digital Services Act (DSA). Leider gilt der DSA jedoch nicht für direkte KI-Systeme, einschließlich Stiftungsmodelle,[26] wodurch eine gefährliche Regelungslücke entsteht.

Daher schlage ich vor, einige der Verpflichtungen der DSA auf die Entwickler generativer KI auszuweiten. Wie könnte das funktionieren? "Wir stellen uns vor, dass es zwei Komponenten hat. Diese Komponenten würden eine zentrale und eine dezentrale Überwachung im Rahmen eines Melde- und Aktionsmechanismus kombinieren (vgl. Artikel 16 DSA).

Die erste Komponente macht sich sozusagen die Weisheit der Masse zunutze, um die LGAIM-Ausgabe zu korrigieren. Die Nutzer sollten in die Lage versetzt werden, problematische Inhalte zu markieren und zu melden. Ein besonderer Status sollte einer bestimmten Gruppe von Nutzern eingeräumt werden, den "Trusted Flaggers" (vgl. Artikel 22 DSA), bei denen es sich um Privatpersonen, technologisch versierte NROs oder freiwillige Programmierer handeln könnte. Nachdem sie sich bei der zuständigen Behörde registriert haben, würden sie im Wesentlichen als dezentrales Team zur Überwachung von Inhalten fungieren. Sie könnten mit verschiedenen Eingabeaufforderungen experimentieren und sehen, ob sie schädliche oder anderweitig problematische Inhalte erzeugen können. Sie könnten auch das Internet nach Werkzeugen durchsuchen, mit denen sich die Richtlinien und Instrumente zur Inhaltsmoderation bei LGAIMs umgehen lassen. Wenn sie etwas finden, würden vertrauenswürdige Flagger eine Meldung mit der Aufforderung und der Ausgabe an einen Check-in-Punkt für die Inhaltsmoderation des jeweiligen LGAIM-Systems senden, der die Meldung an die Entwickler und/oder Bereitsteller weiterleiten würde.

Hier kommt die zweite Komponente ins Spiel, die sich an Techniker richtet, die mit Entwicklern oder Bereitstellern zusammenarbeiten. Sie müssten auf Meldungen reagieren, die von vertrauenswürdigen Markierern eingereicht werden, und diese müssten vom Inhaltsmoderationsteam priorisiert werden. Ihre Aufgabe besteht im Wesentlichen darin, das KI-System zu modifizieren oder seine Ausgabe zu blockieren, so dass die markierte Eingabeaufforderung keine problematischen Ausgaben mehr erzeugt, und generell nach Möglichkeiten zu suchen, einfache Umgehungen zu blockieren, die von böswilligen Akteuren versucht werden. Wenn das LGAIM-System groß genug ist, wäre es außerdem mit der Einrichtung eines umfassenderen Compliance-Systems beauftragt (vgl. Artikel 34-35 DSA). Insgesamt könnte sich eine solche Kombination aus zentraler und dezentraler Überwachung als effektiver und effizienter erweisen als die derzeitigen Systeme, die sich im Wesentlichen auf den guten Willen verlassen, um die erwartete Flut von Hassreden, Fake News und anderen problematischen Inhalten zu bewältigen, die von LGAIMs generiert werden."[27]

[26] Dies ist darauf zurückzuführen, dass Unternehmen, die Foundation/Generative AI-Modelle entwickeln oder einsetzen, keine Vermittler sind, die Inhalte anderer Personen hosten, sondern selbst Inhalte erstellen; siehe ebd., Teil 6.

[27] Hacker, Engel und Mauer, "Regulierung von ChatGPT und anderen großen generativen KI-Modellen", 20.

4. Bislang gibt es eine Reihe von Ideen und Projekten, die von Wasserzeichen bis hin zu Tools reichen, die KI-generierte Texte markieren oder erkennen sollen - beides wird angesichts mangelnder Konsistenz oder Genauigkeit kritisch kommentiert. Wie könnte ein sicheres und effektives Verfahren zur Kennzeichnung von durch generative KI erstellten Inhalten konkret aussehen? Und welche begleitenden Informationen könnten den Nutzern zu Bildungszwecken zur Verfügung gestellt werden?

Wasserzeichen sind in der Tat eine gute Idee, könnten sich aber in einigen Fällen als zu einfach erweisen, um sie zu entfernen. Was die Transparenz betrifft, so würde ich folgende Vorschläge machen:

a) EP-Vorschlag

"i) Die Transparenzverpflichtung bezüglich des Einsatzes von AI (Art. 28b(4) AI-Gesetz EP-Fassung, Art. 52(1) KI-Gesetz) ist ein Schritt in die richtige Richtung. Sie adressiert Pflichten der Anbieter gegenüber den Nutzern von KI-Systemen. Aus unserer Sicht sind darüber hinaus in einigen Fällen auch Pflichten der Nutzer gegenüber den Empfängern gerechtfertigt, um die Verbreitung von Fake News und Fehlinformationen zu bekämpfen." [28]

b) Verpflichtungen für Entwickler und Bereitsteller

"Erstens sollten LGAIM-Entwickler und -Einrichter verpflichtet werden, über die Herkunft und Pflege der Trainingsdaten, die Leistungskennzahlen des Modells und alle Vorfälle und Abhilfestrategien in Bezug auf schädliche Inhalte zu berichten. Idealerweise sollten sie, soweit technisch machbar [54, S. 28, Anhang A], auch die Treibhausgasemissionen des Modells offenlegen, um einen Vergleich und eine Analyse durch Regulierungsbehörden, Überwachungsorganisationen und andere interessierte Parteien zu ermöglichen. Diese Informationen könnten auch als Grundlage für eine AI-Nachhaltigkeitsfolgenabschätzung dienen."[29]

c) Transparenzverpflichtungen für Nutzer

"Zweitens sollten *professionelle Nutzer* dazu verpflichtet werden, offenzulegen, welche Teile ihrer öffentlich zugänglichen Inhalte von LGAIMs generiert oder auf der Grundlage ihrer Ergebnisse angepasst wurden. Konkret bedeutet dies, dass adidas im Beispiel von adidas die Nutzer angemessen darüber informieren muss, dass das Design z. B. mit Stable Diffusion erstellt wurde. Während der Mehrwert einer solchen Information in Verkaufsfällen begrenzt sein mag, ist eine solche Information in allen Fällen, in denen es um Inhalte im Bereich des Journalismus, der akademischen Forschung oder der Bildung geht, von entscheidender Bedeutung. In diesem Fall profitieren die Empfänger von einem Einblick in die Generationspipeline. Sie können eine solche Offenlegung als Warnsignal nutzen und eine zusätzliche Überprüfung der Fakten vornehmen oder zumindest den Inhalt *als gegeben* hinnehmen. Schließlich können wir uns vorstellen, zwischen spezifischen Anwendungsfällen zu unterscheiden, in denen die Transparenz von KI-Output gegenüber den Empfängern gerechtfertigt ist (z. B. Journalismus, akademische Forschung oder Bildung), und anderen, in denen eine solche Offenlegung auf der Grundlage weiterer Analysen und Marktuntersuchungen möglicherweise nicht gerechtfertigt ist (z. B. bestimmte Vertriebs-, Produktions- und B2B-

[28] Hacker, Engel und Mauer, "Regulierung von ChatGPT und anderen großen generativen KI-Modellen", 7.
[29] Hacker, Engel und Mauer, "Regulierung von ChatGPT und anderen großen generativen KI-Modellen", 18.

Szenarien). Zum jetzigen Zeitpunkt würden wir jedoch eine allgemeine Offenlegungspflicht für professionelle Nutzer befürworten, um weitere Informationen und Einblicke in die Rezeption solcher Offenlegungen durch andere Marktteilnehmer oder Empfänger zu erhalten.

Umgekehrt sind wir der Meinung, dass *nicht-professionelle Nutzer* nicht verpflichtet sein sollten, über den Einsatz von KI zu informieren. In dem Geburtstagsbeispiel müsste also ein Elternteil die Eltern nicht darüber informieren, dass die Einladung oder das gesamte Design der Geburtstagsfeier z. B. durch Luminous oder ChatGPT von Aleph Alpha ermöglicht wurde. Dagegen könnte man sich wehren, wenn es um die private Nutzung sozialer Medien geht, insbesondere um schädliche Inhalte, die mit Hilfe von LGAIMs erzeugt wurden. Allerdings würde jede Vorschrift zur Offenlegung von KI-generierten Inhalten wahrscheinlich von böswilligen Akteuren, die schädliche Inhalte veröffentlichen wollen, missachtet werden. Letztendlich könnte man jedoch in Erwägung ziehen, Szenarien in den sozialen Medien in den Anwendungsbereich der Transparenzregel einzubeziehen, wenn die KI-Erkennungswerkzeuge ausreichend zuverlässig sind. In diesen Fällen könnten böswillige Beiträge aufgedeckt werden, und die Akteure wären nicht nur mit den herkömmlichen zivil- und strafrechtlichen Anklagen konfrontiert, sondern auch mit der Durchsetzung des KI-Gesetzes, was finanziell erheblich sein könnte (Geldbußen) und somit einen noch größeren Anreiz schaffen würde, die Transparenzvorschrift einzuhalten oder von der Verbreitung schädlicher Inhalte abzusehen.

Da die Durchsetzung jeder nutzerorientierten Transparenzregel mühsam ist, muss sie durch technische Maßnahmen wie die Verwaltung digitaler Rechte und Wasserzeichen unterstützt werden, die durch das Modell aufgedruckt werden. Das Europäische Parlament denkt derzeit über eine Wasserzeichenpflicht für generative KI nach. Wichtig ist, dass mehr interdisziplinäre Forschung notwendig ist, um Markierungen zu entwickeln, die leicht zu benutzen und zu erkennen, aber von durchschnittlichen Nutzern schwer zu entfernen sind. Dies sollte mit Forschungen zur Erkennung von KI-Inhalten gekoppelt werden, um solche Ausgaben hervorzuheben, bei denen Wasserzeichen versagen[30]".[31]

5. Derzeit kursieren zahlreiche Vorschläge, um die regulatorischen Herausforderungen generativer KI-Anwendungen in den EU-Gesetzesvorhaben für eine KI-Verordnung und eine KI-Haftungsrichtlinie genau zu verankern: Ist der risikobasierte Ansatz zur Regulierung generativer KI überhaupt geeignet oder brauchen wir z.B. eine systemische Risikoanalyse analog zum Risikoanalyse- und -minimierungsmechanismus im DSA?

Meine Antwort darauf wäre eine doppelte. Einerseits sollten wir als Gesellschaft und als Forscher eine systemische Risikoanalyse von Stiftungsmodellen durchführen und dabei auch die möglichen langfristigen Auswirkungen auf die Beschäftigung, das Steueraufkommen und die Nutzung durch böswillige Akteure in Betracht ziehen.

Ich bin jedoch der festen Überzeugung, dass die Regulierung im Hinblick auf das KI-Gesetz und die KI-Haftungsrichtlinien eine anwendungsfallspezifische Architektur in Bezug auf Risikobewertung, -minderung und -management beibehalten sollte. Umfassende und

[30] Siehe auch https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text/.
[31] Hacker, Engel und Mauer, "Regulierung von ChatGPT und anderen großen generativen KI-Modellen", 18.

systemische Risikoanalysen könnten von der KI-Behörde, nationalen Regulierungsbehörden oder engagierten NGOs durchgeführt werden.

KI-Entwickler sind wohl kaum in der Lage, eine solche Bewertung vorzunehmen. "Die Einrichtung eines [umfassenden systemischen Risikomanagement-]Systems scheint angesichts der Vielseitigkeit von LGAIMs an das Unmögliche zu grenzen. [Der allgemeine Ansatz des Rates würde die LGAIM-Anbieter dazu verpflichten, alle "bekannten und vorhersehbaren Risiken für Gesundheit, Sicherheit und Grundrechte" in Bezug auf alle möglichen risikoreichen Verwendungen des LGAIM zu identifizieren und zu analysieren (Artikel 9(2)(a), 4b(6) AI Act Council Version). Auf dieser Grundlage müssen Strategien zur Abschwächung all dieser Risiken entwickelt und umgesetzt werden (Artikel 9 Absatz 2 Buchstabe d und Absatz 4 AI Act Council Version). Anbieter von LGAIMs wie ChatGPT müssten daher die Risiken für jede einzelne, mögliche Anwendung in jedem einzelnen, in den Anhängen II und III enthaltenen Hochrisikofall hinsichtlich Gesundheit, Sicherheit und aller möglichen Grundrechte analysieren.

Ebenso müssen Leistungs-, Robustheits- und Cybersicherheitstests für alle möglichen risikoreichen Verwendungszwecke durchgeführt werden (Artikel 15 Absatz 1, 4b Absatz 6 AI Act Council Version). Dies erscheint nicht nur fast unerschwinglich kostspielig, sondern auch kaum durchführbar. Die gesamte Analyse müsste auf einer abstrakten, hypothetischen Untersuchung beruhen und mit - wiederum hypothetischen - Risikominderungsmaßnahmen gekoppelt sein, die in vielen Fällen von der konkreten Anwendung abhängen, die zum Zeitpunkt der Analyse per definitionem noch nicht erfolgt ist. Hinzu kommt, dass viele dieser möglichen Anwendungsfälle letztlich gar nicht realisiert werden, weil sie wirtschaftlich, politisch oder strategisch unrentabel sind. Daher würde eine solche Regelung wahrscheinlich "viel Lärm um nichts" bedeuten, mit anderen Worten: eine Verschwendung von Ressourcen. Ironischerweise stellt die Konzeption der Artikel 4a-4c des KI-Gesetzes in der derzeit vorgeschlagenen Form eine sehr hohe und wohl auch unangemessene Belastung für die Anbieter wirklich universeller KI-Systeme dar. Es ist sehr unwahrscheinlich, dass diese Anbieter in der Lage sein werden, das KI-Gesetz einzuhalten, da die Vielseitigkeit ihres Modells einfach zu viele Szenarien zulässt, die in Betracht gezogen werden müssen. In Verbindung mit der vorgeschlagenen Regelung für die KI-Haftung, die Schadensersatzansprüche bei Verstößen gegen das KI-Gesetz erleichtert, setzt dies die LGAIM-Anbieter auch einem erheblichen Haftungsrisiko aus [...].

[Solche Vorschriften würden sich wahrscheinlich sehr nachteilig auf das Wettbewerbsumfeld für LGAIMs auswirken. Die Definition des KI-Gesetzes schließt ausdrücklich Open-Source-Entwickler als LGAIM-Anbieter ein, von denen es mehrere gibt.[32] Einige von ihnen erforschen LGAIMs nicht aus kommerziellen, sondern aus philanthropischen oder Forschungsgründen. Stable Diffusion wurde beispielsweise im Rahmen eines Forschungsprojekts an der LMU München entwickelt. Während das KI-Gesetz gemäß Artikel 2 Absatz 7 nicht für (wissenschaftliche, siehe Erwägungsgrund 12b des KI-Gesetzes) Forschungs- und Entwicklungstätigkeiten in Bezug auf KI-Systeme gilt, gilt diese Forschungsausnahme wohl nicht mehr, sobald das System in die freie Wildbahn entlassen wird, da eine öffentliche Freigabe wahrscheinlich nicht die wissenschaftliche Forschung und Entwicklung als "alleinigen Zweck" (Erwägungsgrund 12b des KI-Gesetzes) hat, insbesondere wenn, wie dies häufig der Fall ist,

[32] Siehe z. B. https://www.kdnuggets.com/2022/09/john-snow-top-open-source-large-language-models.html.

ein kommerzieller Partner eintritt, um die Haftung zu begrenzen und die notwendige Feinabstimmung vorzunehmen.

Infolgedessen müssen alle Unternehmen - ob groß oder klein -, die LGAIMs entwickeln und auf den Markt bringen, dieselben strengen und risikoreichen Auflagen erfüllen. Angesichts der Schwierigkeit, diese einzuhalten, ist zu erwarten, dass nur große, finanzstarke Akteure (wie Google, Meta, Microsoft/Open AI) die Kosten für die Veröffentlichung eines LGAIM aufbringen können, das in etwa dem AI Act entspricht. Für Open-Source-Entwickler und viele KMU wird die Einhaltung des Gesetzes wahrscheinlich unerschwinglich teuer sein. Daher könnte das KI-Gesetz die unbeabsichtigte Folge haben, eine weitere wettbewerbswidrige Konzentration auf dem LGAIM-Entwicklungsmarkt zu fördern. Dies steht in direktem Widerspruch zum Geist von Erwägungsgrund 61 Satz 5 des KI-Gesetzes, der im Zusammenhang mit der Normung ausdrücklich eine angemessene Beteiligung von KMU fordert, um Innovation und Wettbewerbsfähigkeit im Bereich der KI in der Union zu fördern (siehe auch Artikel 40 Absatz 2 Buchstabe b und Artikel 53 Absatz 1b Buchstabe a des KI-Gesetzes). Ähnliche Auswirkungen wurden bereits im Zusammenhang mit der Datenschutz-Grundverordnung festgestellt . In diesem Sinne droht das KI-Gesetz die Bemühungen des Gesetzes über digitale Märkte[33] zu untergraben, einen funktionierenden Wettbewerb in den Kern der digitalen und Plattform-Wirtschaft zu bringen."[34]

6. Sind neue Phänomene und Fragestellungen im Hinblick auf einen negativen Einfluss von Anwendungen generativer KI auf den demokratischen Meinungsbildungsprozess zu erwarten und wie können Medienfreiheit und Meinungsvielfalt im Zeitalter generativer KI rechtlich und politisch gestärkt werden, auch - aber nicht nur - im Hinblick auf eine angemessene Vergütung von Journalisten, Künstlern und Kreativen, und wo sehen Sie einen möglichen Anpassungsbedarf, etwa im Urheberrecht?

Was die demokratischen Prozesse und die öffentliche Meinung betrifft, siehe meine Antwort auf Frage 3.

Was das Urheberrecht betrifft, so hat die EU bereits einen Mechanismus mit Art. 3 und 4 C-DSM-Richtlinie. Hier muss dringend das Vetorecht der Rechteinhaber nach Art. 4 Abs. 3 C-DSM-Richtlinie vereinheitlicht werden, damit KI-Entwickler beurteilen können, ob die Werke, an denen sie das Modell trainieren wollen, für das KI-Training außerhalb der Forschung verwendet werden können.

Was die Entlohnung betrifft, so habe ich keine besonderen Erkenntnisse zu diesem Punkt.

[33] Verordnung (EU) 2022/1925 des Europäischen Parlaments und des Rates vom 14. September 2022 über wettbewerbsfähige und faire Märkte im digitalen Bereich, ABl. L265/1 (DMA).
[34] Hacker, Engel und Mauer, "Regulierung von ChatGPT und anderen großen generativen KI-Modellen", 5-6.

7. **Welche rechtlichen Ansatzpunkte gibt es im EU-Recht (z.B. KI-Gesetz, Wettbewerbsrecht, Urheberrechtsrichtlinie) und im nationalen Recht (z.B. UWG, Medienstaatsvertrag), um eine Kennzeichnungspflicht für KI-generierte Inhalte (z.B. Videos, Bilder oder Texte) und Entscheidungen möglichst ohne Umgehungsmöglichkeiten durchzusetzen - und welche technischen Ansatzpunkte sind denkbar, um solche Pflichten in digitalen Diensten effektiv um- und durchzusetzen?**

Auf EU-Ebene wäre der logische Ort für die Umsetzung das AI-Gesetz. Dies würde jedoch bedeuten, dass die Regelung wahrscheinlich erst ab 2024 oder 2025 verbindlich wäre. Daher wäre es ratsam, eine solche Regelung rasch in nationales Recht umzusetzen. Ein guter Ansatzpunkt könnte hier der Medienstaatsvertrag sein oder einfach ein neuartiges Bundesgesetz, das dann irgendwann durch das EU-KI-Gesetz abgelöst wird.

8. **Welche technischen und organisatorischen Maßnahmen halten Sie für geeignet, um Minderjährige zu schützen - sowohl im Hinblick auf die Einbeziehung ihrer personenbezogenen Daten in die Trainings- und Lernumgebung der generativen KI als auch im Hinblick auf die tatsächliche Nutzung von Anwendungen, die KI-basierte Texte, Videos oder Bilder generieren?**

Dies ist eine ausgezeichnete Frage. Ich würde die folgende Architektur vorschlagen:

- Tools zur Altersüberprüfung, wie sie von OpenAI offenbar als Reaktion auf die Anforderungen der italienischen Datenschutzbehörde eingeführt wurden.
- obligatorische und minderjährigenspezifische Tools zur Moderation von Inhalten, abgestimmt auf die BIK+ Strategie,[35] ausgelöst durch bestimmte Altersgrenzen
    - Drei Kernpunkte von BIK+:
        - "Sichere digitale Erfahrungen, Schutz von Kindern vor schädlichen und illegalen Online-Inhalten, -Verhalten und -Risiken und Verbesserung ihres Wohlbefindens durch ein sicheres, altersgerechtes digitales Umfeld.
        - Digitale Befähigung, damit Kinder die notwendigen Fähigkeiten und Kompetenzen erwerben, um sachkundige Entscheidungen zu treffen und sich in der Online-Umgebung sicher und verantwortungsvoll auszudrücken.
        - Aktive Beteiligung, Achtung der Kinder, indem man ihnen ein Mitspracherecht in der digitalen Umgebung einräumt, mit mehr von Kindern geleiteten Aktivitäten, um innovative und kreative, sichere digitale Erfahrungen zu fördern".
- Die personenbezogenen Daten von Kindern und Jugendlichen sind bereits durch die DSGVO geschützt.
- Darüber hinaus muss Artikel 28b der Richtlinie über audiovisuelle Mediendienste konsequent durchgesetzt werden.

[35] https://digital-strategy.ec.europa.eu/en/policies/strategy-better-internet-kids.

9. **Welche KI-getriebene Wirtschaftsentwicklung prognostizieren Sie für die deutsche und europäische Wirtschaft kurz-, mittel- und langfristig im Hinblick auf ihre jeweilige spezifische Struktur und gehen Sie von einer positiven oder negativen Entwicklung hinsichtlich der Implikationen für die realwirtschaftliche Leistungsfähigkeit dieser Volkswirtschaften im globalen Vergleich aus, auch abhängig von der Regulierung?**

Ich bin kein Experte in diesen Fragen.

10. **Was halten Sie von dem Brief des Future of Life Institute, der von vielen anerkannten KI-Experten unterzeichnet wurde? Inwieweit teilen Sie die darin geäußerten Bedenken und halten Sie die darin formulierten Forderungen für sinnvoll?**

Die Regulierung sollte sich auf die unmittelbaren Risiken der KI konzentrieren (Datenschutz, Diskriminierung, Haftung, Nachhaltigkeit, Moderation von Inhalten, Qualität). Natürlich sollten wir unsere Augen nicht vor den langfristigen Risiken verschließen. Das vorgeschlagene Moratorium ist in diesem Sinne jedoch nicht hilfreich.

Generell denke ich, dass ein Moratorium eine gute Idee sein kann, aber nur, wenn a) wirklich alle relevanten Akteure einbezogen werden und b) es wahrscheinlich ist, dass die Dauer des Moratoriums eine sinnvolle Regulierung ermöglicht. Angesichts der aktuellen geopolitischen Lage mit Russland und China und dem harten Wettbewerb im "Westen" halte ich a) nicht für realistisch; angesichts der Uneinigkeit über die Regulierung von GenAI in der EU und darüber hinaus scheint auch b) keine Option zu sein. Insbesondere nicht in 6 Monaten.

Selbst wenn man b) als gegeben hinnimmt - um der Analyse willen -, müssten wir intensiv darüber nachdenken, was ein Moratorium unter "westlicher" Führung für die KI-Entwicklung (durch Staaten und Akteure, die sich nicht daran halten), den Einsatz dieser unsicheren Systeme - sowohl in Entwicklungsländern als auch in der westlichen Welt -, die geopolitische Strategie und die Wettbewerbsfähigkeit von Forschung und Industrie in der EU und den USA bedeuten würde. Dies alles sind schwierige Fragen.

Politisch gesehen halte ich ein verbindliches Moratorium weder in der EU noch in den USA für eine realistische Option.

11. **Laut dem Bundesverband KI sind Investitionen von 300 Millionen Euro nötig, um in Deutschland eine Recheninfrastruktur für das Training von Algorithmen aufzubauen. Sollte es Ihrer Meinung nach Aufgabe des Staates sein, durch die (Mit-)Finanzierung einer solchen Infrastruktur eine aktive Industriepolitik zu betreiben, um deutschen Unternehmen das Bestehen auf dem globalen Markt für generative KI zu ermöglichen?**

Ja, und 300 Millionen Euro reichen höchstwahrscheinlich nicht einmal aus, um mit den Investitionen, die beispielsweise in den USA und im Vereinigten Königreich getätigt werden, gleichzuziehen. Der Staat muss eine führende Rolle in dem Bestreben übernehmen, eine europäische Recheninfrastruktur zu ermöglichen, vorzugsweise zusammen mit anderen EU-Staaten. Der Markt ist bereits konzentriert und viele wichtige Entwicklungen finden außerhalb

der EU statt. Die Marktanteile werden wohl schon jetzt vergeben. Wenn man in diesem Rennen mithalten will - was man aufgrund der geopolitischen Souveränität und Unabhängigkeit muss - , brauchen wir diese Ressourcen dringend.

12. Es besteht weitgehende Einigkeit darüber, künstliche Inte ligenz so zu regulieren, dass ihr Einsatz bestimmten Wertvorstellungen folgt. Wie kann dies konkret umgesetzt werden und wo sollte die Grenze zu einer möglichen Überregulierung gezogen werden, bei der künstliche Intelligenz zur künstlichen Ideologie werden könnte?

Diese Grenze zu ziehen, wird im Einzelfall immer schwierig sein, wie die Rechtsprechung zu schädlichen Äußerungen zeigt. Die Antwort muss aber meines Erachtens eine Regulierung für vertrauenswürdige KI und eine inhaltliche Moderation generativer KI sein, wie oben skizziert (Fragen 3 und 6).

13. Bislang kommen fast drei Viertel aller großen KI-Gründungsmodelle aus den USA und weitere fünfzehn Prozent aus China. Welche Maßnahmen sollte die Politik in Deutschland und Europa vor diesem Hintergrund vorrangig ergreifen, um das Ökosystem der generativen KI zu fördern und zu stärken, wenn wir nicht völlig abhängig von außereuropäischen Gründungsmodellen werden und nur am Ende der Wertschöpfungskette als Abnehmer dieser Modelle auftreten wollen?

Siehe die Antwort auf Frage 11.

Außerdem müssen wir den Weg ebnen, um internationale Talente im Bereich der KI anzuziehen und zu halten.

Schließlich könnten wir eine spezifische Unterstützung für europäische KMU benötigen, sowohl in finanzieller als auch in regulatorischer Hinsicht. So könnten wir europäischen KMU beispielsweise mit einem bevorzugten Zugang zu Sandboxen, mit milderen Bestimmungen für die Umsetzung der Anforderungen des KI-Gesetzes und mit finanzieller Unterstützung für Versicherungen helfen.

14. Welche Regelungen braucht es aus Ihrer Sicht im KI-Gesetz für generative KI, insbesondere im Hinblick auf die Verpflichtung der Entwickler von Basismodellen zur Weitergabe von Informationen innerhalb der Lieferkette, welche Vor- und Nachteile sind damit verbunden und ab welcher Schwelle sollten die im KI-Gesetz vorgesehenen Hochrisikoregeln für Anwendungen auf Basis generativer KI greifen?

Siehe die Antwort auf Frage 1.

Was die Informations- und Zugangsrechte in der KI-Wertschöpfungskette im Einzelnen betrifft:

"Einzelne Akteure in der KI-Wertschöpfungskette verfügen möglicherweise einfach nicht über das allumfassende Wissen und die Kontrolle, die erforderlich wären, wenn sie die einzigen

Adressaten von Regulierungspflichten wären . Diese eher abstrakte Beobachtung zeigt auch, dass gemeinsame und sich überschneidende Zuständigkeiten erforderlich sein können.

Unserer Ansicht nach sind Kooperationen zwischen LGAIM-Anbietern, -Einrichtern und -Nutzern im Hinblick auf die Erfüllung der regulatorischen Pflichten der einzige Weg nach vorn, wenn die Regulierungsbehörde dieser (erzwungenen) Zusammenarbeit angemessene Konturen verleiht. Konkret schlagen wir eine Kombination von Strategien vor, die aus der vorgerichtlichen Offenlegung, dem Gesetz über Geschäftsgeheimnisse und der Datenschutz-Grundverordnung bekannt sind. Unter dem aktuellen GA AI-Gesetz des Rates wird eine solche Zusammenarbeit in Artikel 4b(5) gefördert: Anbieter "müssen" mit den Nutzern zusammenarbeiten und ihnen die erforderlichen Informationen zur Verfügung stellen. Ein zentrales Thema, das ebenfalls in dem Artikel erwähnt wird, ist der Zugang zu Informationen, die möglicherweise als Geschäftsgeheimnisse oder Rechte an geistigem Eigentum (IP) geschützt sind . In diesem Zusammenhang verpflichtet Artikel 70 Absatz 1 des AI-Gesetzes alle an der Anwendung des AI-Gesetzes "Beteiligten", "geeignete technische und organisatorische Maßnahmen zu treffen, um die Vertraulichkeit der Informationen und Daten zu gewährleisten, die sie bei der Durchführung ihrer Aufgaben und Tätigkeiten erhalten". Um praktikabel zu sein, bedarf diese Verpflichtung einer weiteren Konkretisierung; dies gilt auch für den Vorschlag des Europäischen Parlaments in dieser Richtung ; Art. 10(6a) AI Act EP Version spricht nur explizit eine Situation an, in der eine solche Zusammenarbeit nicht stattfindet, und ist auf Verstöße gegen Art. 10.

Das Problem des Gleichgewichts zwischen Zusammenarbeit und Offenlegung und dem Schutz von Informationen ist nicht auf das AI-Gesetz beschränkt. Unseres Erachtens hat es eine interne und eine externe Dimension. Intern, d.h. in der Beziehung zwischen der antragstellenden und der zugangsgewährenden Partei, werden Zugangsrechte von der zugangsgewährenden Partei oft mit dem Hinweis auf vermeintlich unüberwindbare Geschäftsgeheimnisse oder Rechte des geistigen Eigentums abgewehrt . Die von der EU-Kommission vorgeschlagenen Haftungsrichtlinien enthalten beispielsweise ausgefeilte Regeln zur Offenlegung von Beweismitteln, die die Entschädigungsinteressen der Geschädigten gegen die Geheimhaltungsinteressen der KI-Entwickler und -Einrichter ausspielen . Artikel 15(4) DSGVO enthält eine ähnliche Bestimmung, die analog auch für das Auskunftsrecht in Artikel 15(1) DSGVO gilt .

Zu diesem Problem gibt es umfangreiche Literatur und praktische Erfahrungen im Bereich des US-amerikanischen Pretrial Discovery Systems . Nach diesem Mechanismus, der teilweise von den vorgeschlagenen EU-Vorschriften zur Offenlegung von Beweismitteln übernommen wurde können Geschädigte Zugang zu Dokumenten und Informationen beantragen, die sich im Besitz des potenziellen Beklagten befinden, noch bevor ein Rechtsstreit eingeleitet wird. Dies wiederum kann dazu führen, dass Wettbewerber unbegründete Anträge auf Zugang stellen. Solche Bedenken sind in der KI-Wertschöpfungskette nicht zu vernachlässigen. Auch hier können Entwickler, Anwender und Nutzer nicht nur Geschäftspartner, sondern auch (potenzielle) Wettbewerber sein. Daher muss der Zugang von Entwicklern und Nutzern begrenzt werden. Umgekehrt muss ein gewisser Informationsfluss ermöglicht werden, um die Einhaltung der risikoreichen Verpflichtungen durch die Entwickler zu gewährleisten.

Um Missbrauch vorzubeugen, schlagen wir eine Reihe von Maßnahmen vor. Einerseits können die Anbieter (und möglicherweise auch die Verleiher) die Nutzung des Modells unter der Bedingung genehmigen, dass die Nutzer *NDAs* und Wettbewerbsverbote unterzeichnen. Die

private Auftragsvergabe sollte bis zu einem gewissen Grad zwischen professionellen Akteuren funktionieren. Andererseits könnte es sich lohnen, Bestimmungen einzuführen, die sich am US-amerikanischen Pretrial-Discovery-System und dem vorgeschlagenen EU-Mechanismus zur Offenlegung von Beweismitteln (Artikel 3 Absatz 4 der Richtlinie über die Haftung von Versicherungsunternehmen, Schutzanordnung). So sollten die Gerichte befugt sein, *Schutzanordnungen* zu erlassen, die Geheimhaltungsvereinbarungen mehr Gewicht verleihen und sie potenziellen Verwaltungssanktionen unterwerfen. Die Anordnung kann auch bestimmte Geschäftsgeheimnisse von der Offenlegung ausnehmen oder den Zugang nur unter bestimmten Bedingungen erlauben (siehe F.R.C.P. Rule 26(c)(1)(G)). Wie die aufsehenerregenden Fälle von Dokumentenprüfungen in den USA im Zusammenhang mit ehemaligen und amtierenden US-Präsidenten zeigen, kann die Ernennung eines *Special Master* letztlich ein Gleichgewicht zwischen dem Zugang zu Informationen und der unzulässigen Aneignung von Wettbewerbsvorteilen herstellen (vgl. F.R.C.P. Rule 53(a)) . Mit diesen Sicherheitsvorkehrungen sollten LGAIM-Entwickler gezwungen und nicht nur ermutigt werden, mit Anwendern und Nutzern in Bezug auf die Einhaltung des KI-Gesetzes zusammenzuarbeiten, wenn sie den Einsatz genehmigt haben.

Was die externe Dimension betrifft, so stellt sich die Frage, wer für die Erfüllung der einschlägigen Pflichten verantwortlich sein und bei Verstößen gegen risikoreiche Vorschriften letztlich für Geldbußen und zivilrechtliche Schäden haften sollte. Hier können wir uns von Artikel 26 der Datenschutz-Grundverordnung inspirieren lassen (siehe auch). Nach dieser Bestimmung können gemeinsam für die Verarbeitung Verantwortliche intern eine maßgeschneiderte Aufteilung der Pflichten nach der Datenschutz-Grundverordnung vereinbaren (Artikel 26 Absatz 1 DSGVO), haften aber weiterhin gesamtschuldnerisch (Artikel 26 Absatz 3 DSGVO). Der Grund für diese Regelung ist die Erleichterung der Entschädigung der betroffenen Personen, die nicht befürchten müssen, von beiden für die Verarbeitung Verantwortlichen abgewiesen zu werden, wobei jeder die Schuld auf den anderen schiebt. Darüber hinaus muss das Wesen der internen Zuweisung zur Einhaltung der Vorschriften offengelegt werden (Artikel 26 Absatz 2 DSGVO). Dieser Mechanismus könnte mutatis mutandis auf die KI-Wertschöpfungskette übertragen werden. Auch hier ist eine Zusammenarbeit erforderlich und sollte schriftlich dokumentiert werden, um die nachträgliche Rechenschaftspflicht zu erleichtern. Die Offenlegung der wichtigsten Teile des Dokuments unter Wahrung von Geschäftsgeheimnissen dürfte potenziellen Klägern die Auswahl der richtigen Partei für etwaige anschließende Anträge auf Offenlegung von Beweismitteln im Rahmen der KI-Haftungsregelung erleichtern. Schließlich gewährleistet die gesamtschuldnerische Haftung die Zusammenarbeit und dient den Entschädigungsinteressen der Geschädigten. Intern können Parteien, die von geschädigten Personen haftbar gemacht werden, sich dann umdrehen und von anderen in der KI-Wertschöpfungskette Entschädigung verlangen. Wenn beispielsweise die Entwickler über ein API-Vertriebsmodell im Wesentlichen die Kontrolle behalten, wird die interne Haftungslast häufig auf sie fallen. Die Haftung von Entwicklern und Anbietern muss jedoch dort enden, wo ihr Einfluss auf das eingesetzte Modell endet. Jenseits dieses Punktes sollten nur die Nutzer Gegenstand von Regulierung und zivilrechtlicher Haftung sein (und umgekehrt, z. B. in Fällen der Kontrolle über eine API): Handlungsanreize sind nur dann sinnvoll, wenn derjenige, der einen Anreiz erhält, auch tatsächlich in der Lage ist, zu handeln . Im Rahmen der Datenschutz-Grundverordnung wurde dies vom EuGH in der Rechtssache Fashion ID entschieden (EuGH, C-40/17, Rn. 85). Die alleinige Verantwortung der Nutzer für bestimmte Bereiche sollte dann auch in die offengelegte Vereinbarung aufgenommen werden, um potenzielle Kläger zu informieren und nicht

einklagbare Ansprüche gegen den Entwickler und Bereitsteller auszuschließen. Ein solches System würde unseres Erachtens einen angemessenen Interessen- und Machtausgleich zwischen LGAIM-Entwicklern, -Einrichtern, -Nutzern und -Betroffenen herstellen.

Insgesamt enthält die EP-Fassung des KI-Gesetzes nun zu Recht Regeln für die KI-Wertschöpfungskette . Diese müssen jedoch, wie in den vorangegangenen Abschnitten dargelegt, präzisiert werden, um effektiv zu funktionieren. Letztlich ist die Zuordnung von Verantwortung und Haftung entlang der Wertschöpfungskette entscheidend, wenn das KI-Gesetz seinen Geist eines technologiespezifischen Instruments bewahren will, das aber nicht Modelle per se, sondern primär Modelle in konkreten Anwendungsfällen regelt."[36]

15. Welche Initiativen gibt es, insbesondere für große Sprachmodelle (LLM), zur Entwicklung europäischer Modelle und wie beurteilen Sie die Möglichkeiten und Grenzen von Private Public Partnerships in diesem Bereich?

In Bezug auf konkrete Modelle und Unternehmen scheinen mir die folgenden Punkte wichtig zu sein:

- Aleph Alpha
- Mistral
- Blüte

Öffentlich-private Partnerschaften können wohl hilfreich sein, wenn der Staat bestimmte wichtige Teile der Infrastruktur kofinanziert, siehe die Antwort auf Frage 11.

16. Was sind Ihrer Einschätzung nach die nächsten Entwicklungsstufen der generativen KI, nach Sprach- und Videomodellen (Stichworte KI-Agenten, Embodied AI etc.) und wo liegen hier die größten Chancen für unsere Gesellschaft und Wirtschaft?

Andere sind besser in der Lage, diese Frage zu beantworten.

[36] Hacker, Engel und Mauer, "Regulierung von ChatGPT und anderen großen generativen KI-Modellen", 10-11.

17. Inwieweit unterscheidet sich die Verteilung von Vor- und Nachteilen durch GPAI zwischen verschiedenen Bevölkerungsgruppen (sowohl innerhalb nationaler Gesellschaften als auch global betrachtet mit Blick auf den globalen Süden/Norden) aufgrund der nachfolgend aufgeführten Aspekte: - Unterschiede im Zugang zu Technologie (z.B. aufgrund unterschiedlicher technischer, materieller, bildungsbezogener und anderer Voraussetzungen) ; Unterschiedliche Repräsentation in Trainingsdaten (z.B., ; Unterschiedliche Belastung durch stereotype Zuschreibungen und Diskriminierung (z.B. aufgrund von Geschlecht oder ethnischer Zugehörigkeit) ; Unterschiedliche Belastung durch den Ressourcenverbrauch von KI-Systemen ; und wie wäre eine gerechtere Verteilung von Vor- und Nachteilen zu erreichen?

"Darüber hinaus schlagen wir vor, dass abweichend von der Fokussierung auf LGAIM-Anwender bestimmte Pflichten zur Datenkuration, z. B. Repräsentativität und annähernde Ausgewogenheit zwischen geschützten Gruppen (vgl. Artikel 10 AI Act), für LGAIM-Entwickler gelten sollten. Das Risiko der Diskriminierung ist wohl zu groß, um es an die Nutzer zu delegieren, und muss während der Entwicklung und des Einsatzes angegangen werden. Wo immer es möglich ist, sollten diskriminierende KI-Systeme an der Wurzel angegangen werden (oft die Trainingsdaten) und nicht die ML-Pipeline oder die KI-Wertschöpfungskette hinunter verbreitet werden. Schließlich sollten diskriminierende Ergebnisse unserer Ansicht nach in allen Anwendungsfällen vermieden werden, selbst auf Geburtstagskarten. Die regulatorische Belastung muss jedoch an das abstrakte Risikoniveau und die Compliance-Kapazitäten (d. h. typischerweise die Größe) des Unternehmens angepasst werden. Zum Beispiel sollten LGAIM-Entwickler den Schulungsdatensatz proaktiv auf falsche Darstellungen geschützter Gruppen prüfen müssen, und zwar in einer Weise, die ihrer Größe und der Art des Schulungsmaterials (kuratierte Daten im Vergleich zu aus dem Internet gescrapten Twitter-Feeds) angemessen ist, und praktikable Abhilfemaßnahmen umsetzen. Zumindest sollten die realen Schulungsdaten durch synthetische Daten ergänzt werden, um die in den Online-Quellen enthaltenen historischen und gesellschaftlichen Verzerrungen auszugleichen. Beispielsweise könnten Inhalte zu Berufen, die traditionell einem Geschlecht vorbehalten sind (Krankenschwester, Arzt), automatisch kopiert und alle weiblichen Vornamen oder Bilder durch männliche ersetzt werden und umgekehrt, um einen Trainingskorpus mit geschlechtsneutraleren Berufen für die Text- und Bilderzeugung zu schaffen."[37]

[37] Hacker, Engel und Mauer, "Regulierung von ChatGPT und anderen großen generativen KI-Modellen", 19.

18. Sollte generative KI als Mehrzweck-KI grundsätzlich als Hochrisiko-KI im Sinne der europäischen KI-Verordnung eingestuft werden, um höheren Standards zu genügen, und für wie sinnvoll/realisierbar halten Sie regulatorische Optionen für generative KI wie Transparenzverpflichtungen zu Trainingsdaten und Trainingsprozessen, die Verpflichtung zur Risikobewertung durch Anbieter einer GPAI und deren Veröffentlichung, sichtbare oder unsichtbare Kennzeichnung aller oder bestimmter KI-generierter Inhalte, das Recht auf Überprüfbarkeit der Nicht-Diskriminierung und Zugang für Forscher: Innere und andere diskutierte Optionen?

Ich denke, es wäre falsch, GPAIS/Gründungsmodelle/generative KI als risikoreich einzustufen. Siehe Frage 1.

Vielmehr besteht die Architektur des KI-Gesetzes darin, risikoreiche Verpflichtungen für bestimmte risikoreiche Anwendungsfälle festzulegen. Diese Architektur sollte für generative KI nicht auf den Kopf gestellt werden.

III. Wichtigste Referenzen

Shavell S, *Grundlagen der wirtschaftlichen Analyse des Rechts* (Harvard U Press 2004)

Schmidt-Wudy F, 'Art. 15 DSGVO' in Wolff HA and Brink S (eds), *BeckOK Datenschutz* (42 edn, 2023)

Bertuzzi L, 'AI Act: Europaabgeordnete wollen Grundrechtsprüfungen, Verpflichtungen für Hochrisikonutzer' EURACTIV <https://www.euractiv.com/section/artificial-intelligence/news/ai-act-meps-want-fundamental-rights-assessments-obligations-for-high-risk-users/> Zugriff am 30. Januar 2023

Bertuzzi L, 'Führende EU-Gesetzgeber schlagen Verpflichtungen für Allzweck-KI vor' EURACTIV <https://www.euractiv.com/section/artificial-intelligence/news/leading-eu-lawmakers-propose-obligations-for-general-purpose-ai/>

Bertuzzi L, 'MEPs seal the deal on Artificial Intelligence Act' EURACTIV <https://www.euractiv.com/section/artificial-intelligence/news/meps-seal-the-deal-on-artificial-intelligence-act/>

Liang P und andere, 'The time is now to develop community norms for the release of foundation models' CRFM <https://crfm.stanford.edu/2022/05/17/community-norms.html> Zugriff am 3. Februar 2023

Bommasani R und andere, 'On the opportunities and risks of foundation models' (2021) arXiv preprint arXiv:210807258

Calvin N und Leung J, "Wem gehört die künstliche Intelligenz? Eine vorläufige Analyse von Unternehmensstrategien zum geistigen Eigentum und warum sie wichtig sind" (2020) Future of Humanity Institute, Februar

Daniel PF, 'Protecting Trade Secrets from Discovery' (1994) 30 Tort & Ins LJ 1033

Deeks A, "The judicial demand for explainable artificial intelligence" (2019) 119 Columbia Law Review 1829

Drexl J u.a., "Künstliche Intelligenz und geistiges Eigentumsrecht - Stellungnahme des Max-Planck-Instituts für Innovation und Wettbewerb vom 9. April 2021 zur aktuellen Debatte" (2021) Max-Planck-Institut für Innovation und Wettbewerb Research Paper

Edwards L, "Regulierung von KI in Europa: vier Probleme und vier Lösungen" (2022) 2022

Geradin D, Karanikioti T und Katsifis D, "GDPR Myopia: how a well-intended regulation ended up favouring large online platforms" (2021) 17 European Competition Journal 47

Grinbaum A und Adomaitis L, 'The Ethical Need for Watermarks in Machine-Generated Language' (2022) arXiv preprint arXiv:220903118

Gutierrez CI und andere, 'A Proposal for a Definition of General Purpose Artificial Intelligence Systems' (2022) Working Paper, https://ssrncom/abstract=4238951

Hacker P, "Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic discrimination under EU law" (2018) 55 Common Market Law Review 1143
Hacker P, 'The European AI Liability Directives - Critique of a Half-Hearted Approach and Lessons for the Future' (2022) Working Paper, https://arxivorg/abs/221113960
Hacker P, Engel A und Mauer M, 'Regulating ChatGPT and other Large Generative AI Models' (2023) ACM Conference on Fairness, Accountability, and Transparency (FAccT '23, in Vorbereitung) https://arxiv.org/abs/2302.02337
Hacker P und Passoth J-H, 'Varieties of AI Explanations under the Law. From the GDPR to the AIA, and Beyond' (2022) International Conference on Extending Explainable AI Beyond Deep Models and Classifiers 343
Kiela D und andere, 'The hateful memes challenge: Detecting hate speech in multimodal memes" (2020) 33 Advances in Neural Information Processing Systems 2611
Kirchenbauer J und andere, 'A Watermark for Large Language Models' (2023) arXiv preprint arXiv:230110226
Kötz H, "Zivilrechtssysteme in Europa und den Vereinigten Staaten" (2003) 13 Duke J Comp & Int'l L 61
McKown JR, "Entdeckung von Geschäftsgeheimnissen" (1994) 10 Santa Clara Computer & High Tech LJ 35
Meyers JM, "Künstliche Intelligenz und Geschäftsgeheimnisse" (2018) 11 Landslide 17
Mitchell E und andere, 'DetectGPT: Zero-Shot Machine-Generated Text Detection using Probability Curvature' (2023) arXiv preprint arXiv:230111305
Roberts J, "Zu wenig, zu spät: Unwirksamer Beistand eines Anwalts, die Pflicht zur Ermittlung und die vorgerichtliche Offenlegung in Strafsachen" (2003) 31 Fordham Urb LJ 1097
Shavell S, "Über Haftung und Versicherung" (1982) 13 Bell Journal of Economics 120
Shepherd GB, 'An empirical study of the economics of pretrial discovery' (1999) 19 International Review of Law and Economics 245
Spindler G, "Die Vorschläge der EU-Kommission zu einer neuen Produkthaftung und zur Haftung von Herstellern und Betreibern Künstlicher Intelligenz" (2022) Computer und Recht 689
Subrin SN, 'Discovery in Global Perspective: Sind wir verrückt?" (2002) 52 DePaul L Rev 299
Wachter S, 'Die Theorie der künstlichen Unveränderlichkeit: Der Schutz algorithmischer Gruppen unter dem Antidiskriminierungsgesetz" (2022) arXiv preprint arXiv:220501166
Wagner G, 'Liability Rules for the Digital Age - Aiming for the Brussels Effect' (2023) European Journal of Tort Law (in Vorbereitung)

Widder DG und Nafus D, 'Dislocated Accountabilities in the AI Supply Chain: Modularity and Developers' Notions of Responsibility' (2022) arXiv preprint arXiv:220909780
Zuiderveen Borgesius FJ, "Stärkung des Rechtsschutzes gegen Diskriminierung durch Algorithmen und künstliche Intelligenz" (2020) 24 The International Journal of Human Rights 1572
Hacker P, Engel A und List T, 'Understanding and regulating ChatGPT, and other large generative AI models' (2023) <https://verfassungsblog.de/blog/> Zugriff am 20. Januar 2023