Der europäische Ansatz für das Vertrauen in KI*

Nikolaus Forgó, Wien

Die europäische Regulierung von KI, seit Jahren diskutiert, hoch umstritten und ständig im Fluss, zum Zeitpunkt der Verfassung dieses Beitrags immer noch nicht finalisiert,[1] geht von Grundannahmen aus, die epistemischen Grundwahrheiten zu entsprechen scheinen. Solche Grundannahmen sind beispielsweise: die Überzeugung, es sei mit der menschlichen Würde nicht vereinbar, Objekt einer automatisierten Entscheidung zu werden – dem entsprechend verbietet schon Art 22 DSGVO[2] automatisierte Entscheidungen; die Erwartung, das Risiko einer KI-Anwendung vorhersehen und in Kategorien klassifizieren zu können – dem entsprechend geht der AI Act bekanntlich von einem „risikobasierten" Ansatz aus; die Idee, es sei für den individuellen menschlichen Umgang mit KI relevant zu wissen, ob man mit einer KI oder einem Menschen interagiert – dem entsprechend sieht Art 52 des Kommissionsentwurfs zum AI Act vor „dass KI-Systeme, die für die Interaktion mit natürlichen Personen bestimmt sind, so konzipiert und entwickelt werden, dass natürlichen Personen mitgeteilt wird, dass sie es mit einem KI-System zu tun haben, es sei denn, dies ist aufgrund der Umstände und des Kontexts der Nutzung offensichtlich."[3]

Nun scheint es aber so zu sein, dass diese Grundannahmen doch keine solchen sind, sondern vielmehr kulturell oder politisch abhängige Variablen, die andernorts, zu anderer Zeit oder durch andere Menschen anders gesehen werden. Als plastisches Beispiel mag hier nur genannt werden, dass manche das Risiko von

---

* Zusammenfassung eines in freier Rede gehaltenen Vortrags unter Bezugnahme auf einige tagespolitische Ereignisse bis zum Januar 2024.

1 Zwar wurde im Dezember 2023 mit großem Brimborium verkündet, es habe eine politische Einigung beim AI Act gegeben, jedoch liegt der Text hierzu immer noch nicht vor und wird weiterhin „technisch" verhandelt. Vgl etwa https://www.euractiv.com/section/artificial-intelligence/news/european-union-squares-the-circle-on-the-worlds-first-ai-rulebook/ (abgefragt 22.1.2024); https://www.euractiv.com/section/artificial-intelligence/news/ai-act-eu-policymakers-nail-down-rules-on-ai-models-butt-heads-on-law-enforcement/ (abgefragt 22.1.2024).

2 VO (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. 4. 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung), ABl L 2016/119, 1.

3 Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz (Gesetz über Künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union, COM (2021) 206 final.

(general purpose) AI als existentiell einordnen[4] – also als geeignet, die Menschheit insgesamt auszurotten. Es ist gleichzeitig recht offensichtlich, dass dieses Risko, bestünde es, inakzeptabel hoch wäre und zu einem sofortigen Stopp aller Aktivitäten führen müsste, was aber von kaum jemandem gefordert und von niemandem praktiziert wird. Ebenso offensichtlich ist, dass die AI-Applikationen, die 2023 und 2024 in aller Munde sind (mit ChatGPT an der ersten Stelle) ganz ohne irgendeine AI-spezifische Regulierung und damit insbesondere auch ohne systematische (staatliche) Risikobewertung produziert und millionenfach (milliardenfach) verbreitet und genutzt werden.

Es scheint für manche – viele? – (junge) Menschen gar nicht besonders relevant zu sein, ob sie mit einer KI oder einem Menschen kommunizieren. Sie verhalten sich auch bei der KI ganz so, als wäre sie ein Mensch. Auf Instagram, zum Beispiel, finden wir InfluencerInnen, die offensichtlich künstlich sind und über ein „Leben" berichten, das hypothetisch algorithmisch ist. Sie heißen *Millasofia*, *Noonoouri* oder *Miquela* und haben hunderttausende Follower, die ihr computergeneriertes Outfit und ihre computergenerierten Tagespläne ganz so mit Herzchen versehen, teilen und kommentieren, als handelte es sich um eine „echte" Freundin, mit der man interagiere. *Miquela* beschreibt sich selbst als „19-year-old Robot living in LA 💔"[5] und generiert trotzdem mit ihren Bildchen Reaktionen, wie „This is amazing :)" oder „You're my inspiration 💕"[6].

Mit dem Auftreten von ChatGPT in den Allgemeinmedien ist eine Veränderung in unserem medienvermittelten Alltag für viele offensichtlich geworden, die wichtig ist: Während wir uns schon seit Längerem daran gewöhnt haben, dass das Erleben unserer Wirklichkeit das Ergebnis unserer Aktivitäten auf Plattformen ist (wir fahren da entlang, wo Google Maps uns hinführt, wir beantworten Tweets, die uns in unsere Timeline gespült werden, wir vertrauen den Videoempfehlungen auf YouTube usw), und während wir uns auch daran gewöhnt haben, dass die allermeisten dieser Plattformen nicht europäische sind, sondern, bestenfalls, von US-amerikanischen Unternehmen betrieben werden, die über Niederlassungen in Europa verfügen, und während wir auch zunehmend damit vertraut sind, dass es nicht mehr wie früher einen Unterschied macht, ob wir etwas online oder offline tun, sondern wir alle immer mehr „always on" sind (und so auch die Dinge,

4 Zu nennen ist hier insb *Yuval Noah Harari*, vgl etwa https://www.youtube.com/watch?v=LWiM-LuRe6w (abgefragt 22.1.2024); vgl auch den von hunderten Führungspersönlichkeiten unterschriebenen Offenen Brief „Pause Giant AI Experiments: An Open Letter", https://futureoflife.org/open-letter/pause-giant-ai-experiments/ (abgefragt 22.1.2024).

5 https://www.instagram.com/lilmiquela/ (abgefragt 22.1.2024).

6 https://www.instagram.com/p/C1KrTmHNjnE/?utm_source=ig_web_button_share_sheet&igsh=MzRlODBiNWFlZA== (abgefragt 22.1.2024).

die uns umgeben)[7], ist nun noch etwas Neues hinzugetreten. Wir lernen gerade, dass auch der Inhalt selbst maschinell erzeugt werden kann, ohne dass dies für uns einen Unterschied machen muss. Die YouTube-Creatorin oder Twitter-Posterin oder Instagram-Influencerin kann – schon jetzt – ohne Weiteres durch einen autonom agierenden Avatar ersetzt werden – und ihnen werden vielleicht schon bald manche JournalistInnen, RechtsanwältInnen oder ServicemitarbeiterInnen folgen; vielleicht machen ja auch in gar nicht ferner Zeit virtuelle PolitikerInnen Politik für uns – denn wer will sich denn die schon antun?

Dass die Erstellung nun auch der Inhalte selbst (weiter) automatisiert und (künstlich) intelligent gemacht werden kann, erzeugt allerdings, ebenfalls schon jetzt spürbare, unangenehme Folgeeffekte. Zunächst sind die Plattformen selbst mit bisher unbekannten Mengen automatisiert generierter Inhalte konfrontiert, die manchmal, wie in den oben präsentierten Instagram-Beispielen, vermutlich dem Geschäftsmodell entsprechen, manchmal aber, etwa im Falle politischer Manipulationen großen Ausmaßes, auch nicht. Letzteres mag der Fall sein bei Twitter-Postings, die behaupten, Fotos ausgebombter Säuglinge in Trümmern im Gaza-Streifen zu zeigen, in Wirklichkeit aber das Produkt künstlicher Intelligenz mit dem Ziel politischer Manipulation sind, was sich – noch – beweisen lässt (im konkreten Beispiel an dem Umstand, dass das Kind sechs Finger an der Hand hat).[8] Manche Unternehmen reagieren auf diese Herausforderung in Ermangelung funktionierender öffentlichrechtlicher Vorgaben bereits mit privatrechtlichen Mitteln. So hat etwa YouTube im November 2023 Regeln zum Umgang von InhalteerstellerInnen mit KI auf der Plattform erlassen,[9] an deren erster Stelle – wieder einmal – Transparenz steht. Künstlich generierte Inhalte müssen als solche erkennbar sein.[10]

Erst recht werden wir, die UserInnen, zunehmend mit Inhalten konfrontiert, die nicht „echt" sind, was uns, siehe oben, nicht immer wichtig ist, aber unser Vertrauen darin, was und wem wir trauen können, weiter unterwandern mag. Allerdings geraten auch UserInnen, die von der Authentizität des Materials – und der Beurteilbarkeit dieser Authentizität – abhängig sind, weil sie über diese, zB als JournalistInnen, berichten, zunehmend in Schwierigkeiten: Noch nie, so liest man,

7 Vgl nur zB https://headphonesaddict.com/teen-kids-screen-time-statistics/ (abgefragt 22.1.2024); https://www.weforum.org/agenda/2022/08/social-media-internet-online-teenagers-screens-us/ (abgefragt 22.1.2024).

8 https://twitter.com/AyaIsleemEn/status/1714022506070266263 (abgefragt 22.1.2024).

9 https://blog.youtube/inside-youtube/our-approach-to-responsible-ai-innovation/ (abgefragt 22.1.2024).

10 https://blog.youtube/inside-youtube/our-approach-to-responsible-ai-innovation/ (abgefragt 22.1.2024): „To address this concern, over the coming months, we'll introduce updates that inform viewers when the content they're seeing is synthetic. Specifically, we'll require creators to disclose when they've created altered or synthetic content that is realistic, including using AI tools."

war es so schwierig, in den Kriegen in der Ukraine und im Gaza zu erkennen, was „tatsächlich" gerade geschieht.[11]

Das mag auch daran liegen, dass es sich bei den Quellen der Manipulation um sehr potente handeln kann – Geheimdienste, halbstaatliche Paramilitärs, organisierte Hackergruppen usw –, die – allein – durch eine sich vorgeblich verschärfende Rechtslage nicht weiter beeindruckt werden können. Die European Union Agency for Cybersecurity (ENISA) sieht in der Manipulation der anstehenden Wahlen in Europa durch AI-basierte Desinformation ein zentrales Sicherheitsproblem der kommenden Monate (und Jahre).[12]

Was kann, was soll Europa also tun? Darauf kann man, mit einem Blick in die Geschichte, zunächst, erstens, antworten, dass die Europäische Union eine inzwischen lange Tradition darin hat, die Digitalisierung durch „hard law" steuern zu wollen. Schon in der Mitte der 1990er Jahre, mit dem Auftreten der „Internet-Bubble", setzten intensive Regulierungsbemühungen ein, die in Frequenz und Breite von den heutigen gar nicht so verschieden sind: Vom Fernabsatz bis E-Commerce, von elektronischer Signatur bis Datenschutz ergingen Richtlinien, die zT wörtlich (etwa zum Datenschutz in der elektronischen Kommunikation aus 2002), zT zumindest konzeptionell weitgehend unverändert (zB zum E-Commerce) weiterhin in Kraft stehen, ohne dass diese allerdings etwas an der Dominanz nichteuropäischer Technologien und Unternehmen geändert hätten – im Gegenteil. Auch die deutlich jüngere Rede (einer Autorin europäischer Herkunft) vom „Brussels effect"[13] – also der weltweiten Führungs- und Vorbildfunktion der EU in der Regulierung der Digitalisierung – mag vielleicht vor allem in Brüssel selbst überzeugend gefunden werden.

Zweitens ist zu sagen, dass trotz (oder gerade wegen?) dieses Befunds die Kommission *von der Leyen* in eine bisher unbekannte Regulierungsaktivität durch eine Vielzahl von Verordnungen verfallen ist, die nun als Gesetze bezeichnet werden (und auch so verstanden werden wollen) und die genannten Probleme (und noch mehr) beseitigen sollen: DSA (Digital Services Act),[14] DMA (Digital Markets Act),[15]

11 Vgl zB https://www.wired.com/story/x-israel-hamas-war-disinformation/ (abgefragt 22.1.2024); https://twitter.com/IntelCrab/status/1711130580081901643 (abgefragt 22.1.2024).

12 https://www.enisa.europa.eu/news/eu-elections-at-risk-with-rise-of-ai-enabled-information-manipulation (abgefragt 22.1.2024).

13 *Bradford*, The Brussels Effect: How the European Union Rules the World, 2020.

14 VO (EU) 2022/2065 des Europäischen Parlaments und des Rates vom 19. 10. 2022 über einen Binnenmarkt für digitale Dienste und zur Änderung der Richtlinie 2000/31/EG (Gesetz über digitale Dienste), ABl L 2022/277, 1.

15 VO (EU) 2022/1925 des Europäischen Parlaments und des Rates vom 14. 9. 2022 über bestreitbare und faire Märkte im digitalen Sektor und zur Änderung der Richtlinien (EU) 2019/1937 und (EU) 2020/1828 (Gesetz über digitale Märkte), ABl L 2022/265, 1.

DGA (Data Governance Act),[16] DA (Data Act),[17] Chips Act,[18] European Health Data Space Act[19] usw heißen die bekannteren, neben dem AI Act. Es ist hier nicht der Ort und nicht die Zeit, um deren Inhalte und deren – sehr kompliziertes – Verhältnis zueinander und zur DSGVO nachzuzeichnen, eins wird allerdings schon bei ihrer bloßen Aufzählung deutlich: Es ist sehr viel neues, hartes Recht, das dazu dienen soll, das Versprechen *von der Leyens*, das sie zu Beginn ihrer Amtszeit in ihrer „Agenda for Europe" abgegeben hat, einzulösen: „It may be too late to replicate hyperscalers, but it is not too late to achieve **technological sovereignty** in some critical technology areas."[20]

Drittens ist zu erkennen, dass sich auch dieses viele neue Recht vor allem als Instrument der Bekämpfung von Missständen und Risiken versteht. Das wird durchaus breitenwirksam verhandelt, wenn etwa EU-Kommissar *Breton* anlässlich des Gaza-Kriegs und der mit ihm einhergehenden Desinformation den Twitter-Eigentümer *Elon Musk* ausgerechnet auf Twitter an (behauptete) Verpflichtungen der Plattform aus dem DSA erinnert,[21] was von *Musk* eher belustigt zur Kenntnis genommen wird.[22] („I still don't know what they're talking about! Maybe it's in the mail or something."[23]) Es wird aber auch deutlich, wenn man zählt, dass „Risiko" bzw „Risiken" in Präsentation und Begründung des AI Act durch die Kommission gezählte 435 Mal vorkommt, und auch schon gleich im ersten Absatz.[24] Das mag wenig verwunderlich sein, wenn man bedenkt, wie vergleichsweise „risikobewusst" Europa sich dem Phänomen AI im Vergleich zu anderen Gegenden der Welt nähert. Während eine sehr deutliche Mehrheit der Bevölkerung in Asien

16 VO (EU) 2022/868 des Europäischen Parlaments und des Rates vom 30. 5. 2022 über europäische Daten-Governance und zur Änderung der Verordnung (EU) 2018/1724 (Daten-Governance-Rechtsakt), ABl L 2022/152, 1.

17 VO (EU) 2023/2854 des Europäischen Parlaments und des Rates vom 13. 12. 2023 über harmonisierte Vorschriften für einen fairen Datenzugang und eine faire Datennutzung sowie zur Änderung der Verordnung (EU) 2017/2394 und der Richtlinie (EU) 2020/1828 (Datenverordnung), ABl L 2023/2854.

18 VO (EU) 2023/1781 des Europäischen Parlaments und des Rates vom 13. 9. 2023 zur Schaffung eines Rahmens für Maßnahmen zur Stärkung des europäischen Halbleiter-Ökosystems und zur Änderung der Verordnung (EU) 2021/694 (Chip-Gesetz), ABl L 2023/229, 1.

19 Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates über den europäischen Raum für Gesundheitsdaten, COM (2022) 197 final.

20 https://op.europa.eu/en/publication-detail/-/publication/43a17056-ebf1-11e9-9c4e-01aa75ed71a1 (abgefragt 22.1.2024).

21 https://twitter.com/ThierryBreton/status/1711808891757944866 (abgefragt 22.1.2024).

22 https://twitter.com/elonmusk/status/1712091639131353202 (abgefragt 22.1.2024).

23 https://twitter.com/elonmusk/status/1712098678364656066 (abgefragt 22.1.2024).

24 COM(2021) 206 final, 1: „Dieselben Faktoren und Techniken, die für den sozioökonomischen Nutzen der KI sorgen, können aber auch neue Risiken oder Nachteile für den Einzelnen oder die Gesellschaft hervorbringen."

die Entwicklung von AI als etwas gesellschaftlich Erwünschtes sieht, ist Europa hier um Vieles skeptischer.[25]

Die Ziele der Risiko- und Missstandsbekämpfung führen in all den genannten Akten regelmäßig zu neuen Behörden, neuen Bußgelddrohungen und neuen organisatorischen Anforderungen an die Regulierten;[26] nicht jedoch zu korrespondierenden Verpflichtungen auf Behördenseite – schon gar nicht über den einzelnen Gesetzesakt und die einzelnen Behörden hinaus. Es wird deshalb der Rechtswissenschaft, wie, vor allem, den Behörden selbst und den Gerichten vorbehalten bleiben, die Abgrenzungen vorzunehmen, insbesondere auch zur DSGVO, die von all dem neuen Recht weiter weitgehend unberührt bleiben soll. Die dafür zu veranschlagende Zeit und Energie werden vielleicht anderswo fehlen – etwa bei der Einschätzung und Bekämpfung der entstandenen Risiken. „Throughout the latter part of 2022 and the initial half of 2023, there was a notable escalation in cybersecurity attacks, setting new benchmarks in both the variety and number of incidents, as well as their consequences.“, schreibt die ENISA.[27] Die vorläufig entstehende Zuständigkeitskaskade wird europäischen Innovationen im Bereich der KI nicht förderlich sein.

Es ist vor diesem Hintergrund bedauerlich, dass der Trilog zur Verabschiedung des AI Act, der gerade stattfindet, während dieses Referat gehalten wird, so gehetzt ist, weil er unbedingt noch während des spanischen Ratsvorsitzes beendet sein soll. Dieser Zeitdruck ist extern, nämlich durch die anstehende Europawahl, motiviert, hat mit dem Thema selbst nichts zu tun und wird den Druck auf Kompromisse ob der Kompromisse willen erhöhen.

Es ist deshalb besonders bedenklich, dass (auch) die interessierte Fach-Öffentlichkeit von diesem Trilog kaum etwas mitbekommt, weil die vielhundertseitigen „Spaltendokumente“, in denen die verschiedenen Versionen von Kommission, Parlament und Rat nebeneinander gestellt werden, nicht offiziell veröffentlicht, sondern nur durch „Leaks“ verbreitet werden. Die offizielle Dokumentation der

25 https://www.pewresearch.org/wp-content/uploads/2020/12/ft_2020.12.15_artificial-intelligence_01.png?w=537 (abgefragt 22.1.2024).

26 Vgl etwa die komplexen Nachweise zu Cybersicherheitskonzepten wie sie sich aus der NIS-2-Richtlinie, RL (EU) 2022/2555 des Europäischen Parlaments und des Rates vom 14. Dezember 2022 über Maßnahmen für ein hohes gemeinsames Cybersicherheitsniveau in der Union, zur Änderung der Verordnung (EU) Nr 910/2014 und der Richtlinie (EU) 2018/1972 sowie zur Aufhebung der Richtlinie (EU) 2016/1148 (NIS-2-Richtlinie), ABl L 2022/333, 80, ergeben.

27 https://www.enisa.europa.eu/publications/enisa-threat-landscape-2023 (abgefragt 22.1.2024).

Vorgänge auf Eur-Lex ist um mehrere Monate verspätet: Das jüngste dort im Januar 2024 einsehbare Dokument[28] stammt aus dem Dezember 2022 [sic!].[29]

Und es ist ebenso symptomatisch wie alarmierend, dass nach jahrelanger Diskussion elementare Fragen der Regulierung, soweit man hört, weiterhin nicht (einheitlich und eindeutig) beantwortet werden können – nämlich schon die Frage, worum es sich bei AI überhaupt handelt, welche Nutzungsformen jedenfalls verboten sein sollen und wie man mit „Foundation Models" wie zB ChatGPT umgehen soll (und wie diese wiederum zu definieren sind). Die vorgeschlagenen, bemühten Unterscheidungen zwischen Machine Learning und AI mögen JuristInnen erfreuen,[30] im Lichte der Höhe drohender Bußgelder bei Fehlqualifikation wird diese Freude freilich nicht von allen geteilt werden.[31] Ob und wie Foundation Models in den risikobasierten Ansatz einzufügen sind und ob eine eigene Klasse für systemische Risiken geschaffen werden soll, wird gerade – neu – verhandelt, weil Bedenken europäischer Unternehmen lanciert wurden, dass eine Klassifizierung ihrer Produkte als hochrisikobewertet nicht nur den amerikanischen Wettbewerbern, sondern ihnen selbst schaden könnte,[32] so insbesondere dem französischen Unternehmen Mistral: „If the burden that is on the shoulder of the foundation model provider, both from the bureaucratic or legal liability point of view, is too heavy, then it could kill Mistral."[33]

Gleichzeitig werden jedoch die weit fundamentaleren Fragen, ob ein risikobasierter Ansatz überhaupt sinnvoll versucht werden kann, wenn Technologieentwicklung so disruptiv verläuft,[34] ob dieser nicht Risiken systematisch über- und Chancen systematisch unterbewertet[35] und ob überhaupt ein derart umfangreiches,

28 Document ST_15698_2022_INIT, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=consil%3AST_15698_2022_INIT (abgefragt 22.1.2024).

29 https://eur-lex.europa.eu/legal-content/EN/HIS/?uri=celex:52021PC0206 (abgefragt 22.1.2024).

30 https://twitter.com/RAinDiercks/status/1668933434415230978 (abgefragt 22.1.2024).

31 https://twitter.com/TimWybitul/status/1668942877542559744 (abgefragt 22.1.2024).

32 https://www.trendingtopics.eu/mistral-ai-ll-ai-act-kill/ (abgefragt 22.1.2024).

33 https://sifted.eu/articles/eu-ai-act-kill-mistral-cedric-o (abgefragt 22.1.2024).

34 Vgl dazu etwa den Offenen Brief an die Vertreter der Europäischen Kommission, des Europäischen Parlaments und des Europäischen Rats [sic!], https://drive.google.com/file/d/1Gw9cFC8MAcLfhfLNWRQ6UbR2oMNKyXU9/view (abgefragt 22.1.2024). Zu diesem etwa hier: https://www.nzz.ch/technologie/europaeische-unternehmensvertreter-kritisieren-eu-regulierung-von-kuenstlicher-intelligenz-in-offenem-brief-ld.1744956 (abgefragt 22.1.2024).

35 Zu dieser Kritik im Offenen Brief etwa: „Wie die Erfindung des Internets oder der Durchbruch der Siliziumchips wird generative KI als Technologie die Leistungsfähigkeit und damit die Bedeutung ganzer Regionen maßgeblich prägen: Staaten mit den leistungsfähigsten Large Language Modellen werden einen entscheidenden Wettbewerbsvorteil haben. Der Einfluss dieser Modelle ist aber noch viel bedeutender: Indem sie beispielsweise Suchmaschinen ersetzen und sich als Assistenten unseres täglichen

Risikoklassen schaffendes, zahlreiche Dokumentations- und Klassifikationspflichten bringendes, „hartes“ Gesetz erlassen werden sollte,[36] zwar von beteiligten Verkehrskreisen sehr deutlich vernehmbar gestellt, aber nicht systematisch rechtspolitisch adressiert und beantwortet, weil unter allen Umständen ein Kompromiss erzielt werden muss.

Umso wichtiger ist es, dass in einer Veranstaltung wie dieser (und in einem Sammelband wie diesem) diese Fragen gestellt werden können. Es wird sich zeigen, ob wir damit noch rechtzeitig sind und welche Grundannahmen sich werden halten können.

privaten und beruflichen Lebens etablieren, werden sie zudem mächtige Werkzeuge sein, die nicht nur unsere Wirtschaft, sondern auch unsere Kultur prägen. Europa kann es sich nicht leisten, hier ins Hintertreffen zu geraten.“

36 Zu dieser Kritik im Offenen Brief etwa: „Die Regulierung generativer KI in einem Gesetz verankern zu wollen und nach einer starren Compliance-Logik vorzugehen, ist jedoch ein ebenso bürokratischer wie ineffizienter Ansatz.“

Deus ex Machine Learning

Sozio-technische Grundlagen und Spannungsfelder[1]

Paola Lopez, Wien

Übersicht:

I. Künstliche Intelligenz als *deus ex machina*

In Geschichten gibt es verschiedene Möglichkeiten, mit den zahlreichen Problemen und vermeintlichen Unlösbarkeiten umzugehen, die sie auch interessant machen. Eine Variante ist das plötzliche Auftauchen von Göttern, die die Lage lösen – oder die Lage nicht lösen und stattdessen ausweglos machen und so die Geschichte vorantreiben. Über Götter kann man gut schreiben: Man schreibt über sie und schon sind sie Teil der Geschichte, schon sind sie anwesend. Weniger gut funktioniert der glaubhafte schauspielerische Auftritt von Göttern. Um eine Gottheit glaubhaft anwesend erscheinen zu lassen, vor allem im Rahmen einer Theatervorstellung mit Publikum, braucht es mehr als das geschriebene Wort. Im antiken griechischen Theater erachtete man es als nicht ausreichend, Gottheiten spielende Schauspieler einfach auf die Bühne gehen zu lassen. Mit einer technischen Gerätschaft, einer Art Kran (*mekhane*), wurden Schauspieler als Gottheiten von oben auf die Bühne herabgelassen. Aus dem Nichts wurden sie damit in die Geschichte platziert und konnten dort tun, was Götter eben tun: wüten oder gütig sein, rachsüchtig oder milde. Gerade dieses Aus-dem-Nichts-Kommen der Götter mittels des *mekhane* – oder moderner: aus der Maschine – ist, was heutzutage einen *deus ex machina* beschreibt. Eine Geschichte ist ausweglos, in sich widersprüchlich oder festgefahren und plötzlich passiert etwas oder erscheint etwas und löst die Situation. Zehn Menschen sind in einer Grotte tief unter der Erde verschüttet und das Grundwasser steigt, sodass sie zu sterben drohen? Plötzlich kann einer von ihnen zaubern. Plötzlich erscheint ein Außer-

1 Stand: Jänner 2024.

irdischer. Plötzlich klingelt der Wecker und es war alles nur ein Traum. Genauso zielführend wie unrealistisch sind sie, die *dei ex machina* – und vor allem unbefriedigend.

Dramen von größerem Maßstab, die man sich in der Heterogenität und Dringlichkeit nicht hätte ausdenken können, sind die Klimakrise, der steigende Bedarf an medizinischer und altersbedingter Versorgung, die Pandemie, die noch nicht vorbei ist, und das vermeintlich unlösbare Missverhältnis zwischen dem, was es an finanziellen, medizinischen, sozialstaatlichen, planetaren Ressourcen gibt, und dem, was alle wollen – sowie der Verwaltungsaufwand von alldem. Wie verteilen wir unsere Ressourcen? Wie treffen wir wichtige Entscheidungen, im Großen wie im Kleinen? In diesem Aufsatz geht es um einen bestimmten *deus ex machina*: Künstliche Intelligenz (KI). So soll KI gegen die Klimakrise eingesetzt werden,[2] oder transformativ auf das Gesundheitswesen wirken.[3] KI soll effizient und in großen Maßstäben einsetzbar sein. Dieser Aufsatz behandelt sozio-technologische Grundlagen von Künstlicher Intelligenz. Er führt eine grundlegende Unterscheidung zwischen zwei Arten von KI-Systemen ein und diskutiert einige Problemfelder, die aus der Anwendung von KI hervorgehen können.

II. Zwei Arten von KI-Anwendungen

Wenn derzeit von KI-Systemen gesprochen wird, dann sind zwei Typen von Anwendungen gemeint. Zum einen gibt es Systeme, die mit Prognosen und Klassifizierungen arbeiten. Das Ziel ist dabei, mit Methoden der KI Informationen über ein vorliegendes Phänomen zu erhalten. Die Fragestellungen können dabei ganz unterschiedlich sein: Wie gefährlich ist ein*e Straftäter*in?[4] Ist auf diesem Lungenröntgen eine Covid-Erkrankung sichtbar?[5] Welches Produkt des Online-Shops ist als nächstes interessant?[6] Welche Sprache ist auf dieser Audioaufzeichnung zu hören?[7] Wie wahrscheinlich ist es, dass ein*e Kreditnehmer*in den Kredit

2 Vgl *Skene*, Artificial intelligence and the environmental crisis: can technology really save the world? (2020).

3 Vgl *Noorbakhsh-Sabet/Zand/Zhang/Abedi*, Artificial Intelligence Transforms the Future of Health Care, The American Journal of Medicine 2019, 795.

4 Vgl *Travaini/Pacchioni/Bellumore/Bosia/De Micco*, Machine Learning and Criminal Justice: A Systematic Review of Advanced Methodology for Recidivism Risk Prediction, International Journal of Environmental Research and Public Health 2022, 10594.

5 Vgl *Harmon* et al, Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets, Nature Communications 2020, 4080.

6 Vgl *Park/H. K. Kim/Choi/J. K. Kim*, A literature review and classification of recommender systems research, Expert Systems with Applications 2012, 10059.

7 Vgl *Kotsakis/Matsiola/Kalliris/Dimoulas*, Investigation of Spoken-Language Detection and Classification in Broadcasted Audio Content, Information 2020, 211.