EuZ
ZEITSCHRFIT FÜR EUROPARECHT

AUSGABE:
01|2022

LEITARTIKEL 1:

Angela Müller
„Der Artificial Intelligence Act der EU: Ein risikobasierter Ansatz zur Regulierung von Künstlicher Intelligenz“

LEITARTIKEL 2:

Rolf H. Weber
Künstliche Intelligenz: Regulatorische Überlegungen zum „Wie“ und „Was“

Der Artificial Intelligence Act der EU: Ein risikobasierter Ansatz zur Regulierung von Künstlicher Intelligenz – mit Auswirkungen auf die Schweiz

Angela Müller*

A. Einleitung

Prüfungen der Kreditwürdigkeit durch private Unternehmen, Berechnung der Arbeitsmarktintegrationschancen für Arbeitslose, die Bewertung von Bewerbungsunterlagen im Rekrutierungsprozess, Online-Recherchen mittels Suchmaschinen, die Prognose des Rückfallrisikos bei Inhaftierten oder personalisierte Werbung auf Social Media Plattformen – unsere Lebenswelt ist bereits heute geprägt von algorithmischer Entscheidungsfindung (ADM, für algorithmic decision-making), oft prominent diskutiert unter dem Schlagwort „Künstliche Intelligenz" (KI).[1] Unternehmen, aber zunehmend auch öffentliche Stellen, erkennen grosse Potentiale in der Verwendung solcher Systeme, von der Personalisierung des Angebots von Produkten und Dienstleistungen bis zur Effizienzsteigerung bei der Abwicklung von Massenverwaltungsaufgaben. ADM-Systeme breiten sich weltweit rasant aus. Unsere Gesellschaft ist automatisiert.

Das alles hat auch für das Recht Konsequenzen. Algorithmenbasierte Automatisierungsprozesse scheinen bestehende rechtliche Paradigmen zumindest herauszufordern. Die Interaktion zwischen Mensch und Maschine, die Kon-

---

* Dr. iur. des. Angela Müller leitet das Policy & Advocacy Team bei AlgorithmWatch und ist Senior Policy & Advocacy Managerin bei AlgorithmWatch Schweiz. Sie hat ein Doktorat in Rechtswissenschaft und einen MA in Political and Economic Philosophy. Ihre Dissertation hat sie im Bereich des internationalen Menschenrechtsschutzes an der Universität Zürich verfasst, wo sie auch Mitglied der Digital Society Initiative ist. Sie war Visiting Researcher an der Columbia University, New York, und der Hebrew University, Jerusalem. Zuvor war Angela Müller bei einem Think Tank, einer universitären Innovationsplattform sowie beim Schweizerischen Aussendepartement EDA tätig. Sie engagiert sich zudem als Vize-Präsidentin der Gesellschaft Schweiz-UNO.

[1] Unter „ADM-System" wird ein umfassendes soziotechnologisches System verstanden, von der Entwicklung von Prozessen zur Datenerfassung bis hin zur automatisierten Handlung. Im Folgenden wird generell von „ADM-Systemen" gesprochen, im Zusammenhang mit dem Verordnungsentwurf der EU allerdings von „KI-Systemen", da dies seiner Terminologie entspricht. Für weitere Aspekte bezüglich Begriffsabgrenzung, siehe unten, C.III.

zepte von Verantwortlichkeit, Rechenschaft und Haftung auf den Kopf zu stellen scheint; die technologische Entwicklung und die dadurch bedingte dynamische Natur des Regulierungsobjekts; der virtuelle Raum mit seiner grenzüberschreitenden Natur, der den üblichen räumlichen Geltungsbereich von Rechtsnormen transzendiert; oder die auffallende Intransparenz beim Einsatz von ADM-Systemen, die eine demokratische Kontrolle erschwert – alle diese Aspekte werfen neue Fragen für Forschung, Gesetzgebung und Rechtsprechung auf, die angegangen werden müssen.

Dazu gehören Fragen zu den Risiken, die mit dem Einsatz von ADM-Systemen einhergehen: Wir haben typischerweise nicht nur wenig Einblick in ihre Funktionsweise (*Black Box*-Problematik), sondern auch wenig Informationen dazu, wo, von wem und wozu sie eingesetzt werden. Die Systeme können einerseits durch Verzerrungen (*biases*) in Trainingsdaten oder in ihren Modellen bestehende gesellschaftliche Diskriminierungsmuster übernehmen sowie verstärken, können aber andererseits auch darüber hinaus – also selbst wenn der Einfluss dieser Verzerrungen in Daten und Modell minimiert werden könnte – ungerechtfertige Ungleichbehandlungen hervorrufen und zementieren:[2] Vor dem Hintergrund des gesellschaftlichen Kontextes, in dem sie angewendet werden, können sie beispielsweise mit sich selbst verstärkenden Rückkoppelungsschlaufen einhergehen. Weiter können ADM-Systeme Menschen in ihrer individuellen Selbstbestimmung befördern, indem sie Effizienzgewinn versprechen oder neue Handlungsoptionen eröffnen, können diese aber ebenso einschränken. Durch ihre prägende Rolle in der öffentlichen Sphäre können sie Teilhabemöglichkeiten sowohl erschaffen als auch reduzieren. Letzteres zeigt sich gerade am Beispiel der algorithmischen Steuerung von Online-Plattformen mit grosser – zunehmend auch öffentlich diskutierter – Deutlichkeit.

Vor dem Hintergrund dieser Risiken ist es zentral, Rahmenbedingungen für einen verantwortungsvollen Einsatz von ADM-Systemen zu schaffen. Während diese *Governance* hier in einem umfassenden Sinne verstanden wird und Massnahmen in verschiedensten Bereichen beinhalten muss, wird die rechtliche Regulierung zweifellos eine zentrale Dimension davon darstellen – eine Dimension, die im letzten Jahr auf den politischen Agenden weltweit an Bedeutung gewonnen hat. Am 21. April 2021 hat die Europäische Kommission

[2] Siehe unten, C.IV.3.

ihren Entwurf für eine Verordnung zur Künstlichen Intelligenz[3] – der Draft „Artificial Intelligence Act", im Folgenden der „KI-Verordnungsentwurf" – vorgelegt, bei der es sich um die weltweit erste rechtlich bindende horizontale Regulierung von KI-Systemen handeln würde. Sie hat damit sowohl politisch ein Zeichen gesendet als auch substanziell einen Standard gesetzt.

Der KI-Verordnungsentwurf, der derzeit im Europäischen Parlament und im Rat der EU behandelt wird, soll im Folgenden näher erläutert sowie kritisch beurteilt werden. Anschliessend wird diskutiert, inwiefern die Verordnung für die Schweiz relevant würde und – darüber hinaus – welche Fragen sich für die Schweiz beim Umgang mit ADM-Systemen stellen. Ein Fazit schliesst diese Auslegeordnung ab.

B. Der Entwurf der KI-Verordnung

I. Die Form des Verordnungsentwurfs

Der KI-Verordnungsentwurf ist eines der jüngsten Resultate eines ganzen Pakets von Gesetzesvorschlägen und Initiativen der Europäischen Kommission, vorgelegt im Rahmen der europäischen Digitalstrategie. Nebst dem „Digital Markets Act",[4] dem „Data Governance Act",[5] der Maschinenrichtlinie[6] oder der angekündigten Revision der Produkthaftung in Bezug auf KI[7] gehört dazu auch

---

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 vom 21. April 2021, COM(2021) 206 final.

4 Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates über bestreitbare und faire Märkte im digitalen Sektor (Gesetz über digitale Märkte) vom 15. Dezember 2020, COM(2020) 842 final.

5 Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates über europäische Daten-Governance (Daten-Governance-Gesetz) vom 25. November 2020, COM(2020) 767 final.

6 Richtlinie (EU) des Europäischen Parlaments und des Rates vom 17. Mai 2006 über Maschinen und zur Änderung der Richtlinie 95/16/EG (Neufassung), ABl L 157 vom 9. Juni 2006, 24 ff.

7 Richtlinie 85/374/EWG des Rates vom 25. Juli 1985 zur Angleichung der Rechts- und Verwaltungsvorschriften der Mitgliedstaaten über die Haftung für fehlerhafte Produkte, ABl L 210 vom 7. August 1985, 29 ff.; Informationen zum Revisionsprozess sind abrufbar unter <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12979-Civil-liability-adapting-liability-rules-to-the-digital-age-and-artificial-intelligence_en>.

der „Digital Services Act“,[8] der seinem Geltungsbereich entsprechend ebenfalls zentral sein wird im Umgang mit algorithmischen Entscheiden – nämlich jenen, die von Online-Plattformen vorgenommen werden.

Mit Veröffentlichung des Entwurfs zur KI-Verordnung hat die EU-Kommission weltweit die Debatte rund um die Governance von KI allerdings nochmals deutlich vorangetrieben – auch wenn selbstverständlich in anderen Rechtssystemen, Institutionen und Netzwerken das Thema ebenfalls verfolgt wird. Exemplarisch seien hier beispielsweise die OECD Empfehlungen zu KI[9] oder die Empfehlungen zur Ethik von KI der UNESCO[10] erwähnt. Zudem werden auch im Rahmen des Europarats derzeit Verhandlungen zu einem rechtlich bindenden Instrument zur Regulierung von KI vorbereitet, die voraussichtlich im Mai nächsten Jahres formell aufgenommen werden. Das „Ad Hoc Committee on Artificial Intelligence“ (CAHAI) war über die letzten zwei Jahre damit beauftragt, die Machbarkeit und mögliche Elemente eines rechtlichen Rahmens zu KI innerhalb des Europarats zu untersuchen, um einen Ausgangspunkt für die ministeriellen Verhandlungen zu schaffen.[11] Hier ergeben sich trotz der unterschiedlichen Verankerung der rechtlichen Prozesse interessante Bezugspunkte zur KI-Verordnung der EU.

Dem KI-Verordnungsentwurf der EU ging ein Weissbuch voraus, das unter Mitwirkung einer hochrangigen Expertengruppe verfasst, im Februar 2020 publiziert und anschliessend einer öffentlichen Konsultation unterzogen wurde.[12] Der Verordnungsvorschlag ist ein Instrument des Binnenmarktes: Er stützt sich auf Art. 114 AEUV,[13] der der Union die Kompetenz zur Rechtsanglei-

---

8 Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates über einen Binnenmarkt für digitale Dienste (Gesetz über digitale Dienste) und zur Änderung der Richtlinie 2000/31/EG vom 15. Dezember 2020, COM/2020/825 final; vgl. dazu etwa Buiten Miriam C., Der Digital Services Act (DSA): Vertrautes Haftungsregime, neue Verpflichtungen, EuZ 3/2021, 102 ff.

9 Recommendation of the Council on Artificial Intelligence by the OECD vom 22. Mai 2019, OECD/LEGAL/0449.

10 Recommendation on the Ethics of Artificial Intelligence by UNESCO General Conference vom 24. November 2021, 41 C/73.

11 Die Autorin hat die Diskussionen als Vertreterin von AlgorithmWatch verfolgt, das als zivilgesellschaftliche Organisation Beobachterstatus im CAHAI innehatte.

12 Weissbuch der Europäischen Kommission Zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen vom 19. Februar 2020, COM(2020) 65 final.

13 Konsolidierte Fassung des Vertrags über die Arbeitsweise der Europäischen Union, ABl C 326 vom 26. Oktober 2012, 47 ff.

chung mit Blick auf das Funktionieren des Binnenmarktes verleiht. Diesem Ansatz gemäss ist er entsprechend stark an der Logik von Normen zur Produktsicherheit orientiert, was sich auch in Text und Struktur deutlich niederschlägt.

II. Der Inhalt des Verordnungsentwurfs

1. Geltungsbereich

Als allererstes stellt sich die Frage nach dem sachlichen Geltungsbereich, insbesondere dem Regulierungsobjekt. Um die Verordnung möglichst zukunftstauglich zu halten, wird gemäss der Kommission eine möglichst technologieneutrale Definition gewählt, die sich im Wesentlichen auf die Definition der OECD stützt.[14] Zusätzlich verweist der KI-Verordnungsentwurf jedoch auch auf eine Reihe von technologischen Ansätzen, die einem System zugrunde liegen müssen, damit es von der Verordnung erfasst würde.[15] Diese Liste und damit die Definition scheinen sehr breit gefasst zu sein. Es wird sich jedoch im Detail weisen, ob und inwiefern dies tatsächlich der Fall sein wird.[16]

Gemäss Entwurf wäre die KI-Verordnung *ratione personae* insbesondere auf Anbieter·innen und Nutzer·innen von KI-Systemen anwendbar. Als „Anbieter" werden dabei jene Akteure verstanden, die ein System entwickeln und in Verkehr bringen,[17] während unter „Nutzer" diejenigen Stellen fallen, die ein System unter ihrer Verantwortung verwenden, wobei der persönliche, nichtberufliche Bereich ausgenommen ist.[18] Weitere Bestimmungen gelten für Akteure, die direkt mit obigen in Verbindung stehen, wie „Bevollmächtigte", „Importeure" oder „Händler". Konsument·innen, Endnutzer·innen und andere von Ergebnissen der Systeme betroffene natürliche oder juristische Personen sind *ratione personae* nicht erfasst und ihnen würden durch die Verordnung weder direkte Rechte noch Pflichten zugeschrieben.

---

14 Begründung 5.2.1; Art. 3(1) COM(2021) 206 final; vgl. OECD/LEGAL/0449.

15 „(a) Konzepte des maschinellen Lernens, mit beaufsichtigtem, unbeaufsichtigtem und bestärkendem Lernen unter Verwendung einer breiten Palette von Methoden, einschließlich des tiefen Lernens (Deep Learning); (b) Logik- und wissensgestützte Konzepte, einschließlich Wissensrepräsentation, induktiver (logischer) Programmierung, Wissensgrundlagen, Inferenz- und Deduktionsmaschinen, (symbolischer) Schlussfolgerungs- und Expertensysteme; (c) Statistische Ansätze, Bayessche Schätz-, Such- und Optimierungsmethoden" Anhang I COM(2021) 206 final.

16 Siehe unten, C.III.

17 Art. 3(2) COM(2021) 206 final.

18 Art. 3(4) COM(2021) 206 final.

Der räumliche Geltungsbereich des Verordnungsentwurfes wird in Art. 2(1) festgelegt. Ihre Geltung wäre *ratione loci* demnach gegeben, wenn (i) ein KI-System innerhalb der EU zur Anwendung kommt oder (ii) wenn „das vom System hervorgebrachte Ergebnis" innerhalb der EU verwendet wird[19] – und bliebe damit unabhängig davon, wo sich Anbieter·innen oder Nutzer·innen der Systeme befinden. In anderen Worten: Auch in Drittstaaten können sie von der Reichweite der Verordnung erfasst werden. Dieser Ansatz der extraterritorialen Gesetzgebung folgt der Logik der „territorialen Extension", wie man sie aus anderen Kontexten, etwa im Bereich des Wettbewerbsrechts, des Umweltschutzes oder der Privatsphäre von der EU kennt – prominent beispielsweise von der 2018 in Kraft getretenen Datenschutzgrundverordnung.[20] Die EU macht dabei gesetzliche Regelungen im Ausland anwendbar, wobei jedoch ein territorialer Bezugspunkt gegeben ist.[21]

Für Drittstaaten – insbesondere jene mit engem Bezug zum EU-Binnenmarkt, wie etwa die Schweiz – kann dies massive Auswirkungen haben: Auch Anbieter·innen und Nutzer·innen in der Schweiz würden in den oben genannten Fällen direkt der KI-Verordnung unterworfen werden. Aufgrund der Bedeutung des EU-Binnenmarktes wird dies voraussichtlich ein beachtlicher Teil der KI-Anbieter·innen in der Schweiz betreffen.

2. Der risikobasierte Ansatz

Zur Regulierung von KI-Systemen wählt die EU-Kommission einen *risikobasierten Ansatz*. Das heisst, sie klassifiziert KI-Systeme entsprechend dem Grad ihrer Risiken für Gesundheit, Sicherheit und Grundrechte, mit denen sie einhergehen. Es ergeben sich vier Risikokategorien: KI-Systeme mit unzulässigem Risiko; hohem Risiko; begrenztem Risiko; und minimalem Risiko.[22]

---

19 Art. 2(1) COM(2021) 206 final.

20 Z.B. Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 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 119 vom 4. Mai 2016, 1-88; vgl. EuGH, Urteil vom 3. Oktober 2019 in der Rechtssache C-18/19, ECLI:EU:C:2019:8216 – Eva Glawischnig-Piesczek v. Facebook, Rz. 50 ff.

21 Z.B. Scott Joanne, Extraterritoriality and Territorial Extension in EU Law, AJCL 2014, 90, 94 ff.; vgl. dazu auch Bradford Anu, Brussels Effect: How the European Union Rules the World, New York 2020

22 Die Terminologie wird erläutert unter <https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/excellence-trust-artificial-intelligence_de>.

Als erstes werden in Art. 5 jene vier Praktiken erwähnt, die mit *unzulässigen Risiken* verbunden sind und entsprechend verboten werden sollen. Politisch wie medial derzeit im Scheinwerferlicht steht das Verbot der Verwendung von biometrischen Identifizierungssystemen zu Strafverfolgungszwecken im öffentlich zugänglichen Raum, bei denen die Identifikation in Echtzeit und aus der Ferne geschieht.[23] Gleichzeitig wäre dieses Verbot mit einer Reihe von Ausnahmen verbunden, wenn nämlich die Verwendung der Systeme „unbedingt erforderlich" ist für bestimmte festgesetzte Ziele, wie beispielsweise in Verbindung mit Straftaten, die einen bestimmten Schweregrad erreichen.[24] Es gilt jedoch das Erfordernis der vorgängigen Genehmigung – von der allerdings in dringenden Fällen vorerst abgesehen werden kann –, wobei Notwendigkeit und Verhältnismässigkeit gegeben sein müssen.[25] Mitgliedstaaten können schliesslich die Erlaubnis zur eigentlich verbotenen Anwendung biometrischer Identifizierungssysteme im nationalen Recht festschreiben, wenn sie dabei die erwähnten Grenzen beachten.[26]

Bei der biometrischen Identifikation fällt nur die entsprechende *Verwendung* solcher Systeme unter den vom Verbot erfassten sachlichen Geltungsbereich.[27] Anders ist dies bei den anderen drei mit unzulässigem Risiko klassifizierten Praktiken, bei denen jeweils „das Inverkehrbringen, die Inbetriebnahme oder die Verwendung" der entsprechenden Systeme verboten werden soll. Dies umfasst einerseits KI-Systeme, die Personen durch unterschwellige Beeinflussung manipulieren[28] oder bestimmte Vulnerabilitäten ausnutzen,[29] und zwar in einer Weise, die den Betroffenen physischen oder psychischen Schaden zufügen kann, sowie andererseits die Verwendung von KI-basierten sogenannten „Social Scoring" Systemen durch öffentliche Behörden.[30]

Die zweite Kategorie umfasst *Hochrisiko-KI-Systeme*.[31] Dazu gehören einerseits all jene Systeme, die ein von bestimmten anderen EU-Verordnungen er-

---

23 Art. 5(1)d COM(2021) 206 final.

24 Art. 5(1)d(i)-(iii) COM(2021) 206 final.

25 Art. 5(3) COM(2021) 206 final.

26 Art. 5(4) COM(2021) 206 final.

27 Zur Würdigung dieses Aspektes, siehe unten, C.IV.2.

28 Art. 5(1)a COM(2021) 206 final.

29 Art. 5(1)b COM(2021) 206 final.

30 D.h. Systeme „zur Bewertung oder Klassifizierung der Vertrauenswürdigkeit natürlicher Personen über einen bestimmten Zeitraum auf der Grundlage ihres sozialen Verhaltens oder bekannter oder vorhergesagter persönlicher Eigenschaften oder Persönlichkeitsmerkmale", wenn dies zu unverhältnismässiger, ungerechtfertigter und/oder kontextfremder sozialer Benachteiligung führt, Art. 5(1)c COM(2021) 206 final.

31 Art. 6 ff. COM(2021) 206 final.

fasstes Produkt sind oder als Sicherheitskomponente in einem solchen Produkt eingesetzt werden, wenn dieses einer Konformitätsbewertung durch Dritte unterworfen ist.[32] Andererseits fallen bestimmte „stand-alone“ KI-Systeme aus den Bereichen biometrische Identifizierung und Kategorisierung, kritische Infrastrukturen, Bildung, Arbeitsplatz, Zugang zu Dienstleistungen, Strafverfolgung, Migration und Asyl sowie Rechtspflege und demokratische Prozesse in diese Kategorie.[33]

Bei der Hochrisiko-Kategorie handelt es sich um das Kernstück des Verordnungsentwurfs. Er sieht für die entsprechenden Systeme bestimmte Anforderungen vor, die im Rahmen eines Risikomanagementsystems überwacht werden müssen, darunter zu Datenqualität und -governance, technischer Dokumentation und Aufzeichnung, Bereitstellung von Informationen, menschlicher Aufsicht sowie zu Genauigkeit, Robustheit und Sicherheit.[34] Gemäss dem Ansatz des *New Legislative Framework*, auf dem der Verordnungsentwurf basiert, werden die Anbieter·innen von Hochrisiko-Systemen verpflichtet, im Rahmen einer Konformitätsbewertung selbst zu prüfen, dass die Systeme die erwähnten Anforderungen erfüllen, bevor sie diese auf den Markt bringen oder sie in Betrieb genommen werden.[35] Wenn die Konformitätsbewertung erfüllt ist, wird dies mit der CE-Konformitätskennzeichnung ausgezeichnet,[36] was Zugang zu und freie Bewegung auf dem EU-Binnenmarkt sicherstellt. Zudem muss das System von den Anbieter·innen in einer EU-weiten Datenbank eingetragen werden.[37] Falls es zu einem späteren Zeitpunkt substanzielle Änderungen erfährt, muss die Bewertung erneut vorgenommen werden. Weiter unterliegen Anbieter·innen auch Pflichten zur Überwachung eines Systems nach seinem Inverkehrbringen.[38] Für Nutzer·innen, die KI-Systeme einsetzen, bestehen ebenfalls Pflichten, wenn auch deutlich weniger weitgehende. Dazu gehört, das System gemäss Instruktion zu verwenden, menschliche Aufsicht sicherzustellen oder Risiken kontinuierlich zu überwachen.[39]

---

32 Art. 6(1) COM(2021) 206 final.

33 Art. 6 Anhang III COM(2021) 206 final.

34 Art. 8 ff. COM(2021) 206 final.

35 Für Systeme, die Sicherheitskomponenten von Produkten sind, die bereits durch entsprechende sektorielle Verordnungen reguliert und darin einer Konformitätsbewertung unterworfen sind, ist vorgesehen, dass diese Konformitätsbewertung der KI-Verordnung in die bereits bestehende Konformitätsbewertung integriert wird, Art. 43(3) COM(2021) 206 final.

36 Art. 19, 43, 48, 49 COM(2021) 206 final.

37 Art. 51, 60 COM(2021) 206 final.

38 Art. 61 f. COM(2021) 206 final.

39 Art. 29 COM(2021) 206 final.

Die dritte Kategorie des „begrenzten Risikos“ umfasst gemäss Verordnungsentwurf KI-Systeme, die mit natürlichen Personen interagieren (wie etwa Chatbots). Sie werden Transparenzpflichten unterworfen: Den Personen muss mitgeteilt werden, dass sie mit einem KI-System interagieren, wenn dies nicht aus dem Kontext offensichtlich wird. Eine Ausnahme gilt für Strafverfolgungszwecke. Spezifisch werden Informationspflichten für Systeme zur Emotionserkennung und biometrischen Kategorisierung sowie für „Deepfakes“, also manipulierte Bild-, Ton- oder Videoinhalte, erwähnt.[40]

Als letzte Kategorie ergeben sich sodann implizit KI-Systeme mit „minimalem Risiko“, die vom Geltungsbereich der Verordnung ausgenommen wären. Gleichzeitig sollen Verhaltenskodizes gefördert werden, unter denen sich Anbieter·innen selbst verpflichten, die an Hochrisiko-Systeme gerichteten Anforderungen sowie weitere Anforderungen zu erfüllen.[41]

3. Gouvernanz und Durchsetzung

Bezüglich der Gouvernanzstruktur ist wichtig zu betonen, dass es sich bei der Konformitätsbewertung für Hochrisiko-Systeme um eine Selbsteinschätzung der Anbieter·innen handelt. Zudem ist vorgesehen, dass die Kommission Standardisierungsorganisationen damit beauftragen kann, harmonisierte Normen zur Erfüllung der erwähnten Anforderungen für Hochrisiko-KI-Systeme zu erstellen. Wenden Anbieter·innen harmonisierte Standards an, wird eine Konformität mit den Anforderungen der Verordnung vermutet.[42]

Auf Mitgliedstaatsebene werden notifizierende Behörden geschaffen, die Konformitätsbewertungsstellen benennen, notifizieren und überwachen. Sobald letztere – es handelt sich dabei um von den Anbieter·innen unabhängige Akteure, d.h. klassischerweise um private Zertifizierungsunternehmen – notifiziert sind, sind sie berechtigt, Konformitätsbewertungsverfahren durchzuführen. Gleichzeitig bleibt die Bedeutung dieser notifizierten Stellen im Verordnungsentwurf gering: Ihr Einbezug ist nur für einen Hochrisikobereich vorgesehen, nämlich für Systeme zur biometrischen Identifizierung und Ka-

40 Art. 52 COM(2021) 206 final.

41 Art. 69 COM(2021) 206 final.

42 Art. 40 ff. COM(2021) 206 final.

tegorisierung, und kann auch da ausgelassen werden, wenn harmonisierte Normen befolgt werden. Für alle anderen Hochrisikobereiche sind notifizierte Stellen nicht vorgesehen.[43]

Als Gouvernanzstruktur ist auf europäischer Ebene vorgesehen, einen „Europäischen Ausschuss für künstliche Intelligenz“ (EAKI) zu schaffen, der die Kommission in der Umsetzung der Verordnung unterstützt.[44] Auf nationaler Ebene werden zuständige Behörden geschaffen oder benannt, darunter die nationale Aufsichtsbehörde. Diese agiert auch als Marktüberwachungsbehörde und ist Teil des EAKI. Als Marktüberwachungsbehörde erhält sie Zugang zu Daten und Dokumentationen der Anbieter·innen, kann auf Verdacht eines bestehenden Risikos hin das System und seine Konformität mit der Verordnung überprüfen, Korrekturmassnahmen anordnen und – bei deren Nichtbefolgung – die Verwendung eines Systems einschränken oder unterbinden.[45]

Der Verordnungsentwurf enthält zudem Massnahmen zur Innovationsförderung, inklusive der Möglichkeit zur Schaffung von KI-Reallaboren („Sandboxes“) oder der Förderung von kleineren Unternehmen und Start-ups.[46]

Zur Durchsetzung der Verordnung können gemäss Entwurf Sanktionen ergriffen werden. Bei Verstössen gegen die in Art. 5 gelisteten Verbote oder gegen Anforderungen der Hochrisiko-Systeme bezüglich Datengouvernanz sind Geldbussen von bis zu 30 Millionen Euro oder 6% des weltweiten Jahresumsatzes – je nachdem, welche Summe höher ist – vorgesehen. Bei weiteren Verstössen sind abgestuft geringere, aber unter Umständen ebenfalls substanzielle Geldstrafen vorgesehen.[47]

---

43 Art. 40 ff., Art. 43(1) und (2) COM(2021) 206 final; vgl. Veale Michael/Borgesius Frederik Zuiderveen, Demystifying the Draft EU Artificial Intelligence Act: Analysing the good, the bad, and the unclear elements of the proposed approach, Computer Law Review International 2021, Rz. 49 ff., 58 ff.

44 Art. 56 ff. COM(2021) 206 final.

45 Art. 59, 63 ff. COM(2021) 206 final.

46 Art. 53 ff. COM(2021) 206 final.

47 Art. 71 COM(2021) 206 final.

C. Würdigung

I. Vorbemerkungen

Der Vorschlag der Kommission rückt die Thematik zweifellos auf den politischen Agenden weltweit um einige Positionen nach oben und entfacht damit die Debatte zur Governance von KI-Systemen auch ausserhalb des engsten Expert·innenkreises. Zudem setzt sie auch substanziell einen Standard, der die Debatten massgeblich prägen wird: Entscheidungsträger·innen aus Drittstaaten oder internationalen Organisationen werden sich am von der EU gesetzten Standard als Ausgangspunkt orientieren beziehungsweise sich rechtfertigen, wenn sie von diesem abweichen. Dies zeigt sich beispielhaft bereits jetzt in den entsprechenden Prozessen im Europarat, ist aber auch in der Schweiz zu erwarten.

Im Folgenden soll eine Würdigung einzelner Aspekte des KI-Verordnungsentwurfs vorgenommen werden, ohne dass diese Anspruch auf Vollständigkeit erhebt. Sie gibt die Position der Autorin wieder, die von ihr aber gleichzeitig auch als politische Forderungen im Rahmen ihrer Tätigkeit bei AlgorithmWatch, einer zivilgesellschaftlichen Forschungs- und Advocacy-Organisation mit Fokus auf den Einsatz von ADM-Systemen, vertreten werden.[48]

Auch vor diesem Hintergrund scheint es wichtig, vorab transparent zu machen, auf welchen Prinzipien die folgende Würdigung basiert. Trotz der rechtlichen Herausforderungen, die der Einsatz von ADM-Systemen mit sich bringt, soll an dieser Stelle deutlich gemacht werden, dass auch er sich an grundsätzlichen Werten orientieren muss, auf die wir uns als demokratische Gesellschaften geeinigt haben: An individueller Autonomie und Freiheit, Gerechtigkeit, Teilhabe und Gemeinwohl, auf Normebene insbesondere reflektiert in Grundrechten und weiteren basalen – oft auf Verfassungsebene geschützten – demokratischen und rechtsstaatlichen Prinzipien. Ziel muss es sein, dass dieser Einsatz tatsächlich Individuen und Gesellschaft einen Nutzen bringt – dass er also tatsächlich individuelle Autonomie erhöht, statt sie zu reduzieren; dass er Teilnahmemöglichkeiten eröffnet oder befördert, statt sie einzuschränken;

---

[48] Zur der folgenden Würdigung zugrundeliegenden Position zum KI-Verordnungsentwurf siehe AlgorithmWatch, Submission to the European Commission's Consultation on the Draft Artificial Intelligence Act, 2021, abrufbar unter <https://algorithmwatch.org/de/eu-ki-verordnung-einreichung-2021/>; EDRi et al., An EU Artificial Intelligence Act for Fundamental Rights – A Civil Society Statement, 2021, abrufbar unter <https://algorithmwatch.org/en/eu-artificial-intelligence-act-for-fundamental-rights/>.

und dass dieser Nutzen gerecht verteilt wird. Dass dies auch in Europa alles andere als Selbstverständlichkeiten sind, zeigt sich in Dokumentationen zum Einsatz von ADM-Systemen.[49]

II. Der Regulierungsansatz

Der Ansatz der Verordnung als Instrument der Binnenmarktregulierung könnte bereits als Quelle einer gewissen Spannung gedeutet werden: Das verfolgte Ziel, der Schutz von Unionswerten, Grundrechten, Gesundheit und Sicherheit, soll mit dem Mittel der Harmonisierung des digitalen Binnenmarkts erreicht werden – der gleichzeitig auch Rechtssicherheit gewährleisten und Innovation ermöglichen soll.[50]

Erstens muss es bei jeglichen Regulierungsansätzen im Bereich der neuen Technologien genau um dieses Anliegen gehen, nämlich ihren Einsatz sowie die Innovationsförderung mit Grundrechtsschutz in Einklang zu bringen. Gleichzeitig zeigt sich in der Praxis das Ausmass dieses Spannungsverhältnisses deutlich. Die KI-Verordnung versucht, mittels eines Binnenmarktinstruments Grundrechte zu schützen; ob dies aus einer Grundrechtslogik heraus überhaupt ein effektives Regulierungsmittel darstellen kann, kann zumindest hinterfragt werden. Zweitens orientiert sich der Ansatz an jenen zur Regulierung der Produktsicherheit und setzt somit stark auf technische Lösungen, Standardisierungsprozesse und Selbsteinschätzungen. Ob diese in Verbindung mit KI-Systemen, die wesentliche grundrechtliche Problemdimensionen mit sich bringen können, effektiven Schutz zu leisten vermögen, wird sich weisen. Drittens machen der gewählte Regulierungsansatz sowie die Komplexität des Vorschlages diesen für Expert·innen und zivilgesellschaftliche Organisationen, die im Bereich von KI und/oder des Grundrechtsschutzes tätig sind (aber nicht besondere Expertise im Bereich der Produktsicherheit oder der Binnenmarktregulierung aufweisen), wenig zugänglich. Dies kann zur Folge haben, dass die Inklusion wichtiger Stimmen untergraben wird – im Gesetzgebungsverfahren, aber darüber hinaus beispielsweise auch in den für diesen Regulierungsansatz wichtigen Standardisierungsprozessen, denen ja der Entwurf ein grosses Gewicht zuschreibt, indem die Verwendung von Standards eine Konformitätsvermutung zur Folge haben kann. Standardisierungsprozesse sind aus formalen und praktischen Gründen nicht einfach zugänglich für die Zivilgesellschaft, sondern geprägt von Stimmen aus dem Privatsektor.[51] Gerade

---

49 Chiusi Fabio et al. (Hrsg.), Automating Society Report, Berlin/Gütersloh 2020, 5 ff., abrufbar unter <https://automatingsociety.algorithmwatch.org/>.

50 Begründung, Kap. 1.1; Erwägung 1 COM(2021) 206 final.

51 Dazu Veale/Borgesius, Rz. 50 ff.

wenn es um Standards geht, die schlussendlich einen Beitrag zum Ziel des Grundrechtsschutzes leisten sollen, ist es zentral, dass nicht nur technische Expertise, sondern auch Expertise im Bereich der Grundrechte einbezogen wird.

Dies ist ein Aspekt, der insbesondere problematisch erscheint, da er das Ungleichgewicht der Einflussmöglichkeiten zugunsten von grossen Technologieunternehmen verschärft: Diese haben unbestritten die Ressourcen, ganze Teams von Jurist·innen zu beschäftigen und sich so bis in die rechtlichen Detailfragen durch engagiertes Lobbying oder in Standardisierungsprozesse einzubringen, was vielen zivilgesellschaftlichen und wissenschaftlichen Anliegen verwehrt bleiben wird.

III. Der Geltungsbereich

Die Frage nach dem sachlichen Geltungsbereich des KI-Verordnungsentwurfs und der Eignung der gewählten Definition ist verbunden mit grundsätzlichen, über den Entwurf herausgehenden Fragen zum Begriff „Künstliche Intelligenz", der sich in der öffentlich-politischen Debatte zum Schlagwort entwickelt hat, dessen Unschärfe jedoch für die rechtliche Regulierung Schwierigkeiten mit sich bringt. Wenn es um grundrechtliche Risiken geht, wäre es demnach angemessener, den Fokus auf jene Automatisierungsprozesse zu legen, die in ihren Auswirkungen auf Individuen und Gesellschaft Risikosignale aufweisen. In anderen Worten: Das Kriterium zur Bestimmung des sachlichen Geltungsbereichs sollte in diesen Auswirkungen liegen – und nicht in der technologischen Art und Weise, wie diese Auswirkungen zustande gekommen sind. Aus dieser Perspektive bietet der Begriff der algorithmischen Entscheidungsfindung (ADM) Vorteile gegenüber dem unscharfen Konzept der „KI". ADM-Systeme werden nach diesem Verständnis umfassend als soziotechnologische Systeme betrachtet, wobei der Fokus auf den Entscheidungen der Systeme und ihrer Relevanz für Individuum und Gesellschaft liegt. Dies berücksichtigt zwangsläufig den gesellschaftlichen Kontext mit, in dem Systeme eingesetzt werden, was für die Risikobewertung unerlässlich ist.[52] Eine Abstützung auf einer expliziten Liste verschiedener Technologien, wie sie der KI-Verordnungsentwurf vorschlägt – auch wenn diese mittels delegierten Rechtsakten von der Kommission angepasst werden kann – scheint darüber hinaus auch der Eindeutigkeit des sachlichen Geltungsbereiches nicht dien-

[52] Siehe oben, A.

lich.[53] Es ist absehbar, dass der Nachweis der Definitionselemente in der Praxis Schwierigkeiten – und mögliche Schlupflöcher – mit sich bringt. Nichtsdestotrotz ist grundsätzlich die breite Auffassung des Begriffs „KI“ und damit des sachlichen Geltungsbereichs der Verordnung zu begrüssen. Es zeichnet sich jedoch ab, dass diese Breite in den nun laufenden Verhandlungen umstritten sein wird, so dass davon auszugehen ist, dass der Geltungsbereich in dieser Hinsicht eher eingeschränkt wird.

Hinsichtlich des Geltungsbereichs *ratione loci* wird deutlich, dass die oben beschriebene extraterritoriale Wirkung der Verordnung wesentliche Nebeneffekte mit sich bringen würde: Erstens ist damit zu rechnen, dass viele Anbieter·innen und Nutzer·innen, die sowohl auf dem schweizerischen als auch dem europäischen Markt tätig sind, ihre KI-Systeme der KI-Verordnung konform ausgestalten werden, auch wenn diese in der Schweiz zur Anwendung kommen – ganz einfach, da es unmöglich oder mit einem unverhältnismässigen Aufwand verbunden wäre, zwischen Anwendungen in der EU und der Schweiz zu differenzieren. Zweitens ist nicht auszuschliessen, dass private Anbieter·innen Druck auf den schweizerischen Gesetzgeber ausüben werden, die Rechtslage der EU-Regelung konform und äquivalent auszugestalten. All dies hat, drittens, zur Folge, dass das Inkrafttreten der KI-Verordnung in der EU auch für die Schweizer Bevölkerung Nebeneffekte mit sich bringen könnte: Falls die KI-Verordnung sich tatsächlich als imstande erweist, grundrechtlichen Schutz zu gewährleisten, werden auch Endnutzer·innen in der Schweiz teilweise in den Genuss dieses Schutzes kommen.

Aus einer grundrechtlichen Perspektive ist prinzipiell zu begrüssen, dass mit diesem Fokus auf den Ort der Anwendung von Systemen eine kohärente Regulierung ermöglicht wird, die einerseits auf dem EU-Territorium Schlupflöcher vermeidet und andererseits in Drittstaaten das Schutzniveau mitprägt. Gleichzeitig hat dieser Fokus gewisse andere Schlupflöcher zur Folge, wie sich in Verbindung mit den im Verordnungsentwurf vorgesehenen Verboten zeigen wird.

---

[53] Vgl. dazu etwa Townsend Bev, Decoding the Proposed European Union Artificial Intelligence Act, insights 2021, 3, abrufbar unter <https://www.asil.org/insights/volume/25/issue/20>; Smuha Nathalie et al., How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act, 2021, 14 f., abrufbar unter <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3899991>.

IV. Der risikobasierte Ansatz und die vier Risikokategorien

1. Die Kategorisierung anhand Risikolevels

Die Anerkennung, dass der Einsatz von KI-Systemen potenziell grossen Nutzen mit sich bringt, aber gleichzeitig mit Risiken verbunden sein kann, ist grundsätzlich sehr begrüssenswert. Gleichzeitig kann mit guten Gründen hinterfragt werden, ob eine *ex ante* Kategorisierung von Systemen in bestimmte Risikoklassen diesem Umstand gebührend Rechnung trägt. Die Risiken, die mit einem System einhergehen, hängen nämlich wesentlich ab von seiner Verwendung zu einem bestimmten Zweck und in einem bestimmten Kontext. Die Idee, KI-Systeme *a priori* einer Risikokategorie zuzuordnen ist insofern problematisch, als dadurch genau diese Faktoren – Zweck und Kontext der Anwendung eines Systems – nicht im Einzelfall berücksichtigt werden. Erst eine individuelle Folgenabschätzung kann überzeugend aufzeigen, mit welchen (und welchem Grad an) potenziellen Risiken ein bestimmter Einsatz verbunden ist. Um die Praktikabilität eines solchen Ansatzes zu gewährleisten, bietet sich ein zweistufiges Verfahren an: In einem ersten Schritt soll mit verhältnismässig geringem Aufwand eine Triage erfolgen, die es erlaubt, Risikosignale zu erkennen. Je mehr Risikosignale bezüglich des Einsatzes eines Systems erkannt werden, desto umfassender werden die Transparenzpflichten, denen die entsprechenden Akteure unterworfen werden.[54]

Eine weitere Folge der KI-Verordnung und ihres risikobasierten Ansatzes wäre, dass, indem sie gewisse Systeme als Hochrisiko definiert, sie diese zwar Transparenzpflichten unterwirft, sie aber gleichzeitig auch legalisiert und legitimiert. Viele dieser Systeme und ihre Verwendungen waren aber noch nicht im eigentlichen Sinne Objekt einer gesellschaftlichen Debatte. In diesem Sinne greift Anhang III des Verordnungsentwurfs der demokratischen Auseinandersetzung dazu, ob und bis zu welchem Grad wir die Verwendung von KI-basierten Systemen in bestimmten Kontexten überhaupt zulassen wollen, vor.

Gleichzeitig ist davon auszugehen, dass die Risikoklassifizierung anhand der vier Stufen in der endgültigen Fassung der Verordnung so oder ähnlich enthalten bleiben wird. Vor diesem Hintergrund müsste zumindest sichergestellt werden, dass für alle Kategorien konsistente Aktualisierungsmechanismen

---

[54] Loi Michele et al., Automated Decision-Making Systems in the Public Sector: An Impact Assessment Tool for Public Authorities, Berlin/Zürich 2021, abrufbar unter <https://algorithmwatch.ch/de/adms-impact-assessment-public-sector-algorithmwatch/>. Gerade im öffentlichen Sektor sollte gemäss diesem Ansatz eine Folgenabschätzung für jedes eingesetzte ADM-System zwingend vorgenommen werden.

vorhanden sind, um die Liste der jeweiligen Systeme um neue Fälle zu ergänzen. Bisher ist nur eine Aktualisierung der Liste der Hochrisiko-Systeme vorgesehen, die die Kommission mittels delegierter Rechtsakte vornehmen kann – und auch da betrifft dies nur die Liste der einzelnen Systeme, nicht aber die Liste der acht Hochrisikobereiche. Für das Update dieser Hochrisikobereiche sowie aller anderen Kategorien (Systeme mit unakzeptablem Risiko; Systeme mit begrenztem Risiko) ist im Verordnungsentwurf selbst bisher kein Aktualisierungsmechanismus vorgesehen, so dass eine Aktualisierung nur durch eine Änderung der Verordnung an sich vorgenommen werden könnte. Dies widerspricht zudem dem Anspruch der Verordnung, zukunftstauglich gestaltet zu sein.

2. Verbotene KI-Systeme

Die im Verordnungsentwurf vorgesehenen Verbote werfen eine Reihe von Fragen auf. Es bleibt beispielsweise offen, wie das Erfordernis des „psychischen oder physischen Schadens" in Art. 5(1)a und 5(1)b aus einer Grundrechtsperspektive gerechtfertigt werden kann. Aus dieser Perspektive entscheidend ist die Tatsache des *Grundrechtseingriffs* und nicht der sich materialisierende *Schaden*. Ähnlich fragwürdig scheint aus dieser Perspektive, dass das Verbot von manipulativen KI-Systemen in Art. 5(1)b begrenzt bleibt auf Systeme, die Personen aufgrund ihres Alters oder ihrer körperlichen oder geistigen Beeinträchtigung ausnutzen, und damit weitere Dimensionen von „Vulnerabilität" unberücksichtigt bleiben.

Schliesslich bleibt unklar, als wie effektiv sich das Verbot zur Verwendung biometrischer Erkennungssysteme zu Strafverfolgungszwecken in Art. 5(1)d in der Praxis erweisen wird. Die Notwendigkeit, beim Einsatz solcher Systeme im öffentlich zugänglichen Raum rote Linien zu ziehen, ergibt sich, da sie eine Massenüberwachung oder diskriminierend gezielte Überwachung ermöglichen, die mit der Idee einer demokratisch organisierten Gesellschaft in Widerspruch stehen: Wenn Personen im öffentlichen Raum erfasst, identifiziert und verfolgt werden können, verletzt dies nicht nur an sich Grundrechte wie das Recht auf Privatsphäre, inklusive der informationellen Selbstbestimmung,[55] sondern wird auch abschreckende Effekte („chilling effects"[56]) haben auf das Wahrnehmen der Rechte auf Meinungsäusserungs- oder Versammlungsfreiheit.

---

[55] Vgl. Stellungnahme des Bundesrates vom 11. August 2021 zu Interpellation 21.3580 vom 5. Mai 2021, Glättli Balthasar, abrufbar unter <https://www.parlament.ch/de/ratsbetrieb/suche-curia-vista/geschaeft?AffairId=20213580>.

[56] EGMR, Entscheidung vom 16. Dezember 1982 in der Rechtssache 9228/80, Rz. 1.

Die Begrenzung von Art. 5(1)d auf Strafverfolgungszwecke einerseits und Fern- und Echtzeitsysteme[57] andererseits sowie die weitgehenden Ausnahmeregelungen werfen zumindest Fragen auf, ob dieses Verbot alle Formen von grundrechtsinkompatibler Massenüberwachung verhindern kann. Inwiefern beispielsweise eine nachträgliche Identifizierung von Demonstrationsteilnehmenden mittels Gesichtserkennungstechnologie und auf Basis vorhandener Videodaten aus grundrechtlicher Perspektive weniger problematisch sein soll als ein Echtzeiteinsatz, bleibt unklar. Mit Blick auf aktuelle Rechtsentwicklungen fragwürdig scheint zudem der Fokus dieses Verbots auf die *Verwendung* der Systeme, wo sich eine Diskrepanz zu den weiteren drei verbotenen Systemen in Art. 5(1)a-c zeigt. Während diese Beschränkung in Art. 5(1)d mit Blick auf die Liste der Ausnahmen und Erfordernisse in Art. 5(1)d(i-iii) und 5(2-4) angezeigt ist, kann gleichzeitig kritisiert werden, dass nicht in einem zusätzlichen Unterabsatz auch das *Inverkehrbringen* und die *Inbetriebnahme* dem Verbot unterworfen werden. Damit werden Anbieter·innen von biometrischen Erkennungssystemen insofern vom Verbot ausgenommen, als diesen den Raum gelassen wird, ihre Systeme weiterhin zu entwickeln, zu verkaufen und zu exportieren. Wiederum zeigt sich ein gewisses Spannungsverhältnis zu einem konsistenten Grundrechtsschutz, da so der Einsatz von in der EU entwickelten biometrischen Erkennungssystemen zu Strafverfolgungszwecken in Drittstatten – wie er innerhalb der EU verboten wäre – weiterhin möglich bleibt. Hier droht ein extraterritoriales grundrechtliches Schutzvakuum.

3. KI-Systeme mit hohem Risiko

Die Pflichten, denen Hochrisiko-Systeme unterworfen würden, sind in dem Sinne begrüssenswert, da sie Transparenz schaffen, was im Lichte der eingangs erwähnten *Black Box*-Problematik zentral erscheint.[58] Lücken zeigen sich jedoch bei den Pflichten, denen die *Nutzer·innen* der Systeme unterworfen würden. Wie oben ausgeführt ist es nicht in erster Linie eine spezifische Technologie, die mit Risiken einhergeht, sondern die Verwendung dieser Technologie in einem bestimmten Kontext und zu einem bestimmten Zweck. Diese Faktoren hängen jedoch in erster Linie von den Unternehmen oder Be-

---

57 So ist es zum Beispiel fragwürdig, inwiefern eine nachträgliche Identifizierung von Personen anhand biometrischer Erkennungssysteme (z.B. basierend auf vorhandenen Videoaufnahmen) aus grundrechtlicher Perspektive weniger problematisch sein sollte. Vgl. zur Kritik etwa AccessNow, Submission to the European Commission's adoption consultation on the Artificial Intelligence Act, 2021, abrufbar unter <https://www.accessnow.org/cms/assets/uploads/2021/08/Submission-to-the-European-Commissions-Consultation-on-the-Artificial-Intelligence-Act.pdf>.

58 Siehe oben, A.

hörden ab, die diese einsetzen – dem „Nutzer“, nach der Terminologie des KI-Verordnungsentwurfs. Um diesem Umstand Rechnung zu tragen, wäre es demnach notwendig, diese Nutzer·innen stärker in die Pflicht zu nehmen: einerseits durch eine verpflichtende Folgenabschätzung, die die Risiken des Einsatzes eines Systems beurteilt, und andererseits, indem Transparenzanforderungen auf Nutzer·innen ausgeweitet würden, inklusive einer Nachweispflicht zum Einsatz eines KI-Systems in der EU-weiten Datenbank.

Weiter bietet sich eine punktuelle Ausweitung der im KI-Verordnungsentwurf vorgesehenen Transparenzanforderungen an, beispielsweise aus Nachhaltigkeitsperspektive in Hinblick auf den Ressourcenverbrauch von KI-Systemen – unabhängig von ihrem Risikolevel.

Gleichzeitig ist wichtig zu betonen, dass Transparenz notwendige, aber nie hinreichende Bedingung sein kann auf dem Weg zu einem verantwortungsvollen Einsatz von KI-Systemen. Um das zentrale Ziel des Grundrechtsschutzes zu erreichen, braucht es zusätzlich Mechanismen, anhand derer Diskriminierungsfreiheit und andere Dimensionen von Gerechtigkeit sichergestellt und Verantwortung und Rechenschaftspflichten zugeschrieben werden können. Ob dies im Rahmen dieser KI-Verordnung überhaupt sinnvoll gemacht werden kann oder ob es dafür ergänzende Rechtsakte, möglicherweise in anderen Rechtsbereichen, benötigt, kann diskutiert werden. Eine wichtige, wenn auch nicht ausreichende[59], Ergänzungsmöglichkeit bieten beispielsweise die demnächst erwarteten Vorschläge der EU-Kommission zu Haftungsfragen bei KI-Systemen, unter anderem im Zusammenhang mit der geplanten Revision der Produkthaftungsrichtlinie.[60]

Sicher ist, dass die Erfüllung der im KI-Verordnungsentwurf festgelegten Transparenzpflichten nicht verhindern wird, dass Systeme so eingesetzt werden, dass sie diskriminierende Wirkung entfalten oder andere Arten von ungerechtfertigten Ungleichbehandlungen mit sich bringen. Die einzigen Hebel, die der Verordnungsentwurf an dieser Stelle ansetzen könnte, sind die Erfordernisse bezüglich Risikomanagement[61] und Datenqualität[62]. Es zeigt sich jedoch, dass ein enger Fokus auf eine Verzerrung (*bias*) in den verwendeten Daten der vielschichtigen Weise, in der KI-Systeme zu Ungerechtigkeiten beitragen können, nicht Rechnung trägt: Diese können einerseits auch durch die zugrundeliegenden (von Menschen gemachten) Entscheidungs- und Interpretationsmodelle hervorgerufen werden, andererseits entspringen sie darüber

---

59 Siehe unten, C.V.

60 Siehe oben, Fn. 7.

61 Art. 9 COM(2021) 206 final.

62 Art. 10 COM(2021) 206 final.

hinaus oft dem gesellschaftlichen Kontext, in dem das System eingesetzt wird: So kann ein System mit sich selbst verstärkenden Rückkoppelungsschlaufen einhergehen, die nicht durch einen engen Fokus auf die Verbesserung der Datenqualität verhindert werden können. Massnahmen auf technologischer Ebene können zwar notwendig sein, sind aber nicht hinreichend, wenn es um Anforderungen hinsichtlich Nicht-Diskriminierung und Gerechtigkeit von ADM-Systemen geht.[63]

V. Die Perspektive der Grundrechtsträger·innen

Aus einer grundrechtlichen Perspektive erscheint besonders irritierend, dass der KI-Verordnungsentwurf zwar grundrechtliche Ziele verfolgt, allerdings die Träger·innen dieser Grundrechte – also die Personen, deren Rechte von KI-Systemen potenziell berührt werden – unerwähnt bleiben: Der Entwurf schreibt ihnen weder prozedurale noch substanzielle Rechte zu.[64] Um Individuen vor negativen Auswirkungen von KI-Systemen zu schützen – das erklärte Ziel des Verordnungsentwurfs – benötigen sie einerseits Zugang zu Rechtsmitteln, um sich individuell oder kollektiv zur Wehr zu setzen und Wiedergutmachung zu erwirken. Die Möglichkeit, rechtliche Mittel zu ergreifen, setzt wiederum voraus, dass auch gegenüber Betroffenen Transparenz gewährt wird: Sie benötigen beispielsweise Zugang zu Information über eine von einem KI-System getroffene Entscheidung, die sie betrifft, inklusive einer Erklärung des grundsätzlichen Entscheidungsprozesses. Weiter scheint es angezeigt, dass auch Individuen und ihre Vertreter·innen die Möglichkeit haben, bei nationalen Aufsichtsbehörden Beschwerde einzureichen und dadurch Untersuchungen auszulösen. Andererseits fehlt im KI-Verordnungsentwurf auch das substanzielle Recht, einem KI-System nicht unterworfen zu sein, das mit einem unakzeptablen Risiko einhergeht oder das die Erfordernisse der Verordnung nicht erfüllt.

In dieser Hinsicht zeigt sich denn auch eine Diskrepanz zur Datenschutzgrundverordnung, die eine Reihe von Rechten für Individuen vorsieht, wenn deren Daten bearbeitet werden.[65] Der derzeitige KI-Verordnungsentwurf weist hier wesentliche Lücken auf.

---

63 Balayn Agathe/Gürses Seda, Beyond Debiasing: Regulating AI and its Inequalities, European Digital Rights (Hrsg.), Brussels 2021, abrufbar unter <https://edri.org/our-work/if-ai-is-the-problem-is-debiasing-the-solution/>.

64 Diese Lücken werden denn auch in verschiedenen Stellungnahmen kritisiert, vgl. etwa Smuha et al., 44 ff., 50 ff.

65 Art. 12 ff. Verordnung (EU) 2016/679; Smuha et al., 50 ff.

VI. Durchsetzung

Zentral für die Wirksamkeit der KI-Verordnung in der Praxis wird sein, dass Durchsetzungsstrukturen vorhanden und sinnvoll ausgestaltet sind. Die Abstützung auf die Selbsteinschätzung der Anbieter·innen, die nur moderate Bedeutung von notifizierten Stellen und die Unklarheit bezüglich Kompetenzen und Zuständigkeiten – all dies sind Aspekte, bei welchen die Gesetzgeber·innen nun die Möglichkeit haben, Verbesserungen einzuführen. Eine besondere Bedeutung kommt in diesem Zusammenhang auch den nationalen Aufsichtsbehörden zu: Erstens empfiehlt es sich, diese mit umfassenden Aufsichts- und Initiativbefugnissen auszustatten und, wie oben erwähnt, Anlaufstellen und Beschwerdemechanismen für betroffene natürliche oder juristische Personen sicherzustellen. Zweitens sieht der Entwurf zwar explizit vor, dass „die zuständigen nationalen Behörden mit angemessenen finanziellen und personellen Ressourcen ausgestattet werden“[66] und verweist dabei auch auf die notwendige Expertise. Gleichzeitig wird in seiner Begründung ausgeführt, dass dafür eine bis 25 Vollzeitstellen pro Mitgliedstaat vorgesehen wären.[67] Um die Aufgaben einer sinnvollen und umfassenden Aufsicht wahrzunehmen, ist dies deutlich ungenügend. Nicht zuletzt die Herausforderungen im Zusammenhang mit der Umsetzung der Datenschutzgrundverordnung[68] haben gezeigt, dass es sich lohnt, dies mit der entsprechenden Sorgfalt anzugehen.

Abschliessend lässt sich sagen, dass die EU-Kommission mit der KI-Verordnung eine horizontale Regulierung mit nicht zu unterschätzender Ausstrahlkraft vorgeschlagen hat. Sie enthält wichtige Hebel, um die Transparenz zu risikobehafteten KI-Systemen zu fördern, was grundsätzlich zu begrüssen und anzuerkennen ist. Gleichzeitig scheint, dass ihr Ansatz, Grundrechtsschutz via Binnenmarktregulierung, Selbsteinschätzung und Transparenzpflichten sicherzustellen, verschiedene Spannungsverhältnisse mit sich bringt, die Fragen bezüglich der Wirksamkeit der Regulierung in der Praxis aufwerfen.

Das Narrativ der EU-Kommission, einen Rahmen für den Einsatz von KI zu schaffen, dem die Menschen vertrauen, scheint noch immer geprägt von der impliziten Vorstellung, dass das Vertrauen der Bevölkerung Mittel zum Zweck ist, um Innovation zu befördern. Innovation ist essenziell für unsere Gesellschaft – doch sie ist kein Zweck an sich. Auch Innovation soll letztlich der Menschheit dienen. Entsprechend kann das Ziel, das Vertrauen der Menschen zu erhöhen, nur darüber zu erreichen sein, dass der Einsatz neuer Techno-

---

66 Art. 59(4) COM(2021) 206 final.

67 Begründung, Kap. 4 COM(2021) 206 final.

68 Verordnung (EU) 2016/679.

logien eben *vertrauenswürdig* gestaltet wird. Die Pfeiler, an denen sich Innovation und die Verwendung neuer Technologien zu orientieren haben, sind jene, auf die wir uns als Gesellschaft geeinigt haben: Demokratie, Grundrechte, Rechtsstaatlichkeit.

D. Auswirkungen auf die Schweiz

Wie im vorhergehenden Kapitel erwähnt, würde die KI-Verordnung direkte extraterritoriale Wirkung entfalten: Anbieter·innen und Nutzer·innen in der Schweiz wären an sie gebunden, wenn sie ihre Systeme in der EU anbieten oder das Ergebnis ihrer Systeme in der EU verwendet wird. Während es sich bei Anbieter·innen regelmässig um private Unternehmen handeln wird, könnten als Nutzer·innen nebst privaten auch öffentliche Stellen davon erfasst werden – nämlich dann, wenn sie Systeme einsetzen, deren Ergebnisse innerhalb der EU verwendet würden. Denkbar ist beispielsweise ein Chatbot, der auf einer Website der Schweizer Bundesverwaltung von in der EU wohnhaften Auslandschweizer·innen zu Informationszwecken genutzt würde – auch dieser müsste demnach zukünftig die Transparenzpflichten gemäss Art. 52(1) KI-Verordnungsentwurf erfüllen. Eine Ausnahme sieht der Entwurf vor für Behörden, die KI-Systeme nutzen, wenn dies im Rahmen eines Übereinkommens mit der EU oder einem ihrer Mitgliedstaaten im Bereich Strafverfolgung und Justiz geschieht.[69] Dies würde etwa eine Ausnahme bedeuten für schweizerische Behörden, die im Rahmen des Übereinkommens mit Europol KI-Systeme einsetzen.

Nebst diesen direkten rechtlichen extraterritorialen Auswirkungen lässt sich jedoch bereits jetzt feststellen, dass der Vorschlag der EU auch einen Einfluss auf politische Debatten und Prozesse in der Schweiz haben wird. Die Schweiz ist aufgerufen, sich zur Problematik zu positionieren – einerseits, weil es angezeigt ist, Rahmenbedingungen für den verantwortungsvollen Umgang mit KI zu schaffen. Die Verwendung von KI-Systemen wirft neue Fragen auf, kann mit Risiken für individuelle Autonomie, Gerechtigkeit und Gemeinwohl einhergehen und scheint punktuell bestehende rechtliche Konzepte zumindest herauszufordern. Eine reflektierte und evidenzbasierte Debatte, wie wir diesen Einsatz gestalten wollen und wie wir ihn so gestalten können, dass er tatsächlich zum Nutzen der Menschen geschieht, ist unabdingbar. Dies beinhaltet sowohl ein breites gesellschaftliches Engagement, das den verschiedensten Teilen der Gesellschaft zugänglich ist, als auch den politischen Aufruf an den Gesetzgeber, aktiv zu werden. Andererseits wird die Schweiz von der EU-Rechtssetzung

---

[69] Art. 2(4) COM(2021) 206 final.

geprägt, ist eng mit dem EU-Binnenmarkt verknüpft und vom Marktzugang abhängig. Auch wenn dies nicht bedeutet, dass die Schweiz die EU-Regeln nachbilden muss, wird sich mit Inkrafttreten der KI-Verordnung – womit frühestens 2024 zu rechnen ist – der Handlungsbedarf zweifellos noch verstärken.

Vor diesem Hintergrund ist es angezeigt, die Reflektion und die Arbeit dazu jetzt aufzunehmen. Umso bedauerlicher scheint es, dass der Bundesrat bisher zuwartet, auch wenn ein proaktiverer Ansatz von parlamentarischer Seite bereits eingefordert wird.[70]

Das bedeutet gleichzeitig nicht, dass die Schweiz aufgefordert ist, es der EU gleich zu tun. Eine horizontale Regulierung von KI im Sinne des EU-Vorschlags bringt gleichzeitig das Risiko mit sich, dass zu wenig sorgfältig auf den Kontext eines Systems eingegangen und der Blick auf den grösseren Zusammenhang verschleiert wird. Die Verwendungsweise von KI-Systemen variiert enorm, von der Steuerung von Laufbändern in Fabriken bis hin zur Vergabe von Sozialleistungen. Es kann mit guten Gründen hinterfragt werden, ob im Lichte dieser enormen Varietät ein horizontaler Regulierungsansatz geeignet erscheint.

Die Alternative ist, da rechtsetzend tätig zu werden, wo sich rechtliche Lücken zeigen. Führende Schweizer Rechtswissenschaftler·innen haben im Detail analysiert und aufgezeigt, wo diese Lücken bestehen und wie sie zu schliessen wären.[71] Sie verstehen denn auch ihren Beitrag „als Anstoss für eine vertiefte Diskussion und als Aufruf an den schweizerischen Gesetzgeber, die Erarbeitung eines Rechtsrahmens zur Erfassung der Herausforderungen von KI zeitnah anzugehen“.[72]

In diesem Zusammenhang ist erstens wichtig zu betonen, dass es für die Schaffung der erwähnten Rahmenbedingungen für einen verantwortungsvollen Einsatz von KI, der der Gesellschaft zugutekommt, verschiedenste Massnahmen braucht – sei dies im Bereich der Forschung, der unabhängigen Aufsicht, der technologischen Entwicklung, des zivilgesellschaftlichen Engagements, der Bildung und Förderung der *Algorithmic Literacy* und digitaler Kompetenzen, der öffentlichen Debatte oder eben der rechtlichen Regulierung. Zweitens wird es bei dieser rechtlichen Regulierung notwendig sein, ge-

---

70 Vgl. dazu Stellungnahme des Bundesrates vom 25. August 2021 zu Motion 21.3676 vom 10. Juni 2021, Bellaiche Judith, abrufbar unter <https://www.parlament.ch/de/ratsbetrieb/suche-curia-vista/geschaeft?AffairId=20213676>.

71 Braun Binder Nadja et al., Künstliche Intelligenz: Handlungsbedarf im Schweizer Recht, Jusletter, 28. Juni 2021a.

72 Braun Binder et al., 2021a, 1.

setzgeberisch tätig zu werden und durch generelle und sektorielle Normen Lücken zu schliessen sowie bestehende Normen zu ergänzen und anzupassen. Dies ist angezeigt etwa im Bereich des Diskriminierungsschutzes im Verhältnis zwischen Privaten,[73] zur Förderung der Transparenz etwa durch Informationspflichten für Betroffene oder öffentliche Register der im öffentlichen Sektor eingesetzten KI-Systeme[74] oder wenn es darum geht, rote Linien zu ziehen und gewisse Anwendungen von KI-Systemen zu verbieten. Einsatzweisen von KI-Systemen, die inhärent mit Grundrechten in Konflikt stehen, wie die Verwendung von biometrischen Erkennungssystemen im öffentlichen Raum, sollten in demokratischen Gesellschaften nicht erlaubt sein. Zudem existieren in verschiedenen Sektoren bereits Regelungen für die Nutzung von algorithmenbasierten Technologien, sei es beispielsweise im Finanzsektor oder in der Medizin. Auch hier ist es zentral, kontinuierlich Lücken zu identifizieren und zu schliessen.

Drittens kann es aber bei der rechtlichen Regulierung nicht einzig um den Aspekt der Rechtsetzung gehen, sondern müssen auch die Anwendung und Auslegung bestehender Normen in den Blick genommen werden.[75] In erster Linie müssen diese konsistent und konsequent auf Anbieter·innen und Nutzer·innen von KI-Systemen sowie auf die von oder mit Hilfe von KI-Systemen getroffenen Prognosen und Entscheidungen *angewendet* werden. Die Verwaltung ist weiterhin an Grundrechte und rechtsstaatliche Prinzipien, an das Legalitätsprinzip und Verfahrensgarantien gebunden, unabhängig davon, ob sie ihre Entscheide mit oder ohne automatisierte Unterstützung trifft.[76] Zudem prägen auch bereits bestehende technologieunabhängige Regelungen den Einsatz von KI-Systemen und setzen ihm Leitplanken.

---

73 Thouvenin Florent et al., Ein Rechtsrahmen für Künstliche Intelligenz, Positionspapier Digital Society Initiative (DSI) Strategy Lab, Balsthal/Zürich 2021, abrufbar unter <https://www.dsi.uzh.ch/dam/jcr:3a0cb402-c3b3-4360-9332-f800895fdc58/dsi-strategy-lab-21-de.pdf>.

74 Thouvenin et al.; Loi et al., 2; AlgorithmWatch, Submission on the European Commission's „White Paper on Artificial Intelligence – a European approach to excellence and trust", 2020, 3, abrufbar unter <https://algorithmwatch.org/en/response-european-commission-ai-consultation/>.

75 Braun Binder et al., 2021a, Rz. 55.

76 Braun Binder Nadja et al., Einsatz künstlicher Intelligenz in der Verwaltung: rechtliche und ethische Fragen, Schlussbericht vom 28. Februar 2021, Zürich 2021b, abrufbar unter <https://www.zh.ch/content/dam/zhweb/bilder-dokumente/themen/politik-staat/kanton/digitale-verwaltung-und-e-government/projekte_digitale_transformation/ki_einsatz_in_der_verwaltung_2021.pdf>.

Darüber hinaus muss sich – ganz zentral – die *Auslegung* bestehender Normen so weiterentwickeln, dass sie der technologischen Entwicklung und den Herausforderungen von ADM-Systemen gerecht wird. Um diese zentrale Aufgabe anzuleiten und der Komplexität der Problematik gerecht zu werden, ist eine interdisziplinäre und multisektorale Herangehensweise angezeigt. Der Umgang mit KI wirft Fragen auf, deren Beantwortung die Expertise aus verschiedenen Disziplinen – darunter Rechtswissenschaft, Informatik, Data Science, Philosophie oder Kommunikationswissenschaften – und verschiedenen Sektoren – Wissenschaft, Verwaltung, Privatsektor und Zivilgesellschaft – bedingen. Vorschläge wie jene der Einrichtung einer Expert·innenkommission[77] erscheinen in dieser Hinsicht vielversprechend. Gleichzeitig muss sichergestellt werden, dass dieses Engagement breit abgestützt ist und insbesondere auch die Perspektiven von Personen oder Gruppen, die von den Auswirkungen von KI-Systemen auf besondere oder überproportionale Weise betroffen sind, einbezogen werden. Wenn Rechtsetzung, -anwendung und -auslegung evidenzbasiert sein sollen, brauchen wir fundierte Expertise, die aber auch informiertes Wissen dazu, wie KI-Systeme sich real und konkret auf Mensch und Gesellschaft auswirken, berücksichtigt.

E. Fazit

Rahmenbedingungen für den Einsatz von KI-Systemen – inklusive ihrer rechtlichen Regulierung – müssen zum Ziel haben, dass dieser Einsatz so gestaltet werden kann, dass er den Einzelnen und der Gesellschaft tatsächlich nutzt, statt ihnen zu schaden, und diesen Nutzen gerecht verteilt. Was wir dafür brauchen, ist Transparenz zur Verwendung und Funktionsweise der Systeme, um eine evidenzbasierte und breite gesellschaftliche Debatte zu entfachen. Dies ist Grundlage, damit wir als Individuen selbstbestimmt Kontrolle über den Einfluss von KI-Systemen auf uns ausüben können und wir als Gesellschaft die Möglichkeit haben, auf demokratische Weise evidenzbasiert Kontrolle auszuüben. Nicht zuletzt muss sichergesellt werden, dass Möglichkeiten der Zuschreibung von Verantwortung gegeben sind, um die am Einsatz eines Systems beteiligten Akteure auf sinnvolle Weise zur Rechenschaft zu ziehen, wenn das System Auswirkungen hat, die gegen Rechte von Einzelnen verstossen oder der Gesellschaft schaden.

---

[77] Braun Binder et al., 2021a, Rz. 57; Thouvenin et al., 7.

Auf EU-Ebene haben derzeit Parlament und Rat die Möglichkeit, den KI-Verordnungsentwurf an wesentlichen Stellen nachzubessern, um diesen Ansprüchen gerecht zu werden. In der Schweiz sind Politik, Wissenschaft, Privatsektor und Zivilgesellschaft aufgerufen, sich an einem inklusiven Dialog und Engagement zur Thematik zu beteiligen, damit die Gestaltung dieser Rahmenbedingungen angegangen werden kann.

Künstliche Intelligenz: Regulatorische Überlegungen zum „Wie“ und „Was“

Rolf H. Weber

Neue technologische Entwicklungen stellen das Recht regelmässig vor Herausforderungen; diese Einschätzung gilt auch für die Künstliche Intelligenz. Die EU hat mit einem sehr detaillierten risikoorientierten Regulierungsvorschlag reagiert. Ein solches Konzept dürfte für die Schweiz nicht zielführend sein. Erfolgversprechender ist die Ausarbeitung eines technologieneutralen Rahmengesetzes, unter dessen „Schirm“ die regulierungsbedürftigen Themen einer Normierung durch Anpassungen bestehender Gesetze zugeführt werden.

Inhalt

A. Einleitung

Die Künstliche Intelligenz (KI) ist zu einem regulatorischen „Hot Topic" geworden. Der technologische Fortschritt im Digitalisierungsbereich rief eine Debatte über die Notwendigkeit der Regulierung von Algorithmen hervor; in einer ersten Phase haben insbesondere internationale Organisationen (OECD, Europarat, EU, UNESCO) mit Leitlinien und politischen Erklärungen die Diskussionen geprägt,[1] nunmehr liegen angesichts des Rufes nach weiteren Normierungen auch erste gesetzliche Vorschläge vor.

Phänomenologisch lassen sich verschiedene Erscheinungsformen von Algorithmen unterscheiden:[2] Die ursprünglichen deterministischen Algorithmen sind mehr und mehr durch modifizierende selbstlernende Algorithmen, die darauf abzielen, statistische Muster in Datensätzen zu erkennen und sich diese zunutze zu machen, verdrängt worden. Solche Algorithmen, die zur Technologie der Künstlichen Intelligenz gehören, bereiten die menschlichen Entscheidungen vor oder ersetzen sie durch technische Abläufe, obschon ein Algorithmus (im ethisch gehaltvollen Sinne) selbst keine Entscheidung treffen kann.[3]

Mit dem Entwurf für einen „Artificial Intelligence Act" (AIA) ist die EU-Kommission im April 2021 regulatorisch vorgeprescht.[4] Das sehr detaillierte Normengefüge basiert auf einem differenzierten risikobasierten Ansatz und reguliert vornehmlich die risikoreichsten Anwendungen.[5] Die Schweiz ist dadurch auch unter Handlungsdruck gekommen, selbst wenn eine Interdepartemen-

---

1 Zwischenzeitlich ist die Zahl der KI-Deklarationen von Institutionen, Organisationen und Verbänden auf über 150 gestiegen; angesichts des schon vorhandenen Schrifttums wird dieses Thema vorliegend nicht vertieft.

2 Vgl. *Rolf H. Weber/Simon Henseler*, Regulierung von Algorithmen in der EU und in der Schweiz, EuZ 2020, 28, 29; soweit nachfolgend themenspezifische Regulierungsaspekte zur Diskussion gelangen, wird auf diesen materialreichen Beitrag verwiesen, ohne die zitierten Quellen umfassend zu wiederholen.

3 *Weber/Henseler*, 29 m.w.V.

4 Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, April 21, 2021, COM(2021) 206 final.

5 Vgl. *Angela Müller*, Der Artificial Intelligence Act der EU: Ein risikobasierter Ansatz zur Regulierung von Künstlicher Intelligenz – mit Auswirkungen auf die Schweiz, EuZ 01/2022, A 3 ff.; *Philipp Roos/Caspar Alexander Weitz*, Hochrisiko-KI-Systeme im Kommissionsentwurf für eine KI-Verordnung, MMR 2011, 844 ff.

tale Kommission des Bundes im Dezember 2019 noch die Meinung vertrat, die Schweiz sei an sich gut gerüstet, um die neuen technologischen Herausforderungen normativ zu bewältigen.[6]

Das Ziel der nachfolgenden Überlegungen besteht nicht darin, die Chancen und Risiken, die sich für die Gesellschaft aus dem Einsatz von KI ergeben, erneut im Einzelnen darzustellen; die entsprechenden Einschätzungen im publizierten Schrifttum weisen keine erheblichen Diskrepanzen auf. Ebenso wenig soll der AIA-Vorschlag einer detaillierten Analyse der umfangreichen Bestimmungen unterzogen werden.[7] Vielmehr geht der Beitrag der Frage nach, welcher Regulierungsansatz im Kontext der Künstlichen Intelligenz sich für die Schweiz als sachgerecht erweist.[8] Aus diesem Grunde sind vorerst die wesentlichen Elemente möglicher Regulierungskonzepte und -modelle aus einer theoretischen Perspektive darzustellen bzw. deren Charakteristiken gegeneinander abzuwägen.

B. Regulierungskonzepte und -modelle

I. Gestaltungsaufgabe des Rechts

Das Recht ist ein strukturelles System, das auf organisierten oder untereinander verbundenen regulatorischen Elementen beruht. Die Funktion des Rechts besteht darin, durch verhaltensbezogene Vorgaben und Institutionen das Zu-

---

6 Interdepartementale Arbeitsgruppe Künstliche Intelligenz, Herausforderungen der Künstlichen Intelligenz, Bericht an den Bundesrat, 13. Dezember 2019, 10 ff. Verabschiedet hat der Bundesrat immerhin Leitlinien „Künstliche Intelligenz" für den Bund, Orientierungsrahmen für den Umgang mit künstlicher Intelligenz in der Bundesverwaltung, 25. November 2020; weiter hat der Bundesrat am 25. August 2021 beschlossen, ein „Kompetenznetzwerk Künstliche Intelligenz, KNW KI" in der ersten Jahreshälfte 2022 ins Leben zu rufen.

7 Vgl. dazu statt Vieler etwa *Müller*, A 5 ff.; *Roos/Weitz*, 844 ff.; *Gerald Spindler*, Der Vorschlag der EU-Kommission für eine Verordnung zur Regulierung der Künstlichen Intelligenz (KI-VO-E), CR 2021, 361 ff.; *Michael Veale/Frédéric Zuidereveen Borgesius*, Demystifying the Draft EU Artificial Intelligence Act, Comp. L. Rev. Int. 2021, 97 ff.

8 Diese Frage ist auch der Ausgangspunkt des Beitrags von *Nadja Braun Binder/Thomas Burri/Melinda Florina Lohmann/Monika Simmler/Florent Thouvenin/Kerstin Noëlle Vokinger*, Künstliche Intelligenz: Handlungsbedarf im Schweizer Recht, Jusletter 28. Juni 2021, Rz. 4; detailliert diskutiert werden die relevanten Themenbereiche, nicht aber das regulierungstheoretische Konzept, weshalb sich der vorliegende Beitrag als komplementäre Ergänzung verstehen lässt.

sammenleben in der Gesellschaft zu normieren und Konflikte zu lösen, die sich bei der Anwendung der Regeln ergeben.[9] Die Schaffung von Recht vermag dabei auf verschiedenen, nachfolgend anzusprechenden Quellen zu beruhen.

Gewisse Grundfragen stellen sich aber für jede Rechtsordnung:[10] Wer darf regulieren? In welchem Interesse? Durch welche Mechanismen? Mit welchen Zielen? Ungeachtet der konkreten Beantwortung dieser Fragen erscheint indessen als unbestritten, dass dem Recht eine Gestaltungsaufgabe mit Blick auf die zwischenmenschlichen Beziehungen in der Gesellschaft zukommt.[11]

II. Staatliche und kooperative Regulierung

Die traditionellen regulatorischen Instrumente finden sich – für grenzüberschreitende Erscheinungen – in den internationalen Staatsverträgen und in den einzelstaatlichen Rechtsordnungen (Verfassung, Gesetz). Diese Rechtsquellen kommen in einem geordneten (oft demokratischen) Verfahren zustande; die einzelnen Normen lassen sich auch obrigkeitlich mit Zwang durchsetzen.[12] Gerade mit Blick auf schnelle technologische Veränderungen hat sich in den letzten Jahrzehnten aber gezeigt, dass staatliches Recht meist nur relativ langsam entsteht und hernach zur Anwendung gelangt.[13] Andere normative Instrumente erweisen sich deshalb zur Ergänzung des Regulierungsumfelds als notwendig.

Seit Jahren ist anerkannt, dass neben dem staatlichen „Hard Law“ auch das „Soft Law“ eine gewichtige Rolle zu spielen vermag. Unter den Begriff des Soft Law fallen die verschiedenartigsten Formen von Selbstregulierungen, die auf Initiative der betroffenen Branchen, Unternehmen oder zivilgesellschaftlichen Vereinigungen entstehen.[14] Der Vorteil von Selbstregulierungen liegt darin, dass die entsprechenden Bestimmungen gestützt auf Know How und Erfahrungen der Betroffenen zeitgerecht und zu tiefen Kosten entwickelt werden,

---

9 Für einen neueren Überblick vgl. *Rolf H. Weber*, Internet Governance at the Point of No Return, Zürich 2021, 13 f. m.w.V.; *Amnon Reichman/Giovanni Sartor*, Algorithms and Regulation, in: H.-W. Micklitz et al. (eds.), Constitutional Challenges in the Algorithmic Society, Cambridge 2022, 131 ff.

10 Vgl. *Rolf H. Weber*, Realizing a New Global Cyberspace Framework, Zürich 2014, 34-36.

11 Grundlegend dazu *Herbert L.A. Hart*, The Concept of Law, 2nd ed., Oxford 1997, 55 ff.

12 Allgemein zum Recht des Cyberspace vgl. *Chris Reed*, Making Laws for Cyberspace, Oxford 2012, 70-73, 105/06.

13 Im Schrifttum wird oft von „regulatory lag“ gesprochen: vgl. *Weber/Henseler*, 32.

14 Für einen ganz neuen Überblick vgl. *Rolf H. Weber*, Integrity in the ‚Infinite Space‘ – New Frontiers for International Law, ZaöRV 81 (2021), 601, 619 f.

der Nachteil hingegen darin, dass sie sich obrigkeitlich nicht durchsetzen lassen.[15] Soft Law als Form einer kooperativen Regulierung schneidet indessen qualitativ keineswegs schlechter ab als Hard Law.[16]

Als Zwischenform hat in den letzten Jahren das Modell der „Co-Regulierung" (bzw. der „regulierten Selbstregulierung") an Bedeutung gewonnen. Dieses Regulierungskonzept verstärkt Selbstregulierungen durch eine staatliche „Anerkennung" bzw. sogar eine Überwachung der Beachtung von Verhaltenspflichten.[17] Vielfältige Beispiele sind aus den Finanz-, Medien- und Internetmärkten bekannt.[18] Dank der Mitwirkung des traditionellen Gesetzgebers bei der Durchsetzung der normativen Bestimmungen verstärkt sich deren Wirkung.[19] Das Modell der Co-Regulierung entspricht dem im Kontext der Internet Governance intensiv diskutierten und teilweise auch angewendeten Multistakeholder-Konzept; die Entwicklung eines erfolgreichen normativen Umfelds setzt voraus, dass alle betroffenen Akteure (z.B. Regierungen/Verwaltungen, Unternehmen, Zivilgesellschaft, Akademie) gemeinsam an der Regelbildung mitwirken.[20] Diese Vorzüge sind gerade für den KI-Bereich fruchtbar zu machen.

III. Horizontale und vertikale Regulierung

Mit dem Entwurf für einen Artificial Intelligence Act (AIA) hat die EU-Kommission einen horizontalen Regulierungsansatz gewählt; grundsätzlich ist das risikobasierte Modell für alle Wirtschaftssegmente und Branchen relevant.[21] Der Vorteil der horizontalen Regulierung liegt darin, dass eine normative Segmen-

---

15 *Rolf H. Weber*, Artificial Intelligence *ante portas*: Reactions of Law, J 2021(4), 486, 488.

16 *Weber/Henseler*, 33 m.w.V.

17 Eingehender dazu *Weber*, 621 f.

18 Vgl. dazu *Chris T. Marsden/Trisha Meyer/Ian Brown*, Platform values and democratic elections: How can the law regulate digital disinformation? CLSR 36 (2020), 105373, 1, 9 ff.

19 Vgl. auch *Myriam Senn*, Non-State Regulatory Regimes, Understanding Institutional Transformation, Berlin/Heidelberg 2011, 43, 139-148, 230.

20 Für eine umfassende Darstellung des Multistakeholder-Konzepts vgl. *Rolf H. Weber*, Legal foundations of multistakeholder decision-making, ZSR 135 (2016) I 247-267; zur Anwendung des kooperativen Regulierungsansatzes im KI-Kontext vgl. *Weber*, 487 f. und *Weber/Henseler*, 33.

21 Vgl. zum horizontalen risikobasierten Ansatz auch *Müller*, A 6 ff.

tierung vermieden werden kann; nachteilig wirkt sich hingegen aus, dass die Berücksichtigung von branchenspezifischen Besonderheiten nur schwer möglich ist.[22]

Ein vertikaler Regulierungsansatz würde es erlauben, für die betroffenen Wirtschaftssegmente eine möglichst massgeschneiderte Regulierung zu verwirklichen. Der Gesundheitssektor bedarf z.B. nicht zwingend derselben Vorgaben wie die Autobranche. Angesichts der vielfältigen KI-Einsatzmöglichkeiten weist eine sektorspezifische Regulierung deutliche Vorzüge auf; sie vermag den normativen Herausforderungen präziser Rechnung zu tragen und vermeidet, dass allgemeine Anordnungen teilweise ungeeignet sind. Für den KI-Bereich hat dieses Konzept im Vergleich zu einem nur horizontalen Ansatz nicht zu unterschätzende Vorteile.

IV. Regulierungstiefe und -dichte

Der AIA-Entwurf, den die EU-Kommission vorgelegt hat, zeichnet sich durch eine sehr hohe Regulierungstiefe und -dichte aus. Auf über 100 Seiten finden sich viele detaillierte Einzelregeln, deren Umsetzung sich nicht immer leicht vornehmen lässt, was für die Rechtsanwendung und -durchsetzung als nicht unproblematisch erscheint.

Ein anderer Regulierungsansatz würde darin bestehen, konkret zu prüfen, welche Sachverhalte im Interesse des Staates und der Zivilgesellschaft tatsächlich einer Regulierung bedürfen. Dieses Vorgehen hat der Bundesrat mit Bezug auf den durch die Distributed Ledger Technology (DLT) verursachten Handlungsbedarf gewählt. Zwar erklärte der Bundesrat in der Botschaft zum DLT-Gesetz (November 2019), bestmögliche Rahmenbedingungen für DLT-Geschäftsmodelle schaffen zu wollen, damit sich die Schweiz als ein führender, innovativer und nachhaltiger Standort für Blockchain-Unternehmen etablieren und weiterentwickeln könne.[23]

---

22 Im Finanzmarktrecht hat die FINMA bzw. später der Bundesrat beabsichtigt, die traditionelle vertikale Regulierung durch ein horizontales Modell zu ersetzen. Bis zu einem gewissen Grade ist dies mit dem FinfraG und dem FIDLEG auch geschehen; im Laufe der Ausarbeitung der neuen Finanzmarktarchitektur hat sich aber gezeigt, dass gewisse vertikale Regulierungen in Kraft bleiben müssen (z.B. neben dem FiniG das BankG, das VAG, das KAG), um branchengerechte Normierungen zu erreichen (vgl. *Thomas Jutzi/Ksenia Wess/Damian Sieradzki*, Die neue Finanzmarktarchitektur im europäischen Regulierungskontext, AJP 2020, 572, 583 m.w.V.).

23 Botschaft des Bundesrates vom 27. November 2019 zum Bundesgesetz zur Anpassung des Bundesrechts an Entwicklungen der Technik verteilter elektronischer Register, BBl 2020, 233, 239.

Materiell hat der Bundesrat indessen auf den Vorschlag für ein umfassendes Blockchain-Gesetz, wie z.B. im Fürstentum Liechtenstein realisiert, verzichtet, und den bewährten und ausgewogenen Rechtsrahmen der Schweiz nicht grundsätzlich in Frage gestellt.[24] Vielmehr nimmt das zwischenzeitlich vom Parlament (einstimmig) verabschiedete DLT-Gesetz lediglich diejenigen Anpassungen in einzelnen Gesetzen vor, die erforderlich gewesen sind, um den DLT-Anwendungen nicht normative Lücken oder Hindernisse in den Weg zu stellen.[25]

Eine ähnliche Vorgehensweise erscheint auch mit Bezug auf die Künstliche Intelligenz als denkbar und sachgerecht. Nach einer Identifikation der relevanten Themen und der Analyse des konkreten Normierungsbedarfs ist zu beurteilen, welcher theoretische Regulierungsansatz sinnvoll ist. In der Umsetzung liesse sich dann – ähnlich wie im Fall des DLT-Gesetzes – ein „Bundesgesetz zur Anpassung des Bundesrechts an Entwicklungen der Künstlichen Intelligenz" ausarbeiten. Im Gegensatz zum DLT-Gesetz dürfte es indessen nicht ausreichen, in der Einleitung lediglich festzuhalten, die „nachstehenden Erlasse werden wie folgt geändert", sondern zu einzelnen Themen (z.B. Nichtdiskriminierung, IT-Sicherheit) sind ggf. „neue" Gesetze vorzubereiten.

C. Struktureller und themenbezogener Regulierungsansatz

I. Ausgangslage: Kooperative Regulierung mit Rahmengesetz

Die Künstliche Intelligenz ist eine moderne und sich noch stark im Fluss befindliche Technologie. Ein gewisser Handlungsbedarf lässt sich aber bereits identifizieren. Die Wahrscheinlichkeit ist überdies erheblich, dass zusätzliche Fragen und Probleme erst im Laufe der nächsten Monate und Jahre auftreten. Aus diesem Grunde drängt sich ein flexibles normatives Regelwerk auf. Im Sinne der vorangehenden theoretischen Überlegungen sind auch die Vorteile der vertikalen (sektorspezifischen) gegenüber der horizontalen Regulierung fruchtbar zu machen.

Wie erwähnt ist die Erarbeitung und Inkraftsetzung von staatlichen Rechtsinstrumenten regelmässig zeitintensiv; neue Regelungen hinken deshalb den Technologien nach, wie die bereits einsetzende Diskussion zur sog. tech-

---

24 Botschaft, 239 f.

25 Für einen Überblick vgl. *Hans Kuhn/Rolf H. Weber*, Einleitung, in: R.H. Weber/H. Kuhn, Entwicklungen im Schweizer Blockchain-Recht, Basel 2021, Kap. 1, Rz. 1 ff.

nologieneutralen Umschreibung des Begriffs der Künstlichen Intelligenz im AIA-Vorschlag zeigt.[26] Aus diesem Grunde erscheint es als wünschenswert, Soft Law-Komponenten im KI-Regulierungsumfeld ebenfalls zu berücksichtigen. Diese Vorgehensweise lässt sich am ehesten verwirklichen, wenn sich die staatliche Mitwirkung an der Normsetzung auf ein Rahmengesetz beschränkt und ein kooperativer Regulierungsansatz verwirklicht wird.

In überzeugender Weise werden im Europarat mit seinen 47 Mitgliedern mögliche Modelle für eine Regulierung der Künstlichen Intelligenz diskutiert. Das am 11. September 2019 eingesetzte Ad hoc Committee on Artifical Intelligence (CAHAI) hat im umfangreichen Bericht vom 17. Dezember 2020[27] nicht nur die Chancen und Risiken der Künstlichen Intelligenz analysiert und insbesondere die Relevanz von Menschenrechten und Demokratie für KI-Regulierungen diskutiert, sondern auch die möglichen Formen von rechtlichen Instrumenten erläutert. Das Kapitel „Mapping of Instruments Applicable to Artificial Intelligence“[28] hält zwar dafür, dass die bestehenden Ethik-Leitlinien regulatorisch nicht ausreichen würden, aber doch verschiedene Optionen verbleiben, um eine rechtliche Ordnung für KI sinnvoll festzulegen.

Insbesondere bringt der CAHAI-Report die Idee einer „Framework Convention“ auf.[29] Im Gegensatz zum AIA-Vorschlag der EU, der eine umfassende und detaillierte Regulierung postuliert, öffnet der Europarat zutreffend den Weg zu einer differenzierten rechtlichen Strukturierung der normativen Vorgaben. Dieser Vorschlag überlässt den Mitgliedstaaten einen nicht unerheblichen Freiraum, den auch die Schweiz mit einem eigenen Ansatz ausnutzen

---

26 Zu diesem vorliegend nicht zu vertiefenden Problem vgl. *Jonas Fischer/Mathias Fuchs*, Brauchen wir eine Legaldefinition für künstliche Intelligenz? Jusletter 8. November 2021, Rz. 18 ff.

27 Council of Europe, Ad hoc Committee on Artificial Intelligence, Feasibility Study, CAHAI(2020) 23, https://www.coe.int/en/web/artificial-intelligence/cahai (CAHAI-Report).

28 CAHAI-Report, 18 ff.

29 CAHAI-Report, 46 ff. CAHAI hat an diesem Konzept im Final Meeting vom 29.11.–2.12.2021 ausdrücklich festgehalten („Possible elements of a legal framework on artificial intelligence“).

könnte,[30] und weist zudem Ähnlichkeiten zum erwähnten Modell eines „Rahmengesetzes“ auf, den der Bundesrat mit Blick auf die DLT-Entwicklungen gewählt hat.[31]

Ein entsprechendes Regulierungskonzept erweist sich auch für den künftig in der Schweiz normativ zu bewältigenden KI-Bereich als sachgerecht. Das Modell des Rahmengesetzes wird im Schrifttum als mögliche Regulierungsform anerkannt und gerade in komplexen Materien für sinnvoll gehalten.[32] Statt ein horizontales umfassendes „KI-Gesetz“ zu erlassen, können in ein KI-Rahmengesetz allgemeine und spezifische Normen, jeweils – soweit möglich – in Abänderung und Ergänzung bestehender Rechtsquellen, aufgenommen werden.[33]

II. Probleme bei grenzüberschreitenden Sachverhalten

Die Angebote für durch Künstliche Intelligenz geprägte Produkte und Dienstleistungen bleiben nicht auf die Schweiz beschränkt. Vielmehr ist vorauszusehen, dass auch EU-Märkte bedient werden sollen. Der Entwurf der EU für die AIA-Verordnung sieht ausdrücklich vor, dass bei Angeboten an EU-Verbraucher deren Bestimmungen zu beachten sind.[34]

Die AIA-Verordnung wird also das Schutzniveau in der Schweiz mitprägen und der Schweizer Bevölkerung einen grundrechtlichen Schutz anbieten, sofern sich Unternehmen in der EU einzelner AI-Vorkehren in der Produktion von Gütern und der Bereitstellung von Dienstleistungen bedienen.[35] Eine gewisse materielle Kohärenz in der Ausgestaltung der Regulierungen wäre deshalb erwünscht, um zu vermeiden, dass der Zugang von Produkten und Dienstleis-

---

30 So auch *Florent Thouvenin/Markus Christen/Abraham Bernstein/Nadja Braun Binder/Thomas Burri/Karsten Donnay/Lena Jäger/Mariela Jaffé/Michael Krauthammer, Melinda Lohman/Anna Mätzener/Sophie Mützel/Liliane Obrecht/Nicole Ritter/Matthias Spielkamp/Stephanie Volz*, Digital Society Initiative, Positionspapier – Künstliche Intelligenz, November 2021, 2, https://www.itsl.uzh.ch/en/Knowledge-transfer/Publications.html.

31 Vgl. vorne Ziff. B.IV.

32 *Kuhn/Weber*, Rz. 17-19; für das Finanzmarktrecht vgl. *Pascal Zysset*, Selbstregulierung im Finanzmarktrecht, Diss. Bern, Zürich 2017, 207 f. m.w.V.

33 Dieser Regulierungsansatz wird nun auch von *Thouvenin et al.*, 2, vertreten.

34 Ähnlich wie bei der EU-DSGVO kommt nach Art. 2(1) des AIA-Vorschlags das EU-Recht zur Anwendung, wenn ein KI-System innerhalb der EU eingesetzt wird oder wenn das vom System hervorgebrachte Ergebnis innerhalb der EU verwendet wird.

35 Vgl. auch *Müller*, A 5 f.

tungen zum EU-Binnenmarkt erschwert wird.[36] Eine gesetzgeberische Lösung in der Schweiz müsste somit versuchen, sachlich vergleichbare bzw. EU-kompatible KI-Standards zu realisieren.

Äquivalenz bedeutet aber nicht Identität der Regulierungen. Diese Einschätzung zeigt sich seit Jahren in der Beurteilung des vergleichbaren Datenschutzniveaus; ausländische Datenschutzgesetze müssen nicht genau gleiche Schutzprinzipien und -mechanismen aufweisen, sondern „lediglich" in einer „äquivalenten" Weise die Vertraulichkeit der Daten von Schweizer Personen sicherstellen.[37] Zudem sollte die Schweiz nicht vorschnell auf einen durch ein weniger dichtes Regulierungsmodell bewirkten Wettbewerbsvorteil verzichten, nicht zuletzt auch mit Blick auf die vielen Abnehmer ausserhalb der EU, denn KI-Systeme lassen sich global vermarkten.

III. Identifikation der relevanten Themen

Wie erwähnt ist ein horizontaler Regulierungsansatz regelmässig gezwungen, eine hohe Regulierungsdichte zu realisieren, und hat dennoch den Nachteil, dass sektorspezifische Besonderheiten nicht berücksichtigt werden können. Wird von einem solchen umfassenden horizontalen Regulierungsansatz abgesehen, erweist es sich indessen als notwendig, die relevanten Themen der Künstlichen Intelligenz, die einen Handlungsbedarf auslösen, zu identifizieren. Nach zwischenzeitlich verbreiteter Meinung stehen folgende Themen im Vordergrund:[38]

- Nichtdiskriminierung und Fairness;
- Transparenz und Erklärbarkeit;
- Manipulation;
- Datenschutz und Datensicherheit;
- Haftung und Verantwortlichkeit.

Abgesehen von der Identifikation der relevanten Themen hat der Gesetzgeber auch die regulatorischen Instrumente festzulegen, die zum Einsatz kommen sollen. Der stärkste Eingriff würde in einem Verbot bestehen; eine Alternative besteht in der Festlegung von Zulassungsbedingungen, die der Anbieter von KI-Produkten oder -Dienstleistungen einzuhalten hat. Verbote sind grund-

---

36 Ob solche Handelshemmnisse ggf. einen Verstoss des WTO-Rechts zur Folge hätten, wäre gesondert zu prüfen; zum Ganzen auch *Braun Binder et al.*, Rz. 4.

37 Art. 16 Abs. 1 des neuen DSG verlangt einen „angemessenen Schutz" im Ausland, nicht das Vorliegen identischer Datenschutzgrundsätze.

38 Zu den einzelnen Themen vgl. nachfolgend Ziff. D.

sätzlich nur gerechtfertigt, wenn die Sicherheitsrisiken als so erheblich erscheinen, dass eine sachgerechte Minimierung mit möglichen Vorsichtsmassnahmen nicht in Frage kommt;[39] jedes Verbot führt nämlich dazu, dass die Angebote für entsprechende KI-Produkte und -Dienstleistungen völlig vom Markt verschwinden. Zulassungsverfahren sind insbesondere geeignet, die Realisierung von Sicherheitsvorkehren zu gewährleisten; im Einzelnen ist jeweils zu prüfen, ob bereits bestehende Zulassungsverfahren erweitert werden können oder ob es sich aufdrängt, neue Zulassungsverfahren zu schaffen.[40]

D. Regelungspunkte bei KI-relevanten Themen

Die nachfolgend angesprochenen, im KI-Kontext relevanten Themen betreffen besondere politik- und marktrelevante Bereiche. Potentielle KI-Risiken vermögen sich aber auch auf grundlegende Rechtsprinzipien auszuwirken; zutreffend thematisiert der Europarat im CAHAI-Report z.B. die Menschenwürde, die Grundfreiheiten, die Demokratie und die „Rule of Law".[41] Die materielle Diskussion der „ausgewählten" fünf wichtigen Themenbereiche erfolgt bewusst knapp, weil vertiefendes Schrifttum bereits vorhanden ist[42] und die rechtliche Rahmenordnung in einem Forschungsprojekt während der nächsten drei Jahre detailliert analysiert werden soll.[43]

I. Nichtdiskriminierung und Fairness

Die zentralste Herausforderung bei der KI-Regulierung betrifft wohl den Bereich der Nichtdiskriminierung bzw. Gleichbehandlung. Beispiele, die zeigen, dass gewisse menschliche Merkmale zu unfairen oder gar diskriminierenden Behandlungen führen können, weil mit einzelnen Merkmalen ein „Bias" ver-

---

39 Mögliche Beispiele für ein Verbot sind der Einsatz von Gesichtserkennung und anderen biometrischen Fernerkennungsverfahren im öffentlichen Raum zum Zwecke der Massenüberwachung oder der Einsatz von Social Scoring mit dem Ziel, eine Zugangsregulierung zu grundlegenden sozialen Diensten zu bewirken; vgl. *Thouvenin et al.*, 6 f.

40 Vgl. *Thouvenin et al.*, 6.

41 CAHAI-Report, 27 ff.; zur Menschenwürde, zu den Grundfreiheiten und zu den rechtstaatlichen Prinzipien eingehend *Rolf H. Weber*, Automatisierte Entscheidungen: Perspektive Grundrechte, SZW 2020, 18 ff. m.w.V.

42 Für weiterführende Überlegungen vgl. *Weber/Henseler*, 33 ff. und *Braun Binder et al.*, Rz. 8 ff.; beide Beiträge enthalten umfangreiche weitergehende Hinweise, auf deren erneute Erwähnung im Interesse einer „schlanken" Darstellung verzichtet wird.

43 Für einen Bericht über die Eröffnungsveranstaltung vom 10. November 2021 vgl. *Fabienne Graf/Liliane Obrecht*, Rechtliche Rahmenbedingungen für Künstliche Intelligenz in der Schweiz, Jusletter 29. November 2021, Rz. 1 ff.

bunden sein kann, dürfen als weitherum bekannt gelten.[44] Betroffen vom Grundsatz der Nichtdiskriminierung sind nicht nur private Unternehmen, sondern ebenso die öffentliche Verwaltung.[45]

Das Gebot der rechtsgleichen Behandlung ist verfassungsrechtlich verankert (Art. 8 Abs. 1 BV); immerhin bedarf dieser Grundsatz einer Verfeinerung, um sinnvoll anwendbar zu sein.[46] Zu beachten bleibt weiter, dass „Fairness" und Bias" in einzelnen Disziplinen weiter gefasst werden.[47] Besonders problematisch ist die indirekte bzw. verdeckte Diskriminierung, die sich ggf. lediglich in einzelnen Auswirkungen von Einschätzungen zeigt.[48]

Im Gegensatz zu anderen Themenbereichen dürfte sich zur Sicherstellung von Nichtdiskriminierung und Fairness ein horizontaler (allgemeiner) Gesetzeserlass, der öffentliche Verwaltungen und weitergehend als heute insbesondere private Unternehmen betrifft, als notwendig erweisen,[49] da die Ergänzung einer grossen Zahl von bestehenden Gesetzen kaum sinnvoll zu bewerkstelligen wäre.

II. Transparenz und Erklärbarkeit

Der Einsatz von Künstlicher Intelligenz muss transparent erfolgen; die betroffenen Personen sind zudem in die Lage zu versetzen, die KI-Vorgehensweise zu verstehen, was deren Erklärbarkeit und Interpretierbarkeit voraussetzt. Die Transparenz und Nachvollziehbarkeit von KI-Vorgängen sind adressatengerecht auszugestalten, abhängig vom Verständnishorizont des Betroffenen und der Bedeutung der Entscheidung für die erfasste Person; die der automatisierten Entscheidung zugrunde liegende Logik muss verständlich sein und die notwendigen Informationen enthalten, um einen ausreichenden Grad an Nachvollziehbarkeit zu erreichen.[50]

---

44 Im Einzelnen vgl. *Weber/Henseler*, 31 f. und 39 ff. sowie *Braun Binder et al.*, Rz. 21 ff., je m.w.V.

45 Für ein (österreichisches) Beispiel vgl. *Braun Binder et al.*, Rz. 23.

46 Zum sekundärrechtlichen bzw. einfachgesetzlichen Diskriminierungsschutz vgl. *Weber/Henseler*, 39 f.

47 *Braun Binder et al.*, Rz. 25 f.

48 *Braun Binder et al.*, Rz. 24 und 27 ff. m.w.V.

49 In diese Richtung gehen wohl auch *Thouvenin et al.*, 4; ob eine „organische Weiterentwicklung" des heutigen Rechts ausreicht (so *Braun Binder et al.* [Fn. 8], Rz. 30), erscheint mit Bezug auf den Nichtdiskriminierungsgrundsatz indessen als zweifelhaft.

50 *Thouvenin et al.*, 3; zu den spezifischen Transparenzanforderungen des EU-Rechts, das weiter entwickelt ist als das Schweizer Recht (z.B. im Gesundheitsbereich), vgl. *Weber/Henseler*, 37 f.

Denkbar wäre weiter die Einführung einer Kennzeichnungspflicht bei der Verwendung von algorithmischen Systemen. Um die Erkennbarkeit des KI-Einsatzes für die interessierte Öffentlichkeit zu gewährleisten, würde auch die Möglichkeit bestehen, ein öffentlich zugängliches Register zu schaffen, das aufzeigt, in welchen Bereichen die öffentliche Verwaltung einzelne KI-Systeme einsetzt.[51]

Im Datenschutz- und im Verbraucherrecht (stärker in der EU als in der Schweiz) sind erste Regulierungsansätze bereits vorhanden. So enthalten z.B. die DSGVO und das neue DSG der Schweiz eine spezifische Informationspflicht und ein Auskunftsrecht bei automatisierten Entscheidungen; offenzulegen sind aussagekräftige Informationen über die involvierte Logik sowie die Tragweite und die angestrebten Auswirkungen einer KI-Verarbeitung.[52] Zwar sind Inhalt und Tragweite der gesetzlichen Informationspflichten bzw. Auskunftsrechte im Einzelnen umstritten, doch hat sich zwischenzeitlich die Meinung durchgesetzt, die Grundannahmen des Algorithmus seien offenzulegen, nicht aber der Algorithmus selbst.[53]

Entsprechende rechtliche Regelungen erweisen sich insbesondere im Gesundheitsbereich als angebracht. Der AIA-Vorschlag erklärt den Einsatz von KI in medizinischen Instrumenten als mit hohem Risiko behaftet.[54] Für solche Systeme gelten spezifische Transparenzpflichten; offenzulegen ist, mit welchen Daten die KI trainiert wurde und wie sie funktioniert und wirkt.[55] Die Transparenzschaffung kann aber in Widerspruch zu den Geheimhaltungsinteressen der die KI-Systeme entwickelnden Unternehmen geraten, was einen konkreten Interessenabwägungsprozess im Lichte der gegebenen Umstände erforderlich macht.[56]

Einschränkungen mit Bezug auf die Transparenzschaffung sind auch beim raum- bzw. zeitbezogenen „Predictive Policing“ unumgänglich, um zu vermeiden, dass die vorausschauende Polizeiarbeit schon frühzeitig offengelegt werden muss.[57] Mit dem Predictive Policing lassen sich anhand statistischer Pro-

---

51 *Thouvenin et al.*, 3.

52 Art. 13 Abs. 2 lit. f und Art. 14. Abs. 2 lit. g DSGVO; Art. 21 Abs. 1 und Art. 25 Abs. 2 lit. f rev. DSG.

53 *Weber/Henseler*, 35, und *Braun Binder et al.*, Rz. 12, je m.w.V.

54 Art. 5 i.V.m. Anhang 2A des AIA-Vorschlags.

55 Im Einzelnen dazu *Braun Binder et al.*, Rz. 9 und 15.

56 *Braun Binder et al.*, Rz. 10 m.w.V.

57 Vgl. auch *Weber*, 24 f.

gnosen wahrscheinliche Vorfälle identifizieren, was der Polizei das Ergreifen präventiver (aus Sicherheitsgründen als sinnvoll erscheinender) Massnahmen ermöglicht.[58]

III. Manipulation

Künstliche Intelligenz lässt sich einsetzen, um menschliches Verhalten zu beeinflussen. Ein solcher Eingriff in die Autonomie der betroffenen Person vermag unbewusst auf das Denken und Handeln einer Person einzuwirken. Besonders problematisch ist die verdeckte Beeinflussung (z.B. durch Verwendung von Empfehlungsalgorithmen auf Plattformen der sozialen Medien).[59]

Manipulationen können auch im Kontext der demokratischen Prozesse vorfallen und damit z.B. den Schutz der freien Willensbildung beeinträchtigen.[60] Weltweit diskutiert wird die Manipulation insbesondere im Vorfeld von Wahlen und Volksabstimmungen; immerhin lässt sich nicht übersehen, dass bereits heute gewisse rechtliche Rahmenbedingungen bestehen, um – unabhängig von den genutzten Informationskanälen – gegen schwerwiegende Desinformationen vorzugehen.[61] Allgemein betrachtet steht das Recht indessen bei der Erfassung von Manipulationen und Desinformationen noch am Anfang.

Gewisse rechtliche Regelungen, deren Zweck darin besteht, Manipulationen zu vermeiden, sind hingegen im privatrechtlichen Bereich vorhanden: Das Verbreiten von manipulativer Information fällt in den Anwendungsbereich des strafrechtlichen Ehrverletzungsschutzes (Art. 173 ff. StGB) und des allgemeinen Persönlichkeitsrechts (Art. 28 ZGB). Diese rechtlichen Regelungen dürften geeignet sein, die wesentlichsten Desinformationen durch KI zu erfassen.[62]

Weiter enthält das Gesetz gegen den unlauteren Wettbewerb verschiedene Bestimmungen, die im Rahmen der Verbreitung marktrelevanter Desinformationen von Bedeutung sind (insbesondere Art. 3 Abs. 1 lit. b, lit. d und lit. i UWG sowie allgemein Art. 2 UWG).[63] Dennoch bleibt genauer zu analysieren, inwieweit eine Konkretisierung des Manipulationstatbestandes im UWG nicht sinnvoll wäre.

---

58 *Braun Binder et al.*, Rz. 13.

59 Vgl. auch *Thouvenin et al.*, 4.

60 *Thouvenin et al.*, 4 f.

61 Vgl. BGE 140 I 338, E.5.3; *Braun Binder et al.*, Rz. 36.

62 Vgl. dazu den kürzlich erschienenen BAKOM-Bericht, Intermediäre und Kommunikationsplattformen – Auswirkungen auf die öffentliche Kommunikation und Ansätze einer Governance, Bericht vom 17. November 2021.

63 *Braun Binder et al.*, Rz. 35 und 38.

Der AIA-Vorschlag sieht ein Verbot für gewisse Formen des manipulativen Gebrauchs von KI und für KI-Anwendungen vor (Art. 5 Abs. 1 lit. a und b). Die Formulierung des Tatbestandes ist aber sehr offen umschrieben und auch ungenügend zielgerichtet, weshalb sich eine Übernahme in das Schweizer Recht nicht aufdrängt; vielmehr erscheint die hiesige Rechtsordnung als hinreichend flexibel, um die relevanten Probleme zu erfassen, die durch Manipulationen mittels KI entstehen könnten.[64]

IV. Datenschutz und Datensicherheit

Der Schutz der Privatsphäre war historisch betrachtet das erste Thema, das im Kontext der Künstlichen Intelligenz heiss diskutiert worden ist.[65] Die Schaffung der DSGVO in der EU sowie des neuen DSG in der Schweiz scheint den Herausforderungen aber an Brisanz genommen zu haben.

Insbesondere die bereits erwähnten Bestimmungen zu den automatisierten Entscheidungen, aber auch die Anordnungen zu den Datenschutz-Folgeabschätzungen, nehmen den Regelungsbedarf, der durch den Einsatz von Algorithmen entsteht, detailliert auf. Einzelne Verfeinerungen lassen sich diskutieren[66] und sind auch angebracht; deren Realisierung dürfte aber nicht auf unüberwindbare Hindernisse stossen.

Unabhängig von diesen (theoretischen) Überlegungen bleibt indessen im Auge zu behalten, dass insbesondere die Vorgaben des Datenschutzrechts bereits sehr weit gediehen sind und es voraussichtlich in vielen Einzelkonstellationen weniger um legislatorische Ergänzungen wegen Regelungslücken als um die konkrete Umsetzung vorhandener Normierungen geht.[67]

Die Datensicherheit ist schon heute im digitalen Umfeld von grosser Bedeutung. Mit KI-Anwendungen liegt die Messlatte für Sicherheitsstandards aber noch höher. Insbesondere stellt sich die Frage, ob die Schweiz – ähnlich wie andere Länder – ein allgemeines IT-Sicherheitsgesetz erlassen sollte.[68] Im Sinne einer kooperativen Regulierung ist es indessen durchaus denkbar, dass private Organisationen sich bemühen, angemessene IT-Sicherheitsstandards

---

64 Vgl. auch *Braun Binder et al.*, Rz. 39.

65 Für einen Überblick vgl. *Weber/Henseler*, 34 ff. m.w.V.

66 Vgl. dazu *Braun Binder et al.*, Rz. 20.

67 Vgl. *Weber/Henseler*, 35 f.

68 *Thouvenin et al.*, 6.

zu entwickeln, wie dies insbesondere durch die International Standardisation Organisation (ISO) und auch schweizerische Vereinigungen bereits geschehen ist.[69]

V. Haftung und Verantwortlichkeit

In der Öffentlichkeit intensiv diskutiert wird die Problematik der zivilrechtlichen Haftung und der strafrechtlichen Verantwortlichkeit für die Verursachung von Schäden im Falle des Einsatzes von KI im Strassenverkehr.[70] Ohne Zweifel ergibt sich in diesem Bereich ein Handlungsbedarf, und zwar nicht zuletzt angesichts der Tatsache, dass die traditionellen Haftungsregeln (insbesondere diejenigen des Produkthaftungsrechts) an die physischen Güter und nicht an die körperlosen Dienstleistungen bzw. an die Software anknüpfen.[71]

Die EU ist mit ihren Vorbereitungsarbeiten zur Anpassung der Produkthaftpflicht-Richtlinie schon recht weit gediehen; zu klären ist immerhin noch die Einordnung digitaler Dienste.[72] Im Vordergrund steht die Frage, ob eine isolierte Steuerungssoftware als (Teil)Produkt qualifiziert werden kann, eine Annahme, die auch in der Schweiz vermehrt Zustimmung findet.[73] In dieser Betrachtungsweise liesse sich Software in jeglicher Form als „typische Erscheinung der fortschreitenden Technisierung" unter den Anwendungsbereich des Produkthaftpflichtgesetzes subsumieren.[74] Weiter klärungsbedürftig ist die Konkretisierung des Begriffs der Fehlerhaftigkeit im Falle von KI-Anwendungen.[75] Immerhin lässt sich nicht übersehen, dass die Schweizer Rechtsprechung die deliktische Produzentenhaftung durch eine extensive Auslegung der ausservertraglichen Norm von Art. 55 OR bereits seit Jahrzehnten erheblich erweitert hat.[76]

Ein besonderes Thema ist die Risikoprävention. Nach dem schon seit Jahren bekannten Konzept der Zuordnung von Risikosphären ist zu prüfen, welche

---

69 Vgl. dazu schon *Rolf H. Weber/Annette Willi*, IT-Sicherheit und Recht, Zürich 2006, 67 ff. und 74 ff. m.w.V.

70 Vgl. *Melinda F. Lohmann*, Automatisierte Fahrzeuge im Lichte des Schweizer Zulassungs- und Haftungsrechts, Baden-Baden 2016, 211 ff.

71 *Melinda F. Lohmann*, Ein zukunftsfähiger Haftungsrahmen für Künstliche Intelligenz. Warum die Schweiz ihr Produkthaftungsrecht aktualisieren muss, HAVE 2021, 111 ff.

72 Vgl. auch *Braun Binder et al.*, Rz. 41 f. m.w.V.

73 *Walter Fellmann*, Haftpflichtrecht im Zeichen der Digitalisierung, HAVE 2021, 105, 109.

74 *Fellmann*, 107; *Braun Binder et al.*, Rz. 43.

75 Vgl. dazu *Braun Binder et al.*, Rz. 44 f. m.w.V.

76 Diese Rechtsprechung hat vor fast 40 Jahren mit dem sog. Schachtrahmen-Entscheid des Bundesgerichts (BGE 110 II 456 ff.) ihren Anfang genommen.

Partei in der Lieferkette für gewisse Risiken am besten einzustehen vermag.[77] Grundsätzlich lässt sich sagen, dass die KI-Herstellerin am ehesten die Gefährdung durch sorgfältiges Programmieren zu beherrschen vermag; indessen verringert sich bei lernfähigen KI-Anwendungen, die nach dem Inverkehrbringen aufgrund der Nutzung trainiert werden, die Kontrolle.[78] Entlastungsmöglichkeiten ergeben sich aus der unsachgemässen Änderung eines KI-Systems in der kundenseitigen Anwendung oder im Falle von sog. Entwicklungsrisiken.[79]

Gesetzgeberischer Handlungsbedarf besteht somit im Haftungsrecht, selbst wenn einzelne bestehende Haftungsnormen (z.B. Art. 55 OR) durchaus analog auch im KI-Kontext angewendet werden können und die Umschreibung von *ex ante* Sorgfaltspflichten im traditionellen rechtlichen Umfeld als möglich erscheint. Für eine Schweizer Regulierung lässt sich indessen auf die bereits schon weit vorangeschrittenen Diskussionen im EU-Raum zurückgreifen: In Frage kommen z.B. die Einführung sektorspezifischer Gefährdungshaftungsnormen aufgrund der Herstellung oder Nutzung von KI-Systemen oder die Schaffung einer allgemeinen Gefährdungshaftung, kombiniert mit einer Versicherungspflicht.[80]

E. Ausblick

Der Gesetzgeber steht auch in der Schweiz vor neuen Herausforderungen im KI-Kontext. Die Notwendigkeit der Schaffung konkreter Regulierungen, die als unausweichlich erscheinen, um die KI-Risiken sachgerecht in den Griff zu bekommen, bedeutet aber nicht, dass zwingend der sehr detaillierte und teilweise problematische AIA-Vorschlag der EU „ungekürzt" übernommen werden muss. Vielmehr erweist es sich für die Schweiz als sinnvoll, gestützt auf einen kooperativen Regulierungsansatz in der Form eines Rahmengesetzes punktuell Anpassungen der bestehenden Normen in den betroffenen Rechtsbereichen vorzunehmen, soweit ein Handlungsbedarf vorliegt.[81]

Das Modell eines Rahmengesetzes, das – ähnlich wie das DLT-Gesetz – als „Schirm" die verschiedenartigen normativen Anpassungen zusammenhält, erweist sich als eine sinnvolle Vorgehensweise. Konkrete gesetzliche Ergän-

---

77 Zum Konzept der Risikosphären vgl. *Rolf H. Weber*, Smart Contracts: Vertrags- und verfügungsrechtlicher Regelungsbedarf? sic! 2018, 291, 297 f.

78 *Lohmann*, 120; *Braun Binder et al.*, Rz. 46.

79 Art. 5 Abs. 1 lit. b und lit. e PrHG; *Braun Binder et al.*, Rz. 47 f.

80 Im Einzelnen dazu *Braun Binder et al.*, Rz. 50 m.w.V.

81 Vgl. auch *Braun Binder et al.*, Rz. 55.

zungen mit horizontaler Wirkung sind im Datenschutz-, Persönlichkeits- und Lauterkeitsrecht erforderlich, doch wäre deren Ausmass gut überblickbar. Genauere vertikale Regelungen drängen sich im Gesundheitsrecht auf. Etwas komplexer ist die Situation im Haftungsrecht (erweiterte „Produkthaftung"), weil eine Kombination von horizontalen und vertikalen Regeln ins Auge zu fassen ist.

In zwei Bereichen wird – weitergehender als im DLT-Gesetz – zu analysieren sein, ob nicht eine umfassende horizontale Regulierung anzustreben wäre, nämlich beim Thema der Nichtdiskriminierung im privaten Bereich und beim Thema der IT-Sicherheit. Zwar wäre dieser Regulierungsansatz aufwendiger, aber im gesamtgesellschaftlichen Interesse an diesen beiden Materien wohl sachgerecht.

Bei der konkreten Ausgestaltung der neuen Gesetzesbestimmungen sind die erwünschte Äquivalenz zum EU-Recht und das Erfordernis der technologieneutralen Erfassung von KI-System im Auge zu behalten. Zur Verwirklichung dieser beiden Anliegen muss aber nicht der AIA-Vorschlag der EU „kopiert" werden, sondern es bleibt ausreichender Spielraum für eine den Schweizer Bedürfnissen entsprechende KI-Gesetzgebung, z.B. gestützt auf ein materielles, in kooperativer Regelbildung zu verwirklichendes Konzept, das stärker auf die Vermeidung negativer Entwicklungen (Diskriminierung, Manipulation) als auf die Einhaltung risikodeterminierter Standards ausgerichtet ist.[82]

[82] Vgl. vorne Ziff. C.II.