Freiburger Informationspapiere
zum Völkerrecht und Öffentlichen Recht

Ausgabe 3/2021

UNI FREIBURG

Neue Wege zur adaptiven Regulierung von Hochrisiko-KI-Technologien: Schutz von Rechten und Gemeinwohl

Thorsten Schmidt und Silja Vöneky

Silja Vöneky (Hrsg.)

Freiburger Informationspapiere zum Völkerrecht und Öffentlichen Recht

Ausgabe 3/2021

Neue Wege zur adaptiven Regulierung von Hochrisiko-KI-Technologien: Schutz von Rechten und Gemeinwohl

Thorsten Schmidt und Silja Vöneky

V. i. S. d. P.: Prof. Dr. Silja Vöneky
Institut für Öffentliches Recht, Abteilung 2
Rechtswissenschaftliche Fakultät, Alberts-Ludwigs-Universität Freiburg
Werthmannstraße 4, 79098 Freiburg im Breisgau

voelkerrecht@jura.uni-freiburg.de
www.fiponline.de
ISSN 2192-6077

Inhalt

Neue Wege zur adaptiven Regulierung von Hochrisiko-KI-Technologien: Schutz von Rechten und Gemeinwohl

A. Einleitung

Risiken, die von Systemen, Produkten und Dienstleistungen ausgehen, die auf Künstlicher Intelligenz (KI) beruhen oder von dieser gesteuert werden, beruhen auf von Menschen programmierten Algorithmen und wir als Menschen sind daher dafür verantwortlich, wenn sich ein solches Risiko verwirklicht. Dies ist der zentrale Grund, warum Staaten und die internationale Gemeinschaft insgesamt der verantwortungsvollen Steuerung und Regulierung[1] dieser Technologien Priorität einräumen sollten, zumindest wenn mit KI-basierten Produkten oder Dienstleistungen hohe Risiken verbunden sind. Da die Entwicklung neuer KI-gesteuerter Systeme, Produkte und Dienstleistungen in erster Linie durch Unternehmen, also private Akteure, vorangetrieben wird, die stetig neue Produkte und Methoden auf den Markt bringen,[2] sollte es der Kern und das Ziel eines Governance- und Regulierungssystems sein, verantwortungsvolle Innovationen dieser Akteure zwar nicht zu behindern, aber dennoch Risiken durch KI-Systeme für das Gemeinwohl so weit wie möglich zu minimieren und Verletzungen individueller Rechte und Werte – insbesondere von Grund- und Menschenrechten – zu verhindern. Jedenfalls der Schutz derjenigen Menschenrechte, die Teil des Völkergewohnheitsrechts sind, ist dabei eine zentrale Verpflichtung für jeden Staat[3] und hängt weder von dem jeweiligen verfassungsrechtlichen Rahmen des Staates ab noch von den die Staaten bindenden internationalen Verträgen.[4]

In dem vorliegenden Beitrag werden Kernelemente eines Regulierungssystems für KI-gestützte Hochrisikoprodukte und -dienstleistungen erörtert, das die Nachteile von Ansätzen, die auf präventiven Genehmigungs- oder Konsultationsverfahren beruhen und bzw. oder die haftungszentriert sind, vermeidet.[5] Es hat sich in jüngster Zeit in verschiedenen Bereichen gezeigt, dass beide regulatorischen Ansätze

---

1 Unseren Forschungsansatz fassen wir unter dem Begriff „Verantwortliche KI – Responsible AI" zusammen. Im Folgenden konzentrieren wir uns auf einen Regulierungsansatz für Produkte, die auf KI basieren oder durch KI gesteuert werden; wir schließen jedoch – für eine Regulierung *mutatis mutandis* – KI-basierte und KI-gesteuerte Dienstleistungen mit ein. Der vorliegende Beitrag ist eine deutsche Fassung des Beitrages der Autor*innen in Voeneky et al. (Hrsg), *The Cambridge Handbook of Responsible AI*, CUP, 2022 (erscheint demnächst). Silja Vöneky dankt für die Förderung als Senior Fellow durch das FRIAS, Universität Freiburg, im Rahmen der interdisziplinären Saltus-Forschungsgruppe *Responsible AI* von 2018–2021 und die Förderung im Rahmen des Projektes AI-Trust durch die Baden-Württemberg Stiftung (seit 2020). Thorsten Schmidt dankt für die Förderung durch die FRIAS/USIAS Fellowships „Linking Finance and Insurance: Theory and Apllications" 2017/2018 und Ernst Eberlein für seine stetige Unterstützung.

2 Dazu Beckert/Bronk, 'An Introduction to Uncertain Futures' in Beckert/Bronk (Hrsg.), *Uncertain Futures – Imaginaries, Narratives, and Calculation in the Economy* (2018), die dies nur mit dem kapitalistischen System in Verbindung bringen, was jedoch ein zu enger Ansatz ist.

3 Rechtlich verbindliche Menschenrechte enthalten *keine* direkten Verpflichtungen für nichtstaatliche Akteure, wie Unternehmen, aber Staaten sind aufgrund ihrer Pflicht, die Menschenrechte zu achten und zu schützen, auch dazu verpflichtet, Aktivitäten dieser Akteure zu regulieren, wenn diese besondere Risiken bergen; so können bspw. auch Due-diligence-Pflichten im Bereich des Menschenrechtsschutzes bestehen, vgl. Monnheimer, *Due Diligence Obligations in International Human Rights Law* (2021) 13 f., 49 f., 204 f. Im Hinblick auf eine menschenrechtlich begründete Pflicht der Staaten, existentielle und katastrophale Risiken zu vermeiden, die auf Forschung und technologischer Entwicklung beruhen, vgl. Voeneky, 'Human Rights and Legitimate Governance of Existential and Global Catastrophic Risks' in Voeneky/Neuman (Hrsg.), *Human Rights, Democracy and Legitimacy in a World of Disorder* (2018) 151 ff.

4 Es ist hierbei aber nach wie vor umstritten, ob Staaten auch verpflichtet sind, Unternehmen außerhalb ihres Staatsgebietes zu regulieren, vgl. Monnheimer, *Due Diligence Obligations in International Human Rights law* (2021) 307 ff. Dafür Voeneky, 'Human Rights and Legitimate Governance of Existential and Global Catastrophic Risks' in Voeneky/Neuman (Hrsg.), *Human Rights, Democracy and Legitimacy in a World of Disorder* (2018) 155 ff.

5 Siehe unten E.

nicht überzeugend sind, auch wenn sie kombiniert werden, wenn es um die Wahrung zentraler Rechte, wie dem Recht auf Leben und körperliche Unversehrtheit, und Gemeingüter, wie der Umwelt, geht.

In unserem Beitrag legen wir dar, dass – ähnlich wie bei der Regulierung von Risiken im Bankensystem – Risiken, die auf Hochrisiko-KI-Produkten und -Dienstleistungen beruhen, verringert werden können, wenn das Unternehmen, welches das Produkt oder die Dienstleistung entwickelt, herstellt und verkauft, nach der Entwicklung des Produkts, aber vor der Markteinführung einen anteiligen Geldbetrag als finanzielle Garantie in einen Fonds einzahlen muss. Wir argumentieren, dass es für einen Staat, die Europäische Union und auch die internationale Gemeinschaft zentral ist, Regeln einzuführen, die Unternehmen zu dieser Hinterlegung von Kapital verpflichten, um Genehmigungsverfahren und Haftungsnormen zu ergänzen. Zudem soll gezeigt werden, welcher Geldbetrag – auf der Grundlage einer Ex-ante-Bewertung der mit dem risikoreichen KI-Produkt oder der KI-gestützten Dienstleistung verbundenen Risiken – als verhältnismäßig angesehen werden kann, um die technologische Entwicklung nicht zu behindern, sondern um verantwortungsvolle Innovationen und damit das Gemeinwohl zu fördern. Schließlich wird analysiert, welche Art von begleitender Regulierung für die Umsetzung dieses Ansatzes erforderlich ist, die unter anderem eine Risikobewertung des KI-basierten Produkts oder der Dienstleistung durch eine Gruppe unabhängiger Experten und die Erhebung von Daten über die Auswirkungen der Technologie in der Welt vorsieht.

Unser Beitrag soll zeigen, dass und warum die von uns vorgeschlagene Art der „adaptiven" Regulierung mit verschiedenen Rechtssystemen und Verfassungen vereinbar ist. Unser Vorschlag kann daher auch als Vorschlag für einen internationalen Vertrag oder eine internationale *soft law*-Deklaration dienen. Auch wenn die Kommission der Europäischen Union (EU) 2021 einen Vorschlag für eine risikobasierte KI-Regulierung vorgelegt hat,[6] ist es nicht Ziel dieses Beitrages, diesen Entwurf zu erörtern, sondern einen Ansatz aufzuzeigen, der mit diesem, aber auch mit anderen Regulierungsansätzen vereinbar ist, um Lücken zu schließen und Nachteile auszugleichen.

Der Begriff der KI wird im Folgenden weit gefasst. Er erfasst die jüngsten KI-Systeme, die auf komplexen statistischen Modellen der Welt und der Methode des *machine learning* beruhen, besonders selbstlernende Systeme; er umfasst aber auch Systeme der klassischen KI, d. h. der KI-Systeme, die auf Software beruhen, denen grundlegende physikalische Konzepte einprogrammiert sind (*preprogrammed reasoning*), bspw. als eine *symbolic-reasoning engine*.[7] KI kann in ihren unterschiedlichen Ausprägun-

6 Europäische Kommission, Vorschlag für eine Verordnung des Europäischen Parlamentes und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Gesetz über künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union (EU-KI-VO) vom 21. April 2021, COM(2021) 206 final.

7 Vgl. zu diesem Ansatz jüngst Bhatt/Suchan/Varadarajan, 'Commonsense Visual Sensemaking for Autonomous Driving: On Generalised Neurosymbolic Online Abduction Integrating Vision and Semantics' (2021), 299 Artificial Intelligence Journal, <https://doi.org/10.1016/j.artint.2021.103522>. Hier geht es u.a. um das Konzept der Objektpermanenz, also der Idee, dass 'discrete objects continue to exist over time, that they have spatial relationships with one another (such as in-front-of and behind)', dass also Objekte, wie ein Auto, nicht aufhören zu existieren, nur weil sie hinter einem anderen Hindernis verschwinden; vgl. dazu auch 'Is a self-driving car smarter than a seven-month-old?' (*The Economist*, 4. September 2021), <https://www.economist.com/science-and-technology/is-it-smarter-than-a-seven-month-old/21804141>.

gen bereits heute für viele Zwecke genutzt werden (als *multi-purpose tool*) und ist ein sich schnell entwickelndes Schlüsselelement neu entstehender, disruptiver Technologien.[8] Ein Beispiel ist die Verbindung von biologischer Forschung und KI mittels eines KI-gesteuerten, selbstlernenden Programms, mit dessen Hilfe 3D-Formen von Proteinen bestimmt werden können.[9] Darüber hinaus gibt es Anwendungen von KI-Produkten und auch von KI-gestützten Dienstleistungen nicht nur im Bereich der Spracherkennung und Robotik, sondern auch in der Medizin, im Finanzwesen oder bei (halb-)autonomen Fahrzeugen, Schiffen, Flugzeugen, Drohnen etc. KI-getriebene Produkte und diese Dienstleistungen prägen heute schon die unterschiedlichsten Bereiche, von der Kunst bis hin zur Waffenentwicklung.

Es liegt auf der Hand, dass mit diesen KI-basierten Produkten und Dienstleistungen für eine Gesellschaft das Problem zu lösen ist, neue Risiken, die sich aus der Nutzung dieser Produkte und Dienstleistungen ergeben, einzuhegen und zu minimieren, ohne die Chancen und Vorteile der Anwendung zu minimieren oder zu verhindern. Risiken können dabei durch externe Akteure verursacht werden, die KI-gesteuerte Technologie missbrauchen können.[10] Risiken und Schäden können aber auch durch die Unvorhersehbarkeit nachteiliger Auswirkungen entstehen (also durch nicht beabsichtigte sog. Off-target-Effekte),[11] selbst wenn das KI-gesteuerte System für den ursprünglich vorgesehenen Zweck eingesetzt wird. Risiken und Schäden können zudem auf Fehlfunktionen, falschen oder unklaren Eingabedaten, fehlerhafter Programmierung usw. beruhen.[12]

In einigen Bereichen werden KI-Dienstleistungen oder -Produkte zudem neue systemische Risiken verstärken oder schaffen: Bei Finanzanwendungen[13] auf der Grundlage von selbstlernenden KI-Systemen[14] kann KI beispielsweise als kostensparendes und hocheffizientes Instrument in immer größerem Umfang eingesetzt werden. Die Ungewissheit darüber, wie das KI-System in einem unvorhergesehenen

---

8 'AI is relevant to any intellectual task; it is truly a universal field', vgl. Russel/Novig, *Artificial Intelligence – A Modern Approach* (2016), 1; Vöneky, 'Key Elements of Responsible Artificial Intelligence – Disruptive Technologies, Dynamic Law' (2020) 1 OdW 9, 10–11 m. w. N. <https://ordnungderwissenschaft.de/wp-content/uploads/2020/03/2_2020_voeneky.pdf>; Rahwan et al., 'Machine behaviour' (2019) Nature 568, 477–486 (2019) <https://www.nature.com/articles/s41586-019-1138-y>; für die verschiedenen Anwendungsbereiche vgl. auch Wendel, 'The Promise and Limitations of Artificial Intelligence in the Practice of Law' (2019) 72 Oklahoma Law Review 21, 21–24, <https://digitalcommons.law.ou.edu/olr/vol72/iss1/3/>.

9 Dies könnte ein Weg sein, das so genannte Proteinfaltungs-Problem zu lösen, vgl. Callaway, 'It will change everything: DeepMind's AI makes gigantic leap in solving protein structures' (2020) 588 Nature 203, <https://www.nature.com/articles/d41586-020-03348-4>.

10 Brundage et al., 'The Malicious Use of Artificial Intelligence' (2018), 17 f., <https://maliciousaireport.com/>.

11 Zu diesem Begriff im Bereich der Biotechnologie vgl. Zhang et al., 'Off-target Effects in CRISPR/Cas9-mediated Genome Engineering' Molecular Therapy—Nucleic Acids (2015) 4, e264; Costa et al., 'Genome Editing Using Engineered Nucleases and Their Use in Genomic Screening' in Markossian et al. (Hrsg.), *Assay Guidance Manual* (20. November 2017); Reh, *Enhancing Gene Targeting in Mammalian Cells by the Transient Down-Regulation of DNA Repair Pathways* (2010) 22.

12 Vgl. den Beitrag von Wendehorst in Voeneky et al. (Fn. 1).

13 Wie dem Hochfrequenzhandel, „Deep Calibration", „Deep Hedging" und Risikomanagement; unter Hochfrequenzhandel, versteht man den automatisierten Handel von Wertpapieren mit Reaktionszeiten weit unter einer Sekunde; unter „Deep Calibration" versteht man die Anpassung eines Modells an Daten (calibration) mit Hilfe tiefer neuronaler Netze (deep neural networks) und unter „Deep Hedging" das Absichern von Positionen und Derivaten (hedging) mit Hilfe tiefer neuronaler Netze.

14 Um einige Beispiele aus diesem rapide voranschreitendem Bereich zu nennen, vgl. Sirignano/Cont, 'Universal features of price formation in financial markets: perspectives from deep learning' (2019) 19.9 Quantitative Finance 1449–1459; Buehler/Gonon/Teichmann/Wook, 'Deep hedging' (2019) 19.8 Quantitative Finance 1271-129; Horvath/Muguruza/Tomas, 'Deep learning volatility: a deep neural network perspective on pricing and calibration in (rough) volatility models' (2021) 21.1 Quantitative Finance 11–27.

und nicht getesteten Szenario reagiert, führt hierbei zu neuen Risiken und die Einführung neuer Algorithmen oder die Verbesserung bestehender Algorithmen verstärken bereits existierende Risiken, wie zum Beispiel systemische Risiken, operationelle Risiken etc. Gleichzeitig können KI-Systeme das Potential haben, das gesamte Finanzsystem zu destabilisieren,[15] was zu dramatischen Wertverlusten führen kann.

Vorliegend wollen wir zudem auch die Risiken nicht ausblenden, die entstehen, weil es möglich erscheint, dass aufgrund von Lernmechanismen, die sich ohne menschliche Interaktion und ohne regelbasierte Programmierung selbst verbessern können,[16] ein KI-System selbst ein verbessertes KI-System schafft, das die Tür öffnet zu einer Art künstlicher Superintelligenz oder übermenschlicher KI, d. h. zur Singularität (*the Singularity*).[17] Eine übermenschliche KI könnte – jedenfalls nach Ansicht mancher – ein globales katastrophales oder existenzielles Risiko für die Menschheit darstellen.[18] Auch wenn einige Experten dies als irreales Szenario betrachten, prognostizieren andere, dass eine KI mit übermenschlicher Intelligenz bereits bis 2050 Wirklichkeit werden wird.[19] Zudem wird argumentiert, dass eine Intelligenzexplosion zu einem dynamisch instabilen System führen könnte, da intelligentere Systeme es leichter hätten, sich selbst intelligenter zu machen,[20] und da es einen Punkt geben kann, ab dem keine zuverlässigen Vorhersagen mehr gemacht werden könnten.[21] Vor dem Hintergrund dieser unsicheren Zukunft[22] sei es möglich, wird argumentiert, dass Vorhersagen fehlschlagen und Risiken sich schneller als erwartet oder auf unerwartete Weise realisieren können.[23] Daraus folgt, dass übermenschliche KI ein Szenario mit geringer Eintrittswahrscheinlichkeit, aber katastrophalen Auswirkungen (*low probability – high risk/high impact*) darstellt.[24] Staaten und die internationale Gemeinschaft sollten daher auch dieses Risiko nicht *per se* ignorieren, wenn sie über die Regulierung von KI-Produkten und -Dienstleistungen nachdenken und entsprechende Normen vereinbaren.

---

15 Danielsson/Macrae/Uthemann, 'Artificial intelligence and systemic risk' (*Systemic Risk Centre*, 24. Oktober 2019), <https://www.systemicrisk.ac.uk/publications/special-papers/artificial-intelligence-and-systemic-risk>.

16 Siehe LeCun et al., 'Deep Learning' (2015) 521 Nature 436–44 <http://www.nature.com/nature/journal/v521/n7553/full/nature14539.html>.

17 Der Begriff „Singularität" wurde 1993 von *Vernon Vinge* geprägt; er argumentierte, dass '[*w*]*ithin thirty years, we will have the technological means to create superhuman intelligence*', woraus er schlussfolgerte: 'I think it's fair to call this event a singularity ("the Singularity" for the purpose of this paper).' Siehe Vinge, 'The Coming Technological Singularity: How to Survive in the Post-Human Era' in Landis (Hrsg.), *Vision-21: Interdisciplinary Science and Engineering in the Era of Cyberspace* (1993) 11, 12.

18 Anderer Ansicht vgl. Hawking, 'Will Artificial Intelligence Outsmart Us?' in Hawking, *Brief Answers to the Big Questions* (2018), 181; Russel/Novig, *Artificial Intelligence – A Modern Approach* (2016), 1036 ff.; Bringsjord/Govindarajulu, 'Artificial Intelligence' in Zalta (Hrsg.), *The Stanford Encyclopedia of Philosophy* (2020) Abschnitt 9, <https://plato.stanford.edu/archives/sum2020/entries/artificial-intelligence/>; Eden et al., *Singularity Hypotheses: A Scientific and Philosophical Assessment* (2013); Al-Imam/Motyka/Jędrzejko, 'Conflicting opinions in connection with digital superintelligence' (2020) Vol. 9, No. 2 IAES IJ-AI 336–348 <https://ijai.iaescore.com/index.php/IJAI/article/view/20433>; Bostrom, *Superintelligence* (2014), insb. 75 ff.

19 Siehe bspw. Kurzweil, *The Singularity is Near* (2005) 127; für weitere Vorhersagen siehe Bostrom, *Superintelligence* (Fn. 18) 19–21.

20 Yudkowsky, 'Artificial Intelligence as a positive and negative factor in global risk' in Bostrom/Ćirković (Hrsg.), *Global Catastrophic Risks* (2011) 341.

21 Tegmark, 'Will There Be a Singularity within Our Lifetime?' in Brockman (Hrsg.), *What Should We Be Worried About?* (2014) 30, 32.

22 Siehe Beckert/Bronk, 'An Introduction to Uncertain Futures' in Beckert/Bronk (Hrsg.), *Uncertain Futures – Imaginaries, Narratives, and Calculation in the Economy* (2018) 1–38.

23 Wie in Yudkowsky, 'There's no fire alarm for artificial general intelligence' (*Machine Intelligence Research Institute*, 13. Oktober 2017) <https://intelligence.org/2017/10/13/fire-alarm/>.

24 Voeneky, 'Human Rights and Legitimate Governance of Existential and Global Catastrophic Risks' in Voeneky/Neuman (Hrsg.), *Human Rights, Democracy and Legitimacy in a World of Disorder* (2018), 150.

B. Begriffsbestimmungen: Soft Law, Governance und Risiko

Bevor auf die Nachteile und Lücken der derzeitigen Normen für Hochrisiko-KI-gestützte Produkte und solche Dienstleistungen eingegangen wird, ist es sinnvoll, die für diesen Beitrag weiteren, neben der KI, relevanten Schlüsselbegriffe, d. h. *soft law, Governance* und *Risiko* näher zu bestimmen.

Im Zusammenhang mit Governance- und Regulierungsfragen erscheint es wichtig, zwischen rechtlich verbindlichen Regeln einerseits (rechtlichen Normen im engen Sinne) und nicht verbindlichem *soft law* und Kodizes andererseits zu unterscheiden. Nur erstere sind Teil des Rechts *sensu strictu*. Unter dem Begriff internationales *soft law* werden in diesem Beitrag Normen verstanden, die keiner formalen Rechtsquelle des Völkerrechts zuzuordnen und daher nicht unmittelbar rechtsverbindlich sind, aber von Völkerrechtssubjekten (d. h. Staaten, internationalen Organisationen), die grundsätzlich auch Völkerrecht setzen könnten, vereinbart wurden.[25] Aufgrund der Rechtsetzungsbefugnis der Akteure, die internationale *soft law*-Regeln vereinbaren, besitzen diese Regeln dennoch eine besondere normative Kraft und können als relevante Leitlinien für das künftige Verhalten der Staaten gelten, die die Regeln vereinbart haben.[26] Daher sind die Normen des internationalen *soft law* Teil der staatlichen Normsetzung (*top-down*). Sie dürfen nicht mit privater Rechtsetzung durch Unternehmen oder andere private Akteure (*bottom-up*) verwechselt werden.

Letztere, also die private Rechtsetzung, ist *ein* Element der Selbstregulierung privater Akteure, insbesondere von Unternehmen (einschließlich der dort vorhandenen, zahlreichen KI-bezogenen Verhaltenskodizes, wie z. B. der Google AI Principles[27]) Interessengruppen und Nichtregierungsorganisationen.

Diese wird vom Begriff der *Governance* umfasst. Governance fungiert im Folgenden als Oberbegriff und umfasst Normen, die Teil der staatlichen Rechtssetzung im weitesten Sinne darstellen, und auch Normen, die von privaten Akteuren im Wege der Selbstregulierung vereinbart werden.[28] *Regulierung* wird hingegen eng verstanden und umfasst nur staatliche Normsetzung.

Auch der Begriff des *Risikos* ist für die Frage einer adaptiven Regulierung zentral. Obwohl dieser Begriff unterschiedliche Bedeutungen hat und es beispielsweise im Völkerrecht keine allgemein konsentierte Definition gibt – es ist unklar, wie und ob sich ein Risiko von einer Gefahr oder einer Gefährdung unterscheidet[29] –, stützt sich unser Beitrag auf folgende Definition: Demnach ist ein Risiko ein

---

25 Für eine ähnliche Definition siehe Thürer, 'Soft Law', in Wolfrum (Hrsg.), *Max Planck Encyclopedia of Public International Law* (2012) Vol. 9, 271, Rn. 8.

26 Zu den Vor- und Nachteilen von Leitlinien und Standards im Vergleich zur Regulierung siehe Tait/Banda, 'Proportionate and Adaptive Governance of Innovative Technologies: The Role of Regulations, Guidelines, Standards' (*BSI*, 2016) 14.

27 AI Google, 'Artificial Intelligence at Google: Our principles' <https://ai.google/principles/>.

28 Es würde den Rahmen dieses Beitrages sprengen, von Unternehmen oder NGOs ausgearbeitete *„bottom-up"*-Regeln im Bereich der KI zu diskutieren.

29 Siehe Wilson, 'Minimizing Global Catastrophic and Existential Risks from Emerging Technologies through International Law' (2013) 31 Va. Envtl. L.J. 307, 310. Zum Teil wird nicht zwischen Bedrohung, Gefahr und Risiko unterschieden, siehe OECD, *Recommendations of the Council on the Governance of Critical Risks* (6. Mai 2014) <http://www.oecd.org/gov/risk/Critical-Risks-Recommendation.pdf>. Näher Voeneky, 'Human Rights and Legitimate Governance of Existential and Global Catastrophic Risks' in Voeneky/Neuman (Hrsg.), *Human Rights, Democracy and Legitimacy in a World of Disorder* (2018), 140 ff.

unerwünschtes Ereignis, das eintreten kann oder nicht,[30] d. h. ein unerwünschtes hypothetisches zukünftiges Ereignis. Dieser Begriff schließt Situationen der Ungewissheit ein, in denen keine Wahrscheinlichkeiten für den Schadenseintritt bestimmt werden können.[31]

Ein globales katastrophales Risiko wird vorliegend definiert als ein hypothetisches zukünftiges Ereignis, das den Tod einer großen Anzahl von Menschen oder bzw. und die Zerstörung eines großen Teils der Erde verursachen kann; ein existenzielles Risiko wird verstanden als ein hypothetisches zukünftiges Ereignis, welches das Aussterben der Menschen auf der Erde verursachen kann.[32]

Bei KI-gesteuerten Produkten und Dienstleistungen, die mit hohen Risiken verbunden sind, verstehen wir unter Hochrisikoprodukten und -dienstleistungen (*high-risk products and services*) solche, die das Potenzial haben, große Schäden an geschützten individuellen Werten, Rechten oder Interessen oder Allgemeingütern, wie Leben und körperliche Unversehrtheit, Umwelt oder finanzieller Stabilität eines Staates, zu verursachen.[33] Welche spezifischen KI-Systeme bzw. KI-Produkte oder Dienstleistungen solche Hochrisikosysteme darstellen ist umstritten. Die EU-Kommission hat 2021 dazu in ihrem KI-

---

30 Siehe Hansson, 'Risk' in Zalta (Hrsg.), Stanford Encyclopedia of Philosophy <https://plato.stanford.edu/entries/risk/>. In einem quantitativen Sinne kann das Risiko durch Risikomaße definiert werden (sei es unter Verwendung von Wahrscheinlichkeiten oder ohne). Typische Beispiele spezifizieren das Risiko als die Wahrscheinlichkeit eines unerwünschten Ereignisses, das eintreten oder nicht eintreten kann (*value-at-risk*), oder als die Erwartung eines unerwünschten Ereignisses, das eintreten oder nicht eintreten kann (*expected shortfall*). Die Erwartung eines Verlustes, der durch das unerwünschte Ereignis eintritt, ist das Produkt aus seiner Größe in mehreren Szenarien und der Wahrscheinlichkeit dieser Szenarien und gibt somit einen durchschnittlichen Verlust bei Eintritt des unerwünschten Ereignisses an. Es gibt viele Arten von Risikomaßen, siehe zum Beispiel McNeil/Frey/Embrechts, *Quantitative risk management: concepts, techniques and tools-revised edition* (2015). Adaptive Modelle stützen sich auf bedingte Wahrscheinlichkeiten, deren Theorie auf *Bayes* zurückgeht, vgl. Bayes, 'An Essay Towards Solving a Problem in the Doctrine of Chances' (1764) 53 Phil. Transactions 370. Im Bereich des Völkerrechts hat die Völkerrechtskommission (ILC) festgestellt, dass das *'risk of causing significant transboundary harm'* sich auf die kombinierte Wirkung der Eintrittswahrscheinlichkeit eines Unfalls und des Ausmaßes seiner schädigenden Auswirkungen bezieht, siehe ILC, 'Draft Arts. on Prevention of Transboundary Harm from Hazardous Activities' (2001) Vol. II, Part Two, Y.B. Int'l L. Comm. 152.

31 Für einen anderen, engeren Risikobegriff, der Situationen der Unsicherheit ausschließt (*uncertainty versus risk*), siehe Sunstein, *Risk and Reason: Safety, Law and the Environment* (2002) 129; Sunstein, *Worst-Case Scenarios* (2007) 146–47; Posner, *Catastrophe* (2004) 171. Ein Richter des Internationalen Gerichtshofs (IGH) hat dagegen *„uncertain risks"* unter den Risikobegriff gefasst, siehe *Pulp Mills on the River of Uruguay (Argentina v. Uruguay)*, Sep. Op. of Judge Cançado Trindade [2010] ICJ Rep 135, 159, 162; für einen ähnlichen Ansatz (*risk as 'unknown dangers'*) siehe Peel, *Science and Risk Regulation in International Law* (2010) 1.

32 Für eine leicht abweichende Definition, siehe Bostrom, *Superintelligence* (Fn. 17) 115: '[A]n existential risk is one that threatens to cause the extinction of Earth-originating intelligent life or to otherwise permanently and drastically destroy its potential for future desirable development'; vgl. außerdem Bostrom/Ćirković, 'Introduction' in Bostrom/Ćirković (Hrsg.), *Global Catastrophic Risks* (2008), die argumentieren, dass ein globales katastrophales Risiko ein hypothetisches zukünftiges Ereignis ist, das das Potenzial hat *'to inflict serious damage to human well-being on a global scale'*.

33 Für eine Liste von Produkten, die vom Europäischen Parlament (EP) als risikoreiche KI-Produkte eingestuft wurden, vgl. EP, 'Legislative Entschließung des EP vom 20. Oktober 2020 mit Empfehlungen an die Kommission zu dem Rahmen für die ethischen Aspekte von künstlicher Intelligenz, Robotik und damit zusammenhängenden Technologien' (2020/2021(INL)), Rn. 14: „vertritt in diesem Zusammenhang die Auffassung, dass künstliche Intelligenz, Robotik und damit zusammenhängende Technologien als mit hohem Risiko behaftet betrachtet werden sollten, wenn ihre Entwicklung, ihr Einsatz und ihre Nutzung ein erhebliches Risiko der Verletzung oder Schädigung von Einzelpersonen oder der Gesellschaft unter Verletzung der Grundrechte und unter Verstoß gegen die Sicherheitsvorschriften, die im Unionsrecht verankert sind, mit sich bringen; ist der Ansicht, dass bei der Beurteilung der Frage, ob KI-Technologien ein solches Risiko bergen, der Bereich, in dem sie entwickelt, eingesetzt oder genutzt werden, ihre spezifische Verwendung oder ihr spezifischer Zweck sowie die Schwere der zu erwartenden Verletzung oder Schädigung berücksichtigt werden sollten; vertritt die Auffassung, dass das erste und das zweite Kriterium, nämlich der Bereich und die spezifische Verwendung oder der spezifische Zweck, kumulativ betrachtet werden sollten." Abrufbar unter <https://www.europarl.europa.eu/doceo/document/TA-9-2020-10-20_DE.html#sdocta8>.

Verordnungsentwurf (EU-KI-VO) einen Vorschlag vorgelegt.[34] Nach Annex III des Entwurfes dieser EU-KI-VO gelten als Hochrisiko-KI System insbesondere auch menschenrechtsrelevante KI-Systeme, also KI-Systeme, die biometrische Echtzeit-Fernidentifizierung erlauben, KI-Systeme zur Einstellung oder Auswahl natürlicher Personen, KI-Systeme, die Kreditwürdigkeitsprüfungen übernehmen sollen und KI-Systeme, die als Lügendetektor verwendet werden sollen. Weiterhin sind versorgungsrelevante KI-Systeme zu nennen, also KI-Systeme, die im Bereich der Verwaltung und des Betriebs kritischer Infrastrukur eingesetzt werden, sowie rechtsstaatsrelevante KI-Systeme, die Justizbehörden bei der Ermittlung und Auslegung von Sachverhalten und Rechtsvorschriften unterstützen sollen.

Als KI-Hochrisikoprodukte könnten davon abweichend, denkt man an das Schädigungspotential, *prima facie* jedoch auch bestimmte KI-gesteuerte medizinische Produkte – wie die unten erwähnten Gehirn-Computer-Schnittstellen (BCIs) oder KI-gesteuerte Finanzhandelssysteme – gelten. Unter KI-Hochrisikoprodukte fallen zudem jedenfalls auch autonome Waffen, die allerdings einen Sonderfall darstellen.[35] Einen weiteren Sonderfall, sollten sie in Zukunft entwickelt werden, würden KI-Produkte darstellen, die KI-Systeme mit übermenschlicher Intelligenz beinhalten.

C. Defizite der Regulierungsansätze bezüglich Hochrisikotechnologien und -produkten

Zur Beantwortung der drängendsten Regulierungs- und Governance-Fragen im Zusammenhang mit KI-gesteuerten Hochrisikoprodukten und -dienstleistungen wird hier ein Ansatz für eine verantwortungsvolle *Governance* vorgestellt, der die bestehenden Normen in verschiedenen Staaten ergänzen soll. Dieser Ansatz ist weder von einem bestimmten Rechtssystem noch von einem bestimmten verfassungsrechtlichen Rahmen eines Staates abhängig. Er kann sich in verschiedene Rechtskulturen integrieren und in verschiedenen Staaten eingeführt und umgesetzt werden, unabhängig von der Rechtsgrundlage oder dem vorherrschenden Regulierungsansatz. Dies ist wichtig, da KI-gesteuerte Hochrisikoprodukte und -dienstleistungen bereits jetzt und in naher Zukunft in noch größerem Umfang auf den verschiedenen Kontinenten genutzt werden und sich die derzeitigen Regulierungsansätze noch erheblich voneinander unterscheiden oder sogar gänzlich fehlen.

Für die Zwecke dieses Artikels soll die folgende vereinfachende Darstellung allgemeine Regulierungsunterschiede verdeutlichen. So verankern einige Staaten für Hochrisikotechnologien gesetzlich vorrangig einen präventiven Ansatz und legen Genehmigungs- oder ähnliche Verfahren für neue Produkte und Technologien fest; sie beziehen dabei zum Teil das eher risikoaverse sog. Vorsorgeprinzip (*precautionary principle*) ein, so bspw. nach dem Recht der EU.[36] Das Vorsorgeprinzip zielt darauf, die

34 Vgl. oben Fn. 6.

35 Diese werden auch nicht von dem EU-KI-VO-Entwurf erfasst, vgl. dort Art. 2 Abs. 3 EU-KI-VO.

36 Siehe zum Vorsorgeprinzip als Teil des EU-Rechts Art. 191 Abs. 2 Vertrag über die Arbeitsweise der Europäischen Union, ABl. 2016 C 202, 47, sowie die Europäische Kommission, *Mitteilung der Kommission die Anwendbarkeit des Vorsorgeprinzips*, COM(2000) 1 final. Das Vorsorgeprinzip spiegelt sich im internationalen Recht in Grundsatz 15 der Rio-Erklärung wider: 'In order to protect the environment, the precautionary approach shall be widely applied by States according to their capabilities. Where there are threats of serious or irreversible damage, *lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures* to prevent environmental degradation.' (Hervorhebung durch d. Verf.), Konferenz der Vereinten Nationen über Umwelt und Entwicklung 'Rio Declaration on Environment and Development' (14. Juni 1992) UN Doc. A/CONF. 151/26/Rev. 1 Vol. I, 3; vgl. auch Schröder, 'Precautionary Approach/Principle' in Wolfrum (Hrsg.), *Max Planck Encyclopedia of Public International Law* (2012) Vol. 8, 400, Rn. 1–5. In der Philosophie wurde dieser Grundsatz in jüngster Zeit eingehend analysiert und

Staaten zu verpflichten, insbesondere die Umwelt (und ggf. andere Gemeingüter)[37] auch in Fällen wissenschaftlicher Unsicherheit zu schützen. Andere Staaten, wie z. B. die USA, vermeiden dagegen in vielen Bereichen Genehmigungsverfahren insgesamt oder Verfahren mit hohen Zulassungshürden bzw. verzichten auf strikte Implementierung dieser Verfahren. Stattdessen greifen sie primär auf Haftungsregeln zurück, die Verbrauchern oder anderen geschädigten Akteuren die Möglichkeit geben, ein Unternehmen zu verklagen und eine Entschädigung zu erhalten, wenn durch ein Produkt oder eine Dienstleistung ein Schaden entstanden ist.

Beide Regulierungsansätze – die Verankerung von Genehmigungsverfahren oder Haftungsregelungen zur Entschädigung von Konsumenten und anderen Akteuren nach Schadenseintritt – weisen jedoch grundsätzliche Defizite auf, selbst wenn sie miteinander kombiniert werden und nicht in Reinform erscheinen, und sind ergänzungsbedürftig: *Einerseits* ist die Einhaltung der relevanten Standards bei präventiven Genehmigungsverfahren oft schwer durchzusetzen und kann – gerade im Bereich einer neuen Technologie und einem Wissensvorsprung der Produzenten – leicht umgangen werden,[38] wie u. a. der Dieselskandal[39] bei Kraftfahrzeugen in Deutschland gezeigt hat. Wenn Standards umgangen oder nicht implementiert werden, werden Schäden, die durch Produkte nach ihrem Markteintritt verursacht werden, jedoch gerade nicht vermieden; deshalb ist es zusätzlich notwendig, individuelle Rechte, Werte und Gemeinschaftsgüter zu schützen.

Andererseits haben Haftungsregeln, die es den durch ein Produkt oder eine Dienstleistung geschädigten Personen ermöglichen, Schadensersatz zu fordern, den Nachteil, dass unklar ist, inwieweit sie Unternehmen davon abhalten, unsichere Produkte oder Dienstleistungen anzubieten.[40] Die Unternehmen könnten (und scheinen) eher motiviert (zu sein), das geringe und unklare Risiko, in der Zukunft verklagt zu werden, gegen die Chance abzuwägen, durch den Einsatz einer risikoreichen Technologie oder den

---

verteidigt, vgl. Steel, *Philosophy and the Precautionary Principle – Science, Evidence, and Environmental Policy* (2014).

37 Es wird auch vertreten, dass dieser Grundsatz in allen Fällen wissenschaftlicher Unsicherheit und nicht nur zum Schutz der Umwelt angewendet werden sollte, vgl. Phoenix/Treder, 'Applying the Precautionary Principle to Nanotechnology' (*CRN,* 2003/2004) <http://crnano.org/precautionary.htm>; Bostrom, 'Ethical Issues in Advanced Artificial Intelligence' (2003) Abschnitt 2 <https://nickbostrom.com/ethics/ai.html>.

38 Dies hat sich in den letzten Jahren in Bereichen gezeigt, die neue Technologien umfassen, wie bspw. bei den Vorfällen hinsichtlich der Zulassung der Boeing MAX 737, vgl. Sgobba, 'B-737 MAX and the crash of the regulatory system' (2019) 6/4 Journal of Space Safety Engineering 299; Scharper, 'Congressional Inquiry Faults Boeing And FAA Failures For Deadly 737 Max Plane Crashes' (*NPR news*, 16. September 2020) <https://www.npr.org/2020/09/16/913426448/congressional-inquiry-faults-boeing-and-faa-failures-for-deadly-737-max-plane-cr>, entscheidende Fehler im Regulierungsprozess waren 'excessive trust on quantitative performance requirements, inadequate risk-based design process, and lack of independent verification by experts.' Es wird argumentiert, dass ähnliche Fehler oft auftreten können, siehe bspw. Johnston/Rozi, 'The Boeing 737 MAX Saga: Lessons for Software Organizations' (2019) 21(3) Software Quality Professional, 4.

39 Oliver et al., 'Volkswagen emissions scandal exposes EU regulatory failures' (*Financial Times*, 30. September 2015) <https://www.ft.com/content/03cdb23a-6758-11e5-a57f-21b88f7d973f>; Potter, 'EU seeks more powers over national car regulations after VW scandal' (*Reuters*, 27. Januar 2017) <https://www.reuters.com/article/us-volkswagen-emissions-eu-regulations-idUSKCN0V51IO>.

40 Vgl. zu den Nachteilen des US-Haftungsansatzes auch Scherer, 'Regulating Artificial Intelligence' (2016), 29 Harvard Journal of Law & Technology 353 (388, 391).

Verkauf eines risikoreichen Produkts oder einer solchen Dienstleistung in der Gegenwart größere Gewinne zu erzielen. Dies wurde u. a. in den Fällen der Opiatderivatkrise[41] in den USA deutlich.[42] Darüber hinaus wird die Verantwortungslücke weiter vergrößert, wenn Unternehmen, die Produkte oder Dienstleistungen im Bereich der neuen Technologien verkaufen und große Schäden verursacht haben, berechtigte Entschädigungszahlungen durch Vergleiche oder Konkursanmeldungen vermeiden oder begrenzen können.[43]

D. Defizite der KI-Regulierung heute

Betrachtet man die bestehenden spezifischen Regulierungen und Regulierungsansätze für KI-gesteuerte Produkte und (seltener) Dienstleistungen genauer, werden konkrete Defizite auf nationaler, supranationaler und internationaler Ebene deutlich. Es würde den Rahmen dieses Beitrages sprengen, darauf im Detail einzugehen,[44] daher sollen nur einige KI-regulierende Normen und Vorschriften (ohne die oft diskutierten Datenschutzbestimmungen[45]) und deren Lücken und Defizite aufgezeigt werden.

---

41 Die Opiatderivatkrise in den USA zeigt auf erschreckende Weise, dass eine unzureichende Regulierung, die die Verschreibung und den Verkauf eines Hochrisikoprodukts ohne vernünftige Grenzen erlaubt, *ex post* nicht durch eine Haftungsregelung aufgewogen werden kann, selbst wenn geschädigte Akteure Schadensersatz fordern und die Unternehmen verklagen, die den Schaden verursacht haben, vgl. District Court of Cleveland County, *State of Oklahoma, ex rel. Hunter v Purdue Pharma L.P.*, Case No. CJ-2017-816 (2019).

42 Weitere Beispiele sind Erdölbohrungen und -förderungen, da diese Technik als Hochrisikotechnologie angesehen werden kann: British Petroleum (BP) hat bei der Deepwater-Horizon-Katastrophe im Jahr 2010 eine Ölpest im Golf von Mexiko verursacht. Im Jahr 2014 entschied ein US-Gericht, dass BP sich der groben Fahrlässigkeit und des vorsätzlichen Fehlverhaltens gemäß dem US Clean Water Act (CWA) schuldig gemacht hat, da das Unternehmen „recklessly" gehandelt habe, siehe *Oil Spill by the Oil Rig "Deepwater Horizon" in the Gulf of Mexico on April 20, 2010*, Findings of Fact and Conclusion of Law, Phase One Trial, Case 2:19-md-02179-CJB-SS, 4. September 2014, 121–122 <https://www.uscourts.gov/courts/laed/9092014RevisedFindingsofFactandConclusionsofLaw.pdf>. Auch Royal Dutch Shell (Shell) wurde verklagt, weil eine Tochtergesellschaft in Nigeria weitreichende Umweltzerstörungen verursacht hatte; das Berufungsgericht in Den Haag ordnete 2021 an, dass Shell den Bewohnern der Region Entschädigungen zahlen und mit der Reinigung der verseuchten Gewässer beginnen müsse, vgl. Gerechtshof Den Haag, *de Vereniging Milieudefensie v. Royal Dutch Shell PLC and Shell Petroleum Development Company of Nigeria LTD/Shell Petroleum Development Company of Nigeria LTD v. Friday Alfred Akpan*, 29. Januar 2021, ECLI:NL:GHDHA:2021:134 <https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:GHDHA:2021:134>; siehe Peltier/Moses, 'A Victory for Farmers in a David-and-Goliath Environmental Case' (*The New York Times*, 29. Januar 2021) <https://www.nytimes.com/2021/01/29/world/europe/shell-nigeria-oil-spills.html>.

43 Auch dies haben bspw. die Fälle der Opiatkrise in den Vereinigten Staaten gezeigt. Vgl. Hoffmann, 'Purdue Pharma Tentatively Settles Thousands of Opioid Cases' (*New York Times*, 11. September 2019): 'Purdue Pharma (…) would file for bankruptcy under a tentative settlement. Its signature opioid, OxyContin, would be sold by a new company, with the proceeds going to plaintiffs.' <https://www.nytimes.com/2019/09/11/health/purdue-pharma-opioids-settlement.html>. Ein im September 2021 unter Vorbehalt akzeptierter gerichtlicher Vergleich sieht zwar eine Vergleichssumme von ca. 10 Milliarden US-Dollar vor, gewährt jedoch zugleich der Eigentümerfamilie Immunität im Hinblick auf künftige Zivilklagen; vgl. hierzu 'Purdue Pharma Is Dissolved and Sacklers Pay $4.5 Billion to Settle Opioid Claims' (*New York Times,* 1. September 2021). Mehrere US-Bundesstaaten haben angekündigt, gegen die Entscheidung Berufung einzulegen, vgl. 'What is the bankruptcy "loophole" used in the Purdue Pharma settlement?' (*The Economist*, 3. September 2021). Vgl. hierzu auch das Statement des Washingtoner Generalstaatsanwalts B. Ferguson: 'This order lets the Sacklers off the hook by granting them permanent immunity from lawsuits in exchange for a fraction of the profits they made from the opioid epidemic — and sends a message that billionaires operate by a different set of rules than everybody else'.

44 Siehe Vöneky (Fn. 8) 9 ff.

45 Die Europäische Datenschutzgrundverordnung (DSGVO) bezweckt den Schutz personenbezogener Daten natürlicher Personen (Art. 1 Abs. 1 DSGVO) und gilt für die vollständig automatisierte Verarbeitung dieser Daten (Art. 2 Abs. 1 DSGVO), siehe 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, in Kraft getreten am 25. Mai 2018, ABl. 2016 L119, 1. Zu den dort enthaltenen Ansätzen eines „right to explanation" vgl. Wischmeyer, 'Artificial Intelligence and Transparency: Opening the Black Box' in: Wischmeyer/Rademacher, *Regulating Artificial Intelligence* (2019), 75 (89).

I. Regulierung KI-gesteuerter Medizinprodukte

Ein erstes Beispiel ist die Verordnung der EU über Medizinprodukte (MDR),[46] die bestimmte KI-gesteuerte Apps im Gesundheitsbereich und andere KI-gesteuerte Medizinprodukte, etwa im Bereich der Neurotechnologie (wie *Brain-Computer-Interfaces*), regelt.[47] Die geänderte MDR wurde 2017 verabschiedet und ist im Mai 2021[48] in Kraft getreten. Sie sieht nur für Hochrisikoprodukte (bestimmte Produkte der Klasse III) ein sogenanntes Prüfverfahren[49] vor, bei dem es sich um ein Konsultationsverfahren vor der Marktzulassung handelt. Die geänderte MDR normiert unter anderem KI-gesteuerte Medizinprodukte zur Hirnstimulation. Sie fallen unter die MDR, auch wenn *keine* medizinische Zweckbestimmung vorliegt (Anhang XVI, Nr. 6).[50] Damit erfasst die MDR auch Geräte der Verbraucher-Neurotechnologie. Es ist allerdings ein Nachteil, dass unter anderem KI-gesteuerte Neurotechnologien durch die MDR reguliert werden, die MDR aber gerade *kein echtes Genehmigungsverfahren* zur Gewährleistung von Sicherheitsstandards vorsieht, sondern nur das sog. Konsultationsverfahren. In dieser Hinsicht unterscheidet sich die Regulierung KI-gesteuerter Geräte bspw. zur Hirnstimulation in der EU deutlich von den Vorschriften über die Entwicklung von Medikamenten und Impfstoffen in der EU, die wesentlich höhere Sicherheitsstandards vorsehen, einschließlich klinischer Studien mit Versuchen am Menschen.[51] Betrachtet man die Risiken, die durch das Produkt für Menschen und ihre Gesundheit und körperliche Unversehrtheit entstehen können, ist jedoch unklar, warum dieser Regelungsunterschied besteht. Dies gilt umso mehr, wenn die Verwendung der Neurotechnologie bei Menschen keine besondere medizinische Rechtfertigung besitzt, sondern sie als „bloße" Verbrauchertechnologie genutzt wird.

Auf internationaler Ebene fehlt eine Regelung dieser Technologie ganz, so dass bisher kein internationales Abkommen die Staaten verpflichtet, die mit KI-gesteuerter Neurotechnologie, die von Menschen genutzt wird, verbundenen Risiken zu minimieren oder einzuhegen.[52]

---

46 Verordnung (EU) 2017/745 des Europäischen Parlaments und des Rates vom 5. April 2017 über Medizinprodukte, zur Änderung der Richtlinie 2001/83/EG, der Verordnung (EG) Nr. 178/2002 und der Verordnung (EG) Nr. 1223/2009 und zur Aufhebung der Richtlinien 90/385/EWG und 93/42/EWG des Rates, ABl. 2017 L 117, 1. Zudem werden KI-basierte Medizinprodukte unter die neue KI-EU-VO (oben Fn. 6) fallen, vgl. Art. 6 Abs. 11 i. V. m. Annex II (11) und Verordnung (EU) 2017/745. Zu dieser näher unten bei D.III.2.

47 Gemäß Art. 2 MDR bezeichnet „Medizinprodukt" „(…) ein Instrument, einen Apparat, ein Gerät, eine Software, ein Implantat, ein Reagenz, ein Material oder einen anderen Gegenstand, das dem Hersteller zufolge für Menschen bestimmt ist und allein oder in Kombination einen oder mehrere der folgenden spezifischen medizinischen Zwecke erfüllen soll: (…)". Für Ausnahmen siehe Art. 1 Abs. 6 MDR.

48 Die geänderte MDR trat im Mai 2017 in Kraft, aber für Medizinprodukte gilt eine Übergangsfrist von drei Jahren, um die neuen Anforderungen zu erfüllen. Diese Übergangsfrist wurde aufgrund der Covid19-Pandemie bis 2021 verlängert, vgl. Verordnung (EU) 2020/561 des Europäischen Parlaments und des Rates vom 23. April 2020 zur Änderung der Verordnung (EU) 2017/745 über Medizinprodukte hinsichtlich des Geltungsbeginns einiger ihrer Bestimmungen.

49 Vgl. Art. 54, 55 und Art. 106 Abs. 3, Annex IX Abschnitt 5.1, Annex X Abschnitt 6 MDR.

50 Annex XVI lautet: „(…) 6. Geräte zur transkraniellen Stimulation des Gehirns durch elektrischen Strom oder magnetische oder elektromagnetische Felder zur Änderung der neuronalen Aktivität im Gehirn. (…)".

51 §§ 21 ff. Arzneimittelgesetz (AMG), BGBl. 2005 I 3394; Art. 3 Abs. 1 Verordnung (EG) Nr. 726/2004 des Europäischen Parlaments und des Rates vom 31. März 2004 zur Festlegung von Gemeinschaftsverfahren für die Genehmigung und Überwachung von Human- und Tierarzneimitteln und zur Errichtung einer Europäischen Arzneimittel-Agentur, ABL. 2004 L 136, 1.

52 Auch die von der OECD ausgearbeiteten AI Recommendations (siehe unten Fn. 65) sind aufgrund ihres unverbindlichen Charakters in dieser Hinsicht nicht ausreichend, vgl. hierzu im Detail Vöneky (Fn. 8), 17 f.
Zumindest einige Staaten wie Chile und Frankreich versuchen jedoch, diesen Bereich der KI-Technologie zu regulieren. So soll im Rahmen der chilenischen Verfassungsreform der derzeitige Art. 19 der *Carta Fundamental* um einen zweiten Absatz ergänzt werden, der die psychische und physische Integrität vor technischer Manipulation schützt (vgl. zum aktuellen Stand des Gesetzgebungsverfahrens: Cámara dediputadas y diputados, *Boletín No. 13827-19*,

II. Regulierung autonomer Fahrzeuge

Ein zweites Beispiel für eine sektorspezifische Regulierung für KI-gesteuerte Produkte, die bereits in Kraft ist und deutliche Nachteile aufweist, sind Normen zu semi-autonomen Fahrzeugen. In Deutschland wurde das entsprechende nationale Gesetz bereits 2017[53] geändert, um neue automatisierte KI-basierte Fahrsysteme einzubeziehen.[54] Der Gesetzgebungsprozess wurde abgeschlossen, bevor eine zuständige Bundesethikkommission ihren Bericht veröffentlicht hatte. [55] Der relevante § 1a Abs. 1 StVG besagt, dass der Betrieb eines Kraftfahrzeugs mit einer hoch- oder vollautomatisierten (aber nicht autonomen)[56] Fahrfunktion zulässig ist, sofern die Funktion *bestimmungsgemäß* genutzt wird:

> „Der Betrieb eines Kraftfahrzeugs mittels hoch- und vollautomatisierter Fahrfunktion ist zulässig, wenn die Funktion *bestimmungsgemäß verwendet* wird."[57]

Es ist bemerkenswert, dass die Bedeutung des Ausdrucks „bestimmungsgemäß" nicht durch das Gesetz selbst oder eine Rechtsverordnung festgelegt ist, sondern von dem Automobilunternehmen definiert werden kann.[58] Daher ist diese Normsetzung ein Verweis auf die private Standardsetzung von Unternehmen, die KI-gesteuerte Autos herstellen und verkaufen. Die Bestimmung enthält damit eine Öffnungsklausel zur Selbstregulierung, ist aber selbst nicht hinreichend klar.[59] Dies ist ein Beispiel für einen Regulierungsansatz, der im Bereich eines hochriskanten KI-Produkts keine hinreichenden Standards vorgibt. Es kann argumentiert werden, dass dies gegen das Rechtsstaatsprinzip als Teil des Grundgesetzes verstößt, [60] das besagt, dass rechtliche Regeln klar und verständlich für diejenigen sein müssen, die sie betreffen.[61]

<https://www.camara.cl/legislacion/ProyectosDeLey/tramitacion.aspx?prmID=14384&prmBOLETIN=13827-19>). Überdies ist die Implementierung spezifischer Neurorechte geplant, vgl. das Vorhaben *Boletín No. 13828-19*. Das Anfang August 2021 in Kraft getretene französische Bioethik-Gesetz (Loi n° 2021-1017 du 2 août 2021 relative à la bioéthique) erlaubt den Einsatz von brain-imaging Techniken nur zu Medizin- und Forschungszwecken (Art. 18 und 19), vgl. <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000043884384/>.

53 Art. 1 Achtes Gesetz zur Änderung des Straßenverkehrsgesetzes (8. StVGÄndG), BGBl. 2017 I 1648.

54 §§ 1a, 1b und 63 StVG. Für einen Überblick über die wichtigsten internationalen, europäischen und nationalen Vorschriften für autonome oder automatisierte Fahrzeuge, vgl. Böning/Canny, 'Easing the brakes on autonomous driving' (FIP 1/2021), <http://www.jura.uni-freiburg.de/de/institute/ioeffr2/downloads/online-papers/FIP_2021_01_Boening-Canny_AutonomousDriving_Druck.pdf>.

55 Deutschland, Bundesministerium für Verkehr und digitale Infrastruktur, Ethikkommission, *Automated and Connected Driving* (*BMVI*, Juni 2017), <https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.html>.

56 In Deutschland wird derzeit ein neues Gesetz zu vollautonomen Autos erarbeitet, vgl. Gesetzentwurf der Bundesregierung, Drs. 19/27439, vom 9. März 2021: Entwurf eines Gesetzes zur Änderung des Straßenverkehrsgesetzes und des Pflichtversicherungsgesetzes (Gesetz zum autonomen Fahren), <https://dip21.bundestag.de/dip21/btd/19/274/1927439.pdf>.

57 § 1a Abs. 1 StVG.

58 Böning/Canny, 'Easing the brakes on autonomous driving' (Fn. 54).

59 Dies dürfte auch trotz der Vorgabe des Gesetzgebers gelten, dass die Beschreibung des Verwendungszwecks und des Automatisierungsgrads „eindeutig" sein soll, vgl. Deutscher Bundestag, 'Entwurf eines … Gesetzes zur Änderung des Straßenverkehrsgesetzes' (2017) Drucksache 18/11300, 20: „Die Systembeschreibung des Fahrzeugs muss über die Art der Ausstattung mit automatisierter Fahrfunktion und über den Grad der Automatisierung unmissverständlich Auskunft geben, um den Fahrer über den Rahmen der bestimmungsgemäßen Verwendung zu informieren." Abrufbar unter <https://dip21.bundestag.de/dip21/btd/18/113/1811300.pdf>.

60 Grundgesetz für die Bundesrepublik Deutschland (GG), BGBl. 1949 I 1, letzte Änderung 29. September 2020, BGBl. 2020 I 2048.

61 Grzeszick, 'Art. 20' in Maunz/Dürig (Hrsg.), *Grundgesetz-Kommentar* (August 2020), Rn. 99.

III. Allgemeine KI-Vorschriften und -Grundsätze – Europäische Regulierung und internationales *soft law*

Es stellt sich die Frage, ob die oben erwähnten Schwächen und Lücken auf nationaler und europäischer Ebene in bestimmten Bereichen der KI-Regulierung durch Regeln des internationalen Rechts (1) oder auf europäischer Ebene (2) geschlossen werden können.

1. Internationale Regulierung? Internationales *soft law*!

Es gibt bisher keinen internationalen Vertrag, der KI-Systeme, -Produkte oder -Dienstleistungen regelt. Ein solcher wird auch nicht verhandelt – zu unterschiedlich sind noch die Interessen der einzelnen Staaten, die ihre eigenen Unternehmen und ihre nationalen Interessen im Blick haben. Diese Ausgangslage unterscheidet sich von der bestehenden Regulierung im Bereich der Biotechnologie, einer vergleichbaren innovativen, aber potentiell disruptiven Technologie. Biotechnologie ist international durch das Cartagena-Protokoll als Vertrag verbindlich reguliert und auf das Vorsorgeprinzip[62] hin ausgerichtet. Da über 170 Staaten Vertragsparteien sind,[63] kann man von einer nahezu universalen Regulierung sprechen, auch wenn die USA als wichtiger Akteur nicht gebunden sind. Selbst in eindeutigen Hochrisikobereichen der KI-Entwicklung, wie der Entwicklung und dem Einsatz autonomer Waffen, fehlt jedoch bisher ein internationaler Vertrag. Auch dies steht im Gegensatz zu anderen Bereichen der Entwicklung hochriskanter Waffen, wie etwa biologischer Waffen.[64]

Es findet sich jedoch zumindest internationales *soft law*, das erste allgemeine Grundsätze für KI-Systeme auf internationaler Ebene formuliert. So hat der Rat für künstliche Intelligenz der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) 2019 KI-Empfehlungen, die *OECD AI Recommendations*, ausgearbeitet. [65] Über 50 Staaten haben erklärt, sich an diese Grundsätze zu halten, darunter Staaten, die für die KI-Forschung und -Entwicklung besonders relevant sind, wie die USA, das Vereinigte Königreich, Japan und Südkorea. In den OECD-Empfehlungen zur KI werden fünf komplementäre, wertebasierte Prinzipien genannt und ausgeführt.[66] Diese sind inklusives Wachstum, nachhaltige Entwicklung und Lebensqualität (IV. 1.1.), menschenzentrierte Werte und Fairness (IV. 1.2.), Transparenz und Erklärbarkeit (IV. 1.3.), Robustheit und Sicherheit (i. S. v. *security and safety*) (IV. 1.4.) und Rechenschaftspflichten (IV. 1.5.). Darüber hinaus sollten KI-Akteure – d. h. diejenigen, die eine aktive Rolle im Lebenszyklus von KI-Systemen spielen, einschließlich Organisationen und Einzelpersonen, die KI einsetzen oder betreiben[67] – Menschenrechte und demokratische Werte achten (IV. 1.2. lit. a). Dazu gehören Freiheit, Würde und Selbstbestimmung, Schutz der Privatsphäre und Datenschutz,

62 Dazu oben Fn. 36.

63 Cartagena-Protokoll über die biologische Sicherheit der Konvention über biologische Vielfalt (verabschiedet 29. Januar 2000, in Kraft getreten am 11. September 2003) 2226 UTNS 208.

64 Konvention über das Verbot der Entwicklung, Herstellung und Lagerung bakteriologischer (biologischer) Waffen und Toxinwaffen sowie über die Vernichtung solcher Waffen (verabschiedet am 10. April 1972, in Kraft getreten am 26. März 1975) 1015 UNTS 163.

65 OECD, *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449, beschlossen am 22. Mai 2019, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>.

66 Ein KI-System wird definiert als 'a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.', siehe OECD AI Recommendations (Fn. 65), I.

67 OECD AI Recommendations (Fn. 65), ibid.

Nichtdiskriminierung und Gleichbehandlung, Vielfalt, Fairness, soziale Gerechtigkeit und international anerkannte Arbeitsrechte.

Allerdings sind diese Prinzipien ausgesprochen weich formuliert ('should respect'). Selbst die OECD-KI-Empfehlung zu Transparenz und Erklärbarkeit (IV. 1.3.) hat wenig Substanz. Sie besagt, dass

> '[…] [AI Actors][68] should provide meaningful information, appropriate to the context, and consistent with the state of art: […]
>
> (iv) to enable those adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information on the factors, and the logic that served as the basis for the prediction, recommendation or decision.'

Geht man davon aus, dass Diskriminierung und ungerechtfertigte Vorurteile zu den Hauptproblemen von KI-Systemen gehören,[69] scheint die Forderung nach einem „Risikomanagement-Ansatz" (*systematic risk management approach*) zur Lösung dieser Probleme, wie in Grundsatz IV. 1.4 (c) der OECD-Empfehlungen für KI benannt,[70] als Standard für die Sorgfaltspflicht der KI-Akteure ebenfalls nicht ausreichend zu sein.

Darüber hinaus wird in den KI-Empfehlungen der OECD nicht auf eine rechtlich verankerte Haftung der Unternehmen oder eine rechtliche Verantwortung eingegangen. KI-Akteure „sollten rechenschaftspflichtig sein"[71]. Dies bedeutet, dass diese Akteure Bericht erstatten und bestimmte Informationen über ihre Tätigkeiten bereitstellen sollten, um „das ordnungsgemäße Funktionieren von KI-Systemen" und „die Einhaltung der oben genannten Grundsätze" zu gewährleisten (IV. 1.5). Dies impliziert jedoch gerade keine rechtliche Verpflichtung zur Erfüllung dieser Standards und keine rechtliche Haftung, wenn ein Akteur diesen Anforderungen nicht entspricht.

Schließlich wird in den KI-Empfehlungen der OECD nicht die Verantwortung der Regierungen für den Schutz der Menschenrechte im Bereich der KI betont. Sie enthalten lediglich fünf Empfehlungen für die politischen Entscheidungsträger der Staaten (Abschnitt 2), die in der nationalen Politik und der internationalen Zusammenarbeit im Einklang mit den oben genannten Grundsätzen umgesetzt werden sollen. Dazu gehören Investitionen in KI-Forschung und -Entwicklung (V. 2.1), die Förderung eines digitalen Ökosystems für KI (V. 2.2), die Gestaltung und Ermöglichung eines politischen Umfelds für KI (V. 2.3), der Aufbau personeller Kapazitäten und die Vorbereitung auf die Transformation des Arbeitsmarktes (V. 2.4) sowie die internationale Zusammenarbeit für vertrauenswürdige KI (V. 2.5). Selbst wenn man sich auf die OECD-Empfehlungen zur KI stützen will, bleibt also unklar, welche staatlichen Verpflichtungen sich aus den Menschenrechten in Bezug auf KI-Governance ergeben.

Darüber hinaus wird das Problem, wie mit der Herausforderung einer möglichen Entwicklung einer übermenschlichen KI, die zwar eine geringe Wahrscheinlichkeit hat, aber mit hohem Risiko verbunden ist, umgegangen werden soll, in den KI-Empfehlungen der OECD nicht einmal erwähnt.[72]

---

68 KI-Akteure sind dabei 'those who play an active role in the AI system lifecycle, including organisations and individuals that deploy or operate AI', siehe OECD AI Recommendations (Fn. 65), I.

69 Siehe Datenethikkommission, *Gutachten* (2019), 194, <https://www.bmjv.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_EN_lang.pdf?__blob=publicationFile&v=3>.

70 'AI actors should, based on their roles, the context, and their ability to act, apply a *systematic risk management approach* to each phase of the AI system lifecycle on a continuous basis to address risks related to AI systems, including privacy, digital security, safety and bias.' Vgl. IV. 1.4. c) OECD AI Recommendations (Fn. 65).

71 'AI actors should be accountable for the proper functioning of AI systems and for the respect of the above principles, based on their roles, the context, and consistent with the state of art.' Vgl. IV. 1.5. OECD AI Recommendations (Fn. 65).

72 Siehe oben A.

2. Entwurf einer Europäischen KI-Verordnung

Der oben genannte[73] von der EU-Kommission vorgelegte Entwurf einer KI-VO, die harmonisierte Vorschriften für KI festlegen soll, stellt eine erste umfassende Regulierung von Hochrisiko-KI-Systemen für die Mitgliedstaaten der EU in Aussicht. Der Entwurf sieht Kriterien für die Konstruktion und die Entwicklung solcher Systeme vor, ohne sich auf bestimmte Sektoren zu beschränken. Dabei wird ein risikobasierter präventiver Regulierungsansatz verfolgt.

Der Begriff des KI-Systems wird dabei weit verstanden (Art. 3 Abs. 1 EU-KI-VO).[74] Darüber hinaus richtet sich die Verordnung an alle Anbieter[75], die „KI-Systeme in der Union in den Verkehr bringen oder in Betrieb nehmen" (Art. 2 Abs. 1 lit. a i. V. m. Art. 3 Abs. 2 EU-KI-VO), sowie an alle Nutzer von KI-Systemen, die sich in der Union befinden (Art. 2, Art. 3 Abs. 2 EU-KI-VO).

Welche Arten von KI-Systemen hochriskant sind, wird in den Art. 6 und 7 EU-KI-VO allgemein festgelegt und in Anhang II und Anhang III des Entwurfs näher ausgeführt. Die oben erwähnte Liste in Anhang III kann von der EU-Kommission bei Bedarf ergänzt werden, so dass es möglich ist, auch auf kurzfristige Entwicklungen zu reagieren und somit eine gewisse Flexibilität bei der Regulierung zu wahren.[76]

Die KI-VO berücksichtigt insbesondere die möglichen negativen Auswirkungen des Einsatzes von KI-Systemen im Hinblick auf den Schutz der Menschenrechte und betont dabei zentrale Werte und Rechte wie den Schutz der Menschenwürde sowie das Recht auf Leben und körperliche Unversehrtheit. Daher sind gem. Art. 5 EU-KI-VO bestimmte Nutzungen bei KI-Systemen gänzlich verboten, insbesondere wenn sie von staatlichen Behörden eingesetzt werden. Dazu gehört unter anderem der Einsatz bestimmter KI-Systeme, die Techniken zur „unterschwelligen Beeinflussung außerhalb des Bewusstseins einer Person" einsetzen, sofern dies dieser Person oder einer anderen Person einen physischen oder psychischen Schaden zufügt oder zufügen kann.

Das Gleiche gilt, wenn KI-Praktiken Personen Schaden zufügen, indem sie die Schwächen einer bestimmten Gruppe aufgrund ihres Alters oder ihrer Behinderung ausnutzen. Auch der Einsatz eines biometrischen Fernidentifizierungssystems im Rahmen der Strafverfolgung ist grundsätzlich verboten,

---

73 Vgl. Fn. 6.

74 Siehe zum Vergleich die Definition der OECD in Fn. 66. Unter einem KI-System versteht der europäische Entwurf „eine Software, die mit einer oder mehreren der in Anhang I aufgeführten Techniken und Konzepte entwickelt worden ist und im Hinblick auf eine Reihe von Zielen, die vom Menschen festgelegt werden, Ergebnisse wie Inhalte, Vorhersagen, Empfehlungen oder Entscheidungen hervorbringen kann, die das Umfeld beeinflussen, mit dem sie interagieren", siehe Art. 3 Abs. 1; Anhang I führt hierzu näher aus: „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."

75 Unter den Begriff des „Anbieters" fallen nicht nur Private, sondern sämtliche natürlichen oder juristischen Personen, Behörden, Einrichtungen oder sonstige Stellen, vgl. Art. 3 Abs. 2 EU-KI-VO.

76 Allerdings könnte diese Flexibilität auf Kosten der demokratischen Legitimation gehen, die in derartigen Verfahren kaum noch gewährleistet werden kann. Die sich aus der Problematik ergebenden Fragen können hier nicht näher erörtert werden; festzuhalten bleibt daher lediglich, dass es unklar ist, ob der Vorteil, mehr Flexibilität bei der Regulierung einer schnelllebigen Technologie zu gewinnen, den angesprochenen Nachteil der fehlenden demokratischen Legitimation tatsächlich überwiegt.

auch wenn Art. 5 des Entwurfs für diesen Fall Ausnahmen vorsieht. Ferner weisen die Transparenzpflichten einen deutlichen Menschenrechtsbezug auf; die Transparenzpflicht in Art. 52 EU-KI-VO findet etwa Anwendung, wenn das KI-System dazu bestimmt ist, mit natürlichen Personen zu interagieren und kann als Schutz der Autonomie und Menschenwürde verstanden werden. Auch Art. 62 EU-KI-VO nimmt ausdrücklich auf Individualrechte Bezug und bestimmt die Pflicht, „schwerwiegende Vorfälle oder Fehlfunktionen (...) [zu melden], die einen Verstoß gegen die Bestimmungen des Unionsrechts zum Schutz der Grundrechte darstellen".

Abgesehen von den gerade erwähnten Verboten und Pflichten muss jedes Hochrisiko-KI-System spezifische Anforderungen erfüllen (vgl. Art. 8 EU-KI-VO). Diese Anforderungen sehen unter anderem vor, dass Risikomanagementsysteme eingerichtet und aufrechterhalten werden müssen (Art. 9 EU-KI-VO) und dass Trainingsdatensätze bestimmten Qualitätskriterien entsprechen müssen (Art. 10 EU-KI-VO). Darüber hinaus werden Kriterien für die technische Dokumentation von Hochrisiko-KI-Systemen festgelegt (Art. 11 und Anhang IV EU-KI-VO) und es muss bei der Konzeption der KI-Systeme sichergestellt werden, dass sie in der Lage sind, Ereignisse automatisch aufzuzeichnen, dass ihr Betrieb „hinreichend transparent" ist (Art. 12 und 13 EU-KI-VO) und dass sie von Menschen wirksam beaufsichtigt werden können (Art. 14 EU-KI-VO).

Eine weitere Besonderheit ist, dass der Entwurf nicht nur für Entwickler und Anbieter von Hochrisiko-KI-Systemen (Art. 16 ff. EU-KI-VO) sowie deren Importeure und Händler (Art. 26 und 27 EU-KI-VO) Pflichten vorsieht, sondern auch für die Nutzer. Zu den Nutzern zählen Unternehmen, die als Kreditinstitute hochriskante KI-Systeme einsetzen (Art. 3 Abs. 4 i. V. m. Art. 28 und 29 EU-KI-VO). Diese müssen beispielsweise dafür sorgen, dass „die Eingabedaten der Zweckbestimmung des Hochrisiko-KI-Systems entsprechen"; zudem werden Überwachungs- und Protokollpflichten festgelegt (Art. 29 EU-KI-VO).

Da der Entwurf keine näheren Haftungsvorschriften enthält, ist er ein klares Beispiel für einen präventiven Regulierungsansatz.[77] Allerdings sieht der Entwurf kein Zulassungsverfahren vor, sondern lediglich ein Konformitätsbewertungsverfahren (Art. 48 und Anhang V EU-KI-VO), das entweder auf einer internen Kontrolle beruht (Anhang VI EU-KI-VO) oder die Einschaltung einer notifizierten Stelle vorsieht (Art. 19 und 43, Anhang VII EU-KI-VO). Die notifizierten Stellen müssen die Konformität von Hochrisiko-KI-Systemen überprüfen (Art. 33 EU-KI-VO). Es ist jedoch Sache der EU-Mitgliedstaaten, eine solche notifizierende Behörde (Art. 30 EU-KI-VO) gemäß den Anforderungen des Entwurfs einzurichten, und eine notifizierte Stelle darf zudem bestimmte Aufgaben an Unterauftragnehmer vergeben (Art. 34 EU-KI-VO). Die EU-Kommission kann als Aufsichtsbehörde Fälle untersuchen, in denen „begründete Zweifel daran bestehen, dass eine notifizierte Stelle die in Art. 33 festgelegten Anforderungen erfüllt" (Art. 37 EU-KI-VO). Außerdem können Ausnahmen vom Konformitätsbewertungsverfahren genehmigt werden, wenn „außergewöhnliche Gründe" vorliegen; derartige Gründe umfassen gem. Art. 47 EU-KI-VO solche „der öffentlichen Sicherheit, des Schutzes des Lebens und der Gesundheit von Personen, des Umweltschutzes" und sogar (sic!) „des Schutzes wichtiger Industrie- und Infrastrukturanlagen".

77 Vgl. zu dieser Unterscheidung oben III. Eine genauere Darstellung findet sich bei Wendehorst, in Voeneky et al. (Hrsg), *The Cambridge Handbook of Responsible AI*, CUP, 2022 (erscheint demnächst).

Letztendlich obliegen jedoch in erster Linie den Anbietern die durch die Verordnung vorgesehenen Verpflichtungen, wie zum Beispiel die Dokumentationspflichten (Art. 50 EU-KI-VO), die Beobachtungspflichten nach dem Inverkehrbringen (Art. 61 EU-KI-VO) oder die Pflicht zur Registrierung des Systems in der EU-Datenbank (Art. 51 und 60 EU-KI-VO).

Die Verordnung sieht nur Geldbußen „bis zu" einem bestimmten Betrag vor (zwischen 10 000 000 und 30 000 000 EUR), und es ist Sache der Mitgliedstaaten, über die Schwere des Verstoßes zu entscheiden. Darüber hinaus sind die gegen die Organe, Einrichtungen und sonstige Stellen der Union verhängbaren Geldbußen wesentlich niedriger (vgl. Art. 72 EU-KI-VO: bis zu 250 000 EUR bzw. bis zu 500 000 EUR).[78]

Eine ausführlichere Analyse und Bewertung des EU-KI-Verordnungsentwurfs muss einem späteren Beitrag vorbehalten bleiben.[79] Auffällig ist jedoch, dass die Regulierung von Hochrisiko-KI kein strenges Zulassungsverfahren beinhaltet. Somit legt diese Verordnung im Hinblick auf Hochrisiko-KI-Systeme Maßnahmen fest, die im Vergleich zur Regulierung anderer risikoreicher Produkte, wie etwa der Entwicklung von Medikamenten und Impfstoffen in der EU, deutlich niedriger ausfallen. Ohne eine überzeugende Regulierung der mit der KI zusammenhängenden Schadensersatz- und Haftungsprobleme ist es daher fraglich, ob die großen Risiken, die KI-Systeme mit sich bringen können, durch diese Verordnung ausreichend gemildert werden können.

IV. Zwischenfazit

Aus den Ausführungen oben folgt *erstens*, dass es bei der Regulierung neu entstehender Technologien und insbesondere von KI-Systemen große Lücken und Defizite gibt, auch wenn zumindest in einigen Bereichen Normen bestehen. *Zweitens* gibt es keine kohärente, allgemeine oder universelle internationale Regulierung von KI oder KI-Produkten; die EU hat zwar einen Vorschlag für eine allgemeine KI-Regulierung vorgelegt, dieser ist allerdings bisher noch nicht in Kraft getreten und hat, auch wenn er sinnvolle Regelungen enthält, als rein präventiver Ansatz noch Lücken und Defizite.

Es besteht jedoch weitgehende Einigkeit, auch außerhalb der EU, darüber, dass zumindest für Hochrisiko-KI-Produkte und -Dienstleistungen eine angemessene Regulierung erforderlich ist. Betrachtet man die zahlreichen Bereiche, in denen KI-gesteuerte Systeme derzeit eingesetzt werden und in Zukunft eingesetzt werden könnten, sowie die mit diesen Systemen und Produkten verbundenen Vorteile und Risiken, überrascht es nicht, dass selbst Unternehmer, die KI-gesteuerte Produkte verkaufen, die Notwendigkeit der Regulierung von KI-Systemen, -Produkten und -Dienstleistungen gefordert haben.[80] Die Anfälligkeit automatisierter Handelssysteme auf dem Finanzmarkt mag nur als ein Beispiel dienen,

78 Vgl. zur Durchsetzung Art. 63 ff.; zu den Sanktionen Art. 71.

79 Für eine Darstellung vgl. Burri, in Voeneky et al. (Hrsg), *The Cambridge Handbook of Responsible AI*, CUP, 2022 (erscheint demnächst).

80 *Bill Gates*, *Sundar Pichai* und *Elon Musk* haben bspw. die Regulierung von KI gefordert, s. Pichai, 'Why Google thinks we need to regulate AI' (*Financial Times*, 20. Januar 2020) <https://www.ft.com/content/3467659a-386d-11ea-ac3c-f68c10993b04>; Mack, 'Bill Gates says you should worry about Artificial Intelligence' (*Forbes*, 28. Januar 2015) <https://www.forbes.com/sites/ericmack/2015/01/28/bill-gates-also-worries-artificial-intelligence-is-a-threat/>; Gibbs, 'Elon Musk: Regulate AI to Combat "Existential Threat" before It's Too Late' (*The Guardian*, 17. Juli 2017) <https://www.theguardian.com/technology/2017/jul/17/elon-musk-regulation-ai-combat-existential-threat-tesla-spacex-ceo>.

um die enormen Auswirkungen intelligenter Systeme zu verdeutlichen: 2010 wurden automatische Verkäufe ausgelöst, wodurch US-Aktien im Wert von fast 1000 Mrd. USD für mehrere Minuten verschwanden.[81]

Daher stimmen wir denjenigen zu, die argumentieren, dass hochriskante KI-Produkte und -Dienstleistungen, auch weil letztere oft vernachlässigt werden, als neu entstehende Technologien reguliert werden müssen. Unserer Ansicht nach ist eine verantwortungsvolle, d. h. angemessene und möglichst robuste Regulierung hochriskanter KI-Produkten und -Dienstleistungen bereits heute erforderlich, um zu verhindern, dass wir diese erst regulieren, wenn bereits Schäden aufgetreten sind und die Regulierung damit, auch zum Schutz der Menschenrechte und zum Schutz des Gemeinwohls, zu spät kommt.[82]

E. Adaptive Regulierung KI-gesteuerter Hochrisikoprodukten und -dienstleistungen

I. Ein neuer Weg

Wir argumentieren in diesem Beitrag, dass ein neuer Ansatz zur Regulierung KI-gesteuerter Produkten erforderlich ist, um die oben genannten Defizite auf nationaler und internationaler Ebene zu vermeiden oder auszugleichen. Ziel ist es, durch den hier vorgeschlagenen Regulierungsansatz präventive Genehmigungsverfahren zu ergänzen und gleichzeitig die Lücken haftungsbasierter Ansätze verschiedener Rechtssysteme zu schließen. Dieser Ansatz soll global anwendbar sein und im nationalen, supranationalen und bzw. oder internationalen Recht verankert werden können. Unser Vorschlag zielt auf ein Regulierungssystem ab, das adaptiv in dem Sinne ist, dass es flexibel und risikosensitiv ist und einen Anreiz zur Senkung und Bewertung von Risiken durch diejenigen Unternehmen bietet, die KI-gesteuerte Produkte entwickeln und verkaufen. Der Kern des Vorschlags besteht darin, dass ein Betreiber oder Unternehmen einen anteiligen Geldbetrag (im Folgenden als *regulatorisches Kapital* bezeichnet) als finanzielle Sicherheit für künftige Schäden zahlen muss, *bevor* ein KI-basiertes Produkt oder eine KI-basierte Dienstleistung auf den Markt kommt. Um eine Überregulierung zu vermeiden, konzentrieren wir uns auf KI-basierte Produkte, die zu einer Klasse von Hochrisikoprodukten und -dienstleistungen gehören. Im Folgenden wird, wegen der besonderen Gefahren dieser Hochrisiko-KI-Systeme, auch ein Regulierungsrahmen für die mögliche Entwicklung einer sog. übermenschlichen KI erörtert.

Der Sonderfall autonomer Waffen, die ebenfalls KI-Hochrisikoprodukte sind, soll zudem erwähnt werden. Im Hinblick auf die spezifischen Probleme der Entwicklung (halb-)autonomer Waffen argumentieren viele Autoren und Staaten, dass ein Verbot dieser Waffen aus ethischen und rechtlichen Gründen zwingend sei.[83] Dies könnte bedeuten, dass jede Art von adaptiver Regulierung, die hier vorgeschlagen wird, nicht diskutiert werden sollte, da eine solche Regulierung den Markteintritt solcher Waffen

---

81 Vgl. Mackenzie/van Duyn, '"Flash crash" was sparked by single order' (*Financial Times*, 1. Oktober 2010) <https://www.ft.com/content/8ee1a816-cd81-11df-9c82-00144feab49a>. Dazu auch der Beitrag von Paul in Voeneky et al. (Hrsg), *The Cambridge Handbook of Responsible AI*, CUP, 2022 (erscheint demnächst).

82 Dazu schon *Elon Musk* (vgl. Fn. 80).

83 Wie bspw. die Regierungen Österreichs, Brasiliens und Chiles, Working Paper Gruppe der Regierungsexperten des Übereinkommens über konventionelle Waffen über tödliche autonome Waffensysteme, 'Proposal for a mandate to negotiate a legally-binding instrument that addresses the legal, humanitarian and ethical concerns posed by emerging technologies in the area of lethal autonomous weapons systems (LAWS)', 8 August 2018, CCW/GGE.2/2018/WP.7, <https://undocs.org/CCW/GGE.2/2018/WP.7>. Vgl. zudem den am 20. November 2020 von den Grünen und der ÖVP

rechtfertigen könnte. Unser Argument für die Einbeziehung (halb-)autonomer Waffen in die Diskussion über verantwortungsvolle und adaptive Regulierung bedeutet jedoch nicht, dass wir die Entwicklung, die Produktion oder den Verkauf (halb-)autonomer Waffen befürworten – das Gegenteil ist richtig. Gegenwärtig erscheint es allerdings (leider!) wenig wahrscheinlich, dass sich die Staaten, die solche Waffen entwickeln, produzieren oder verkaufen, darauf einigen können, einen internationalen Vertrag zu unterzeichnen, der diese Produkte in sinnvoller Weise verbietet oder einschränkt.[84] Der hier vorgeschlagene Regulierungsansatz sollte daher zumindest die Verantwortungslücke schließen, die entsteht, wenn solche Waffen entwickelt und eingesetzt werden. Dies erscheint ebenfalls dringend erforderlich, da die Regeln des herkömmlichen humanitären Völkerrechts (*ius in bello*),[85] des Völkerstrafrechts,[86] und die internationalen Regeln über die Staatenverantwortung[87] erhebliche Lücken aufweisen. Es besteht daher die Gefahr, dass aufgrund dieser Lücken und der fehlenden Anpassung der herkömmlichen Regeln an die neuen Entwicklungen der KI-Waffentechnik ein Staat keine Entschädigung zahlen muss, wenn beispielsweise eine von ihm genutzte autonome Waffe Zivilisten völkerrechtswidrig angreift und tötet.

II. Kernelemente adaptiver Regulierung

Wir argumentieren, dass der adaptive Regulierungsansatz für KI-Hochrisikoprodukte und -dienstleistungen aus den folgenden Elementen bestehen soll:

Erstens soll das Risikopotenzial des KI-gesteuerten Produkts von einer unabhängigen Expertenkommission bewertet werden. Der Schwellenwert, ab dem eine solche Bewertung stattfinden muss, hängt davon ab, ob das KI-basierte Produkt oder eine solche Dienstleistung bei einer *Prima-facie-Klassifizierung* in eine Hochrisikokategorie fällt.[88] Mögliche Zukunftsszenarien bilden zusammen mit den verfügbaren Daten über frühere Erfahrungen (mit dem bewerteten Produkt oder einem ähnlichen) die Grundlage für die Bewertung. Handelt es sich bei dem evaluierten Produkt oder der Dienstleistung um eine völlige Neuentwicklung, sollte eine bestimmte Anzahl von Testfällen, die von der Expertenkommission vorgeschlagen werden, die nicht vorhandenen Daten auf der Grundlage früherer Anwendungen ersetzen.

Zweitens: Nachdem die Expertenkommission bewertet hat, ob das spezifische KI-gesteuerte Produkt oder die spezifische KI-gesteuerte Dienstleistung im oben genannten Sinne hochrisikoreich ist und

---

eingereichten Entschließungsantrag zum Verbot von autonomen Waffensystemen ohne menschliche Kontrolle („Killer-Robotern“) (1116/A(E)), der am 24. Februar 2021 durch den Österreichischen Nationalrat angenommen wurde, abrufbar unter <https://www.parlament.gv.at/PAKT/VHG/XXVII/E/E_00136/index.shtml#>.

84 Zu den verschiedenen staatlichen Positionen siehe das Übereinkommen über das Verbot oder die Beschränkung des Einsatzes bestimmter konventioneller Waffen, die übermäßige Leiden verursachen oder unterschiedslos wirken können (CCW), Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, Bericht der Sitzung 2019, CCW/GGW.1/2019/3, 25. September 2019, <https://undocs.org/en/CCW/GGE.1/2019/3>. Zur Diskussion dieser Ansichten vgl. Vöneky (Fn. 8), 15–16.

85 Siehe Genfer Konventionen (verabschiedet am 12. August 1949, in Kraft getreten am 21. Oktober 1950) 75 UNTS 31, 85, 135, 287; Zusatzprotokoll zu den Genfer Konventionen vom 12. August 1949 über den Schutz von Opfern internationaler bewaffneter Konflikte (Protokoll I) (verabschiedet am 8. Juni 1977, in Kraft getreten am 7. Dezember 1978) 1125 UNTS 3; Zusatzprotokoll zu den Genfer Konventionen vom 12. August 1949 über den Schutz von Opfern nicht-internationaler bewaffneter Konflikte (Protokoll II) (verabschiedet am 8. Juni 1977, in Kraft getreten am 7. Dezember 1978) 1125 UNTS 609.

86 Römisches Statut des Internationalen Strafgerichtshofs (beschlossen am 17. Juli 1998, in Kraft getreten am 1. Juli 2002) 2187 UNTS 3.

87 ILC, Materialien zur Verantwortung von Staaten für völkerrechtswidrige Handlungen, ST/LEG/SER.B/25 (2012).

88 Siehe dazu unten in diesem Abschnitt unter IV.

unter den neuen Regulierungsansatz fällt, und diese Frage bejaht wurde, entwickelt die Expertenkommission Risikoszenarien, in denen die möglichen Verluste und die damit verbundenen Wahrscheinlichkeiten für das Eintreten der Szenarien spezifiziert werden.

Drittens legen die Experten, auf Basis der in Punkt Eins getroffenen Risikoeinschätzung, auch unter Einbeziehung der aktuellen finanziellen Situation des entwickelnden oder produzierenden Unternehmens, ein zu zahlendes regulatorisches Kapital fest.[89] Sie erarbeiten zudem ein Evaluierungssystem, das es ermöglicht, auftretende Schäden zu messen und zu bewerten, die durch die Einführung oder den Betrieb des KI-gesteuerten Produkts oder Dienstes entstehen.

Viertens ist die Errichtung eines Fonds erforderlich, in den das regulatorische Kapital eingezahlt werden muss. Dieses Kapital soll zur Deckung von Schäden verwendet werden, die durch das KI-gesteuerte Produkt oder die KI-gesteuerte Hochrisikodienstleistung verursacht werden. Nach einem angemessenen Zeitraum, z. B. nach fünf bis zehn Jahren, soll das Kapital an das Unternehmen zurückgezahlt werden, wenn keine Verluste oder Schäden durch das Produkt oder die Dienstleistung verursacht wurden.

Fünftens muss das Unternehmen nach der Markteinführung eines Produkts oder einer Dienstleistung die Leistung und Auswirkungen des Produkts oder der Dienstleistung überwachen, indem es Daten darüber im Rahmen des Evaluierungssystems und in einer Überwachungsphase sammelt. Diese Daten dienen als Grundlage für die künftige Bewertung der Risikobehaftetheit des Produkts durch die Expertenkommission. Insbesondere wenn es sich um ein neues Produkt oder eine neue Dienstleistung handelt und nur wenige Daten verfügbar sind, ist das Evaluierungssystem von entscheidender Bedeutung, denn es dient als Datenbasis für künftige Entscheidungen über die Höhe des regulatorischen Kapitals und die Notwendigkeit einer künftigen Überwachung des Produkts oder der Dienstleistung.

Sechstens: Ein weiteres Element des vorgeschlagenen Governance-Systems ist, dass das Unternehmen aufgefordert werden soll, geeignete Testmechanismen zu entwickeln. Ein Testmechanismus ist ein transparentes Verfahren, das die Sicherheit des KI-gesteuerten Produkts gewährleistet. So muss beispielsweise ein selbstfahrendes Fahrzeug eine ausreichende Anzahl von Testfällen bestehen, um sicherzustellen, dass sich diese Fahrzeuge sicher verhalten und einen angemessenen Standard erfüllen.[90] Ein solcher Standard und ein Testmechanismus sollen von der Expertenkommission festgelegt werden. Ohne Testmechanismus sollte demnach kein Markteintritt möglich sein. Anhand der Daten aus der Überwachungsphase kann die Expertenkommission das Produkt bewerten; der Testmechanismus hat zusätzliche Vorteile, da er von dem Unternehmen selbst zur kontinuierlichen Bewertung des Produkts verwendet werden kann. Zudem kann er die im Folgenden erwähnte Reevaluierung unterstützen. Es wird darüber hinaus der Regulierungsbehörde helfen, automatisierte Testmechanismen für die kontinuierliche Überwachung und Bewertung der Technologie bereitzustellen, auch in ähnlichen Szenarien, etwa bei der Zulassung ähnlicher Technologien.

Siebtens: Die Expertenkommission muss das KI-gesteuerte Produkt oder die KI-gesteuerte Dienstleistung regelmäßig, d. h. jedes Jahr, neu bewerten. Sie kann ihre Entscheidung über die angemessene Höhe des für die Risiken erforderlichen regulatorischen Eigenkapitals ändern, indem sie sich auf neue

89 Unten F.

90 Siehe bspw. Menzel/Bagschik/Maurer, 'Scenarios for development, test and validation of automated vehicles' (2018) IEEE Intelligent Vehicles Symposium (IV).

Informationen stützt und die gesammelten Daten auswertet. Das oben erwähnte Evaluierungssystem wird zuverlässige Daten für die relevanten Entscheidungen liefern.[91]

III. Vorteile adaptiver Regulierung

Aus diesem Ansatz der KI-Regulierung ergeben sich entscheidende Vorteile. Er vermeidet eine Überregulierung von Hochrisiko-KI-Produkten und -Dienstleistungen, insbesondere in Fällen, in denen die Technologie neu und die damit verbundenen Risiken *ex ante* unklar sind. Denn Regulierungsansätze, die präventive Genehmigungsverfahren vorsehen, könnten einerseits den Markteintritt solcher Produkte ganz verhindern (wenn Zulassungsschwellen zu hoch gesetzt sind) oder andererseits den Markteintritt eines unsicheren Produkts ermöglichen (wenn die Schwellen zu niedrig gesetzt sind oder umgegangen werden bzw. nicht implementiert werden), ohne dass Schäden abgesichert sind.[92] Mit dem hier vertretenen Ansatz wird es dagegen möglich sein, abzusichern, dass ein neues Hochrisiko-KI-Produkt oder eine neue KI-basierte Hochrisikodienstleistung auf den Markt gebracht werden kann, wenn es hinreichend sicher ist, während ausreichend regulatorisches Kapital mögliche zukünftige Schäden abdeckt. Das hinterlegte Kapital wird an das Unternehmen (erst) zurückgezahlt werden, wenn sich das Hochrisiko-KI-Produkt oder die Dienstleistung nach einem Bewertungszeitraum als risikoarm erweist, wofür die während dieses Zeitraums gemäß dem Evaluierungssystem gesammelten Daten herangezogen werden.

1. Flexibilität

Der hier vertretene Ansatz ermöglicht es daher, schnell und flexibel auf neue technologische Entwicklungen zu reagieren. Da nur die Kernelemente der Regulierung *a priori* gesetzlich und rechtlich festgelegt sind und die Einzelheiten von einer Expertenkommission von Fall zu Fall angepasst werden, kann der spezifische Rahmen für ein KI-Hochrisikoprodukt eine solche Dienstleistung je nach den verfügbaren Informationen und Daten geändert werden. Eine regelmäßige Neubewertung des Produkts oder der Dienstleistung gewährleistet, dass neue Informationen berücksichtigt werden können und die Entscheidung auf dem neuesten Stand der Technik beruht.

2. Risikosensibilität

Der Ansatz ist nicht nur risikosensibel im Hinblick auf das neu entwickelte KI-basierte Hochrisikoprodukt oder die entsprechende KI-basierte Dienstleistung, sondern berücksichtigt auch die unterschiedlichen Risikoniveaus, die von verschiedenen Gesellschaften und Rechtskulturen akzeptiert werden. Es ist davon auszugehen, dass verschiedene Staaten und Gesellschaften bereit sind, je nach dem erwarteten Nutzen des KI-Systems unterschiedlich hohe Risiken im Zusammenhang mit KI-Produkten und -dienstleistungen zu akzeptieren. Wenn eine Gesellschaft – beispielsweise aufgrund einer alternden Bevölkerung und Defiziten im öffentlichen Verkehrssystem – auf autonome Fahrzeuge besonders angewiesen ist, könnte sie geneigt sein und in einer Demokratie in den dort vorgesehenen Verfahren vereinbaren, höhere Risiken im Zusammenhang mit diesen Fahrzeugen zu akzeptieren, um einen früheren Markteintritt für diese KI-basierten Produkte zu ermöglichen. Diesem Ziel entsprechend könnte eine

91 Wie bereits erwähnt, sollte das Kapital nach einem angemessenen Zeitraum an das Unternehmen zurückgezahlt werden, wenn keine Verluste oder Schäden durch das Produkt oder die Dienstleistung entstanden sind.

92 Letzteres trifft auch auf den Regulierungsansatz auf der neuen EU-KI-VO zu, dazu oben IV. 3. D.III.2.

rechtlich im Rahmen des Genehmigungsverfahrens festgelegte Schwelle für den Markteintritt gesenkt werden, wenn gleichzeitig das regulatorische Kapital in den Fonds eingezahlt wird und sicherstellt, dass Schäden ausgeglichen werden. Das Gleiche gilt beispielsweise für KI-gesteuerte medizinische Geräte oder andere KI-basierte Hochrisikoprodukte oder -dienstleistungen, die aufgrund bestimmter Umstände für das Allgemeinwohl besonders wichtig sein könnten. Als Grundsatz sollte hier jedoch gelten, dass KI-systeme, die Menschen ersetzen, bspw. als Fahrer, mindestens genauso gut sein müssen wie ein entsprechender menschlicher Fahrer, soll das KI-System zum Nutzen für die Gesellschaft sein.

3. Potenzielle Universalität und mögliche Regionalisierung

Da es sich bei KI-Systemen um eine Technologie handelt, die auf allen Kontinenten eingesetzt werden kann, könnten die Expertenkommission und ihre Entscheidungen auf internationales Recht gestützt sein. Ein internationaler Vertrag, der diesen Ansatz völkerrechtlich verankert, kann Lücken schließen, die durch unzureichende nationale Zulassungsverfahren entstehen. Die Empfehlungen oder (rechtsverbindlichen) Entscheidungen der Kommission könnten, sobald sie veröffentlicht sind, bei gleichem Risikobewusstsein in den verschiedenen nationalen Rechtsordnungen umgesetzt werden und als Ergänzung zum nationalen Zulassungsverfahren dienen.

Wenn es in verschiedenen Staaten jedoch unterschiedliche Risikoeinstellungen gegenüber einem KI-gesteuerten Hochrisikoprodukt oder einer solchen Dienstleistung gibt, kann bei der Umsetzung des hier dargelegten Regulierungsvorschlags auf nationaler oder regionaler Ebene zudem ein kultureller *bias* der Risikoabneigung (oder Risikoneigung) berücksichtigt werden. Dies ermöglicht die Flexibilität eines Staates, um eine unzureichende Regulierung oder eine Überregulierung zu vermeiden und gleichzeitig das Allgemeinwohl zu fördern, z. B. den Schutz der Gesundheit und der Umwelt oder den Schutz bestimmter finanzieller Vermögenswerte.

Solche Anpassungen können insbesondere in einer demokratischen Gesellschaft notwendig sein, wenn sich die Risikowahrnehmung der Bevölkerung im Laufe der Zeit ändert und Gesetzgeber und Regierungen auf die veränderten Einstellungen reagieren müssen. Hierzu hatte bereits das Bundesverfassungsgericht festgestellt, dass Hochrisikotechnologien (in dem entschiedenen Fall die Atomenergie) wegen der potenziell schweren Schäden bei ihrer Nutzung besonders auf die Akzeptanz der Bevölkerung in der demokratischen Gesellschaft angewiesen sind. Das Gericht betonte, dass im Falle einer veränderten Wahrnehmung einer Hochrisikotechnologie in der Bevölkerung eine Neubewertung durch den nationalen Gesetzgeber auch dann gerechtfertigt sei, wenn keine neuen Tatsachen vorliegen.[93]

4. Risikoüberwachung

Es ist davon auszugehen, dass ein Unternehmen in den meisten Fällen *a priori* von der Sicherheit seines Produkts oder seiner Dienstleistung überzeugt ist und argumentieren wird, dass sein KI-gesteuertes Produkt oder seine KI-gesteuerte Dienstleistung ohne relevante Risiken verwendet werden kann,

93 BVerfG, Urteil des Ersten Senats vom 6. Dezember 2016 – 1 BvR 2821/11 Rn. 308, <http://www.bverfg.de/e/rs20161206_1bvr282111.html>. Eine der Fragen in dem Verfahren war, ob der Gesetzgeber in Deutschland den Atomausstieg rechtfertigen kann, der nach dem Reaktorunfall in Fukushima, Japan, beschlossen wurde. Dieser wurde zum Teil als „irrationale" Änderung der Gesetze kritisiert, da der Reaktorunfall in Fukushima an sich die Risikofaktoren im Zusammenhang mit den in Deutschland befindlichen Kernreaktoren nicht verändert habe.

während diese Meinung möglicherweise nicht von allen Experten auf diesem Gebiet geteilt wird. Daher ist die Erhebung von Daten über die Leistung des Produkts in der realen Welt durch das Unternehmen im Rahmen des Evaluierungssystems ein wichtiger Teil des hier vorgestellten Ansatzes.

Einerseits können diese Daten dem Unternehmen helfen, nach einem bestimmten Bewertungszeitraum nachzuweisen, dass sein Produkt – wie behauptet – ein risikoarmes Produkt ist, und so begründen, dass das regulatorische Kapital reduziert oder zurückgezahlt werden kann.

Andererseits können die gesammelten Daten, falls das KI-gesteuerte Produkt Schäden verursacht, dazu beitragen, das Produkt zu verbessern und künftige Probleme bei der Nutzung der Technologie zu beheben. Die Daten können auch als wichtige Informationsquelle dienen, wenn es darum geht, ähnliche Produkte zu bewerten und ihre Risiken abzuschätzen. Eine Überwachungsphase ist daher ein wichtiges Element des Vorschlags, da zuverlässige Daten über die Leistung des Produkts erhoben werden, die wichtig sein können, um nachzuweisen, dass die Technologie tatsächlich so risikolos ist, wie das Unternehmen zu Beginn vertreten hat.

5. Demokratische Legitimitation und *Expertokratie*?

Der hier dargelegte Ansatz ist nicht von der Verfassung eines demokratischen, auf den Menschenrechten basierenden Staates abhängig, aber er ist mit einer Demokratie und den Zielen des Schutzes zentraler Menschen- und Verfassungsrechte, wie Leben und Gesundheit, sowie von Allgemeingütern, wie der Umwelt, kompatibel. Um eine ausreichende und legitimierte Grundlage zu haben, sollten die von der Expertenkommission umgesetzten Regeln und die Regeln zur Einrichtung der Expertenkommission ihrerseits auf einem Parlamentsgesetz beruhen. Solche rechtlich verankerten Expertenkommissionen bestehen bereits in verschiedenen Bereichen im Rahmen der Regulierung disruptiver risikoreicher Produkte. Sie sind ein entscheidender Bestandteil von Genehmigungsverfahren bei der Entwicklung neuer Arzneimittel, wie es zum Beispiel im deutschen Arzneimittelgesetz (AMG) vorgesehen ist.[94] Ein weiteres Beispiel der rechtlichen Verankerung einer mitbewertenden Kommission, allerdings aufgrund einer Verordnung und nicht durch Parlamentsgesetz, ist der Bereich der Biotechnologie.[95]

Solange die wichtigsten Anforderungen an die Kommission, wie das Verfahren zur Ernennung ihrer Mitglieder, ihre Anzahl, ihr wissenschaftlicher Hintergrund und das Verfahren zur Ausarbeitung von Empfehlungen oder Beschlüssen, auf einem Parlamentsgesetz beruhen, ist ein ausreichendes Maß an demokratischer Rückbindung und Legitimation gegeben.[96] Dies vermeidet in einer Demokratie die Fallstricke einer Expertokratie ohne ausreichende Anbindung an das Parlament. Eine gesetzliche Grundlage entspricht zudem den Anforderungen menschenrechts- und demokratiebasierter Verfassungen, wie

94 §§ 40 Abs. 1, 42 Abs. 1 AMG (Fn. 51). Für mehr Details vgl. Vöneky, *Recht, Moral und Ethik* (2010) 584–635, insb. 594–606.

95 Vgl. hier die interdisziplinäre Zentrale Kommission für Biologische Sicherheit (ZKBS), deren Errichtung ursprünglich auf Grundlage der „Richtlinien zum Schutz vor Gefahren durch in vitro neu kombinierte Nukleinsäuren" erfolgte, bei denen es sich (wohl) um eine Verwaltungsvorschrift handelt, vgl. Hirsch/Schmidt-Didczuhn, 'Gentechnik-Gesetz — ein Schritt in gesetzgeberisches Neuland' (1989), 22 Zeitschrift für Rechtspolitik 12, 458. Mit der Verabschiedung des GenTG (s. Fn. 99) im Jahre 1990 wurde die Einrichtung des Gremiums schließlich in einem formellen Gesetz geregelt (vgl. § 4 GenTG).

96 Vöneky (Fn. 94), 564 ff.

dem deutschen Grundgesetz, die verlangen, dass die wesentlichen verfassungsrechtlich relevanten Entscheidungen auf vom Parlament beschlossenen Regelungen beruhen müssen.[97]

6. Unabhängigkeit vom Versicherungsmarkt

Der hier vertretene Ansatz vermeidet schließlich den Rückgriff auf ein privatwirtschaftlich verankertes Versicherungssystem. Ein Ansatz, der sich auf ein Versicherungssystem bezieht, das die herstellenden Unternehmen verpflichtet, eine Versicherung für ihre KI-basierten Hochrisikoprodukte abzuschließen, würde von der Verfügbarkeit solcher Versicherungen durch die Versicherungsunternehmen abhängen. Dies könnte jedoch aus praktischen oder strukturellen Gründen scheitern. Zudem könnte eine Versicherung für die Entwicklung neuer Hochrisiko-KI-Produkte und -Dienstleistungen nicht praktikabel sein, wenn und weil in diesem Fall nur eine begrenzte Menge an Daten und Erfahrungen zur Verfügung steht.[98] Außerdem können Szenarien mit geringer Wahrscheinlichkeit oder hohem Risiko und unklarer Wahrscheinlichkeit kaum angemessen durch Versicherungen abgedeckt werden, da eine Risikoteilung durch den Versicherer unmöglich oder nur schwer zu erreichen ist. Schließlich würde der Rückgriff auf eine Versicherung bedeuten, dass höhere Kosten von einem KI-Produkte produzierenden Unternehmen getragen werden müssten, da die Versicherungsgesellschaft für ihr Versicherungsprodukt entlohnt werden muss und finanzielle Nachteile durch die Unterbewertung von Risiken vermieden werden müssen.

Auf nationaler Ebene gibt es ein Beispiel dafür, dass der Versuch, eine disruptive Technologie, in diesem Fall die Biotechnologie, auf der Grundlage einer Versicherungspflicht zu regulieren, gescheitert ist, da diese Pflicht weder von der Regulierungsbehörde noch von der Versicherungsbranche umgesetzt wurde.[99] Auch auf internationaler Ebene stellt sich die Pflicht zum Abschluss einer Versicherung für die Betreiber als ein wesentliches Hindernis für die Ratifizierung und das Inkrafttreten eines internationalen Abkommens zur Haftung für Umweltschäden im Bereich der Antarktis dar.[100]

---

97 Das sog. „Wesentlichkeitsprinzip“, das aus dem Grundgesetz hergeleitet wird, hängt von der verfassungsrechtlichen Ausgestaltung ab und ist kein notwendiges Element *jeder* auf den Menschenrechten basierenden liberalen Demokratie. In den USA beispielsweise ist es verfassungsmäßig, dass der Präsident der Vereinigten Staaten Exekutivanordnungen erlässt, die für die Ausübung der verfassungsmäßigen Rechte des Einzelnen von großer Bedeutung sind, ohne dass eine besondere Rechtsgrundlage des Parlaments erforderlich ist. Zum „Wesentlichkeitsprinzip“ nach dem deutschen Grundgesetz vgl. Vöneky (Fn. 94), 214–218 m. w. N.; Grzeszick (Fn. 61), Rn. 105.

98 Dies ist das Problem, das im Zusammenhang mit der Pflicht zum Abschluss einer Versicherung für einen Betreiber besteht, der in der Antarktis Umweltkatastrophen verursachen kann, wie es im Annex zum Antarktis-Vertrag festgelegt ist, Annex VI des Umweltschutzprotokolls zum Antarktis-Vertrag über die Haftung bei umweltgefährdenden Notfällen (beschlossen am 14. Juni 2005, noch nicht in Kraft getreten), vgl. IGP&I Clubs, *Annex VI to the Protocol on Environmental Protection to the Antarctic Treaty: Financial Security* (2019), <https://documents.ats.aq/ATCM42/ip/ATCM42_ip101_e.doc>.

99 Vgl. die hierfür vorgesehene Regelung im dt. Gentechnikgesetz (GenTG), BGBl. 1993 I 2066: Gem. § 36 GenTG sollte die deutsche Bundesregierung die Versicherungspflicht mit Zustimmung des Bundesrates durch eine Rechtsverordnung umsetzen. Eine solche Rechtsverordnung wurde allerdings bis jetzt nicht verabschiedet, vgl. Deutscher Ethikrat, *Biosicherheit – Freiheit und Verantwortung in der Wissenschaft: Stellungnahme* (2014) 264, <https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-biosicherheit.pdf>.

100 Siehe Fn. 98.

IV. Herausforderungen adaptiver Regulierung

1. Nichtvorhandene finanzielle Mittel?

Ein erstes Gegenargument gegen den Vorschlag kann sein, dass (anders als bei Finanzinstituten, insbesondere Banken) die Unternehmen, die risikoreiche KI-Produkte oder -Dienstleistungen entwickeln und verkaufen, nicht über das Kapital verfügen, um einen bestimmten Geldbetrag als Garantie für mögliche künftige, durch die Produkte verursachte Schäden zu zahlen. Dieses Argument überzeugt jedoch schon nicht, wenn an etablierte große Technologieunternehmen wie Facebook, Google oder Apple usw. gedacht wird, die KI-Produkte und -Dienstleistungen entwickeln oder diese Entwicklung an ihre Tochterunternehmen auslagern.

Andererseits entwickeln auch junge Unternehmen KI-gesteuerte Produkte und Dienstleistungen, die in den Hochrisikobereich fallen können. Auch diese Unternehmen erhalten allerdings oftmals Kapital von privaten Investoren und erreichen erst mehrere Jahre nach der Entwicklung, Produktion und Markteinführung eines Produkts oder einer Dienstleistung die Gewinnschwelle.[101] Wenn ein Investor, häufig ein Risikokapitalgeber, Kenntnis davon hat, dass die regulatorische Anforderung darin besteht, einen bestimmten Kapitalbetrag in einen Fonds einzuzahlen, der als finanzielle Sicherheit dient und nach einer bestimmten Zeit an das Unternehmen zurückgezahlt wird, wenn das Produkt oder die Dienstleistung keine Schäden verursacht, würde diese Verpflichtung die Finanzierung des Unternehmens im Vergleich zu anderen Anforderungen, beispielsweise im Rahmen eines Genehmigungsverfahrens, nicht behindern. Im Gegenteil: Die Festlegung eines Schwellenwerts für ein bestimmtes regulatorisches Kapital als notwendige Bedingung für den Markteintritt eines KI-basierten Hochriskoprodukts (nicht für das Forschungs- oder Entwicklungsstadium des Produkts) oder einer KI-basierten Dienstleistung ist eine Möglichkeit für den Investor, die Risiken zu berücksichtigen, die das Unternehmen selbst möglicherweise zu niedrig einschätzt.

Für den Fall, dass ein Staat davon überzeugt ist, dass ein bestimmtes KI-getriebenes Produkt oder eine bestimmte KI-getriebene Dienstleistung das Gemeinwohl seiner Gesellschaft in einem solchen Maße fördert, dass es gute Gründe gibt, den Markteintritt zu unterstützen, und private Investoren zögern, dies zu tun, weil mit dem Produkt oder der Dienstleistung große oder unklare Risiken verbunden sind, besteht zudem die Möglichkeit, dass der jeweilige Staat ein Unternehmen mit seinen finanziellen Mitteln unterstützt. Auch in anderen Fällen der Entwicklung von Hochrisikotechnologien oder der Entwicklung von Hochrisikoprodukten wurde und wird in unterschiedlicher Form finanzielle Unterstützung gewährt.[102]

---

101 Bspw. hat der Automobilhersteller Tesla, der sich mit der Entwicklung (teil-)autonomer Autos befasst, erst 2020 die Gewinnzone erreicht, vgl. 'Tesla Has First Profitable Year but Competition is Growing' (*The New York Times*, 27. Januar 2021) <https://www.nytimes.com/2021/01/27/business/tesla-earnings.html>.

102 Dies geschieht auch in anderen Szenarien; so wurden z. B. während der Covid19-Pandemie bestimmte Unternehmen in Deutschland, die Impfstoffe entwickelten, von der Bundesregierung und der EU unterstützt; z. B. hat die Kreditanstalt für Wiederaufbau, eine Förderbank, die im Auftrag des Bundes eine Minderheitsbeteiligung an der CureVac AG erworben, vgl. KfW, 'KfW acquires minority interest in CureVac AG on behalf of the Federal government' (*KfW*, 6. August 2020) <https://www.kfw.de/KfW-Group/Newsroom/Latest-News/News-Details_600640.html>; außerdem wurde die Hochrisikotechnologie der Atomkraftwerke in Deutschland seit ihrer Gründung mit verschiedenen Mitteln finanziell unterstützt; die Haftung nach § 31 Atomgesetz (AtG) ist grds. unbegrenzt, wird jedoch durch § 34 AtG effektiv auf 2,5 Mrd. Euro begrenzt; ab diesem Betrag (in bestimmten Fällen bereits bei geringeren Beträgen) ist der Betreiber von der Haftung freizustellen und der Staat haftet, vgl. auch §§ 25 ff., 31, und 38 AtG, BGBl 1985 I 1565.

2. Unklarheit und Überregulierung?

Ein weiteres Gegenargument, das gegen den hier vertretenen Regulierungsansatz vorgebracht werden könnte, ist, dass zu unklar sei, welche KI-gesteuerten Produkte oder Dienstleistungen als Hochrisikoprodukte oder Hochrisikodienstleistungen angesehen werden müssen, und es daher eine inhärente Verzerrung geben könne, die zu einer Überregulierung führe, da die Kategorie der Hochrisikoprodukte oder -dienstleistungen nicht ohne Grauzonen und nicht eng genug bestimmt werden könne. Diesem Argument kann jedoch entgegengehalten werden, dass die Kategorie der risikoreichen KI-Produkte und -Dienstleistungen, die die Expertenkommission bewerten soll, nach einem Prozess, der den Diskurs mit verschiedenen Akteuren, wie Unternehmen, Entwicklern, Forschern usw., über diese Fragen einschließt, in nationalem, supranationalem oder internationalem Recht festgelegt wird.[103] Kriterien für eine solche Einstufung sollten die möglichen Schäden sein, die auftreten können, wenn ein bestimmtes, mit dem Produkt oder der Dienstleistung verbundenes Risiko eintritt. Um eine Überregulierung zu vermeiden, sollte die Gruppe der KI-gesteuerten Hochrisikoprodukte und Hochrisikodienstleistungen auf die offensichtlichen beschränkt werden; dies wiederum kann sich an der Risikoneigung oder dem Risikobewusstsein einer Gesellschaft eines Staates orientieren, wenn kein internationaler Konsens besteht.

3. Zu frühe Regulierung?

Die Regulierung neu entstehender Technologien sowie von Produkten und Dienstleistungen, die auf diesen Technologien basieren, ist eine besondere Herausforderung, da oftmals das Argument vorgebracht wird, dass eine Regulierung zu früh geschehe, weil das Endprodukt in diesem Entwicklungsstadium noch zu ungewiss sei. Dies ist oft mit dem Argument verbunden, dass die Regulierung neu entstehender Technologien eine unvermeidliche Überregulierung dieser Technologien bedeuten würde. Gegen das Argument spricht, dass eine Gesellschaft und jeder Staat und die globale Gemeinschaft als Ganzes vor allem vermeiden sollten, in die „It-is-too-early-until-it-is-too-late"-Regulierungsfalle zu tappen. Gerade dynamische Entwicklungen im Hochrisikobereich zeichnen sich dadurch aus, dass sinnvolle Regulierung zu spät kommen kann, da Gesetzgebungsprozesse oftmals langwierig sind. Der Vorteil adaptiver Regulierung ist, dass trotz Regulierung eine flexible auf den spezifischen Fall und die Risikoentwicklung angepasste Normierung möglich ist.

4. Nichtverfügbarkeit unabhängiger Experten?

Wie schon erwähnt ist der Einsatz von Expertenkommissionen oder interdiziplinären Gremien, wie Ethikkommissionen, in verschiedenen Bereichen im Rahmen der Regulierung von disruptiven, risikoreichen Produkten seit vielen Jahren rechtlich verankert und hat sich bewährt.[104] Es ist nicht ersichtlich, warum im Fall der KI-Regulierung anderes gelten sollte. Die Abhängigkeiten von Experten können durch Transparenzvorschriften offengelegt werden. Zudem verhindert eine plurale Besetzung dieser Gremien eine einseitige Ausrichtung.

103 Vgl. die bereits erwähnte Liste des Europäischen Parlaments im Hinblick auf risikoreiche KI-Produkte in Fn. 33.
104 Für den Bereich der Biotechnologie, vgl. bspw. die ZKBS, dazu oben bei Fn. 95.

5. Unzulässige Mithaftung von Unternehmen?

Kein Argument gegen die Fondskonstruktion ist es auch, dass Unternehmen, die KI-basierte Produkte oder Dienstleistungen vertreiben, die sich später als risikoarm herausstellen, unzulässig in Mithaftung genommen werden für die Unternehmen, die KI-basierte Produkte oder Dienstleistungen herstellen und vertreiben, die sich später als risikoreich herausstellen und Schäden verursachen. Ziel der Einrichtung des Fonds ist es, dass konkrete Schadensersatzforderungen gegen ein bestimmtes Unternehmen X nach einem Schadensfall zunächst aus dem Fonds beglichen werden, und zwar aus der Summe, die das schadensverursachende (sic!) Unternehmen X mit seinen risikoreichen KI-Produkten oder Dienstleistungen gerade für diese Schadensfälle hinterlegt hat; sollte die Schadenssumme darüber hinausgehen, sollten die weiteren Schäden von diesem Unternehmen X zunächst selbst getragen werden. Anders als bei Fonds, die ein Gesamtkapital enthalten, das sich erschöpft, wenn Schadensersatzzahlungen in großer Höhe geleistet werden, würde also sichergestellt werden, dass grundsätzlich der Fonds aus den für jedes Unternehmen getrennten finanziellen Reserven bestehen bleibt.

Sollte dagegen vereinbart werden, dass der gesamte Fonds bei einem Schadensfall haftet, müsste der Staat, dessen Staatszugehörigkeit das Unternehmen besitzt, dessen KI-Produkte risikoarm sind, eine Ausfallhaftung übernehmen, um eine Rückzahlung des Kapitals an das Unternehmen zu garantieren. Der Staat wäre dann verpflichtet, das eingezahlte regulative Kapital einem Unternehmen zu ersetzen, wenn sich ein KI-Produkt, anders als nach Expertenansicht erwartet, als risikoarm herausstellt und das regulative Kapital an das Unternehmen zurückzuzahlen ist, der Fonds aber aufgrund anderer Schadensfälle keine finanziellen Mittel dafür besitzen sollte.

F. Bestimmung des regulatorischen Kapitals

Zentral für die hier vorgeschlagene adaptive Regulierung ist die Bestimmung der Höhe des regulativen Kapitals. In diesem Abschnitt stellen wir einen formalen Ansatz vor, der probabilistische Methoden verwendet, wie sie bereits in der Wissenschaft vorgeschlagen werden.[105] Im *ersten Beispiel* betrachten wir ein Unternehmen, das die Möglichkeit hat, in zwei konkurrierende neue KI-Produkte zu investieren, von denen eines wesentlich risikoreicher ist als das andere. Selbst wenn wir davon ausgehen, dass dieses rational (im Sinne eines nutzenmaximierenden[106] Unternehmens)[107] handelt, gibt es gute Gründe für die Annahme, dass Risiken, die das Vermögen des Unternehmens übersteigen, im Entscheidungsprozess des Unternehmens nicht in vollem Umfang berücksichtigt werden, da ein Eintreten der Risiken den Insolvenzfall des Unternehmens nach sich ziehen würde. Obwohl es auf den ersten Blick vernünftig erscheint, dass für die Leitung des Unternehmens die Verringerung dieser Art von Risiken Priorität haben sollte, da sie die Existenz des Unternehmens bedrohen, wird das gegenteilige Verhalten incentiviert. Die hohen oder sogar existenziellen Risiken werden vom Unternehmen vernachlässigt, wenn es keine Regelung gibt, die das Unternehmen dazu verpflichtet, sie zu berücksichtigen: Das Unternehmen wird

105 Acharya et al., 'Measuring systemic risk' (2017) 30.1, The review of financial studies 2–47.

106 Für diese erste Aussage ist es nicht erforderlich, dass der Nutzen in einem monetären Maßstab gemessen wird. Später, wenn es um die Bestimmung des regulatorischen Kapitals geht, werden wir jedoch davon ausgehen, dass der Nutzen in Form von Vermögen gemessen wird.

107 Das bedeutet, dass zukünftige Gewinne und Verluste mit einer Nutzenfunktion gewichtet und dann durch die Erwartung gemittelt werden. Siehe zum Beispiel Kreps, *A course in microeconomic theory* (1990) oder Mas-Colell/Whinston/Green, *Microeconomic theory* Vol. 1 (1995).

risikoreiche Investitionen anstreben, da die höhere Rendite gerade nicht ausreichend durch die erwarteten Verluste abgeschwächt wird, da diese auf die Höhe des regulatorischen Kapitals begrenzt sind.[108]

Erstes Beispiel: Zwei konkurrierende KI-Technologien oder -Produkte

Betrachten wir ein Unternehmen mit einer anfänglichen Kapitalausstattung w_0. Das Unternehmen kann sich entscheiden, in zwei verschiedene Produkte oder Technologien zu investieren, die (zufällige) Renditen r und r' für die Investition von einer Währungseinheit bieten. Die erste Technologie ist die risikoärmere, während die zweite risikoreicher ist. Wir nehmen an, dass es zwei Szenarien gibt: Das erste Szenario (der günstigste Fall (*best case*), bezeichnet mit $+$) ist, wenn das Risiko de facto nicht eintritt. Dieses Szenario tritt mit einer gewissen Wahrscheinlichkeit p auf. In diesem Szenario bietet die riskantere Strategie eine höhere Rendite, d. h. $r(+) < r'(+)$.

Im zweiten Szenario (dem ungünstigsten Fall (*worst case*), der mit $-$ bezeichnet wird und eine Wahrscheinlichkeit von $1 - p$ hat, führt die risikoreichere Technologie zu größeren Verlusten, so dass wir $0 > r(-) > r'(-)$ annehmen, wobei beide Werte negativ sind (d. h. zu Verlusten führen).

Zusammenfassend lässt sich sagen, dass, wenn das Unternehmen das Anfangskapital w_0 in die Strategie investiert, das Vermögen am Ende der betrachteten Periode (z. B. zum Zeitpunkt 1) $w_1 = w_0 \cdot r$ ist, wenn es in die *erste* Technologie investiert, oder ${w_1}' = w_0 \cdot r'$, wenn es in die *zweite*, riskantere Technologie investiert. Der Insolvenzfall tritt ein, wenn $w_1 < 0$ bzw. ${w_1}' < 0$ ist.

Wir nehmen an, dass das Unternehmen den erwarteten Nutzen maximiert: Der erwartete Nutzen der ersten (risikoärmeren) Strategie ist $EU = E\left[u(w_1)1_{\{w_1>0\}}\right]$ und derjenige der zweiten (risikoreicheren) Strategie gerade $EU' = E[u({w_1}')1_{\{{w_1}'>0\}}]$. Dabei ist u eine Nutzenfunktion (*utility function*)[109] (wir nehmen an, dass sie wachsend ist), E bezeichnet den Erwartungswert und $1_{\{w_1>0\}}$ ist die Indikatorfunktion, die gleich eins ist, wenn $w_1 > 0$ (kein Konkurs), und andernfalls gleich null. Das Unternehmen wählt diejenige Strategie, die den größeren erwarteten Nutzen verspricht, entscheidet sich als für die erste Strategie, falls $EU > EU'$ und für die zweite, falls $EU' > EU$. Bei Gleichheit werden typischerweise weitere Merkmale betrachtet um die bessere Strategie auszuwählen. Dies ist im Standardfall eine rationale Strategie.

Betrachten wir zwei disruptive Technologien, so ist zu erwarten, dass im *worst case*-Szenario eine Insolvenz ausgelöst wird. Dies verändert das Bild dramatisch: Im *worst case*-Szenario ist der Indikator $1_{\{w_1>0\}}$ bzw. $1_{\{{w_1}'>0\}}$ jeweils gleich Null und fällt in der Berechnung weg. So ergibt sich $EU = p \cdot u(w_0 \cdot r(+))$ für die erste Technologie und $EU' = p \cdot u(w_0 \cdot r'(+))$ für die zweite, riskantere Technologie. Da die Rendite der risikoreicheren Technologie im *best case* höher ist, wird das Unternehmen diese Technologie stets (!) bevorzugen. Am wichtigsten ist, dass dies weder von der Wahrscheinlichkeit des schlimmsten Falles noch von der Höhe der auftretenden Verluste abhängt. Es ergibt sich, dass ein Unternehmen, welches seinen Nutzen maximieren will, Verluste, die über einen Konkurs hinausgehen, in seiner Strategie nicht berücksichtigt.

108 Siehe etwa Eberlein/Ernst/Madan, 'Unbounded liabilities, capital reserve requirements and the taxpayer put option', Quantitative Finance 12.5 (2012), 709–724 m. w. N.

109 Eine Nutzenfunktion ordnet verschiedenen Alternativen eine Zahl (den Nutzen) zu. Je höher die Zahl (der Nutzen) ist, desto stärker wird die Alternative bevorzugt. Bspw. hat 1 EUR für eine Person, die Millionär ist, einen anderen Wert als für eine Person, die arm ist. Die Nutzenfunktion ist in der Lage, solche (und andere) Effekte zu erfassen. Siehe Föllmer/Schied, *Stochastic finance: an introduction in discrete time* (2011) für weitere Nachweise.

Zusammenfassend lässt sich daher feststellen, dass das Ergebnis dieser Analyse die Bedeutung der Regulierung hervorhebt, die Anreize für das Unternehmen schafft, übermäßig riskante Strategien zu vermeiden.

Das erste Beispiel zeigt, dass ein nutzenmaximierendes Unternehmen überraschend leicht große Risiken akzeptieren wird. Insbesondere hat die genaue Höhe der disruptiven Verluste keinen Einfluss auf den rationalen Entscheidungsprozess, denn wenn die Verluste hoch genug sind, führt dies zur Insolvenz, da die Verluste auf die Höhe des Insolvenzfalls begrenzt sind. Es kann davon ausgegangen werden, dass es für das Unternehmen unerheblich ist, wie hoch die Verluste sind, wenn die Insolvenz eingetreten ist. Dies begünstigt insbesondere eine risikoreiche Strategie der relevanten Unternehmen, da Strategien mit durchschnittlich höherem Risiko im Allgemeinen höhere Gewinne versprechen. Eine adaptive Regulierung kann jedoch die Gemeinwohlinteressen fördern, indem sie darauf zielt, große Verluste zu vermeiden. Wir werden im Folgenden zeigen, dass die vorgeschlagene Regulierung große Verluste wieder in das Nutzenmaximierungsverfahren einbezieht, indem sie hohe Verluste mit hohen Regulierungskosten koppelt und damit „bestraft" und so dazu beiträgt, diese zu vermeiden.

Betrachtet man das oben erwähnte Problem der sog. Superintelligenz, zeigt sich eine besondere Herausforderung: Sobald Superintelligenz von einem Unternehmen entwickelt wird, wird der realisierte Nutzen potentiell unermesslich sein. Dementgegen wird argumentiert, dass eine Superintelligenz nicht mehr kontrolliert werden kann und somit eine existenzielle Bedrohung nicht nur für das Unternehmen darstellt.[110] Die potenziellen Verluste sind damit enorm und es handelt sich um eine disruptive Technologie. Wie in dem oben genannten Beispiel anhand einer rationalen Strategie erläutert, werden Unternehmen die Entwicklung einer superintelligenten KI anstreben (insbesondere wenn sie sich nicht rational verhalten), da die potentiellen Vorteile sehr verlockend sind. Die potentiellen Risiken und deren Höhe werden dabei ignoriert. Das obige Beispiel verdeutlich aber zusätzlich, dass die Ignoranz disruptiver Risiken durchaus rational sein kann; bei der erwähnten Strategie der Optimierung des erwarteten Nutzens wurde gerade die risikoreichere, disruptive Strategie bevorzugt. Dieser Fall macht einmal mehr deutlich, dass gerade bei disruptiven Technologien eine Regulierung notwendig ist. Ein besonders spektakulärer Fall ist die Entwicklung und Erforschung eines superintelligenten Systems. Die schwer abzuschätzenden Risiken und die immensen möglichen Profite charakterisieren diesen Fall, welcher damit direkt in den Kontext des Beispiels 1 fällt – selbstverständlich folgt, dass Leitlinien für die Kontrolle der Entwicklung solcher KI-Systeme dringend benötigt werden.

Zusammenfassend zeigt das Beispiel 1: Wenn KI-Produkte mit hohem Risiko zu großen Verlusten (und Schäden) führen – auch wenn die Wahrscheinlichkeit ex ante gering oder sehr gering ist – müssen sie von den jeweiligen Gesellschaften und Staaten ausgeglichen werden, da das Unternehmen im Konkurs nicht mehr in der Lage ist, sie zu decken. Daher ist eine Regulierung erforderlich, um eine Verantwortungslücke zu vermeiden. Das nun folgende Beispiel wird zeigen, dass eine vernünftige Regulierung eine effiziente Maximierung des Gesamtwohlstands und damit des Gemeinwohls (im Vergleich zu einem Fall ohne Regulierung) fördert.

110 Dazu oben unter A.

Zweites Beispiel: Ein stilisierter Rahmen für die Regulierung

In dem zweiten Beispiel wird das regulatorische Kapital einbezogen. Eine adaptive Regulierung kann den Gesamtwohlstand maximieren, relevante Risiken minimieren, große Verluste vermeiden und das Allgemeinwohl fördern, indem sie angemessene Kapitalanforderungen stellt.

Nehmen wir an, es gibt I Unternehmen: Jedes Unternehmen i verfügt über ein Anfangsvermögen $\overline{w}_0^i$, wobei ein Teil $\overline{w}_0^i - w_0^i$ zunächst konsumiert wird und der andere Teil w_0^i investiert wird (wie im obigen Beispiel). Das Unternehmen i zahlt ein regulatorisches Kapital r^i und strebt daher die folgende Maximierung an:

$$\boldsymbol{max}\left[c\cdot\left(\overline{w}_0^i - w_0^i - r^i\right) + E[u(w_1^i 1_{\{w_1^i>0\}})]\right]$$

Die zu betrachtende Regulierung zielt darauf ab, den Gesamtwohlstand zu maximieren: Im Falle des Konkurses eines Unternehmens, sagen wir i, müssen die öffentliche Hand und andere Akteure, wie die Konsumenten, die nicht entschädigt werden, die Verluste decken. Wir nehmen an, dass dies proportional zu den entstandenen Verlusten ist, $g\cdot w_1^i 1_{\{w_1^i<0\}}$. Die Gesamtwohlstandsfunktion $P^1 + P^2$ besteht aus zwei Teilen: Der erste Teil ist einfach die Summe des Nutzens der Unternehmen,

$$P^1 = \sum_{i=1}^{I} c\cdot\left(\overline{w}_0^i - w_0^i - r^i\right) + E[u(w_1^i 1_{\{w_1^i>0\}})].$$

Der zweite Teil,

$$P^2 = \sum_{i=1}^{I} \boldsymbol{E}\left[g\cdot w_1^i 1_{\{w_1^i<0\}}\right],$$

sind die erwarteten Kosten im Falle eines Konkurses der Unternehmen. In der Literatur wird argumentiert,[111] dass man ein effizientes Ergebnis, d. h. die Maximierung des Gesamtvermögens bzw. des Gemeinwohls, erreicht, wenn man das regulatorische Kapital festlegt als

$$\rho^i = \frac{g}{c}\cdot P\left(w_1^i<0\right)\cdot ES^i; \qquad (1)$$

Der erwartete Fehlbetrag ist hier gegeben durch $ES^i = -E\left[w_1^i 1_{\{w_1^i<0\}}\right]$. Durch die Auferlegung dieses regulatorischen Kapitals werden die Unternehmen Verluste über den Konkurs hinaus berücksichtigen, was dazu beiträgt, einen maximalen Gesamtwohlstand zu erreichen. Wie in der Literatur dargelegt, kann man zusätzlich zu den systemischen Effekten noch weitere berücksichtigen, die wir hier der Einfachheit halber nicht in Betracht ziehen.[112]

Hier stützt sich der adaptive Regulierungsansatz auf Erwartungen und geht daher davon aus, dass Wahrscheinlichkeiten bewertet werden können, auch wenn sie von Experten(-kommissionen) geschätzt[113] oder vorgeschlagen werden müssen.

Die Projektion unbekannter zukünftiger Risiken kann, wie in der Literatur erwähnt,[114] mit Hilfe der Extremwerttheorie formalisiert werden. Zentral ist daher für den hier vertretenen Vorschlag, dass

111 Siehe Acharya et al. (Fn. 105).
112 Siehe Acharya et al. (Fn. 105).
113 Pitera/Schmidt, 'Unbiased estimation of risk' (2018) 91 Journal of Banking & Finance 133–145.
114 Siehe bspw. De Haan/Ferreira, *Extreme value theory: an introduction* (2007).

angepasste Methoden verwendet werden, um eingehende Daten – die aufgrund des oben genannte Überwachungsprozesses oder anderer Quellen erhoben werden – einzubeziehen; die relevanten mathematischen Werkzeuge sind hierfür vorhanden.[115]

Bei großer Unsicherheit ist dies möglicherweise nicht mehr möglich. In einem solchen Fall kann man sich auf das Konzept der Unsicherheit von *Frank Knight*[116] stützen und sog. nichtlineare Erwartungswerte nutzen. Die genaue Bewertung in diesem Rahmen geht aber deutlich über den hier vorgestellten Kontext hinaus.

G. Dissens und Experten

Im Hinblick auf die Expertenkommission ist zu erwarten, dass unterschiedliche Meinungen aufeinandertreffen werden. Eine Möglichkeit wäre hier, die *worst case*-Meinung zu berücksichtigen, d. h. die risikoaverseste Sichtweise. Eine im Vergleich zu *best/worst case*-Szenarien oder ähnlichen Schätzungen vorzugswürdige Alternative ist es jedoch, sich auf die Glaubwürdigkeit der zugrunde liegenden Schätzungen zu verlassen. Dieser Ansatz basiert auf der so genannten Glaubwürdigkeitstheorie, einer Theorie zur Zusammenführung von Schätzungen, internen Schätzungen und verschiedenen Expertenmeinungen im versicherungsmathematischen Kontext.[117] Im Folgenden wird gezeigt, warum dies für die vorgeschlagene Regulierung relevant ist.

Drittes Beispiel: Regulierung auf der Grundlage der Glaubwürdigkeitstheorie

Zur Illustration dieses Ansatzes kehren wir zu Beispiel 1 zurück und betrachten zwei Expertinnen, von denen die eine p_1 und die andere p_2 für die (unbekannte) Wahrscheinlichkeit des *best case* vorschlägt. Die zugehörigen Werte des mit Gleichung (1) berechneten regulatorischen Kapitals werden mit ρ_1 bzw. ρ_2, bezeichnet.

Die Idee ist, ρ_1 und ρ_2 für die Schätzung des regulatorischen Kapitals zu gewichten und wie folgt aufzuaddieren:

$$\rho^{credible}(\theta) = \theta \cdot \rho_1 + (1-\theta) \cdot \rho_2.$$

Hierbei wird das Gewicht θ in einem geeigneten Sinne optimal gewählt. Wenn wir davon ausgehen, dass es bereits Erfahrungen mit den Schätzungen der beiden Experten gibt, können wir die Varianzen $v_1$und v_2 aus ihrer Schätzungshistorie ermitteln. Das Gewicht, welches zu minimaler Varianz des regulatorischen Kapitals führt, erhält man durch die Wahl von

$$\theta_{opt} = \frac{v_2}{v_1 + v_2}.$$

115 Siehe bspw. Jazwinski, *Stochastic processes and filtering theory* (1970); Frey/Schmidt, 'Filtering and Incomplete Information' in Bielecki/Brigo (Hrsg.), *Credit Risk Frontiers* (2011); Fadina/Neufeld/Schmidt, 'Affine processes under parameter uncertainty' (2019) 4.1 Probability, uncertainty and quantitative risk 1–35.

116 Knight, *Risk, uncertainty and profit* (1921). Eine Anwendung dieser Theorie in modernem Kontext findet sich bspw. in Fadina/Neufeld/Schmidt (Fn. 115).

117 Siehe den Überblick von Norberg, 'Credibility theory', Encyclopedia of Actuarial Science 1 (2004), 398–406 oder die sehr einflussreiche Arbeit von Bühlmann, 'Experience rating and credibility', ASTIN Bulletin: The Journal of the IAA 4.3 (1967), 199–207.

Bei unterschiedlichen Expertenmeinungen kann mit Hilfe der Glaubwürdigkeitstheorie ein valides Verfahren zur Kombination der vorgeschlagenen Modelle angewendet werden. Hierbei werden systematisch Experten bevorzugt, welche in der Vergangenheit bessere Schätzungen geliefert haben. Als Alternative bietet sich auch die Auswahl der Schätzung mit dem höchsten Kapital (oder geringstem), was allerdings leichter zu manipulieren ist. Robustere Varianten dieses Verfahrens etwa auf Basis von Quartilen existieren ebenfalls.

H. Fazit

In diesem Beitrag wurde ein adaptives Regulierungsmodell für Hochrisko-KI-Produkte und -Dienstleistungen vorgestellt, das auf Basis von Expertenmeinungen ein regulatorisches Kapital zur Hinterlegung in einen Fonds verlangt und global, regional oder auch national umgesetzt werden kann. Zum einen erlaubt es diese adaptive Regulierung, potentiell auftretende Schäden zu begleichen; zum anderen motiviert sie Unternehmen, zu große Risiken zu vermeiden. Beides führt dazu, Menschenrechte, wie das Recht auf Leben und Unversehrtheit, zu schützen und das Gemeinwohl zu fördern. Da das regulatorische Kapital wieder an die Unternehmen zurückgezahlt wird, wenn ein KI-Hochrisiko-Produkt oder eine solche Dienstleistung sicher ist und sich Risiken über Jahre hinweg nicht realisieren, entstehen durch diese Art der Regulierung keine unnötig hohen Barrieren für die Entwicklung, den Verkauf und die Nutzung neuer und wichtiger KI-basierter Technologien.

UNI
FREIBURG