Daniel Feuerstack

*Menschenrechtliche Vorgaben an die Transparenz KI-basierter Entscheidungen und deren Berücksichtigung in bestehenden Regulierungsansätzen**

I. Einleitung

Entscheidungen, die früher ausschließlich von Menschen getroffen wurden, werden zunehmend an Systeme der künstlichen Intelligenz (KI-Systeme) delegiert.[1] Wo immer KI-Systeme – indirekt als Entscheidungsunterstützungssysteme oder direkt als autonome Entscheidungsträger – über Menschen entscheiden, ist es für die Betroffenen bisher schwierig nachzuvollziehen, wie es zu der konkreten Entscheidung kam. Diese Intransparenz[2] KI-basierter Entscheidungen resultiert daraus, dass die Nutzer[3] und Entwickler von KI-Systemen entweder nicht willens oder nicht in der Lage sind, die tragenden Gründe für eine KI-basierte Entscheidung offenzulegen.[4] Aus menschenrechtlicher Sicht problematisch ist dies, wenn KI-Systeme in politisch und sozial sensiblen Bereichen eingesetzt werden, in denen Menschenrechtsverletzungen, insbesondere ungerechtfertigte Diskriminierungen, möglich sind. In den USA werden beispielsweise KI-Systeme zur Bewertung des Rückfallrisikos straffälliger Personen eingesetzt oder zur Prognose künftiger Straftaten im Rahmen des sog. *Predictive Policing*.[5] In Großbritannien und in den USA unterstützen KI-Systeme zudem Arbeitgeber bei der Personalauswahl, indem sie eigenständige Vorauswahlen treffen.[6] Viel Aufsehen erregte zudem der 2020 während der COVID-19-Pandemie in Großbritannien eingesetzte Algorithmus, der Schülerinnen und Schüler auf Basis ihrer bisherigen Noten bewerten sollte.[7] In Deutschland wird der Einsatz KI-basierter automatischer Gesichtserkennung im Rahmen der Strafverfolgung erwogen,[8] während andere europäische Staaten wie Großbritannien und Frankreich solche Systeme bereits einsetzen.[9] Zu nennen

* Ich danke Frau Prof. Dr. *Silja Vöneky* sowie *Tobias Crone* und *Daniel Becker* für ihre wertvollen Anmerkungen. Des Weiteren danke ich der Baden-Württemberg Stiftung für die Förderung im Rahmen des Projektes „*AI Trust*".

1 *Kroll et al.*, Accountable Algorithms, University of Pennsylvania Law Review 165 (2017), 633, 636.

2 Das Problem wird auf Englisch häufig mit dem Ausdruck „*opacity*" (dt. Undurchsichtigkeit/Opazität) umschrieben, vgl. *Burrell*, How the machine 'thinks': Understanding opacity in machine learning algorithms, Big Data & Society, 2016, 1; *de Laat*, Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?, Philosophy & technology 2018, Vol. 31, 525, 536.

3 Sofern in diesem Aufsatz das generische Maskulinum verwendet wird, und sich dieses nicht ausschließlich auf juristische Personen bezieht, sind alle Geschlechter umfasst.

4 Zusammenfassend *Burrell*, How the machine 'thinks': Understanding opacity in machine learning algorithms, Big Data & Society, 2016.

5 *Brayne/Christin*, Technologies of Crime Prediction: The Reception of Algorithms in Policing and Criminal Courts, Social Problems, 2020, 1, 2.

6 *Lischka/Klingel*, Wenn Maschinen Menschen bewerten – Internationale Fallbeispiele für Prozesse algorithmischer Entscheidungsfindung, Arbeitspapier der Bertelsmann Stiftung, 05.2017, 22 ff.

7 *Kaun,* Suing the algorithm: the mundanization of automated decision-making in public services through litigation, Communication and Society 2021, 2.

8 *Martini*, Gesichtserkennung im Spannungsfeld zwischen Sicherheit und Freiheit, NVwZ Extra 2022, 2 f.

9 *Martini*, Ebd., 3.

sind auch der Einsatz von Algorithmen zur Entscheidung über die Gewährung von Sozialhilfe in Österreich[10] und Schweden[11] sowie der Einsatz von KI-Systemen zur Entscheidung über den Zugang zu Universitäten in Frankreich.[12] Die deutsche Wirtschaftsauskunftei *Schufa* bewertet mithilfe algorithmischer Verfahren die Kreditwürdigkeit einzelner Personen.[13] Ein von dem Unternehmen Amazon in den USA genutztes KI-System entlässt sogar autonom Arbeitskräfte.[14]

In diesen Fällen werden Menschen von KI-Systemen beurteilt und ihnen drohen auf Grundlage dieser Bewertung Nachteile, etwa die Ablehnung einer Bewerbung für einen Ausbildungs- oder Arbeitsplatz, die Auflösung des Arbeitsverhältnisses, die Verwehrung eines Kredits oder die Attestierung eines hohen Rückfallrisikos.[15] Anders als bei KI-gestützten Empfehlungssystemen wie Musik-, Streaming- oder Kaufempfehlungen, bei Spracherkennungs-, Übersetzungs- oder Navigationssystemen nutzen die Betroffenen nicht selbst das KI-System und können sich diesen Entscheidungen daher auch nicht durch Nichtnutzung entziehen.[16]

Die statistik- und wahrscheinlichkeitsbasierte Entscheidungsfindung von KI-Systemen ermöglicht zwar genauere und effizientere Entscheidungen,[17] birgt jedoch auch das Risiko statistischer Diskriminierungen.[18] Sofern den Betroffenen weder die tragenden Gründe für die Entscheidung noch sonstige Anhaltspunkte mitgeteilt werden, können sie Diskriminierungen nicht nachweisen und in der Regel nicht vor Gericht geltend machen.[19] Zwar sind auch menschliche Entscheidungen intransparent. Menschen sind jedoch in der Lage, ihre Entscheidungen zu begründen und Betroffenen die tragenden Gründe für die Entscheidung mitzuteilen. Dadurch kann eine hinreichende Entscheidungstransparenz hergestellt werden.[20] Nach hier vertretener Ansicht ist diese Form der Entscheidungstransparenz in manchen Bereichen durch die völkerrechtlich verbindlichen Menschenrechte vorgeschrieben. Dies gilt einerseits für staatliche Entscheidungen, insbesondere der Exekutive und der Judikative. Andererseits können aber auch Private verpflichtet sein, Betroffenen die tragenden Gründe einer Entscheidung mitzuteilen, wenn sie eine staatsähnliche Funktion besitzen und Dritte von grundlegenden Leistungen, etwa im Bereich der Daseinsvorsorge, ausschließen können.

Diese sich aus den völkerrechtlich verbindlichen Menschenrechten ergebenden Informations- bzw. Begründungspflichten gelten auch, wenn Entscheidungen von KI-Systemen getroffen werden oder auf diesen basieren. Wenn und soweit eine Begründungspflicht für menschliche Entscheidungen besteht, müssen auch KI-basierte Entscheidungen begründet werden und damit grds. auch begründbar sein.

Mit dem Fokus auf Entscheidungstransparenz durch Begründungen zeigt dieser Beitrag eine konkrete Lösung für das Problem der Intransparenz KI-basierter Entscheidungen aus menschenrechtlicher Perspektive auf. Gleichzeitig wird der in diesem Kontext vielfach beschworene aber doch häufig abstrakt bleibende Begriff der Transparenz[21] konkretisiert und rechtlich fassbarer gemacht. Die in der Diskussion um transparente KI verwendeten Begriffe der Erklärbarkeit, Verstehbarkeit, Interpretierbarkeit und Nachvollziehbarkeit[22] werden damit nicht abgelehnt. Sie werden vielmehr um

10 *Szigetvari*, Beruf, Ausbildung, Alter, Geschlecht: Der Algorithmus des AMS, Der Standard, 15.10.2018, https://derstandard.at/2000089325546/Beruf-Ausbildung-AlterGeschlecht-Das-sind-die-Zutaten-zum-neuen.

11 *Kaun,* Suing the algorithm: the mundanization of automated decision-making in public services through litigation, Communication and Society 2021, 2.

12 *Martini*, Automatisch Erlaubt? Fünf Anwendungsfälle algorithmischer Systeme auf dem juristischen Prüfstand, Arbeitspapier der Bertelsmann Stiftung, Mai 2020, 12 ff.

13 Siehe hierzu den Vorlagebeschluss des VG Wiesbaden, ZD 2022, 121, 122.

14 *Crispin*, Welcome to dystopia: getting fired from your job as an Amazon worker by an app, The Guardian, 05.07.2021, https://www.theguardian.com/commentisfree/2021/jul/05/amazon-worker-fired-app-dystopia?CMP=fb_a-technology_b-gdntech.

15 Bericht der Enquete-Kommission Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale (2020), BT-Drucks. 19/23700, 2020, 63.

16 *Beining*, Wie Algorithmen verständlich werden, 2019, Bertelsmann-Stiftung, 15.

17 Vgl. *Kleinberg et al.*, Discrimination in the Age of Algorithms, Journal of Legal Analysis 2018, Vol. 10, 113, 160.

18 Vgl. *Hinsch*, Differences that Make a Difference – Computational Profiling and Fairness to Individuals, in *Vöneky et al.*, The Cambridge Handbook of Responsible Artificial Intelligence – Interdisciplinary Perspectives (erscheint 2022). Statistische Diskriminierungen sind ungerechtfertigte Ungleichbehandlungen mithilfe von Ersatzinformationen, vgl. *Orwat*, Diskriminierungsrisiken durch Verwendung von Algorithmen, Antidiskriminierungsstelle des Bundes, 2019, 5, mit Nachweisen.

19 Vgl. *UN-Ausschuss für die Beseitigung der Rassendiskriminierung*, General Comment No. 36, CERD/C/GC/36 (2020), 7.

20 *Wischmeyer*, Regulierung intelligenter Systeme, AöR 143 (2018) 1, 55.

21 *Zum Begriff der Transparenz allgemein vgl. Meijer*, Transparency, in Bovens/Goodin/Schillemans, The Oxford Handbook of Public Accountability, 2014, 507. Zur Transparenz von KI-Systemen vgl. u.a. *Floridi et al.*, AI4People: An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations, Minds and Machines 28 (2018), 689, 700; *Wischmeyer*, Artificial Intelligence and Transparency: Opening the Black Box, in *Wischmeyer/Rademacher*, Regulating Artificial Intelligence (2019), 76. Zur Kritik der Abstraktheit vgl. *Krishnan*, Against Interpretability: a Critical Examination of the Interpretability Problem in Machine Learning, Science & Philosophy 2019, Vol. 33, 487.

eine (menschen-)rechtliche Perspektive ergänzt. Zunächst ist dabei eine Definition von KI-Systemen (II.) und eine Einführung in das Problem der Intransparenz KI-basierter Entscheidungen erforderlich (III.). Sodann wird untersucht, welche Anforderungen an die Transparenz KI-basierter Entscheidungen sich aus den völkerrechtlich verbindlichen Menschenrechten, wie sie im Internationalen Pakt über bürgerliche und politische Rechte (IPBPR)[23] verankert sind, ableiten lassen (IV.). Der daraus abgeleitete konkrete Regulierungsvorschlag soll mit bestehenden Regulierungsansätzen kritisch verglichen werden (V.). Schließleich werden die gefundenen Ergebnisse zusammengefasst (VI.).

II. Definition von KI-Systemen

Dieser Beitrag orientiert sich an der Definition von KI-Systemen, wie sie in dem von der EU-Kommission im April 2021 veröffentlichten Vorschlag für eine KI-Verordnung[24] enthalten ist. Demnach ist ein KI-System „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".

Die im genannten Anhang I aufgeführten Techniken umfassen sowohl Ansätze des maschinellen Lernens, einschließlich *Deep Learning*, als auch logik- und wissensbasierte Ansätze, inklusive Expertensysteme, sowie verschiedene statistische Ansätze, wie etwa Bayessche Netze. Dem Vorschlag zufolge kann die im Anhang enthaltene Liste von der EU-Kommission erweitert und angepasst werden, wenn neue Techniken auftauchen.

Diese Definition vermeidet eine Auseinandersetzung mit dem juristisch kaum zu definierenden Begriff der Intelligenz und beschreibt stattdessen die wesentlichen Eigenschaften und Funktionsweisen von KI-Systemen. Die Definition ist zudem hinreichend bestimmt und gleichzeitig offen und flexibel hinsichtlich neuer technologischer Entwicklungen im Bereich der KI.[25]

III. Das Problem der Intransparenz KI-basierter Entscheidungen und dessen rechtliche Konsequenzen

1. Ursachen der Intransparenz KI-basierter Entscheidungen

Wenn in diesem Beitrag von der Intransparenz KI-basierter Entscheidungen gesprochen wird, so ist die fehlende Kenntnis der Gründe für die KI-basierte Entscheidung durch Betroffene gemeint. Diese erfahren nicht, warum die sie betreffende Entscheidung gerade so und nicht anders ausgefallen ist.[26] Sie können daher nicht feststellen, aus welchen Gründen ihnen ein KI-System beispielsweise ein hohes Rückfallrisiko oder eine geringe Kreditwürdigkeit attestiert hat, weshalb sie von einem KI-System als Gefährder eingestuft wurden oder weshalb ihre Bewerbung für einen Ausbildungs- oder Arbeitsplatz abgelehnt wurde. Diese Entscheidungsintransparenz hat verschiedene Ursachen, die in der Regel nicht isoliert, sondern kumulativ auftreten.

Informationen, wie etwa die Gewichtung einzelner Daten oder die relevante Vergleichsgruppe, anhand derer die Rückfallgefahr, Kreditwürdigkeit oder Gefährdereigenschaft der betroffenen Person beurteilt werden, sind häufig als Geschäftsgeheimnis geschützt.[27] Denn KI-Systeme werden in der Regel von privaten Unternehmen entwickelt, die ein – zum Teil auch verfassungs-[28] und völkerrechtlich[29] geschütztes – Interesse an der Geheimhaltung dieser Informationen haben. So verneinte bspw. der BGH in seinem *Schufa-Urteil* 2014 in Bezug auf das deutsche Recht einen Anspruch auf Auskunft gegen die *Schufa* über die genaue Gewichtung der Daten bei der Erstellung des Scores, da diese Information als Geschäftsgeheimnis geschützt sei.[30]

Ein weiterer Grund für die Intransparenz KI-basierter Entscheidungen ist deren hochkomplexes, für Menschen schlechthin nicht mehr nachvollziehbares Zustandekommen. Das gilt nicht für symbolische Ansätze oder lernende Systeme, die ihre Outputs auf lineare Regressi-

22 *Bartneck et al.*, An Introduction to Ethics in Robotics and AI, 2020, 36.

23 International Covenant on Civil and Political Rights*, 19.12.1966, 999 UNTS 171.*

24 *Europäische Kommission*, Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Gesetz über künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union COM(2021) 206 final (KI-VO).

25 Kritik jedoch in *Müller*, Der Artificial Intelligence Act der EU: Ein risikobasierter Ansatz zur Regulierung von Künstlicher Intelligenz, EuZ 1 (2022), A 14, wonach der Begriff der algorithmischen Entscheidungsfindung (ADM) sachgerechter gewesen wäre.

26 Vgl. *Burrell*, How the machine 'thinks': Understanding opacity in machine learning algorithms, Big Data & Society, 2016, 1.

27 Hierauf beruft sich bspw. die *Schufa* bzgl. des von ihr verwendeten Algorithmus, AG Wiesbaden, ZD 2022, 121, 123 Rn. 29.

28 Für die deutsche Verfassung mit zahlreichen Nachweisen vgl. *Martini*, Blackbox Algorithmus – Grundfragen einer Regulierung Künstlicher Intelligenz, 2019, 37 ff.

29 *Siehe unten unter III.3.*

30 BGH NJW *2014, 1235, 1237.*

onen oder Entscheidungsbäume stützen,[31] wohl aber für solche KI-Systeme, die auf tiefen neuronalen Netzen und der Methode des Deep Learning basieren.[32] Die nichtlineare Interaktion der verschiedenen neuronalen Knoten, die Gewichtung und Kombination einer Vielzahl abstrakter Variablen sowie die riesigen Datenmengen, die von den Systemen verarbeitet werden, machen es auch für die Entwickler unmöglich, die genauen Gründe für einen konkreten Output eines KI-Systems festzustellen.[33] Zudem ändern sich die Verknüpfungen der verschiedenen Knoten und damit die Gewichtung und Kombination der Variablen mit jedem Lernprozess.[34] Dadurch lässt sich kaum noch bestimmen, welche Variablen zum Zeitpunkt einer bestimmten Ausgabe signifikant waren, da sich deren Klassifizierung möglicherweise bereits geändert hat. Aus diesem Grund werden solche KI-Systeme häufig als „Black Boxes"[35] bezeichnet, deren Innenleben für uns „opak" ist.[36] Ob das System bei einer bestimmten Entscheidung an unzulässige, diskriminierende Parameter angeknüpft hat, können insofern weder die Betroffenen noch die Entwickler selbst feststellen.

2. Rechtliche Konsequenzen der Intransparenz KI-basierter Entscheidungen

Die Intransparenz KI-basierter Entscheidungen hat zur Folge, dass sich ungerechtfertigte Diskriminierungen in einzelnen Fällen nicht nachweisen lassen.[37] Bewertet ein KI-System Menschen, so besteht eine nicht zu unterschätzende Gefahr statistischer Diskriminierungen.[38] Menschen werden anhand bestimmter Eigenschaften einer Vergleichsgruppe zugeordnet und auf Grundlage von dieser Gruppe anhaftenden Wahrscheinlichkeitswerten beurteilt.[39] Dabei können geschützte Merkmale wie ethnische Herkunft, Geschlecht, Religionszugehörigkeit oder Alter direkt oder indirekt[40] als Variablen in die Vergleichsgruppenbildung einfließen. Führt die derartige Anknüpfung an ein geschütztes Merkmal zu einer für die betroffene Person nachteiligen Entscheidung, so kann dies eine *prima facie* verbotene (statistische) Diskriminierung darstellen.

Ein Beispiel für eine indirekte Diskriminierung aufgrund der Hautfarbe ist die Software *COMPAS*. Das System wird in den USA von Gerichten verwendet, um das Rückfallrisiko von straffälligen Personen zu bewerten.[41] Die für die Risikoeinschätzung erforderlichen Informationen bezieht *COMPAS* aus den Strafakten und aus Selbstauskünften der Angeklagten.[42] Obwohl *COMPAS* bei der Erstellung des Scores keine Informationen über die ethnische Zugehörigkeit erhält, wurde Personen mit dunkler Hautfarbe fast doppelt so häufig ein hohes Rückfallrisiko attestiert wie Personen mit heller Hautfarbe, ohne dass sich dieses realisierte.[43] Da sich der Entwickler über die interne Logik und Funktionsweise des Systems

31 *Bibal* et al, 'Legal requirements on explainability in machine learning' (2021) 29 Artificial Intelligence and Law 149, 149, 161.

32 *Burrell*, How the machine 'thinks': Understanding opacity in machine learning algorithms, Big Data & Society, 2016, 5. Diese Form der KI ist in ihrer Funktion dem menschlichen Gehirn nachempfunden. Sie bestehen aus mehreren Schichten miteinander verknüpfter sog. künstlicher neuronaler Netze. Jeder einzelne neuronale Knoten kombiniert einen bestimmte Inputwert, um einen Outputwert zu erzeugen, der wiederum an die andere nachgeschaltete Einheit weitergegeben wird, vgl. etwa *Kelleher*, Deep Learning (2019), 65 ff.

33 *Martini*, Blackbox Algorithmus – Grundfragen einer Regulierung Künstlicher Intelligenz, 2019, 41, *Zednik*, Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence, Philosophy & Technology, Philosophy & Technology (2019), 3.

34 *Burrell*, How the machine 'thinks': Understanding opacity in machine learning algorithms, Big Data & Society, 2016, 5.

35 Vgl. *de Laat*, Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?, Philosophy & technology 2018, Vol. 31, 525, 536. Zum Begriff der Black Box aus techniksoziologischer Perspektive vgl. *Miebach*, Soziologische Handlungstheorie, 5.Aufl. 2022, 613.

36 *Krishnan*, Against Interpretability: a Critical Examination of the Interpretability Problem in Machine Learning, Science & Philosophy, Philosophy and Technology 2019, Vol. 33, 487, 488.

37 *UN-Ausschuss für die Beseitigung der Rassendiskriminierung*, General Comment No. 36, CERD/C/GC/36 (2020), 7.

38 *Barocas/Selbst*, Big Data's Disparate Impact, California Law Review 2016, Vol. 104, 671.

39 *Martini*, Algorithmen als Herausforderung für die Rechtsordnung, JZ 2017, 1017, 1018.

40 Eine indirekte Einbeziehung geschützter Merkmale liegt vor, wenn Variablen, wie bspw. der Wohnort oder das Einkommen stark mit dem geschützten Merkmal korrelieren.

41 *Lischka/Klingel*, Wenn Maschinen Menschen bewerten – Internationale Fallbeispiele für Prozesse algorithmischer Entscheidungsfindung, Arbeitspapier der Bertelsmann Stiftung, 05.2017, 9.

42 *Supreme Court of Wisconsin*, State v. Lomis, 881 N.W.2d 749 (Wis. 2016), Rn. 13.

43 *Angwin et al.*, Machine Bias, ProPublica, 23.05.2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. Dieses Ergebnis wurde jedoch vom Entwickler Northpointe bestritten und den Autoren wurden statistische Fehler vorgeworfen, Dietrich/Mendoza/Brennan, COMPAS Risk Scales:
Demonstrating Accuracy Equity and Predictive Parity, 2016, 1, 20 ff. In diesem Streit wurden unterschiedliche, miteinander nicht vereinbare Fairnesskonzepte zugrunde gelegt. Dass am Ende jedenfalls ein System steht, dass bei dunkelhäutigen Betroffenen doppelt so häufig wie bei hellhäutigen fälschlicherweise eine hohe Rückfallgefahr annimmt, wird von Northpointe allerdings nicht bestritten. Die diskriminierenden Tendenzen bei der KI-basierten Bewertung des Rückfallrisikos wurden zudem auch bei anderen Modellen nachgewiesen, vgl. etwa *Tolan et al.*, Why Machine Learning May Lead to Unfairness: Evidence from Risk Assessment for Juvenile Justice in Catalonia, ICAIL '19, 17.-21. Juni 2019.

bedeckt hält, konnte dies erst nach statistischer Auswertung von über 11 000 Fällen festgestellt werden.[44] Welche Entscheidung im Einzelfall tatsächlich diskriminierend war, lässt sich über solche statistischen Methoden nicht feststellen, was jedoch aus juristischer Sicht erforderlich ist.[45] Gerechtigkeit ist individualisiert[46]: Nur durch die Möglichkeit, eine im konkreten Fall sie betreffende Diskriminierung festzustellen, können die jeweils Betroffenen ihre subjektiven Rechte effektiv wahrnehmen.[47]

3. Qualitative Unterschiede zu anderen „Black Boxes"

Problematisch ist die Intransparenz KI-basierter Entscheidungen besonders dort, wo sie sich qualitativ von anderen Transparenzproblemen unterscheidet und bestehende rechtliche Lösungskonzepte daher nicht unmittelbar anwendbar sind bzw. angepasst werden müssen. So benutzen Menschen im Alltag häufig komplexe Technologien, deren genaue Funktionsweise sie nicht verstehen, seien dies Autos, Computer oder Smartphones. Auch bei vielen Arzneimitteln kann zwar festgestellt werden, dass sie wirken, nicht jedoch erklärt werden, warum dies der Fall ist.[48] In diesen Bereichen nehmen die Nutzer Intransparenz jedoch in Kauf, da der Nutzen die Risiken überwiegt.

Eine solche Risiko-Nutzen-Abwägung überzeugt bei KI-basierten Entscheidungen, denen Menschen unfreiwillig ausgesetzt und bei denen die Betroffenen nicht die Nutzer des Systems sind, jedoch nicht. Ein solch utilitaristischer Ansatz würde dem Grundsatz der Universalität und Unveräußerlichkeit der Menschenrechte[49] widersprechen.[50]

Vor allem geht es bei Arzneimitteln und anderen teilweise in ihrer Wirkungsweise intransparenten Produkten und Technologien nicht um die Intransparenz von Entscheidungen. KI-Systeme sind die einzige Technologie, die autonom Outputs erzeugen und damit aufgrund eigener Schlussfolgerungen diskriminieren kann.

Die sich insofern von anderen Formen der Intransparenz qualitativ unterscheidende Intransparenz von Entscheidungen existiert zwar auch, wenn Menschen Entscheidungen treffen. Auch Menschen sind in diesem Sinne „Black Boxes".[51] Sie sind jedoch in der Lage, Betroffenen die tragenden Gründe für die Entscheidung mitzuteilen und können dadurch Entscheidungstransparenz herstellen. Zwar sind dies nicht immer die ursächlichen Gründe, sie versetzen Betroffene und Gerichte jedoch in die Lage zu prüfen, ob die Entscheidung auf diskriminierenden Parametern oder unzutreffenden Annahmen basiert. Sofern jedoch weder Nutzer noch Entwickler in der Lage oder willens sind, die tragenden Gründe für eine KI-basierte Entscheidung offenzulegen, sind diese keiner derartigen Überprüfung zugänglich.

IV. Menschenrechtliche Vorgaben an die Transparenz KI-basierter Entscheidungen

Die völkerrechtlich verbindlichen Menschenrechte sind sowohl in universellen Menschenrechtsverträgen wie dem IPBPR mit 173 Vertragsparteien[52], als auch in regionalen Abkommen wie der Europäischen Menschenrechtskonvention (EMRK)[53] sowie zum Teil auch völkergewohnheitsrechtlich[54] verankert. Aufgrund dieser Verbreitung und Verankerung stellen sie eine hinreichend legitimierte Basis für eine verhältnismäßige KI-Regulierung dar.[55] Die für menschliche Entscheidungen aus den Menschenrechten ableitbaren Transparenzanforderungen lassen sich auch auf KI-basierte Entscheidungen anwenden. Dadurch kann die in der Literatur häufig abstrakt bleibende Forderung nach Transparenz von KI-Systemen angemessen und interessensgerecht konkretisiert werden. Nachfolgend werden exemplarisch einige relevante Rechte, wie sie im IPBPR enthalten sind erörtert. Zu unterscheiden ist zwischen Entscheidungen staatlicher und nichtstaatli-

44 *Larson/Mattu et al.*, How We Analyzed the COMPAS Recidivism Algorithm, ProPublica, 23.05.2016, https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

45 *Kleinberg et al.*, Discrimination in the Age of Algorithms, Journal of Legal Analysis 2018, Vol. 10, 113, 129.

46 Vgl. *Zu Art. 14 IPBPR Menschenrechtsausschuss*, General Comment No. 32, CCPR/C/GC/32, 23.08.2007, Rn. 9.

47 Zu der Frage, ob eine solche Diskriminierung aus sachlichen Gründen gerechtfertigt ist, ist damit freilich noch nichts gesagt. Die Ebene der Rechtfertigung kann jedoch erst erörtert werden, wenn bekannt ist, ob überhaupt eine per se diskriminierende Entscheidung vorliegt. Insofern ist die Forderung nach Entscheidungstransparenz der Frage der Rechtfertigung algorithmischer Ungleichbehandlungen vorgelagert.

48 Vgl. *Zittrain*, Intellectual Debt: With Great Power Comes Great Ignorance, Berkman Klein Center Collection, 2019, https://medium.com/berkman-klein-center/from-technical-debt-to-intellectual-debt-in-ai-e05ac56a502c.

49 Vgl. *World Conference on Human Rights*, Vienna Declaration and Programme of Action, A/CONF.157/23, I.5: „All human rights are universal, indivisible and interdependent and interrelated."

50 Vgl. *Vöneky*, Key Elements of Responsible AI, OdW 2020, 9, 19.

51 *Wischmeyer*, Regulierung intelligenter Systeme, AöR 143 (2018) 1, 54. Vgl. auch die systemtheoretische Begründung *Luhmanns*, wonach psychische Systeme Black Boxes sind, zu denen nur sehr eingeschränkter Zugang besteht, vgl. *Miebach*, Soziologische Handlungstheorie, 5. Aufl. 2022, 689.

52 International Covenant on Civil and Political Rights, United Nations Treaty Collection, https://treaties.un.org/Pages/ViewDetails.aspx?src=TREATY&mtdsg_no=IV-4&chapter=4&clang=_en.

53 Convention for the Protection of Human Rights and Fundamental Freedoms, 4.11.1950, ETS No. 005.

54 *Crawford*, Brownlie's Principles of Public International Law, 9. Aufl. 2019, 618.

55 Für eine menschenrechtsbasierte KI-Regulierung argumentiert auch *Vöneky*, Key Elements of Responsible AI, OdW 2020, 9, 19.

cher (privater) Akteure, da sich bei diesen Fallgruppen jeweils verschiedene menschenrechtliche Vorgaben ergeben.

1. Staatliche KI-basierte Entscheidungen

Nutzen Staaten KI-Systeme, so sind sie bei der Nutzung stets an die Menschenrechte gebunden, die für sie vertraglich oder gewohnheitsrechtlich gelten. Dass die Entscheidung direkt oder indirekt durch ein KI-System und nicht von einem Menschen getroffen wird, ändert an der Anwendbarkeit dieser Menschenrechte nichts.[56]

a) Begründungspflicht aus dem Willkürverbot

Wird bei einer KI-gestützten Bewertung bspw. der Rückfallgefahr oder der Gefährdereigenschaft einer Person direkt oder indirekt an Kriterien wie Hautfarbe, Herkunft, Geschlecht oder Religionszugehörigkeit angeknüpft, kommt ein Verstoß gegen das Recht auf Nichtdiskriminierung aus Art. 26 IPBPR in Betracht.[57] Ebenso verhält es sich, wenn Personen aufgrund der genannten Kriterien staatliche Leistungen versagt[58] oder sie von einer Gesichtserkennungssoftware falsch identifiziert und dadurch fälschlicherweise verdächtigt werden.[59] Sofern eine staatliche Entscheidung jedoch mit einem hohen, durch ein KI-System berechneten Risikowert begründet wird, ohne dessen Zustandekommen zu begründen, kann allein aufgrund dieses Scores im Einzelfall weder durch die Betroffenen noch durch die Gerichte ermittelt werden, ob die Entscheidung diskriminierend war oder nicht.

Solche intransparenten Entscheidungen staatlicher Stellen können bereits für sich genommen einen Verstoß gegen das in Art. 26 S. 1 IPBPR verankerte Recht auf Gleichheit vor dem Gesetz darstellen: Dem Gleichheitssatz lässt sich ein allgemeines Willkürverbot für Entscheidungen der Gerichte und der Verwaltung entnehmen.[60] Das Willkürverbot ist verletzt, wenn eine nachteilige Entscheidung nicht auf vernünftigen und objektiven Gründen beruht und sich daher nicht rechtfertigen lässt.[61] Sofern sich eine für den Betroffenen nachteilige staatliche Entscheidung ausschließlich auf den Output eines intransparenten KI-Systems stützt, dürfte dies eine Verletzung des Willkürverbotes darstellen, da das Vorliegen vernünftiger und objektiver Gründe im Einzelfall nicht überprüfbar ist. Die Verwendung intransparenter KI-Systeme durch die Gerichte sowie die Eingriffs- und Leistungsverwaltung bei der Entscheidungsfindung verstößt folglich gegen den Gleichheitssatz.[62]

b) Begründungspflicht aus dem Recht auf wirksame Beschwerde in Verbindung mit dem Diskriminierungsverbot

Ein Recht auf Information über die tragenden Gründe einer Entscheidung kann sich nach hier vertretener Auffassung zudem aus dem Recht auf eine wirksame Beschwerde (engl. *effective remedy*) in Verbindung mit dem Diskriminierungsverbot ergeben. Art. 2 Abs. 3 lit. a IPBPR bestimmt, dass „jeder der in seinen Rechten verletzt worden ist, das Recht hat, eine wirksame Beschwerde einzulegen". Entgegen dem Wortlaut von Art. 2 Abs. 3 lit. a) IPBPR muss aber eine Menschenrechtsverletzung nicht bereits stattgefunden haben, damit das Recht auf wirksame Beschwerde zum Tragen kommt.[63] Vielmehr genügt bereits die glaubhafte Geltendmachung – sog. *arguable claim*[64] – einer Menschenrechtsverletzung.[65] Wann eine Beschwerde *wirksam* ist, kann nur in Bezug auf die Verletzung eines anderen im IPBPR enthaltenen Rechts beurteilt werden.[66] Geht es um Verletzungen des Diskriminierungsverbotes aus Art. 26 IPBPR, so ist eine Beschwerde im Einzelfall nur wirksam, wenn der betroffenen Person diejenigen Informationen mitgeteilt werden, die sie benötigt, um eine Diskriminierung nachzu-

56 *Fundamental Rights Agency*, Getting the future right – Artificial intelligence and fundamental rights, 2020, https://fra.europa.eu/sites/default/files/fra_uploads/fra-2020-artificial-intelligence_en.pdf, 75; *Kriebitz/Lütge*, Artificial Intelligence and Human Rights: A Business Ethical Assessment, Business and Human Rights Journal 5 (2020), 84, 89.

57 Vgl. *Barfield/Pagallo*, Law and Artificial Intelligence, 2020, 25.

58 *Martini*, Blackbox Algorithmus – Grundfragen einer Regulierung Künstlicher Intelligenz, 2019, 70.

59 *UN-Menschenrechtsrat*, Surveillance and human rights – Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, A/HRC/41/35, 28.05.2019, Rn. 12.

60 *Schabas*, Nowak's CCPR Commentary, 3. Ed. 2019, Article 26, Rn. 16.

61 *Schabas*, Ebd., Rn. 17; *Menschenrechtsausschuss*, Borzov v. Estonia, Communication No. 1136/2002, CCPR/C/81/D/1136/2002, 25.08.2004, Rn. 7.2; zur deutschen Verfassung vgl. u.a. *BVerfG* NJW 1954, 1153, 1153.

62 Ähnlich argumentiert auch die die UN-Sonderberichterstatterin für gegenwärtige Formen des Rassismus: Sie entnimmt dem Recht auf Gleichbehandlung und Nichtdiskriminierung aus der UN-Rassendiskriminierungskonvention die Pflicht der Staaten, im öffentlichen Sektor bei der Nutzung neuer digitaler Technologien Transparenz zu gewährleisten, indem nur überprüfbare Systeme verwendet werden, *UN-Menschenrechtsrat*, Racial discrimination and emerging digital technologies: a human rights analysis – Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance, A/HRC/44/57 (2020), Rn. 57.

63 *Menschenrechtsausschuss*, Kazantis v Cyprus, Communication No. 972/2001 (2003), Rn. 6.6.

64 *EGMR*, Guide on Article 13 of the Convention – Right to an effective remedy (last updated 31.08.2021) https://www.echr.coe.int/Documents/Guide_Art_13_ENG.pdf, Rn. 10 ff.

65 *Menschenrechtsausschuss*, Kazantis v Cyprus, Communication No. 972/2001 (2003), Rn. 6.6.

66 *Schabas*, Nowak's CCPR Commentary, 3. Ed. 2019, Article 2 CCPR, Rn. 73.

weisen und geltend zu machen. Nach hier vertretener Auffassung müssen sich Betroffene demnach über die tragenden Gründe der sie betreffenden Entscheidung informieren können.[67]

Eine solche Informations- bzw. Begründungspflicht ergibt sich aus Art. 2 Abs. 3 lit. a IPBPR in Verbindung mit Art. 26 IPBPR freilich nicht ausdrücklich. Jedoch können völkerrechtliche Verträge nach dem Grundsatz der „*necessary implication*"[68] auch nicht explizit garantierte Rechte enthalten, wenn dies zur Erreichung des Regelungsziels erforderlich ist.[69] Dies ergibt sich auch aus dem völkerrechtlich weithin anerkannten Effektivitätsgrundsatz[70], wonach Verträge so auszulegen sind, dass sie ihren Zweck bestmöglich erreichen.[71] Bei Anwendung dieser Auslegungsgrundsätze gelangt man zu einer staatlichen Pflicht, Betroffenen in Bereichen, in denen Diskriminierungen möglich sind, Informationen über die tragenden Gründe der Entscheidung mitzuteilen. Ohne eine solche Pflicht wäre kein effektiver Schutz des Diskriminierungsverbotes und des Rechts auf eine wirksame Beschwerde möglich, da Diskriminierungen im Einzelfall weder identifizierbar noch nachweisbar wären.[72] Der Annahme einer Begründungspflicht ließe sich zwar entgegenhalten, dass eine wirksame Beschwerde bereits bei einer Beweislastumkehr hinsichtlich der Diskriminierung möglich wäre.[73] Da die Widerlegung einer Diskriminierung jedoch nur durch die Darlegung der tragenden Gründe für die Entscheidung möglich wäre, liefe eine solche Beweislastumkehr letztlich ebenfalls auf eine Begründungspflicht hinaus.

Dieses Recht auf Information über die tragenden Gründe einer Entscheidung muss auch gelten, wenn Staaten für die Entscheidungsfindung KI-Systeme nutzen. Dies ergibt sich aus einer dynamischen Vertragsauslegung[74] und unter Berücksichtigung des Sinns und Zwecks des IPBPR, möglichst effektiven Menschenrechtsschutz zu gewährleisten.[75] Denn gegen eine ungerechtfertigte Diskriminierung durch ein KI-System können Betroffene nur vorgehen, wenn ihnen Informationen zur Verfügung gestellt werden, die die Identifikation oder jedenfalls die glaubhafte Geltendmachung einer Diskriminierung ermöglichen. Bei staatlichen KI-basierten Entscheidungen müssen sich diese Informationen auf das konkrete Zustandekommen des Outputs beziehen. Die Betroffenen müssen wissen, welcher Vergleichsgruppe sie aufgrund welcher Eigenschaften zugeordnet wurden.[76] Da es sich bei den genannten Informationen um die tragenden Gründe der Entscheidung handelt, kann insofern auch hier von einer Begründungspflicht gesprochen werden.[77]

2. Nichtstaatliche KI-basierte Entscheidungen

Die Hauptentwickler und -verwender von KI-Systemen sind jedoch keine staatlichen, sondern private Akteure, insbesondere multinationale Großkonzerne.[78]

a) Keine indirekte Bindung an die Menschenrechte

Inwieweit Private und insbesondere Wirtschaftsunternehmen an die Menschenrechte gebunden sind, ist im Einzelnen umstritten. Nach ganz überwiegender Auffassung sind die Menschenrechte nicht direkt horizontal anwendbar.[79] Eine unmittelbare Bindung transnationaler Unternehmen (engl. *transnational corporations*, kurz TNCs) an die Menschenrechte scheitert zum einen an deren begrenztem Völkerrechtssubjektstatus[80], zum anderen am Fehlen einer universellen völkervertragli-

67 *Zur Herleitung einer Begründungspflicht aus Art. 13 EMRK vgl. EGMR*, Al-Nashif v. Bulgaria, No. 50963/99, § 13. Zur Begründungspflicht aus Art. 47 EU-Grundrechtecharta vgl. *EuGH*, Rs. C-372/09, 373/09 BeckRS 2011, 80245 Rn. 63. Zur Herleitung einer Begründungspflicht aus dem Recht auf effektiven Rechtsschutz aus Art. 19 Abs. 4 GG vgl. BVerfGE 103, 142, 160 f.

68 *IGH*, Reparation for Injuries Suffered in the Service of the United Nations (Gutachten) ICJ Rep. 1949, 174 (182).

69 *v. Arnauld*, Völkerrecht, 4. Aufl. 2019, Rn. 231.

70 *Fitzmaurice*, Interpretation of Human Rights Treaties, in *Shelton*, The Oxford Handbook of International Human Rights Law, 751 f.

71 *v. Arnauld*, Völkerrecht, 4. Aufl. 2019, Rn. 231. Mit dieser Methode entnahm der EGMR in *Golder v United Kingdom* dem Recht auf ein faires Verfahren aus Art. 6 EMRK das Recht auf Zugang zu einem Gericht, vgl. *Smith*, International Human Rights Law, 10. Aufl. 2022, 188.

72 Dass Personen, in deren Rechte die Verwaltung eingreift, nur durch Begründungen ihre Rechte sachgemäß verteidigen können, vertritt auch *BVerwG*, DVBl 1982, 198 f.

73 So etwa *Martini*, Algorithmen als Herausforderung für die Rechtsordnung, JZ 2017, 1017, 1023 f.

74 EGMR, Tyrer v. UK, BeckRS 1978, 108297, Rn. 31; *Menschenrechtsausschuss*, Judge v. Canada, 13.8.2003, Communication No. 829/1998, § 10.3.

75 *Vgl. Shelton*, Advanced Introduction to International Human Rights Law, 2. Aufl. 2020, 114 ff.

76 In diese Richtung auch *Martini*, Blackbox Algorithmus – Grundfragen einer Regulierung Künstlicher Intelligenz, 2019, 72, allerdings nur zum deutschen Verfassungsrecht.

77 Zu einer sich aus Art. 3 GG ergebenden Begründungspflicht in Bezug auf algorithmenbasierte Entscheidungen vgl. *Kischel* in *Epping/Hillgruber*, BeckOK GG, Art. 3 Rn. 218c.

78 *UN-Menschenrechtsrat*, Racial discrimination and emerging digital technologies: a human rights analysis – Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance, A/HRC/44/57, 18.06.2020, Rn. 15.

79 *Menschenrechtsausschuss,* General Comment No. 31, CCPR/C/21/Rev.1/Add. 13, 26.05.2004, Rn. 8.

80 *Mucholinsky*, Corporations in International Law, MPEPIL, 2014, Rn. 30; *Weilert*, Transnationale Unternehmen im rechtsfreien Raum? Geltung und Reichweite völkerrechtlicher Standards, ZaöRV 2009, 883 910.

chen oder -gewohnheitsrechtlichen Regel, die TNCs unmittelbar an die Menschenrechte bindet.[81]

Zwar existieren auf internationaler Ebene inzwischen zahlreiche Verhaltenskodizes, die sich direkt an TNCs richten, allen voran die *UN Guiding Principles on Business and Human Rights*,[82] der von der UN initiierte *Global Compact*[83], die *OECD-Guidelines for Multinational Enterprises*[84] sowie die *Tripartite Declaration* der ILO.[85] Diese Verhaltenskodizes sind als internationales *soft law* jedoch für TNCs nur moralisch,[86] nicht aber rechtlich verbindlich.[87] In jüngster Zeit gab es zwar verstärkt Bemühungen durch Gerichte, teilweise über die *UN Guiding Principles*[88], teilweise über bilaterale Investitionsschutzabkommen[89], TNCs rechtsverbindliche Menschenrechtspflichten aufzuerlegen. Diese Bemühungen konnten jedoch bislang keine (völkergewohnheits-) rechtskonstituierende Kraft entfalten.

b) Staatliche Schutz- und Sorgfaltspflichten

Es besteht jedoch eine mittelbare Bindung transnationaler Unternehmen an die Menschenrechte, die sich aus staatlichen Schutzpflichten ergibt.[90] Wie sich auch aus Art. 2 Abs. 1 und 2 IPBPR ergibt, enthalten die Menschenrechte nicht nur negative, sondern auch positive Pflichten.[91] Die Staaten sind verpflichtet, Menschenrechtsverletzungen durch nichtstaatliche Akteure im Rahmen der gebotenen Sorgfalt (*due diligence*[92]) zu verhindern.[93]

Bei der Wahrnehmung dieser Schutzpflicht haben die Staaten bzgl. der geeigneten Mittel zwar einen großen Ermessensspielraum.[94] Dieser ist jedoch durch das Untermaßverbot begrenzt.[95] Jedenfalls dort, wo gar keine oder nur offenkundig ineffektive Maßnahmen zum Schutz der Menschenrechte ergriffen werden, verletzen die Staaten ihre Schutzpflichten.[96]

Auch aus dem Wortlaut von Art. 26 S. 2 IPBPR ergibt sich, dass die Staaten auch vor Diskriminierungen durch Private wirksamen Schutz gewährleisten müssen. Es besteht jedoch Einigkeit, dass sich diese Schutzpflicht nicht auf die Verhinderung jedweder Diskriminierung im Alltag beziehen kann.[97] So gehört es zu den Freiheiten jeder Person, nach eigenen Präferenzen zu bestimmen, mit wem sie unter welchen Bedingungen Verträge abschließen will.[98] Effektive Schutzmaßnahmen gegen Diskriminierungen Privater müssen die Staaten jedoch nach herrschender Ansicht im sog. „quasi-öffentlichen Sektor" ergreifen.[99] Dies betrifft solche Bereiche, in denen Private die Entscheidungsmacht besitzen, Dritte von existenziellen Leistungen und der Teilhabe an wichtigen Ressourcen des täglichen Lebens auszuschließen.[100] In einer solchen Stellung befinden sich unter anderem Arbeitgeber.[101] Auch Wirtschaftsauskunfteien wie die *Schufa*, die durch ihre Bewertungen Menschen faktisch von der Gewährung eines Kredits ausschließen können, dürften darunter fallen. Werden KI-Systeme von Arbeitgebern bei der Einstellung von Arbeitskräften oder von Kredit-

81 *Mucholinsky*, Ebd., Rn. 31; *Crawford*, Brownlie's Principles of Public International law, 9. Aufl. 2019, 630.

82 *UN-Menschenrechtsrat*, Report of the Special Representative of the Secretary General on the issue of human rights and transnational corporations and other business enterprises, John Ruggie, A/HRC/17/31, 21.03.2011.

83 *UN-Wirtschafts- und Sozialrat*, Norms on the responsibilities of transnational corporations and other business enterprises with regard to human rights, E/CN.4/Sub.2/2003/12/Rev.2, 26.8.2003.

84 *OECD*, OECD-Guidelines for Multinational Enterprises, 2011, http://www.oecd.org/daf/inv/mne/48004323.pdf.

85 *ILO*, Tripartite declaration of principles concerning multinational enterprises and social policy (MNE declaration), 5. Aufl. 2017.

86 *Clapham*, Non-State Actors, in *Moeckli/Shah/Sivakumaran*, International Human Rights Law, 3. Aufl. 2018, 557, 569.

87 *Shaw*, International Law, 9. Aufl. 2021, 198; *Krajewski*, Wirtschaftsvölkerrecht, 3. Aufl. 2017, Rn. 63.

88 Vgl. *Rechtbank Den Haag*, Milieudefensie v. Royal Dutch Shell PLC, C/09/571932/HA ZA 19-379 (engelse versie), Rn. 4.4.11. ff.

89 Urbaser S.A. and Consorcio de Aguas Bilbao Bizkaia, Bilbao Biskaia Ur Partuergoa v The Argentine Republic, ICSID Case No. ARB/07/26, Award, 08.12.2016, Rn. 1194 ff.; *Krajewski*, A Nightmare or a Noble Dream?, Business and Human Rights Journal 5:1 (2020), 105, 121 ff.

90 *v. Arnauld*, Völkerrecht, 4. Aufl. 2019, Rn. 633.

91 Vgl. *Menschenrechtsausschuss*, General Comment No. 31, CCPR/C/21/Rev.1/Add. 13, 26.05.2004, Rn. 6–8.

92 Zum genauen Inhalt von menschenrechtlichen Sorgfaltspflichten vgl. *Monnheimer*, Due Diligence Obligations in International Human Rights Law, 2021, 204 ff.

93 *Menschenrechtsausschuss*, General Comment No. 31, CCPR/C/21/Rev.1/Add. 13, 26.05.2004, Rn. 8; *UN-Menschenrechtsrat*, Report of the Special Representative of the Secretary General on the issue of human rights and transnational corporations and other business enterprises, John Ruggie, UN Doc. A/HRC/17/31 (2011), Rn. 1 ff.

94 *v. Arnauld*, Völkerrecht, 4. Aufl. 2020, Rn. 665.

95 Ebd., Rn. 665.

96 *Stahl*, Schutzpflichten im Völkerrecht – Ansatz einer Dogmatik, 2012, 315.

97 *Schabas*, Nowak's CCPR Commentary, 3. Ed. 2019, Article 26, Rn. 100.

98 *So das BVerfG in seiner Entscheidung zum Stadion-Verbot, vgl. BVerfG* NVwZ 2018, 813, 816; diese Argumentation gilt auch für den IPBPR, vgl. *Schabas*, Nowak's CCPR Commentary, 3. Ed. 2019, Article 26, Rn. 100 mit Nachweisen.

99 *Englisch:* „quasi-public sector", *vgl. Menschenrechtsausschuss*, Nahlik v. Australia, Communication No. 608/1995 (1996), Rn. 8.2; *Schabas*, Nowak's CCPR Commentary, 3. Ed. 2019, Article 26, Rn. 100 f.

100 Vgl. BVerfG, NVwZ 2018, 813, 815 Rn. 33.

101 *Schabas*, Nowak's CCPR Commentary, 3. Ed. 2019, Article 26, Rn. 100 f.

instituten bei der Bewertung von Kreditbewerbern verwendet, so entfaltet das Diskriminierungsverbot mittelbare Bindungswirkung. Folgerichtig muss auch in diesen Fällen eine wirksame Beschwerde gegen Diskriminierungen möglich sein. Dies kann einerseits ebenfalls aus dem Recht auf eine wirksame Beschwerde aus Art. 2 Abs. 3 lit. a IPBPR abgeleitet werden. Andererseits ergibt es sich aus der Pflicht aus Art. 26 S. 2 IPBPR, wirksamen Schutz gegen Diskriminierungen im quasi-öffentlichen Sektor zu gewährleisten.

Begründet man eine mittelbare Drittwirkung des Diskriminierungsverbots mit einer quasi-staatlichen Funktion, so kann daran auch eine Informations- bzw. Begründungspflicht bei intransparenten Entscheidungen angeknüpft werden.[102] Ohne Mindestinformationen über die tragenden Gründe einer Entscheidung wären Betroffene andernfalls gänzlich schutzlos gestellt. Der Beweis einer Diskriminierung, auch der Beweis von Indizien für eine solche, wäre kaum möglich.[103] Freilich ist hier der souveränitätsbedingt weite staatliche Ermessensspielraum, sowie die abgestufte, mittelbare Bindung der privaten Akteure zu beachten. Dies kann berücksichtigt werden, indem in einfachgesetzlichen Konkretisierungen nur das geringste Mindestmaß an Informationen über das konkrete Zustandekommen der Entscheidung gefordert wird.[104]

3. Rechtfertigung

Einschränkungen der Menschenrechte können jedoch gerechtfertigt sein, wenn sie zur Erreichung eines legitimen Ziels erforderlich und angemessen sind.[105] In der Diskussion um KI-Systeme wird häufig argumentiert, dass deren Einsatz schnellere, genauere und objektivere Ergebnisse hervorbringe. Allein dieses Argument vermag die Intransparenz KI-basierter Entscheidungen und die Hinnahme etwaiger Diskriminierungen jedoch nicht zu rechtfertigen.

Zum einen wird die Behauptung, KI-basierte Prognoseentscheidungen seien genauer und objektiver als menschlichen Entscheidungen berechtigterweise in Zweifel gezogen. So ergab eine Studie etwa, dass *COMPAS* nicht genauer entscheidet als zufällig ausgewählte Internetnutzer.[106] Jedenfalls darf die Intransparenz einer Entscheidung zugunsten von mehr Effizienz nicht dazu führen, dass Betroffene gar keine wirksame Beschwerde gegen KI-basierte Diskriminierungen erheben können. Dies wäre eine unverhältnismäßige Einschränkung des Rechts auf wirksame Beschwerde.

Bei der Nutzung von KI-Systemen durch Private hat der Staat bei der Ergreifung angemessener Schutzmaßnahmen indes auch die Menschenrechte und sonstige rechtlich geschützte Interessen der Nutzer bzw. Entwickler zu berücksichtigen. Dem Interesse der Betroffenen, die Gründe für die Entscheidung zu erfahren, um dagegen rechtlich vorgehen zu können, steht das Interesse der Entwickler gegenüber, die genaue Funktionsweise aus wirtschaftlichen Gründen geheim zu halten. Je nach Konstellation können KI-Systeme bspw. als Investition in den Schutzbereich bilateraler oder multilateraler Investitionsschutzabkommen fallen.[107] Computerprogramme und Geschäftsgeheimnisse werden zudem explizit im TRIPS-Abkommen[108] geschützt.[109] Der Schutz von Geschäftsgeheimnissen kann jedoch nicht so schwer wiegen, dass er komplette Intransparenz zu rechtfertigen vermag. Vielmehr sind die kollidierenden Interessen miteinander in Ausgleich zu bringen.[110] Die Pflicht der vollständigen Offenlegung des Quellcodes gegenüber der betroffenen Person dürfte dabei unverhältnismäßig, aber auch nicht zielführend sein: Da die meisten Menschen nicht in der Lage sind, Computercodes zu lesen und zu verstehen, dürfte dies ohnehin für die wenigsten Betroffenen von Nutzen sein. Vielmehr genügt es, wenn Betroffene über die vom System verwendeten Daten, deren Gewichtung im Einzelfall und die Einordnung in die jeweilige Vergleichsgruppe informiert werden. Zwar werden auch hierdurch bereits sensible Details über die Funktionsweise des verwendeten KI-Systems offenbart.[111] Dies ist jedoch hinzunehmen, da andernfalls für

102 So auch das *BVerfG* in seiner Entscheidung zum Stadionverbot, NVwZ 2018, 813, 816 Rn. 45.

103 So müssen Betroffene nach § 22 AGG nur Indizien für eine Diskriminierung beweisen. Indizien für eine ungleiche Bezahlung aufgrund des Geschlechts können allerdings nur bewiesen werden, wenn die betroffene Person Zugang zu Informationen über die Gehälter der Mitarbeiter hat. Hier ergibt sich in Deutschland ein Auskunftsanspruch aus dem Entgelttransparenzgesetz.

104 Zur Rechtfertigung aufgrund rechtlich geschützter Geschäftsgeheimnisse siehe unten IV.3.

105 *Menschenrechtsausschuss*, General Comment No. 31, CCPR/C/21/Rev.1/Add. 13, 26.05.2004, Rn. 5.

106 *Dressel/Farid*, The accuracy, fairness, and limits of predicting recidivism, Science Advances 2018 Vol. 4: eaao5580, 1.

107 In Betracht kommt auch der Schutz von Geschäftsgeheimnissen über das Recht auf (geistiges) Eigentum gem. Art. 15 Abs. 1 lit. c IPWSKR oder Art. 1 des ersten Zusatzprotokolls zur EMRK.

108 Agreement on Trade-Related Aspects of Intellectual Property Rights (as amended on 23 January 2017), https://www.wto.org/english/docs_e/legal_e/31bis_trips_01_e.htm.

109 Vgl. Art. 10 zu Computerprogrammen und Art. 39 zum Schutz von Geschäftsgeheimnissen; *Barfield/Pagallo*, Law and Artificial Intelligence, 2020, 172.

110 Problematisch ist freilich, dass das TRIPS keine allgemeine Schrankenklausel enthält. Inwiefern Einschränkungen von Art. 10 und 39 TRIPS damit überhaupt noch möglich sind und welche Implikationen sich hieraus für den Schutz mit dem Geheimnisschutz kollidierender Menschenrechte ergeben, scheint deshalb fraglich.

111 So die Argumentation der *Schufa*, die deshalb diese Informatio-

die Betroffenen kein Menschenrechtsschutz möglich und infolgedessen das Untermaßverbot verletzt wäre. Außerdem lässt sich ein angemessener Interessensausgleich bspw. auch dadurch erzielen, dass die Offenlegung nur gegenüber zur Geheimhaltung verpflichteten unabhängigen Sachverständigen erfolgt.[112]

4. Zusammenfassung

Die Menschenrechte schreiben Entscheidungstransparenz sowohl bei staatlichen als auch bei nichtstaatlichen Entscheidungen im quasi-öffentlichen Sektor vor. Die tragenden Gründe für die Entscheidung müssen den Betroffenen dann, aber auch nur dann, mitgeteilt werden, wenn andernfalls kein wirksamer Rechtsschutz möglich ist. Diese Voraussetzung ist bei intransparenten KI-basierten Entscheidungen in Bereichen, in denen Diskriminierungen möglich sind, erfüllt: Ohne Kenntnis der tragenden Gründe für eine Entscheidung kann im Einzelfall eine Diskriminierung nicht festgestellt werden.[113]

Aus dieser Erwägung lässt sich folgender Regulierungsvorschlag ableiten: Entscheidungen eines KI-Systems oder auf Basis eines KI-Systems müssen so transparent sein, dass der betroffenen Person diejenigen wesentlichen tatsächlichen und rechtlichen Gründe für die Entscheidung mitgeteilt werden können, deren Kenntnis für den effektiven Schutz der Rechte dieser Person notwendig ist.

Insofern kann auch von der Notwendigkeit der Begründbarkeit KI-basierter Entscheidungen gesprochen werden. Zentral ist die Frage nach dem „Warum" – nicht in einem kausalen, sondern in einem rechtfertigenden Sinn.[114] So geben Menschen, wenn sie eine Entscheidung begründen, nicht notwendig die *wirklich* ursächlichen Gründe für die Entscheidung an. Dies würde etwa die Verknüpfung bestimmter Neuronen beinhalten. Vielmehr beinhaltet eine juristisch tragfähige Begründung nur diejenigen Gründe, die aus Sicht des Entscheidungsträgers die Entscheidung tragen.[115] Nicht erforderlich – und jedenfalls für Laien auch nicht sinnvoll[116] – ist es, in diesen Bereichen Betroffenen den Quellcode des Systems offenzulegen. Vielmehr müssen die tragenden Gründe für den Output für die betroffene Person verständlich und nachvollziehbar sein und sie in die Lage versetzen, eine mögliche Rechtsverletzung zu identifizieren und vor Gericht glaubhaft und substantiiert darzulegen. Hierzu kann es bspw. erforderlich sein, die betroffene Person darüber zu informieren, auf welchen Daten die Klassifikation durch das algorithmische System basiert, welcher Vergleichsgruppe sie durch das System zugeordnet wurde, weshalb diese Zuordnung erfolgt ist und wie einzelne Variablen bei der Klassifikation gewichtet wurden.[117]

Nicht in diesem Sinne transparent müssen indes KI-basierte Entscheidungen sein, bei denen der effektive Schutz der Menschenrechte auch anderweitig erreicht werden kann. Dies betrifft etwa materielle Schäden, bei denen lediglich Kausalitätszusammenhänge nicht durch die betroffene Person nachgewiesen werden können, der Schaden selbst jedoch feststeht. In diesen Fällen können eine erleichterte Beweis- und Darlegungslast[118] oder eine Gefährdungshaftung[119] Abhilfe schaffen.

V. Transparenz in bestehenden Regulierungsansätzen

Es fragt sich, inwiefern die soeben herausgearbeiteten, sich aus den Menschenrechten ergebenden Vorgaben an die Transparenz von KI-Systemen in bestehenden Regulierungsansätzen bereits berücksichtigt werden.

1. Die EU-Datenschutzgrundverordnung

Die Intransparenz von KI-Systemen wird in der juristischen Literatur hauptsächlich im Zusammenhang mit dem sich aus der EU-Datenschutzgrundverordnung

nen unter Verschluss hält, vgl. VG Wiesbaden, ZD 2022, 121.

112 Vgl. *de Laat*, Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?, Philosophy & technology 2018, Vol. 31, 525, 536; *Citron/Pasquale*, The Scored Society: Due Process for Automated Predictions, Washington Law Review 2014, Vol. 89, 1, 28.

113 Zu diesem Schluss kommt auch *Martini* in Blackbox Algorithmus – Grundfragen einer Regulierung Künstlicher Intelligenz, 2019, 71.

114 Vgl. *Krishnan*, Against Interpretability: a Critical Examination of the Interpretability Problem in Machine Learning, Science & Philosophy, Philosophy and Technology 2019, Vol. 33, 487, 492.

115 Vgl. zu § 39 VwVfG, *Schüler-Harms* in *Schoch/Schneider*, Verwaltungsrecht, 2021, § 39 VwVfG, Rn. 56.

116 *Wischmeyer*, Artificial Intelligence and Transparency: Opening the Black Box, in *Wischmeyer/Rademacher*, Regulating Artificial Intelligence (2019), 77.

117 Vgl. auch *Martini*, Blackbox Algorithmus – Grundfragen einer Regulierung Künstlicher Intelligenz, 2019, 72. A.A. der BGH in seiner *Schufa*-Entscheidung 2014, der einen solchen Anspruch aufgrund des Schutzes des Betriebs- und Geschäftsgeheimnisses ablehnte, vgl. *BGH* NJW 2014, 1235.

118 Etwa am Vorbild von § 34 GenTG oder § 6 UmwHG, vgl. *Zech*, Künstliche Intelligenz und Haftungsfragen, ZfPW 2019, 198, 218.

119 *Zech*, Ebd., 214 f; *Wendehorst,* Liability for Artificial Intelligence – the Need to Address both Safety Risks and Fundamental Rights Risks, in *Vöneky et al.*, The Cambridge Handbook of Responsible Artificial Intelligence – Interdisciplinary Perspectives (erscheint 2022).

(DSGVO)[120] ergebenden Recht auf Erklärung diskutiert.[121] Im Falle einer ausschließlich auf einer automatisierten Verarbeitung personenbezogener Daten beruhenden Entscheidung gem. Art. 22 Abs. 1 DSGVO haben betroffene Personen gem. Art. 13 Abs. 2 lit. f, 14 Abs. 2 lit. g, 15 Abs. 1 lit. h DSGVO ein Recht auf „aussagekräftige Informationen über die involvierte Logik sowie die Tragweite und die angestrebten Auswirkungen einer derartigen Verarbeitung".

Diese in Art. 15 DSGVO als subjektives Recht ausgestaltete Informationspflicht genügt den sich aus den Menschenrechten ergebenden Anforderungen an Entscheidungstransparenz jedoch nicht. Abgesehen davon, dass sie nur bei ausschließlich automatisierten Entscheidungen ohne menschliche Beteiligung besteht[122], bezieht sich das behauptete Recht auf Erklärung nur auf abstrakte, nicht jedoch auf konkrete Informationen. Eine Pflicht, Betroffenen die tragenden Gründe für die Entscheidung im Einzelfall offenzulegen, lässt sich daraus nicht ableiten. Dies ergibt sich neben dem Wortlaut der Art. 13 Abs. 2 lit. f, Art. 14 Abs. 2 lit. gund Art. 15 Abs. 1 lit. h DSGVO auch aus deren Systematik. Denn Art. 13 und 14 DSGVO beziehen sich auf den Zeitpunkt der Datenerhebung, zu dem eine Begründung der konkreten Entscheidung noch gar nicht möglich ist und fordern daher nur eine *Ex-ante*-Erklärung.[123] Nichts anderes kann somit für das Auskunftsrecht in Art. 15 Abs. 1 lit h DSGVO gelten, das zwar auch nach der Datenerhebung besteht, jedoch mit dem Wortlaut der Art. 13 Abs. 2 lit. f und Art. 14 Abs. 2 lit. g DSGVO identisch ist.[124]

Andere versuchen, ein Recht auf Erklärung aus Art. 22 Abs. 3 DSGVO abzuleiten.[125] Demnach müssen die Verwender des KI-Systems „angemessene Maßnahmen [treffen], um die Rechte und Freiheiten sowie die berechtigten Interessen der betroffenen Person zu wahren, wozu mindestens das Recht auf Erwirkung des Eingreifens einer Person seitens des Verantwortlichen, auf Darlegung des eigenen Standpunkts und auf Anfechtung der Entscheidung gehört." Bereits der Begriff „mindestens" suggeriert, dass es sich in der Vorschrift lediglich um einen Mindeststandard handelt.[126] Ein Recht auf Erklärung ergibt sich hieraus nicht. Etwas Anderes kann sich auch nicht in Zusammenschau mit ErwGr 71 der DSGVO ergeben, wonach bei automatisierten Entscheidungen auch eine „Erläuterung der nach einer entsprechenden Bewertung getroffenen Entscheidung" garantiert werden „sollte". Die Erwägungsgründe können zwar trotz fehlender Verbindlichkeit bei der Auslegung des operativen Teils der Verordnung von Bedeutung sein.[127] In Bezug auf die vorliegende Frage spricht hiergegen jedoch der entgegenstehende Wille des Verordnungsgebers. Denn das in ErwGr 71 formulierte Recht auf Erklärung war auch ursprünglich in Art. 22 DSGVO enthalten, wurde jedoch im Laufe des Gesetzgebungsverfahrens aus der Vorschrift entfernt.[128]

Die DSGVO enthält folglich keine Vorgaben an die Transparenz KI-basierter Entscheidungen, die den Menschenrechten hinreichend Rechnung tragen. Denn sie ermöglicht es Betroffenen nicht, die für eine Entscheidung ursächlichen Gründe zu erfahren und ggf. Diskriminierungen festzustellen und vor Gericht zu beweisen.

2. Der Entwurf der EU-KI-Verordnung

In dem im April 2021 von der EU-Kommission verabschiedeten Entwurf zu einer KI-VO findet sich in Art. 13 eine explizite Vorgabe an die Transparenz von Hochrisiko-KI-Systemen. Diese müssen demnach so konzipiert und entwickelt werden, „dass ihr Betrieb hinreichend transparent ist, damit die Nutzer die Ergebnisse des Systems angemessen interpretieren und verwenden können."

Auch diese Transparenzpflicht ist gemessen an den – oben dargelegten – sich aus den Menschenrechten ergebenden Vorgaben an die Transparenz von KI-Systemen jedoch unzureichend. Insbesondere bleibt völlig unklar, wie der aus juristischer Sicht neue Begriff der Interpre-

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

121 *Wischmeyer*, Regulierung intelligenter Systeme AöR 143 (2018), 49 ff.; *Wachter/Mittelstadt/Floridi*, Why a Right to Explanation of Automated Decision-Making Does Not Exist in the GDPR, International Privacy Law 7 (2017), 76.

122 *Wischmeyer*, Artificial Intelligence and Transparency: Opening the Black Box, in *Wischmeyer/Rademacher*, Regulating Artificial Intelligence (2019), 83, mit Nachweisen. Die Reichweite des Anwendungsbereichs von Art. 22 DSGVO ist im Einzelnen umstritten, vgl. auch die aktuelle Vorlage des VG Wiesbaden, ZD 2022, 121.

123 *Wachter/Mittelstadt/Floridi*, Why a Right to Explanation of Automated Decision-Making Does Not Exist in the GDPR, International Privacy Law 7 (2017), 76, 82.

124 *Wachter/Mittelstadt/Floridi*, Ebd., 82.

125 *Wachter/Mittelstadt/Floridi*, Why a Right to Explanation of Automated Decision-Making Does Not Exist in the GDPR, International Privacy Law 7 (2017), 76, 79, mit Nachweisen.

126 *Wachter/Mittelstadt/Floridi*, Ebd., 80.

127 *EuGH*, NVwZ 1998, 269, 270.

128 Vgl. *Bibal et al.*, Legal requirements on explainability in machine learning Artificial intelligence and Law, 29 (2021), 149, 152; *Wachter/Mittelstadt/Floridi*, Why a Right to Explanation of Automated Decision-Making Does Not Exist in the GDPR, International Privacy Law, 7 (2017), 76, 81.

tierbarkeit zu verstehen ist. Die Kommission übernimmt einen aus der ethisch-philosophischen Debatte[129] zu KI stammenden Begriff, ohne diesen zu definieren. Es bleibt zudem unklar, ob lediglich das Ergebnis an sich interpretierbar sein muss oder ob sich die Transparenzpflicht auch auf das Zustandekommen des spezifischen Ergebnisses bezieht. Der Hinweis auf den Zweck der Regelung in Art. 13 Abs. 1 S. 2 KI-VO, wonach Transparenz es den Nutzern von KI-Systemen unter anderem ermöglichen soll, Anbieter über potenzielle Risiken für den Schutz der Grundrechte von Personen zu informieren, kann die Unklarheiten bei der Auslegung des Begriffs der Interpretierbarkeit nicht beseitigen.

Es handelt sich zudem ausschließlich um eine Pflicht der Anbieter gegenüber den Nutzern. Ein Recht von Betroffenen, die ursächlichen Gründe für die sie betreffende Entscheidung zu erfahren und dieses Recht gegebenenfalls gerichtlich durchzusetzen, fehlt demgegenüber und lässt sich auch durch Auslegung nicht herleiten. Hier fällt die neue KI-VO sogar hinter dem Standard der (ebenfalls unzureichenden) DSGVO zurück, deren Recht auf Erklärung jedenfalls ein subjektives und durchsetzbares Recht darstellt. Es bleibt zu hoffen, dass diese Missstände bis zur geplanten Verabschiedung der Verordnung im September 2022[130] behoben werden.

3. OECD-Empfehlungen zu KI

Einen ersten internationalen Regulierungsansatz stellen die Empfehlungen der OECD zu KI[131] dar. In Prinzip 1.3 heißt es unter der Überschrift „Transparency and explainability":

„AI Actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of art […] to enable those affected by an AI system to understand the outcome, and, […] 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."[132]

Zwar fehlt ein Verweis auf die tragenden Gründe für die konkrete KI-basierte Entscheidung. Positiv zu bewerten ist jedoch der deutliche Fokus auf die Rechte der von einer Entscheidung negativ betroffenen Personen.

Negativ zu bewerten ist jedoch der unverbindliche Wortlaut („should") sowie die Tatsache, dass es sich nur um eine völkerrechtlich unverbindliche Empfehlung, also *soft law*, handelt, die zudem auch nur für die 38 Mitgliedstaaten der OECD.[133] Als *soft law* könnten die Empfehlungen jedoch einen Ausgangspunkt für einen universellen völkerrechtlichen Vertrag zu KI darstellen.

VI. Zusammenfassung und Ausblick

Die Transparenz von Entscheidungen ist jedenfalls in bestimmtem Umfang durch die bürgerlichen und politischen Menschenrechte vorgeschrieben, wenn sie notwendige Voraussetzung für deren effektiven Schutz ist. Entscheidungstransparenz ergibt sich insbesondere als eine Art „Hintergrundrecht"[134] aus dem Recht auf eine wirksame Beschwerde und dem Diskriminierungsverbot. Diese Vorgaben lassen sich auch auf Fälle des Einsatzes von KI-Systemen übertragen: Entscheidungen, die auf dem Output eines KI-Systems basieren, müssen begründet werden, wenn dies für den Schutz verbindlicher Menschenrechte erforderlich ist. Dies gilt insbesondere dort, wo andernfalls Diskriminierungen unentdeckt blieben und wo die Gefahr besteht, dass sich Vorurteile zu sozialen Tatsachen verfestigen.

Entscheidungstransparenz muss dabei nicht nur bei staatlichem Handeln im Rahmen der Eingriffs- und Leistungsverwaltung, sondern auch bei privatem Handeln im sogenannten quasi-öffentlichen Sektor gewährleistet werden. Bei der Verwendung von KI-Systemen durch Private besteht jedoch ein großer regulativer Handlungsspielraum der Staaten. Wenn Staaten nicht in der Lage oder nicht willens sind, die erforderlichen gesetzlichen Maßnahmen zu ergreifen, kann dies zu erheblichen Rechtsschutzlücken für Betroffene führen, ohne dass die handelnden Unternehmen juristisch zur Verantwortung gezogen werden können.

Die DSGVO, der Entwurf der neuen KI-VO und die Empfehlungen der OECD richten sich auch an private

129 *Burri*, The New Regulation of the European Union on Artificial Intelligence – Fuzzy Ethics Diffuse into Domestic Law and Sideline International Law, in *Vöneky et al.*, The Cambridge Handbook of Responsible Artificial Intelligence – Interdisciplinary Perspectives (erscheint 2022).

130 Zurzeit wird der Entwurf noch im Ausschuss für Binnenmarkt und Verbraucherschutz (IMCO) und im Ausschuss für bürgerliche Freiheiten, Justiz und Inneres (LIBE) diskutiert. Das KI-Gesetz soll Ende September von den beiden Ausschüssen gemeinsam verabschiedet werden, vgl. https://www.europarl.europa.eu/news/de/press-room/20220429IPR28228/kunstliche-intelligenz-eu-soll-weltweit-standards-setzen.

131 *OECD*, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, 22.05.2019, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449; Zur (inoffiziellen) deutsche Übersetzung vgl. "Empfehlung des Rats zu künstlicher Intelligenz" http://www.oecd.org/berlin/presse/Empfehlung-des-Rats-zu-kuenstlicher-Intelligenz.pdf.

132 *OECD*, Ebd.

133 Vgl. hierzu auch umfassender *Vöneky*, Key Elements of Responsible AI, OdW 2020, 9, 17 f.

134 *Yeung/Lodge*, Algorithmic Regulation, 2019, 72.

Akteure und schließen damit in der EU verbindlich und für die OECD-Staaten als *soft law* zum Teil bestehende Normierungslücken. Aus menschenrechtlicher Perspektive sind jedoch zumindest die in der DSGVO und im Entwurf der KI-VO enthaltenen Vorgaben an die Transparenz von KI-Systemen nicht geeignet, um die entstehenden Rechtsschutzlücken auf Seiten der von KI-basierten Entscheidungen betroffenen Personen zu schließen.

Der Autor ist akademischer Mitarbeiter am Institut für öffentliches Recht (Abt II: Völkerrecht, Rechtsvergleichung) der Albert-Ludwigs-Universität Freiburg. Er ist dort tätig im Teilprojekt „Ethical, Legal and Societal Analysis of the AI-based Assistive System" (Teilprojektleitung: Prof. Dr. Silja Vöneky, Dr. Philipp Kellmeyer, Prof. Dr. Oliver Müller) des Projektes „AI-Trust: Interpretable Artificial Intelligence Systems for Trustworthy Applications in Medicine" (Projektleitung: Dr. Philipp Kellmeyer) der Baden-Württemberg Stiftung. Er promoviert bei Prof. Dr. Silja Vöneky zum Thema „Menschenrechtliche Vorgaben an die Transparenz KI-basierter Entscheidungen".