DeutscherAnwaltVerein

Stellungnahme

des Deutschen Anwaltvereins durch den Ausschuss Europa und den Ausschuss Zivilrecht

zur Konsultation der Europäischen Kommission zum Weißbuch Künstliche Intelligenz COM(2020) 65 final

Stellungnahme Nr.: 40/2020 Berlin/Brüssel, im Juni 2020

Mitglieder des Ausschusses Europa

- Rechtsanwältin Dr. Claudia Seibel, Frankfurt am Main, (Vorsitzende)
- Rechtsanwältin Béatrice Deshayes, Paris, (Berichterstatterin)
- Rechtsanwalt Prof. Dr. Christian Duve, Frankfurt am Main, (Berichterstatter)
- Rechtsanwalt Prof. Dr. Thomas Gasteyer, LL.M., Frankfurt am Main
- Rechtsanwalt Prof. Dr. Hans-Jürgen Hellwig, Frankfurt am Main
- Rechtsanwalt Dr. Ulrich Karpenstein, Berlin
- Rechtsanwältin Gül Pinar, Hamburg
- Rechtsanwalt Prof. Dr. Dirk Uwer, Düsseldorf
- Rechtsanwalt Michael Jürgen Werner, Brüssel

Zuständig in der DAV-Geschäftsführung und Ansprechpartnerin in Brüssel:

- Rechtsanwältin Eva Schriever, LL.M.

Deutscher Anwaltverein
Littenstraße 11, 10179 Berlin
Tel.: +49 30 726152-0
Fax: +49 30 726152-190
E-Mail: dav@anwaltverein.de

Büro Brüssel
Rue Joseph II 40, Boîte 7B
1000 Brüssel, Belgien
Tel.: +32 2 28028-12
Fax: +32 2 28028-13
E-Mail: bruessel@eu.anwaltverein.de
EU-Transparenz-Registernummer:
87980341522-66

www.anwaltverein.de

Mitglieder des Ausschusses Zivilrecht

- Rechtsanwalt Dr. Christian Bereska, Celle (Vorsitzender und Berichterstatter)
- Rechtsanwalt Dr. Rupert Bellinghausen, Frankfurt (Berichterstatter)
- Rechtsanwalt Dr. Markus Beaumart, Köln
- Rechtsanwalt Dr. Tobias Heinrich Boecken, Berlin
- Rechtsanwältin Petra Heinicke, München
- Rechtsanwältin Dr. Sylvia Kaufhold, Maître en droit, Dresden
- Rechtsanwalt Dr. Dr. h.c. Georg Maier-Reimer, LL.M., Köln
- Rechtsanwalt (BGH) Dr. Michael Schultz, Karlsruhe

Zuständig in der DAV-Geschäftsführung

- Rechtsanwältin Christine Martin

Verteiler

Europa

Europäische Kommission

- Generaldirektion Justiz und Verbraucher
- Generaldirektion Kommunikationsnetze, Inhalte und Technologien

Europäisches Parlament

- Ausschuss für Bürgerliche Freiheiten, Justiz und Inneres
- Rechtsausschuss
- Ausschuss für Binnenmarkt und Verbraucherschutz

Rat der Europäischen Union

Ständige Vertretung der Bundesrepublik Deutschland bei der EU

Justizreferenten der Landesvertretungen

Rat der Europäischen Anwaltschaften (CCBE)

Vertreter der Freien Berufe in Brüssel

DIHK Brüssel

BDI Brüssel

Deutschland

Bundesministerium der Justiz und für Verbraucherschutz

Bundesministerium für Wirtschaft und Energie

Ausschuss für Recht und Verbraucherschutz im Deutschen Bundestag

Ausschuss für Wirtschaft und Energie im Deutschen Bundestag

Ausschuss Digitale Agenda im Deutschen Bundestag

Ausschuss für die Angelegenheiten der Europäischen Union

Arbeitsgruppen Recht der im Deutschen Bundestag vertretenen Parteien

Justizministerien und –senatsverwaltungen der Länder

Rechtsausschüsse der Landtage

Europäische Kommission - Vertretung in Deutschland

Bundesrechtsanwaltskammer

Bundesnotarkammer

Bundesverband der Freien Berufe
Deutscher Richterbund
Deutscher Notarverein e.V.
Deutscher Steuerberaterverband
Bundesverband der Deutschen Industrie (BDI)
GRUR
BITKOM
DGRI
Gewerkschaft der Polizei (Bundesvorstand)
Deutsche Polizeigewerkschaft im DBB
Ver.di, Recht und Politik
Stiftung neue Verantwortung e.V.
DAV-Vorstand und Geschäftsführung
Vorsitzende der DAV-Gesetzgebungs- und Geschäftsführenden Ausschüsse des DAV
Vorsitzende der DAV-Landesverbände
Vorsitzende des FORUMs Junge Anwaltschaft

Presse

Frankfurter Allgemeine Zeitung
Süddeutsche Zeitung GmbH
Berliner Verlag GmbH
Redaktion NJW
Juve-Verlag
Redaktion Anwaltsblatt
Juris
Redaktion MultiMedia und Recht (MMR)
Redaktion heise online
JurPC

Der Deutsche Anwaltverein (DAV) ist der freiwillige Zusammenschluss der deutschen Rechtsanwältinnen und Rechtsanwälte. Der DAV versammelt mehr als 62.000 Rechtsanwältinnen und Rechtsanwälte sowie Anwaltsnotarinnen und Anwaltsnotare, die in 252 lokalen Anwaltvereinen im In- und Ausland organisiert sind. Er vertritt die Interessen der deutschen Anwaltschaft auf nationaler, europäischer und internationaler Ebene.

A. Vorwort[1]

1 Die EU-Kommission hat zu ihrem "Weißbuch Künstliche Intelligenz (KI) - Ein europäischer Ansatz für Exzellenz und Vertrauen" eine Konsultation durchgeführt.

2 Im Zusammenhang mit der Konsultation hat das Referat Grundrechte der Generaldirektion Justiz und Verbraucherschutz der Europäischen Kommission in einer E-Mail vom 2. April 2020 Fragen an den DAV gerichtet. Diese beantwortet der DAV mit dieser Stellungnahme.

3 Die von der EU-Kommission aufgeworfenen allgemeinen Fragen hat der DAV im Konsultationsfragebogen beantwortet.

B. Wesentliche Ergebnisse und Empfehlungen

4 Im Bewusstsein um die zunehmende Bedeutung künstlicher Intelligenz in der modernen Gesellschaft und die möglichen Vorteile, die diese Technologie auch für das Justizwesen schaffen kann, empfiehlt der DAV der Kommission, bei der Ausarbeitung eines neuen Rahmenwerks zur künstlichen Intelligenz die folgenden wesentlichen Erkenntnisse zu berücksichtigen:

5 1) Die Einführung von KI-Systemen im Bereich der Justiz ist mit besonders hohen Grundrechtsrisiken verbunden und sollte daher strengen Anforderungen unterworfen werden.

6 2) Gerichtliche und ähnlich eingriffsintensive verbindliche Entscheidungen staatlicher Instanzen dürfen niemals vollständig automatisiert werden.

7 3) In jedem Fall müssen umfassende und sinnvolle Transparenzpflichten eingehalten werden.

---

[1] Diese Stellungnahme entstand unter Mitwirkung von Frau Viola Zollitsch, der unser Dank gilt.

8 4) Darüber hinaus müssen die Haftungsregeln auf EU-Ebene in Bezug auf KI erweitert werden. Ebenso müssen wirksame Rechtsbehelfs- und Kontrollmechanismen für den Einsatz von KI im Bereich der Justiz und der öffentlichen Verwaltung geschaffen werden.

9 5) Um den von der EU verfolgten menschenzentrierten Ansatz zu gewährleisten, müssen die EU und ihre Mitgliedstaaten, dafür sorgen, dass die zunehmende Automatisierung von Dienstleistungen nicht zu einem Abbau von Arbeitsplätzen im Justizsektor führt, sondern zusätzliche Ausbildungsangebote und einen verstärkten Wissensaustausch für Angehörige von Rechtsberufen im Bereich KI schaffen.

C. Einführung

10 Der Futurist Ray Kurzweil hat vorausgesagt, dass künstliche Intelligenz bis 2029 das Niveau der menschlichen Intelligenz erreichen oder übersteigen könnte. Dabei spielt es keine Rolle, ob der Zeitpunkt der Vorhersage zutrifft. Entscheidend ist, wie wir mit einer Technologie umgehen, die das Potenzial hat, die menschliche Entwicklung zu übertreffen. Daher sind die Herausforderungen, die dieser Konsultation zugrunde liegen, von existenzieller Natur, und es bedarf einer vorausschauenden Regulierung, um eine humane Gesellschaft zu erhalten und die Menschenrechte zu schützen.

11 Heute können wir die rasante Entwicklung von selbstfahrenden Autos oder Robotern im Gesundheitswesen beobachten. Was wir noch nicht in gleichem Maße erfahren haben, ist, wie das menschliche Urteilsvermögen von KI übernommen wird. Wenn wir eine humane Gesellschaft erhalten wollen, in welcher der Mensch weiterhin die endgültigen Entscheidungen trifft, müssen wir jedoch sicherstellen, dass der Mensch die Kontrolle behält. Diese Überlegungen gelten insbesondere für die Bereiche Justiz, Strafverfolgung und öffentliche Verwaltung. Auch in diesen Bereichen, die für das Funktionieren jeder demokratischen Gesellschaft von zentraler Bedeutung sind, schreitet die Digitalisierung – wenn derzeit auch noch in einem frühen Stadium – voran.

12 Die Betonung der Bedeutung einer menschlichen Gesellschaft bedeutet nicht, die Vorteile von Innovation und Fortschritt zu verkennen. Studien haben zum Beispiel gezeigt, dass in manchen Rechtsordnungen weniger als 50% der

Bevölkerung Zugang zum Rechtssystem haben.[2] Technologie – einschließlich KI-basierter Instrumente – kann dazu beitragen, diesen Zugang aufgrund der geringeren Kosten und des einfachen Zugangs zu erweitern. Intelligente Systeme könnten zum Beispiel eingesetzt werden, um die Einreichung von Schriftsätzen und die Ausfertigung von Gerichtsbeschlüssen in Zivilverfahren weitgehend zu automatisieren. Sobald jedoch KI-basierte Technologie im Gerichtssaal oder in Entscheidungsprozessen angewandt wird, könnten Grundrechte ernsthaft beeinträchtigt werden.

13 Anwälte werden ihre Arbeitsmethoden anpassen und hierbei neue Technologien einsetzen, doch werden sie sich weiterhin als Anwälte der Schwächeren und als Hüter der Rechtsstaatlichkeit als übergeordnetes Prinzip von Freiheit und Demokratie, begreifen. Aufgrund dieses Auftrags ist die Anwaltschaft dazu berufen, auf Entwicklungen hinzuweisen, die sich negativ auf die Rechtsstaatlichkeit auswirken könnten.

I. Frage 1: In welchen konkreten Situationen erhöht oder führt der Einsatz von KI-Anwendungen zu Grundrechtsrisiken einschließlich eines hohen Verbraucherschutzniveaus?

14 Solange KI die Organisation von Verwaltungs- oder Gerichtsverfahren oder das gezielte Sammeln relevanter Informationen erleichtert, kann sie denjenigen, die Orientierung suchen oder ein Gerichtsverfahren durchlaufen, sicherlich das Leben erleichtern. Wesentlich problematischer ist es jedoch, wenn KI zur Identifizierung und Beschaffung von Informationen im Rahmen von streitigen Verfahren, insbesondere im Rahmen von Entscheidungsprozessen, eingesetzt wird. Ebenso könnten Grundrechte beeinträchtigt werden, wenn KI-basierte Technologie im Gerichtssaal eingesetzt würde, etwa wenn Gesichtszüge gleich einem Lügendetektor ausgelesen, oder menschliches Verhalten anderweitig interpretiert würde.

15 In den folgenden Abschnitten werden zunächst einige Beispiele für KI-Instrumente, die im Justizsektor eingesetzt werden, untersucht (1). Wir werden

[2] Marr, Bernard, The Future of Lawyers: Legal Tech, AI, Big Data and Online Courts, Forbes, 17. Januar2020, abrufbar unter: <https://www.forbes.com/sites/bernardmarr/2020/01/17/the-future-of-lawyers-legal-tech-ai-big-data-and-online-courts> [abgerufen am 22. Mai 2020].

erklären, warum die letztendliche Entscheidungsgewalt bei einem Richter bzw. einer Richterin liegen muss. Wir werden außerdem darlegen, warum jeder Versuch, einen menschlichen Richter bzw. eine menschliche Richterin durch KI zu ersetzen, das Grundrecht auf ein unabhängiges Gericht und ein faires Verfahren verletzen würde. In einem zweiten Abschnitt werden wir erläutern, dass ähnliche Erwägungen dort angebracht sind, wo Richter oder Richterinnen die Entscheidungsgewalt behalten, sich aber weitgehend auf KI-Systeme stützen. Unter solchen Umständen würde es ihnen sehr viel schwerer fallen, ihr eigenes, unabhängiges Urteil zu fällen. Beispiele für den Einsatz von Predictive justice-Instrumenten in der Strafzumessungspraxis der Vereinigten Staaten und der Niederlanden zeigen in einem dritten Abschnitt, wie real die Bedrohungen der Rechtsstaatlichkeit in Gerichtsbarkeiten außerhalb und innerhalb der EU bereits geworden sind. Im Hinblick auf die Justizverwaltung zeigt ein Beispiel aus Polen, wie die Unabhängigkeit des Justizsystems auch durch den Einsatz von Algorithmen in der Geschäftsverteilung beeinträchtigt werden kann.

16 Außerdem skizzieren wir weitere Beispiele für KI-Systeme, die in der Strafverfolgung (2) und in der öffentlichen Verwaltung (3) eingesetzt werden, und die Risiken für die Grundrechte darstellen und sich direkt oder indirekt auf die Rechtspraxis auswirken können, da diese Systeme entweder in Gerichtsverfahren eingesetzt werden können (z.B. videobasierte Lügendetektoren) oder als Beweismittel in Gerichtsverfahren eine Rolle spielen, da sie die Grundlage für Verwaltungs- oder Vollzugsentscheidungen bilden können (z.B. intelligente Videoüberwachung und Kreditpunktesysteme).

1. In der Justiz verwendete KI-Systeme

17 Während KI-Tools bisher vor allem von Anwälten genutzt wurden, werden einige Anwendungen allmählich auch von der Justiz eingeführt. Solche Anwendungen können grundlegende Elemente des Rechtssystems und damit der Rechtsstaatlichkeit beeinflussen, insbesondere das Recht auf ein faires Verfahren, Rechenschaftspflichten, Unparteilichkeit, Nichtdiskriminierung, Autonomie und das Recht auf ein rechtsstaatliches Verfahren. In den folgenden drei Anwendungsfeldern kann KI von Angehörigen der Justizberufe eingesetzt werden:

18 Erstens kann KI als Hilfsmittel zur Vorhersage eines bestimmten Ausgangs eines Verfahrens eingesetzt werden (sogenannte Predictive justice-Instrumente). Im Extremfall könnte KI dazu verwendet werden, die Entscheidung eines menschlichen Richters bzw. einer Richterin zu ersetzen.

19 Zweitens können KI-Instrumente vor einer Verhandlung als Analyseinstrument eingesetzt werden, z.B. um Datenbanken oder andere Dokumente nach relevanten Informationen zu durchsuchen und auf der Grundlage dieser Ergebnisse Teile schriftlicher Urteile zu erstellen.

20 Drittens kann KI in der Justizverwaltung eingesetzt werden, z.B. als intelligentes Verteilungssystem oder als Werkzeug, um der Öffentlichkeit über Chatbots grundlegende Informationen über Gerichtsverfahren oder mündliche Verhandlungen zu vermitteln.

21 Im Folgenden werden die Risiken dieser verschiedenen Kategorien von KI-Anwendungen untersucht, beginnend mit dem weitreichendsten und damit risikobehafteten Praxisbeispiel, nämlich der Möglichkeit, einen menschlichen Richter bzw. eine Richterin durch KI-Technologie zu ersetzen.

a) KI als Ersatz für eine menschliche Entscheidung vor Gericht

22 In Estland wurde im Rahmen eines Pilotprojekts ein "Roboter-Richter" geschaffen, der über Streitigkeiten mit geringem Streitwert von weniger als 7.000 EUR entscheidet und sich insbesondere auf Vertragsstreitigkeiten konzentriert. Das Konzept sieht vor, dass das KI-System völlig autonom eine Entscheidung trifft, die ausschließlich auf von den Parteien hochgeladenen Dokumenten beruht. Erst in der Berufungsinstanz wird der Fall wird von einem menschlichen Richter entschieden. [3]

23 In ähnlicher Weise wurde durch das so genannte "Cyber-Gericht" in China vor drei Jahren das gesamte Verfahren für die Bearbeitung eines Falles (- von der Einreichung der Klage bis zur Veröffentlichung des Urteils -) in ein Online-Format übertragen. [4] Darüber hinaus führt der Oberste Volksgerichtshof seit Anfang 2019 ein Pilotprogramm für "mobile Gerichte" durch. In diesem "mobilen Gericht"

[3] Niller, Eric, Can AI Be a Fair Judge in Court? Estonia Thinks So, WIRED,25. März 2020, <https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so> [abgerufen am 12. Mai 2020].
[4] Feng, Zhen/Xia, Helen, China: Three Cyberspace Courts now online and open for business, 16.Oktober2018, verfügbar unter: <https://www.jdsupra.com/legalnews/three-cyberspace-courts-now-online-and-91459> [abgerufen am 12. Mai 2020].

verwaltet ein KI-gesteuerter Chatbot-"Richter" Zivilverfahren über die landesweite Social-Media-Plattform WeChat, alle beweisrelevanten Unterlagen werden in eine Blockchain eingegeben. Allerdings liegt die Letztentscheidungsgewalt in diesem Fall wohl noch bei einem menschlichen Richter.[5]

24 Die Ersetzung eines menschlichen Richters bzw einer Richterin durch KI-gesteuerte Software berührt eine Reihe von Grundrechten:

25 Zunächst kann ein Softwareprogramm nach dem heutigen Verständnis von Rechtsstaatlichkeit nicht das Recht jedes Einzelnen erfüllen, von einem unparteiischen und unabhängigen Gericht gehört zu werden, wie es in Art. 47 Abs. (2) der EU-Grundrechtecharta (GRCh) festgelegt ist. Ebenso verlangt das deutsche Grundgesetz (GG), dass niemandem sein "gesetzlicher Richter" entzogen werden darf (Art. 101 Abs. 1 Satz 2 GG).

26 Die Gesetzgebung zeigt, dass es aus der Perspektive des Rechtsstaats stets als erforderlich angesehen wurde, dass Entscheidungen von menschlichen Akteuren getroffen werden. Dies veranschaulichen bereits einfachgesetzliche Regelungen, wie etwa die Bestimmung, dass der Status eines Richters einer Person nur "auf Lebenszeit" verliehen werden kann, (§ 27 Abs. 1 Deutsches Richtergesetz, DRiG). Die Erwartung des Gesetzgebers ebenso wie die Wahl der Sprache setzten somit eindeutig voraus, dass der Richter ein Mensch ist.[6]

27 Als die Schöpfer der Verfassung und des Rechts über den Richter nachdachten, hatten sie mehr im Sinn als eine analytische Maschine. Nach ihrer Vorstellung sollten Menschen von Menschen unter Einsatz ihrer Intuition und Erfahrung sowie ihrer Kenntnisse und Fähigkeiten beurteilt werden. Indem sie sich mit den Argumenten der Parteien auseinandersetzen, erleichtern Richter es den Betroffenen zudem, auch nachteilige Entscheidungen zu akzeptieren. Dies ist eine notwendige Voraussetzung, um Vertrauen in die Justiz zu schaffen.

28 Selbst wenn Art. 47 GRCh anders ausgelegt würde und die Vorschrift grundsätzlich auch einem „Nichtmenschen“ erlauben würde, richterliche Funktionen wahrzunehmen, würde ein solcher keine mündliche Verhandlung im

[5] Cui, Yadong, Artificial Intelligence and Judicial Modernization, Shanghai: Springer 2020, p. 26; Harris, Briony, Could an AI ever replace a judge in court?, 11. Juli 2018, verfügbar unter: <https://www.worldgovernmentsummit.org/observer/articles/could-an-ai-ever-replace-a-judge-in-court> [abgerufen am 22 Mai 2020].
[6] Enders, Peter, Einsatz künstlicher Intelligenz bei juristischer Entscheidungsfindung, JA 2018, 721 (723).

Rechtssinne durchführen können. Das Recht auf rechtliches Gehör, das in Art. 47 Abs. (2) GRCh verankert ist, dient dazu, den Einzelnen in die Lage zu versetzen, an der Entscheidungsfindung im gerichtlichen Verfahren teilzunehmen und diese insofern zu beeinflussen, als er in einen Dialog mit der Entscheidungsinstanz tritt. [7] Ein solcher Dialog, der von gegenseitigem Austausch und Interaktion sowie von Rede und Gegenrede geprägt ist, kann derzeit jedoch mit einer Maschine noch nicht stattfinden. Der Film "*Her*" veranschaulicht allerdings, wie ein solcher Austausch in Zukunft funktionieren könnte. Dementsprechend könnte eine solche mündliche Verhandlung technisch durchaus denkbar sein. Im Fall der Übernahme der Verhandlung durch eine Maschine würde diese den Bereich der Kommunikation von Mensch zu Mensch, von Abwägung und Kontrolle jedoch verlassen.

29 Schließlich wäre es zum jetzigen Zeitpunkt praktisch unmöglich, zu überprüfen, ob ein KI-Richter "unparteiisch" im Sinne des Art. 47 Abs. (2) GRCh ist. Da es unmöglich ist festzustellen , auf welche Kriterien die Maschine ihr Ergebnis stützt, und mit Hilfe welcher – möglicherweise unzureichenden oder voreingenommenen – Informationen zu Sachverhalt und Rechtsprechung diese trainiert wurde, wäre es auch unmöglich, die Unparteilichkeit des Systems zu überprüfen. Darüber hinaus erfordert das Kriterium der Unparteilichkeit auch, dass der entscheidende Richter als Akteur identifiziert werden kann. KI-Systeme haben jedoch keine Rechtspersönlichkeit und könnten daher nicht für eine Entscheidung verantwortlich gemacht werden.

30 Daher würde die Ersetzung einer gerichtlichen Entscheidung durch KI-Technologie eine Verletzung des Rechts auf rechtliches Gehör durch ein unabhängiges und unparteiisches Gericht darstellen.

31 Aus systemischer Sicht könnte der Schutz von Menschenrechten grundlegend gefährdet sein, wenn die Rolle der Anwaltschaft durch eine zunehmende Digitalisierung der Gerichtsverfahren geschwächt würde. Wenn die Rolle der Rechtsanwälte nicht ausreichend geschützt wird, könnte dies schwerwiegende Folgen für Demokratien und Rechtsstaatlichkeit haben, etwa wenn weniger demokratische Regime Algorithmen zum Nachteil kritischer Bürger einsetzten.

[7] Hillebrand Pohl, Jens, The Right to Be Heard in European Union Law and the International Minimum Standard- Due Process, Transparency and the Rule of Law, 8. Juni 2018, verfügbar unter <https://ssrn.com/abstract=3192858>, S. 3 [abgerufen am 12. Mai 2020].

Wenn ein Verfahren etwa ohne mündliche Verhandlung und ohne eine unabhängige menschliche Entscheidung endgültig entschieden würde, wären die bürgerlichen Grundfreiheiten in Gefahr.

b) KI zur Unterstützung des Entscheidungsprozesses

32 Auch wenn kein Zweifel daran bestünde, dass die letzte Entscheidungsgewalt bei einem menschlichen Richter liegt, so könnte doch die Möglichkeit des Einsatzes von KI-Systemen zur Unterstützung des Entscheidungsprozesses in Betracht kommen.

(1) KI-Werkzeuge, die von Richtern und Staatsanwälten zur Unterstützung eingesetzt werden

(2) Einsatz von sogenannten Predictive justice-Instrumente in Gerichtsverfahren

33 Aus grundrechtlicher Sicht sind solche KI-Werkzeuge am eingriffsintensivsten, die darauf abzielen, den Ausgang eines bestimmten Falles vorherzusagen (sogenannte Predictive justice-Instrumente). Solche KI-Instrumente befinden sich derzeit noch in einem frühen Stadium der Entwicklung. Sie werden in erster Linie für Anwälte entwickelt, könnten in Zukunft aber auch von der Justiz angewandt werden. In Frankreich beispielsweise haben bereits zwei zivile Berufungsgerichte in Rennes und Douai im Frühjahr 2017 ein Predictive justice-Instrument (namens "Prédictice") getestet. [8]

34 Im Lichte der obigen Feststellungen liegen die grundrechtlichen Risiken auch hier auf der Hand. Würde die Anwendung eines solchen Systems zu einer automatischen Übernahme der Entscheidung führen, würden Richter Gefahr laufen, nichts weiter als ein Vehikel für maschinengenerierte Entscheidungen zu werden. Wenn die letztendliche Entscheidung zu einer reinen Formalität verkommen würde, würde das Recht auf rechtliches Gehör unangemessen eingeschränkt. Die Zulässigkeit eines solchen Instruments hängt daher davon ab, ob es dem betreffenden Richter[9] einen ausreichenden Ermessensspielraum lässt, eine autonome, unparteiische und unvoreingenommene Entscheidung zu

[8] Ronsin, Xavier/Lampos, Vasileios, Appendix I – In-depth study on the use of AI in judicial systems, notably AI applications processing judicial decisions and data, in: European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment. Strasbourg, CEPEJ - Commission Européene pour l'Efficacité de la Justice, 2018, verfügbar unter: <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>, S. 42 [abgerufen am 12. Mai 2020].

[9] Aus Gründen der Lesbarkeit wurde im Text die männliche Form gewählt, die Angaben beziehen sich jedoch auf Angehörige beider Geschlechter.

treffen. Folglich sollte die Entscheidung eines Richters auf einer Begründung beruhen, die von der Vorhersage des KI-Instruments hinreichend unabhängig ist, so dass eine klare Unterscheidung zwischen beiden gewährleistet wird. Mit anderen Worten sollte ein Richter seine eigene Argumentation klar darlegen, indem er nachprüfbare Gründe für die Befolgung (oder Ablehnung) einer vom KI-System vorhergesagten Entscheidung angibt.

35 Doch selbst in Situationen, in denen Predictive justice-Instrumente nicht automatisch zu einer vordefinierten Entscheidung führt, kann ein Richter durch Daten, die er weder kennt noch beurteilen kann, unangemessen beeinflusst werden. Denn der Richter wird höchstwahrscheinlich nicht die Gründe kennen, aufgrund derer das System zu einer bestimmten Schlussfolgerung gelangt ist, und nicht in der Lage sein, die Daten zu überprüfen, die das System gesammelt und/oder ausgewertet hat. Dementsprechend könnte die letztliche Entscheidung unbemerkt verzerrt sein und somit zu einem diskriminierenden Ergebnis für Einzelpersonen oder Gruppen führen.

36 Wenn sich ein Richter auf Daten und Algorithmen verlässt, die er nicht verifizieren kann, muss er zumindest sicherstellen, dass sein Urteil nicht durch das vorhergesagte Ergebnis beeinflusst wird, bevor er sich seine eigene Meinung bildet. Um dieses Risiko zu mindern, muss ein Richter daher seine Entscheidung treffen, bevor er das Ergebnis der KI konsultiert. Um dies zu gewährleisten, könnten Richter im Laufe des Verfahrens verpflichtet sein, eine formelle Erklärung darüber abzugeben, inwieweit sie das KI-generierte Ergebnis zum Gegenstand ihrer Entscheidung gemacht haben.[10]

37 Die Notwendigkeit, einen unabhängigen menschlichen Gedankengang für eine gerichtliche Entscheidung zu entwickeln, wird den Ehrgeiz, die Kreativität und die Logik des menschlichen Geistes erhalten. Und die Bewahrung der menschlichen Urteilskraft ist über die individuelle Ebene hinaus von entscheidender Bedeutung: Als eine demokratische, auf Rechtsstaatlichkeit basierende Gesellschaft bauen wir auf einen Meinungspluralismus und auf die ständige Änderung und Verfeinerung unserer Anschauungen. In der Welt des Rechts spiegeln sich diese Entwicklung und dieser Fortschritt, der sich im Laufe der Zeit den sich stets verändernden Einstellungen anpasst, in der ständigen Weiterentwicklung der

[10] *Vgl. Fn*. 6.

Rechtsfortbildung wider. Eine solche Rechtsprechung kann im Laufe der Zeit zu neuen Gesetzen führen und dient dazu, das Rechtssystem am Leben zu erhalten. Der Rückgriff auf KI-Systeme, die ihre Analyse notwendigerweise auf die bestehende Rechtsprechung und formelle Analysemuster stützen, könnte indes zu einer strukturellen Beschränkung der Initiative der Gerichte führen, die Rechtsfortbildung aktiv voranzutreiben, indem sie bestimmte, für den jeweiligen Sachverhalt zentrale Fragen stellen und die allgemeine juristische Methodenlehre anwenden. Dies gilt insbesondere für Situationen, die von dem typischen, vom Gesetz vorgesehenen Fall abweichen, und daher einen Widerspruch zwischen dem Gesetzestext und seiner ratio legis hervorrufen können. Solche Rechtslücken werden in der Regel durch die Anwendung anerkannter Auslegungsmethoden geschlossen, wie z.B. durch eine historische Analyse, eine systematische Auslegung oder eine teleologische Reduktion oder Erweiterung. Wenn jedoch die Predictive justice-Instrumente übermäßig genutzt werden, könnte sich die weitere Entwicklung des Gesetzes verlangsamen oder gänzlich zum Erliegen kommen.

(i) Einsatz von Instrumenten der *predictive justice* nach der Urteilsverkündung

38 Ein weiteres praktisches, häufig diskutiertes Beispiel für ein Predictive justice-Instrument, das von der Justiz eingesetzt werden könnte, betrifft das Verhalten eines Straftäters nach seiner Verurteilung. Instrumente zur Rückfallprognose bestimmen das Risiko, dass dieser rückfällig wird und können zur Bestimmung der Haftdauer des Straftäters verwendet werden.

39 Ähnliche Anwendungen wie das bekannteste Beispiel in diesem Bereich namens COMPAS, das in mehreren US-Bundesstaaten eingesetzt wird, gibt es auch in den Strafverfolgungssystemen der EU-Mitgliedstaaten, so z.B. das KI-Tool ProKid, das in den Niederlanden verwendet wird. ProKid zielt darauf ab, die Rückfallgefahr von zwölfjährigen Kindern zu ermitteln, die zuvor von der Polizei einer Straftat verdächtigt wurden. [11] Ein ähnliches Tool ("SAVRY") wird von den spanischen Behörden eingesetzt. [12]

[11] Algorithm Watch, Automating Society – Taking Stock of Automated Decision-Making in the EU, Januar 2019,S. 100, verfügbar unter: <http://www.algorithmwatch.org/automating-society> [abgerufen am 12. Mai 2020].
[12] *Vgl. Fn*. 11, S. 122.ff.

40 In diesem Zusammenhang ergeben sich im Wesentlichen die gleichen Grundrechtsrisiken wie in den oben zitierten Fällen: Erstens könnte das Recht auf ein faires Verfahren verletzt werden, wenn das System mit voreingenommenen und diskriminierenden Daten geschult wird. Dieser Effekt wird in Fällen, in denen die Funktionsweise des algorithmischen Systems nicht öffentlich zugänglich gemacht wird, sogar noch verschärft, da es keine Möglichkeit gibt, eine auf diesen Daten basierende Entscheidung anzufechten. Selbst wenn die hierfür erforderlichen Angaben frei zugänglich wären, müssten die betroffenen Parteien die kostspielige und zeitintensive Last der Datenanalyse selbst tragen. Eine solche Belastung würde ihre Situation im Verfahren erheblich verschlechtern und die Grundsätze eines fairen Verfahrens verletzen.

(ii) Einsatz intelligenter juristischer Recherche-Tools

41 Intelligente Tools zur juristischen Recherche sind ein weiteres praktisches Beispiel für KI, die von Juristen verwendet wird. Das italienische Programm TOGA zum Beispiel wird als intelligente Datenbank für Staatsanwälte (und Rechtsanwälte) verwendet. [13]

42 Der Einsatz dieser Instrumente ist generell zu begrüßen. Nichtsdestotrotz könnte aus einer Grundrechtsperspektive zu bedenken sein, dass die Auswahl des Suchinstruments das Ergebnis der Entscheidung eines Staatsanwalts vordefinieren könnte, da das System einige gesuchte Schlüsselwörter stärker gewichten könnte als andere. Dadurch könnte es den Entscheidungsprozess beeinflussen und so unbemerkt eine Vorentscheidung herbeiführen.

(3) Von Anwälten verwendete Predictive justice-Instrumente

43 Rechtsanwälte und Versicherer verlassen sich zunehmend insbesondere auf solche KI-Anwendungen, die dazu dienen, Gerichtsentscheidungen vorherzusagen. Ein typisches Beispiel hierfür ist Jurimetria, eine statistische und Entscheidungen „antizipierende“ rechtswissenschaftliche Software, die Juristen in Spanien bei der Analyse ihrer Fälle hilft. Sie systematisiert und extrahiert Inhalte aus mehr als 10 Millionen Gerichtsentscheidungen, die aus allen Instanzen und Gerichtsbeschlüssen in Spanien[14] stammen. Ein weiteres prominentes Beispiel ist Casecruncher Alpha, das im Oktober 2017 einen

[13] *Vgl.* TOGA, abrufbar unter: <https://toga.cloud/>.
[14] *Vgl.* Jurimetría,abrufbar unter: <https://jurimetria.laleynext.es/content/Inicio.aspx>.

einwöchigen Wettbewerb gegen menschliche Wirtschaftsanwälte mit einer Genauigkeit von 86,6% der gemachten Vorhersagen gewann.[15]

44 Auf den ersten Blick scheint die Nutzung der geschilderten prognostischen („predictive“) Analyseinstrumente von Anwälten den Zugang zur Justiz nicht zu behindern. Hierbei darf jedoch nicht außer Betracht bleiben, dass die Arbeit von Anwälten weit über die Übermittlung binärer juristischen Antworten auf eine einfache Frage hinausgeht.

45 Die Arbeit eines Rechtsanwalts ist viel facettenreicher als die bloße Bereitstellung rechtlicher Analysen. KI-Technologie kann eine juristische Analyse unter Umständen schneller und genauer als ein Anwalt durchführen, da sie auf einen umfangreichen Datenbestand zurückgreifen und diesen innerhalb kürzester Zeit auswerten kann. Ein Mandant, der nur binäre Antworten erhält, wird jedoch in vielen Situationen nicht unbedingt die gewünschte Beratung erhalten. Zu Beginn eines Austausches zwischen einem Anwalt und einem Mandanten ist oft selbst Letzterem nicht bewusst, worin das eigentliche Problem liegt. Und oft treten erst im Laufe der Zeit die für die Beilegung eines Rechtsstreits tatsächlich relevanten Fakten zutage. Neben rechtlichen Fragen können insbesondere auch persönliche oder wirtschaftliche Interessen eine große Rolle spielen. Deshalb sollte jeder die Möglichkeit haben, sich umfassend von kompetenten, erfahrenen Juristen beraten zu lassen, die es verstehen, die richtigen Fragen zu stellen und die Fakten zu erforschen, um mit dem Mandanten die beste Lösung zu entwickeln.

46 Während Rechtsanwälte KI-Ressourcen unterstützend einsetzen können, müssen sie auch sicherstellen, dass sie ihre Fähigkeit bewahren, die relevanten Fragen zu stellen, den Sachverhalt mit den Mandanten zu erforschen und gemeinsam geeignete Lösungen zu erarbeiten. Andernfalls könnte eine negative Vorhersage Betroffene vorzeitig davon abhalten, einen Fall allein auf der Grundlage der Meinung einer Maschine gerichtlich zu verfolgen. Dies könnte nicht nur in Fällen problematisch sein, in denen die Maschine ein schlichtweg falsches Ergebnis produziert, sondern mehr noch in Fällen, in denen sich nur geringfügige, von der KI unbemerkte, aber möglicherweise entscheidende

[15] Hill, Caroline, ‘Machine beats man’ in Casecrunch lawyer challenge, Legal IT Insider, 30. Oktober 2017, abrufbar unter: <https://legaltechnology.com/machine-beats-man-in-casecrunch-lawyer-challenge/> [abgerufen am 12. Mai 2020].

Merkmale eines Falles von scheinbar ähnlichen Fällen unterscheiden. Im Extremfall könnte sich ein systematischer Einsatz solcher Maschinen auch nachteilig auf die weitere Entwicklung der Rechtsfortbildung seitens der Gerichte auswirken. Eine solche Entwicklung der Rechtsprechung war jedoch immer Teil des gesellschaftlichen Wandels und der sich daran anpassenden Rechtssysteme.

c) In der Justizverwaltung verwendete KI

47 Auch der Einsatz von KI in der Gerichtsverwaltung könnte bei gezielter Anwendung Grundrechtsrelevanz haben. Solche Bedenken wurden real, als das Justizministerium in Polen ein Algorithmus-basiertes System zufälliger Geschäftsverteilung einführte. Das digitale System ordnet Fälle landesweit einmal pro Tag bestimmten Richtern zu. Unproblematisch wäre ein solches System wohl, wenn es tatsächlich nach einem reinen Zufallsprinzip funktionierte und keinen Ermessensspielraum zuließe. Im konkreten Fall kam jedoch die Vermutung auf, dass der polnische Generalstaatsanwalt den Prozess unangemessen beeinflussen konnte, da er neben seiner Rolle als potentieller Beteiligter in einem Strafverfahren zugleich Teil des Justizministeriums ist und hierdurch möglicherweise kontrollieren konnte, wie die Verfahren zugewiesen wurden. Wenn eine solche Einflussnahme stattgefunden hätte, wäre eine Verletzung des Rechts auf Gewährleistung eines fairen Verfahrens naheliegend.[16] Die Bedenken in diesem Beispiel wurden durch die Tatsache verschärft, dass das Ministerium nicht bereit war, die Funktionsweise des für das System verwendeten Algorithmus offenzulegen.[17]

2. In der Strafverfolgung verwendete KI

48 Elementare Grundsätze der Rechtsstaatlichkeit könnten ebenfalls durch den Einsatz von KI-Tools in der Strafverfolgung gefährdet werden. Werkzeuge, die dort bereits eingesetzt werden, könnten entweder direkt im Gerichtssaal angewandt werden oder indirekt als Grundlage für eine in einem Gerichtsverfahren angefochtene Entscheidung eine Rolle spielen. Die besondere Problematik ergibt sich in diesem Zusammenhang aus der Tatsache, dass die

[16] Matczak, Marcin, 10 Facts on Poland for the Consideration of the European Court of Justice, 13. Mai 2018, verfügbar unter: <https://verfassungsblog.de/10-facts-on-poland-for-the-consideration-of-the-european-court-of-justice/>, unter Bezugnahme auf den Fall Daktaras gg. Litauen – Beschwerdenr. 42095/98 [abgerufen am 12. Mai 2020].
[17] *Vgl.* Fn. 11**Fehler! Textmarke nicht definiert.**, S. 107-108.ff.

Betroffenen in der Regel nicht wissen, dass entsprechende Werkzeuge zu ihrem Nachteil eingesetzt werden. Darüber hinaus wird die Polizei typischerweise nicht veröffentlichen, welche Kriterien für das Ergebnis des Systems herangezogen, wie es diese gewichtet und welche Trainingsdaten den Algorithmen des Systems zu Grunde gelegt werden. Solche Systeme verhindern den Zugang zum Recht, da die Betroffenen in den meisten Fällen weder erkennen noch beweisen können, ob sie Gegenstand einer fehlerhaften oder ungerechten Entscheidung sind. Risiken entstehen auch dadurch, dass die Systeme beträchtliche Datenmengen sammeln, die gehackt werden können und zu schwerwiegenden Verletzungen des Datenschutzes und der Privatsphäre führen.

49 Ein besonders kritisches Beispiel ist das von der EU finanzierte iBorderCtrl-Projekt (Intelligent Border Control System), dessen Software darauf abzielt, Personen aufzuspüren, die bei Grenzkontrollen lügen[18]: Drittstaatsangehörige werden gebeten, Fragen eines computeranimierten Grenzschutz-Avatars zu beantworten, der die Mikrogestik Einreisender analysiert, um herauszufinden, ob der Befragte lügt.[19] Laut einer Analyse von Algorithm Watch birgt das System ein erhebliches Risiko rassistischer Voreingenommenheit, da es hauptsächlich auf Grundlage von Daten weißer europäischer Männer trainiert wurde und eine hohe Fehlerquote von 25%aufwies.[20] Es wirft ferner Bedenken hinsichtlich der Gewährleistung eines fairen Verfahrens auf, da berechtigte Zweifel an der wissenschaftlichen Genauigkeit und Zuverlässigkeit eines Lügendetektors bestehen. Solche Geräte könnten technisch und hypothetisch auch im Rahmen von Gerichtsverhandlungen eingesetzt werden, was hinsichtlich der Gewährleistung eines fairen Verfahrens noch problematischer wäre.

50 Weitere kritische Beispiele für den Einsatz von KI in der Strafverfolgung sind die intelligente Videoüberwachung oder Instrumente auf Grundlage von Gesichtserkennungstechnologie, die in der gesamten EU zunehmend eingesetzt werden. In der deutschen Stadt Mannheim wurde beispielsweise ein Experiment durchgeführt, das die KI-gestützte Erkennung sozialer Alltagssituationen auf Grundlage automatischer Bildverarbeitung ermöglichte. Das Kamerasystem

[18] *Vgl.* iBorderCtrl, available at: <https://www.iborderctrl.eu/>.
[19] *Vgl.* iBorderCtrl, Intelligent Portable Control System, Projektpräsentation abrufbar unter: <https://www.iborderctrl.eu/sites/default/files/publications/iBorderCtrl%20global%20presentation%20v5.pdf> [abgerufen am 12. Mai 2020].
[20] Vgl. Fn. 11,S. 36-37.ff.

informiert die Polizei, wenn es Handlungen erkennt, die als Tätlichkeit oder Diebstahl gewertet werden könnten. So ist es möglich, die beteiligten Personen über das gesamte Kamerasystem zu verfolgen. [21] Solche Verhaltensscanner üben einen starken Konformitätsdruck aus und sind anfällig für Fehlalarme. Im Hinblick auf den Zugang zum Recht ergibt sich das Hauptrisiko aus der Tatsache, dass das System nicht offenbart, auf welche "unnatürlichen Bewegungen" die Algorithmen trainiert sind zu reagieren.

3. In der öffentlichen Verwaltung verwendete KI

51 Ein weiterer Bereich, in dem Rechtsanwälten eine besondere Aufgabe zukommt, die Einhaltung der Rechtsstaatlichkeit zu gewährleisten, ist der Einsatz von KI-Technologie durch die öffentliche Verwaltung. Die Intransparenz solcher Systeme - im Folgenden anhand praktischer Beispiele veranschaulicht - zeigt, dass diskriminierende oder anderweitig verzerrte Ergebnisse rechtsstaatliche Verfahren gefährden, da sie schwer zu erkennen und vor einem Richter anzufechten sind.

52 Ein besonders kritisches Beispiel sind das Profiling und Punktevergabesysteme. In der dänischen Stadt Gladsaxe wurde zum Beispiel im Januar 2018 im Rahmen des sogenannten Ghetto-Plans des Landes ein Ermittlungsinstrument eingeführt, um Kinder unter sie gefährdenden Umständen frühzeitig zu erkennen. Die Stadtverwaltungen waren angehalten, Informationen über Kinder aus verschiedenen öffentlichen Quellen zu sammeln, auszuwerten und nach bestimmten "Risikoindikatoren" zu kategorisieren. Das System wies der Familie dann eine Punktzahl zu, die auf Informationen wie der Wahrnehmung von Arztterminen, dem Beschäftigungs- und Familienstatus, der psychischen Gesundheit und ähnlichen Kriterien basierte.

53 Im Dezember 2018 kam es in der Gemeinde Gladsaxe zu einem Datenleak, wodurch Daten von mehr als 20.000 Bürgerinnen und Bürgern offengelegt wurden, darunter Angaben über das Geschlecht, Alter, Sozialleistungen und Besonderheiten der Familie.[22] Dieser Fall veranschaulicht die typischen

[21] Mannheim testet verhaltensbasierte Videoüberwachung, Heise Online, 3. Dezember 2018, abrufbar unter: <https://www.heise.de/newsticker/meldung/Mannheim-testet-verhaltensbasierte-Videoueberwachung-4239279.html> [abgerufen am 12. Mai 2020].
[22] *Vgl.* Fn.11, S. 36–37*;* s. ebenfalls Enforcement Tracker, abrufbar unter: <https://www.enforcementtracker.com>.

Auswirkungen, die mit dem Profiling einhergehen: Solche Programme bergen nicht nur erhebliche Risiken für die Privatsphäre und den Datenschutz, sie können auch auf diskriminierende Weise verwendet werden. Die meisten Menschen waren sich nicht einmal bewusst, dass ihre Daten durch das Programm verarbeitet wurden, und konnten daher auch nicht gegen das Programm vorgehen.

54 In anderen Verwaltungssystemen wurden Verwaltungsaufgaben automatisiert. Solche Anwendungen können grundsätzlich als positive Anwendungsfälle von KI gelten. Bei falscher Programmierung oder unzureichendem Training mit Daten können solche Anwendungen mitunter jedoch bestimmte soziale Gruppen in größerem Maße betreffen als andere und hierdurch soziale Ungleichheiten verstärken. So traf etwa ein in vielerlei Hinsicht fehlerhaftes Steuererhebungssystem in Australien im Jahr 2018 Menschen aus schwächeren sozialen Schichten finanziell wesentlich stärker als andere.[23] Daher muss auch das Risiko zunehmender sozialer Ungleichheiten im Zusammenhang mit dem Einsatz von KI berücksichtigt werden.

II. Frage 2: Welche Situationen sehen Sie aus grundrechtlicher Sicht als Hochrisikosituationen an? Wie würden Sie diesbezüglich Hochrisikosituationen definieren?

55 Der derzeitige Regulierungsrahmen des KI-Weißbuchs sieht zwingende Anforderungen nur für den Fall von Hochrisikoanwendungen vor. Der DAV schlägt einen nuancierteren Ansatz vor. Aus Sicht des DAV sollte ein Konzept entwickelt werden, das mindestens fünf Risikostufen umfasst. Alle anderen KI-Anwendungen müssen jedoch, nach ihrer Eingriffsintensität differenziert, gewisse Transparenz-, Sicherheits- und Kontrollanforderungen erfüllen. Wesentlich für die Einstufung sollten hierbei der Einsatzbereich und das individuelle Gefährdungspotenzial der jeweiligen KI-Anwendung sein. Im Folgenden werden zwei konkrete Beispiele für Hochrisikoanwendungen vorgestellt. Anschließend wird anhand konkreter Kriterien ein Konzept zur Risikobewertung entwickelt.

[23] Djeffal, Christian, Artificial Intelligence and Public Governance Normative Guidelines for Artificial Intelligence in Government and Public Administration, in: Wischmeyer, Thomas/Rademacher, Timo (edt) Regulating Artificial Intelligence, Cham: Springer 2019, S. 281.

Schließlich wird auf der Grundlage der vorangegangenen Erkenntnisse eine Risikomatrix entworfen.

1. Situationen mit hohem Risiko

56 Es gibt eine Vielzahl von Situationen, die aus grundrechtlicher Sicht als Hochrisikosituationen bezeichnet werden können. Im Folgenden werden zwei beispielhafte Situationen im Zusammenhang mit der Justiz und der Strafverfolgung skizziert.

a) Situation 1: Predictive justice-Instrumente

57 Wie bereits unter Frage (1) ausgeführt, sind KI-Systeme, die innerhalb des Justizsektors, der Strafverfolgung und der Verwaltung eingesetzt werden, besonders kritisch. Das höchste Risiko besteht dort, wo eine rechtsverbindliche Entscheidung ausschließlich auf einer autonomen Entscheidung beruht und somit das menschliche Judiz ersetzt. Es handelt sich um eine Situation mit hohem Risiko, da die Entscheidung typischerweise intransparent ist, voreingenommen sein kann und das Potential besitzt, einen Schaden hohen Ausmaßes zu verursachen , da sie - letztendlich - zu schweren finanziellen Schäden oder gar zur Inhaftierung unschuldiger Personen führen kann.

b) Situation 2: Biometrische Identifizierungssysteme

58 Eine weitere hochrisikoreiche Situation innerhalb der Justiz, Strafverfolgung und Verwaltung besteht darin, dass rechtsverbindliche Entscheidungen in erster Linie auf biometrischen Identifizierungssystemen beruhen. Der Einsatz solcher Instrumente ist mit einem hohen Risiko verbunden, da die Wahrscheinlichkeit schwerwiegender Verletzungen von Grundrechten wie der Privatsphäre und des Grundsatzes der Nichtdiskriminierung sehr hoch ist und weitreichende Folgen für die betroffene Person haben kann. Die Systeme neigen zudem häufig zu hohen Fehlerquoten und sind anfällig für Manipulationen.

2. Definition von hohem Risiko - Stufenansatz erforderlich

59 Das Weißbuch verfolgt einen zweifachen Risikobewertungsansatz und teilt KI-Anwendungen in Hochrisiko- und Niedrigrisikokategorien ein. Anwendungen mit einem "hohem Risiko" werden als solche definiert, die sowohl in einem Sektor als auch bei der beabsichtigten Verwendung mit erheblichen Risiken verbunden sind.

60 Die DAV ist der Ansicht, dass die Unterteilung von KI-Systemen in lediglich zwei Kategorien, nämlich Anwendungen mit hohem und niedrigem Risiko anhand einer erschöpfenden Liste, nicht die Komplexität und Vielfalt realer Anwendungsfälle widerspiegelt. Es ist zudem wahrscheinlich, dass sie zu einer künstlichen Aufspaltung der Anwendungen führt. Dies könnte entweder empfindliche regulatorische Lücken verursachen oder andererseits zu einer Überregulierung führen, wenn aus Gründen der gebotenen Vorsicht zu viele Anwendungen in die Hochrisikokategorie aufgenommen würden.

61 Die DAV schlägt daher vor, eine Risikomatrix zu entwickeln. Eine solche Risikomatrix könnte einen Stufenansatz verfolgen und - dem Vorschlag der Deutschen Datenethikkommission folgend – einzelne Anwendungen in mindestens fünf verschiedene Risikostufen einteilen.[24] Jede Risikokategorie würde dann zu unterschiedlichen rechtlichen Anforderungen von unterschiedlicher Regelungstiefe führen.[25]

62 Bei der Bewertung potenzieller Risiken sollten sowohl der betroffene Sektor als auch einzelne Aspekte der konkreten KI-Maßnahme berücksichtigt werden. Im Folgenden werden die Sektoren identifiziert, die für das systemische Funktionieren eines Staates von größter Bedeutung sind (a). Im darauffolgenden Abschnitt werden Kriterien für eine konkrete Risikobeurteilung einer einzelnen Anwendung entwickelt (b).

a) Hochrisikosektoren

63 Angesichts der Bedeutung der unter Frage (1) identifizierten Risiken sollte das Justizsystem als systemrelevant und besonders kritisch für die Erhaltung von Demokratien gekennzeichnet werden, die auf der Idee des Machtgleichgewichts der Institutionen aufbauen. Wenn Entscheidungen in diesem Bereich voll

[24] Datenethikkommission der Bundesregierung, Gutachten der Datenethikkommission, 22. Januar 2020, abrufbar unter: <https://www.bmjv.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_EN_lang.html>, S. 177 [abgerufen am 12. Mai 2020].

[25] Der Risikobewertungsansatz und die Risikomatrix sind angelehnt an: Gutachten der Datenethikkommission vom 22. Januar 2020, s. Fn. 24; Martini, Mario, Grundlinien eines Kontrollsystems für algorithmenbasierte Entscheidungsprozesse, abrufbar unter: <https://www.uni-speyer.de/fileadmin/Lehrstuehle/Martini/2019_Gutachten_GrundlageneinesKontrollsystemendgueltig.pdf>; Zweig, Katharina, Algorithmische Entscheidungen: Transparenz und Kontrolle, Januar 2019, verfügbar unter: <https://www.kas.de/documents/252038/4521287/AA338+Algorithmische+Entscheidungen.pdf/533ef913-e567-987d-54c3-1906395cdb81?version=1.0&t=1548228380797 >[abgerufen am 12. Mai 2020].

automatisiert, aber nicht transparent sind, besteht das Risiko, dass eine Regierung, die die Kontrolle über die Technologie hat, entscheidende Demokratieelemente missachtet. Dementsprechend könnten KI-Instrumente von autoritären Regimen genutzt werden, um Kontrolle über Gesellschaften auszuüben und ihre repressiven Fähigkeiten zu verstärken.

64 Andere Sektoren, die für das systemische Funktionieren eines Staates relevant sind, wie der Gesundheits- und Energiesektor, müssen besonders geschützt werden. Dies wurde sicherlich während des Ausbruchs der Covid-19-Pandemie deutlich. Die genannten Sektoren müssen als besonders risikobehaftet eingestuft werden, um sicherzustellen, dass die Grundbedürfnisse der Bevölkerung erfüllt werden können - mit oder ohne den Einsatz von KI.

b) Hochrisikoanwendungen

65 Dem Weißbuch fehlt es derzeit an klaren Leitlinien, was eine konkrete kritische Anwendung darstellt. Zwei Hauptfaktoren sollten berücksichtigt werden, nämlich die Eintrittswahrscheinlichkeit eines identifizierten Risikos (i) und die Schwere eines potentiellen Schadens (ii).

(1) Eintrittswahrscheinlichkeit eines identifizierten Risikos

66 In diesem Abschnitt ist die klassische Risikobewertung hinsichtlich des Kausalzusammenhangs zwischen einer Handlung und einer wahrscheinlichen Folge zu bewerten. Dabei spielt auch eine entscheidende Rolle, ob der Endnutzer die Möglichkeit hat, die automatisierte Entscheidung neu zu treffen und seine Folgen damit abzumildern.

(2) Schwere eines potenziellen Schadens

67 Bei der Beurteilung der Schwere eines potenziellen Schadens müssen sowohl seine Art (a) als auch sein wahrscheinliches Ausmaß (b) berücksichtigt werden.

68 Was die Art des Schadens (a) betrifft, so müssen Kriterien wie die Frage, ob der beabsichtigte Schaden reversibel oder irreversibel sind, in die Bewertung einbezogen werden. Ein weiterer entscheidender Aspekt ist die Art des betroffenen Grundrechts: In diesem Zusammenhang ist die potenzielle körperliche Schädigung und damit eine Verletzung des Rechts auf körperliche Unversehrtheit (wie in Art. 3 GRCh verankert) auf höchster Ebene anzusiedeln.

69 Im Hinblick auf das Ausmaß des potenziellen Schadens (b) sind Kriterien wie die Anzahl möglicher betroffener Personen, die Auswirkungen auf andere Grundrechte, die Umstände, die Häufigkeit und die Dauer der Beeinträchtigung zu berücksichtigen.

(3) Risiko-Matrix

70 Die Einstufung einer bestimmten Anwendung in eine der Risikokategorien sollte zu unterschiedlichen rechtlichen Anforderungen führen. Diese könnten auf der folgenden Risikomatrix basieren:

(i) Risikostufe (1):

71 Auf der untersten Interventionsebene könnten lediglich Ex-post-Transparenzpflichten vorgesehen werden. Dies würde bedeuten, dass keine permanenten Kontrollmechanismen installiert werden müssten. Auf Anfrage müsste eine Ad-hoc-Analyse durchgeführt und die Risikobewertung wiederholt werden. KI-Systeme, die z.B. dazu dienen, Verbrauchern Produkte vorzuschlagen oder sie in einer bestimmten Reihenfolge in sozialen Netzwerken darzustellen, könnten in diese Kategorie fallen.

(ii) Risikostufe (2):

72 Auf einer zweiten Risikostufe sollte eine allgemeine Kontrolle des Systems stattfinden. Um die Überwachung zu erleichtern, sollten Betreiber des Systems verpflichtet werden, die Qualitätsmaßnahmen und den Lernprozess des Systems anzugeben und offenzulegen, wie sich das System zur letztendlichen Entscheidung verhält, d.h. inwieweit die Entscheidung beeinflusst wird oder auf den Ergebnissen des Systems beruht. Die Kontrolle könnte von externen Auditoren durchgeführt werden. Dynamische oder personalisierte Preisgestaltungstechniken könnten unter diese Kategorie fallen.

(iii) Risikostufe (3):

73 Auf einer dritten Ebene könnten zusätzlich zur Kontrolle umfassende Transparenzpflichten gesetzlich vorgeschrieben werden. Diese sollten die Offenlegung aller relevanten Faktoren und Kriterien, die als Grundlage für die automatisierte Entscheidung dienen, sowie aller Ausbildungsdaten umfassen, jedoch ohne zwingend die Offenlegung weiterer technischer Funktionen des Systems zu erfordern.

74 Die Überwachung durch Dritte kann mit staatlichen Informations- und Inspektionsrechten einhergehen. Es können auch Aufzeichnungsanforderungen gelten. Als Regulierungsmodell könnten die Vorschriften des deutschen Wertpapierhandelsgesetzes (WpHG) über den algorithmischen Handel mit Finanzinstrumenten dienen (§ 6 Abs. 4 WpHG).

75 Darüber hinaus sollte das Unternehmen, das das System in Anspruch nimmt, verpflichtet werden, eine verantwortliche Person innerhalb des Unternehmens für das diesbezügliche Risikomanagement zu benennen. Diese Person könnte gegenüber Dritten für Systemfehler verantwortlich gemacht werden.

76 Was die Nutzung von KI durch die Regierung anbelangt, so könnten rein informative Instrumente, die innerhalb der Verwaltung eingesetzt werden, wie z.B. der Chatbot "Bobbi", der von der Berliner Verwaltung verwendet wird[26], in diese Kategorie fallen.

(iv) Risikostufe (4):

77 Auf der nächsten Stufe wäre eine vollständige Erklärung des Systems erforderlich. Wenn in dieser Stufe etwas schief geht, wären die möglichen Schäden so hoch, dass KI-Systeme mit einer Lernkomponente nur erklärbare Methoden des maschinellen Lernens verwenden dürfen. Eine solche Erklärung könnte z.B. durch Erklärungs- oder Entschlüsselungsalgorithmen erfolgen. Wenn Geschäfts- oder Betriebsgeheimnisse die vollständige Offenlegung der Logik eines Algorithmus behindern würden, müsste die Geschäftseinheit sie zumindest gegenüber einer autorisierten staatlichen Stelle oder Behörde offenlegen. Darüber hinaus könnte ein präventives Zulassungsverfahren vorgeschrieben werden, um die Einhaltung des einschlägigen Rechts vor seiner Verwendung zu gewährleisten.

78 Darüber hinaus müssten laufende dynamische Betreiberpflichten erlassen werden. Dadurch wäre der Betreiber auch nach der Marktzulassung für die Ergebnisse der Entscheidungen und die verfahrenstechnische Korrektheit des Systems verantwortlich.

79 Solche Anforderungen könnten für KI gelten, die in Bereichen wie Justiz und Strafverfolgung oder für die Prüfung und Zuweisung von Leistungen verwendet

[26] *Vgl.* Chatbot Bobbi, verfügbar unter: <https://service.berlin.de/chatbot/chatbot-bobbi-606279.php>.

wird. Eine Vorabkontrolle kann auch bei Anwendungen angebracht sein, die erhebliche Auswirkungen auf diejenigen Lebensbereiche haben können, die für das Funktionieren des Staates und die freiheitliche demokratische Grundordnung von systemischer Bedeutung sind (insbesondere bei Wahlen oder im Bereich der öffentlichen Meinungsbildung).

(v) Risikostufe (5):

80 Die höchste Interventionsebene sollte nur für außergewöhnliche Anwendungen gelten, wie z.B. automatisierte tödliche Waffen. Zusätzlich zu den zuvor aufgeführten Verpflichtungen wäre zu prüfen, ob der KI-Einsatz in diesem Bereich beispielsweise auf nicht lernende KI-Systeme, d.h. auf Systeme, die auf linearen Entscheidungsbäumen beruhen, beschränkt werden sollte oder ob andere zuverlässige, sichere Beschränkungen gefunden werden können, um eine vollständige Überwachung und menschliche Kontrolle der Anwendung zu gewährleisten.

III. Frage 3: Kennen Sie wirksame Mittel zur Bewältigung der Risiken, die Sie in Ihrer Antwort auf die obigen Fragen genannt haben?

81 Der DAV begrüßt die Arbeit an den Grundprinzipien, welche die Expertenkommission für Künstliche Intelligenz(High Level Expert Group, HLEG) in ihren Richtlinien für vertrauenswürdige KI identifiziert hat. Die sieben Hauptanforderungen an "vertrauenswürdige KI" - nämlich 1) menschliche Handlungsfähigkeit und Aufsicht, 2) technische Robustheit und Sicherheit, 3) Datenschutz und Datenverwaltung, 4) Transparenz, 5) Vielfalt, Nichtdiskriminierung und Fairness, 6) gesellschaftliches und ökologisches Wohlergehen und 7) Rechenschaftspflicht - sollten in den Mittelpunkt künftiger KI-Regulierungen gestellt werden.

82 Auf der Grundlage dieser Prinzipien sollten die folgenden Mittel in Betracht gezogen werden, um den im vorigen Abschnitt identifizierten Risiken wirksam zu begegnen:

83 Zunächst sollten rechtsverbindliche (d.h. nicht anfechtbare) Entscheidungen eines Gerichts und ähnlich eingriffsintensive Entscheidungen staatlicher Akteure

niemals ausschließlich auf einer autonomen Entscheidung beruhen, sondern stets von einem Menschen getroffen werden (1).

84 Zweitens, sollte dort, wo dies nicht möglich oder nicht unbedingt notwendig ist, ein hohes Maß an Transparenz, das die Identifizierbarkeit und Anfechtbarkeit einer KI-basierten Entscheidung gewährleistet, zwingend vorgeschrieben werden (2).

85 Um drittens den Zugang zum Recht zu erleichtern und sicherzustellen, dass von einer fehlerhaften oder schädlichen KI-Anwendung Betroffene, angemessen entschädigt werden, sind effiziente und umfassende Rechtsbehelfe und ein Haftungssystem erforderlich (3).

86 Konkret könnten die folgenden Maßnahmen beschlossen werden:

1. Der Vorrang der menschlichen Entscheidung

87 Rechtlich bindende Gerichts- oder ähnlich eingreifende Behördenentscheidungen sollten nicht ausschließlich auf einem automatisierten Entscheidungssystem beruhen. Wie im vorigen Abschnitt dargelegt, ist das Recht, von einem unabhängigen und unparteiischen Gericht gehört zu werden, nur dann erfüllt, wenn der entscheidende Richter ein Mensch und keine Software ist. Ein zukünftiger EU-Rahmen zur künstlichen Intelligenz sollte dieses Prinzip bekräftigen, um Rechtsklarheit zu gewährleisten.

88 Dieser Grundsatz spiegelt sich auch in Art. 22 der Datenschutz-Grundverordnung ("DSGVO") wider, der festlegt, dass jedermann das Recht hat, nicht einer ausschließlich auf einer automatisierten Verarbeitung – einschließlich Profiling – beruhenden Entscheidung unterworfen zu sein, die ihr gegenüber rechtliche Wirkung entfaltet oder sie in ähnlicher Weise erheblich beeinträchtigt. Dieser Grundsatz gilt umso mehr für Situationen, die zu einer rechtsverbindlichen Gerichtsentscheidung führen können.

89 Die Forderung nach "menschlicher" Entscheidungsfindung als Grundvoraussetzung für ein funktionierendes Justizsystem kann nur dann erfüllt werden, wenn weiterhin ausreichend und angemessen ausgebildetes menschliches Fachwissen vorhanden ist. Mit anderen Worten: Die fortschreitende Digitalisierung und Automatisierung darf nicht zu einem Personalabbau führen. Vielmehr werden Angehörige der Rechtsberufe

zusätzliche Fortbildungen und Schulungen zum Erlernen technischer und digitaler Fähigkeiten benötigen. Zu diesem Zweck würden sich ein EU-weiter Wissensaustausch und die Förderung von Schulungen als hilfreich erweisen. Darüber hinaus können spezielle Einheiten für KI-bezogene Inhalte gebildet werden.

90 Ferner sind angemessene Sicherheitsmaßnahmen und "analoge" Auffangpläne erforderlich, um das Funktionieren der Einheit im Falle eines Systemausfalls zu gewährleisten.

2. Transparenz

91 Ein weiteres entscheidendes Element, das in den ethischen Richtlinien der HLEG hervorgehoben wurde, ist die Schaffung transparenter Systeme.

92 Damit das Postulat der Transparenz nicht zur leeren Worthülse wird, müssen zunächst klare Anforderungen und Kriterien formuliert werden. Anschließend werden die praktischen Mittel zu ihrer Umsetzung skizziert.

a) Der erforderliche Grad an Transparenz

93 Moderne maschinelle Lernsysteme erzeugen komplexe Modelle, welche die Nachvollziehbarkeit, warum und wie sie einen bestimmten Output erzeugen, stark erschweren. Und selbst Systeme, die Algorithmen verwenden, deren zugrundeliegende Arbeitsweise und Logik erklärt werden können (z.B. weil sie auf Entscheidungsbäumen basieren), geben ihren Entscheidungsprozess möglicherweise nicht gegenüber ihren Nutzern an. Und selbst wenn Informationen über ein System verfügbar sind, werden Endnutzer aufgrund der Informationsmenge und der Komplexität dieser Systeme oft nicht in der Lage sein, diese zu verstehen oder einzuordnen. Der kombinierte Effekt dieser Faktoren wird häufig als "Black-Box"-Phänomen bezeichnet.[27]

94 Die Opazität selbstlernender algorithmischer Systeme trägt nicht nur dazu bei, potenziell voreingenommene oder fehlerhafte Entscheidungen zu verbergen. Sie kann auch den Zugang zum Recht behindern, wenn aus einer individuellen Entscheidung keine verwertbaren Beweise gewonnen werden können. Gleichzeitig könnte sich die Forderung nach vollständiger Offenlegung aller

[27] Yeung, Karen, Responsibility and AI, Studie des Europarats, September 2019, abrufbar unter: <https://rm.coe.int/responsability-and-ai-en/168097d9c5>, S. 21, abgerufen am 12. Mai 2020].

technischen Funktionen eines Algorithmus auch als kontraproduktiv erweisen. Ähnliche Regelungsansätze wie z.B. die Pflicht zur Einwilligung in die Verarbeitung personenbezogener Daten nach der DSGVO haben die Gefahr einer Informationsüberlastung aufgezeigt: Wenn Betroffene zu viele Informationen über eine Transaktion erhalten, die sie nicht nachvollziehen können, besteht die Gefahr einer noch stärkeren Intransparenz.

95 Im Lichte der Gewährleistung eines rechtsstaatlichen Verfahrens besteht der Hauptzweck von Transparenz darin, dem Einzelnen zu ermöglichen, eine bestimmte Entscheidung zu verstehen, zu bewerten und anzufechten. Im Zusammenhang mit dem Einsatz von KI in staatlichem Handeln ist Transparenz auch ein Mittel zur Erfüllung der gesetzlichen Begründungspflicht für Verwaltungsakte, wie sie in Artikel 41 Absatz 2 GRCh festgelegt ist.

96 In Anbetracht dieser Feststellungen sollte das Prinzip der Transparenz drei Schlüsselkriterien erfüllen:

97 Transparenzanforderungen sollten folgende drei Elemente gewährleisten

1) Identifizierbarkeit,

2) Anfechtbarkeit und

3) angemessene Aufsicht.

(1) Identifizierbarkeit

98 Jede KI-Anwendung sollte für Endnutzer leicht und klar als solche identifizierbar sein. Der Einzelne sollte ein Recht darauf haben zu erfahren, ob er einem autonomen Entscheidungsmechanismus unterliegt oder nicht. Betreiber sollten ferner den beabsichtigten Zweck der Nutzung eines KI-Systems offen legen.

99 Betroffene sollten auch in klarer und verständlicher Sprache darüber informiert werden, ob die von den Instrumenten der künstlichen Intelligenz angebotenen Lösungen bindend sind oder nicht, und ob sie alternative Optionen haben. Darüber hinaus sollten sie darüber informiert werden, wenn sie das Recht auf die Einholung von Rechtsrat und das Recht auf Zugang zu einem Gericht haben. Sie müssen auch klar über jede vorherige Bearbeitung eines Falles durch künstliche Intelligenz vor oder während eines Gerichtsverfahrens informiert werden und das Recht haben, hiergegen Einspruch einzulegen.

(2) Anfechtbarkeit

100 Um das Recht auf Anfechtung einer Entscheidung zu gewährleisten, sollte Auskunft über die grundlegenden Merkmale der betreffenden Einzelentscheidung gegeben werden, nämlich

- die verwendeten Kriterien,

- ihre Gewichtungen und die

- Trainingsdaten des selbstlernenden Algorithmus.

(3) Angemessene Aufsicht

101 Eine angemessene Aufsicht bedeutet, dass insbesondere öffentliche Einrichtungen, die KI-Technologie nutzen, unabhängigen Dritten die Möglichkeit geben sollten, unabhängige Tests der Technologie durchzuführen.

102 Wenn ein Gericht beispielsweise ein bestimmtes Predictive justice-Instrument anwenden wollte, müsste es seinen Technologieanbieter bitten, die technischen Möglichkeiten zur Verfügung zu stellen, um angemessene, unabhängige und vernünftige Tests bezüglich der Anwendungsgenauigkeit zu ermöglichen und so auch das Risiko möglicher Diskriminierungen verschiedener Bevölkerungsgruppen aufzudecken.

103 Darüber hinaus sollte der Technologieanbieter alle Beschwerden oder Berichte über Ungleichbehandlungen seines Diensts offen legen. Er sollte ferner verpflichtet werden, Informationen über die Fehlerquote des Systems zu veröffentlichen.

104 Außerdem sollten permanente Kontroll- und Überwachungspflichten eingeführt werden. Der Betreiber sollte auch sicherstellen, dass sein Technologieanbieter dynamischen Betreiberpflichten unterliegt, die ihn für die Ergebnisse der Entscheidungen und die Verfahrenskorrektheit des Systems verantwortlich machen.

b) Mittel zur Umsetzung von Transparenzaspekten: Regulierung „durch und im Design“ (sog „Regulation by and in Design“)

105 Zur Umsetzung der oben genannten identifizierten Transparenzanforderungen könnte der Ansatz der "Regulation by and in Design" verfolgt werden. Der

Grundgedanke hinter "Regulierung durch Design" ist, dass relevante Normen in der Technologie selbst verankert werden.[28] Das Konzept lehnt sich an Artikel 25(1) DSGVO an, der von den für die Verarbeitung Verantwortlichen verlangt, nur die Daten zu verarbeiten, die für jeden spezifischen Zweck absolut erforderlich sind. Diese Idee wurde weiterentwickelt und wird nun unter dem Begriff "Ethik durch oder im Design" fortgeführt, wonach auch weitere relevante Anforderungen in das System selbst eingebettet werden sollen.[29] Wenn die oben genannten Kriterien standardmäßig in die Architektur eines KI-Systems integriert werden, würde die Transparenz des Systems erhöht, da diese im System klar identifizierbar und nachvollziehbar wären.

c) Mittel zur Gewährleistung der Transparenz bezüglich der zugrunde liegenden Technologie

106 Die Erklärung der technischen Funktionsweise und Logik von Anwendungen, die auf neuronalen Netzen basieren, ist mit hohen Hürden verbunden. Die aktuelle Forschung konzentriert sich insbesondere auf Werkzeuge, die darauf abzielen, neuronale Systeme mit Hilfe von Entschlüsselungsformeln erklärbar zu machen (1). Eine weitere Möglichkeit ist der Einsatz der Blockchain-Technologie (2).

(1) Erklärende KI-Modelle

107 Explainable AI (XAI) ist ein Konzept, das auf der Idee beruht, dass Algorithmen Erklärungen für ihre eigenen Entscheidungen liefern.[30] Es geht auf die "Explainable AI"-Initiative zurück, die 2016 von der US Defense Advanced Research Projects Agency ins Leben gerufen wurde. Diese Initiative mehrerer Organisationen und Unternehmen zielt auf die Entwicklung von Methoden zur Entschlüsselung von Algorithmen mit einem sogenannten tiefen Lernprozess (Deep Learning) ab.

---

[28] Buchholtz, Gabriele, Artificial Intelligence and Legal Tech: Challenges to the Rule of Law, in: Wischmeyer, Thomas/Rademacher, Timo (edt) Regulating Artificial Intelligence, Cham: Springer 2019, S. 192.

[29] Vgl. Fn. 24, S. 74.

[30] Nassar, Mohamed/Salah, Khaled/ur Rehman Muhammad Habib/Svetinovic, Davoc Blockchain for explainable and trustworthy artificial intelligence, WIREs Data Mining Knowl. Discov., 10(1),17. Oktober 2019, abrufbar unter: <https://doi.org/10.1002/widm.1340>, S.1 [abgerufen am 12. Mai 2020].

108 Das "Quantitative Input Influence"-Modell beispielsweise ist ein System, das in der Lage ist, den Grad des Einflusses von den Eingabedaten bis zur Ausgabe zu messen.[31] Die *'Layer-wise relevance propagation' XAI* ist ein System, das es ermöglicht, den Denkprozess neuronaler Systeme rückwärts laufen zu lassen und so zu erkennen, welche Neuronen bestimmte Entscheidungen verursacht haben und wie diese zum Ergebnis beigetragen haben.[32] Die *'local interpretable model agnostic explanation'* arbeitet als kontrafaktisches Modell und identifiziert auf der Grundlage von Tausenden von Tests, in denen jeweils minimale Varianten verändert werden, welcher Faktor für eine Entscheidung ausschlaggebend war.[33] Das "*Generalized Additive Model*" findet lineare Trends in Datensätzen und könnte möglicherweise auch auf komplexe Datensätze angewandt werden.[34]

109 Um praktische Lösungen zur Erfüllung der Transparenzanforderungen zu entwickeln, empfiehlt der DAV der Kommission, mehr in Gründerinitiativen und Start-ups zu investieren, die Instrumente zur Erklärung von KI-Systemen entwickeln.

(2) Blockchain-Technologie

110 Ein weiteres Mittel zur Verbesserung der Transparenz von KI-Systemen ist die Nutzung der so genannten Blockchain-Technologie. Bei der Blockchain-Technologie handelt es sich um dezentral verteilte „Register“ (sogenannter „distributed ledger“), welche die Herkunft eines digitalen Vermögenswerts aufzeichnet. Durch den Einsatz der Blockchain-Technologie stehen unveränderliche Aufzeichnungen aller Daten, Variablen und Prozesse zur Verfügung. Diese Aufzeichnungen könnten als Beweismittel verwendet werden und so helfen, eine darauf basierte Entscheidung vor Gericht anzufechten.

---

[31] Datta, Anupam/Sen, Shayak/Zick, Yair, Algorithmic Transparency via Quantitative Input Influence: Theory Experiments with Learning Systems, verfügbar unter: <https://www.andrew.cmu.edu/user/danupam/datta-sen-zick-oakland16.pdf>, S. 1, [abgerufen am 12. Mai 2020].

[32] Montavon, Grégoire/Binder, Alexander/Lapuschkin, Sebastian/Samek, Wojciech/Müller, Klaus-Robert, Layer-Wise Relevance Propagation: An Overview, in: Samek, Wojciech/Montavon, Grégoire/Vedaldi, Andrea/Hansen, Lars Kai/Müller, Klaus-Robert (edt) Explainable AI: Interpreting, Explaining and Visualizing Deep Learning, Lecture Notes in Computer Science, 11700, Cham: Springer, 2019,abrufbar unter: <https://link.springer.com/chapter/10.1007/978-3-030-28954-6_10>, S. 193 [abgerufen am 12. Mai 2020].

[33] Barredo Arieta, Alejandro *u.a.,* Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI, Information Fusion 58, Juni 2020, abrufbar unter: <https://doi.org/10.1016/j.inffus.2019.12.012>, S. 94 [abgerufen am 12. Mai 2020].

[34] Ibid, S. 91.

111 Somit könnte die Blockchain-Technologie einen wesentlichen Beitrag zur Schaffung von KI-Transparenz leisten. Bislang wird dies jedoch nur mit Zurückhaltung als Lösung vorgeschlagen. Ein Grund dafür ist, dass die Blockchain-Technologie einen hohen Energieeinsatz erfordert. Ein weiteres Problem stellt sich im Hinblick auf die DSGVO: Aufgrund der dezentralen Struktur und Funktionsweise der Blockchain ist sie nicht mit dem Grundsatz der DSGVO vereinbar, wonach jede Datenverarbeitung einem Datenverantwortlichen zugewiesen werden können muss. Zudem ist die Tatsache, dass Transaktionen in der Blockchain kaum verändert werden können und daher als Hacking-resistent gelten, nur schwer mit dem Recht auf Vergessen zu vereinbaren, das in Art. 17 DSGVO kodifiziert ist. Nach einer Studie des Europäischen Parlaments[35] sind die festgestellten Spannungen jedoch in erster Linie eine Folge mangelnder Klarheit darüber, wie bestimmte Begriffe der DSGVO auszulegen sind.

112 Angesichts der skizzierten Rechtsunsicherheiten der Blockchain in Bezug auf die DSGVO empfiehlt der DAV der Kommission, diesbezüglich dringend benötigte Richtlinien zu erarbeiten, um die benannten Konflikte aufzulösen.

3. Haftung

113 Eines der Kernprobleme der Haftung für KI liegt darin, dass der Nachweis der Kausalität durch die Automatisierung und beschränkte Nachvollziehbarkeit *ex post* maßgeblich erschwert, wenn nicht sogar verhindert wird. Der Geschädigte hat nahezu keine Möglichkeiten der eigenen Ursachenforschung. Es ist zwar grundsätzlich möglich, sämtliche Entscheidungsprozesse aufzuzeichnen (sog. „Logging“), angesichts der großen Datenmengen gibt es jedoch praktische Kapazitätsprobleme. Dieser Befund wird durch die Verbindung von KI als Software mit Maschinen als Hardware bzw. die Vernetzung mehrerer KIs und Hardware („internet of things“, IoT) noch verstärkt. Zusätzlich ist ein verantwortungsvoller Umgang mit den Daten im Sinne geltender Datenschutzbestimmungen einzufordern.

114 Der hohe Anpassungsbedarf spricht dagegen, den in Deutschland und anderen Mitgliedstaaten bestehenden Dualismus der Haftungsregime der

[35] European Parliamentary Research Service, Blockchain and the General Data Protection Regulation: Can distributed ledgers be squared with European data protection law?, abrufbar unter: <https://www.europarl.europa.eu/RegData/etudes/STUD/2019/634445/EPRS_STU(2019)634445_EN.pdf>, S. 97 [abgerufen am 12. Mai 2020].

verschuldensabhängigen Haftung (hier insbesondere § 823 BGB) und der verschuldensunabhängigen Haftung (hier insbesondere § 1 Produkthaftungsgesetz) auf den Bereich der KI auszudehnen. Dies gilt für die Tatbestandsmerkmale, die Beweislastverteilung und die Haftungsausschlüsse. Um den Besonderheiten der KI Rechnung zu tragen und die absehbaren Fortentwicklungen mit zu erfassen, sollte ein neues Haftungsregime im Sinne einer Gefährdungshaftung nach dem Vorbild der KFZ-Haftung eingeführt werden.

a) Definition

115 Zuzustimmen ist zunächst der Forderung der EU-Kommission nach einer klaren Definition von KI, die einerseits flexibel genug ist, um der fortschreitenden technologischen Entwicklung gerecht zu werden, und andererseits im Interesse der Rechtssicherheit ausreichend präzise ist.[36] Die Details sollten aber Experten mit technischem und juristischem Sachverstand überlassen bleiben.

116 Angesichts der vielfältigen Anwendungsszenarien ist eine präzise Definition nicht einfach. KI könnte selbstständig als reines Softwareprodukt, eingebaut in (fremde) Hardware, oder im Zusammenhang mit einer Dienstleistung auf den Markt gebracht werden. Zum anderen sind auch mögliche Modifikationen durch Dritte, beispielsweise durch Software-Updates oder Add-Ins, zu berücksichtigen.

117 Eine starre Definition wäre für eine derart neue Technologie, die ständiger Innovation unterliegt, wenig hilfreich. In gewissem Maße kann man auf die Rechtsprechung bzw. richterliche Rechtsfortbildung vertrauen, die Gerichte werden angesichts der Komplexität aber mitunter überfordert sein.

b) Haftungsadressaten

118 Eine Unterscheidung zwischen Herstellern bzw. Entwicklern einerseits und Verwendern (nach der Definition der Expertengruppe sog. Operators) andererseits als Adressaten potenzieller Haftungsansprüche ist sinnvoll. Es wird dort auch einer weiteren Unterdifferenzierung bedürfen.

(1) In Bezug auf die Lieferkette

119 Wenn KI nur als reine Software in Verkehr gebracht wird, sollte der Entwickler der einzige Akteur und somit unproblematischer Haftungsadressat sein.

[36] Vgl. etwa die Definitionen in: *Zech* ZfPW 2019, S. 198 (199 f.); Hochrangige Expertengruppe für Künstliche Intelligenz, Eine Definition der KI, S. 8.; *Dettling* PharmR 2019, S. 633 (634)

120 Wenn ein Drittunternehmen die KI mittels Software modifiziert, erweitert oder aktualisiert und dadurch Schäden entstehen, erscheint es zunächst logisch, das Drittunternehmen haften zu lassen. Im bisherigen Recht sieht Art. 3 Abs. 1 der Produkthaftungsrichtlinie (ProdHaftRL) vor, dass neben dem eigentlichen Hersteller auch die Zulieferer für das von ihnen hergestellte Grund- oder Teilprodukt haften, nach Art. 5 ProdHaftRL als Gesamtschuldner. Es fällt aber bereits schwer, den Zuliefererbegriff auf Softwaremodifikationen zu übertragen. Unabhängig davon erscheint die gesamtschuldnerische Haftung problematisch, weil dadurch in vielen Konstellationen auch Gering- oder gar Nichtverursacher haften werden. Die wahre Ursachenforschung wird damit häufig auf den Gesamtschuldnerausgleich verlagert, der dafür kaum geeignet erscheint. Streitigkeiten zwischen Gesamtschuldnern dürften in der Folge enorm zunehmen, weil sehr oft unter den Gesamtschuldnern einer sein wird, der gar keinen oder nur einen sehr geringen und wertungsmäßig zu vernachlässigenden Verursachungsbeitrag erbracht hat. Durch diverse Streitverkündungen bringt dies erhöhte Kostenrisiken für die Prozessbeteiligten mit sich, die eine effiziente Rechtsdurchsetzung erschweren könnten. Die Durchsetzbarkeit von Ansprüchen kann durch viele Streitverkündete und Streithelfer verkompliziert und damit sehr erschwert werden. Insgesamt würde jede Inanspruchnahme einem „Streuschuss“ gleichen, weil jeder, der auch nur entfernt mit der schadensstiftenden KI in Kontakt gekommen ist, als Haftungsadressat einbezogen würde. Eine solche Ausdehnung der Haftung auf „Nicht-Verursacher“ sollte deshalb vermieden werden.

121 Man sollte versuchen, zwar am Grundprinzip der Verursacherhaftung festzuhalten, aber stärkere Haftungsschwerpunkte zu bilden. Angesichts der Komplexität der Materie und der damit zusammenhängenden Beweisschwierigkeiten ist es für den Geschädigten oftmals nahezu unmöglich, in dieser Konstellation den richtigen Haftungsadressaten auszumachen. Ob nun die Softwaremodifikation des Drittunternehmens, die ursprüngliche KI-Software oder ein fehlerhaftes Zusammenwirken beider den Schaden verursacht hat, lässt sich u.a. wegen der Autonomie und Lernfähigkeit von KI und einem regelmäßig technisch begrenzten Logging wahrscheinlich kaum belegen.

122 Falls der Hersteller aber entsprechende Schnittstellen freigibt bzw. den Zugang oder Zugriff auf die Software im Ganzen oder zumindest teilweise ermöglicht,

dann sollte man den Haftungsschwerpunkt bei ihm setzen. Auf eine Haftung von Drittanbietern, die Softwareupdates, Add-Ins o.ä. für die bestehende KI-Software bereitstellen, könnte man dann verzichten. Der Hersteller kann im eigenen Ermessen externe Einflussmöglichkeiten festlegen bzw. beschränken. Er kennt die potenziellen Fehlerquellen und hat die besten Möglichkeiten der Ursachenforschung. Der Geschädigte hätte in diesen Fällen den großen Vorteil eines zentralen und mit vertretbarem Aufwand zu ermittelnden Anspruchsgegners. Falls die Schädigung dann tatsächlich durch die Modifikation des Drittanbieters verursacht wurde, könnte der insoweit beweisbelastete Hersteller Regress bei diesem nehmen. Das wäre dann kein Gesamtschuldnerausgleich, sondern ein klassischer Drittregress.

123 Dasselbe sollte für Einwirkungen auf die KI-Software durch das IoT gelten, wenn die KI mit anderen elektronischen Geräten kommuniziert und auf der Grundlage der Kommunikation ein bestimmtes Verhalten an den Tag legt. Der KI-Entwickler gibt derartige Schnittstellen bewusst frei, um den Nutzern mehr Funktionen oder Services bieten zu können. Er kann die Einflussmöglichkeiten der anderen vernetzten Geräte im Vorhinein festlegen bzw. begrenzen. Er muss auch, wie in der klassischen Produkthaftung, den Markt beobachten. Wäre der Nutzer hier auf eine Feststellung des tatsächlichen Schädigers bzw. des schädigenden vernetzten Gerätes angewiesen, wäre die Verfolgung eines Entschädigungsanspruchs enorm erschwert. Selbst wenn sich das ursächliche, vernetzte Gerät bestimmen ließe, wäre es immer noch fraglich, ob dieses Gerät oder das Reaktionsverhalten der KI fehlerhaft war. Auch hier sollten also Vorschriften etabliert werden, die eine unkomplizierte Entschädigung durch den KI-Entwickler ermöglichen, ohne dass weitere Akteure oder Zusammenhänge erforscht werden müssen. Einem Regress des Entwicklers bei dem Hersteller des jeweiligen vernetzten Gerätes stünde in einem zweiten Schritt nichts entgegen.

124 Wenn KI in Hardware von anderen Herstellern angewendet wird, die ihrerseits Fehler hat und dadurch Schäden verursachen kann, ist die Haftungslage übersichtlicher. Die Ermittlung der Schadensursache dürfte in diesen Fällen leichter sein. Schließlich ergibt sich hier, anders als bei Software-Updates oder der Vernetzung von elektronischen Geräten mit rein softwareinternen Gefahrenpotenzialen, eine Wechselwirkung zwischen Soft- und (greifbarer)

Hardware. Es geht dann nicht um die datenbasierte Modifikation einer bereits vorhandenen und komplexen, weil sich weiterentwickelnden Software, sondern lediglich um die Verwendung einer Steuerungselektronik für eine bereits vorhandene Hardware. Diese Konstellation ist nicht neu (vgl. Betriebssysteme für Computer, Roboter etc.). Es fehlt an der die potenzielle Gefährlichkeit von KI charakterisierenden Intransparenz und Komplexität durch Autonomie und Selbstlernfähigkeit.

125 Möglicherweise ist auch gar nicht ersichtlich, wessen KI-Software in einer Hardware eingesetzt wird. In solchen Fällen erscheint es nach dem Vorbild des „Quasi-Herstellers“ im Sinne der Produkthaftungsrichtlinie sinnvoll, denjenigen als Haftungsadressaten zu etablieren, der nach außen (mit seiner Hardware) als Hersteller auftritt und die Software für die Funktion seiner Hardware benötigt.

(2) In Bezug auf die Nutzer

126 Wenn durch die Nutzung von KI-Systemen Schäden entstehen, müssen auch die Nutzer für einen angemessenen Ausgleich des Geschädigten sorgen. Insoweit erscheint die von der Expertengruppe angedeutete Differenzierung zwischen professionellen und privaten Nutzern von KI aber nicht überzeugend, weil die entstehenden Risiken und potentiellen Schäden nicht davon abhängen, ob ein Nutzer professionell Software nutzt oder privat Die Regeln sollten insoweit identisch sein.

c) Regelungscharakter

127 *Wie im* Weißbuch vorgeschlagen sollten die Regelungen sowohl präventiven Charakter im Sinne einer Gefahrenreduzierung bei Markteinführung haben als auch die Rechtsdurchsetzung im Schadensfall erleichtern. Im deutschen Recht etwa ist dies bereits gängige Praxis (vgl. ProdSG und ProdHaftG).

d) Mögliche Regulierungsoptionen

Im Weißbuch werden verschiedene Optionen zur Regulierung erörtert, zu denen nachfolgend Stellung genommen wird.

(1) Freiwilliges Labeling

128 Das im Weißbuch vorgeschlagene Labeling ist als Ergänzung zu den Haftungsregimen sinnvoll. Auf freiwilliger Basis erscheint es aber kaum

wirkungsvoll. Ein verpflichtendes Labeling für in der EU in Verkehr gebrachte Produkte nach dem Vorbild der CE-Kennzeichnung wäre vorzugswürdig.

(2) Obligatorische risikoabhängige Anforderungen jedenfalls für Hochrisiko-Anwendungen

129 Eine Differenzierung zwischen Hoch- und Niedrigrisiko-Applikationen für neu zu etablierende Haftungsregime, wie eine verschuldensunabhängige Gefährdungshaftung, ist zwar grundsätzlich sachgerecht, sie dürfte aber zu schwierigen Abgrenzungsfragen führen. Die Einzelfallrechtsprechung könnte überfordert sein, für die Marktteilnehmer wäre das Ergebnis wenig vorhersehbar.

130 Sachgerechter wäre die oben vorgeschlagene Risikomatrix (vgl. Randziffer 71 ff.). Für die Einstufung sollten der Einsatzbereich und das Gefährdungspotenzial wesentliche Faktoren sein.

131 Jedenfalls für hochriskante KI-Systeme sollte dann neben das verschuldensunabhängige Haftungsregime der Produkthaftungsrichtlinie und die verschuldensabhängigen, nicht harmonisierten Haftungsregime der einzelnen Mitgliedsstaaten die von der Kommission vorgeschlagene harmonisierte verschuldensunabhängige Haftung als Gefährdungshaftung treten, wie sie zum Beispiel bei KFZ gilt. Unterwirft man nur solche KI-Systeme einer verschuldensunabhängigen Haftung nach dem Vorbild der KFZ- oder Tierhalterhaftung, die ein vergleichbares Gefahrenpotenzial aufweisen, dann vermeidet man eine Überregulierung und ein denkbares und nicht gewünschtes Ausbremsen der Innovationskraft. Die vorgeschlagene Klassifizierung durch die kumulative Bewertung des Einsatzbereichs als Risikobereich (z.B. Gesundheitswesen, Verkehr, Verwaltung etc.) einerseits und des Gefahrenpotenzials der Anwendung andererseits knüpft an die richtigen Kriterien an. Wo genau aber in Bezug auf den Sektor und die Gefährlichkeit entsprechende Grenzen zu ziehen sind, müsste interdisziplinär unter Einbeziehung der Wissenschaft beantwortet werden, noch dazu regelmäßig. So befasst sich auch die bisherige Rechtsprechung laufend mit dem Stand der Technik und leistet mithilfe von Sachverständigen gute Arbeit. Rechtsklarheit für die Entwickler und Verwender von KI bringt die Differenzierung aber nicht. Unternehmen müssten KI wohl immer unter dem Risiko Gefährdungshaftung versichern, um angesichts vielfältiger Einsatzgebiete auf der sicheren Seite zu

sein. Daher könnte man auch von vorneherein eine Gefährdungshaftung für alle Bereiche, in denen KI eingesetzt wird, für den richtigen regulatorischen Ansatz halten.

132 Es ist sinnvoll, eine bereits für andere Haftungsregime etablierte gesetzliche Versicherungspflicht in Betracht zu ziehen. So wird nach den harmonisierten Vorschriften zur KFZ-Versicherung eine Versicherungspflicht für den Halter vorgeschrieben, die einerseits den Geschädigten schützt, indem für eine reibungslose Abwicklung ein solventer Anspruchsgegner (Pflichtversicherung) zur Verfügung steht, und andererseits den Schädiger, der sich möglicherweise besonders hohen Schadensersatzforderungen ausgesetzt sieht. Weder für Nutzer von KI noch für Hersteller von Produkten generell gibt es eine solche Versicherungspflicht bisher auf Unionsebene. Man sollte dies jedenfalls für Hochrisikoanwendungen, vielleicht aber auch für KI insgesamt ernsthaft erwägen.

(3) Sicherheits- und Haftungsregime

133 Wie bereits angesprochen, sind Anpassungen der bisher etablierten Rechts- bzw. Haftungssysteme notwendig, um den Besonderheiten von KI zu begegnen.

134 Modifizieren müsste man auch zwei Ausschlussgründe der Produkthaftungsrichtline: Nach aktuellem Recht können die Hersteller den Ausschlussgrund des später, also nach Inverkehrbringen des Produkts, aufgetretenen Fehlers geltend machen. Und sie können sich auf das Entwicklungsrisiko berufen, wonach eine Haftung ausgeschlossen ist, wenn der Fehler nach dem damaligen Wissensstand (zum Zeitpunkt des Inverkehrbringens) nicht vorhersehbar war. In Anbetracht der durch Autonomie bedingten Lern- und Veränderungsfähigkeit von KI können beide Ausschlussgründe keine oder zumindest nur noch sehr eingeschränkte Anwendung finden. Es wäre widersprüchlich einem Entwickler von KI, der um die Lernfähigkeit seiner Software weiß und diese auch bewusst hervorruft, wegen nach Inverkehrbringen erfolgter Veränderungen, die auf dem Lernprozess beruhen, die Haftung zu erlassen, zumal der Lernprozess immer auf ein bestimmtes Ziel ausgerichtet programmiert ist.

135 In vielen Fällen kann der Entwickler später erkannte Fehler mit Software-Updates ausräumen. Deshalb ist auch die Annahme eines gewissen Mitverschuldens der Nutzer zu begrüßen, wenn sie (sicherheitsrelevante) Updates nicht innerhalb einer angemessenen Frist vornehmen.

e) In Bezug auf die Beweislastverteilung

136 Grundsätzlich ist das Verhalten von KI für den Nutzer nur sehr schwer oder gar nicht nachvollziehbar. Dies liegt nicht nur an der Autonomie und der Lernfähigkeit solcher Systeme. Durch die Verwendung komplexer Techniken wie Algorithmen oder künstlicher neuronaler Netze sind KI-Systeme schon in ihrer „Grundkonfiguration“ äußerst komplex. Vieles bleibt verborgen, weil es um wertvolle Betriebsgeheimnisse geht. Wenn sich solche Systeme aufgrund ihrer Lernfähigkeit und der äußeren Einflüsse zusätzlich verändern (möglicherweise sogar abhängig von den äußeren Bedingungen und anderen, zufälligen Parametern) und dadurch ihre „Grundkonfiguration“ modifizieren, wird dieses Phänomen noch verstärkt.

137 Aufgrund dieser Umstände wird der Nachweis eines Produktfehlers oder Verschuldens, ganz zu schweigen vom Kausalzusammenhang, für den Geschädigten erschwert. Hier müssten aufgrund des notwendigen Fachwissens und der Analysekapazitäten Sachverständige zum Einsatz kommen, was regelmäßig mit hohen Kosten verbunden ist. Solche Kosten könnten viele Geschädigte von der Geltendmachung ihrer Ansprüche abhalten.

138 Zusätzlich ergeben sich durch unterschiedliche Einsatzmöglichkeiten von KI weitere Probleme: Wird KI nicht ausschließlich als bloße Software vermarktet, wird sie regelmäßig Anwendung in Hardware von anderen Herstellern finden, die ihrerseits Fehler haben und dadurch Schäden verursachen kann. Die Ursachenfindung für durch KI verursachte Schäden kann ferner auch durch Software-Updates bzw. Add-Ins von Drittanbietern, ähnlich den sogenannten „Apps“ auf Smartphones, erschwert werden. Dasselbe gilt für das IoT, wenn die KI mit anderen elektronischen Geräten kommuniziert und sogar Befehle von diesen annimmt oder zumindest auf der Grundlage der Kommunikation ein bestimmtes Verhalten zeigt.

139 Geschädigten müssen gewisse Beweiserleichterungen für bestehende Haftungsregime in Bezug auf das Verschulden bzw. den Fehler und den

Ursachenzusammenhang zugesprochen werden. Ebenso ist ein Auskunftsanspruch erforderlich, der schon im Vorfeld einer Inanspruchnahme greifen müsste. Idealerweise sollte das nur für diejenigen Geschädigten gelten, die tatsächlich mit solchen Schwierigkeiten konfrontiert sind.

140 Bei der von der EU-Kommission in Erwägung gezogenen verschuldensunabhängigen Haftung von Entwicklern und Nutzern von Hochrisikoanwendungen sollte es wie bei der KFZ-Haftung ausreichend sein, wenn der Geschädigte beweist, den Schaden beim Betrieb der KI bzw. des mit der KI verbundenen Produkts bzw. der Dienstleistung erlitten zu haben. Ein Verschulden des Schädigers, einen Fehler und einen Kausalzusammenhang sollte der Geschädigte nicht nachweisen müssen. Die Darlegung des Schadens unterliegt auch bei KI-Systemen keinen besonderen Schwierigkeiten.

141 Außerhalb von Hochrisikoanwendungen sollten aber die bisherigen Grundsätze des Haftungsrechts gelten: Möglichen Beweisschwierigkeiten sollte hier mithilfe des sog. „Loggings“ begegnet werden: Die Hersteller müssen verpflichtet werden, die entsprechend relevanten Daten aufzuzeichnen und zur Verfügung zu stellen, also eine besondere Ausprägung der Befundsicherungspflicht aus der Produkthaftung. Für den Fall, dass eine KI diese Daten nicht, falsch oder unvollständig aufzeichnet oder abspeichert, sollte eine Beweislastumkehr für Fehler und Kausalität zu Lasten des Herstellers zur Anwendung kommen. Allerdings könnte es mit Blick auf die fortschreitende Entwicklung wohl immer schwieriger werden, derartige Datenmengen auf Dauer aufzubewahren.

142 Gegenüber Nutzern sollte außerhalb von Hochrisikoanwendungen dem geschädigten Dritten eine Beweiserleichterung im Hinblick auf das Verschulden des Nutzers zukommen. Der Nutzer der KI müsste sein Verschulden dann widerlegen. Hier sollte keine Differenzierung zwischen privaten und professionellen Nutzern von KI stattfinden, weil die Sachlage vergleichbar ist.

IV. Frage 4: Welche einzelnen Akteure oder Gruppen von Akteuren sind am besten in der Lage, den von Ihnen identifizierten Risiken zu begegnen?

143 Drei verschiedene Gruppen von Akteuren können für die Bewältigung von Risiken von KI-Anwendungen identifiziert werden.

144 Zunächst sollte der Gesetzgeber - sei es auf nationaler oder auf EU-Ebene - dafür sorgen, dass bestehende Regelungen auf KI-Systeme Anwendung finden und dass Regelungslücken (wie im Falle der Haftung, die im obigen Abschnitt weiter ausgeführt wird) geschlossen werden und dass diese Gesetze effizient durchgesetzt werden. Rechtsanwälte müssen als freie und unabhängige Berater in allen Rechtsangelegenheiten und als unabhängige Organe der Justizverwaltung eine zentrale Rolle bei der Gewährleistung des Schutzes der Rechtsstaatlichkeit und der Verteidigung der Grundrechte spielen.

145 Zweitens könnten quasi-staatliche Akteure mit Zertifizierungs- und Prüfungsaufgaben betraut werden, wenn es um bestimmte risikoreiche KI-Anwendungen geht. Der Vorteil quasi-staatlicher Stellen gegenüber unabhängigen Dritten bestünde darin, dass es für Unternehmen schwieriger wäre, sich ihnen gegenüber auf Geschäftsgeheimnisse oder Urheberrechte zu berufen. Folglich wären Transparenzvorschriften der in besonders kritischen Bereichen verwendeten Algorithmen leichter zu überprüfen.

146 Drittens könnten unabhängige Dritte eine wichtige Rolle spielen, bestimmte Anforderungen zu überprüfen und etwa freiwillige oder verpflichtende Labels in Kategorien mit geringem Risiko zu vergeben. Die Anforderungen und Verfahren für die Kennzeichnung von KI als „vertrauenswürdige KI“ sollten auf EU-Ebene definiert und nicht den einzelnen Mitgliedstaaten überlassen werden. Die Erfahrungen mit der Einführung eines Labels für die Einhaltung der Anforderungen der DSGVO haben gezeigt, dass ein solches Label wenig praktischen Nutzen hat, wenn es nicht EU-weit einheitlich angewandt wird.[37] Daher ist es entscheidend, bei der Schaffung eines EU-weiten Labels für "vertrauenswürdige KI" einheitliche und klare Regeln auf EU-Ebene zu erarbeiten.

[37] Gasparotti, Alessandro/Harta, Lukas, Europäische Strategie zur Künstlichen Intelligenz, 11. Februar 2020, abrufbar unter: https://www.cep.eu/fileadmin/user_upload/cep.eu/Studien/cepAdhoc_Europaeische_Strategie_zur_kuenstlichen_Intelligenz/cepAdhoc_Europaeische_Strategie_zur_kuenstlichen_Intelligenz.pdf>, S. 5 [abgerufen am 12. Mai 2020].

V. Frage 5: Welche Situationen sind Ihnen bekannt, in denen die Verwendung von KI-Anwendungen die effektive Einhaltung oder Durchsetzung der geltenden Rechtsvorschriften erschwert hat?

147 KI-basierte Anwendungen auf Social-Media Plattformen stellen eine besondere Herausforderung für die Rechtsdurchsetzung dar.

148 Obwohl der Straftatbestand der Beleidigung, Urheberrechtsverletzungen und andere typische online-relevante Delikte auf EU- oder nationaler Ebene geregelt sind, finden viele KI-basierte Handlungen in einer gewissen "Grauzone" zwischen rechtmäßigem und unrechtmäßigem Verhalten statt. Folglich ist die Überwachung der Einhaltung dieser Gesetze besonders schwierig. Darüber hinaus können die Akteure ihre Identität leicht hinter ihren Anwendungen verbergen.

149 Aufgrund der begrenzten Regulierungsmöglichkeiten in diesem Bereich ist es insbesondere Aufgabe der Gerichte und Anwälte, die widerstreitenden Interessen miteinander in Einklang zu bringen. Im Folgenden sollen zwei besonders kritische KI-Anwendungen, nämlich Social Bots und Deep Fakes, skizziert werden. Anschließend werden die Hauptschwierigkeiten dieser Instrumente und mögliche Lösungen analysiert.

1. Beispiel 1: Social Bots

150 Ein kritisches KI-Tool in diesem Zusammenhang, das weder eindeutig rechtmäßig noch unrechtmäßig ist, ist der Einsatz von Social Bots, d.h. von Konten, die vollständig durch Software kontrolliert werden. Diese Konten sind kritisch, da sie zur Platzierung von Werbung, aber auch zur Verbreitung von Informationen oder Desinformationen eingesetzt werden können. Der Einsatz solcher Social Bots hat in Ländern wie den Vereinigten Staaten oder Brasilien, aber auch in vielen kleineren Nationen eine grundlegende Rolle im Wahlkampf gespielt. Was sie risikobehafteter macht als von Menschen kontrollierte Benutzerkonten, ist die Fähigkeit, den Eindruck zu erwecken, eine große Anzahl von Benutzern teile eine bestimmte Meinung, sodass sie als Multiplikatoren bei der öffentlichen Meinungsbildung eingesetzt werden können.[38] Die

[38] Krönke, Christoph, Social Media and Artificial Intelligence, in: Wischmeyer, Thomas/Rademacher, Timo (edt) Regulating Artificial Intelligence, Cham: Springer 2019, S. 149.

Rechtsdurchsetzung wird zusätzlich durch die Tatsache erschwert, dass Nutzer von Social Bots ihre Identität oft hinter ihren Werkzeugen verbergen können.

2. Beispiel 2: Deep fakes

151 Eine weitere sehr kritische KI-Anwendung auf Social Media ist die der sogenannten "Deep Fakes". Deep Fake-Videos nutzen Deep-Learning-Algorithmen und kombinieren diese mit der Eingabe von großen Videobilddateien, um neue visuelle Produkte herzustellen. Dabei handelt es sich um Produkte von mindestens zwei KI-Algorithmen, einem "Generator" und einem "Diskriminator-Algorithmus", die in einem "generativen feindlichen Netzwerk" zusammenarbeiten.[39]

152 Die Gefahren dieser Technologie sind vielfältig: „Deep-faked“ Nachrichtenberichte könnten den Ruf einzelner Personen beschädigen, falsche oder erfundene Ereignisse (z.B. einen vorgetäuschten Terroranschlag) vortäuschen oder Wahlkampagnen beeinflussen. Langfristig könnten sie gezielt eingesetzt werden, um das Vertrauen in politische Institutionen zu untergraben und die Polarisierung zwischen gesellschaftlichen Gruppen zu vertiefen. Daher gefährdet das Wesen von "deep fakes" die Grundrechte sowie die Grundprinzipien liberaler Demokratien. [40]

153 Diese Gefahren werden noch greifbarer, wenn man sich aktuelle Studien ansieht, die darauf hindeuten, dass Menschen typischerweise nicht nur ihre Fähigkeit überschätzen, Wahres von Falschem zu trennen, sondern auch einseitig solche politische Nachrichten, die mit ihren Überzeugungen übereinstimmen, mehr Gewicht geben und dagegen Nachrichten, die ihren Überzeugungen widersprechen, weniger Bedeutung beimessen.[41]

3. Maßnahmen zur Verbesserung der Durchsetzung

a) Regulatorische Schwierigkeiten

154 Die Regulierung im Bereich der sozialen Medien weist in vielerlei Hinsicht Schwierigkeiten auf:

---

[39] Chivers*, Tom,* What do we do about deepfake videos, The Guardian,23. Juni 2019, verfügbar unter: <https://www.theguardian.com/technology/2019/jun/23/what-do-we-do-about-deepfake-video-ai-facebook> [abgerufen am 12. Mai 2020].
[40] Meskys, Edvinas/Liaudanskas, Aidas/Kalpokiene, Julija/Jurcys, Paulius., Regulating deep fakes: legal and ethical considerations, Journal of Intellectual Property Law and Practice 2020, 15 (1), S.. 24, 31.
[41] Ebenda.

155 Das Filtern und Blockieren von Inhalten in sozialen Medien kann das individuelle Recht der Nutzer auf Zugang zu Informationen, das im Recht auf freie Meinungsäußerung enthalten ist, gefährden. Andererseits kann das Recht der Anbieter auf freie Meinungsäußerung infolge einer zu starken Regulierung der sozialen Medien beeinträchtigt werden. [42] Darüber hinaus könnten bestimmte Anwendungen, die darauf abzielen, die politische Debatte zu beeinflussen, das dem Recht auf freie Meinungsäußerung innewohnenden Ziel, ein günstiges Umfeld für eine pluralistische öffentliche Debatte zu schaffen, negativ beeinträchtigen.

156 Dies wirft auch die Frage auf, ob die Hauptverantwortung für die Überwachung der Einhaltung bei den Betreibern von sozialen Plattformen oder bei den Staatsanwaltschaften liegen sollte. Die Übertragung der Verantwortung für die Entfernung illegaler Inhalte auf die Anbieter birgt das Risiko, die Verantwortung für die Strafverfolgung an private Unternehmen abzugeben. Das Fehlen einer sinnvollen und wirksamen staatlichen Aufsicht könnte auch Bedenken hinsichtlich der Rechtsstaatlichkeit aufwerfen.[43]

157 Darüber hinaus sind die Algorithmen, die zur Aufdeckung dieser illegalen Handlungen verwendet werden, derzeit noch nicht in der Lage, ironische oder kritische Beiträge als solche zu erkennen. Das Filtern und Beseitigen schädlicher Inhalte durch Algorithmen birgt daher ein hohes Risiko der Überblockierung und Entfernung von Inhalten, die einen positiven Beitrag zur öffentlichen Debatte leisten könnten.[44]

158 Von regulatorischer Seite empfiehlt der DAV der Kommission daher, die Besonderheiten von KI-Anwendungen beim Entwurf des neuen Gesetzes über digitale Dienste zu berücksichtigen, das neue Haftungsregeln für Plattformbetreiber auf EU-Ebene einführen wird.

b) Mögliche Vorgehensweisen

159 Um die Anwendung und Durchsetzung der bestehenden Gesetze in diesem Bereich zu verbessern, sollten die folgenden Überlegungen berücksichtigt werden:

[42] Vgl. Fn. 27, S.. 31.
[43] Ebenda.
[44] Ebenda.

160 Zunächst sollte anerkannt werden, dass die Schaffung neuer öffentlicher Sphären für die öffentliche Meinungsbildung auf sozialen Plattformen nicht per se verboten werden kann. Tatsächlich gibt es weder den Prototyp einer demokratischen Öffentlichkeit noch ein festes Modell öffentlicher Kommunikation. Mit der Digitalisierung sind neue Formen der Kommunikation entstanden und haben ihre klassischen normativen Konzepte ersetzt.[45] Politische Aktivitäten auf sozialen Medien als solche stellen daher offensichtlich keine Bedrohung der Demokratie dar. In Verbindung mit Technologie können sie jedoch nicht nur Meinungen, sondern auch die Integrität eines rechtsstaatlichen Verfahrens und die Rechtsstaatlichkeit beeinträchtigen.

161 Die notwendige und kritische Einzelabwägung wird daher nicht auf der Regulierungs-, sondern auf der Vollzugsebene gelöst: Hier ist es insbesondere die Aufgabe von Anwälten und Gerichten, die widerstreitenden Interessen auszugleichen. So sollten einige KI-Anwendungen gar nicht erst von dem Schutzbereich der Grundrechte erfasst werden. Eine solche Ausnahme könnte z.B. im Hinblick auf Anbieter diskutiert werden, die mit Social Bots unter Pseudonymen operieren und damit fälschlicherweise vorgeben, dass die Konten tatsächlich von (vielen) Menschen geführt werden. Darüber hinaus könnten auch Nutzer, die gezielte (Fehl-)Informationskampagnen auf der Basis von KI durchführen, unter diesen Ausschluss fallen.[46]

162 Um die Durchsetzung von KI-basierten Rechtsverletzungen in sozialen Medien zu erleichtern, sollten die Mitgliedstaaten ermutigt werden, in Staatsanwaltschaften und Gerichten stärker spezialisierte Einheiten zur Bekämpfung der Cyberkriminalität zu schaffen. Dies könnte zusätzlich zu einer richtungsweisenden Rechtsprechung führen.

163 Aus technischer Sicht sollte auf EU-Ebene die Entwicklung von Instrumenten gefördert werden, um Fake-Accounts und Fake-News besser herauszufiltern und Social Bots aufzuspüren. Im Hinblick auf "Deep Fakes" könnten beispielsweise neuronale Netze eingesetzt werden, um Augenblinzeln in Videos zu erkennen,

[45] Vgl. Fn. 38, S. 156.
[46] Vgl. Fn. 38, S. 154.

da dies ein typisches Anzeichen eines gefälschten Videos darstellt.[47] Weitere Studien zitieren Blockchain als eine Lösung für die Erstellung manipulationssicherer Inhalte.[48]

164 Technische Lösungen können jedoch immer nur ein Teil der Problemlösung sein. Wie das Beispiel der Deep Fakes zeigt, wird es immer wieder neue Erfindungen geben, die jeweils neue Herausforderungen mit sich bringen. Umso wichtiger ist es, dass Rechtsanwälte und Behörden ihre Mitarbeiter weiterhin für den Umgang mit neuen Risiken schulen und ausbilden. Dies verdeutlicht letztlich auch die überragende Bedeutung menschlicher Expertise im Rahmen schwieriger Abwägungsprozesse. Diese können und sollten nicht durch Technologie ersetzt werden.

Bibliografie:

Algorithm Watch, Automating Society – Taking Stock of Automated Decision-Making in the EU, A Report by AlgorithmWatch in Zusammenarbeit mit der Bertelsmann Stiftung, unterstützt von der Open Society Foundation, 1. Auflage, Januar 2019, S. 100, abrufbar unter: <http://www.algorithmwatch.org/automating-society> [abgerufen am 12. Mai 2020].

Barredo Arieta, Alejandro *u.a.*, Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI, Information Fusion 58, Juni 2020, abrufbar unter: <https://doi.org/10.1016/j.inffus.2019.12.012>, S. 82 [abgerufen am 12. Mai 2020].

Buchholtz, Gabriele, Artificial Intelligence and Legal Tech: Challenges to the Rule of Law, in: Wischmeyer, Thomas/Rademacher, Timo (edt) Regulating Artificial Intelligence, Cham: Springer 2019, S. 192.

Chivers, Tom, What do we do about deepfake videos, The Guardian, 23 Juni 2019, abrufbar unter: <https://www.theguardian.com/technology/2019/jun/23/what-do-we-do-about-deepfake-video-ai-facebook> [abgerufen am 12. Mai 2020].

Cui, Yadong, Artificial Intelligence and Judicial Modernization, Shanghai: Springer 2020, S. 26.

Datta, Anupam/Sen, Shayak/Zick, Yair, Algorithmic Transparency via Quantitative Input Influence: Theory Experiments with Learning Systems,

[47] Li, Yuezun/Chang, Ming-Ching/Lyu, Siwei, In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking, 2018 IEEE International Workshop on Information Forensics and Security (WIFS), 2018, abrufbar unter: <https://arxiv.org/pdf/1806.02877.pdf> [abgerufen am 12. Mai 2020].
[48] *Vgl. Fn*. 40, S.. 30.

abrufbar unter: <https://www.andrew.cmu.edu/user/danupam/datta-sen-zick-oakland16.pdf>, S. 1, [abgerufen am 12. Mai 2020].

Datenethikkommission der Bundesregierung, Gutachten der Datenethikkommission, 22. Januar 2020, abrufbar unter: <https://www.bmjv.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_EN_lang.html>, S. 177 [abgerufen am 12. Mai 2020].

Djeffal, Christian, Artificial Intelligence and Public Governance Normative Guidelines for Artificial Intelligence in Government and Public Administration, in: Wischmeyer, Thomas/Rademacher, Timo (edt) Regulating Artificial Intelligence, Cham: Springer 2019, S. 281.

Enders, Peter, Einsatz künstlicher Intelligenz bei juristischer Entscheidungsfindung, JA 2018, S. 721.

European Parliamentary Research Service, Blockchain and the General Data Protection Regulation: Can distributed ledgers be squared with European data protection law?, abrufbar unter: <https://www.europarl.europa.eu/RegData/etudes/STUD/2019/634445/EPRS_STU(2019)634445_EN.pdf>, S. 97 [abgerufen am 12. Mai 2020].

Feng, Zhen/Xia, Helen, China: Three Cyberspace Courts now online and open for business, October 16, 2018, abrufbar unter: <https://www.jdsupra.com/legalnews/three-cyberspace-courts-now-online-and-91459> [abgerufen am 12. Mai 2020].

Gasparotti, Alessandro/Harta, Lukas, Europäische Strategie zur Künstlichen Intelligenz – Eine Bewertung des Entwurfs eines Weißbuchs der EU-Kommission zur KI, 11. Februar 2020, abrufbar unter:<https://www.cep.eu/fileadmin/user_upload/cep.eu/Studien/cepAdhoc_Europaeische_Strategie_zur_kuenstlichen_Intelligenz/cepAdhoc_Europaeische_Strategie_zur_kuenstlichen_Intelligenz.pdf>, S. 5 [abgerufen am 12. Mai 2020].

Harris, Briony, Could an AI ever replace a judge in court?, 11. Juli 2018, abrufbar unter:<https://www.worldgovernmentsummit.org/observer/articles/could-an-ai-ever-replace-a-judge-in-court> [abgerufen am 12. Mai 2020].

Hill, Caroline, ‘Machine beats man’ in Casecrunch lawyer challenge, Legal IT Insider, 30. Oktober 2017, abrufbar unter <https://legaltechnology.com/machine-beats-man-in-casecrunch-lawyer-challenge/> [abgerufen am 12. Mai 2020].

Hillebrand Pohl, Jens, The Right to Be Heard in European Union Law and the International Minimum Standard- Due Process, Transparency and the Rule of Law, 8. Juni 2018, abrufbar unter:<https://ssrn.com/abstract=3192858>, S. 3 [abgerufen am 12. Mai 2020].

Hochrangige Expertengruppe für Künstliche Intelligenz: Eine Definition der KI, abrufbar unter: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60664

iBorderCtrl, Intelligent Portable Control System, Projektpräsentation, abrufbar unter: <https://www.iborderctrl.eu/sites/default/files/publications/iBorderCtrl%20global%20presentation%20v5.pdf> [abgerufen am 12. Mai 2020].

Krönke, Christoph, Social Media and Artificial Intelligence, in: Wischmeyer, Thomas/Rademacher, Timo (edt) Regulating Artificial Intelligence, Cham: Springer 2019, S. 149.

Li, Yuezun/Chang, Ming-Ching/Lyu, Siwei, In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking, 2018 IEEE International Workshop on Information Forensics and Security (WIFS), 2018, abrufbar unter: <https://arxiv.org/pdf/1806.02877.pdf> [abgerufen am 12. Mai 2020].

Marr, Bernard, The Future of Lawyers: Legal Tech, AI, Big Data and Online Courts, Forbes,17. Januar 2020, abrufbar unter: <https://www.forbes.com/sites/bernardmarr/2020/01/17/the-future-of-lawyers-legal-tech-ai-big-data-and-online-courts> [abgerufen am 12. Mai 2020].

Martini, Mario, Grundlinien Eines Kontrollsystems für algorithmenbasierte Entscheidungsprozesse, abrufbar unter: <https://www.uni-speyer.de/fileadmin/Lehrstuehle/Martini/2019_Gutachten_GrundlageneinesKontrollsystemendgueltig.pdf> [abgerufen am 12. Mai 2020].

Matczak, Marcin, 10 Facts on Poland for the Consideration of the European Court of Justice, 13. Mai 2018, abrufbar unter <https://verfassungsblog.de/10-facts-on-poland-for-the-consideration-of-the-european-court-of-justice/>, unter Bezugnahme auf den Fall Daktaras gg. Litauen – Beschwerdenr. 42095/98 [abgerufen am 12. Mai 2020].

Meskys, Edvinas/Liaudanskas, Aidas/Kalpokiene, Julija/Jurcys, Paulius., Regulating deep fakes: legal and ethical considerations, Journal of Intellectual Property Law and Practice 2020, 15 (1), S. 24.

Montavon, Grégoire/Binder, Alexander/Lapuschkin, Sebastian/Samek, Wojciech/Müller, Klaus-Robert, Layer-Wise Relevance Propagation: An Overview, in: Samek, Wojciech/Montavon, Grégoire/Vedaldi, Andrea/Hansen, Lars Kai/Müller, Klaus-Robert (edt) Explainable AI: Interpreting, Explaining and Visualizing Deep Learning, Lecture Notes in Computer Science, 11700, Cham: Springer, 2019, S. 193.

Nassar, Mohamed/Salah, Khaled/ur Rehman Muhammad Habib/Svetinovic, Davoc Blockchain for explainable and trustworthy artificial intelligence, WIREs Data Mining Knowl. Discov., 10(1), 17. Oktober 2019 <https://doi.org/10.1002/widm.1340> [abgerufen am 12. Mai 2020].

Niller, Eric, Can AI Be a Fair Judge in Court? Estonia Thinks So, WIRED, 25. März 2020, verfügbar unter: <https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so> [abgerufen am 12. Mai 2020].

Ronsin, Xavier/Lampos, Vasileios, Appendix I – In-depth study on the use of AI in judicial systems, notably AI applications processing judicial decisions and data, in: European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment. Strasbourg, CEPEJ - Commission Européene pour l'Efficacité de la Justice, 2018, abrufbar unter: <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>, S. 42 [abgerufen am 12. Mai 2020].

Yeung, Karen, Responsibility and AI, Council of Europe Study, September 2019, abrufbar unter: <https://rm.coe.int/responsability-and-ai-en/168097d9c5>, S. 21 [abgerufen am 12. Mai 2020].

Zweig, Katharina, Algorithmische Entscheidungen: Transparenz und Kontrolle, Januar 2019, abrufbar unter: <https://www.kas.de/documents/252038/4521287/AA338+Algorithmische+Entscheidungen.pdf/533ef913-e567-987d-54c3-1906395cdb81?version=1.0&t=1548228380797>, abgerufen am 12. Mai 2020].

Unbekannter Verfasser, Mannheim testet verhaltensbasierte Videoüberwachung, Heise Online, 3. Dezember 2018, abrufbar unter: <https://www.heise.de/newsticker/meldung/Mannheim-testet-verhaltensbasierte-Videoueberwachung-4239279.html>, [abgerufen am 12. Mai 2020].