Risikoregulierung der KI: normative Herausforderungen und politische Entscheidungen

Stellungnahme zum Weißbuch der Europäischen Kommission „Zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen“

von
Dr. Carsten Orwat, Anja Folberth M.A., Jascha Bareis M.A.,
Dr. Jutta Jahnel, Christian Wadephul M.A.[1]

Karlsruher Institut für Technologie (KIT),
Institut für Technikfolgenabschätzung und Systemanalyse (ITAS)

14. Juni 2020

Zusammenfassung:

Wir bedanken uns für die Möglichkeit, Stellung zum Weißbuch „Zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen“ (COM(2020) 65 final, vom 19.2.2020) nehmen zu können. Die nachfolgenden Ausführungen beziehen sich auf die deutschsprachige Version und adressieren überwiegend den Teil zur Risikoregulierung. Es werden die Herausforderungen für eine Risikoregulierung der KI dargestellt, die sich vor allem in der normativen Ambiguität bei der Risikobestimmung und -bewertung zeigt. Sie erfordert umfangreiche politische Prozesse, bevor eine Risikoregulierung eingerichtet und betrieben werden kann. Des Weiteren werden einzelne Vorschläge zum Ausbau der Risikoregulierung unterbreitet.

1 Einleitung

Mittlerweile finden sich zahlreiche Anhaltspunkte, dass existierende oder künftige Anwendungen künstlicher Intelligenz (KI) und algorithmischer Entscheidungssysteme (AES) Risiken und konkrete Beeinträchtigungen von Menschen- und Verfassungsrechten und weiteren gesellschaftlichen Grundwerten, wie Demokratie und Rechtsstaatlichkeit, haben.[2] Zum Schutz der Grundrechte und -werte sind zahlreiche Vorschläge zur Regulierung von KI und AES unterbreitet worden, wobei

[1] Die Stellungnahme ist im Rahmen des Forschungsprojekts „Governance von und durch Algorithmen“ entstanden, das vom Bundesministerium für Bildung und Forschung (BMBF) gefördert wird (Förderkennzeichen 01IS19020B). Die hier geäußerten Meinungen sind allein die der Autor*innen und in keiner Weise die des BMBF oder eine offizielle Stellungnahme des KIT oder ITAS. Die Autor*innen werden teilweise über die Projektförderung und teilweise durch die institutionelle Förderung des ITAS über die Helmholtz-Gemeinschaft, die von Bund und Ländern getragen wird, finanziert. An dieser Stelle sei dankend auf die Förderungen hingewiesen.

[2] So auch im Weißbuch, Europäische Kommission (2020: 12-14).

gegenwärtig Vorschläge zur risiko-basierten oder risiko-adäquaten Regulierungsansätzen hervortreten.[3] Dabei stellt die *risiko-basierte Regulierung* nur eine von vielen Formen der Risikoregulierung[4] dar. Mit der risiko-basierten Regulierung wird insbesondere auf eine Priorisierung der Regulierungsaktivitäten abgezielt, die die Ressourcen der Regulierungsbehörde schont und sich auf diejenigen Regulierungsobjekte fokussiert, denen ein hohes Risiko zugeschrieben wird.[5]

2 Vorteile des Vorschlags der Europäischen Kommission

Im Weißbuch wird besonders treffend betont, dass eine Besonderheit von KI- und AES-Anwendungen die schwierige und teilweise gar unmögliche Nachvollziehbarkeit der Entscheidungen ist. Dies gilt auch für die Charakterisierung, der sich daraus ergebenden Schwierigkeiten, die Ansprüche der Betroffenen auf rechtlichem Wege durchzusetzen (S. 14). Des Weiteren berücksichtigt das Weißbuch explizit, dass auch Risiken für Grundrechte und Grundwerte jenseits der genannten (S. 12f.) einbezogen werden sollten. Dadurch könnten auch heute noch nicht genau abschätzbare „neue“ Risiken (z.B. psychische Auswirkungen, S. 18) und weitere mögliche Grundrechts- und Grundwertverletzungen in dem Regulierungsrahmen adressiert werden.

Zudem kann der Vorschlag so verstanden werden, dass auch gegebenenfalls neue Risikotypen, die nicht punktuell durch einzelne Systeme verursacht werden, sondern eher systemischen Charakter haben[6], in der Risikoregulierungskonzeption Beachtung finden können. Die Kommission schlägt ferner weitere Kriterien für die Kategorisierung von „hohen Risiken“ vor (S. 20) und es ist die Absicht erkennbar, dass das Risikoverständnis über bisherige Ansätze erweitert werden kann.[7] Auch verdeutlicht der Vorschlag treffend die Auffassung, dass der bestehende Rechtsrahmen angesichts der neuen Herausforderungen überprüft und existierende Regelungen, z.B. zur Produktsicherheit und -haftung, angepasst und erweitert werden sollten. Hier geht die Kommission davon aus, dass auch eine spezifisch auf die KI zugeschnittene zusätzliche Regulierung notwendig sei, wobei gleichzeitig die Überprüfung und mögliche Anpassung des Rechtsrahmens angestrebt wird (S. 15-17). Das kann die Dringlichkeit und Bedeutung, mit der die Kommission den Regulierungsdarf sieht, verdeutlichen.

---

[3] So im Weißbuch, Europäische Kommission (2020: 20ff.), ebenso AI HLEG (2019b), Datenethikkommission (2019) oder Council of Europe (2020).

[4] Übersichten z.B. in Hood et al. (2001), Renn (2008), Black (2010b) oder Alemanno (2016).

[5] Nach Black (2010a: 330-332) und Black (2010a).

[6] Das kann aus der Ausführung zur Produkthaftung „Einige dieser Risiken sind zwar nicht auf KI-gestützte Produkte und Dienstleistungen beschränkt, allerdings kann der Einsatz von KI die Risiken erhöhen oder verschärfen.“ (S. 14 des Weißbuchs) interpretiert werden.

[7] So kann Fußnote 49 des Weißbuchs („In den EU-Rechtsvorschriften können »Risiken« je nach dem betreffenden Bereich, z. B. Produktsicherheit, anders eingestuft werden als hier beschrieben.“) interpretiert werden, Europäische Kommission (2020: 20).

3 Wissensgrundlagen für die Risikoregulierung

Für die Gestaltung der Risikoregulierung ist vor allem entscheidend, wie die Risiken charakterisiert[8] sind und welches *Wissen über Risiken* bei welchen Akteur*innen vorliegt oder erzeugt werden kann. Dabei ist zu beachten, dass sich die Risikocharakteristika und Wissengrundlagen der Risikoregulierung von KI- und AES-Anwendungen stark von denen bestehender Risikoregulierungen unterscheiden. Übliche Risikoregulierung basiert in vielen Fällen auf naturwissenschaftlichem Wissen zur möglichst objektiven Bestimmung von Risiken.[9] Als wissenschaftliches Wissen ist es nachvollziehbar und anfechtbar und damit fähig zur Weiterentwicklung und kann zu konsolidiertem Wissen[10] ausgearbeitet werden, sodass Risikoabschätzungen mit einem relativ hohen Grad an Eindeutigkeit und Überprüfbarkeit möglich sind.[11] Selbst in diesen „harten" naturwissenschaftlichen Bereichen wird zur Vorsicht bei der Umsetzung von risiko-basierten Regulierungsansätzen gemahnt.[12] Im Vergleich zu diesen meist naturwissenschaftlich geprägten Bereichen der Risikoregulierung, sind die Wissensgrundlagen bei KI- und AES-Anwendungen unvollständig, gesellschaftlich stark asymmetrisch verteilt und durch mangelnde Eindeutigkeit und Unsicherheiten bei den Bewertungsmaßstäben geprägt.

Zunächst geht es um Wissen über die für einzelne Anwendungen relevanten Wahrscheinlichkeiten von adversen Ereignissen, die durch Anwendungen von KI und AES verursacht werden können, und die mögliche Anzahl von betroffenen Personen (z.B. Zahl der Kund*innen oder Bewerbende). Dieses Wissen ist weitgehend in den Händen der Betreibenden. Hierzu kann eine relativ geringe Unsicherheit über Ursache-Wirkungszusammenhänge vermutet werden, da bei deterministischen Algorithmen die Ursache-Wirkungszusammenhänge in AES eindeutig sind und bei nicht-deterministischen Algorithmen durch Testen und empirische Untersuchungen bei entsprechendem Aufwand ermittelt werden könnten. Forschungen zur Nachvollziehbarkeit von KI („Explainable AI") versprechen hier, zur Minderung der Unsicherheiten beizutragen.

Jedoch bestehen Unklarheit, Unbestimmtheit und Uneindeutigkeit vor allem bei Wissen, das zur Einschätzung und Bewertung der Risiken erforderlich ist. Denn die oft abstrakten Grundrechte und Grundwerte werden in der Gesellschaft sehr unterschiedlich interpretiert, kontrovers diskutiert und liegen nicht als klare und eindeutige Bewertungskriterien für Risikoabschätzungen vor. Zudem dürfte bei vielen Anwendungen die gesellschaftliche Größenordnung der Risiken nicht klar sein, da nicht bekannt ist, wie oft eine bestimmte Anwendung von KI und AES in der Praxis eingesetzt wird. Ferner gibt es Risiken, bei denen die Verursachung von Schäden für die Gesellschaft nicht einzelnen Verfahren zugeschrieben werden kann, wie z.B. indirekt wirkende Abschreckungseffekte für ganze Bevölkerungsgruppen (siehe auch im Folgenden).

---

8 Z.B. Renn (2008: 173-200).
9 Vgl. z.B. Fisher (2012: 420).
10 Vgl. Grunwald (2019: 28).
11 Siehe dazu z.B. Hansson & Aven (2014).
12 Siehe dazu Lloyd-Bostock & Hutter (2008).

In den Risikoabschätzungen müssten zudem die konkreten potentiellen Schädigungen in Betracht gezogen werden, die durch eine hohe Multidimensionalität gekennzeichnet sind und deren Beziehung zu den Schutzzielen der Grundrechte und Grundwerte verdeutlicht werden müssen. Abgesehen von Datenschutzskandalen mit Massen von Betroffenen, liegen eher sehr viele, „kleinteiligere“[13] und sehr unterschiedliche Schädigungen vor, die als Einschränkungen von vielen verschiedenen Freiheitskonzeptionen auftreten: wie äußere Handlungsfreiheiten und Selbstbestimmung in der Persönlichkeitsentfaltung, innere Freiheit der noch weitgehend selbstbestimmten Identitätsbildung und der unbefangenen Nutzung von digitalen Diensten und Produkten. Zudem handelt es sich um mögliche ungerechtfertigte Ungleichbehandlungen durch falsche oder ungerechte Bildung von und Zuordnung zu Personenklassen oder individuellen Zuschreibungen. Damit verbunden sind beispielsweise Schäden wie Verletzung der Menschenwürde durch Behandlungen allein als bloßes Objekt, Stigmatisierungen, Stereotypisierungen, Rufschädigungen, Missbrauch von Informationsmacht und struktureller Überlegenheit, Konformitätszwänge durch Überwachungsdruck, Einschüchterungseffekte bzw. „chilling effects“, Enttäuschung von Vertraulichkeitserwartungen, die erhöhte Dauerhaftigkeit der Schädigungen durch die Permanenz der Datenverarbeitungen.[14] Sie werden hauptsächlich als Schädigungen auf individueller Ebene betrachtet, allerdings bleiben hier die Schädigungen auf kollektiver bzw. Makroebene bzw. für Gemeinwohlorientierungen und Gerechtigkeitsvorstellungen noch weitgehend unskizziert. Die Vielfalt der Schädigungen, die teils abstrakt wirken, teils sehr subjektiv wahrgenommen werden, teils noch nicht ausreichend erforscht sind, steht der Bestimmung von Schadensausmaßen oder Eintrittswahrscheinlichkeiten und der Bildung von Risikoklassen entgegen. So wäre z.B. die Bildung einer Stufe „unbedenklich“ oder „nicht-hohes“ Risiko immer noch mit großen Unsicherheiten für die Regulierenden verbunden.

Es liegt daher eher eine Situation der Risikoregulierung mit wissenschaftlich nicht konsolidiertem Wissen vor, bei der jedoch Hinweise auf hohe gesellschaftliche Schädigungspotentiale bestehen, etwa angedeutet durch einzelne Untersuchungen von Forschenden zu Diskriminierungsrisiken[15] und Manipulationsrisiken oder Skandale, die Journalist*innen aufgedeckt haben. Die vorliegende Situation wissenschaftlichen Wissens und des Schädigungspotentials deutet auf die Notwendigkeit der Anwendung des Vorsorgeprinzips hin. So fordert die Hochrangige Expertengruppe zu KI demensprechend auch die Anwendung des prinzipien-basierten Vorsorgeprinzips.[16] Die im Weißbuch vorgeschlagenen Anforderungen an KI-Anwendungen (S. 22-26) können als Umsetzung des prinzipien-orientierten Regulierungsansatzes interpretiert werden.

[13] Zu einer solchen Risikotypeneinteilung siehe z.B. van der Heijden (2019: 12f.) mit weiteren Nachweisen.
[14] Übersicht zu den potentiellen Schäden der Datenverarbeitung z.B. in Drackert (2014).
[15] Übersicht zu Beispielsfällen und den Forschungsarbeiten z.B. in Orwat (2019) (Stand Mitte 2019).
[16] Siehe AI HLEG (2019b: 38).

4 Notwendige Konkretisierung und Operationalisierungen von Risiken

Effektive Formen der Risikoregulierung bedürfen ein klares Risikoverständnis und eine eindeutige und justiziable Konkretisierung der Risiken, auf die sich die Risikoabschätzungen und das Risikomanagement der Risikoregulierung beziehen sollen.[17] Die erforderliche Effektivität der Risikoregulierung bemisst sich dabei an dem tatsächlichen Erreichen der „richtigen" Schutzziele bzw. der Vermeidung von Schädigungen, zu denen Regulierende Rechenschaft abgeben müssen.[18] Eindeutig konkretisierende, der Risikoregulierung zu Grunde liegende Kriterien und Prinzipien über das, was ein Risiko bzw. möglicher Schaden darstellt, dienen zur Vermeidung abiträrer Bewertungen. Zudem gewähren sie Rechtssicherheit bei der Risikoabschätzung oder der Einteilung von Anwendungen in Risikostufen, -graden oder -klassen bzw. bei der Beurteilung der Übereinstimmung mit Prinzipien oder Anforderungen und den daraus abgeleiteten Regulierungsmaßnahmen (z.B. Verbote, Moratorien, Zulassungsprüfungen). Ferner ermöglichen sie die Absicherung der Regulierenden gegenüber Einwänden und Klagen gegen vermeintlich falsche oder ungenaue Beurteilungen. Nicht zuletzt bilden sie die Grundlage für die Rechenschaft der Regulierenden über das Erreichen von Schutzzielen. Jedoch sind die mögliche Probleme der risiko-basierten Regulierung eine ungeeignete Auswahl und Priorisierung von Risiken sowie ungeeignete Interpretationen und Operationalisierungen von Schutzzielen im Regulierungsprozess, die im schlechtesten Falle zum Verfehlen der eigentlichen Schutzziele führen können.

Empfehlung: Wir empfehlen deshalb, die Risikokriterien noch weiter auszuarbeiten und zu konkretisieren. Dabei sollten auch verschiedene Konkretisierungskonzepte, d.h. neben Risikoklassen auch Indikatoren, Prinzipien, Anforderungen u.a., in Betracht gezogen werden. Eine besondere Herausforderung kann darin bestehen, die verschiedenen Risikokonzeptionen unterschiedlicher disziplinärer Herkunft zu integrieren. Hier scheinen Risikokonzeptionen aus den Bereichen der Sicherheit und Konzeptionen zur Gefährdung von Grundrechten nicht ohne großen Forschungsaufwand integrierbar zu sein.

Es scheint, als ob durch die mittlerweile zahlreichen internationalen ethischen Richtlinien[19] Anhaltspunkte für die Konkretisierung der Risikobewertung existieren. Doch gehen diese teilweise nicht auf europäische oder nationale Besonderheiten ein. Dies betrifft insbesondere die historische Entwicklung bestimmter Verfassungswerte und deren Stellung im Wertegefüge einer Gesellschaft. Zudem unterscheiden sie sich so stark hinsichtlich beinhalteter Werte und Prinzipien, dass die jeweilige Auswahl der berücksichtigten oder unberücksichtigten Werte und Prinzipien nicht nachvollzogen werden kann.[20] Zudem spielen sie eine bisher nicht geklärte Rolle bei der Regulierung

[17] So auch im Weißbuch, Europäische Kommission (2020: 20).

[18] Z.B. nach Black (2010b: 332-336).

[19] Übersichten z.B. in Jobin et al. (2019), Hagendorff (2019) oder Fjeld et al. (2020).

[20] Ausnahmen sind die Richtlinien, die systematisch mit den Katalogen der Menschenrechte arbeiten. Allerdings sind Menschenrechte und Verfassungswerte nicht immer deckungsgleich.

von KI und AES. Denn teilweise ist unklar, ob die Richtlinien überhaupt in ihren Anforderungen über bestehende rechtliche Vorgaben hinausgehen oder diese sogar unterschreiten und damit lediglich einem „ethical washing“ dienen würden.

Auch der bestehende sekundärrechtliche Rahmen, insbesondere die im nationalen Kontext besonders relevante Datenschutzgrundverordnung (DSGVO) und das Allgemeine Gleichbehandlungsgesetz (AGG), liefert nur wenige konkrete Orientierungspunkte, da er selbst mit zu vielen Auslegungsspielräumen und Unklarheiten versehen ist. Die abstrakten Sätze der Menschen- und Verfassungsrechte sowie die Rechtsprechung dazu liefern schon eher wertvolle Anhaltspunkte der Interpretation und Konkretisierung von Grundrechten und -werten, die in ihren idealen Schutzzielen nachvollzogen werden können. Ebenso liefern die umfassenden ethischen Forschungen zu KI und AES Anhaltspunkte: Allerdings besteht hierbei die Herausforderung, mit der Pluralität ihrer Ansätze und Weltsichten umzugehen. Der Bezug zu den konkreten Auslegungsnotwendigkeiten und die Konkretisierung für die Grundlage der Risikoregulierung von KI und AES steht bei beiden Ansatzpunkten noch aus.

Anhand einiger Beispiele von betroffenen Grundrechten, -werten und Schutzzielen, die keinen Anspruch auf Vollständigkeit erheben, kann verdeutlicht werden, dass die zur Risikoregulierung notwendige Operationalisierung bzw. Konkretisierung ein aufwendiges Vorgehen von Wissenschaft und Politik und letztlich politisch-normative Abwägungen und Entscheidungen erfordern.

So ist zum Beispiel der Schutz der *Menschenwürde* als gesellschaftlich höchster Wert anerkannt und die Gefährdung durch KI und AES skizziert.[21] Auf KI basierende AES-Anwendungen können die Menschenwürde verletzen. Die meisten Anwendungen des maschinellen Lernens verwenden Daten über Personengruppen, um Modelle für die algorithmische Entscheidungsfindung und Prognose zu bilden. In vielen Fällen geschieht dies mit Hilfe von Korrelationen und der Konstruktion von Persönlichkeitstypen, wie bei Anwendungen für die Bestimmung von Kreditwürdigkeit, Rückfallwahrscheinlichkeit von Straftätern oder der Passfähigkeit zu Personalstellen. Solche Datenanalysen nehmen nicht das Individuum mit seinen individuellen Subjektqualitäten in Betracht, sondern sortieren Personen zu vorfabrizierten Personenkategorien anhand von einer oder einigen Variablen, die bei den Personen wiedererkannt werden. Die Verletzung der Menschenwürde kann daraus resultieren, dass Menschen als bloßes Objekt, Instrument oder Mittel für die Zwecke anderer behandelt werden. Jedoch sind Konzepte, die den Schutz der Menschenwürde konkretisieren – wie die Objektformel und das Instrumentalisierungsverbot, Voraussetzung der Zustimmung, Vorhandensein von Erniedrigung oder Achtung der individuellen Subjektqualitäten – noch nicht ausreichend auf KI und AES bezogen worden, als das daraus tragfähige Kriterien oder Prinzipien für Abschätzungen der Risiken für die Menschenwürde abgeleitet werden könnten.[22]

---

[21] Siehe z.B. die Gutachten der Hochrangigen Expertenkommission für KI, siehe AI HLEG (2019a: 10), der Datenethikkommission (2019: 43) oder des Europarats durch Yeung (2019: 33-34).

[22] Beispielsweise können KI-Systeme auch zur Erzeugung und Analyse von umfassenden Personenprofilen genutzt werden. Dabei geht zwar das Bundesverfassungsgericht davon aus, dass die Menschenwürde und Selbstbestimmung u.a. dann gefährdet ist, wenn umfassende Profile über die Betroffenen erzeugt und verwendet

Dies zeigt auch das Beispiel der *informationellen Selbstbestimmung*[23], die im Datenschutzrecht weitgehend als Kontrolle über personenbezogenen Daten durch die Betroffenen umgesetzt worden sein sollte. Eine andere, ebenso legitime Interpretation setzt noch direkter am Schutz der Menschenwürde und der freien Entfaltung der Persönlichkeit an und fordert die Wahrung von weitgehend selbstbestimmten Handlungsmöglichkeiten sowie die freie, noch als eigene empfundenen Identitätsbildung, die Wahrung einer unbefangenen Nutzung und das Vermeiden von Abschreckungseffekten.[24] Vor allem aus letzterer Interpretation kann eine stärkere Regulierung von algorithmen- und datenbasierten Entscheidungen bzw. AES begründet werden. Die regulatorische Risikobestimmung würde dann nicht nur auf einen möglichen Kontrollverlust an personenbezogenen Daten abzielen, sondern müsste auch stärker auf die Möglichkeit von Einschränkung der freien Selbstbestimmung, des Rechts auf Selbstdarstellung und der eigenen Identitätsbildung abstellen und diese anhand von Kriterien und Prinzipien konkretisieren. Daneben existieren noch vielfältige Konzeptionen der Privatheit, teilweise mit Elementen der informationellen Selbstbestimmung überlappend.[25] Im Recht hat sich vor allem eine Konzeption der Privatheit etabliert, die vor allem einen besonders schützenswerten Kernbereich privater Lebensgestaltung (wie die Wohnung) vorsieht. Diese kann u.U. dann zu kurz greifen, wenn Datenverarbeitungen bzw. Anwendungen der KI und AES außerhalb dieser engen Privatsphäre stattfinden und für Entscheidungen herangezogen werden, die Konsequenzen für die Lebensführung und Entwicklungschancen der Betroffenen haben. Aus dieser Sicht wäre die Risikokonkretisierungen beispielsweise sinnvoller direkt an den KI- und AES-basierten Entscheidungen und ihren Konsequenzen anzusetzen.

Ein weiteres Beispiel ist *Antidiskriminierung und Gerechtigkeit*. Zwar zeigen sich bereits hinsichtlich des bestehenden Antidiskriminierungsrechts vielfältige Diskriminierungsrisiken durch AES und KI.[26] Es liegen jedoch viele Auslegungsspielräume und Unklarheiten im AGG vor[27], so dass eine Orientierung am AGG bei der Risikoabschätzung lediglich eine grobe Grundlage sein kann und Auslegungsspielräume oder Regelungslücken durch weitergehende Prinzipien und Kriterien geschlossen werden müssen. So genannte „Fairnessdefinitionen“ oder „Fairnessmaße“[28] wollen eine Quantifizierung dieser Risiken erreichen und teilweise über die rechtliche Situation hinausge-

werden (hier nach Britz (2010)), jedoch ist wenig konkret, ab wann ein Profil so „umfassend“ genannt werden kann, so dass die Menschenwürde tatsächlich beeinträchtigt ist. Auch im Gutachten der Datenethikkommission (2019), die den Schutz der Menschenwürde relativ detailliert auf KI und AES bezieht (S. 43), werden keine konkreten Kriterien für die Risikoabschätzung abgeleitet und in Verbindung mit den Vorschlägen zur risikoadäquaten Regulierung gebracht.

23 Gefordert in Bundesregierung (2018: 10, 16, 29 etc.) oder Council of Europe (2020: Appendix, Point B.2.1).

24 Vgl. Britz (2010).

25 Siehe für Übersichten z.B. Tavani (2007) oder Koops et al. (2016).

26 Siehe Barocas & Selbst (2016), Zuiderveen Borgesius (2018), Hacker (2018) oder Orwat (2019).

27 Siehe z.B. Orwat (2019: 107-114).

28 Siehe Übersichten z.B. in Schweighofer et al. (2018) oder Verma & Rubin (2018).

hen. Doch hinter den jeweiligen Konzeptionierungen stehen höchst unterschiedliche Weltanschauungen und Gerechtigkeitsvorstellungen[29], die offengelegt und einem gesellschaftlichen Diskurs zugeführt werden müssen. Zudem bilden viele der Maße lediglich Relationen zwischen Fehlerraten ab, sodass sie wie Maße von Restrisiken zu behandeln sind, die den Betroffenen zugemutet werden. Ferner zielen sie grundsätzlich auf die gerechtere Behandlung von Gruppen und berücksichtigen nicht die Einzelfallgerechtigkeit, d.h. die zu einer gerechten Behandlung von Personen notwendige Betrachtung des individuellen Falls bzw. der individuellen Person und ihrer jeweiligen Situation, die in der Regel aus Generalisierungs- und Effizienzgründen bei AES aufgegeben wird.[30] Möchte man Risiken von Anwendungen der KI und AES für gesellschaftliche Gerechtigkeit bestimmen, wären zuvor von der Gesellschaft eine Reihe von Punkten zu entscheiden. Wiederum ohne Anspruch auf Vollständigkeit könnten das beispielhaft folgende Fragen sein: Soll Verdienst bei der Verteilung schwerer wiegen als Bedürfnis? Welcher Kontext muss geschaffen werden, um Menschen zu einem erfüllten Leben zu befähigen? Wie soll die Bewertung von Prozess und Outcome gewichtet werden (Verfahrensgerechtigkeit oder soziale Gerechtigkeit)? In welchem Maße muss die „Lotterie der Natur“ (Rawls) ausgeglichen werden, um Chancengleichheit zu schaffen?

Auch das Beispiel des grundlegenden Prinzips der *Gemeinwohlorientierung* zeigt, dass diese zwar in der Entwicklung und Anwendung von KI eingefordert wird,[31] es jedoch noch weitgehend unklar ist, wie dieses Prinzip für eine Risikobeurteilung konkretisiert werden kann. Dabei stehen eine Reihe von möglichen Konkretisierungen zur Verfügung, wie beispielsweise über die adversen Ergebnisse für ganze Gruppierungen, die nicht durch den hauptsächlich auf das Individuum bezogenen Rechtsrahmen abgedeckt sind. Diese können auch durch Vertrauensverluste, Einschüchterungs- oder Abschreckungseffekte („chilling effects“) ausgelöst sein, die durch die abstrakte Ungewissheit über die (Weiter-) Verarbeitung von personenbezogenen Daten und über die tatsächlich bestimmenden Kriterien bei algorithmenbasierten Entscheidungen hervorgerufen werden können. Sie können einen Rückzug von der Nutzung von digitalen Diensten und Produkten durch bestimmte Bevölkerungsgruppen bewirken. Das kann zu sozialer Segregierung und Schäden an Prozessen der demokratischen Willensbildung führen, da digitale Dienste und Produkte zunehmend der Gemeinschaftsbildung und dem politischen Austausch dienen. Des Weiteren ermöglichen und beschleunigen KI und AES die Personalisierung und Individualisierung von Diensten und Produkten, die üblicherweise auf Solidarität bzw. Gemeinschaftsleistungen beruhende Praktiken verdrängen können. Dies betrifft das Angebot vieler öffentlicher bzw. kollektiver Güter, wie Gesundheit, Bildung, Personennahverkehr etc. Nicht zuletzt sollten Fragen der gesellschaftlichen Verteilung von Risiken sowie die Verteilung der Effizienzgewinne durch Automatisierungen mit KI und AES unter dem Gemeinwohlgesichtspunkt behandelt werden. Insgesamt zeigen adverse Effekte auf das Gemeinwohl, dass sich auch die Eigenschaften der Risiken ändern können, d. h. von punktuellen Risiken einzelner Systeme hin zu Risiken, die stärker systemischen Charakter haben.

---

[29] Siehe z.B. Mulligan et al. (2019).

[30] Grundlegend dazu Britz (2008), Gandy Jr. (2010) und auf KI und Algorithmen bezogen Binns (2020: 519).

[31] Vgl. Bundesregierung (2018: 7, 9, 10, 45, 47).

5 Normative Ambiguität

An Hand dieser wenigen Beispiele zeigt sich, dass unterschiedliche Interpretationen und Operationalisierungen von gesellschaftlichen Normen, Grundwerten und -rechten zu stark abweichenden Anforderungen, Kriterien und Prinzipien für die Risikoabschätzung führen. Zudem sind bei der Risikoregulierung von KI und AES meist genau die gesellschaftlichen Werte betroffen, die Gegenstand jahrzehntelanger, teils heftiger gesellschaftlicher Kontroversen waren und sind. Folglich herrscht eine hohe *normative Ambiguität*[32] vor und eine wesentliche Komponente für möglichst objektive Risikoabschätzungen fehlt, nämlich die klare und nachvollziehbare Bewertungsgrundlage.

Risikoregulierung von KI und AES ist dadurch *kein nach objektiven Standards strebender, wissenschaftlich fundierter Prozess*, der auch an diejenigen Einrichtungen delegiert werden könnte, die zu einem objektiv, wissenschaftlichen Prozess nachprüfbar fähig wären.[33] Stattdessen ist dieser vielmehr ein *politischer Prozess*, in dem es um die Abwägung zwischen den Realisierungen oder Einschränkungen verschiedener Grundrechte und -werte geht, wie Recht auf Leben, Menschenwürde, Selbstbestimmung, freie Meinungsbildung, Gerechtigkeit und weitere und in der Regel konfliktträchtig dazu stehende Werte wie vertragliche Privatautonomie und Streben nach Effizienz durch Automatisierung. Darüber hinaus haben KI- und AES-Anwendungen nicht nur Folgen für individualrechtlich fokussierte Grundrechtsprinzipien, sondern auch solche mit gesamtgesellschaftlichen Dimensionen, wie Gerechtigkeits-, Gemeinwohl- oder Gleichheitsvorstellungen, zu denen es üblicherweise öffentliche politische und ideologische Auseinandersetzungen gibt.

Empfehlung: Aufgaben der Risikoregulierung sollten nicht delegiert werden (etwa an Test- oder Zertifizierungseinrichtungen), solange Risikoabschätzungen und -bewertungen nicht auf politisch legitimierten, nach objektiven Standards strebenden Kriterien und Verfahren beruhen, die von außen untersuchbar, anfechtbar und verifizierbar sind.

6 Politische Prozesse für normative Entscheidungen

Hieraus folgt, dass bei der Risikoregulierung der KI und AES zahlreiche *normative Entscheidungen*[34] anfallen, die für eine effektive und von der Gesamtgesellschaft als legitim angesehene Regulierung zu berücksichtigen sind:

(1) Bereits die Auswahl und Priorisierung, welche Grundwerte bzw. Schutzgüter und damit Risiken in die Risikoregulierung aufgenommen oder weggelassen werden, ist eine normative Entscheidung und beeinflusst stark die konkrete Ausgestaltung der Prozesse.[35]

---

[32] Nach Renn (2008: 150-156).

[33] Im Beispiel der REACH-Regulierung erfolgt unter der Voraussetzung, dass eine objektive Risikoabschätzung möglich ist, die Risikoabschätzung durch private Akteure. Vgl. Fisher (2012: 421f.).

[34] Vgl. Cranor (1997) oder Baldwin & Black (2016).

[35] Vgl. Baldwin & Black (2016) und Ansell & Baur (2018).

(2) Die Festlegung und der Zuschnitt des Regulierungsobjekts, ob z.B. jedes einzelne System, nur Prototypen oder generische Anwendungsformen (wie z.B. „Profiling“ an sich) reguliert wird, können auch als normative Entscheidung angesehen werden, da daraus höchst unterschiedliche Regulierungsformen (z.B. ob als Marktzulassungsverfahren oder Selbstbeschränkung mit behördlichen Stichprobenkontrollen) erwachsen, die normsetzend in der Gesellschaft wirken.
(3) Ebenso ist, wie dargelegt, die Konkretisierung bzw. Operationalisierung von gesellschaftlichen Grundrechten und -werten ein normativer Entscheidungsprozess, sei es die Operationalisierung von Kriterien, Prinzipien, Indikatoren, quantitative Größen oder ähnlichem. Diese betreffen nicht nur die Regulierung von KI und AES und den Schutz der Rechte Einzelner, sondern beinhalten wichtige Grundsatzentscheidungen darüber, wie die Gesellschaft als Gesamtgefüge funktionieren und das Zusammenleben gestaltet werden soll (siehe z.B. hinsichtlich der Umsetzung von Gerechtigkeitsvorstellungen).
(4) Auch der Umgang mit unvermeidlichen Wertekonflikten erfordert normative Abwägungen über die Notwendigkeit der Eingriffe in bzw. Beschränkung der Grundrechte und -werte, das Fehlen milderer Alternativen oder die Verhältnismäßigkeit der Verletzung von Grundwerten und -rechten.
(5) Insbesondere erfordern die Bestimmung des „Restrisikos“ bzw. des Risikoausmaßes (wie z.B. Fehlerraten), die einer Gesellschaft aufgebürdet bzw. für akzeptabel gehalten werden sollen, normative Abwägungen.
(6) Nicht zuletzt ist die resultierende gesellschaftliche Verteilung der Risiken, vor allem in dem Sinne, ob bestimmte Bevölkerungsgruppen besonders schwer betroffen sind, Gegenstand normativer und politischer Entscheidungen. Insbesondere die Berücksichtigung der Risikoverteilung zeichnet gesellschaftliche oder ethische Risikoabschätzungen aus.[36]

Bei den normativen Entscheidungen kann die Wissenschaft nur Beratungsleistungen abgeben. Die normativen Entscheidungen der Risikoregulierung, einschließlich der erforderlichen Operationalisierung und Konkretisierung von Grundrechten und -werten für die Risikobestimmung und -bewertung, müssen in *politischen Prozessen der gesellschaftlichen Abwägung* stattfinden, die demokratisch legitimiert und kontrolliert werden müssen.

Empfehlung: Zu allen genannten Punkten der grundlegenden normativen Entscheidungen sind politische Prozesse zu empfehlen, die die zugrundeliegenden normativen Ambiguitäten adäquat adressieren und die notwendigen politischen Entscheidungen mit adäquaten Beteiligungsverfahren – vor allem mit Einschluss der von KI- und AES-Anwendungen Betroffenen – herbeiführen, bevor eine Risikoregulierung betrieben wird. Dabei sollte vermieden werden, dass normative Entscheidungen nicht an Akteure delegiert werden, die dazu nicht legitimiert sind.

[36] Siehe Grunwald (2019: 26-28) und Hansson (2018).

7 Einzelne Vorschläge zur Ausgestaltung der Risikoregulierung

7.1 Untersuchungen zu den gesellschaftlichen Dimensionen der Risiken

Von Forschenden oder Journalist*innen aufgedeckte oder gerichtlich festgestellte Fälle von algorithmen-basierten Diskriminierungen oder Datenschutzskandale mit Nutzungen von KI-Anwendungen (wie z.B. der Clearview-Fall) umreißen nur für einzelne Systeme oder Anwendungen jeweils die Anzahl der tatsächlich oder potentiell Betroffenen. Sie sagen nichts über die tatsächliche gesellschaftliche Häufigkeit der Nutzung bestimmter Anwendungen der KI und AES aus, geschweige denn über das gesamtgesellschaftliche Ausmaß der Risiken, die von diesen Anwendungen resultieren können.

Empfehlung: Es ist zu empfehlen, dass regelmäßige und systematische Erfassungen und Untersuchungen von unabhängigen Stellen zur Verbreitung der Nutzung von KI- und AES-Anwendungen in den Praktiken der Wirtschaft, Verwaltung, Justizsystem, Polizeiarbeit etc. sowie zu den jeweiligen Anwendungszwecken und den potentiell Betroffenen (Ausmaß und Zusammensetzung) durchgeführt werden, mit denen die Dimensionen bzw. der Umfang potentieller Risiken in und für die Gesellschaft abgeschätzt werden können.

7.2 Forschungen zu Operationalisierung von Grundrechten und -werten

Wie oben skizziert, stehen die Operationalisierungen und Konkretisierungen von Grundrechten und -werten noch weitgehend aus, weshalb sie auch (noch) nicht als klare Kriterien oder Prinzipien der Risikoregulierung genutzt werden können.

Empfehlung: Die Europäische Kommission sollte einen Forschungsprozess anstoßen, der die wissenschaftliche Fundierung der politischen Prozesse über die normativen Entscheidungen liefert. Hierfür sollten vor allem alle für eine Risikoregulierung notwendigen normativen Entscheidungen herausgearbeitet, verdeutlicht und in ihren Abwägungserfordernissen dargestellt und an legitime politische Entscheidungsverfahren weitergereicht werden. Dabei sollte der Forschungsprozess (a) die Pluralität in den Interpretationen von Grundrechten und -werten und den jeweiligen Stellenwert jeweiliger Interpretationen in der Gesellschaft verdeutlichen, (b) die unterschiedlichen Wertevorstellungen, die den verschiedensten bestehenden Konzepten zur Werteumsetzung und der Risikoregulierung teils verdeckt zu Grunde liegen, hervorholen und (c) die vielfältigen Möglichkeiten der Konkretisierungen von Grundrechten, -werten und gesamtgesellschaftlichen normativen Prinzipien liefern. Hierbei sollte sich die Forschung möglichst an den ursprünglichen Schutzidealen, die den Prinzipien, Grundwerten und -rechten zu Grunde liegen, orientieren und diese in der Vielfalt als handhabbare Kriterien, Prinzipien oder Indikatoren ableiten. Durch die Orientierung an

Schutzidealen soll verhindert werden, dass Abwägungen zwischen Werten und Einschränkungen des normativen Gehalts einzelner Werte bereits bei der Operationalisierung bzw. Konkretisierung einzelner Werte erfolgen.

7.3 Forschungen zur Strukturierung des Regulierungsobjekts und zum Verfahren der Risikoregulierung

Die im Weißbuch vorgeschlagene Vorab-Selektion von vermeintlich besonders risikoreichen Branchen bzw. Sektoren (S. 20-21 des Weißbuchs) erscheint unangemessen. So sind beispielsweise algorithmenbasierte Diskriminierungen auch bei Plattformunternehmen[37] aufgetreten, die man eher der Medienbranche zuordnen würde, die aber im Weißbuch noch nicht aufgeführt ist. Dies kann verdeutlichen, dass Branchen bzw. Sektoren keine geeigneten Strukturierungsmerkmale für die Risikobestimmung wären, was auch vor dem Hintergrund einer fortschreitenden Konvergenz von Branchen zu sehen ist.

Empfehlung: Falls eine Vorabauswahl von besonders risikoträchtigen Anwendungsbereichen überhaupt erforderlich sein sollte, wäre eine Strukturierung aus der Perspektive der Grundrechte und Grundwerte sinnvoll: Dies kann z.B. als die Bestimmung von Anwendungsbereichen nach dem Grad der Abhängigkeit von Ressourcen erfolgen, die für die freie Entfaltung der Persönlichkeit (z.B. Arbeit, Wohnen, Kredit, Infrastrukturzugänge) oder für die freie Meinungsbildung erforderlich sind. Ebenso können Anwendungsbereiche identifiziert werden, die besonders anfällig für Verletzungen der Menschenwürde oder die besonders diskriminierungsanfällig sind (z.B. Arbeit, Wohnen, bestimmte Dienstleistungen) oder die für die Herstellung gleichwertiger Lebensverhältnisse besonders relevant sind. Die Strukturierung bedarf jedoch grundlegender Erforschung zur graduell unterschiedlichen potentiellen Betroffenheit der Grundrechte sowie der normativen Prinzipien und deren Festlegungen in politischen Prozessen.

7.4 Ausbau der Anforderungen an KI

Die Kommission schlägt eine Reihe von Anforderungen an Anwendungen der KI vor (S. 22-26), die allerdings nur für Anwendungen mit „hohem“ Risiko gelten sollen. Sie können als Element der Risikoregulierung verstanden werden.

[37] Siehe beispielsweise den Fall der algorithmenbasierten Diskriminierung nach Geschlecht bei Stellenanzeigen über die Plattform des Unternehmens Facebook, bei denen die Bürgerrechtsorganisationen ACLU und weitere Klage erhoben haben und es zu einem Vergleich gekommen ist; siehe Orwat (2019: 39f.) und weitere Beispiele.

7.4.1 Begründung und Anwendungsbereich der Anforderungen

Die Begründung von Anforderungen bzw. Prinzipien lässt sich mit einem Verweis auf die Risikoregulierung des Datenschutzes[38] verdeutlichen. Auch in der DSGVO wirken neben einer Form der risiko-basierten Regulierung, vor allem in Form der Datenschutzfolgenabschätzung nach Art. 35 DSGVO und daraus resultierenden Maßnahmen, noch weiterhin Anforderungen an Datenverarbeitungen in Form der Datenschutzprinzipen insbesondere des Art. 5 DSGVO, die unabhängig von der Risikohöhe der Datenverarbeitung gelten. Hierbei ist die Kritik an den Regularien zur Datenschutzfolgenabschätzung zu beachten. So wird die Unklarheit bei den Richtlinien zur Festlegung der („hohen“) Risiken, die abgeschätzt werden sollen, sowie zu den anzuwendenden Verfahren kritisiert, ebenso die Verlagerung der Durchführung der Datenschutzfolgenabschätzung auf die Betreibenden selbst (im Sinne einer Selbstregulierung), die dadurch in Konflikten zu ihren ökonomischen Eigeninteressen kommen können.[39] Das Festhalten an den Datenschutzprinzipien scheint daher sinnvoll.

Beim vorgeschlagenen Ansatz der risiko-basierten Regulierung bestehen einige Gefahren von Regulierungsversagen. Diese können aus der ungeeigneten Auswahl und Interpretation der Grundrechte und -werte, auf die sich die risiko-basierte Regulierung fokussieren soll, der ungeeigneten Bestimmung der „Höhe“ bei „hohen“ Risiken oder aus der fortbestehende Uneinigkeit und Uneindeutigkeit der normativen Grundlagen und Konkretisierungen der Risikoregulierung resultieren.

Empfehlung: Es scheint daher sinnvoll, neben dem Element der risiko-basierten Regulierung, auch allgemein verbindliche Anforderungen bzw. Prinzipien des vorsorgenden Vorgehens zu entwickeln und anzuwenden, die für KI-Anwendungen gelten sollten, wie es im Weißbuch vorgeschlagen wird (S. 22-26). Dabei wäre zu prüfen, wie weit die Anforderungen und Prinzipien auch für Anwendungen der KI und AES mit „nicht-hohen“ Risiken sinnvoll sind und ausgebaut werden sollten. Dies gilt auch als Alternative, wenn der risiko-basierte Ansatz sich als wenig praktikabel oder ineffektiv zum Erreichen der Schutzziele erweist.

Weiterhin gilt es ein grundsätzliches Problem auszubalancieren: Einerseits ist mit dem risiko-basierten Ansatz eine Konzentration auf „hohe Risiken“ und damit eine regulatorische Vernachlässigung von „nicht-hohen“ Risiken verbunden. Andererseits will der Regulierungsvorschlag dem Anspruch einer KI-Entwicklung „unter uneingeschränkter Achtung der Werte und Rechte der europäischen Bürgerinnen und Bürger“[40] entsprechen und den risikounabhängigen Grundsätzen der Datenverarbeitung sowie den risikounabhängigen Ansprü-

[38] Vgl. Gellert (2016), der betont, dass Datenschutz auch als Risikoregulierung aufzufassen ist.
[39] Z.B. Macenaite (2017), Quelle (2018), Demetzou (2019) oder Gonçalves (2019).
[40] So im Weißbuch, Europäische Kommission (2020: 3).

chen an Grundrechte und Betroffenenrechte folgen. Aus letzterem könnte sich beispielsweise ergeben, dass Anforderungen und Prinzipien auch für Anwendungen von KI gelten, denen ein „nicht-hohes“ Risiko zugeschrieben wird.

7.4.2 Anforderungen an Bereitstellung von Informationen

Bei den Anforderungen an die KI ist die Forderung der *Transparenz* im Sinne der Informierung der Betroffenen zu überdenken (S. 23-24 des Weißbuchs). Fehlende Nachvollziehbarkeit ist nicht allein ein Problem der technischen Komplexität von Anwendungen der KI und AES, sondern die Intransparenz ergibt sich auch durch die Rechtsprechung zum Schutz von Betriebs- und Geschäftsgeheimnissen (siehe die so genannte „Schufa-Entscheidung“[41] oder die „Meister-Entscheidung“[42]), mit denen eine adäquate Erfüllung von Auskunftsansprüchen durch die Betroffenen verhindert wird. Ist regulatorisch die Nachvollziehbarkeit von Entscheidungen durch Betroffene gewünscht, wäre auch eine Revision der Rechtsprechung angeraten.

Informationen für die Betroffenen über lediglich das bloße Vorhandensein von AES-Anwendungen sind oft von geringem Wert. Wenn Informationen an die Betroffenen gehen sollen, wird den Betroffenen Verantwortung für Zustimmung zum Vorgehen und Selbstschutzmaßnahmen zugeschrieben, zu denen sie weitgehend nicht mehr in der Lage sind, da die teils langfristigen oder indirekten Konsequenzen kaum mehr durch sie abschätzbar sein dürften.[43]

Empfehlungen: Wenn Pflichten zur Informierung der Betroffenen eingerichtet werden sollen, dann nicht über das Vorhandensein, sondern vor allem über die den automatisierten Entscheidungen zugrundeliegenden Kriterien und den daraus folgenden möglichen Konsequenzen für die Betroffenen, einschließlich möglichen Ungleichbehandlungen oder Manipulationen.

7.4.3 Anforderung an menschliche Aufsicht

Die Anforderung zum Vorhandensein der *menschlichen Aufsicht* ist begrüßenswert (S. 25 des Weißbuchs). Insbesondere sind sie von Bedeutung, wenn bei stark automatisierten oder sogar autonomen Systemen die körperliche Unversehrtheit sichergestellt werden muss (z.B. autonom fahrende Autos oder medizinische Diagnosen). Forschungen zum so genannten „automation bias“ weisen jedoch darauf hin, dass sich tendenziell die menschliche Aufsicht stark an den automatisierten Empfehlun-

[41] Kennzeichen VI ZR 156/13, *Schufa*, BGH-Urteil vom 28. Januar 2014.

[42] Rechtssache C-415/10, *Meister*, Urteil des Europäischen Gerichtshofs vom 19. April 2012; siehe auch Hacker (2018: 1169f.).

[43] Näheres dazu z.B. in Orwat (2019: 106f.).

gen orientiert und in unnormalen Situationen hohe Anforderungen an die Expertise der menschlichen Aufsicht bestehen, wo die Situation und das Funktionieren des Systems verstanden und bewertet werden muss, um überhaupt noch korrigierend eingreifen zu können. Daraus können sich Qualifizierungsanforderungen an die menschliche Aufsicht ergeben, die regulatorisch einzufassen wären.

Die Anforderungen an die menschliche Aufsicht bei AES-Anwendungen sollte sich in erster Linie aus dem Bedarf zum Schutz der Grundrechte ableiten, insbesondere dem Schutz der Menschenwürde, der freien Entfaltung der Persönlichkeit und der Meinungsfreiheit, aus denen sich der Bedarf nach geeigneten Möglichkeiten der Zustimmung, der Selbstdarstellung und der Anfechtung oder Korrektur ergeben.

Empfehlung: Die menschliche Aufsicht sollte jederzeit den von automatisierten Entscheidungen Betroffenen erklären können, wie Entscheidungen zu Stande kommen, am besten auch bevor eine Entscheidung getroffen wird. Das betrifft beispielsweise die tatsächlich ausschlaggebenden Entscheidungskriterien oder die Gewichtungen zwischen ihnen und die daraus folgenden Konsequenzen. Kann eine menschliche Aufsicht das Zustandekommen einer Entscheidung nicht jederzeit erklären, sollte die Anwendung eines AES in Anwendungsbereichen, die für die freie Entfaltung der Persönlichkeit und für die freie Meinungsbildung essentiell sind, untersagt werden.

Empfehlung: Grundsätzlich wäre zu überlegen, ob die Entscheidungsprozesse und kommunikativen Prozesse zwischen Betreibenden und Betroffenen detaillierter durch konkrete Anforderungen an die Anwendungen von KI und AES geregelt werden könnten. Dabei könnte für besonders grundrechtsrelevante Anwendungsbereiche festgelegt werden, welche Kriterien und Verfahren bei der Entscheidungsfindung in AES verwendet werden dürfen. Dies umfasst z.B. nur Kriterien die in einem wissenschaftlich robusten kausalen Zusammenhang zum Bewertungsgegenstand (wie z.B. Kreditwürdigkeit, Bedürftigkeit) stehen. Die kommunikativen Prozesse sollten so gestaltet sein, dass Betroffene eine Chance haben, ihr Bild von sich selbst in Bewertungen und Entscheidungen adäquat einzubringen.[44]

7.4.4 Besondere Anforderungen - sofortige Moratorien

Das Weißbuch eröffnet die regulatorische Möglichkeit, *besondere Anforderungen* an bestimmte Anwendungen der KI zu stellen und schlägt diese für biometrische Fernidentifikation vor (S. 25-

[44] Näheres in Orwat (2019: 123-125).

26). Darüber hinaus gibt es eine Reihe von KI- und AES-Anwendungen, die ebenso schwerwiegende Risiken für Grundrechte und Grundwerte aufweisen, dass sofortige regulatorische Eingriffe gerechtfertigt erscheinen. Die besonders hohen Risiken resultieren vor allem daraus, dass diese (1) oft mit umfassenden Personenprofilen und Personenkategorisierungen via Proxies arbeiten, (2) oft verdeckt bzw. im Geheimen operieren und (3) oft die Betroffenen keine Möglichkeit haben, die automatisierten Schlussfolgerungen und Konsequenzen zu erkennen, sie zu hinterfragen oder dagegen vorzugehen. (4) Oft ist dabei unklar, wie die Schlussfolgerungen genutzt, weiterverwendet oder weitergegeben werden. (5) Zudem beruhen die Schlussfolgerungen auf dem automatisierten Abgleich mit Kriterien, die aus Datensätzen über andere Menschen gebildet werden, und nehmen nicht das Individuum und seine jeweilige Situation in Betracht.

Moratorien sollten vor allem für die privaten und öffentlichen Anwendungsbereiche ausgesprochen werden, bei denen die Betroffenen in einem starken Abhängigkeitsverhältnis zum erbrachten Dienst, dem Produkt oder der Position stehen (z.B. bei Grenzkontrollen, bei Sozialleistungen, im Personalwesen, auf dem Wohnungsmarkt oder bei der Kreditvergabe). Zudem sollte dies ebenso Anwendungsbereiche betreffen, bei denen die Behandlung von Personen mit der Möglichkeit verbunden sind, dass nicht die Subjektqualitäten des Individuum geachtet werden und es zu Erniedrigungen kommen kann, die die Menschenwürde verletzen (wie z.B. Verweigerung der Einreise, Verweigerung von Sozialleistungen, Erniedrigung von ganzen Bevölkerungsgruppen auf dem Arbeits-, Wohnungs- oder Kreditmarkt oder bei der Zulassung zu Studienplätzen). Das gleiche gilt für Anwendungsbereiche, bei denen es zu gravierenden Einschüchterungs- und Abschreckungseffekten kommen kann, die Menschen davon abhalten können, öffentliche Plätze oder Dienste und Produkte, die der Gemeinschaftsbildung, der freien Persönlichkeitsentfaltung sowie der politischen Meinungsbildung und des politischen Austauschs dienen, zu nutzen. Ganze Bevölkerungsgruppen können sogar davon abgehalten werden, privatwirtschaftlich bereit gestellte Dienste oder Produkte zu nutzen, die der freien Entfaltung ihrer Persönlichkeit dienen und per Chancengleichheit verteilt werden sollten (z.B. auf dem Arbeits-, Wohnungs- oder Kreditmarkt).

Empfehlung: Für KI-Anwendungen, die der Personenerkennung bzw. -identifizierung und der automatisierten Ableitung von Persönlichkeitsmerkmalen dienen, sollten Moratorien verhängt werden.[45] Die Moratorien sollten so lange bestehen bis (1) die Auswirkungen auf Grundrechte und -werte in unabhängigen Untersuchungen umfassend und klar ermittelt sind, (2) bis die erforderlichen Verhältnismäßigkeitsprüfungen und gesellschaftlichen Abwägungen zwischen Nutzen und Risiken in Form von Einschränkungen der Grundrechte

[45] Das sollte über die im Weißbuch (S. 25-26) genannten biometrischen Fernidentifikationsverfahren zur Authentifizierung und Identifizierung von Personen hinausgehen und auch diejenigen Anwendungen umfassen, die der Analyse von Persönlichkeitsmerkmalen, wie z.B. Kreditwürdigkeit, Wert als Kund*in, Leistungsfähigkeit oder Passfähigkeit zum Unternehmen, dienen und dabei nicht nur die Identität feststellen, sondern auch Charaktereigenschaften, emotionale oder psychologische Zustände (psychometrische Analysen) teilweise auf Basis scheinbar trivialer Daten ermitteln. Zu den diesbezüglichen Möglichkeiten der KI siehe z.B. Matz et al. (2019).

und vor allem die Verteilung von Risiken durch öffentliche politische Prozesse[46] entschieden sind, und so lange nicht (3) Regulierungsmechanismen eingerichtet sind, die eine konstante Einhaltung von Prinzipien, Kriterien oder sonstige Vorgaben auch für diejenigen Anwendungen sorgen, die mit einem konstanten Training und Anpassungen von Entscheidungsregeln arbeiten.[47]

Ebenso sollten Moratorien für KI-Anwendungen verhängt werden, die der Deanonymisierung von Daten und der Reidentifizierung von Personen dienen. Die wie auch immer gearteten potentiellen Nutzen scheinen dabei in keinem Verhältnis zu den Verletzungen der Prinzipien des Datenschutzes zu stehen.

7.5 Reform des bestehenden Rechtsrahmens - Antidiskriminierungsrecht

Die Kommission schlägt in ihrem Weißbuch die Überprüfung und mögliche Anpassung des Rechtsrahmens vor (S. 15-17). Wir möchten die Gelegenheit nutzen, um auf Verbesserungsbedarf im nationalen Antidiskriminierungsrecht, vor allem dem Allgemeinen Gleichbehandlungsgesetz (AGG) hinzuweisen, das die Europäischen Antidiskriminierungsrichtlinien umsetzt.

Das AGG deckt auch Diskriminierungen mit Hilfe von algorithmischen Entscheidungssystemen ab. Aber das grundsätzliche Problem, dass ein diskriminierungsrechtliches Vorgehen von der Wahrnehmung einer Ungleichbehandlung durch die betroffene Person und ihrer ersten Initiative zu rechtlichen Schritten ausgehen muss, wird bei Differenzierungsformen mit Anwendungen von KI und AES erschwert. Denn die Differenzierungsentscheidungen erfolgen oft personalisiert und Vergleiche zur Feststellung einer Schlechterstellung sind schwierig. Hier ist die Wahrnehmung einer Ungleichbehandlung sowie die Beibringung der notwendigen Indizien durch die Betroffenen kaum oder gar nicht zu leisten.[48] Das im Weißbuch beschriebene Problem der mangelnden Transparenz und der Schwierigkeit, Schadensersatz einzufordern (S. 16), gilt vor allem auch im Antidiskriminierungsbereich.

Empfehlung: Aus diesen Gründen wären die Stärkungen der Antidiskriminierungsstellen mit Kompetenzen und Befugnissen erforderlich, die für die Erfüllung ihres gesetzlichen Schutzauftrags notwendig sind. Dazu gehören Verbesserungen der Nachweismöglichkeiten und das Verbandsklagerecht, mit denen anstelle der Betroffenen gegen Diskriminierungen vorgegangen werden kann. Auch Zugangsmöglichkeiten zu Dokumentationen, die nach der DSGVO erstellt werden müssen, und zu den Aufzeichnungen und Daten, die im Weißbuch für Anwendungen der KI vorgeschlagen werden (S. 23), sollten dazugehören.

---

[46] Im Sinne der breit angelegten öffentlichen Debatte, die auf S. 26 des Weißbuchs genannt wird.

[47] Auf diese besondere Eigenschaft wird auch im Weißbuch eingegangen, Europäische Kommission (2020: 16).

[48] Vgl. Hacker (2018: 1167-1170) oder Orwat (2019: 107-109).

8 Literatur

AI HLEG (2019a): Ethics Guidelines for Trustworthy AI; Brussels: Independent High-Level Expert Group on Artificial Intelligence (AI HLEG), report published by the European Commission.

AI HLEG (2019b): Policy and Investment Recommendations for Trustworthy AI; Brussels: Independent High-Level Expert Group on Artificial Intelligence (AI HLEG), report published by the European Commission.

Alemanno, Alberto (2016): Risk and Regulation, in: Burgess, Adam, Alberto Alemanno und Jens O. Zinn (Hrsg.): Routledge Handbook of Risk Studies; London, New York: Routlege, S. 191-203.

Ansell, Christopher und Patrick Baur (2018): Explaining Trends in Risk Governance: How Problem Definitions Underpin Risk Regimes, in: Risk, Hazards and Crisis in Public Policy, 9. Jg., H. 4, S. 397-430.

Baldwin, Robert und Julia Black (2016): Driving Priorities in Risk-based Regulation: What's the Problem?, in: Journal of Law and Society, 43. Jg., H. 4, S. 565-595.

Barocas, Solon und Andrew D Selbst (2016): Big data's disparate impact, in: California Law Review, 104. Jg., S. 671-732.

Binns, Reuben (2020): On the Apparent Conflict Between Individual and Group Fairness, FAT* ’20, January 27–30, 2020, Barcelona, Spain.

Black, Julia (2010a): Risk-based Regulation: Choices, Practices and Lessons Being Learnt, in: OECD (Hrsg.): Risk and Regulatory Policy: Improving the Governance of Risk; Paris: Organisation for Economic Co-Operation and Development (OECD), S. 185-236.

Black, Julia (2010b): The Role of Risk in Regulatory Processes, in: Baldwin, Robert, Martin Cave und Martin Lodge (Hrsg.): The Oxford Handbook of Regulation; Oxford: Oxford University Press, S. 302-348.

Britz, Gabriele (2008): Einzelfallgerechtigkeit versus Generalisierung. Verfassungsrechtliche Grenzen statistischer Diskriminierung; Tübingen: Mohr Siebeck.

Britz, Gabriele (2010): Informationelle Selbstbestimmung zwischen rechtswissenschaftlicher Grundsatzkritik und Beharren des Bundesverfassungsgerichts, in: Hoffmann-Riem, Wolfgang (Hrsg.): Offene Rechtswissenschaft; Tübingen: Mohr Siebeck, S. 561-596.

Bundesregierung (2018): Strategie Künstliche Intelligenz der Bundesregierung; Berlin: Bundesregierung.

Council of Europe, European Committee of Ministers (2020): Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems; Strasbourg: Council of Europe.

Cranor, Carl F (1997): The Normative Nature of Risk Assessment: Features and Possibilities, in: Risk: Health, Safety and Environment, 8. Jg., H. Spring, S. 123-136.

Datenethikkommission (2019): Gutachten der Datenethikkommission; Berlin: Datenethikkommission der Bundesregierung.

Demetzou, Katerina (2019): Data Protection Impact Assessment: A tool for accountability and the unclarified concept of ‘high risk’ in the General Data Protection Regulation, in: Computer Law and Security Review, 35. Jg., H. 6, S. 1-14.

Drackert, Stefan (2014): Die Risiken der Verarbeitung personenbezogener Daten. Eine Untersuchung zu den Grundlagen des Datenschutzrechts; Berlin: Dunker & Humblot.

Europäische Kommission (2020): Weißbuch zur Künstlichen Intelligenz - ein europäisches Konzept für Exzellenz und Vertrauen. COM(2020) 65 final, vom 19.2.2020; Brüssel: Europäische Kommission.

Fisher, Elizabeth (2012): Risk and Governance, in: Levi-Faur, David (Hrsg.): Oxford Handbook of Governance; Oxford: Oxford University Press, S. 417-428.

Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy und Madhulika Srikumar (2020): Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. Berkman Klein Center Research Publication No. 2020-1; Cambridge, MA: Berkman Klein Center for Internet & Society at Harvard University.

Gandy Jr., Oscar H. (2010): Engaging rational discrimination: exploring reasons for placing regulatory constraints on decision support systems, in: Ethics and Information Technology, 12. Jg., H. 1, S. 1-14.

Gellert, Raphael (2016): We Have Always Managed Risks in Data Protection Law: Understanding the Similarities and Differences between the Rights-Based and the Risk-Based Approaches to Data Protection, in: European Data Protection Law, 2. Jg., H. 4, S. 481-492.

Gonçalves, Maria Eduarda (2019): The risk-based approach under the new EU data protection regulation: a critical perspective, in: Journal of Risk Research, S. 1-14.

Grunwald, Armin (2019): Technology Assessment in Theory and Practice; London, New York: Routledge.

Hacker, Philipp (2018): Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination Under EU Law, in: Common Market Law Review, 55. Jg., H. 4, S. 1143-1185.

Hagendorff, Thilo (2019): The ethics of AI ethics - an evaluation of guidelines, in: Minds and Machines, 30. Jg., H. 1, S. 99-120.
Hansson, Sven Ove (2018): How to Perform an Ethical Risk Analysis (eRA), in: Risk Analysis, 38. Jg., H. 9, S. 1820-1829.
Hansson, Sven Ove und Terje Aven (2014): Is Risk Analysis Scientific?, in: Risk Analysis, 34. Jg., H. 7, S. 1173-1183.
Hood, Christopher, Henry Rothstein und Robert Baldwin (2001): The government of risk: Understanding risk regulation regimes; Oxford: Oxford University Press.
Jobin, Anna, Marcello Ienca und Effy Vayena (2019): The global landscape of AI ethics guidelines, in: Nature Machine Intelligence, 1. Jg., H. 9, S. 389-399.
Koops, Bert-Jaap, Bryce Clayton Newell, Tjerk Timan, Ivan Skorvanek, Tomislav Chokrevski und Masa Galic (2016): A typology of privacy, in: University of Pennsylvania Journal of International Law, 38. Jg., H. 2, S. 483-575.
Lloyd-Bostock, Sally M. und Bridget M. Hutter (2008): Reforming regulation of the medical profession: The risks of risk-based approaches, in: Health, Risk and Society, 10. Jg., H. 1, S. 69-83.
Macenaite, Milda (2017): The "Riskification" of European Data Protection Law through a two-fold Shift, in: European Journal of Risk Regulation, 8. Jg., H. 3, S. 506-540.
Matz, Sandra C, Ruth E Appel und Michal Kosinski (2019): Privacy in the age of psychological targeting, in: Current opinion in psychology
Mulligan, Deirdre K., Joshua A. Kroll, Nitin Kohli und Richmond Y. Wong (2019): This thing called fairness: Disciplinary confusion realizing a value in technology, in: Proceedings of the ACM on Human-Computer Interaction, 3. Jg., H. CSCW
Orwat, Carsten (2019): Diskriminierungsrisiken durch Verwendung von Algorithmen. Eine Studie, erstellt mit einer Zuwendung der Antidiskriminierungsstelle des Bundes; Berlin: Nomos, abrufbar unter: https://www.antidiskriminierungsstelle.de/SharedDocs/Downloads/DE/publikationen/Expertisen/Studie_Diskriminierungsrisiken_durch_Verwendung_von_Algorithmen.html?nn=6575434.
Quelle, Claudia (2018): Enhancing Compliance under the General Data Protection Regulation: The Risky Upshot of the Accountability- and Risk-based Approach, in: European Journal of Risk Regulation, 9. Jg., H. 3, S. 502-526.
Renn, Ortwin (2008): Risk Governance: Coping with Uncertainty in a Complex World; London, Sterling: Earthscan.
Schweighofer, Erich, Georg Borges, Matthias Grabmair, Daniel Krupka, Burkhard Schäfer, Christoph Sorge und Bernhard Waltl (2018): Technische und rechtliche Betrachtungen algorithmischer Entscheidungsverfahren. Gutachten der Fachgruppe Rechtsinformatik der Gesellschaft für Informatik e.V. im Auftrag des Sachverständigenrats für Verbraucherfragen; Berlin: Sachverständigenrat für Verbraucherfragen beim Bundesministerium der Justiz und für Verbraucherschutz.
Tavani, Herman T. (2007): Philosophical Theories of Privacy: Implications for an adequate online privacy policy, in: Metaphilosophy, 38. Jg., H. 1, S. 1-22.
van der Heijden, Jeroen (2019): Risk governance and risk-based regulation: A review of the international academic literature, State of the Art in Regulatory Governance Research Paper Series; Wellington: Victoria University of Wellington.
Verma, Sahil und Julia Rubin (2018): Fairness definitions explained, at: 2018 IEEE/ACM International Workshop on Software Fairness (FairWare), veröffentlicht von IEEE.
Yeung, Karen (2019): A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework; Strasbourg: Council of Europe, Committee of experts on human rights dimensions of automated data processing and different forms of artificial intelligence (MSI-AUT).
Zuiderveen Borgesius, Frederik (2018): Discrimination, artificial intelligence, and algorithmic decision-making; Strasbourg: Council of Europe, Directorate General of Democracy.