verbraucherzentrale
Nordrhein-Westfalen

Christian Wadephul

DATENSOUVERÄNITÄT DURCH DATENINTERMEDIÄRE?

Eine Kritik am Beispiel der Datenschutzgrundverordnung (DSGVO) und dem Versuch einer risikobasierten Regulierung von KI und automatisierten Entscheidungssystemen (AES)

Vortrag 12 der Reihe „Zu treuen Händen“ | Februar 2022

Eine Online-Vortragsreihe der Verbraucherzentrale NRW e. V.

VERBRAUCHER INFORMATIK

mit Unterstützung durch das
Institut für Verbraucherinformatik der Hochschule Bonn-Rhein-Sieg

Impressum

Verbraucherzentrale Nordrhein-Westfalen e. V.
Kompetenzzentrum Verbraucherforschung NRW
Mintropstraße 27
40215 Düsseldorf
zutreuenhaenden@verbraucherzentrale.nrw

Gefördert durch

Ministerium für Umwelt, Landwirtschaft,
Natur- und Verbraucherschutz
des Landes Nordrhein-Westfalen

ORIGINALBEITRAG

Verbraucherzentrale NRW, Düsseldorf 2022

AUTOR

Christian Wadephul ist Technikphilosoph und hat bis 2021 als Wissenschaftlicher Mitarbeiter am Institut für Technikfolgenabschätzung und Systemanalyse (ITAS) des Karlsruher Instituts für Technologie (KIT) gearbeitet – u.a. im Arbeitskreis Ethik des BMBF-Projekts „Begleitforschung Big Data (ABIDA – Assessing Big Data)“, in den ITA-Projekten „Personalisierte, adaptive kooperative Systeme für automatisierte Fahrzeuge“ (PAKoS) sowie „Abklärung des Verdachts aufsteigenden Bewusstseins in der Künstlichen Intelligenz (KI-Bewusstsein)“ und zuletzt in der Forschungsgruppe „Digitale Technologien und gesellschaftlicher Wandel (DigIT)“ in den Projekten „Governance von und durch Algorithmen (GOAL)“ und „Gesellschaftliches Vertrauen in lernende Systeme (GVLS)“.

DOKUMENTATION „ZU TREUEN HÄNDEN?“
Alle Videos und Paper der Vortragsreihe finden Sie unter
https://www.verbraucherforschung.nrw/zu-treuen-haenden-tagungsreihe-datenintermediaere-datentreuhaender-60831

INHALT

I. ABSTRACT

Obwohl der Entwurf eines Daten-Governance-Gesetzes zur Datennutzung in der EU explizit Datentreuhandmodelle vorsieht, bleibt – gerade nach den vielen kritischen Stimmen in dieser Vortragsreihe (v. a. von Ingrid Schneider 2022 und Jörg Pohle 2022) – fraglich, ob Datenintermediäre überhaupt als Lösung für die Konflikte zwischen Datenschutz- und Datenverwertungsinteressen dienen sollten. Befriedigende Antworten auf folgende dringliche Fragen stehen weiterhin aus: Wie kann die Souveränität derjenigen, die Daten zur Verfügung stellen, gewahrt werden? Welche Regulierung ist notwendig, um informationelle Selbstbestimmung zwischen kommerzieller und einer am Gemeinwohl orientierten Nutzung zu sichern?

II. EINLEITUNG

Das Problem datenschutzrechtlicher Grauzonen ist für eine Regulierung der Datenökonomie insofern von zentraler Bedeutung, als ein erheblicher Angriff auf Grundrechte zu beobachten ist. Deshalb sind zusätzliche regulatorische Vorgaben zum geltenden Recht überfällig, denn der Datenschutz bleibt auch nach Einführung der DSGVO prekär – und das, obwohl sich nicht nur diesem Gesetz, sondern auch dem kürzlich veröffentlichten Verordnungsentwurf zur Regulierung der künstlichen Intelligenz der EU-Kommission Anknüpfungspunkte für ein Recht auf Erklärung bezüglich der investierten (Software-)Verfahren („Algorithmen“) entnehmen lassen. Als andere Seite der ubiquitären Datenverarbeitung enthalten auch die neuartigen (Software-)Verfahren, die sich unter dem Oberbegriff „künstliche Intelligenz“ (KI) zusammenfassen lassen, durch die zugrundeliegenden Algorithmen diejenigen Regelsysteme, die entscheidend dafür sein können, welche Schlussfolgerungen aus Datenverarbeitungen gezogen und welche Bewertungen und Entscheidungen über Personen getroffen werden. Software und ihre Algorithmen dienen dementsprechend nicht mehr allein der Automatisierung der Datenverarbeitung sowie der Informations- und Wissensgenerierung, sondern zunehmend auch der Automatisierung von Entscheidungen.

Mittlerweile finden sich zudem zahlreiche Anhaltspunkte dafür, dass existierende oder künftige Anwendungen von KI und automatisierten Entscheidungssystemen (AES) hohe Risiken haben[1], wie konkrete Beeinträchtigungen von Menschen- und Verfassungsrechten und weiteren gesellschaftlichen Grundwerten, wie Demokratie und Rechtsstaatlichkeit.[2] Zum (auch präventiven) Schutz der Grundrechte und -werte sind zahlreiche Vorschläge zur Regulierung von KI und AES unterbreitet worden, wobei risikobasierte Ansätze überwiegen – etwa im Weißbuch der Europäischen Kommission (2020), ebenso AI HLEG (2019a und 2019b), Datenethikkommission (2019) oder Council of Europe (2020).[3] Dabei stellt die risikobasierte Regulierung nur eine von vielen Formen der Risikoregulierung dar, mit der Ressourcen der Regulierungsbehörde geschont werden sollen, indem sie sich auf diejenigen Regulierungsobjekte fokussiert, denen ein hohes Risiko zugeschrieben wird. Hierbei drängt sich jedoch die Frage auf, ob der risikobasierte Ansatz – im Unterschied etwa zum rechtsbasierten Ansatz (vgl. Or-

[1] Vgl. für Literaturhinweise Orwat et al. 2022, 255, Fn. 2.

[2] Vgl. hierzu Orwat et al. 2020 sowie Orwat et al. 2022, 258ff.

[3] Vgl. für weitere Beispiele Orwat et al. 2022, 255, Fn. 4-7.

wat et al. 2022, 264) – die betroffenen Schutzgüter erfassen kann und welche normativen Entscheidungen über gefährdete Grundrechte und akzeptable Risiken an welcher Stelle getroffen werden (sollten).

Im Folgenden werden die Herausforderungen für eine Risikoregulierung von KI und AES dargestellt, die sich vor allem in der normativen Ambiguität bei der Risikobestimmung und -bewertung zeigt. Es sind umfangreiche politische Prozesse erforderlich, bevor eine Risikoregulierung eingerichtet und betrieben werden kann. Deshalb stehen am Ende sechs konkrete Vorschläge zum Ausbau der Risikoregulierung, um der Gefahr des Regulierungsversagens zu entgehen. So scheint es – auch und vor allem zur Stärkung des Datenschutzes – notwendig zu sein, neben dem Element der risikobasierten Regulierung, auch allgemein verbindliche Anforderungen beziehungsweise Prinzipien – etwa das Vorsorgeprinzip (vgl. Orwat et al. 2022, 275ff.) – zu entwickeln und anzuwenden, die für KI- und AES-Anwendungen gelten sollten.

III. DATENSCHUTZGRUNDVERORDNUNG (DSGVO) – ANSPRUCH UND UMSETZUNG

Am 25. Mai 2018 trat in der EU eine weitreichende Reform des Datenschutzrechts in Kraft: die Datenschutzgrundverordnung (DSGVO). Deren Ziel ist es, die Daten von EU-Bürger:innen zu schützen und ihnen die Souveränität darüber zu geben, ob und wie ihre Daten genutzt werden. So soll eine Balance zwischen gesellschaftspolitischen und ökonomischen Widersprüchen geschaffen werden. Insbesondere der Schutz personenbezogener Daten und eine grundsätzliche Datenminimierung stehen im Konflikt mit der gleichzeitigen Förderung technischer Innovationen der Datenökonomie (Big Data, ML und KI). Dabei sollte eine echte digitale Souveränität vor allem dem Schutz vor digitaler Fremdbestimmung, Manipulation und Diskriminierung dienen. Dies wird gesetzlich etwa durch Betroffenenrechte (wie das Datenverarbeitungsverbot mit Erlaubnisvorbehalt) sowie Auskunftsrechte (insbesondere über Erhebung, Verknüpfung und Analyse personenbezogener Daten, den Zweck der Datenverarbeitung, Empfänger der Daten, Speicherdauer sowie das Recht auf Datenlöschung) zugesichert.

Trotz hoher Geldstrafen bei Nichteinhaltung der DSGVO haben viele Unternehmen jedoch immer noch nicht ausreichende Maßnahmen getroffen, um Daten rechtskonform zu verarbeiten.[4] Daten jedoch, die nicht personenbeziehbar sind wie etwa anonymisierte Statistikdaten, fallen nicht unter den Schutz der DSGVO.

Personenbezogene beziehungsweise -beziehbare Daten sind laut DSGVO Art. 4(1) „alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person [...] beziehen; als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Onlinekennung oder zu einem oder mehreren besonderen Merkmalen [...] identifiziert werden kann.“ Also insbesondere Name, Alter, Familienstand, Geburtsdatum; Anschrift, Telefonnummer, E-Mail-Adresse;

[4] So wurde etwa die Wohnungsgesellschaft „Deutsche Wohnen“ im Jahr 2019 zu einer Geldstrafe in Höhe von 14 Millionen Euro verurteilt, weil das Archivsystem keine Löschmöglichkeit besaß. Vgl. für eine stetig wachsende Auflistung weltweiter Verurteilungen wegen Verstößen gegen den Datenschutz das DSGVO-Bußgeld-Portal (Holzhofer o. D.).

Konto-, Kreditkartennummer; Kraftfahrzeugnummer, Kfz-Kennzeichen; Personalausweis-, Sozialversicherungsnummer; genetische Daten und Krankendaten; Werturteile wie zum Beispiel Zeugnisse; Fotos; aber auch IP-Adressen.

Einige dieser Daten werden auch weiterhin verarbeitet, wie etwa die Ergebnisse folgender zweier Studien zur Umsetzung der DSGVO zeigen.

1. ERGEBNISSE ZWEIER STUDIEN ZUR UMSETZUNG DER DSGVO

Dorfleitner und Hornuf (2018) analysieren die Datenschutzerklärungen deutscher FinTech-Unternehmen nach Einführung der DSGVO.[5] Bezüglich einer notwendigen Anpassung der Datenschutzerklärungen zeigt sich, dass in den meisten Fällen weiterhin Daten wie die E-Mail-Adresse (77 %), der Name (74 %), die Anschrift (55 %) und die Telefonnummer (46 %) verarbeitet werden. Darauf folgen die IP-Adressen (34 %), die Sammelkategorie Sonstige (31 %), Angaben zum Alter (30 %), Konto-und Zahlungsdaten (28 %) und sogar Passwörter (19 %). Obwohl auch IP-Adressen zu den personenbeziehbaren Daten gehören, werden diese bei den meisten Unternehmen (93 %) weiterhin verarbeitet. Als Gründe werden überwiegend Webtracking-Dienste (78 %) angegeben. Aber auch sogenannte Social Plug-ins (26 %), Newsletter sowie Werbedienste. Auffällig dabei ist, dass kein Unternehmen explizit angibt, dass die IP-Adresse nicht verarbeitet wird. Häufig findet sich zudem kein expliziter Hinweis, sondern bloß ein indirekter, etwa durch einen Textbaustein zur Nutzung von Google Analytics (Dorfleitner und Hornuf 2018, 19f. und Abb. 2.20 und 2.21).

In einer weiteren Studie untersuchen Wiebe und Helmschrot (2019) die Umsetzung der DSGVO durch Onlinedienste[6] und stellen einleitend Folgendes klar: „[I]n der Vergangenheit [wurde] immer wieder von den Unternehmen [...] kritisiert, dass die Vorgaben der DSGVO zu hohe Anforderungen für sie bedeuten würden. [...] Zudem würden die [...] rechtlichen Pflichten der Unternehmen [...] die Innovationspotenziale datenbasierter Geschäftsmodelle hemmen oder gar inkompatibel mit den Geschäftsmodellen sein. Vor diesem Hintergrund scheint die Frage nach der *rechtstatsächlichen Umsetzung* der rechtlichen Anforderungen der DSGVO durch Unternehmen besonders drängend.“ (Wiebe und Helmschrot 2019, 12f.)

Bezüglich einer Umsetzung der DSGVO zu „Informationspflichten & Datenschutzerklärung“ (Thema 1) liefert die Studie das Ergebnis, dass 13 von 35 Diensten (= 37 %) die Vorgaben nicht oder nur unzureichend umsetzen. Zehn Dienste setzen die Vorgaben teilweise um. Nur neun Dienste treffen die rechtlichen Anforderungen und klären Verbraucher verständlich über die Datenverarbeitungsvorgänge. Bezüglich einer Umsetzung der DSGVO im „Umgang mit sensiblen Daten“ (Thema 4) zeigt sich ein noch düsteres Bild, denn 19 von 35 (= 54 %) setzen die Vorgaben im Umgang mit sensiblen Daten nicht oder nur unzureichend um. Lediglich zwei Dienste informieren überhaupt über die Verarbeitung sensibler Daten. Insgesamt ist die Situation beim Minderjährigenschutz (v. a. in sogenannten „sozialen“ Netzwerken) kritisch (Wiebe und Helmschrot 2019, 245f.).

Ähnlich kritisch steht es um das vermeintlich durch die DSGVO zugesicherte „Recht auf Erklärung“ bezüglich der investierten (Software-)Verfahren („Algorithmen“).

---

[5] Ein Gutachten, entstanden in dem vom BMBF geförderten Projekt „Assessing Big Data (ABIDA)“, vgl. https://www.abida.de.

[6] Im Auftrag des Bundesministeriums der Justiz und für Verbraucherschutz (Referat VB3: Digitale Kundenbeziehungen, Datensouveränität).

2. RECHT AUF ERKLÄRUNG BEI AUTOMATISIERTEN ENTSCHEIDUNGEN?

Als andere Seite der ubiquitären Datenverarbeitung („Big Data“) enthalten auch die neuartigen (Software-)Verfahren, die sich unter dem Oberbegriff „künstliche Intelligenz“ (KI) zusammenfassen lassen, durch die zugrundeliegenden Algorithmen diejenigen Regelsysteme, die entscheidend dafür sein können, welche Schlussfolgerungen aus Datenverarbeitungen gezogen und welche Bewertungen und Entscheidungen über Personen getroffen werden. Software und ihre Algorithmen dienen dementsprechend nicht mehr allein der Automatisierung der Datenverarbeitung sowie der Informations- und Wissensgenerierung, sondern zunehmend auch zur Automatisierung von Entscheidungen.

Art. 22 (1) der DSGVO bietet auf den ersten Blick ein „Recht auf Erklärung“ bei automatisierten Entscheidungen: „Die betroffene Person hat das Recht, nicht einer ausschließlich auf einer automatisierten Verarbeitung – einschließlich Profiling – beruhenden Entscheidung unterworfen zu werden [...].“ Dies wird durch Art. 15 (1) DSGVO als Auskunftsrecht bestätigt: „Die betroffene Person hat das Recht bei Bestehen einer automatisierten Entscheidungsfindung einschließlich Profiling [...] auf aussagekräftige Informationen über die involvierte Logik sowie die Tragweite und die angestrebten Auswirkungen einer derartigen Verarbeitung [...]“.

Doch es lassen sich (mindestens) zwei Typen algorithmischer Erklärungen unterscheiden: Eine *globale* Erklärung liefert eine Erklärung der „Systemfunktionalität“ – etwa eines Machine-Learning-Algorithmus (z. B. die Funktion und Anzahl der Schichten bei neuronalen Netzen). Eine *lokale* Erklärung wäre jedoch die konkrete Begründung einer individuellen Entscheidung. Also die Antwort auf die Frage, welche Faktoren zu einer bestimmten Entscheidung beigetragen haben, die sich auf eine bestimmte Person ausgewirkt hat. Die DSGVO bietet (bloß) ein Recht auf globale Formen der Erläuterung von Systemfunktionalität statt einer lokalen Erklärung von konkreten Entscheidungen (vgl. Wachter, Mittelstadt und Floridi 2018).

Ein weiteres Problem von ML und KI ist der Trade-off zwischen der Performance und der Erklärbarkeit derartiger Systeme. Das junge Forschungsgebiet der Explainable AI (XAI) widmet sich diesem Problem und versucht sich an einer Standardisierung der Erklärbarkeit der „opaken“ Computersysteme. Aufgrund des zunehmenden Einsatzes von Cloud- und Blockchain-Technologien nehmen die Herausforderungen für eine adäquate Umsetzung der DSGVO zu.

IV. REGULIERUNG VON ALGORITHMEN, KI UND AUTOMATISIERTEN ENTSCHEIDUNGSSYSTEMEN (AES)

Die Datenverarbeitung geschieht mit Softwareverfahren und den ihnen zugrunde liegenden Algorithmen. Algorithmen enthalten diejenigen Regelsysteme, die entscheidend dafür sein können, welche Schlussfolgerungen aus Datenverarbeitungen und -analysen gezogen, welche Bewertungen, Vorhersagen und Entscheidungen über Personen getroffen werden. Sie dienen nicht mehr allein der Automatisierung der Datenverarbeitung sowie der Informations- und Wissensgenerierung, sondern zunehmend auch der Automatisierung von Entscheidungen. Insbesondere Verfahren der künstlichen Intelligenz (KI) und des Machine Learning (ML) bedeuten eine erhebliche Rechts-

unsicherheit und die Praxis zeigt zudem rechtliche Auslegungsprobleme und Grauzonen (etwa der DSGVO), die weitere Anpassungen nötig machen – insbesondere im Bereich Profiling, Scoring und automatisierte Entscheidungssysteme (AES).[7]

Der Vorschlag der Europäischen Kommission zur Regulierung von KI (Weißbuch 2020) zeigt, dass der bestehende Rechtsrahmen angesichts der neuen Herausforderungen überprüft und existierende Regelungen, zum Beispiel zur Produktsicherheit und -haftung, angepasst und erweitert werden sollten.[8] 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 (Weißbuch 2020, 15–17). Das verdeutlicht die Dringlichkeit und Bedeutung, mit der die Kommission den Regulierungsdarf sieht.

Der Vorschlag von Anforderungen an KI-Anwendungen (Weißbuch 2020, 22–26) soll allerdings nur für solche mit „hohem“ Risiko gelten. Eine ebenso risikobasierte Regulierung von KI und AES findet sich auch im Vorschlag des Council of Europe (2020) sowie in den Vorschlägen der Hochrangigen Expertengruppe (AI HLEG 2019a und 2019b) und der Datenethikkommission (2019)[9]. Doch kann ein rein risikobasierter Ansatz die betroffenen Schutzgüter erfassen? Welche normativen Entscheidungen über gefährdete Grundrechte und akzeptable Risiken werden (wo, von wem und warum) getroffen?

1. DATENSCHUTZ ALS RISIKOBASIERTE REGULIERUNG?

Bei einer rein risikobasierten Regulierung besteht die Gefahr von Regulierungsversagen. Diese können aus der ungeeigneten Auswahl und Interpretation der Grundrechte und -werte, auf die sich die risikobasierte 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.[10] Es existieren zwar unterschiedliche Ansätze der Risikoregulierung[11], aber es überwiegt vor allem die risikobasierte Regulierung (vgl. Orwat et al. 2022, 261-263).

In der DSGVO wird die risikobasierte Orientierung vor allem in der Datenschutzfolgenabschätzung nach Art. 35 DSGVO deutlich. Ein kritischer Punkt ist dabei die Unklarheit bei den Richtlinien zur Festlegung der („hohen“) Risiken, die abgeschätzt werden sollen, sowie zu den anzuwendenden Verfahren. Zudem wird eine Durchführung der Datenschutzfolgenabschätzung auf die Betreibenden selbst (im Sinne einer Selbstregulierung) abgewälzt, die dadurch in Konflikte zu ihren ökonomischen Eigeninteressen kommen können.

---

[7] An dieser Stelle zwei Literaturhinweise: Das Science and Technology Options Assessment (STOA) Komitee, ein Ausschuss des Europaparlaments, hat eine Studie veröffentlicht (Sartor und Lagioi 2020), die sich mit der Auswirkung der DSGVO auf KI-Anwendungen befasst. Haag und Risthaus (2022) fragen sich, ob bei einer Regulierung von KI die DSGVO als Leitbild dienen kann. Vgl. Haag und Risthaus 2022, 290-292 insb. zum „Regulierungsbedarf von KI“ sowie „KI und Datenschutz“.

[8] Vgl. zur staatlichen Aufsicht und den Strukturentscheidungen auf europäischer sowie nationaler Ebene Haag und Risthaus 2022, 306-317.

[9] Vgl. für weitere Beispiele Orwat et al. 2022, 255, Fn. 4-7.

[10] Vgl. Orwat et al. 2020 sowie Orwat et al. 2022, 256-258 für eine Kritik an einer rein risikobasierten Regulierung von künstlicher Intelligenz und automatisierten Entscheidungen.

[11] Vgl. etwa zum rechtebasierten Ansatz Orwat et al. 2022, S. 264 sowie zum pflichten- bzw. prinzipienbasierten Ansatz Haag und Risthaus 2022, S. 298-301.

Doch in der DSGVO wirken auch Anforderungen an Datenverarbeitungen in Form der Datenschutzprinzipen (insb. Art. 5 DSGVO), die unabhängig von der Risikohöhe der Anwendung gelten. Bereits dies zeigt, dass neben einer risikobasierten auch prinzipienbasierte Regulierungsansätze eine größere Rolle spielen müssen – etwa durch allgemein verbindliche Anforderungen beziehungsweise Prinzipien des vorsorgenden Vorgehens.[12] Eine zentrale Frage, die durch die aktuelle Risikoregulierung negativ beantwortet wird, ist, ob dies auch für KI-/AES-Anwendungen mit „nicht-hohen" Risiken gelten sollte.

2. RISIKEN UND POTENZIELLE SCHÄDEN DURCH AUTOMATISIERTE DATENVERARBEITUNG

Vor allem KI-Methoden und automatisierte Datenverarbeitung wie AES generieren hohe gesellschaftliche Schädigungspotenziale, etwa Diskriminierungsrisiken (vgl. Orwat 2019) und Manipulationsrisiken, sogar die konkrete Beeinträchtigung von Menschen- und Verfassungsrechten und weiteren gesellschaftlichen Grundwerten wie Demokratie und Rechtsstaatlichkeit (vgl. Orwat et al. 2022, 258–260). Abgesehen von Datenschutzskandalen mit Massen von Betroffenen liegen eher sehr viele, „kleinteiligere" und sehr unterschiedliche Schädigungen vor, wie die Einschränkung einer unbefangenen Nutzung von digitaler Technik.[13] Zudem kann es zu Ungleichbehandlungen durch falsche oder ungerechte Bildung von und Zuordnung zu Personenklassen oder individuellen Zuschreibungen kommen. Damit verbunden sind beispielsweise Schäden wie die Verletzung der Menschenwürde durch Behandlungen allein als bloßes Objekt, Stigmatisierungen, Stereotypisierungen, Rufschädigungen, Missbrauch von Informationsmacht und struktureller Überlegenheit, Konformitätszwänge durch Überwachungsdruck, Einschüchterungseffekte[14] etc.

Die Vielfalt der Schädigungen, die teils abstrakt wirken, teils sehr subjektiv wahrgenommen werden, steht der Bestimmung von Schadensausmaßen oder Eintrittswahrscheinlichkeiten und der Bildung von Risikoklassen (etwa bei Krafft und Zweig 2019) entgegen. So wäre zum Beispiel die Bildung einer Stufe „unbedenklich" oder „nicht-hohes" Risiko immer noch mit großen Unsicherheiten für die Regulierenden verbunden. Sie werden hauptsächlich als Schädigungen auf individueller Ebene betrachtet, allerdings bleiben hier die Schädigungen auf kollektiver wie auch auf Makroebene beziehungsweise für Gemeinwohlorientierungen und Gerechtigkeitsvorstellungen noch weitgehend unskizziert.

3. NOTWENDIGE KONKRETISIERUNG UND OPERATIONALISIERUNG VON RISIKEN

Jegliche risikobasierte Regulierung braucht ein klares Risikoverständnis und eindeutige, justiziable Konkretisierung der Risiken (vgl. Weißbuch 2020, 20). Die Kriterien und Prinzipien über das, was ein Risiko oder auch einen möglichen Schaden darstellt, dienen vor allem zur Vermeidung gegensätzlicher Bewertungen und sie gewähren Rechtssicherheit bei der Risikoabschätzung oder der Einteilung von Anwendungen in Risikostufen, -graden oder -klassen beziehungsweise bei Beurteilung der Übereinstimmung mit Anforderungen und Regulierungsmaßnahmen (wie z. B. Verbote, Moratorien, Zu-

[12] Zum „Vorsorgeprinzip" vgl. Europäische Kommission 2020, 22-26 sowie Orwat et al. 2022, 275-278.

[13] Vgl. zu einer solchen Risikotypeneinteilung Heijden 2019, 12 f.

[14] Vgl. Übersicht in Drackert 2014.

lassungsprüfungen). Doch erst die Integration verschiedener (disziplinärer) Konkretisierungskonzepte kann hier Rechtssicherheit gewährleisten. So sollten neben Risikoklassen auch Indikatoren, Prinzipien und Anforderungen zur Abwehr möglicher Probleme der risikobasierten Regulierung implementiert werden, denn etwa eine ungeeignete Auswahl und Priorisierung von Risiken sowie ungeeignete Interpretationen und Operationalisierungen von Schutzzielen im Regulierungsprozess können im schlechtesten Falle zum Verfehlen der eigentlichen Schutzziele führen. Dabei bemisst sich die Effektivität der Risikoregulierung an dem tatsächlichen Erreichen der „richtigen" Schutzziele oder auch der Vermeidung von Schädigungen, zu denen Regulierende Rechenschaft ablegen müssen.

Selbst in „harten" naturwissenschaftlichen Bereichen mit konsolidiertem Wissen wird zur Vorsicht bei der Umsetzung von risikobasierten Regulierungsansätzen gemahnt (vgl. Lloyd-Bostock und Hutter 2008). Dabei ist zu beachten, dass sich die Risikocharakteristika und Wissensgrundlagen der Risikoregulierung von KI- und AES-Anwendungen deutlich von denen bestehender Risikoregulierungen unterscheiden.

4. UNGENÜGENDE WISSENSGRUNDLAGE FÜR EINE RISIKOREGULIERUNG VON KI UND AES

Übliche Risikoregulierung basiert in vielen Fällen auf naturwissenschaftlichem Wissen zur möglichst objektiven Bestimmung von Risiken. Als wissenschaftliches Wissen ist es nachvollziehbar und anfechtbar und damit fähig zur Weiterentwicklung und kann zu konsolidiertem Wissen ausgearbeitet werden, sodass Risikoabschätzungen mit einem relativ hohen Grad an Eindeutigkeit und Überprüfbarkeit möglich sind (vgl. Grunwald 2019, 28).

Im Vergleich zu diesen meist naturwissenschaftlich geprägten Bereichen der Risikoregulierung sind die Wissensgrundlagen bei KI- und AES-Anwendungen unvollständig, gesellschaftlich sehr asymmetrisch verteilt und durch mangelnde Eindeutigkeit und Unsicherheiten bei den Bewertungsmaßstäben geprägt.[15] Zunächst geht es um Wissen über die für einzelne Anwendungen relevanten Wahrscheinlichkeiten von unerwünschten Ereignissen, die durch Anwendungen von KI und AES verursacht werden können. Dieses Wissen ist weitgehend in den Händen der Betreibenden. Es liegt also nicht nur eine Situation der Risikoregulierung mit wissenschaftlich nicht konsolidiertem Wissen vor, sondern es besteht auch Unklarheit, Unbestimmtheit und Uneindeutigkeit bei demjenigen Wissen, das zur Einschätzung und Bewertung der Risiken erforderlich wäre.

Ferner gibt es Risiken, bei denen die Verursachung von Schäden für die Gesellschaft nicht einzelnen Verfahren zugeschrieben werden kann. So kommt es zu unbeabsichtigten Folgen für ganze Gesellschaftsgruppen, die nicht durch das Individualrecht abgedeckt sind: Vertrauensverluste, Einschüchterungs- oder Abschreckungseffekte („chilling effects"), Rückzug von der Nutzung digitaler Technik (= soziale Segregierung und Schäden an Prozessen demokratischer Willensbildung). 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.

[15] Vgl. Orwat et al. 2022, 270-274 zu Unsicherheiten im wissenschaftlichen Wissen über Risiken.

5. NORMATIVE ENTSCHEIDUNGEN TROTZ UNSICHERHEITEN

Problematisch für eine Risikoregulierung von KI und AES ist vor allem die normative Ambiguität bei der Risikobestimmung und -bewertung (vgl. Orwat et al. 2020 und Orwat et al. 2022, 265–270):

- Bereits die Auswahl und Priorisierung von Grundwerten beziehungsweise Schutzgütern ist eine normative Entscheidung.
- Die Festlegung und der Zuschnitt des Regulierungsobjekts ist ebenfalls eine normative Entscheidung, mit höchst unterschiedlichen Regulierungsformen (z. B. Marktzulassungsverfahren oder Selbstbeschränkung mit Stichprobenkontrollen).
- Die Konkretisierung beziehungsweise Operationalisierung von gesellschaftlichen Grundrechten und -werten ist ein normativer Entscheidungsprozess im Sinne einer gesellschaftlichen Grundsatzentscheidung – zum Beispiel hinsichtlich der Umsetzung von Gerechtigkeitsvorstellungen (Orwat et al. 2022, 265–267).
- Auch der Umgang mit unvermeidlichen Wertekonflikten erfordert normative Abwägungen (etwa über die Notwendigkeit der Eingriffe in beziehungsweise Beschränkung der Grundrechte und -werte sowie deren Verhältnismäßigkeit) (Orwat et al. 2022, 268).
- Insbesondere die Bestimmung von akzeptablen Restrisiken beziehungsweise des Risikoausmaßes (wie z. B. Fehlerraten) bedarf normativer Abwägungen (Orwat et al. 2022, 269).
- Die gesellschaftliche Verteilung der Risiken – etwa ob bestimmte Bevölkerungsgruppen besonders schwer betroffen sind – ist eine normative Entscheidung.

Alle diese normativen Entscheidungen, die trotz Unsicherheiten getroffen werden (müssen), verlangen nach politischen Entscheidungsprozessen, denn ethische Richtlinien etwa können bloß als Orientierung für die Risikobewertung dienen, sind aber rechtlich unverbindlich. Denn teilweise ist unklar, ob die Richtlinien überhaupt in ihren Anforderungen über bestehende rechtliche Vorgaben hinausgehen oder diese sogar unterschreiten und damit lediglich einem sogenannten „ethical washing“ dienen würden.

6. GEMEINWOHLORIENTIERUNG ZUR VERMITTLUNG DES KONFLIKTS ZWISCHEN INDIVIDUELLER UND KOLLEKTIVER EBENE IN DER DATENÖKONOMIE?

Die Personalisierung und Individualisierung von Prozessen, die üblicherweise auf Solidarität beruhen, zum Beispiel öffentliche oder auch kollektive Güter, wie Gesundheit und Bildung, generieren zunehmend gesellschaftliche beziehungsweise politisch-ökonomische Probleme, die sich nicht auf technische Probleme reduzieren lassen. Als Beispiel kann der Datenschutz bei Krankenkassen dienen, bei dem individuelle zu kollektiven Patienteninteressen in direktem Widerspruch stehen. Jeanette Hofmann bringt den Konflikt, der sich durch die Digitalisierung eingestellt hat, wie folgt auf den Punkt: „Die Stärkung und das gleichzeitige Unterlaufen von Menschenrechten bilden einen zentralen Reibungspunkt im Verhältnis von demokratischer Selbstbestimmung und digitalen Geschäftsmodellen.“ (Hofmann 2017, 14) Die kollektive Festlegung gesellschaftlicher Optimierungsziele könnte eine Zielausrichtung jenseits privat-ökonomisch nützlicher Datenverwertung sein, um diesem Konflikt zu begegnen. Dass dies sogar eine Gemeinwohlorientierung für KI und AES und die Suche nach fairen gemeinwohlorientierten Algorithmen bedeutet, machen Stefan Selke et al. (2018) deutlich:

„Einerseits ermöglichen Big-Data-Analysen extrem individuelle Bestimmungen und Kategorisierungen sowie darauffolgend maßgeschneiderte Angebote. Algorithmen wirken sich zunächst trennend (im Wortsinn: diskriminierend) aus. Gleichzeitig betreffen die Folgen die kollektive Ebene der Gesellschaft selbst, schon allein deshalb, weil die technologischen Lösungen in den Märkten kollektiviert werden (müssen). [...] Daher ist es auch in Zukunft sinnvoll, nach Anwendungsbereichen für faire gemeinwohlorientierte Algorithmen zu suchen, die gleichzeitig Diskriminierungen vermeiden. Das bedeutet, von gesellschaftlichen ‚Schieflagen' oder Herausforderungen auszugehen und dann dafür Technologien zu entwickeln – nicht umgekehrt. [...] Stärkung der Gemeinwohlorientierung – ein Topos, der [...] ideologisch belastet ist! [16][...] Dennoch stellt sich die Frage, ob und wie es [...] gelingen kann, Gemeinwohlorientierung als Balance zwischen den unterschiedlichen Interessen nutzbar zu machen. Der Orientierungsrahmen dazu ist vorhanden. Demokratie dient der Aushandlung individueller und kollektiver Perspektiven. [...] Damit würde sich eigentlich die Chance eröffnen, Big Data alternativ zu rahmen und insbesondere solche Anwendungen zu fördern, die eine langfristige, sozialprogressive und nachhaltige Wirkung versprechen. [...] Gemeinwohlorientierung kann auch bedeuten, Daten für übergreifende Zwecke einzusetzen. Datensammeln und Datenanalyse wäre in dieser Einstellung also nicht primär negativ – als Eingriff in die Privatsphäre – besetzt, sondern positiv konnotiert." (Selke et al. 2018, 155 f.)

V. FAZIT UND SECHS VORSCHLÄGE ZUM AUSBAU DER RISIKOREGULIERUNG VON KI UND AES

Risikoregulierung von KI und AES ist 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. Stattdessen ist dieser vielmehr ein politischer Prozess, in dem es um die Abwägung zwischen den Realisierungen oder Einschränkungen verschiedener Grundrechte und -werte geht, wie digitale 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. Deshalb folgen hier abschließend (im Anschluss an Orwat et al. 2020 und Orwat et al. 2022, 275–279) sechs konkrete Vorschläge zum Ausbau der Risikoregulierung von KI und AES – auch als Orientierung für einen progressiven Datenschutz:

(1) Konkretisierung der Risikokriterien inklusive gesellschaftlicher Dimensionen der Risiken

(2) Forschung zur Operationalisierung von Grundrechten und -werten

[16] Vgl. Selke et al. 2018, Kap. 5.4.

(3) politische Prozesse der gesellschaftlichen Abwägung normativer Entscheidungen der Risikoregulierung durch demokratisch legitimierte und kontrollierte Beteiligungsverfahren – vor jeglicher Risikobestimmung und -bewertung

(4) ohne politisch legitimierte, objektiv untersuchbare und verifizierbare Prozesse keine Delegation an Dritte (etwa Test- bzw. Zertifizierungseinrichtungen oder Treuhänder)

(5) Notwendigkeit der Anwendung des prinzipienbasierten Vorsorgeprinzips[17]

(6) gesellschaftliche Verteilung von Risiken und Effizienzgewinnen durch Automatisierung mit KI und AES unter dem Gemeinwohlgesichtspunkt behandeln[18]

Zusammenfassend sei betont, dass die Notwendigkeit umfangreicher politischer Prozesse besteht, bevor eine Risikoregulierung eingerichtet und betrieben werden kann. Erst durch regelmäßige systematische Erfassung und Analyse von unabhängigen Stellen zur Nutzung von KI und AES in allen relevanten Bereichen (Wirtschaft, Verwaltung, Justizsystem, Polizeiarbeit etc.) sowie zu jeweiligen Anwendungszwecken und potenziell Betroffenen (Ausmaß und Zusammensetzung) können die Dimensionen und der Umfang potenzieller Risiken in und für die Gesellschaft abgeschätzt werden. Problematisch bleibt, dass eine Konkretisierung des Gemeinwohlprinzips für eine Risikobeurteilung (noch) weitgehend unklar ist.

VI. LITERATURVERZEICHNIS

AI HLEG (Level Expert Group on Artificial Intelligence) 2019a: Ethics guidelines for trustworthy AI. 8. April. Brüssel: AI HLEG. https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html (Zugriff: 15.02.2022).

AI HLEG (Level Expert Group on Artificial Intelligence). 2019b. Policy and investment recommendations for trustworthy AI. 26. Juni. Brüssel: AI HLEG. https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence (Zugriff: 15.02.2022).

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. 8. April. Strasbourg: Council of Europe. https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016809e1154 (Zugriff: 15.02.2022).

Datenethikkommission. 2019. Gutachten der Datenethikkommission. Oktober. Berlin: Datenethikkommission der Bundesregierung. https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.pdf;jsessionid=0B4ED9E1D1C53F856500FBE8D3D30AC2.1_cid373?__blob=publicationFile&v=6 (Zugriff: 15.02.2022).

[17] Vgl. Europäische Kommission 2020, 22-26 sowie Orwat et al. 2022, 275ff.

[18] Vgl. Selke et al. 2018 sowie Modelle 2 und 3 im Vortrag von Ingrid Schneider (2022).

Dorfleitner, Gregor und Lars Hornuf. 2018. Analyse der Datenschutzerklärungen deutscher FinTech-Unternehmen nach Einführung der DSGVO. ABIDA-Gutachten. o. O:. ABIDA. https://www.abida.de/sites/default/files/ABIDA_Folgegutachten_Fintech_DSGVO.pdf (Zugriff: 15.02.2022).

Drackert, Stefan. 2014. *Die Risiken der Verarbeitung personenbezogener Daten: Eine Untersuchung zu den Grundlagen des Datenschutzrechts*. Schriftenreihe des Max-Planck-Instituts für Ausländisches und Internationales Strafrecht S, Strafrechtliche Forschungsberichte 149. Berlin: Duncker & Humblot.

Europäische Kommission. 2020. Weißbuch zur Künstlichen Intelligenz - ein europäisches Konzept für Exzellenz und Vertrauen. COM(2020) 65 final. 19. Februar. Brüssel: Europäische Kommission. https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_de.pdf (Zugriff: 15.02.2022).

Europäische Kommisssion. 2020. Zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen. COM(2020) 65 final. 19. Februar. Brüssel: Europäische Kommission. https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_de.pdf (Zugriff: 15.02.2022).

Grunwald, Armin. 2019. *Technology assessment in practice and theory*. London: Routledge.

Haag, Matthias und Hendrick Risthaus. 2022. Regulierung von KünstlicherIntelligenz – DS-GVO als Leitbild? In: *Künstliche Intelligenz - Ethik und Recht*, hg. von Thomas Hoeren und Stefan Pinelli, 289–321. Schriftenreihe Information und Recht 87. München: C.H. Beck.

Heijden, Jeroen van der. 2019. Risk governance and risk-based regulation: A review of the international academic literature. *SSRN Electronic Journal*. (20. Juni). https://doi.org/10.2139/ssrn.3406998.

Hofmann, Jeanette. 2017. Demokratie im Datenkapitalismus. *WZB Mitteilungen*, Nr. 155: 14–17. https://bibliothek.wzb.eu/artikel/2017/f-20465.pdf (Zugriff: 15.02.2022).

Holzhofer, Martin. o. D. DSGVO Bußgeld Datenbank. *DSGVO-Portal*. https://www.dsgvo-portal.de/dsgvo-bussgeld-datenbank.php (Zugriff: 15.02.2022).

Krafft, Tobias D. und Katharina Zweig. 2019. Transparenz und Nachvollziehbarkeit algorithmenbasierter Entscheidungsprozesse. Ein Regulierungsvorschlag aus sozioinformatischer Perspektive. 22. Januar. Berlin: vzbv. https://www.vzbv.de/sites/default/files/downloads/2019/05/02/19-01-22_zweig_krafft_transparenz_adm-neu.pdf (Zugriff: 15.02.2022).

Lloyd-Bostock, Sally M. und Bridget M. Hutter. 2008. Reforming regulation of the medical profession: The risks of risk-based approaches. *Health, Risk & Society* 10, Nr. 1: 69–83. https://doi.org/10.1080/13698570701782460.

Orwat, Carsten, Anja Folberth, Jascha Bareis, Jutta Jahnel und Christian Wadephul. 2022. Risikoregulierung von künstlicher Intelligenz und automatisierten Entscheidungen. In: *Künstliche Intelligenz - Ethik und Recht*, hg. von Thomas Hoeren und Stefan Pinelli, 255–287. Schriftenreihe Information und Recht 87. München: C.H. Beck.

Orwat, Carsten, Anja Folberth, Jascha Bareis, Jutta Jahnel und Christian Wadephul. 2020. Risikoregulierung der KI: Normative Herausforderungen und politische Entscheidungen. Stellungnahme zum Weißbuch der Europäischen Kommission „Zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen“. 14. Juni. Karlsruhe. https://doi.org/10.5445/IR/1000121489.

Orwat, Carsten. 2019. *Diskriminierungsrisiken durch Verwendung von Algorithmen: Eine Studie, erstellt mit einer Zuwendung der Antidiskriminierungsstelle des Bundes*. Baden-Baden: Nomos.

Pohle, Jörg. 2022. *Datenschutz: Rechtsstaatsmodell oder neoliberale Responsibilisierung? Warum Datentreuhänder kein Mittel zum Schutz der Grundrechte sind*. Paper der Vortragsreihe „Zu treuen Händen?“ 5. Düsseldorf: Verbraucherzentrale NRW. https://www.verbraucherforschung.nrw/sites/default/files/2022-02/zth-05-pohle-datenschutz-rechtsstaatsmodell-oder-neoliberale-responsibilisierung_0.pdf (Zugriff: 22.02.2022).

Sartor, Giovanni und Francesca Lagioi. 2020. *The impact of the general data protection regulation on artificial intelligence*. Directorate General for Parliamentary Research Services. Luxemburg: Publications Office. https://doi.org/10.2861/293.

Schneider, Ingrid. 2022. *Datentreuhandschaft durch Intermediäre: Chancen, Herausforderungen und Implikationen.* Paper der Vortragsreihe „Zu treuen Händen?“ 2. Düsseldorf: Verbraucherzentrale NRW. https://www.verbraucherforschung.nrw/sites/default/files/2022-02/zth-2-schneider-datentreuhandschaft-durch-intermediaere.pdf (Zugriff: 15.02.2022).

Selke, Stefan, Peter Biniok, Johannes Achatz und Elisabeth Späth. 2018. Ethische Standards für Big Data und deren Begründung. ABIDA-Gutachten. o. O.: ABIDA. https://www.abida.de/sites/default/files/ABIDA%20Gutachten%20Ethische%20Standards.pdf (Zugiff: 15.02.2022).

Wachter, Sandra, Brent Mittelstadt und Luciano Floridi. 2017. Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law* 7, Nr. 2: 76–99. https://doi.org/10.1093/idpl/ipx005.

Wiebe, Andreas und Céline Helmschrot. 2019. Untersuchung der Umsetzung der Datenschutz-Grundverordnung (DSGVO) durch Online-Dienste. o. O. https://www.bmj.de/SharedDocs/Downloads/DE/News/Artikel/112919_DSGVO_Studie.pdf?__blob=publicationFile&v=2 (Zugriff: 15.02.2022).