Typoskript im Sinne des § 38 Abs. 4 UrhG; das Original ist abgedruckt in der JZ 2017, 1017 ff.

Professor Dr. Mario Martini, Speyer*

Algorithmen als Herausforderung für die Rechtsordnung

Algorithmen, die im Maschinenraum moderner Softwareanwendungen werkeln, sind zu zentralen Steuerungsinstanzen der digitalen Gesellschaft avanciert. Immer nachhaltiger beeinflussen sie unser Leben. Ihre Funktionsweise gleicht aber teilweise einer Blackbox. Die in ihr schlummernden Risiken zu bändigen, fordert die Rechtsordnung heraus. Der Autor entwickelt erste Regulierungsideen, mit deren Hilfe sich die Wertschöpfungspotenziale automatisierter digitaler Prozesse mit den Grundwerten der Rechtsordnung, insbesondere der informationellen Selbstbestimmung und Diskriminierungsfreiheit, versöhnen lassen.

I. Algorithmen als Schlüssel digitaler Erkenntniswelten

In immer mehr existenziellen Lebensbereichen fräsen sich algorithmisch gesteuerte Klassifikationen durch die analoge Welt und treffen Differenzierungsentscheidungen, die bislang dem Menschen vorbehalten waren: Als ubiquitäre Leitsysteme der sozialen und individuellen Präferenzordnung wirken Softwareanwendungen[1] und die in ihnen implementierten Algorithmen[2] darauf ein, was wir kaufen, zu welchen Konditionen wir

* *Mario Martini* ist Lehrstuhlinhaber an der Deutschen Universität für Verwaltungswissenschaften Speyer und Leiter des Programmbereichs „Digitalisierung" am Deutschen Forschungsinstitut für öffentliche Verwaltung. Der Aufsatz fasst zentrale (Zwischen-)Erkenntnisse des durch das Bundesministerium der Justiz und für Verbraucherschutz geförderten Drittmittelprojekts „Algorithmenkontrolle im Internet der Dinge" zusammen, die Eingang in eine Monographie finden werden, welche im Jahr 2018 erscheint. Der Autor dankt allen voran *Michael Kolain* und *Jan Mysegades,* ferner den ehemaligen Mitarbeiterinnen *Wiebke Fröhlich* und *Saskia Fritzsche* für die sehr hilfreiche inhaltliche Mitwirkung. Internetquellen sind (soweit nicht anders angegeben) auf dem Stand vom 15.9.2017.

[1] „Softwareanwendung" meint ein von Programmcode gesteuertes Gesamtsystem, das – vermittelt durch ein Endgerät – Außenwirkung gegenüber einem Nutzer entfaltet, sei es auf der Grundlage eines heruntergeladenen Programms oder eines Telemediendienstes, etwa einer E-Commerce-Plattform. Im Gegensatz zu einer „Hardwareanwendung" (bspw. einem Roboter) wirkt sie grundsätzlich nicht als solche physisch in die reale Lebenswelt hinein, sondern beschränkt sich auf Prozesse der Datenverarbeitung und Entscheidungsfindung innerhalb eines informationstechnischen Systems.

[2] Algorithmen sind Schritt-für-Schritt-Anleitungen zur Lösung eines (mathematischen) Problems. Als solche sind sie kein Phänomen des digitalen Zeitalters, vgl. etwa *Barth*, Algorithmik für Einsteiger, 2. Aufl., 2014, S. 10 ff.; *Hoffmann-Riem*, AöR 142 (2017), 1 (2 f.). Wenn der Beitrag den (in der publizistischen und rechtswissenschaftlichen Diskussion bislang vorherrschenden) Begriff „Algorithmus" verwendet, hat er stets *Computer*algorithmen vor Augen – und versteht darunter eine in Programmiersprache(n)

einen Kredit erhalten, welche Meldungen wir lesen und wen wir kennenlernen.[3] Sie gewähren Einblick in die Glaskugel, aus der Datenanalysen unser künftiges Verhalten herauslesen – zugleich machen die Orakel der schönen neuen „smarten" Welt unsere Persönlichkeit und unser Leben auch durchsichtiger.

II. Risiken

1. Monopolisierung von Markt- und Meinungsmacht

Je genauer Amazon und Co. mit Hilfe ihrer prall gefüllten Datensilos die Wünsche potenzieller Kunden voraussagen können, desto zielgerichteter können sie ihnen Waren und Dienstleistungen anbieten. Nicht nur aus Einkaufspräferenzen der Vergangenheit, sondern auch aus Facebook-Likes,[4] ja sogar aus dem Rhythmus des Tastaturanschlags lässt sich unterdessen mithilfe technischer Analyseverfahren auf die Konsumneigung rückschließen.[5] Was viele Verbraucher freut, die nicht mehr lange nach dem passenden Produkt suchen müssen, sondern auf personalisierte Vorschläge sowie Stimuli reagieren können und sogar individuelle Rabatte erhalten, macht ihr Verhalten aber auch leicht berechenbar und setzt mit steigenden Nutzerzahlen zugleich eine Markt-Macht-Spirale in Gang. Angetrieben durch für Plattformmärkte typische Netzwerk- und Skaleneffekte[6] beginnt sie sich immer schneller zu drehen.

Schlimmstenfalls befähigt sie Datenkonzerne, den Marktzutritt neuer Akteure zu behindern[7] und damit das reibungsfreie Funktionieren des Wettbewerbs als Entdeckungsverfahren (*Hayek*) zu beeinträchtigen – ferner gesellschaftliche Meinungsbildungsprozesse zu steuern und menschliches Verhalten engmaschig zu überwachen. Schon heute fürchten nicht wenige Bürgerinnen und Bürger eine negative Scoring-Bewertung mehr als staatliche Zwangsmaßnahmen. Dass auch (verdeckte) politische Einflussnahme zu dem Risikoarsenal ubiquitärer algorithmischer Datenanalyse gehört, ahnen wir spätestens seit den Gerüchten um die Rolle von Facebook im US-Wahlkampf.[8]

---

transformierbare Vorgehensweise, nach der ein Rechner eine bestimmte Aufgabe in endlicher Zeit bewältigt.

[3] Dazu bspw. *Coglianese/Lehr*, Georgetown Law Journal 105 (2017), 1147 ff.; *Hoffmann-Riem* (Fn. 2), 4 f.; *Tutt*, Administrative Law Review 67 (2016), 1 (2).

[4] Siehe die empirische Untersuchung bei *Youyoua/Kosinski/Stillwell*, PNAS 112 (2014), 1036 (1037 ff.).

[5] Dazu *Christl*, Kommerzielle digitale Überwachung im Alltag, 2014, S. 21 unter Hinweis auf *Epp/Lippold/Madryk*, CHI 2011, 715 ff.

[6] Siehe bspw. *Bundesministerium für Wirtschaft und Energie*, Weißbuch Digitale Plattformen – Digitale Ordnungspolitik für Wachstum, Innovation, Wettbewerb und Teilhabe, 2017, S. 28, 58.

[7] Dazu gesellt sich häufig auch die Marktmacht, die Bedingungen für die Teilhabe an Infrastrukturleistungen der digitalen Welt sowie den Markterfolg zu diktieren. Wer bspw. die Vertragsbedingungen des *Google Play Store* oder des *Apple Store* nicht akzeptieren möchte, dem bleiben zentrale Märkte der Smartphone-Welt faktisch verschlossen.

[8] *Facebook* sah sich dem Verdacht ausgesetzt, bewusst Nachrichten aus dem konservativen Spektrum zu unterdrücken und damit die sog. *Trending Topics* zugunsten anderer politischer Strömungen zu

2. Intransparenz

Der mathematisch-logische Problembewältigungsmodus eines Algorithmus verheißt Objektivität. Seine in Programmcode transformierten Entscheidungsmaßstäbe bleiben für den Nutzer einer Softwareanwendung[9] aber ein Mysterium: Wie sie zu ihren Arbeitsergebnissen gelangt, liegt im Verborgenen. Die algorithmische Welt gründet ihr Fundament auf Arkan-Formeln.

Eine intransparente und dadurch für Betroffene nicht nachvollziehbare Entscheidungsfindung birgt Gefahren für gesellschaftliche Grundwerte.[10] Die „Blackbox Algorithmus“[11] beschwört nicht nur das dumpfe Gefühl herauf, überwacht zu werden und die Selbstbestimmungsmacht darüber zu verlieren, wer welche persönlichen Daten sammeln, auslesen und daraus Schlüsse ziehen darf. Sie kann auch die Anwandlung auslösen, nach undurchsichtigen Entscheidungskriterien diskriminiert zu werden oder zum Objekt sublimer Steuerung[12] zu degenerieren. Gleichzeitig entwaffnet der fehlende Einblick in das Arsenal einer Softwareanwendung den Verbraucher: Die Rechtmäßigkeit von Entscheidungen kann nur prüfen, wer die Datengrundlage, Handlungsabfolge und Gewichtung der Entscheidungskriterien kennt – und versteht.

3. Diskriminierung

Algorithmen beruhen auf menschlichen Modellierungen, in die auch Ansichten, Neigungen und Wertmuster ihrer Schöpfer einfließen; sie sind nicht per se objektiv oder neutral. Ihre Wertungen folgen strukturell den Zielmustern des Entscheiders und damit typischerweise eher der ökonomischen Rationalität ihrer Schöpfer als Wertvorstellungen des Gemeinwesens. In diesem Entscheidungsraster übersetzen sie soziale Realität in binären Code, der Eingabevariablen nach metrischen Kategorien strukturiert und mit eigenen Mustern abgleicht: Algorithmenbasierte Entscheidungen sind das Ergebnis einer klassifizierenden Rasterung auf der Grundlage stochastischer Rückschlüsse. Diese ermitteln ihrem Wesen nach nur Korrelationen, aber keine Kausalitäten.[13] Eine Softwareanwendung mit Profiling-Algorithmus trifft aufgrund von

---

manipulieren, vgl. etwa *Meier*, Macht Facebook geheime Politik gegen Konservative?, Welt online vom 11.5.2016. Als Antwort auf die Vorwürfe veröffentlichte der Internetkonzern seine internen Auswahlrichtlinien zumindest teilweise.

[9] Das gilt jedenfalls für proprietäre Softwareanwendungen. Bei Open-Source-Software liegt der Programmcode bereits per se für jedermann offen. Ihn zu entschlüsseln, überfordert den durchschnittlichen Verbraucher aber in aller Regel.

[10] Dazu auch *Hoffmann-Riem* (Fn. 2), 32 f.

[11] Den Begriff hat *Pasquale*, The Black Box Society, 2015 (passim) geprägt.

[12] Dazu bspw. *Hildebrandt*, Smart technologies and the end(s) of law, 2016, S. 261, 263; *Hoffmann-Riem* (Fn. 2), 6, 11 ff.

[13] Aus der auf der Grundlage von Massendaten gewonnenen Erkenntnis, dass Jugendliche überdurchschnittlich häufig Schokolade essen und häufig Akne haben, folgt eben noch nicht, dass Schokolade Akne verursacht. Algorithmen sind für derartige Cum-hoc-ergo-propter-hoc-Fehlschlüsse anfällig. Dazu auch *Martini*, DVBl 2014, 1481 (1485).

Gruppenwahrscheinlichkeiten Aussagen über Einzelne – und entscheidet zunehmend autonom, wie sie welche Kriterien gewichtet. Ihre quantifizierende Durchdringung des Sozialen macht dadurch auch den Zugang zu Leistungen von Gruppenzuordnungen abhängig, die in eine selbstreferenzielle Entscheidungsmechanik münden können, welche bereits existierende strukturelle Ungleichheiten verstärken oder zementieren kann. Schon der Wohnort kann dafür ausschlaggebend sein, wer einen Kredit erhält oder auf Rechnung zahlen darf; wer in Neukölln oder am Mainzer Hauptbahnhof wohnt, muss sich auf schlechtere Konditionen einstellen als eine wirtschaftlich gleich leistungsfähige Person in Berlin-Charlottenburg.[14] Auf der Grundlage einer Vergangenheitsdatenanalyse können E-Commerce-Plattformen dem zahlungskräftigen Apple-Kunden für das gleiche Produkt unbemerkt einen anderen Preis unterbreiten als dem preisbewussten Medion-Kunden.[15] Menschen mit Namen wie „Mandy", „Kevin" oder „Mohammed", mit denen ein Algorithmus typischerweise geringe Zuverlässigkeit assoziiert, fallen in automatisierten Bewerbungsverfahren häufiger durch das Ausleseraster: Sie erhalten keine Einladung zu Vorstellungsgesprächen.[16]

Im US-Bundesstaat Wisconsin ist das Vertrauen in die Macht der Algorithmen sogar bereits so weit vorangeschritten, dass die Software *Compas* die Wahrscheinlichkeit dafür berechnet, ob ein Straftäter rückfällig wird. Das System sollte nicht zuletzt den Wertmustern ethnischer Vorurteile entgegenwirken. Genau das Gegenteil trat aber ein: Dunkelhäutigen attestiert die Software ein Rückfallrisiko, das – im Vergleich zu Weißen – deutlich über der tatsächlichen Rückfallquote liegt. Die Abweichung zwischen der Prognose und der tatsächlichen Rückfallhäufigkeit ist bei ihnen doppelt so hoch wie bei Weißen.[17]

Die Beispiele zeigen: Diskriminierungsverbote aus Art. 3 GG bzw. Art. 21 GRCh sind dem binären Wesen einer Software grundsätzlich fremd.[18] In vielen Kontexten sind Algorithmen den Schwankungen und Vorteilen menschlicher Entscheidungsfindung zwar überlegen, diskriminierungsfrei sind sie aber nicht; ihnen fehlt ein ethischer Kompass.

---

[14] Vgl. etwa *Álvarez*, Im Netz hat jeder seinen Preis, Tagesspiegel Online vom 5.4.2016. US-amerikanische Wissenschaftler entwickeln zu diesem Phänomen bereits Gegen- und Aufdeckungsmaßnahmen, vgl. *Hannak/Soeller/Lazer et al.*, Measuring Price Discrimination and Steering on E-commerce Web Sites, in: Williamson/Akella/Taft (Hrsg.), IMC '14, 2014, S. 305 (307 f.); *Vafa/Haigh/Leung et al.*, Price Discrimination in The Princeton Review's Online SAT Tutoring Service, http://techscience.org/a/2015090102/ (22.11.2016).

[15] Vgl. *Mattioli*, In Orbitz, Mac Users Steered to Pricier Hotels, Wall Street Journal vom 23.8.2012.

[16] Zu Risiken und Chancen der Bewerbervorauswahl per Online-Persönlichkeitstests *Lischka/Klingel*, Wenn Maschinen Menschen bewerten, 2017, S. 22 ff.

[17] Zur rassendiskriminierenden Tendenz der *Compas*-Analyse siehe *Angwin/Larson/Mattu et al.*, Machine Bias, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[18] Dazu bereits *Grundmann/Hacker*, ERCL 13 (2017), 255 (278 f.); *Martini/Nink*, NVwZ-Extra 10/2017, 1 (9 f.).

Ihre Kontrolle entpuppt sich bei *lernfähigen Systemen*[19] technisch gleichzeitig immer mehr als Quadratur des Kreises. Denn Methoden maschinellen Lernens arbeiten kein statisches Schema ab: Sie sind nicht nach einem linearen Modell eines Zeilencodes programmiert,[20] sondern passen ihre Wissensbasis und ihre Entscheidungsstrukturen ihrem Handlungsumfeld dynamisch an – und das grundsätzlich autonom.[21] Den Kern ihrer maschinellen Logik bildet ein Dialog zwischen Daten und Modellen: Lernfähige Softwareanwendungen erweitern ihre Analysekraft auf der Grundlage von Trainingsdaten und verallgemeinern die gewonnenen Erkenntnisse durch Lerntransfer. Ihnen gelingt es in der Folge immer besser, Muster und Gesetzmäßigkeiten in den ihnen zur Verfügung stehenden Daten – auch unter Unsicherheitsbedingungen[22] – zu erkennen sowie Wichtiges von Unwichtigem treffsicher zu unterscheiden. Aus den Datenhalden können sie selbstadaptiv Schlussfolgerungen ziehen, ohne explizit auf eine konkrete Form der Problemlösung programmiert zu sein. Dadurch sind sie in der Lage, Aufgaben mit minimalen menschlichen Vorgaben gleichsam intrinsisch lösen.[23] Wie sie zu ihren Ergebnissen gelangen, bleibt von außen nicht einsehbar und mithin auch für Kontrollmechanismen nicht nachvollziehbar. Fehler lassen sich in einem solchen Modell daher nicht auf traditionellem Wege aufspüren und ausmerzen.[24]

III. Regulierungsvorschläge

Den Gefahrenzonen im Maschinenraum des digitalen Zeitalters können Verbraucher nicht ohne Weiteres selbst durch aufgeklärtes, wachsames Verhalten ausweichen. Der Einzelne ist schnell überfordert, seine Privatsphäre gegen hochkomplexe und oftmals neugierige Anwendungen zu verteidigen, deren technische Funktionsweise er nicht zu durchdringen vermag. Nicht selten folgt sein Verhalten auch der Losung des Matthäus-Evangeliums: „Der Geist ist willig, aber das Fleisch ist schwach“. Zwischen der individuellen Wertschätzung für informationelle Selbstbestimmung und der Bereitschaft, sie bspw. durch kritische Lektüre von Einwilligungserklärungen[25] durchzusetzen, tut sich ein nicht immer überbrückter Graben auf (sog. *privacy paradox*[26]). Bei intuitiver Abwägung zwischen verfügbaren zeitlichen Ressourcen, dem

[19] Gemeint sind damit Softwareanwendungen, die sich nicht nur (statischer) linearer Handlungsabfolgen, sondern auch (dynamischer) maschineller Lernverfahren bedienen.
[20] Vgl. etwa die anschauliche Darstellung zur Funktionsweise neuronaler Netze bei *Schlieter*, Die Herrschaftsformel, 2015, S. 24 ff.
[21] Vgl. zu den verschiedenen (mathematischen) Modellen für die Umsetzung *Ertel*, Grundkurs Künstliche Intelligenz, 4. Aufl., 2016, S. 195 ff.
[22] *Ghahramani*, Nature 521 (2015), 452 (452).
[23] Vgl. *Tutt* (Fn. 3), 8.
[24] Dazu auch *Barocas/Selbst*, California Law Review 104 (2016), 671 (692 f.); *Martini/Nink* (Fn. 18), 12.
[25] Siehe auch die Kritik zum Konzept der Einwilligung bei *Hoffmann-Riem* (Fn. 2), 21 ff.
[26] Vgl. dazu etwa *Athey/Catalini/Tucker*, The Digital Privacy Paradox: Small Money, Small Costs, Small Talk, 2017 mit einem Experiment unter 3000 Studierenden, das der Leitfrage verschrieben war: Pizza over privacy?; *Wagner*, in: Wolff/Brink (Hrsg.), BeckOK DatenschutzR, 21. Ed., Stand: 1.8.2017, Grundlagen und bereichsspezifischer Datenschutz - Landesdatenschutz, Rn. 79; ausführlich und

Recherche- und Präparierungsaufwand, der sich mit Instrumenten des Privatsphärenschutzes verbindet, und dem individuellen Nutzen digitaler Dienstleistungen folgt der Nutzer typischerweise dem einladendsten Pfad – im Zweifel siegt die Bequemlichkeit. Unter den Prämissen eines individuellen Opportunitätskosten-Kalküls verspricht der normative Anspruch, Persönlichkeitsschutz durch die individuelle Lektüre seitenlanger Einwilligungserklärungen sicherstellen zu wollen, daher keinen durchschlagenden Erfolg. Sachgerecht ist es dann, Datenschutz stärker (auch) als kollektiven Güterschutz zu konzipieren, namentlich als Schutz eines gemeinschaftlichen Grundwertes, der die Grenzen seiner Geltungskraft auch aus den Ausstrahlungen der Handlungen des Einzelnen auf die Gesellschaft bezieht:[27] Er wirkt Informations- und Durchsetzungsmachtasymmetrien, die zwischen Nutzern und Anbietern bei der Inanspruchnahme eines (Telemedien-)Angebots bestehen, durch gesetzliche Kontrollregeln entgegen. Dazu gehört nicht nur, die Prüfung von Einwilligungserklärungen nicht allein in die Hand des Einzelnen zu legen: Der Staat sollte sie einem strengeren gesetzlichen Kontrollregime unterwerfen, das die Ausgewogenheit der Interessen und strukturelle Durchsetzungsasymmetrien ergänzend zum Einzelnen wirkungsvoll analysiert. Die Vorgaben des Art. 7 DSGVO, insbesondere seines Abs. 4, gehen erste Schritte in die richtige Richtung, bleiben aber hinter dem Wünschenswerten noch zurück.

Zu einem zielgerechten gesetzlichen Kontrollregime gehört auch ein verfahrensrechtlicher Ansatz der Wirkkontrolle über die „Blackbox Algorithmus“: Wenn der Einzelne in einer immer komplexer werdenden digitalen Welt daran scheitert, technische Systeme zu erklären, zu verstehen und zu beherrschen, sollte es sich die Rechtsordnung umso angelegener sein lassen, gesellschaftliche Grundwerte unmittelbar in die Technikgestaltung zu implementieren[28] und wirksame Mechanismen zu ihrer Kontrolle zu entwickeln. Ein solcher Ansatz stellt sich der Frage, wie wir in einer zunehmend algorithmisch gesteuerten Gesellschaft leben wollen – und wie eine passgenaue Regulierung aussehen sollte, die den Asymmetrien der digitalen Welt legislatorisch entgegengewirkt.

Der Gesetzgeber hat dabei einen schwierigen Spagat zu bewältigen: Algorithmenregulierung bewegt sich in in einer komplexen Spannungslage zwischen wirksamem Persönlichkeitsschutz, dem Schutz der Betriebs- und Geschäftsgeheimnisse der Unternehmen und der Förderung digitaler Wertschöpfungspotenziale.

---

differenzierend *Dienlin/Trepte*, European Journal of Social Psychology 45 (2015), 285 (286 f.) m. w. N. zur soziologischen Analyse des Phänomens.

[27] Vgl. auch den Überblick bei *Golla*, Mehr als die Summe der einzelnen Teile? Kollektiver Datenschutz, in: Taeger (Hrsg.), Recht 4.0, 2017, S. 199 (202, 207 f.).

[28] Zu dem Grundgedanken „privacy by design“ siehe Art. 25 I DSGVO vgl. auch *Cavoukian* Privacy by Design – The 7 Foundational Principles, 2011, 3.

Um ihren konkurrierenden Zielen gerecht zu werden, kann der Gesetzgeber an mehreren Punkten auf der Zeitachse ansetzen: präventiv (1.), parallel zum Einsatz einer Softwareanwendung (2.), im Wege begleitender Selbstregulierung (3.) und ex post (4.).

1. Präventive Regulierungsinstrumente

a) Art. 22 DSGVO

An einer Baustelle legislativer Steuerung algorithmenbasierter Verfahren ist der europäische Gesetzgeber bereits tätig geworden: Die neue Datenschutz-Grundverordnung verleiht Betroffenen ein Abwehrrecht gegen vollautomatisierte Entscheidungsverfahren (Art. 22 I DSGVO). Der Vorschrift liegt ein zentraler grundrechtlicher Leitgedanke zugrunde: Algorithmen sollen Menschen nicht zu reinen Objekten der Entscheidung einer Softwareanwendung herabwürdigen dürfen. Das dystopische Narrativ der englischen Comedy-Serie *Little Britain* „Computer says no" stand dem Unionsgesetzgeber dabei als Kontrastfolie eines von einer von einer opaken binären Zahlenlogik determinierten gesellschaftlichen Miteinanders warnend vor Augen.

Der normative Radius des Art. 22 DSGVO ist aber insgesamt enger, als die Überschrift „[…] einschließlich Profiling" vermuten lässt: Zum einen lässt er zahlreiche Durchbrechungen des Grundsatzes zu;[29] zum anderen erfasst er nur Entscheidungen, die *ohne jeglichen menschlichen Einfluss* zustande gekommen sind – nicht aber die algorithmenbasierte *Unterstützung* menschlicher Entscheidungen.[30] Das reine *Scoring* als solches fällt bspw. nicht in den Anwendungsbereich der Vorschrift. Denn die automatisierte Bewertung einer Person, etwa hinsichtlich ihrer Bonität, ist als solche noch keine Entscheidung, sondern bereitet eine solche lediglich vor.[31]

---

[29] Insbesondere im Falle einer Einwilligung oder wenn dies für den Abschluss oder die Erfüllung eines Vertrages erforderlich ist (Art. 22 II lit. a, c DSGVO). Dann muss der Verantwortliche aber Maßnahmen treffen, um die Interessen der betroffenen Person zu wahren (Art. 22 III DSGVO). Zu diesen Schutzmaßnahmen im Einzelnen *Martini/Nink* (Fn. 18), 3 ff.

[30] Kritisch hierzu auch *Hoffmann-Riem* (Fn. 2), 35 f.

[31] Art. 22 DSGVO erfasst diese nur, soweit sie in eine automatisierte Entscheidung mündet, also keine menschliche Entscheidung dazwischentritt. Mitgliedstaatlichen Regelungsspielraum für eine Regelung zum *Scoring*, wie sie § 31 BDSG nF (der im Wesentlichen § 28b BDSG aF entspricht) nunmehr trifft, eröffnet Art. 22 II lit. b DSGVO daher nicht – auch nicht Art. 6 I lit. e DSGVO („Verarbeitung im öffentlichen Interesse"). *Scoring* hat zwar Bedeutung für die Kreditwirtschaft; sie ist gleichsam das Schmieröl der Gesamtwirtschaft. Wären bereits makroökonomische Interessen ausreichend, ließe sich jedoch für sehr viele Verarbeitungen ein öffentliches Interesse rechtfertigen und das System der DSGVO leicht unterlaufen. Lit. e meint vielmehr Verarbeitungen, die spezifischen öffentlichen Aufgaben, nicht aber privatwirtschaftlichen Zwecken verpflichtet sind. Die Kreditvorsorge gehört nicht dazu. § 31 BDSG nF ist daher unionsrechtswidrig.

b) Transparenz

aa) Kennzeichnungspflicht, insbesondere mittels visueller Symbole

Nur wer eine Rechtsverletzung, die ein algorithmenbasiertes Verfahren verursacht, zu erkennen und zu beweisen vermag, kann sich ihrer wirksam erwehren. Dafür muss der Betroffene aber zunächst darum wissen, von einer solchen Entscheidung überhaupt betroffen zu sein.
Findet die Entscheidungsfindung einer Softwareanwendung in einer Blackbox statt, können Transparenzpflichten ein taugliches Regulierungsinstrument sein, um Licht in sensible Entscheidungsvorgänge zu bringen. Als erster Baustein ist daher eine Pflicht empfehlenswert, den Einsatz von (insbesondere lernfähigen) Algorithmen in persönlichkeitssensiblen Feldern[32] zu kennzeichnen.

Die DSGVO kennt im Grundsatz zwar eine Hinweispflicht (Art. 13 II lit. f bzw. Art. 14 II lit. g DSGVO). Sie beschränkt sich – ebenso wie Art. 22 DSGVO – jedoch auf das Bestehen einer „automatisierten Entscheidungsfindung", also auf solche Entscheidungen, die *keinerlei* direktem menschlichen Einfluss unterliegen.

Die Rechtsordnung sollte dabei nicht stehen bleiben, sondern die Kennzeichnungspflicht auf alle algorithmenbasierten Verfahren in persönlichkeitssensiblen Feldern erstrecken. Die Kennzeichnung sollte mithilfe visuell leicht erfassbarer Symbole erfolgen, welche die Bürger tatsächlich erreichen – anderenfalls droht sie Datenschutzhinweise lediglich um einen weiteren (ungelesenen) Absatz zu verlängern.

bb) Begründungspflicht für algorithmenbasierte Entscheidungsverfahren

Als Gegengift gegen die Intransparenz, die algorithmenbasierten Entscheidungsverfahren eigen ist, kann eine Begründungspflicht wirken. Im Idealfall sollte die Aussteuerung einer individuellen Begründung bereits in den Programmcode implementiert sein. In Anlehnung an das Arzneimittelrecht ist ein „digitaler Beipackzettel" denkbar, der das Zustandekommen eines auf der Grundlage persönlichkeitssensibler Algorithmen zustande gekommenen Entscheidungsergebnisses einfach und verständlich[33] erläutert. Die Diensteanbieter müssen algorithmisch generierten Entscheidungen dann Angaben nicht nur zu Vergleichsgruppen und Parametern, sondern auch zu Grundsätzen beigeben, welche die Entscheidungsfindung im Einzelfall leiten.[34] Sie sollten dabei grundsätzlich auch Einblick in die Datengrundlage personenbezogener Informationen geben müssen, die Eingang in die Entscheidung gefunden hat, um eine Richtigkeitsprüfung zu ermöglichen.

---

[32] Welche Bereiche das im Einzelnen sind und ab welcher Erheblichkeitsschwelle eine Regulierung greifen soll, müsste der Gesetzgeber durch hinreichend präzise Klassifikationskriterien konkretisieren.
[33] Vgl. auch die normative Vorgabe des Art. 12 I 1 DSGVO.
[34] Dazu bereits auch *Martini/Nink* (Fn. 18), 11.

Die Implementierung einer (dem Vorbild des § 39 I VwVfG nachgebildeten) Begründungspflicht in Softwareanwendungen wird Programmierer vor Herausforderungen stellen. Insbesondere bei komplexen maschinellen Lernverfahren neuronaler Netze[35] können selbst ihre Schöpfer häufig nur im Nachhinein sagen, *dass* es zu einer Entscheidung gekommen ist – nicht aber, *aus welchen Gründen*. Was technisch schwierig erscheint, muss deshalb aber noch nicht im normativen Sinne unmöglich sein.[36]

Die normativ gebotene Begründungstiefe sollte mit der Größe des Diskriminierungsrisikos und der Persönlichkeitsgefährdung korrespondieren, die dem jeweiligen Entscheidungsgegenstand innewohnt. Eine generelle Pflicht, den Programmcode einer Softwareanwendung[37] offenzulegen, ist aber nicht geboten.[38] Ausreichend, aber auch erforderlich ist eine Information des Betroffenen über die grundsätzliche Entscheidungsstruktur, die den implementierten Algorithmen zugrunde liegt, um die entscheidungsleitenden Prinzipien auch im Einzelfall nachvollziehbar zu machen.[39]

cc) Besondere Transparenzanforderungen an algorithmenbasierte Dienste der Nachrichtenauswahl

Mit der Reichweite von Nachrichtenaggregatoren[40], wie *Google News* oder Facebooks *Newsfeed*, wächst auch deren Ausstrahlungswirkung in die öffentliche Meinungsbildung – und damit ihre Meinungsmacht.[41] Der Schutz unbeeinflusster öffentlicher Meinungsbildung rechtfertigt es, ihnen pro futuro Transparenzpflichten aufzuerlegen,

---

[35] Neuronale Netze arbeiten nach dem Funktionsmechanismus des menschlichen Gehirns, sind also seiner neuronalen Plastizität nachempfunden: Der Lösungsweg, den sie entwickeln, hängt von unzähligen gewichteten Einzelentscheidungen einzelner Netzwerknoten ab. Vgl. auch Fn. 20.

[36] Zu ersten Ansätzen, den Output eines neuronalen Netzes zurückzuverfolgen, vgl. bspw. *Beuth*, Die rätselhafte Gedankenwelt eines Computers, Zeit Online vom 24.3.2017; *Voßberg*, Wie Forscher dem Computer beim Denken zusehen, https://www.ferchau.com/de/de/blog/details/2017/06/19/wie-forscher-dem-computer-beim-denken-zusehen;.

[37] Bei bestimmten Anwendungsformen maschinellen Lernens, bspw. neuronalen Netzen, hilft der Blick auf den Programmcode ohnedies nicht weiter, da er die dynamischen Entscheidungsmuster weder konkret vorzeichnet noch abbildet.

[38] Die DSGVO erstreckt ihre Unterrichtungspflicht in Art. 13 II lit. f bzw. Art. 14 II lit. g zwar auch auf die einem Algorithmus zugrunde liegende Logik; sie beschränkt diese aber grundsätzlich auf Fälle des Art. 22 DSGVO, also auf vollautomatisierte Entscheidungen. Im Hinblick auf den Mindestinhalt der Erläuterung gibt die DSGVO – anders als im Hinblick auf die Hinweispflicht – zwar mit der Wendung „zumindest in diesen Fällen“ eine Öffnung für weitere Anwendungsfelder zu erkennen. Sie lässt aber offen, auf welche weiteren Anwendungsfelder sich die Erläuterungspflicht erstrecken soll.

[39] Erst auf der Stufe aufsichtsrechtlicher oder gerichtlicher Kontrolle kann ein Einblick in den Programmcode bei begründetem Verdacht oder im Wege eines In-camera-Verfahrens notwendig werden. Dazu insbesondere unten III. 2 a.

[40] „Nachrichtenaggregator“ ist ein Anbieter eines Dienstes, der Nachrichten (auf Grundlage eines Algorithmus) automatisch aus verschiedenen Quellen zusammenträgt, listet und ständig aktualisiert. Abgesehen von der Implementierung eines bestimmten Algorithmus verzichtet er im Grundsatz auf jede menschliche Aufbereitung der Beitragsliste nach Relevanz oder Wahrheitsgehalt.

[41] Vgl. auch *Hoffmann-Riem* (Fn. 2), 11 f., 14, 38.

die weiter als bei sonstigen Anbietern reichen. Jedenfalls ab einer kritischen meinungsbildungsrelevanten Größenschwelle sollten sie einen öffentlichen Einblick in ihr technisches Verfahren der Nachrichtenauswahl und -priorisierung gewähren müssen.[42] Mögliche Interessenkonflikte – z. B. wirtschaftliche Verflechtungen, die einen Anreiz auslösen, bestimmte Dienste oder Inhalte gegenüber anderen systematisch herabzustufen oder auszuschließen – müssten sie (bußgeldbewährt) kenntlich machen. Eine solche Transparenz kann ein tragfähiger Baustein eines normativen Gerüsts zum Schutz unbefangener Meinungsbildung in einem freien, demokratischen Gemeinwesen sein. Pluralitäts- oder Must-carry-Pflichten einzuführen,[43] schösse dagegen über das Ziel hinaus: Nachrichtenaggregatoren sollen nach ihrem wesensmäßigen Auftrag gerade eine präferenzgesteuerte Auswahl ermöglichen, nicht ein umfassendes Meinungsspektrum abbilden.

c) Inhaltliche Ex-ante-Kontrolle bei algorithmenbasierten Entscheidungsverfahren in persönlichkeitssensiblen Anwendungsfeldern

Kann *der Einzelne* die Funktionsweise einer Softwareanwendung, die Einfluss auf seine gesellschaftliche Teilhabe ausübt, weder verstehen noch kontrollieren, ist als Instrument eines Ausgleichs informatorischer Wissensasymmetrien ein ergänzende *kollektiver* Kontrollmechanismus sachgerecht: Denkbar ist ein staatliches Kontrollverfahren, das Softwareanwendungen, die Entscheidungen in besonders persönlichkeitssensiblen[44] Anwendungsfeldern steuern, vor ihrem Einsatz durchlaufen müssen.

Zu dem Prüfradius eines solchen „Algorithmen-TÜV“[45] sollten nicht nur der Programmcode deterministischer Verfahren und die korrekte Einbindung der Datenbasis gehören. Er sollte auch kontrollieren, ob die Trainingsprozesse[46] lernfähiger Softwareanwendungen standardisierten Vorgaben entsprechen. Deren Ziel ist es, maschinelle Lernverfahren reguliertem Input auszusetzen, der die spätere

---

[42] Regelungstechnisch ist eine entsprechende Ergänzung des § 54 II RStV sinnvoll. Das Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (NetzDG, vgl. BT-Drs. 18/13013) verfolgt demgegenüber eine andere Zielrichtung: Es sieht eine gesetzliche Berichtspflicht sozialer Netzwerke über den Umgang mit Hasskriminalität und anderen strafbaren Inhalten (§ 2 NetzDG), ein wirksames Beschwerdemanagement (§ 3 NetzDG) sowie die Benennung eines inländischen Bevollmächtigten (§ 5 NetzDG) vor. Kritisch zum NetzDG etwa *Feldmann*, K&R 2017, 292 ff.; *Guggenberger*, NJW 2017, 2577 (2581 f.); *Holznagel*, ZUM 2017, 615 (622 ff.).

[43] Zu dieser Forderung z. B. *Broemel*, MMR 2013, 83 (85); *Hentsch*, MMR 2015, 434 (437); *Kluth/Schulz*, Konvergenz und regulatorische Folgen, Oktober 2014, S. 36 f., 98.

[44] Zur Notwendigkeit, die Einsatzszenarien und Erheblichkeitsschwelle, ab der eine Regulierung greift, näher zu konkretisieren bereits Fn. 32.

[45] *Dräger*, Ein Tüv für Algorithmen, Handelsblatt vom 21.8.2017; *Martini*, DÖV 2017, 443 (453).

[46] Zur nachträglichen, kontinuierlichen Routinevalidierung insbesondere lernfähiger Systeme im Verwaltungsverfahren bereits *Martini/Nink* (Fn. 18), 12 ff.

Musterkennung und -kombination prägt und die algorithmische Intelligenz auf diese Weise gleichsam „erzieht".

Für behördliche Softwareanwendungen sollte eine Vorab-Prüfung im Grundsatz zwingend sein; für nicht-öffentliche Stellen nur dann, wenn Softwareanwendungen in besonders persönlichkeitssensiblen Bereichen zum Einsatz kommen sollen oder wenn schwere Schäden für sonstige wichtige Rechtsgüter drohen.

d) AGG

Die regulatorische Zielsetzung, Diskriminierungsrisiken algorithmenbasierter Verfahren zu begrenzen, geht mit der Schutzmission des Allgemeinen Gleichbehandlungsgesetzes (AGG) Hand in Hand: Beide sollen die Benachteiligung diskriminierungsgefährdeter Menschen – typischerweise Minderheiten – verhindern.

Das AGG schließt zwar softwarebasierte Verfahren schon heute nicht von seinem Anwendungsbereich aus; es ist technologieneutral konzipiert. Zugleich ist es aber auf einen begrenzten Kanon von Lebensbereichen limitiert, namentlich auf Arbeitsverhältnisse, Bildung und Sozialleistungen sowie Leistungen, die der allgemeinen Öffentlichkeit zur Verfügung stehen (§§ 2 und 19 AGG).[47] Für Verträge zwischen Privaten außerhalb des Arbeitsrechts gilt das Gesetz nur bei sog. Massengeschäften und Versicherungen – von der Diskotür bis zur Krankenversicherung, ob analog oder digital.[48]

Zahlreiche spezialisierte Anwendungsfelder softwarebasierter Verfahren erfasst das AGG demgegenüber nicht. De lege ferenda ist eine Ergänzung des Katalogs der Anwendungsfälle des § 2 I AGG um eine Nr. 9 für Ungleichbehandlungen zwischen Privaten erwägenswert, die auf einer algorithmenbasierten Datenauswertung oder einem automatisierten Entscheidungsverfahren beruhen.

2. Begleitende Fehlerkontrolle und Risikomanagement

Ebenso wie nicht für alle Softwareanwendungen eine präventive Zulassungskontrolle geboten ist, reicht sie umgekehrt dort, wo sie ausnahmsweise angezeigt ist, nicht per se aus, um den Risiken einer Lebenswelt wirksam zu begegnen, die von der Dynamik algorithmenbasierter Entscheidungen durchdrungen ist.

---

[47] Zu der umstrittenen Frage, ob § 2 AGG den Anwendungsbereich des § 19 I AGG mit absteckt *Thüsing*, in: Säcker/Rixecker/Oetker et al. (Hrsg.), MüKo-BGB I, 7. Aufl., 2015, § 19 AGG, Rn. 4 (bejahend); a. A. wohl *Wendtland*, in: Bamberger/Roth (Hrsg.), BeckOK BGB, 43. Ed., 2017, AGG § 19, Rn. 3 ff.

[48] Erfasst sind jedenfalls alle Geschäfte, die der Anbieter mit *invitatio ad offerendum* bewirbt, vgl. die Begründung des Gesetzesentwurfs BT-Drs. 16/1780, S. 25 f., 32; *Franke*, in: Däubler/Bertzbach (Hrsg.), AGG, 3. Aufl., 2013, § 2, Rn. 56; *Nickel*, NJW 2001, 2668 (2669); vgl. auch OLG Karlsruhe, NJW 2010, 2668 (2669); a. A. *Thüsing*, NJW 2003, 3441 (3442 f.). Differenziert zum *Scoring Moos/Rothkegel*, ZD 2016, 561 (563 f.).

Für komplexe Softwareanwendungen in persönlichkeitssensiblen Anwendungsfeldern resultiert die Notwendigkeit kontinuierlicher Kontrolle schon daraus, dass sie ihr Verhalten im Laufe ihres Einsatzes häufig wie ein Chamäleon verändern – sei es durch Updates, sei es aufgrund eines maschinellen Lernverfahrens. Ein obsiegendes Urteil gegen eine diskriminierende softwarebasierte Entscheidung, das zwei Jahre nach der Rechtsverletzung Rechtskraft erlangt, ist dann in seiner Aussagekraft längst überholt.

a) Kontrollalgorithmen und Standardisierung maschineller Lernverfahren

Jedenfalls bei lernfähigen Softwareanwendungen lässt sich eine zuverlässige Rechtmäßigkeitskontrolle regelmäßig nur auf Grundlage eines Einblicks in die Lernmechanismen und die zugrunde liegende Datenbasis sowie die Ergebnisse erzielen, die sie hervorbringen – denn von außen sind die autonomen Abläufe kaum nachvollziehbar. Wer maschinelle Lernverfahren in persönlichkeitssensiblen Anwendungsfeldern zum Einsatz bringt, sollte daher fortlaufenden Kontrollpflichten unterliegen.[49] Die Verfahren müssen dafür Sorge tragen, dass eine regelmäßige Überprüfung der technischen und organisatorischen Maßnahmen stattfindet, die eine rechtmäßige algorithmenbasierte Entscheidungsfindung gewährleisten sollen. Der aufsichtsrechtliche Prüfradius muss sich insbesondere auf die Trainingsumgebung, die Validität der Testdaten und die Richtigkeit der Datenbasis lernfähiger Softwareanwendungen erstrecken.

Ein wichtiges Prüfwerkzeug können insoweit Kontrollalgorithmen sein, welche die Entscheidungsergebnisse einer Softwareanwendung systematisch analysieren. Sie folgen dem Leitmotiv: „An ihren Taten sollt ihr sie erkennen." Sie durchleuchten die Entscheidungsergebnisse eines lernfähigen Systems (bzw. einer Software, deren Programmcode nicht offen liegt) auf Auffälligkeiten, insbesondere Diskriminierungen – und nutzen dabei grundsätzlich die gleichen statistischen Mittel wie das überprüfte Verfahren, um herauszufinden, welche Faktoren der Algorithmus besonders stark gewichtet. Den berechtigten Geheimhaltungsinteressen der betroffenen Diensteanbieter lässt sich durch vertraulichkeitswahrende Instrumente, insbesondere In-camera-Verfahren, angemessen Rechnung tragen.[50]

b) Risikomanagementsystem und Veröffentlichung einer Risikoabschätzung

Wer in seine Softwareanwendung Algorithmen implementiert, die das Potenzial erheblicher Persönlichkeits-, insbesondere Diskriminierungsrisiken bergen, sollte grundsätzlich eine *Risikoprognose* erstellen müssen: Er muss dann analysieren und

[49] Auch hier sollte der Gesetzgeber jedoch lernfähige Softwareanwendungen, deren Schadenspotenzial unter einer (noch zu konkretisierenden) Erheblichkeitsschwelle zurückbleibt, von dem Pflichtenkanon (ggf. abgestuft) ausnehmen. Vgl. zum Merkmal der Persönlichkeitssensibilität auch bereits Fn. 32.
[50] Dazu bereits *Martini* (Fn. 13), 1485 f. Zu In-camera-Verfahren im Verwaltungsrecht *Neumann*, DVBl 2016, 473.

offenlegen, inwieweit das digitale System grundrechtlich geschützte Güter gefährdet und welche technischen, organisatorischen und rechtlichen Schutzmechanismen er vorsieht, um Rechtsverletzungen zu vermeiden.[51]

In persönlichkeitssensiblen Einsatzbereichen automatisierter Entscheidungsmechanismen, insbesondere im öffentlichen Bereich (z. B. in automatisierten Verwaltungsverfahren), sollte den Verwender auch die Pflicht treffen, *Risikomanagementsysteme*[52] zu implementieren. Deren Aufgabe ist es, sicherzustellen, dass – insbesondere lernfähige – Softwareanwendungen keine unvorhergesehenen fehlerhaften, insbesondere diskriminierenden Entscheidungen treffen. Im Falle einer Anknüpfung an besonders sensible Daten (z. B. die sexuelle Orientierung oder die ethnische Herkunft) oder strukturell erhöhtem Risiko mittelbarer Diskriminierungen können die Systeme dann eine (menschliche) Kontrolle der automatisierten Entscheidung auslösen.

Bisher kennen weder das geltende (Datenschutz-)Recht noch das künftige Regelungsregime der DSGVO eine Pflicht der Verantwortlichen, Risikomanagementsysteme in ihre Datenverarbeitungsprozesse zu integrieren oder die Ergebnisse einer Folgenabschätzung (Art. 35 DSGVO) zu veröffentlichen. Solche Verpflichtungen für (insbesondere lernfähige) Softwareanwendungen in persönlichkeitssensiblen Anwendungsfeldern normativ zu verankern, ist aber sachgerecht. Sie ermöglichen nicht zuletzt den Betroffenen, auf informierter Grundlage eine selbstbestimmte Entscheidung über die eigene Neigung zu treffen, sich den Risiken auszusetzen, die sich mit der Nutzung einer bestimmten Anwendung verknüpfen.

Bei der Auferlegung eines Risikomanagementsystems und einer Veröffentlichungspflicht für die Folgenabschätzung ist jedoch regulatorisches Augenmaß geboten: Mit Blick auf den mit ihnen verknüpften Ressourcenaufwand ist der Gesetzgeber aufgerufen, einen vertretbaren normativen Ausgleich zwischen wirtschaftlichen und persönlichkeitsrechtlichen Interessen herzustellen.

---

[51] Art. 35 I DSGVO schreibt ab dem 25.5.2018 (Art. 99 II DSGVO) für risikoreiche Verarbeitungen eine Datenschutz-Folgenabschätzung vor. Ihr Prüfradius ist jedoch – entsprechend dem Schutzauftrag der DSGVO (Art. 2 II DSGVO) – grundsätzlich auf den Schutz personenbezogener Daten verengt (der allerdings von zentraler Bedeutung ist).

[52] Zum Risikomanagementsystem im (automatisierten) Besteuerungsverfahren i. S. d. § 88 V S. 3 AO *Braun Binder*, DÖV 2016, 891 (895 f.); *Braun Binder*, NVwZ 2016, 960 (961 f.); *Martini/Nink* (Fn. 18), 8 ff.; *Schmitz/Prell*, NVwZ 2016, 1273 (1273 f.).

c) Protokollierung der Programmabläufe

Die Möglichkeit, etwaige Rechtsverstöße überprüfen, feststellen und beweisen bzw. widerlegen zu können, setzt eine wirksame Beweissicherung voraus – sowohl hinsichtlich der Modellierung der Softwareanwendung als auch der Entscheidungsparameter (und ggf. Lernschritte) der verwendeten Algorithmen. Um dies zu gewährleisten, sollte auch eine umfangreiche Protokollierung der Programmabläufe und Rückkopplungsprozesse Bestandteil eines begleitenden Risikomanagements algorithmenbasierter Verfahren sein.

Das Verzeichnis der Verarbeitungstätigkeiten, das Art. 30 DSGVO dem Verantwortlichen abverlangen wird, bleibt hinter diesen Anforderungen zurück: Es beschränkt sich auf Elementardaten (insbesondere die Kontaktdaten des Verantwortlichen, den Zweck der Verarbeitung sowie die Kategorien betroffener Personen und Empfänger).[53] Programmabläufe und Entscheidungsparameter gehören nicht dazu.

Gerade bei verteilten, hochgradig vernetzten oder lernfähigen Anwendungen kann eine umfassende Systemprotokollierung – sowie deren Auswertung – aber extrem aufwendig sein und sich schnell zu einer unverhältnismäßigen Belastung auswachsen. Die Protokollierungspflicht sollte sich daher in ihrer normativen Reichweite, ihrem Umfang und ihrer Tiefe nach der persönlichkeitsrechtlichen Sensibilität und Skalierungsintensität des Geschäftsmodells bemessen sowie Härtefallklauseln vorsehen.

3. Selbstregulierung: Algorithmic Responsibility Kodex mit Erklärungspflicht

Wer Softwareanwendungen in persönlichkeitssensiblen Anwendungsfeldern anbietet, verfügt nicht nur über überlegenen Sachverstand hinsichtlich der Risiken, die seine Systeme auslösen können. Er kennt am ehesten auch wirksame Problembewältigungsmechanismen zu ihrer Überwindung. Schon mit Blick auf die limitierten staatlichen Kontrollkapazitäten und die Komplexität sowie Dynamik moderner (insbesondere lernfähiger) algorithmenbasierter Systeme ist es erwägenswert, bei dem Versuch, die mit ihnen verbundenen Risiken einzudämmen, auch die Potenziale regulierter Selbstregulierung fruchtbar zu machen. Die Anbieter früh in den Regulierungsprozess und die Kontrollverantwortung einzubeziehen, steigert insbesondere ihre Akzeptanz späterer Handlungsrestriktionen sowie die Bereitschaft zur Regelbefolgung.

[53] Vgl. dazu *Hartung*, in: Kühling/Buchner (Hrsg.), DS-GVO, 2017, Art. 30, Rn. 16 ff.; *Martini*, in: Paal/Pauly (Hrsg.), DS-GVO, 2017, Art. 30, Rn. 2 ff.

Selbstverpflichtungen sind in der Digitalwirtschaft bislang nicht zum wirkmächtigen Erfolgsmodell aufgestiegen.[54] Regulierungsadressaten nutzen sie nicht selten als Feigenblatt: Sie streifen sich die Verkleidung ernsthaften Bemühens über, ohne aber mit ihrem öffentlichkeitswirksamen Versprechen ausreichende Anstrengungen zu verknüpfen oder verknüpfen zu wollen. Davon zeugen paradigmatisch die Bemühungen der EU-Kommission, *Facebook* im Wege der Selbstregulierung zu einem sach- und zeitgerechten Umgang mit Ehrverletzungen und rassistischen Äußerungen zu bewegen.

Die Regulierungserfahrung mit Selbstverpflichtungen zeigt: Ohne steuernde Anreize und Sanktionierungsmechanismen entfalten sie kaum die erhoffte Wirkung. Ein modifiziertes Modell – gleichsam eine Selbstregulierung „mit Zähnen" – kann dem guten Grundgedanken, Betroffene in den Regelvollzug einzubeziehen, womöglich zur wirksamen Entfaltung verhelfen. Als regulatorische Schleifsteine eignen sich insbesondere stärkere Mindestanforderungen an den Kodexinhalt sowie integrierte gesetzliche Erklärungspflichten.

Eine normative Blaupause kann insoweit das hybride Modell des Corporate Governance Kodex i. S. d. § 161 AktG liefern. Bei ihm handelt es sich nicht um staatlich gesetztes Recht. Er bündelt vielmehr das Erfahrungswissen eines privaten Gremiums,[55] das mit sachverständigen Wirtschaftsvertretern besetzt ist.[56] Dessen Expertise erlangt durch den Hebel des § 161 AktG mittelbar normative Wirkung: Der Gesetzgeber verpflichtet die Regulierungsadressaten dazu, jährlich zu erklären, ob sie den Empfehlungen des Kodex entsprechen – insbesondere welchen Empfehlungen sie nicht gefolgt sind und warum nicht.[57]

In Anlehnung an diese Grundphilosophie des „comply or explain" sollte der Gesetzgeber Anbieter besonders persönlichkeitssensibler, insbesondere lernfähiger Softwareanwendungen dazu verpflichten, sich zu einem Algorithmic Responsibility Kodex erklären zu müssen. Im Idealfall erarbeitet ihn eine Regierungskommission, die sich aus gewählten Vertretern verschiedener Gruppen rekrutiert. Ähnlich wie börsennotierte Unternehmen ihren Anteilseignern und potenziellen Investoren (im Interesse der Investitionstransparenz) offenbaren müssen, wie es um das wertschöpfungsrelevante Verhalten des Unternehmens bestellt ist, müssen dann Diensteanbieter, die Softwareanwendungen in persönlichkeitssensiblen Bereichen bereitstellen, Farbe bekennen: Sie müssen öffentlich erklären, ob sie den

---

[54] Dazu *Hoffmann-Riem* (Fn. 2), 39; aus datenschutzrechtlicher Sicht bspw. *Martini*, NVwZ-Extra 3/2016, 1 (9 f.); *Meltzian*, in: Wolff/Brink (Hrsg.), BeckOK DatenschutzR, 21. Ed., Stand: 1.8.2017, § 38a BDSG, Rn. 3; *Petri*, in: Simitis (Hrsg.), BDSG, 8. Aufl., 2014, § 38a, Rn. 16.

[55] Die Kommission heißt zwar (kraft der Berufung ihrer Mitglieder durch den Bundesminister der Justiz und für Verbraucherschutz) „Regierungskommission" (§ 161 I 1 AktG). Das ändert aber nichts daran, dass sie nicht in den Staatsaufbau integriert ist und sich aus Mitgliedern rekrutiert, die ihren Sachverstand (und ihre Interessen) als Private in das Gremium einbringen.

[56] *Hölters*, in: ders. (Hrsg.), AktG, 2. Aufl., 2014, § 161, Rn. 3.

[57] *Vetter*, in: Henssler/Strohn (Hrsg.), GesR, 3. Aufl., 2016, AktG § 161, Rn. 9.

Verhaltensregeln für einen ethisch verantwortbaren Einsatz algorithmenbasierter Verfahren entsprechen. Weichen sie von Grundregeln guter digitaler Ethik ab, die der Kodex formuliert, müssen sie sich jedenfalls dazu äußern, warum und inwieweit sie seinen Regeln nicht folgen.

Das daraus erwachsende öffentliche Bekenntnis zu Mindestgrundsätzen kann nicht nur zum Ausgleich struktureller Wissensasymmetrien zwischen Nutzern und Diensteanbietern beitragen. Die Selbstbindung und der Wahrheitsanspruch, die von der Erklärung ausgehen, üben vor allem im Idealfall einen Befolgungsdruck aus: Ein Algorithmic Responsibility Kodex hält Diensteanbieter präventiv zu einem verantwortungsvollen Umgang mit Grundregeln gesellschaftlich konsentierten Algorithmeneinsatzes an. Decken Kontrollmechanismen auf, dass das tatsächliche Verhalten des Unternehmens der öffentlichen Erklärung widerspricht, löst das im Idealfall nicht nur eine Sanktionswirkung des Marktes aus, mit der ein nachhaltiger Reputationsverlust bei Verbrauchern korrespondiert. Die Erklärungspflicht sollte ergänzend auch Aufnahme in die jeweiligen Kataloge der §§ 5, 5a UWG[58] finden. Sinnvoll kann es auch sein, an eine fehlende oder unrichtige Entsprechenserklärung unmittelbar die Drohung einer bußgeldbewehrten Sanktion zu knüpfen.

Ausnahmslos jedes Unternehmen, das Algorithmen in seine Produkte implementiert, einer Erklärungspflicht zu unterwerfen, schösse jedoch über das Ziel hinaus. Erst wenn Softwareanwendungen eine – gesetzlich näher festzulegende – kritische Grenze digitaler Wirkmacht und persönlichkeitsrechtlichen Risikos überschreiten, ist eine solche Erklärungspflicht angemessen. Dabei sollten auch die Größe des Unternehmens, die Skalierbarkeit seines Geschäftsmodells und die Reichweite seines Angebots Berücksichtigung finden.

Um ihre normative Mission in der digitalen Welt auch tatsächlich zu erreichen, darf die (im Idealfall unionsrechtlich verankerte) Pflicht zur Entsprechenserklärung – genau wie die anderen Pflichten, denen Diensteanbieter unterworfen sind – nicht allein an eine Niederlassung in Deutschland anknüpfen, sondern sollte dem Marktortprinzip folgen.[59]

4. Ex-post-Schutz

Wo die „Blackbox Algorithmus“ der Rechtsdurchsetzung des Verbrauchers im Wege steht, sollten auch das Haftungs- und das Prozessrecht auf die davon ausgehenden Wissensasymmetrien reagieren.

[58] § 5 I S. 1 Nr. 6 UWG nimmt de lege lata nur auf Verhaltenskodizes Bezug, welche *die Unternehmen* vereinbaren, nicht aber auf gesetzliche Erklärungspflichten zu einem *staatlich veranlassten* Kodex; vgl. *Bornkamm/Feddersen*, in: Köhler/Bornkamm/Feddersen (Hrsg.), UWG, 35. Aufl., 2017, § 5, Rn. 7.1.
[59] Für die gesamte digitale Ordnungspolitik auch *Spindler/Thorun*, MMR Beilage 2016, 1 (24).

a) Haftung

aa) Beweislastverteilung

Mangels Einblicks in die Entscheidungsvorgänge einer Softwareanwendung können Verbraucher Verletzungshandlungen, Kausalitätszusammenhänge, eine Verkehrssicherungspflichtverletzung und das Verschulden des Diensteanbieters nur schwer erkennen – geschweige denn beweisen. Diese Gefahr struktureller informationeller Schieflagen teilt die Haftung für Schäden im Gefolge des Einsatzes algorithmenbasierter Systeme – bspw. bei Auswahlentscheidungen wie der Kredit- oder Arbeitsplatzvergabe – mit der Arzt- und Produzentenhaftung:[60] Bei Behandlungs- und Produktfehlern steht der Geschädigte häufig in der Not, die Kausalität einer Handlung für einen eingetretenen Schaden nachzuweisen. Im Falle algorithmenbasierter Verfahren wird es dem Kläger häufig ebenso wenig gelingen, zu beweisen, dass ein Schaden – etwa eine Ungleichbehandlung – auf unzulässigen Entscheidungsparametern (z. B. dem Geschlecht) beruht, weil diskriminierende Faktoren Eingang in den Algorithmus gefunden haben. Ebenso schwierig ist der Nachweis mangelnder Sorgsamkeit bei der Programmierung der eingesetzten Softwareanwendung oder der Aufsicht über sie (also der Verkehrssicherungspflichtverletzung). Denn der Geschädigte kann nahezu keinen Einblick in diese internen Vorgänge des Verwenders algorithmengesteuerter Verfahren nehmen.

Angesichts dieser strukturellen Asymmetrie sollten Gerichte den Nutzern persönlichkeitssensibler Softwareanwendungen im Haftungsprozess – dem Gebot prozessualer Waffengleichheit folgend – mit einem abgestuften System der Beweislastverteilung entgegenkommen: Es genügt dann, dass er Tatsachen vorträgt, die mit überwiegender Wahrscheinlichkeit darauf schließen lassen, dass unzulässige Parameter Eingang in die Entscheidung gefunden haben.[61] Der Anbieter der Softwareanwendung muss dann den Gegenbeweis antreten, indem er sich durch Vorlage protokollierter Programmabläufe,[62] den Nachweis hinreichender Aufsicht über das eingesetzte technische Verfahren oder die anderweitige Erschütterung einer Kausalitätsvermutung von einer Rechtsverletzung freizeichnet. Für eine solche Exkulpation muss der Anbieter freilich – ggf. umfangreich – Interna (bis hin zu Teilen des Programmcodes [in camera][63]) offenlegen – oder haften. Bestehen Indizien dafür,

---

[60] Grundlegend zur Funktionsweise der Produzentenhaftung BGH, NJW 1969, 269 (273 ff.); vgl. auch *Wagner*, in: Säcker/Rixecker/Oetker et al. (Hrsg.), MüKo-BGB VI, 7. Aufl., 2017, § 823 BGB, Rn. 858 m. w. N. Zur Arzthaftung *Wagner* (Fn. 60), § 823 BGB, Rn. 909 f, 915. In den §§ 630a ff. BGB hat der Gesetzgeber die Rechtsprechung des BGH zur Arzthaftung teilweise kodifiziert. § 630h V BGB etabliert insbesondere die Beweislastumkehr für die haftungsbegründende Kausalität bei schweren Behandlungsfehlern.

[61] Vgl. für das AGG auch BAG, NZA 2012, 1345 (1348).

[62] Zur Forderung, die Protokollierung zur gesetzlichen Pflicht zu erheben, oben III. 2. c).

[63] Dazu bereits oben III. 2. a) mit Fn. 50.

dass der Fehler erkennbar war, erweitert sich die Pflichtenlast – und damit auch der Haftungsbereich des Betreibers.

bb) Gefährdungshaftung?

Von Softwareanwendungen geht zwar nicht generell eine Betriebsgefahr aus. In besonders sensiblen Einsatzbereichen, etwa digitalisierten medizinischen Anwendungen oder dem Einsatz von Pflegerobotern, ist aber nach dem Vorbild der Tierhalter-, Straßenverkehrs- und Arzneimittelhaftung eine Gefährdungshaftung für digital automatisierte Prozesse denkbar – jedenfalls dann, wenn sie besonders nachhaltige Schäden für gewichtige Rechtsgüter, insbesondere Leib und Leben, befürchten lassen.

Als Leitregel der haftungsrechtlichen Risikoverteilung gilt dann: Wer von (insbesondere lernfähigen) Softwareanwendungen profitiert, sollte auch für deren Fehler und Risiken einstehen müssen – auch und gerade wenn sich ein System emergent, d. h. nicht vorhersehbar, verhält. Damit der Geschädigte mit seinem Anspruch nicht in eine leere Tasche greift, sondern ein zahlungsfähiges Haftungssubjekt belangen kann, sollte der Gesetzgeber mit der Gefährdungshaftung eine Versicherungspflicht verknüpfen.

b) Erweiterung prozessualer Handlungsräume

aa) Abmahnbefugnisse für Wettbewerber

Um den Schutz der Verbraucher gegen unzulässige, aber zugleich undurchsichtige Softwareanwendungen wirksam durchzusetzen, kann und sollte sich der Staat die Wachsamkeit und Expertise der Wettbewerber zunutze machen: Sie verspüren regelmäßig einen ökonomischen Anreiz, rechtswidrige algorithmenbasierte Verfahren ihrer Konkurrenten zu unterbinden. Der Gesetzgeber sollte die Abmahnbefugnisse der §§ 12 I 1, 8 III Nr. 1, I i. V. m. § 5 I S. 1 und 2 Nr. 6 UWG bzw. i. V. m. § 3 III UWG daher um den Tatbestand diskriminierender oder sonst persönlichkeitsverletzender Softwareanwendungen erweitern.

bb) Verbandsklagerecht der Verbraucherverbände und spezialisierte Schiedsstelle für algorithmische Diskriminierungen und sonstige Persönlichkeitsverletzungen

Wer als Verbraucher zwar in seinen Rechten, dadurch aber nicht auch nachhaltig in seiner gesamten Lebensführung beeinträchtigt ist, ist geneigt, die finanziellen und zeitlichen Risiken, die ihm ein gerichtliches Verfahren abverlangt, nicht auf sich zu nehmen. Die Aussicht auf einen langen Rechtsstreit mit unsicherem Ausgang sowie der verfahrensrechtliche Formalismus gerichtlicher Prozesse schrecken schnell ab. Oftmals werden Betroffene es daher vorziehen, die Streuschäden einer Rechtsverletzung auf sich beruhen zu lassen. Der mit der Rechtsdurchsetzung verbundene Aufwand steht, aus

einer rational kalkulierenden prozessökonomischen Perspektive betrachtet, vielfach in keinem das Risiko rechtfertigenden Verhältnis. Der gesellschaftliche Nutzen solcher Kontrollprozesse ist häufig größer als der individuelle. Sachgerecht ist dann ein dem kollektiven Güterschutz verschriebener Verteidigungsmechanismus.[64]

Als „Schirmherr" des Verbraucherinteresses kann und sollte die Rechtsordnung daher – neben den Wettbewerbern – auch die Verbraucherverbände in die Normdurchsetzung gegen Diskriminierungen und sonstige Persönlichkeitsverletzungen einbeziehen. Die Schwelle, das Risiko eines gerichtlichen Verfahrens auf sich zu nehmen, liegt für spezialisierte Verbände aufgrund von Skaleneffekten regelmäßig niedriger als für den Einzelnen: Sie bündeln im Idealfall den für die Rechtsdurchsetzung erforderlichen technischen sowie rechtlichen Sachverstand. Dadurch können sie diskriminierende Muster wirksamer aufdecken sowie rechtswidrige Geschäftspraktiken mit einem Maßnahmenbündel aus juristischen und anderen öffentlichkeitswirksamen Mitteln adäquat bekämpfen. Es ist daher erwägenswert, das Verbandsklagerecht der Verbraucherverbände auf der Grundlage des UKlaG (vgl. v. a. § 2 Abs. 2 Satz 1 Nr. 11 und Satz 2) generell auf (insbesondere lernfähige) Softwareanwendungen in persönlichkeitssensiblen Anwendungsfeldern auszudehnen. Ein zugelassener Verband kann dann einzelfallunabhängig gegen eine rechtswidrige, insbesondere diskriminierende algorithmenbasierte Entscheidungsfindung vorgehen.[65]

Auch eine staatlich geförderte Schlichtungsstelle als Instanz der alternativen Streitbeilegung,[66] die für algorithmische Verfahrensgegenstände besonderen Sachverstand mitbringt, könnte die Initiierungsschwelle und Kosten einer Rechtsdurchsetzung für Verbraucher senken und dadurch die tatsächliche Rechtsbefolgung verbessern.

cc) Nebenfolgen-Kompetenz für Zivilgerichte

Will der Gesetzgeber die Bevölkerung vor den Risiken eines persönlichkeitsverletzenden, insbesondere diskriminierenden, oder wettbewerbswidrigen Algorithmeneinsatzes wirksam schützen, kann er auch an der Stellschraube „Reichweite der Rechtskraft gerichtlicher Urteile" drehen. Die Erträge umfangreicher rechtsstaatlicher Beweisaufnahmen, die Verfahren vor den Zivilgerichten bereits hervorgebracht haben, sollte das Prozessrecht über die grundsätzlich *inter partes* wirkende Rechtskraft des Zivilprozessrechts (vgl. § 325 I ZPO) hinaus auch Dritten zugänglich machen: Der Gesetzgeber könnte eine Rechtskrafterstreckung

[64] Vgl. auch *Golla* (Fn. 27), S. 204 f.
[65] Vgl. zu anderen Verbandsklagerechten bspw. § 3 UKlaG, §§ 63 f. BNatSchG, § 15 BGG, § 8 III Nr. 2–4 UWG; zur Erweiterung des Verbandsklagerechts im Datenschutzrecht bspw. *Halfmeier*, NJW 2016, 1126 ff.; *Spindler*, ZD 2016, 114 ff.
[66] Vgl. zu anderen Schlichtungsstellen in der Rechtsordnung, bspw. § 214 VVG, *Looschelders*, in: Langheid/Wandt (Hrsg.), MüKo VVG, 2. Aufl., 2016, § 214, Rn. 1 ff.; *Müller*, VuR 2010, 259 ff.

verfügen (vgl. bspw. auch § 11 UKlaG) oder den Zivilgerichten eine Nebenfolgenkompetenz zuerkennen (soweit sich das mit dem Prinzip der formellen Wahrheit im Zivilprozess in Einklang bringen lässt). Sie könnten womöglich z. B. im Anschluss an ein AGG-Verfahren *erga omnes* wirkende Unterlassungspflichten des Anbieters einer nachgewiesen diskriminierenden oder sonst rechtswidrig operierenden Softwareanwendung aussprechen. Erreichen ließe sich dieses prozessökonomische Ziel bspw. auch durch eine Erweiterung des § 23 AGG auf ein Streitbeitrittsrecht für Verbraucherverbände, die dann wiederum im Gerichtsverfahren einen Antrag stellen könnten, die streitgegenständliche algorithmische Praxis zu unterlassen.

IV. Fazit

Wie die Algorithmen funktionieren, denen wir die Organisation unseres Lebens immer stärker anvertrauen, verstehen wir immer weniger. Wie aber wir funktionieren, verstehen umgekehrt die Algorithmen im Maschinenraum moderner Softwareanwendungen immer besser. Auch in einem anbrechenden digitalen Maschinenzeitalter ist der grundrechtsverpflichtete Staat aufgerufen, die Autonomie des Einzelnen gegen Persönlichkeitsbeeinträchtigungen abzuschirmen, die von zunehmend autonom agierenden Systemen ausgehen. Seinem Handlungsauftrag entspricht es, die vielfältigen Einsatzmöglichkeiten moderner (insbesondere lernfähiger) Softwareanwendungen mit einem wirksamen Regelungssystem zu umhegen: Er hat dafür Sorge zu tragen, dass die ethischen Grundwerte unserer Gesellschaft auch für das Wirken automatisierter Systeme gelten.

Ähnlich wie bei anderen Risikotechnologien, etwa der Gen- oder der Nanotechnologie, ist angesichts des unbestreitbaren gesellschaftlichen Mehrwerts algorithmenbasierter Verfahren weder ein reflexartiges Verbot sachgerecht, noch hilft (schon aufgrund der Besonderheiten maschinellen Lernens) eine generelle Pflicht, den Quellcode offenzulegen, der Problematik generell ab; sie schösse mit Blick auf die berührten Betriebs- und Geschäftsgeheimnisse vielmehr über das Ziel hinaus. Der Gesetzgeber sollte sorgsam darauf achten, die Schraube der Regulierung nicht zu überdrehen, um das Innovationspotenzial, das in den Zukunftsbereichen „Künstliche Intelligenz“ und „Big Data“ steckt, nicht auszubremsen: Entwicklungsoffene Start-up-Strukturen sollte er nicht mit einem Regulierungsinstrumentarium überziehen, das ihrer Innovationskraft den erforderlichen Entfaltungsspielraum nimmt. Die Regulierungsintensität muss vielmehr mit der Leistungsfähigkeit der Unternehmen und ihrer Größe korrespondieren.

Zu den geeigneten Regulierungsinstrumenten eines sowohl individuellen als auch kollektiven Schutzes der informationellen Selbstbestimmung gegen die Zauberformeln algorithmischer Alchemie gehören neben einer Kennzeichnungs- und Begründungspflicht für algorithmenbasierte Entscheidungsverfahren und (soweit technisch abbildbar) wirksamen aufsichtsrechtlichen Kontrollmöglichkeiten

(„Algorithmen-TÜV“) eine Pflicht, ein Risikomanagementsystem zu implementieren sowie eine Risikoabwägung durchzuführen und zu veröffentlichen. Auch ein Kodex für den ethischen Umgang mit Algorithmen und maschinellem Lernen (ein „Algorithmic Responsibility Kodex“ nach dem regulatorischen Vorbild des Corporate Governance Kodex), wettbewerbliche Abmahnbefugnisse sowie eine spezialisierte Schiedsstelle können sinnvolle Elemente eines Regulierungsbündels sein.

Im Idealfall sollte der Gesetzgeber keine nationale Insellösung ansteuern, sondern eine einheitliche unionale Regulierungsstrategie verfolgen. In zahlreichen Bereichen – allen voran dem Datenschutzrecht (Art. 16 II 1 AEUV) – fehlt dem nationalen Gesetzgeber ohnedies die Freiheit, als regulierungspolitischer *Robinson Crusoe* nationale Alleingänge zu unternehmen. Bislang sind die Vorahnungen eines Missbrauchs neuer technischer Möglichkeiten durch die Landnahme algorithmischer Invasoren verbreiteter als die tatsächlichen Rechtsverletzungen. So wenig die Gefährdungslagen nur eine Fata Morgana sind, so sehr sollte der Gesetzgeber in der sich immer schneller drehenden digitalen Welt aber auch nicht allein die Furcht vor einem Phänomen zum Leitmotiv seines Handelns machen. Bei der Regulierung dystopisch erscheinender Zukunftsszenarien der Digitalisierung lehrt ihn vielmehr die Weisheit des Schiffbrüchigen *Crusoe*, Besonnenheit und Augenmaß zu bewahren: „Furcht vor Gefahr“ – so lässt *Daniel Defoe* seinen Romanhelden ausrufen – „ist zehntausendmal beängstigender als die Gefahr selbst“.