Daten-Fairness als Daten-Gerechtigkeit by Design

Felix Bieker und Marit Hansen

Zusammenfassung

Dieser Beitrag nimmt die aktuelle Debatte um Chatbots zum Anlass die Risiken algorithmischer Systeme für betroffene Personen und insbesondere Angehörige marginalisierter Gruppen zu untersuchen. Letztere sind vom Einsatz neuer Technik meist überproportional negativ betroffen, obwohl die gesellschaftlichen Machtstrukturen, die Rassismus, Sexismus und Transphobie den Weg bereiten, längst umfassend untersucht sind. Wir analysieren Regelungen des Datenschutzrechts, die Strukturen von Informationsmacht adressieren, blicken auf relevante Regelungen der neuen EU-Datengesetze, ziehen Schlussfolgerungen aus dem Diskurs zum Antidiskriminierungsrecht und erweitern durch den Rückgriff auf das Konzept der Intersektionalität die Perspektive des Rechts. Auf Grundlage von Ansätzen zu *Value Sensitive Design*, *Data Justice* und *Design Justice* zeigen wir mit Bezug insbesondere auf die relevanten datenschutzrechtlichen Regelungen, wie Daten-Gerechtigkeit „by Design" durch Prozesse gewährleistet werden kann.

1. Einführung

Die Bereitstellung von OpenAIs Chatbot ChatGPT zur allgemeinen Nutzung ab Ende 2022 befeuerte den über längere Zeit beständig aufgebauten Hype um sogenannte Künstliche Intelligenz (KI). In der Folge stellten auch die großen Plattformanbieter ihre eigenen Chatbots und verwandte Anwendungen vor. Die mediale und gesellschaftliche Aufmerksamkeit richtete sich dabei zunächst auf die zukünftigen Potenziale und möglichen Anwendungsfelder dieser Technik. Doch zuletzt mehrten sich Berichte über Fehlfunktionen und Ausfälle von Chatbots verschiedener Hersteller. Von ChatGPT, das rassistische Rap-Texte und Computer-Programme ge-

nerierte[1], über Googles Bard, das bei seiner Vorstellung im Brustton der Überzeugung falsche Antworten gab[2], zu einer Test-Version von Microsofts Suchmaschine Bing als Chatbot, die Nutzende davon überzeugen wollte, dass es noch das Jahr 2022 sei[3] oder auch, dass sie ihre Frau verlassen sollten, um eine Beziehung mit dem Chatbot zu führen[4]. Diese Probleme sind keineswegs neu, sie zeigten sich z. B. in ähnlicher Form bereits bei den Bildgeneratoren, die zuvor vorgestellt wurden.[5]

Mit den auf *Large Language Models* basierenden öffentlich zugänglichen Chatbots ist der bisherige Höhepunkt einer Entwicklung erreicht, in der Technik als geheimnisvolle Black Box betrachtet wird, in die eine Vielzahl ungefilterter und auch personenbezogener[6] Daten, oft aus dem Internet, hineingegeben werden und die auf für die Nutzenden meist nicht nachvollziehbare Weise Outputs erzeugt. Wie genau diese Modelle arbeiten, wird dabei nicht dokumentiert oder, soweit dies doch geschieht, nicht transparent offengelegt. Die Modelle werden wahlweise als „magische Instrumente"[7] oder „menschenähnliche Intelligenzen"[8] beschrieben, was in beiden Fällen nicht zutrifft.[9] Das Vorgehen der Anbieter:innen erinnert bei genauerer Betrachtung eher an einen Wursthersteller, der ein wundersames neues Produkt bewirbt, jedoch nicht offenlegt, welche Zutaten darin verarbeitet wurden, ob diese aus zuverlässigen Quellen stammen und auf welche Weise die Herstellung vonstatten geht.

Mit dieser Entwicklung entfernen sich die technische Realität und die verfügbaren Angebote immer weiter von der gesetzlich normierten Vorstellung, dass Technik und die damit einhergehende Datenverarbeitung kontrolliert eingesetzt werden sollen, damit ihre Auswirkungen untersucht und gesteuert werden können. Diese Entwicklung wird durch die schiere Infor-

1 *Perrigo*, Time v. 5. Dez. 2022; *Biddle*, The Intercept v. 8. Dez. 2022.
2 *Schräer*, Heise Online v. 9. Feb. 2023.
3 https://www.reddit.com/r/bing/comments/110eagl/the_customer_service_of_the_new_bing_chat_is/.
4 *Roose*, New York Times v. 16. Feb. 2023.
5 Vgl. z.B. *Johnson*, Wired v. 5. Mai 2022; *Luccioni u.a.*, Stable Bias: Analyzing Societal Representations in Diffusion Models.
6 *Gal*, The Conversation v. 8. Feb. 2023.
7 *Elish*, Don't Call AI "Magic".
8 *Schwartz*, The Guardian v. 25. Jul. 2018.
9 *Bender*, The Guardian v. 14 Jun. 2022. Vgl, auch schon Clarke, Profiles of the Future: An Inquiry into the Limits of the Possible, 1973: „Any sufficiently advanced technology is indistinguishable from magic".

mationsmacht,[10] die Anbieter:innen durch diese Datenpraktiken erlangen können, begünstigt.

Nach unserem Verständnis bedarf es einer umfassenden Betrachtung der Verarbeitung von Daten und ihrer unerwünschten Folgen, die sich nicht nur auf das, was klassischerweise als Datenschutzrecht betrachtet wird, beschränken darf. Vielmehr ist der Blick holistisch und interdisziplinär auf die Informationsmacht und sämtliche unerwünschten Folgen von Datenpraktiken zu lenken – es wäre naiv anzunehmen, dass allein die Datenschutz-Grundverordnung (DSGVO) und ihre Durchsetzung über die Datenschutzaufsicht all diese Probleme lösen könnten. Gleichzeitig ist die DSGVO der rechtliche Ausgangspunkt für jede Verarbeitung personenbezogener Daten. Über das Konzept der Risiken für die Rechte und Freiheiten natürlicher Personen[11] und das Ziel nach Art. 1 Abs. 2 DSGVO, die Grundrechte natürlicher Personen zu schützen, ist sie für viele weitere Gebiete der Technikregulierung anschlussfähig. So greift etwa der aktuelle EU-Entwurf zur Verordnung über Künstliche Intelligenz (KI-VO-Entwurf)[12] Risiken u.a. auch für die Grundrechte auf.[13]

Die Jahreskonferenz 2022 des *Forum Privatheit* wählte mit dem Begriff der Daten-Fairness einen bewusst weiten Begriff, der viele Interpretationsmöglichkeiten eröffnet. Fragen nach Fairness verhandeln stets Formen von Gerechtigkeit, die wir in diesem Beitrag analysieren, in dem wir unser Verständnis von Daten-Fairness als Daten-Gerechtigkeit darlegen.

Im Folgenden erläutern wir zunächst die bekannten Probleme algorithmischer Systeme, die schon vor einiger Zeit dadurch aufgefallen sind, dass sie rassistische, sexistische und transphobe Inhalte generieren (Abschn. 2) und identifizieren die Form des Diskurses, der in zweifacher Hinsicht von Individualisierung geprägt ist, als hinderlich für das Erreichen von Daten-Gerechtigkeit: Zum einen wird der Blick auf einzelne Schuldige gerichtet, zum anderen von Einzelnen erwartet, diese machtvolle Anbieter:innen mit Hilfe ihrer Individualrechte zur Rechenschaft zu ziehen (Abschn. 3). In

10 Rouvroy/Poullet, in: Gutwirth u.a. (Hrsg.), Reinventing Data Protection?, 2009, 45 (69).

11 *Bieker*, DuD 2018, (27-31).

12 Der aktuelle Entwurf des Rates mit neuen Regelungen zu „General Purpose"-Anwendungen ist abrufbar unter: https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf. Für einen Überblick über den KI-VO-Entwurf, vgl. *Guijarro Santos*, ZfDR 2023, 23.

13 Vgl. ErwGr. 32, 42, 43, 47 72a, 79a, Art. 6 Abs. 3, Art. 7 Abs. 1 Buchst. b, Abs. 2, Abs. 3 Buchst. a, Art. 9 Abs. 2 Buchst. a, Art. 13 Abs. 3 Buchst. b (iii), Art. 14 Abs. 2, Art. 36 Abs. 6 Buchst. e, Abs. 7 Buchst. a, Abs. 8 Buchst. a, Art. 43 Abs. 6, Art. 65 Abs. 1, Abs. 2, Art. 67 Abs. 1 u. Annex IV Nr. 3.

Abschn. 4 blicken wir auf die rechtlichen Regelungen, die sich mit Informationsmacht und strukturellen Aspekten von Datenverarbeitung befassen, insbesondere das Datenschutz- und Informationsfreiheitsrecht, und zeigen Bezüge zum Antidiskriminierungsrecht auf. Unter Rückgriff auf die im bestehenden Recht enthaltenen systemischen Regelungen, ergänzt um die Ansätze von *Value Sensitive Design*, *Data Justice* und *Design Justice* zeigen wir, inwieweit sich Daten-Gerechtigkeit im bestehenden Recht finden lässt (Abschn. 5). Schließlich unterbreiten wir in Abschn. 6 unter Rückgriff auf die relevanten rechtlichen Regelungen und vorherigen Erkenntnisse Vorschläge für Prozesse, mit denen Daten-Gerechtigkeit in der Praxis erreicht werden kann. Abschn. 7 gibt einen Überblick über die Erkenntnisse und endet mit einem Ausblick.

2. Neue Technik, alte Probleme

Die oben aufgezeigten, anekdotischen Berichte der Presse offenbaren grundlegende Beschränkungen und systemische Probleme dieser Technik. Allerdings wird der weitere Schritt zu einer Betrachtung dieser strukturellen Ebene meist nicht vollzogen, sondern es bleibt beim Aufzeigen einer Vielzahl von „Einzelfällen". Dies macht es unkritischen Befürwortern leicht, Risiken im Vergleich zu – rein hypothetischen – Vorteilen und den Potenzialen des Einsatzes der Technik kleinzureden.[14]

Dabei würde eine tiefergehende Analyse wenig Überraschendes zutage fördern. Sie würde zeigen, dass die seit Jahren bestehenden Probleme von Technik als Black Box, die von systemischem Rassismus, Sexismus oder Transphobie durchzogen ist und die vor dem teilweise überstürzten, teilweise (bewusst) unfertigen Verbreiten nicht ausreichend getestet und dokumentiert wurde, im Fall dieser speziellen algorithmischen Systeme in besonderem Maße bestehen. Diese und weitere Probleme waren schon vor der Veröffentlichung der Chatbots absehbar.[15] Die Beseitigung dieser sich, wie Simone Brown[16] und Mar Hicks[17] aufgezeigt haben, beständig wiederholenden gesellschaftlichen Probleme und machtvollen Strukturen in

14 *Altman*, Planning for AGI and Beyond.
15 *Bender u.a.*, in: FAccT '21, 2021.
16 *Browne*, Dark Matters, 2015.
17 *Hicks*, in diesem Band; vgl. bereits *Hicks*, in: Mullaney u.a. (Hrsg.), Your Computer is on Fire, 2021, 11 (13).

verschiedenen Gestalten wird von den Verantwortlichen aufgeschoben und an betroffene Personen und Gruppen ausgelagert, damit die Unternehmen vom Boom der Anwendungen profitieren und ihre Monetarisierungsstrategien umsetzen können. Sie spielen die Informationsmacht[18] aus, die sich aus der eingesetzten Technologie, der Marktstellung ihrer Unternehmen, einer fehlenden oder mangelhaften Regulierung und den aus der massenhaften Verarbeitung von Daten gewonnenen Erkenntnissen ergibt.

Die Versprechen der Anbieter:innen über die Effizienz algorithmischer Systeme orientieren sich dabei an fragwürdigen Benchmarks[19], und in der praktischen Umsetzung wird teilweise nicht einmal danach gefragt, ob eine Anwendung überhaupt die gewünschte Funktionalität bietet.[20] Dies hängt auch damit zusammen, dass die Zwecke der Datenverarbeitung durch die Anwendung nicht im Vornherein klar festgelegt sind. Diese „General Purpose"-Anwendungen werden vielmehr frei zugänglich gemacht, sodass sich gar nicht absehen lässt, wofür die für ihre Bereitstellung und die im Rahmen ihrer Nutzung gesammelten Daten eingesetzt werden.

Während im EU-Gesetzgebungsprozess zur KI-Verordnung noch an Details gefeilt wird, setzt die Realität schon Fakten – und dabei ist noch nicht einmal die grundlegende Entscheidung getroffen, wie diese algorithmischen Systeme nach dem Willen des Gesetzgebers in das im KI-VO-Entwurf vorgesehene System von Risikokategorien eingestuft werden sollen.[21]

In den USA hat die Federal Trade Commission inzwischen angekündigt, Werbeversprechen der Anbieter kritisch in Hinblick darauf zu überprüfen, ob die Fähigkeiten der algorithmischen Systeme übertrieben dargestellt werden, den Anbietern die Risiken des Einsatzes bekannt sind und ob in einem bestimmten Produkt überhaupt ein solches System enthalten ist.[22] Während insbesondere der letzte Punkt überraschend wirken mag, gab es bereits verschiedene Fälle, in denen Aufgaben, die vorgeblich durch algorithmische Systeme automatisiert wurden, tatsächlich von menschlichen Crowdworkern zu Niedriglöhnen erledigt wurden.[23] Allerdings gilt auch

18 *Rouvroy/Poullet*, in: Gutwirth u.a. (Hrsg.), Reinventing Data Protection?, 2009, 45 (69).

19 *Raji u.a.*, in: Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks 1 (NeurIPS Datasets and Benchmarks 2021), 2021.

20 *Raji u.a.*, in: FAccT '22, 2022.

21 *Volpicelli*, Politico v. 3. März. 2023; *Heidelberger/Diakopoulos*, Internet Policy Review 2023.

22 *Atleson*, Keep your AI claims in check.

23 *Roberts*, in: Mullaney u.a. (Hrsg.), Your Computer is on Fire.

beim tatsächlichen Einsatz algorithmischer Systeme, dass die zugrundeliegenden Trainingsdaten von Menschen als unsichtbare Arbeiter:innen – und oft unter besonders schlechten Arbeitsbedingungen[24] – von gewaltvollen, rassistischen, sexistischen, transphoben oder anderen unerwünschten Inhalten bereinigt werden sollen.[25] All diese Probleme, die von solchen Datenpraktiken ausgehen und Probleme in der Umsetzung des Datenschutzrechts offenbaren, gleichzeitig aber auch eindeutig über die Materie der herkömmlichen Datenschutzkonzepte hinausgehen, zeigen sich in der aktuellen Entwicklung von Chatbots, Bildgeneratoren[26] und ähnlichen Anwendungen wie unter einem Brennglas.

Die Bezüge zwischen diesen Datenpraktiken und den bestehenden, problematischen Machtstrukturen, insbesondere im Hinblick auf algorithmische Systeme, sind bereits umfassend wissenschaftlich aufgearbeitet worden. Dies betraf etwa Fälle rassistischer, sexistischer und transphober Diskriminierung: Joy Buolamwini und Timnit Gebru haben schon 2018 nachgewiesen, dass die Gesichtserkennungsfunktionalität algorithmischer Systeme die Gesichter Schwarzer Menschen schlechter erkennt als die weißer Personen und dass die Erkennungsrate bei Schwarzen[27] Frauen besonders schlecht ist.[28] Ein von der Stadt Rotterdam eingesetztes algorithmisches System zur Bestimmung des Missbrauchspotenzials von Sozialhilfeempfänger:innen, bewertete das Risiko bei Frauen und People of Color automatisch höher als bei weißen Männern.[29] Ebenso wie bei Schwarzen

24 Vgl. dazu auch den Entwurf einer Richtlinie zur Plattformarbeit der Kommission (https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:52021PC0762) und die Position des Rates (https://data.consilium.europa.eu/doc/document/ST-10107-2023-INIT/de/pdf) sowie *Veale/Silberman/Binns*, European Labour Law Journal 2023.

25 Ebd.; *Jones*, Work without the Worker, Labour in the Age of Platform Capitalism, 2021; *Gray/Suri*, Ghost Work, 2019.

26 *Luccioni u.a.*, Stable Bias: Analyzing Societal Representations in Diffusion Models.

27 Wir schreiben das Wort Schwarz groß, da es sich dabei nicht um die tatsächliche Farbe der Haut oder eine angebliche biologische „Rasse" handelt, sondern „um eine politische Selbstbezeichnung von Menschen in einer bestimmten gesellschaftlichen Position, die mit Rassismuserfahrungen verbunden ist", https://www.journalist.de/startseite/detail/article/das-diversity-lexikon; vgl. auch: *Hasters*, Was weiße Menschen über Rassismus nicht hören wollen aber wissen sollten, 2020, S. 29 f.; aus juristischer Sicht: *Liebscher,* Rasse im Recht – Recht gegen Rassismus, 2021.

28 *Buolamwini/Gebru*, in: Proceedings of Machine Learning Research 81.

29 *Constantaras u.a.*, Wired v. 6. Mar. 2023, vgl. schon *Eubanks*, Harper's Magazine v. Jan. 2018.

Menschen[30] gibt es auch bei trans* Menschen eine lange Geschichte der Nutzung von Datenverarbeitung zu Überwachungszwecken[31]. Wie Mar Hicks und Simone Browne darlegen, sind algorithmischer Bias gegen trans* Frauen und Schwarze Menschen also keine neuartigen Phänomene.[32]

In diesen Aufarbeitungen wird auch deutlich, dass die entsprechenden Personen meist aufgrund mehrerer Kategorisierungen – insbesondere Geschlecht, Rassifizierung und Geschlechtsidentität, aber auch Klasse oder Behinderung[33] – von Diskriminierung betroffen sind. Dieses Phänomen wird vom Konzept der Intersektionalität erfasst, nach dem diese Kategorisierungen sich nicht gegenseitig ausschließen, sondern aufeinander aufbauen und zusammenwirken.[34] Die bereits aufgezeigten Beispiele verdeutlichen, dass die Probleme der aktuellen Datenpraktiken sich nicht nur in einer Dimension von Unterdrückung auswirken, sondern eben etwa Schwarze Frauen oder trans* Frauen, die in mehrere Kategorisierungen fallen, besonders von Diskriminierung betroffen sind. Um die strukturellen Probleme der aktuellen Datenpraktiken anzugehen und nicht nur immer wieder über vermeintliche Einzelfälle zu reden, ist es daher an der Zeit, diese Erkenntnisse auch in der rechtlichen Praxis zu berücksichtigen.

3. Individualisierung von Problemen

Für den notwendigen Schritt in die Praxis ist häufig die Form des Diskurses hinderlich: So findet oft eine Verengung auf einzelne „Bad Actors" oder „Bad Tech" statt, als würde das Problem sich durch das Ansetzen an einer Stelle lösen lassen, wie Anna Lauren Hoffmann in ihrer Kritik des Begriffs der Fairness im Antidiskriminierungsrecht herausgearbeitet hat.[35]

Dies lässt sich auch im Datenschutzdiskurs, zum Beispiel bei der Fokussierung auf einzelne Unternehmen, erkennen. Es wird über konkrete Anbieter berichtet und im selben Zug das weitere Feld einer gesamten Art von

30 Vgl. etwa *Gilliard/Culik*, Digital Redlining, Access, and Privacy.
31 *Hicks*, IEEE Annals of the History of Computing 2019.
32 *Hicks*, in diesem Band; *Hicks*, in: Mullaney u.a. (Hrsg.), Your Computer is on Fire, 2021, 11 (13); Browne, Dark Matters, 2015.
33 Es ist in diesem Kontext bemerkenswert, dass ein kolumbianischer Richter ausgerechnet in einem Fall zur Krankenversicherung eines autistischen Kindes auf ChatGPT zurückgriff, vgl. *Taylor*, The Guardian v. 3. Feb. 2023.
34 *Hill Collins/Bilge*, Intersectionality, 2020, S. 2.
35 *Hoffmann*, Information, Communication & Society 2019, 900 (903-905).

Anwendungen ausgeblendet. Das Problem dieses Fokus auf „Bad Actors" ist, dass es den Blick primär auf die schuldigen Individuen lenkt, die eine konsentierte gesellschaftliche Regel brechen und deren unangemessenes Verhalten zu beseitigen sei.[36] Dies geht zulasten der Problematisierung sozialer und systemischer Ungerechtigkeiten, da in der rechtlichen Debatte sodann nur über Schuld und Verursachung Einzelner diskutiert wird.

Dieses problematische Framing beschränkt sich nicht allein auf menschliche Verursacher: Genauso werden teilweise unbeabsichtigte Biases angeführt, die sich in Systeme „einschleichen". So ist zu beobachten, dass sich Anbieter für entstandene Schäden, zum Beispiel durch eine Gesichtserkennungssoftware, die Schwarze Menschen als Gorillas kennzeichnet, entschuldigen und Besserung geloben, obwohl der Eintritt der Schäden absehbar war und vermutlich bewusst in Kauf genommen wurde.[37] Inzwischen gibt es sogar Anbieter, die sich im Vornherein für den Output des von ihnen eingesetzten Chatbots entschuldigen.[38] Ob – und wenn ja, welche – Maßnahmen zum Vermeiden einer unerwünschten Funktionalität getroffen wurden und warum man sich trotz offensichtlich mangelnder Beherrschbarkeit des selbst produzierten algorithmischen Systems für eine Bereitstellung des Angebots entschieden hat, wird in der Regel nicht thematisiert.

Teilweise wird vorgeschlagen, das oft (fast) ausschließlich weiß und männlich besetzte Entwicklungsteam diverser zu besetzen.[39] Allerdings darf als Abhilfemaßnahme nicht nur die Beseitigung weißer Flecken innerhalb des Entwicklungsteams gefordert werden, da dies Systemfehler auf ein individuelles Problem seiner imperfekten menschlichen Gestalter:innen reduziert.[40] So können pauschale Forderungen nach „mehr Diversität" benutzt werden, um die zugrundeliegenden strukturellen Probleme zu überdecken.[41] Hoffmann fordert stattdessen, der Mentalität, dass man nur eine einzelne Quelle von Problemen beseitigen müsse, entgegenzuwirken, damit die dahinter liegenden systemischen Ungerechtigkeiten adressiert werden können.[42] Unter diesen Voraussetzungen könnten auch die relevanten Akteure identifiziert werden, damit die Technologiefirmen sich nicht

36 *Freeman*, Minnesota Law Review 1978, 1049 (1053-1054).
37 *Mac*, New York Times v. 3. Sep. 2021.
38 *Shanklin*, Engadget v. 27. Feb. 2023.
39 *Hoffmann*, Information, Communication & Society 2019, 900 (904).
40 *Ebd.*
41 *Theilen u.a.*, Internet Policy Review 2021.
42 *Hoffmann*, Information, Communication & Society 2019, 900 (904-905).

kollektiv durch Verweise auf unbeabsichtigte, dem Menschen inhärente Biases der Verantwortung entziehen können. Schließlich verstellt die Frage nach Biases auch den Blick auf ein grundlegenderes Problem: die Frage, ob eine bestimmte Technologie überhaupt eingesetzt werden sollte.[43] Diese Frage fokussiert auf die Risiken und strukturellen Probleme des Einsatzes bestimmter Technik, auf die wir in Abschn. 6 eingehen.

Neben dieser Individualisierung der Quelle von Problemen - bei der das Recht zu einer Verengung des Blicks führen kann und so die weiteren Folgen nicht ausreichend berücksichtigt werden - wird auch die Umsetzung des geltenden Rechts individualisiert, indem die Erwartung formuliert wird, dass betroffene Personen mit Hilfe der datenschutzrechtlichen Betroffenenrechte die Datenpraktiken global agierender Plattformen aufbrechen.[44]

Zwar sind die Betroffenenrechte und weitere Regelungen des Individualdatenschutzes ein wichtiger Teil des Datenschutzrechts[45], dieses enthält jedoch auch zahlreiche Regelungen zum Systemdatenschutz[46], wie wir sie im folgenden Abschnitt vorstellen. Im Angesicht der strukturellen Natur der Ursachen versprechen solch systemische Ansätze, auch weil damit verbundene by-Design-Ansätze bereits vor der Verarbeitung ansetzen, größeren Erfolg. Dass einzelne Organisationen auch über die individuellen Rechte des Datenschutzrechts strukturelle Probleme von Datenpraktiken aufzeigen und in Teilen sogar weitreichende Folgen auslösen können[47], steht dieser Schlussfolgerung nicht entgegen. Vielmehr zeigen der Grad der erforderlichen Organisation und die erheblichen Ressourcen[48], die für solche Formen der Rechtsverfolgung notwendig sind, wie unrealistisch es ist, dass eine einzelne, bereits von den negativen Auswirkungen einer Datenverarbeitung betroffene Person diesen Strukturen begegnen könnte.

Zudem können die Betroffenenrechte erst geltend gemacht werden, wenn die Datenverarbeitung bereits erfolgt. In einem solchen Fall, sind die Rechte der betroffenen Personen jedoch bereits verletzt. Eine Verkürzung

43 *Powles/Nissenbaum*, Medium v. 7. Dez. 2018.

44 Vgl. auch *Matzner u.a.,* in: Gutwirth u.a. (Hrsg.): Data Protection on the Move: Current Developments in ICT and Privacy/Data Protection. 2016, S. 277-305.

45 Vgl. *Kaminski*, Notre Dame Law Review Reflection 2022, 385.

46 *Bieker*, The Right to Data Protection 2022 (186 f.).

47 EuGH, Rs. C-362/14, Urteil v. 6. Okt. 2015, ECLI:EU:C:2015:650 – Schrems I; EuGH, Rs. C-311/18, Urteil v. 16. Jul. 2020, ECLI:EU:C:2020:559 – Schrems II.

48 noyb, Annual Report 2021, S. 20. URL: https://noyb.eu/sites/default/files/2022-07/ANNUAL%20REPORT%202021%2014072022%20interactive.pdf.

des Datenschutzrechts auf die ex-post-Durchsetzungsbemühungen einzelner betroffener Personen oder, im Wege der Verbandsklage[49], auch von Einzelnen, die sich von einer Organisation vertreten lassen, dient also nicht deren „Empowerment", sondern liefert diese der Informationsmacht der großen Plattformen aus.

4. Bestehende gesetzliche Regelungen

Mit ihrer Datenverarbeitung schaffen die Anbieter einseitig Risiken, die sich nicht nur auf die betroffenen Menschen und einzelne Gruppen, sondern auch auf die gesamte Gesellschaft auswirken können.[50] Dem entgegenzutreten, ist auf individueller Ebene meist schon unmöglich, da das Individuum, dessen Daten gesammelt wurden, gerade nicht in der Lage ist, diese ohne Weiteres „zurückzuerlangen". Hier setzen die gesetzlichen Anforderungen an die Datenverarbeitung an: Unmittelbar adressiert das Datenschutzrecht die aufgeworfene Machtfrage[51] und setzt neben dem Schutz von Individuen auch bei den Strukturen von Datenverarbeitung und Gesellschaft an.[52]

Dafür enthält die DSGVO etwa einen eigenen Grundsatz der Fairness[53], der Rechtmäßigkeit und Transparenz (Art. 5 Abs. 1 Buchst. a DSGVO). Nach dem Grundsatz der Zweckbindung, dürfen Daten nur erhoben werden, um ein bestimmtes Ziel zu erreichen oder eine bestimmte Funktion[54] zu erfüllen (Art. 5 Abs. 1 Buchst. b DSGVO). Weiterhin müssen die Datenverarbeitenden im Rahmen ihrer Rechenschaftspflicht nachweisen können, dass sie die datenschutzrechtlichen Regelungen einhalten (Art. 5 Abs. 2 DSGVO). Zudem müssen sie als Verantwortliche die Risiken für Grundrechte[55] schon vor dem Beginn der Verarbeitung (Art. 25 DSGVO) – ggf.

49 Vgl. auch *Bieker*, The Right to Data Protectionm 2022 (190-192).

50 *Steinmüller u.a.*, Grundfragen des Datenschutzes, 1971, BT-Drs. VI/3826 Anlage 1, 5 (36, 40, 82 f.).

51 *Theilen* u.a., Internet Policy Review 2021; *Rouvroy/Poullet,* in: Gutwirth u.a. (Hrsg.), Reinventing Data Protection?, 2009, 45 (69).

52 *Bieker*, The Right to Data Protection, 2022 (186-193).

53 In der deutschen Sprachfassung mit „Treu und Glauben" unglücklich übersetzt.

54 Grundlegend zur Funktionalität algorithmischer Systeme vgl. *Raji u.a.*, in: FAccT '22, 2022.

55 Der KI-VO-Entwurf sieht in Art. 11 Abs. 1 eine technische Dokumentation vor, in der Anbieter:innen nach Annex IV Nr. 3 u.a. Informationen zu den vorhersehbaren unbeabsichtigten Folgen und Risikoquellen bezüglich der Grundrechte und Diskri-

im Rahmen einer Datenschutz-Folgenabschätzung (Art. 35 DSGVO) – bewerten und entsprechende Maßnahmen ergreifen, um ein den Risiken angemessenes Schutzniveau zu gewährleisten (Art. 25 und Art. 32 DSGVO). Dabei muss ein Perspektivwechsel erfolgen, denn die Regelungen dienen, anders als etwa die IT-Sicherheit, gerade nicht den eigenen Interessen des Verantwortlichen, sondern der Personen, denen aufgrund der Verarbeitung Schäden drohen.[56]

Neben dem Datenschutzrecht, das für personenbezogene Daten Anwendung findet, steht das Informationsfreiheitrecht, das ebenfalls dazu dient, Informationsmacht einzuhegen. Die Informationsfreiheit strebt eine Machtbegrenzung durch Herstellung von Transparenz über staatliches Handeln an, ist jedoch in den meisten Anwendungsfällen, insbesondere im Hinblick auf private Anbieter algorithmischer Systeme, nur auf staatliche Akteure anwendbar. Während das Informationsfreiheitsrecht als individuelles Antragsrecht von Personen ausgestaltet ist und hier die informationspflichtigen Stellen bei ihnen vorhandene Daten herausgeben müssen, soweit sie nicht durch gesetzlich vorgegebene private oder öffentliche Interessen daran gehindert sind, sind in einigen Bundesländern Transparenzportale oder Open-Data-Portale aufgebaut worden, die ein proaktives Veröffentlichen vorsehen. Ebenso wie beim Datenschutz „by Design" empfiehlt sich ein planvolles Vorgehen für Informationsfreiheit „by Design", um ohne großen Aufwand die angeforderten Informationen zusammenzustellen und rechtssicher zu beurteilen, welche Informationen in welcher Form herauszugeben sind.

Im Fall von Ausschlussgründen ist es häufig möglich, die Informationen mit geeigneten Schwärzungen bereitzustellen – auch hierbei kann eine entsprechende technische und organisatorische Gestaltung der Prozesse und der Aktenführung helfen.[57] Es lassen sich zwar geeignete Lösungen finden, um die Datenverarbeitenden sowohl im Datenschutz als auch in der Informationsfreiheit zu unterstützen.[58] Doch die heutigen E-Aktensysteme, Transparenzportale oder E-Government-Anwendungen, für deren Hersteller und Betreiber die Anforderung der Umsetzung des Informations-

---

minierung bereitstellen müssen. Ebenso wie die Untersuchung möglicher Biases in den Trainingsdaten nach Art. 10 Abs. 2 Buchst. f ist dies nur für Hochrisikosysteme vorgesehen.

56 *Friedewald u.a.,* White Paper Datenschutz-Folgenabschätzung, 2017 (31).

57 *Hansen/Krasemann* 2022, S. 35 ff.

58 *Hansen/Bieker/Bremert* 2022, S. 287 f.

zugangrechts unter Wahrung der gesetzlich vorgesehenen privaten und öffentlichen Interessen jedenfalls nicht überraschend sein sollte, sind in dieser Hinsicht heutzutage zumindest ausbaufähig.[59]

Neben diesem Datenschutzrecht im engeren Sinne behandeln noch weitere Rechtsnormen die aus der Datenverarbeitung entstehende Informationsmacht. Insbesondere zum Antidiskriminierungsrecht bestehen zahlreiche Bezüge.[60] Dies zeigt sich auch in den Verweisen auf Antidiskriminierungsvorschriften in der DSGVO selbst. Als einer der möglichen Schäden für die Rechte und Freiheiten natürlicher Personen, die der besondere, auf Grundrechtsrisiken abstellende Ansatz der DSGVO verhindern soll, nennen ErwGr. 75 und 85 die Diskriminierung der betroffenen Personen. ErwGr. 71 verweist auf besondere Schutzmaßnahmen, die sicherstellen sollen, dass die besonderen Kategorien personenbezogener Daten nach Art. 9 DSGVO nicht in einer Weise verarbeitet werden, die diskriminierend ist. Insofern umfasst der Verweis auf den Schutz der Grundrechte natürlicher Personen in Art. 1 Abs. 2 DSGVO selbstverständlich auch das Grundrecht auf Nichtdiskriminierung gemäß Art. 21 EU-Grundrechte-Charta.[61]

Die Verbindungen zwischen Datenschutz- und Antidiskriminierungsrecht sind jedoch erst in Ansätzen untersucht,[62] obwohl sich die bestehenden gesellschaftlichen Probleme wie Sexismus, Rassismus oder Transphobie auch – und wie bereits aufgezeigt in verstärktem Maße – bei den problematischen Datenpraktiken, wie dem Einsatz von Chatbots, niederschlagen. Im Antidiskriminierungsrecht ist das Problem intersektionaler Diskriminierungen ein bekanntes Phänomen.[63]

Allerdings ist dieses Problem auch dort noch nicht ausgeräumt: So stellen die verschiedenen, gesetzlich geschützten Kategorisierungen eine Vereinfachung dar, die sich dem Problem der Intersektionalität stellen

59 Ansätze der schleswig-holsteinischen Landesverwaltung siehe *Thomsen,* Sommerakademie, 2022. Bedarf zu einer verbesserten „By-Design"-Umsetzung besteht auch in Bezug auf Archive, siehe *Friedewald u.a.,* Access to Archives: Implementation of Recommendation No. R(2000)13 on a European policy on access to archives, 2023 I.E.

60 Im Sinne von Steinmüller u.a. lässt sich der Teil des Anti-Diskriminierungsrechts, der dazu dient, unerwünschte Auswirkungen von Datenverarbeitung zu adressieren, auch als Datenschutzrecht im weiteren Sinne verstehen, *Steinmüller u.a.*, Grundfragen des Datenschutzes, 1971, BT-Drs. VI/3826 Anlage 1, 5 (44); vgl. dazu *Bieker*, Right to Data Protection 2022 (189-190, 195-199).

61 Vgl. z.B. *Draude/Hornung/Klumbytė*, in: Hepp u.a. (Hrsg.), New Perspectives in Critical Data Studies 2022, 187 (194).

62 Vgl. Ebd. (202-208).

63 Vgl. schon *Crenshaw*, The University of Chicago Legal Forum 1989, 139.

muss, damit die strukturellen Privilegien, die vorwiegend weiße Männer genießen, nicht aus dem Fokus geraten.[64] Dies betrifft insbesondere die Objektivität vermeintlich statischer und gegebener Grundannahmen sozialer Kategorisierung, die zu hinterfragen ist, damit ebendiese bestehenden Strukturen und Privilegien aufgebrochen werden können.[65]

5. Fairness als Gerechtigkeit

In diesem Prozess des Hinterfragens ist es wesentlich, mit Hilfe der gesetzlichen Regelungen die herrschenden Machtstrukturen sichtbar zu machen, um Schieflagen zu erkennen, Teilhabe zu ermöglichen und Macht umzuverteilen. Diese Fragen von Gerechtigkeit können den Blickwinkel des Rechts erweitern und weiße Flecken ausfüllen.

Aus der Perspektive von *Data Justice* sind insbesondere individualistische Ansätze im Datenschutzrecht wenig hilfreich, da sie die dahinterliegenden Strukturen eher verwischen, als dass sie ein strukturelles Korrektiv anbieten.[66] Allerdings sollte das bestehende Datenschutz-Governance-System in seiner Gesamtheit betrachtet werden. Während etwa die DSGVO Regelungen zum Schutz von Individuen, wie etwa die Betroffenenrechte, enthält, deren effektive Umsetzung auch von den Ressourcen der betroffenen Personen selbst abhängt,[67] gibt es auch Regelungen, die gerade darauf setzen, dass Datenschutz „out of the box", von Beginn an, gewährleistet sein muss. Dies sind insbesondere die bereits angesprochenen Regelungen des Systemdatenschutzes,[68] nach denen insbesondere Risiken für die Grundrechte von Individuen bereits in der Planungsphase zu ermitteln und technische und organisatorische Maßnahmen zur Eindämmung dieser Risiken noch vor Beginn der Datenverarbeitung umzusetzen sind (vgl. Art. 25, 32, 35 DSGVO).

Während sich diese Regelungen also an individuellen Grundrechten als Maßstab orientieren, müssen sie nicht von betroffenen Personen durchgesetzt werden, sondern sind von Datenverarbeitenden stets zu berücksichtigen. Dabei ist natürlich problematisch, dass die Datenverarbeitenden

64 *Hoffmann*, Information, Communication & Society 2019, 900 (906).
65 Ebd. (907). *González Hauck*, Zeitschrift für Rechtssoziologie 2022, 153-175.
66 *Hintz*, in: Dencik u.a. (Hrsg.): Data Justice 2022, 89 (95-97).
67 Ebd. (95).
68 S. Abschn. 3; *Bieker*, The Right to Data Protection, 2022, 187 f.

selbst diese Maßnahmen ergreifen und damit einen Perspektivwechsel zugunsten der betroffenen Personen und gegebenenfalls gegen ihre eigenen wirtschaftlichen Interessen vornehmen müssen. In diesem Zielkonflikt liegt den Datenverarbeitenden eine Optimierung im Sinne von Umsatz, Gewinn und Marktanteil deutlich näher als die Beschäftigung mit Grundrechten; das geschäftliche Risiko einer spürbaren Sanktionierung oder im schlimmsten Fall einer Untersagung der konkreten Verarbeitung oder der Basis für das Geschäftsmodells durch die Datenschutzaufsicht wird als niedrig eingeschätzt.[69]

Damit es in der Praxis gelingt, Gerechtigkeit zu gewährleisten, muss Datenverarbeitung holistisch und kleinschrittig betrachtet werden: von den Komponenten und Prozessen über Software-Anwendungen, Betriebssysteme, Hardware bis zu Infrastrukturen wie Libraries und weithin genutzten Code-Repositories, Kommunikationsprotokollen und Plattformen sowie dem weiteren gesellschaftlichen Kontext der Nutzung der Datenverarbeitung. Diese Betrachtungen müssen stetig – in der Planungsphase, während der Gestaltung, beim Inbetriebnehmen und fortlaufend beim Einsatz – erfolgen. Nur durch solche Schutzmaßnahmen „by Design" lässt sich gewährleisten, dass die Risiken einer Datenverarbeitung für Individuen sich nicht in konkreten Schäden realisieren.[70]

Auch in technischen Systemen werden stets bewusste und unbewusste Vorstellungen, Grundannahmen und Werte ihrer Entwicklungsteams eingeschrieben.[71] Dies wird durch die bestehenden Machtstrukturen, die dazu führen, dass bestimmte Gruppen ihre Privilegien nicht hinterfragen müssen, sondern es als gegeben annehmen können nicht diskriminiert zu werden,[72] begünstigt und bildet dadurch diese Strukturen wiederum ab. Bei der Gestaltung dieser Systeme stellt sich also die grundlegende Frage, welche Werte in dem System umgesetzt werden.[73] Ansonsten drohen Werte – sowie bewusste und unbewusste Biases – ohne Bedacht und Reflektion festgeschrieben zu werden. Der Ansatz von *Value Sensitive Design* versucht,

69 Zu den drei Fällen in den Jahren 2019 bis 2022, in denen die Federal Trade Commission eine Löschung des Algorithmus angeordnet hatte, siehe *Goland* 2023, S. 17 ff.

70 *Hansen/Bieker/Bremert* 2022.

71 *Hicks*, in: Mullaney u.a. (Hrsg.), Your Computer is on Fire, 2021, 11 (14 f.). Dabei werden Forderungen nach mehr Diversität in den Entwicklungsteams, wie bereits in Abschn. 3 aufgezeigt, oft verwendet, um von den dahinterliegenden strukturellen Problemen abzulenken.

72 Vertiefend hierzu: *Eggers* u.a., Mythen, Masken und Subjekte, 2020.

73 *Friedman*, Interactions 1996.

diesen Prozess sichtbar zu machen. Dabei geht es in einem dreischrittigen Prozess zunächst in einer konzeptuellen Analyse darum, die direkt und indirekt betroffenen Personen („Stakeholder") und die umzusetzenden Werte und ihre Abwägung zu identifizieren. In einer empirischen Untersuchung wird im Anschluss der soziale Kontext der Technologie betrachtet und kann später auch der Erfolg eines bestimmten Designs untersucht werden. Schließlich wird in einer technischen Untersuchung analysiert, wie die angestrebten Werte durch proaktive Gestaltung unterstützt werden können.[74]

Um die notwendige Erweiterung des Betrachtungsrahmens für die genannten Probleme zu erreichen, lassen sich die Ansätze von *Data Justice* mit *Value Sensitive Design* und „by-Design"-Ansätzen kombinieren. Die daraus hervorgegangene Methodologie von *Design Justice* hinterfragt bestehende Machtverhältnisse, die sich in den einzelnen Bereichen von Datenverarbeitung und ihren Folgen zeigen.[75] Durch die frühzeitige Einbindung der Perspektiven der von Marginalisierung betroffenen Personen und Gruppen, die überproportional von den negativen Folgen, in Form der in Abschn. 2 beispielhaft aufgezeigten Diskriminierungen,[76] der hier besprochenen Datenpraktiken betroffen sind, können Risiken einer Verarbeitung bereits im Gestaltungsprozess erkannt werden.

Auch hier gilt es, Intersektionalität zu beachten und die gelebten Erfahrungen der Angehörigen marginalisierter Gruppen als Expertise über ihre eigene Unterdrückung ernst zu nehmen. Dabei gilt es, wie wir für das Antidiskriminierungsrecht bereits im vorigen Abschnitt festgestellt haben, auch zu berücksichtigen, inwieweit vermeintlich objektive Dritte durch bestehende Machtstrukturen Hegemonie ausüben können. Unter den Design-Prinzipien von *Design Justice* gilt es etwa die Personen, die von den Ergebnissen des Design-Prozesses direkt betroffen sind, in den Vordergrund zu stellen, den potenziellen Auswirkungen einer Design-Entscheidung einen höheren Stellenwert zu geben als den Intentionen der Designenden und Veränderungen nicht als Endpunkt, sondern als Teil eines fortlaufenden Prozesses zu betrachten.[77]

In Anlehnung an das Konzept von *Design Justice* und unter Berücksichtigung der Kritik der Individualisierung von Problemen zeigen wir im fol-

74 *Friedman u.a.*, in: Himma/Tavani (Hrsg.): The Handbook of Information and Computer Ethics, 2008.
75 *Costanza-Chock*, Design Justice, 2023.
76 S. auch ErwGr. 35, 36, 37, 44 KI-VO-Entwurf.
77 Ebd., 190-204.

genden Abschnitt auf, wie sich Daten-Gerechtigkeit in Prozessen verankern lässt, damit ein fortlaufendes Korrektiv aufgebaut wird.

6. Prozesse für Daten-Gerechtigkeit

Der aktuelle Stand der Implementierung von algorithmischen Systemen verdeutlicht bereits die Notwendigkeit der grundlegenden Umsetzung und Durchsetzung der datenschutzrechtlichen Regelungen. Die Einführung von Chatbots hat in besonderem Maße gezeigt, dass diese Vorschriften eine wesentliche Rolle spielen müssen. Die Regelungen zur Risikobewertung, „by-Design"-Ansätzen, Datenschutz-Folgenabschätzung und Rechenschaftsplicht hätten bereits frühzeitig zum Erkennen der bestehenden Probleme geführt. Dazu hätte auch gehört, die Rechtmäßigkeit der Verwendung der aus zahlreichen, großteils gegenüber den Nutzenden nicht offengelegten, Quellen gesammelten Daten in den zugrundeliegenden *Large Language Models* sicherzustellen und sich der Qualität – und der Datensammlungen inhärenten Biases – bewusst zu werden.

Es ist anzunehmen, dass bei einer vernünftigen Analyse der Risiken für Grundrechte eine Freigabe der Anwendungen nicht hätte erfolgen können. Neben grundlegenden Rechtmäßigkeitsfragen hätte die Folgenabschätzung nicht nur Risiken aufgezeigt, sondern auch Maßnahmen zur deren Einhegung oder Abmilderung enthalten.

Um solche Risiken überhaupt zu erkennen, genügt es jedoch nicht, eine Checkliste abzuarbeiten. Dafür sind die möglichen Risiken zu abhängig von den konkreten Kontexten eines Verarbeitungsvorgangs. Vielmehr bedarf es dafür vordefinierter Prozesse, die je nach Kontext angepasst werden. Um den spezifischen Kontext jeweils ausreichend zu berücksichtigen, ist es wiederum maßgeblich, dass der Einsatzzweck einer Anwendung – also letztlich der Zweck der Datenverarbeitung – im Vorwege ausreichend definiert wird.

Anstatt solche Anwendungen zu veröffentlichen und am Rande in den AGB[78] oder Aufsätzen in rudimentärer Weise auf bestehende Risiken hinzuweisen[79], ohne diese zu adressieren oder zu bewältigen, müssen die Risiken einer solchen Verarbeitung für die Rechte von Individuen noch vor dem Beginn der Verarbeitung umfassend beschrieben und bewertet

78 https://openai.com/policies/usage-policies.

79 *Bubeck u.a.*, Sparks of Artificial General Intelligence: Early experiments with GPT-4, 2023.

werden. Nur so kann der Verarbeitungsvorgang entsprechend den Risiken angepasst werden; einschließlich der möglichen Entscheidung, ein *Large Language Model* nicht zur allgemeinen Nutzung freizugeben, solange das nicht bisher beherrschte Risiko besteht, dass diese Anwendungen rassistische, sexistische und transphobe Diskriminierungen perpetuieren.

Es liegt nicht in der Verantwortlichkeit der betroffenen Personen, einen Verarbeitungsvorgang zu testen und Verbesserungsvorschläge zu unterbreiten. Diese Aufgabe ist im Datenschutzrecht, begrifflich und rechtlich, den Verantwortlichen, also den Datenverarbeitenden, zugeordnet. Im Datenschutzrecht ebenfalls Sache der Anbieter ist es im Rahmen ihrer Rechenschaftspflicht, über den ganzen Lebenszyklus einer Anwendung, also bereits vor Beginn über den Betrieb bis nach deren Einstellung, nachzuweisen, dass die datenschutzrechtlichen Anforderungen eingehalten werden.

Diese datenschutzrechtlichen Regelungen sollten jedoch ebenfalls erweitert und über den konkreten Kontext des Datenschutzes hinaus weitergedacht werden. Zum Beispiel ist es ein allgemeiner Grundsatz des Zivilrechts, dass eine Haftung für Tätigkeiten, die andere potenziell gefährden, besteht, wenn diese Gefahren nicht ausreichend durch Verkehrssicherungspflichten begrenzt werden. Demnach müssen durch den Verursacher der Gefahr die Maßnahmen ergriffen werden, die erforderlich sind, um dem Entstehen von Schäden vorzubeugen und negative Auswirkungen im Falle eines Schadenseintritts gering zu halten oder abzumildern. Diese Verkehrssicherungsgrundsätze aus der physischen Welt, z. B. zur Gebäude- oder Wegesicherung, sind auch auf den digitalen Bereich übertragbar. Zudem ist das Treffen von Maßnahmen zur Einhegung möglicher Schäden ein Grundsatz des Produkthaftungsrechts.[80] In Anbetracht der Risiken, die mit algorithmischen Systemen einhergehen und die oft gerade nicht ausreichend untersucht werden, erscheint auch eine Gefährdungshaftung, also eine Haftung ohne den Nachweis konkreten Verschuldens aufgrund der aus der Tätigkeit folgenden erlaubten Gefahr, wie etwa dem Betreiben gefährlicher Anlagen, möglich.[81]

Im Lebenszyklus von Anwendungen gibt es spezifische Risiken, die mit den Änderungen des Einsatzes oder der Umgebung verbunden sind und

80 Dabei sind auch Pflichtversicherungen denkbar. Ein von ChatGPT synthetisch erzeugter Text zu einer möglichen gesetzlichen Regelung findet sich bei *Wagner*, ZDFheute v. 30. Apr. 2023.

81 Vgl. dazu auch den Entwurf der Kommission zu einer KI-Haftungsrichtlinie (https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:52022PC0496) und *de Conca* u.a., May Cause Liability – Use Care When Using the Internet of Things.

für deren Erkennung und Behandlung ein prozessorientierter Ansatz notwendig ist. So besteht die Gefahr eines *function creep*, also der schleichenden Ausweitung der ursprünglich begrenzten Zwecke durch Nutzende oder die einsetzende Organisation selbst. Diese Gefahr besteht natürlich nur, wenn die Zwecke vorab festgelegt sind, wie es in Art. 5 Abs. 1 Buchst. b DSGVO vorgeschrieben ist.

Mit ähnlicher Auswirkung können sich auch die Risiken einer Verarbeitung im Lauf der Zeit ändern oder müssen etwa mit neugewonnenen Erkenntnissen anders beurteilt werden. Beispielsweise ist die Annahme verbreitet, dass die Modelle, die durch Maschinenlernen entstehen, lediglich abstrahierte Muster oder Strukturen enthalten, die keinen unmittelbaren Rückschluss auf die möglicherweise personenbezogenen oder anderweitig sensiblen Trainingsdaten zulassen. Diese Annahme muss aber nicht stimmen: So konnten Forschungsteams für algorithmische Systeme demonstrieren,

1. dass sich zumindest in bestimmten Konstellationen feststellen ließ, ob Datensätze zu einer Person in den Trainingsdaten enthalten waren (Membership Inference Attack, z. B. bei *Large Language Models*, die mit Daten zu bestimmten Krankheiten oder Patient:innen in bestimmten Krankenhäusern trainiert wurden)[82] sowie
2. dass sich aus *Large Language Models* personenbezogene oder urheberrechtlich relevante Daten aus den Trainingsdaten extrahieren ließen.[83]

Diese Erkenntnisse sind aus Datenschutzsicht besonders deswegen relevant, weil *Large Language Models* zumeist wie anonyme Daten eingestuft wurden und man annahm, dass mit dem Maschinenlernen der Effekt einer Anonymisierung einherginge. Anonyme oder anonymisierte Daten unterfallen jedoch – anders als personenbezogene Trainingsdaten – nicht dem Datenschutzrecht. Unter dieser Annahme wurde bisher nach unserer Kenntnis nicht problematisiert, wenn algorithmische Systeme aus der EU in einen Drittstaat weitergegeben wurden. Doch angesichts des Stands der Forschung, dass darin eben doch personenbezogene Daten eincodiert seien und entlockt werden könnten, wäre die Annahme einer vorliegenden Anonymität nicht haltbar; der Anwendungsbereich der DSGVO wäre eröff-

82 *Shokri u.a.* 2017, *Ye u.a.* 2022.
83 *Frederikson u.a.* 2015, *Yeom u.a.* 2020, *Carlini u.a.* 2021, *Tramèr u.a.* 2022, *Yu u.a.* 2023.

net.[84] Dieser Effekt von vermeintlich anonymen Daten kann insbesondere deswegen Schäden für die Grundrechte der betroffenen Personen auslösen, wenn die Verantwortlichen keine Vorsorge für diesen Fall getroffen haben, weil sie sich außerhalb der DSGVO wähnten.[85]

Folglich ist auch für die Überwachung von Risiken ein durchgängiger Prozess notwendig,[86] der eine Anwendung über den gesamten Lebenslauf begleitet. Auch bedarf es einer konstruktiven Fehlerkultur, die nicht durch eine voreilige Entschuldigung versucht, eine Diskussion zu beenden, sondern durch die ein Verantwortlicher die Folgen eines Fehlers begrenzt und Änderungen umsetzt, um diesen Fehler in der Zukunft ausschließen zu können.

In diesem Kontext wird auch deutlich, wie sinnvoll es ist, dass der Risiko-Begriff der DSGVO[87] nicht nur auf die Rechte der betroffenen Personen (also Personen, deren Daten verarbeitet werden), sondern auf sämtliche natürliche Personen abstellt, ganz im Sinne der direkten und indirekten betroffenen Stakeholder im Rahmen des *Value Sensitive Designs*. Wenn etwa ein *Large Language Model* aufgrund der Daten einer bestimmten Gruppe von Personen einen Bias aufweist, kann sich dieser zulasten ganz anderer Personen auswirken.[88]

Auch die Erkenntnisse zur Intersektionalität von Diskriminierungen müssen in diesen Prozessen berücksichtigt werden. Dies gilt etwa konkret für die Bewertung von Risiken und die Erkenntnis, dass Personen, die unter verschiedene, sich überschneidende Kategorisierungen fallen einem besonders hohen Risiko unterfallen. Dieses ergibt sich nicht einfach aus

84 Wiederum gibt es Forschungsansätze zum gezielten „Machine Unlearning", sodass es nicht undenkbar ist, einzelne oder viele Datensätze „herauszulernen" und damit auch einen Schutz gegen „Membership Inference Attacks" oder dem Auslesen der Daten zu schaffen; für einen Überblick siehe *Nguyen et al.* 2022. Es ist noch nicht geklärt, inwieweit mit bestimmten Unlearning-Ansätzen auch das Betroffenenrecht auf Löschung umgesetzt werden kann und unter welchen Umständen dies überhaupt eine realistische Option darstellt.

85 *Bruegger* 2021, S. 103 ff.

86 Vgl. auch *Bender u.a.*, in: FAcct '21, S. 619.

87 Auch nach Artikel 34 Abs. Buchst. b Digitale-Dienste-Verordnung ist eine Bewertung der nachteiligen Auswirkungen der Grundrechte, insbesondere des Rechts auf Nichtdiskriminierung vorgesehen, nach Art. 48 Abs. 4 Buchst. e sind Schutzvorkehrungen zur Vermeideung negativer Auswirkungen vorgesehen.

88 *Bieker*, The Right to Data Protection, 2023, 189.

einer simplen Addition der Risiken, denen etwa ein Schwarzer Mann (Rassismus) und eine weiße Frau (Sexismus) ausgesetzt sind.[89]

Allerdings gilt generell bei der Risikobewertung, dass diese sich nicht hinter vermeintlich objektiver Pseudo-Mathematik verstecken darf.[90] Mit diesem erweiterten Fokus, der auf breiteren Gerechtigkeitsüberlegungen auch außerhalb des Datenschutzrechts fußt, gibt es noch weitere Gruppen, die von algorithmischen Systemen betroffen sind. Dazu zählen insbesondere die Arbeiter:innen, die „unsichtbare Arbeit", etwa an den Chatbots zugrundeliegenden *Large Language Models* selbst, verrichten.[91] Zudem beruhen viele Anwendungen, die sich algorithmischer Systeme bedienen, darauf, dass konkrete Personen die Arbeit, die mit Hilfe der Systeme organisiert werden soll, verrichten. Dies sind in vielen Fällen sogenannte Gig-Worker:innen, also prekär Beschäftigte, die Waren ausliefern oder andere Dienstleistungen, wie zum Beispiel Content-Moderation, erbringen.

Die Einbindung der betroffenen Personen und Endnutzer:innen einer Anwendung, insbesondere Angehöriger bereits marginalisierter Gruppen, ist zudem ein wesentlicher Teil jeder Risikobewertung. Dies ist auch im Rahmen einer Datenschutz-Folgenabschätzung eine Möglichkeit (Art. 35 Abs. 9 DSGVO), von der frühzeitig im Prozess Gebrauch gemacht werden sollte. Dabei ist die Perspektive der betroffenen Personen und Gruppen als Expert:innen der von ihnen erfahrenen Diskriminierungen bereits bei der Identifizierung möglicher, auch intersektionaler, Risiken eine hilfreiche Unterstützung.

Allerdings ist es wichtig, auch hier Probleme nicht zu individualisieren oder etwa in ausbeuterische Muster zu verfallen. So besteht die Gefahr einer Tokenisierung[92] einzelner Angehöriger marginalisierter Gruppen, also der Vereinnahmung einer einzelnen Person als Repräsentant:in einer ganzen Gruppe, um Inklusion vorzutäuschen, oder eine mangelnde Anerkennung der Arbeit betroffener Personen in einem solchen Prozess, die in der Regel zu vergüten ist. Die DSGVO verweist explizit auf Organisationen, die die Interessen der betroffenen Personen vertreten. Es geht also auch hier um eine breite Einbindung von NGOs, Organisationen der Zivilgesellschaft, Gewerkschaften, Kooperativen und Bewegungen, die auch

89 *Crenshaw*, The University of Chicago Legal Forum 1989, 139 (149, 151 f.).
90 *Bieker/Hansen/Friedewald*, RDV 2016, 188 (193).
91 Gray/Suri, Ghost Work, 2019.
92 *Theilen u.a.*, Internet Policy Review 2021, S. 5.

diejenigen einbeziehen sollte, die die „unsichtbare Arbeit" verrichten, wie z. B. Gig-Worker:innen.

Es ist selbstverständlich problematisch, dass ausgerechnet der für die Datenverarbeitung Verantwortliche die gegenläufigen Interessen betrachten und umsetzen muss. Dies ist eine inhärente Begrenzung des Datenschutz- und letztlich auch Antidiskriminierungsrechts, sodass auch der Rückgriff auf dieses das Problem nicht ohne Weiteres lösen kann. Die rechtlichen Korrektive gegen Verstöße sind zudem für die betroffenen Personen mühsam und erfordern von der Beschwerde bei den entsprechenden Stellen bis zum Führen eines langwierigen und teuren Gerichtsverfahrens über mehrere Instanzen erhebliche finanzielle, zeitliche und mentale Ressourcen.[93]

Dabei besteht zudem das Problem, dass sich Datenverarbeitende nach dem Unterliegen in einem Gerichtsverfahren darauf berufen, ihre Praxis inzwischen angepasst zu haben, um sich so den gerichtlichen Anforderungen zu entziehen. Im schlimmsten Fall erfordert dies erneute (gerichtliche) Auseinandersetzungen, damit eine angemessene Umsetzung erfolgt. Dies liegt auch an den geradezu unbegrenzten Ressourcen, die Anbietern zur Verfügung stehen, um ihre Geschäftsmodelle zu verteidigen.

Um dieses Gefälle zwischen den Anbietern, mit ihrer Informationsmacht, und den Einzelnen, die die Risiken dieser Machtakkumulation tragen, zu überbrücken, sind tiefgreifende strukturelle Eingriffe notwendig, die über das bestehende Daten(schutz)recht hinausgehen. Dabei sollte jedoch nicht auf unpassende historischen Präzedenzfälle geblickt werden; die Anbieter von heute agieren global und ähneln daher weniger den nationalen Elektrizitäts- oder Eisenbahnmonopolen des 19. und 20. Jahrhunderts. Insofern ist nicht klar, ob mit kartellrechtlichen Regelungen auf nationaler oder europäischer Ebene die Informationsmacht der Anbieter wirkungsvoll zu regeln ist oder ob dies nur auf internationaler Ebene, etwa in Form von Governance-Modellen wie der ICANN – die wiederum eigene Probleme aufweisen – oder durch die Implementierung von Protokollen auf Infrastruktur-Ebene, zu erreichen ist.[94]

93 So nahm der Rechtsstreit einer betroffenen Person gegen ehrverletzende Suchwortergänzungen zum eigenen Namen, die Google als Autocomplete-Funktion im Jahr 2009 eingeführt hatte, mehrere Jahre bis zum BGH-Urteil vom 14. Mai 2013 – VI ZR 269/12 – in Anspruch. Als Reaktion führte Google ein Formular zum Entgegennehmen von Beschwerden ein. Dass die Anbieter von ChatGPT oder anderen Anwendungen das BGH-Urteil in die Gestaltung ihres Angebots einbezogen hätten, ist nicht ersichtlich.

94 *Keyes*, Wired v. 11. Jan. 2022.

7. Ausblick

Es ist daher fraglich, ob im Rahmen der aktuellen Debatte um den Einsatz algorithmischer Systeme eine Lösung erreicht werden kann, die die fundamentalen Probleme dieser Technik ausreichend durch eine Veränderung der ihrer Anwendung zugrundeliegenden Machstrukturen bewirkt. Notwendig wäre es dafür zunächst den Fokus von Einzelnen abzuwenden und stattdessen die bestehenden gesellschaftlichen Machtstrukturen zu analysieren. Eine solche Analyse muss den Blick jedoch über das, was klassischerweise als Datenschutzrecht angesehen wird, hinaus in benachbarte Rechtsgebiete lenken.

Allerdings liefern auch diese Bereiche, wie etwa das Antidiskriminierungsrecht in Bezug auf die hier beschriebenen intersektionalen Diskriminierungen, keine Patentlösungen. Es ist daher eine Erweiterung der Perspektive jenseits des Rechts auch auf Lösungsansätze nötig, die von Angehörigen marginalisierter Gruppen entwickelt wurden. Mit den grundlegenden Prozessen des bestehenden Datenschutzrechtes lassen sich – durch eine solche Perspektiverweiterung – viele der aktuellen Probleme erfassen, bewerten und Abhilfemaßnahmen ableiten.

Wesentlich ist die Erkenntnis, dass die aktuellen Datenpraktiken nicht alle Menschen gleich, sondern marginalisierte Gruppen überproportional treffen. Auch hier muss sich der Fokus von der Industrie und den Anbietern algorithmischer Systeme abwenden. Nur weil jemand ein Problem schafft, heißt es nicht, dass er, mit all seinen wirtschaftlichen Eigeninteressen,[95] auch zur Problemdefinition und -lösung berufen ist.[96] Vielmehr muss die Deutungshoheit von der Industrie hin zu den betroffenen Personen und der Gesellschaft insgesamt gelenkt werden.

Die bestehenden Strukturen, die es Unternehmen ermöglichen und sogar dafür Anreize bieten, mächtige *Large Language Models* bereitzustellen und Risiken auszulagern, müssen grundlegend verändert werden. Erst wenn Anbieter gezwungen sind, Risiken zu bewältigen, bevor sie mit Anwendungen Geld verdienen oder Marktanteile sichern können, ist es realistisch, dass Regelungen, die ihren wirtschaftlichen Interessen derart entgegenlaufen wie der Schutz marginalisierter Gruppen und Datenschutz, umgesetzt werden.

95 Die sich durch den Einsatz von Anwendungen, die auf *Large Language Models* beruhen, zukünftig noch einfacher und womöglich (teil-)automatisiert vertreten lassen, vgl. *Sanders/Schneier*, Technology Review v. 14 März. 2023.

96 *Dencik*, in: Dencik u.a. (Hrsg.): Data Justice 2022, 123 (133 f.).

Je stärker eine Technik Auswirkungen auf die Menschen und die demokratische Gesellschaft haben kann, desto wichtiger ist eine unabhängige Kontrolle bereits vor dem Inverkehrbringen und auch während ihres Einsatzes. Nur so kann Technik dazu beitragen gesellschaftliche Probleme zu lösen und nicht, um die wirtschaftlichen Interessen weniger zu erfüllen, eine Vielzahl von Problemen für die Gesellschaft insgesamt zu verstärken oder zu verursachen.

Danksagung

Diese Arbeit wurde vom Bundesministerium für Bildung und Forschung im Rahmen des Projekts „Privatheit, Demokratie und Selbstbestimmung im Zeitalter von Künstlicher Intelligenz und Globalisierung" (PRIDS), https://forum-privatheit.de/, gefördert (FKZ 16KIS1376). Wir bedanken uns herzlich bei Jens T. Theilen für die hilfreichen Anmerkungen.

Literatur

Alle Internet-Quellen zuletzt besucht am 08.05.2023

Altman, Sam (24. Feb. 2023): Planning for AGI and beyond. URL: https://openai.com/blog/planning-for-agi-and-beyond

Atleson, Michael (27. Feb. 2023): Keep your AI claims in check. URL: https://www.ftc.gov/business-guidance/blog/2023/02/keep-your-ai-claims-check

Bender, Emily M.; Gebru, Timnit; McMillan-Major, Angelina und Shmitchell, Shmargaret (2021): On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In: *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. S. 610-623. URL: https://doi.org/10.1145/3442188.3445922

Bender, Emily M. (2022): Human-like programs abuse our empathy – even Google engineers aren't immune. *The Guardian* vom 14. Jun. 2022. URL: https://www.theguardian.com/commentisfree/2022/jun/14/human-like-programs-abuse-our-empathy-even-google-engineers-arent-immune

Biddle, Sam (2022): The Internet's New Favorite AI Proposes Torturing Iranians and Surveilling Mosques. *The Intercept* vom 8. Dez. 2022. URL: https://theintercept.com/2022/12/08/openai-chatgpt-ai-bias-ethics/

Bieker, Felix (2022): *The Right to Data Protection: Individual and Structural Dimensions of Data Protection in EU Law*. The Hague: T.M.C. Asser Press. URL: https://doi.org/10.1007/978-94-6265-503-4

Bieker, Felix (2018): Die Risikoanalyse nach dem neuen EU-Datenschutzrecht und dem Standard-Datenschutzmodell. *Datenschutz und Datensicherheit (DuD)*, 42(1), S. 27-31.

Bieker, Felix; Hansen, Marit; Friedewald, Michael (2016): Die grundrechtskonforme Ausgestaltung der Datenschutz-Folgenabschätzung. *Recht der Datenverarbeitung (RDV)*, 32(4), S. 188-197.

Browne, Simone (2015): *Dark Matters*. Durham: Duke University Press.

Bruegger, Bud P. (2021): Towards a Better Understanding of Identification, Pseudonymization, and Anonymization. ULD. https://uld-sh.de/PseudoAnon

Bubeck, Sébastien; Chandrasekaran, Varun; Eldan, Ronen; Gehrke, Johannes; Horvitz, Eric; Kamar, Ece; Lee, Peter; Lee, Yin Tat; Li, Yuanzhi; Lundberg, Scott; Nori, Harsha; Palangi, Hamid; Ribereiro, Marco Tulio; Zhang, Yi (2023): Sparks of Artificial General Intelligence: Early experiments with GPT-4. URL: https://arxiv.org/abs/2303.12712

Buolamwini, Joy; Gebru, Timnit (2018): Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In: *Proceedings of Machine Learning Research 81*, S. 1-15. URL: http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Carlini, Nicholas; Tramèr, Florian; Wallace, Eric; Jagielski, Matthew; Herbert-Voss, Ariel; Lee, Katherine; Roberts, Adam; Brown, Tom; Song, Dawn; Erlingsson, Úlfar; Oprea, Alina; Raffel, Colin (2021): Extracting Training Data from Large Language Models. In: *Proceedings of 30th USENIX Security Symposium*, S. 2633-2650. https://www.usenix.org/conference/usenixsecurity21/presentation/carlini-extracting

Clarke, Arthur C. (1973): Profiles of the Future: An Inquiry into the Limits of the Possible. 2. Aufl. Harper & Row.

Constantaras, Eva; Geiger, Gabriel; Braun, Justin-Casimir; Mehrotra, Dhruv; Aung, Htet (2023): Inside the Suspicion Machine. *Wired* vom 6. Mar. 2023. URL: https://www.wired.com/story/welfare-state-algorithms/

Crenshaw, Kimberle (1989): Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. *The University of Chicago Legal Forum*, 168, 139-167.

de Conca, Silvia; Bratu, Ioana; Leiser, Mark; Cooper, Zac (14. Nov. 2022): May Cause Liability – Use Care When Using the Internet of Things. URL: https://alti.amsterdam/may-cause-liability-part-1/

Dencik, Lina (2022): Data and Social Justice. In: Dencik, Lina; Hintz, Arne; Redden, Joanna; Treré, Emiliano (Hrsg.): *Data Justice*. Los Angeles: Sage. S. 123-137.

Draude, Claude; Hornung, Gerrit, Klumbytė, Goda (2022): Mapping Data Justice as a Multidimensional Concept Through Feminist and Legal Perspectives. In: Hepp, Andreas; Jarke, Juliane; Kramp, Leif (Hrsg.): *New Perspectives in Critical Data Studies*. Cham: Palgrave Macmillan. S. 187-216.

Eubanks, Virginia (2018): The Digital Poorhouse. *Harper's Magazine* von Jan. 2018. URL: https://harpers.org/archive/2018/01/the-digital-poorhouse/

Elish, Madeleine Clare (17. Jan. 2018). Don't Call AI "Magic". URL: https://points.datasociety.net/dont-call-ai-magic-142da16db408

Fredrikson, Matt; Jha, Somesh; Ristenpart, Thomas: Model inversion attacks that exploit confidence information and basic countermeasures. *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS '15)*. S. 1322-1333. https://doi.org/10.1145/2810103.2813677

Freeman, Alan David (1978): Legitimizing racial discrimination through antidiscrimination law: A critical review of Supreme Court doctrine. *Minnesota Law Review*, 62, S. 1049-1120.

Friedewald, Michael; Bieker, Felix; Obersteller, Hannah; Nebel, Maxi; Martin, Nicholas; Rost, Martin; Hansen, Marit (2019): White Paper Datenschutz-Folgenabschätzung. Karlsruhe: Forum Privatheit und selbstbestimmtes Leben in der digitalen Welt. URL: https://www.forum-privatheit.de/download/datenschutz-folgenabschaetzung-3-auflage-2017/

Friedewald, Michael; Székely, Iván; Karaboga, Murat; Runge, Greta; Ebbers, Frank (2023, i.E.): Access to Archives: Implementation of Recommendation No. R(2000)13 on a European policy on access to archives. Study commissioned by the Council of Europe. Karlsruhe: Fraunhofer ISI.

Friedman, Batya (1996): Value-Sensitive Design. *Interactions*, 3(6), S. 17-23. URL: https://dl.acm.org/doi/pdf/10.1145/242485.242493

Friedman, Baty; Kahn, Peter H.; Borning, Alan (2008): Value Sensitive Design and Information Systems. In: Himma, Kenneth Einar; Tavani, Herman T. (Hrsg.): *The Handbook of Information and Computer Ethics*. Hoboken: John Wiley & Sons. URL: https://onlinelibrary.wiley.com/doi/10.1002/9780470281819.ch4

Gal, Uri (2023). ChatGPT is a data privacy nightmare. If you've ever posted online, you ought to be concerned. *The Conversation* vom 8. Feb. 2023. URL: https://theconversation.com/chatgpt-is-a-data-privacy-nightmare-if-youve-ever-posted-online-you-ought-to-be-concerned-199283

Gilliard, Chris; Culik, Hugh (24. Mai 2016): Digital Redlining, Access, and Privacy. URL: https://www.commonsense.org/education/articles/digital-redlining-access-and-privacy

Goland, Joshua A. (2023): Algorithmic Disgorgement: Destruction of Artificial Intelligence Models as The FTC's Newest Enforcement Tool for Bad Data. *Richmond Journal of Law & Technology*, 29(2).

González Hauck, Sué (2022): *Weiße* Deutungshoheit statt Objektivität: Der ‚objektive Dritte' und die systematische Abwertung rassismuserfahrener Perspektiven. *Zeitschrift für Rechtssoziologie*, 42(2), S. 153-175.

Gray, Mary L.; Suri, Siddarth (2019): *Ghost Work*, Boston: Houghton Miffin Harcourt.

Guijarro Santos, Victoria (2023): Nicht besser als nichts, ein Kommentar zum KI-Verordnungsentwurf. *Zeitschrift für Digitalisierung und Recht*, 3, S. 23-42.

Hansen, Marit; Bieker, Felix; Bremert, Benjamin (2022): Datenschutz und Privatheitsschutz durch Gestaltung der Systeme. In: Roßnagel, Alexander; Friedewald, Michael (Hrsg.): *Die Zukunft von Privatheit und Selbstbestimmung*, Wiesbaden: Springer Vieweg. S. 259-300. https://doi.org/10.1007/978-3-658-35263-9_8

Hansen, Marit; Krasemann, Henry (2022): Datenherausgabe und Informationsfreiheit by Design – Was (nicht nur behördliche) Datenschutzbeauftragte wissen sollten. *BvD-News* 2/2022, S. 34-38.

Hasters, Alice (2020): Was weiße Menschen nicht über Rassismus hören wollen aber wissen sollten. München: Hanserblau.

Heidelberger, Natali; Diakopoulos, Nicholas (2023): ChatGPT and the AI Act. *Internet Policy Review*, 12(1). URL: https://policyreview.info/essay/chatgpt-and-ai-act

Hicks, Mar (2019): Hacking the Cis-tem. *IEEE Annals of the History of Computing* 2019, S. 20-33.

Hicks, Mar (2021): When Did the Fire Start? In: Mullaney, Thomas S.; Peters, Benjamin; Hicks, Mar; Philip, Kavita (Hrsg.): *Your Computer is On Fire*. Cambridge, MA: MIT Press. S. 11-26.

Hill Collins, Patricia; Bilge, Sirma (2020): *Intersectionality*. 2. Aufl. Cambridge: Polity.

Hintz, Arne (2022): Data and Policy. In: Dencik, Lina; Hintz, Arne; Redden, Joanna; Treré, Emiliano (Hrsg.): *Data Justice*. Los Angeles: Sage. S. 89-104.

Hoffmann, Anna Lauren (2019): Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication & Society* 22(7), S. 900-915.

Johnson, Khari (2022): DALL-E 2 Creates Incredible Images—and Biased Ones You Don't See. *Wired* vom 5. Mai 2022. URL: https://www.wired.com/story/dall-e-2-ai-text-image-bias-social-media/

Jones, Phil (2021): *Work without the Worker, Labour in the Age of Platform Capitalism*. London: Verso.

Eggers, Maureen Maisha; Kilomba, Grada; Piesche, Peggy; Arndt, Susan (2020): Mythen, Masken und Subjekte – Kritische Weißseinsforschung in Deutschland. 4. Aufl. Münster: Unrast-Verlag.

Kaminski, Margot (2022): The Case for Data Privacy Rights (or "Please, a little Optimism"), *Notre Dame Law Review Reflection*, 97(5), S. 385-399.

Keyes, Os (2022): It Doesn't Make Sense to Treat Facebook Like a Public Utility. *Wired* vom 11. Jan. 2022. URL: https://www.wired.com/story/facebook-public-utility-regulation/

Liebscher, Doris (2021): *Rasse im Recht – Recht gegen Rassismus*. Berlin: Suhrkamp.

Luccioni, Alexandra Sasha; Akiki, Christopher; Mitchell, Margaret; Jernite, Yacinde (2023): Stable Bias: Analyzing Societal Representations in Diffusion Models. URL: https://arxiv.org/abs/2303.11408

Mac, Ryan (2021): Facebook Apologizes after A.I. Puts 'Primates' Label on Video of Black Men, *New York Times* vom 3. Sep. 2021. URL: https://www.nytimes.com/2021/09/03/technology/facebook-ai-race-primates.html

Matzner, Tobias, Masur, Philipp K., Ochs, Carsten, von Pape, Thilo (2016): Do-It-Yourself Data Protection—Empowerment or Burden? In: Gutwirth, Serge, Leenes, Ronald, De Hert, Paul (Hrsg.): *Data Protection on the Move*. Law, Governance and Technology Series. Dordrecht: Springer. https://doi.org/10.1007/978-94-017-7376-8_11

Nguyen, Thanh Tam; Huynh, Thanh Trung; Nguyen, Phi-Le; Liew, Alan Wee-Chung; Yin, Hongzhi; Nguyen, Quoc Viet Hung (2022): A Survey of Machine Unlearning. https://arxiv.org/abs/2209.02299. Siehe auch https://github.com/tamlhp/awesome-machine-unlearning

Perrigo, Billy (2022): AI Chatbots Are Getting Better. But an Interview With ChatGPT Reveals Their Limits. *Time* vom 5. Dez. 2022. URL: https://time.com/6238781/chatbot-chatgpt-ai-interview/

Powles, Julia; Nissenbaum, Helen (2018): The Seductive Diversion of 'Solving' Bias in Artificial Intelligence. *Medium* vom 7. Dez. 2018. URL https://onezero.medium.com/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53

Raji, Inioluwa Deborah; Bender, Emily M.; Paullada, Amandalynne; Denton, Emily; Hanna, Alex (2021). in: *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks 1 (NeurIPS Datasets and Benchmarks 2021)*. URL: https://datasets-benchmarks-proceedings.neurips.cc/paper_files/paper/2021/file/084b6fbb10729ed4da8c3d3f5a3ae7c9-Paper-round2.pdf

Raji, Inioluwa Deborah; Kumar, I. Elizabeth; Horowitz, Aaron; Selbst, Andrew D. (2022): The Fallacy of AI Functionality. In: *FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. URL: https://facctconference.org/static/pdfs_2022/facct22-78.pdf

Roberts, Sarah T. (2021): Your AI is a Human. In: Mullaney, Thomas S. u.a. (Hrsg.): *Your Computer is on Fire*. Cambridge, Mass. und London: MIT Press, S. 51-70.

Roose, Kevin (2023): A Conversation with Bing's Chatbot Left Me Deeply Unsettled. *New York Times* vom 16. Feb. 2023. URL: https://www.nytimes.com/2023/02/16/technology/bing-chatbot-microsoft-chatgpt.html

Rouvroy, Antoinette und Poullet, Yves (2009): The Right to Informational Self-Determination and the Value of Self-Development, Reassessing the Importance of Privacy for Democracy. In: Gutwirth, Serge; Poullet, Yves; De Hert, Paul; de Terwangne, Cécile; Nouwt, Sjaak (Hrsg.): *Reinventing Data Protection?* Dordrecht: Springer, S. 45-76.

Sanders, Nathan E.; Schneier, Bruce (2023): How AI could write our laws. *Technology Review* vom 14 Mar. 2023. URL: https://www.technologyreview.com/2023/03/14/1069717/how-ai-could-write-our-laws/

Schräer, Frank (2023): Google Bard: Fehlerhafte Antwort der KI lässt Experten und Anleger zweifeln, *Heise Online* vom 09. Feb. 2023; abrufbar unter: https://www.heise.de/news/Google-Bard-Fehlerhafte-Antwort-der-KI-laesst-Experten-und-Anleger-zweifeln-7489896.html

Schwartz, Oscar (2018): 'The discourse is unhinged': how the media gets AI alarmingly wrong. *The Guardian* vom 25. Jul. 2018. URL: https://www.theguardian.com/technology/2018/jul/25/ai-artificial-intelligence-social-media-bots-wrong

Shanklin, Will (2023): Snapchat adds OpenAI-powered chatbot and proactively apologizes for what it might say. *Engadget* vom 27. Feb. 2023. URL: https://www.engadget.com/snapchat-adds-openai-powered-chatbot-and-proactively-apologizes-for-what-it-might-say-180507261.html

Shokri, Reza; Stronati, Marco; Song, Congzheng; Shmatikov, Vitaly (2017): Membership Inference Attacks Against Machine Learning Models. *Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP '17)*. S. 3-18. https://doi.org/10.1109/SP.2017.41

Taylor, Luke (2023): Colombian Judge says he used ChatGPT in ruling. *The Guardian* vom 3. Feb. 2023. URL: https://www.theguardian.com/technology/2023/feb/03/colombia-judge-chatgpt-ruling

Theilen, Jens T.; Baur, Andreas; Bieker, Felix; Ammicht Quinn, Regina; Hansen, Marit; González Fuster, Gloria (2021): Feminist Data Protection: An Introduction. *Internet Policy Review*, 10(4). URL: https://policyreview.info/pdf/policyreview-2021-4-1609.pdf

Tramèr, Florian; Shokri, Reza; San Joaquin, Ayrton; Le, Hoang; Jagielski, Matthew; Hong, Sanghyun; Carlini, Nicholas (2022): Truth Serum: Poisoning Machine Learning Models to Reveal Their Secrets. In: *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security (CCS '22)*. S. 2779-2792, https://doi.org/10.1145/3548606.3560554

Veale, Michael; Silberman, Michael; Binns, Reuben (2023): Fortifying the algorithmic management provision in the proposed Platform Work Directive. *European Labour Law Journal*, 14(2). S. 308-322, https://doi.org/10.1177/20319525231167983

Volpicelli, Gian (2023): ChatGPT broke the EU plan to regulate AI. *Politico* vom 3. März 2023. URL: https://www.politico.eu/article/eu-plan-regulate-chatgpt-openai-artificial-intelligence-act/

Wagner, Lukas (2023): Ethischer Umgang mit der Technik: Wie KI in Zukunft reguliert werden soll. *ZDFheute* vom 30. Apr 2023. URL: https://www.zdf.de/nachrichten/politik/ki-regeln-gesetz-ai-act-eu-ethik-experten-100.html

Xiang, Chloe (2022): AI Is Probably Using Your Images and It's Not Easy to Opt Out, *Vice*, 26.09.2022, https://www.vice.com/en/article/3ad58k/ai-is-probably-using-your-images-and-its-not-easy-to-opt-out

Ye, Jiayuan; Maddi, Aadyaa; Murakonda, Sasi Kumar; Bindschaedler, Vincent; Shokri, Reza (2022): Enhanced Membership Inference Attacks against Machine Learning Models. In: *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security (CCS '22)*. S. 3093-3106. https://doi.org/10.1145/3548606.3560675

Yeom, Samuel; Giacomelli, Irene; Menaged, Alan; Fredrikson, Matt; Jha, Somesh (2020): Overfitting, Robustness, and Malicious Algorithms: A Study of Potential Causes of Privacy Risk in Machine Learning. *Journal of Computer Security* 28(1), S. 35-70. https://doi.org/10.3233/JCS-191362

Yu, Weichen; Pang, Tianyu; Liu, Qian; Du, Chao; Kang, Bingyi; Huang, Yan; Lin, Min; Yan, Shuicheng (2023): Bag of Tricks for Training Data Extraction from Language Models. https://arxiv.org/pdf/2302.04460.pdf