Marc Mölders

Zum fairen Gestalten fairer Algorithmen Stakeholder-Verfahren und die Korrektur des *machine bias*

1. Einleitung

Algorithmen wird sowohl als Lösung wie auch als Problem sehr viel zugetraut. Manche sind sogar bereit, von einer Algokratie zu sprechen, von einer Herrschaft der Algorithmen (vgl. Aneesh 2009). Doch Algorithmen wird nicht nur Einflusspotential zugeschrieben, dieses sei auch noch in bestimmter Weise voreingenommen. Zu dieser Debatte hat die Machine Bias-Studie von *ProPublica* viel beigetragen. Mit ihr wurde es möglich, algorithmischer Entscheidungsassistenz, bezogen auf strafrechtliche Urteile, diskriminierende Voreingenommenheit öffentlich zu attestieren. Erst als mit dem gleichen Datensatz andere Resultate – insbesondere Diskriminierung betreffend – belegbar wurden, ließ sich mathematisch zeigen, dass es unterschiedliche und obendrein miteinander inkompatible Verständnisse algorithmischer Fairness gibt (Kapitel 2).

Aus unterschiedlichen Richtungen sind seither Forderungen zu vernehmen, Fragen der Fairness nicht der Informatik oder der Technikentwicklung zu überlassen. Teile der Informatik könnten dem kaum deutlicher zustimmen. Sie erklären Fairness zu einer wissenschaftlich oder mathematisch nicht entscheidbaren Frage (vgl. Berk et al. 2018). Über Fairness solle stattdessen im Sinne einer Wertfrage im Recht bzw. im politischen Prozess befunden werden. Damit ist eine Verfahrensfrage aufgeworfen. Womöglich auch weil in Politik und Recht bisher kaum Ansätze erkennbar geworden sind, auf welchem Wege zu Wertentscheidungen zu kommen wäre, liegen inzwischen durchaus diskussionswürdige Verfahrensvorschläge aus der Informatik vor (Kapitel 3).

Die Suche nach fairen Verfahren zur Gestaltung fairer Algorithmen hat vielleicht die empirisch vorfindliche Kontrollarchitektur aus dem Blick geraten lassen. Es existiert eine ganze Reihe an (zivilgesellschaftlichen) Initiativen, die die Entwicklung von gesellschaftlich relevanten Algorithmen akribisch verfolgen. Gut möglich also, dass Algorithmen unter intensiverer Kontrolle und Beobachtung stehen als viele andere Regulierungsgegenstände. Kapitel 4 zielt dennoch darauf ab, die Community der Technikfolgenab-

schätzung (TA) nicht zuletzt als Verfahrensexpertin in die Suche nach fairen Gestaltungsmöglichkeiten fairer Algorithmen zu integrieren.

2. Die Entdeckung von Fairness im Plural

Spätestens mit der investigativ-journalistischen Studie „Machine Bias. There is software that is used across the county to predict future criminals. And it is biased against blacks" von *ProPublica* hat die Diskussion um „algorithmic fairness" Fahrt aufgenommen (Angwin et al. 2016). Da der Einsatz von Algorithmen häufig durch ihre Neutralität begründet wird, also gerade durch eine Emanzipation von typisch menschlicher Voreingenommenheit, überrascht die diskursive Wucht der Studienergebnisse kaum.

Im Mittelpunkt der Untersuchung stand eine Software namens COMPAS („Correctional Offender Management Profiling for Alternative Sanctions"). Dabei handelt es sich um einen sogenannten Rückfälligkeitsvorhersagealgorithmus (vgl. König/Krafft 2021). Einsatzfeld dieses Instruments, das auf dem Markt nicht singulär ist, sind Vorverhandlungen („Pretrial Hearings"). Vor allem aus dem bisherigen Strafregister soll die Software die Wahrscheinlichkeit dafür errechnen, ob Angeklagte zur Hauptverhandlung erscheinen und bis dahin wieder straffällig werden. Es geht dabei ausdrücklich um Entscheidungsassistenz, die Entscheidung – etwa zu Verwahrung oder Entlassung – bleibt eine richterliche.

Zentrales Ergebnis von „Machine Bias" war, dass sowohl bei den Falsch Positiven (Rückfall vorhergesagt, aber nicht eingetreten) als auch bei den Falsch Negativen (Rückfall nicht vorhergesagt, aber eingetreten) der Anteil von African Americans unverhältnismäßig hoch ausfällt. Die Studie und ihr Kernresultat wurden sowohl medial als auch wissenschaftlich überaus rege aufgenommen; Google Scholar weist allein über 1.500 wissenschaftliche Zitationen aus.

Etwas weniger populär geworden sind Studien, die mit dem gleichen Datensatz zu anderen Schlussfolgerungen gekommen sind (Flores et al. 2016; Fenton/Neil 2018). Aus ihnen ging hervor, dass der Algorithmus keinesfalls unrechtmäßig diskriminiere, sondern vielmehr gleichbehandle. Erst aus dieser Gegenüberstellung stellte sich die Erkenntnis ein, dass es unterschiedliche, mathematisch inkompatible Vorstellungen von Fairness gibt (vgl. Chouldechova 2017; Kleinberg et al. 2016, 2019). Katharina Zweig und Tobias Krafft (2018, S. 224) bemerkten am selben Fall, dass es interessant sei, dass dieser mathematische Widerspruch nicht früher entdeckt worden sei. Im selben Papier wird formuliert, dass solche Probleme bei der Wahl des richtigen Fairnessmaßes zeigen, „dass die dafür notwendigen Diskussionen nicht nur von einem kleinen Team an Informatikern geführt werden

dürfen, sondern eines breiter geführten Diskurses bedürfen" (ebd., S. 218). Das beziehen die aus der Informatik stammenden Verfasser*innen auch auf sich selbst. Diese Forderung ist nicht selten anzutreffen. Ein weiteres Beispiel kommt aus dem Gutachten der von der deutschen Bundesregierung eingesetzten Datenethikkommission: „Welche Kriterien für Nicht-Diskriminierung und Gerechtigkeit in welchem Kontext angemessen sind, ist keine technische, sondern eine gesellschaftliche und politische Frage. Daher dürfen diese Entscheidungen auch nicht allein den Technik-Entwicklern überlassen werden" (Datenethikkommission 2019, S. 169).

Informatik und Technikentwicklung sollen nicht allein über Fragen algorithmischer Fairness, über Diskriminierung und Gerechtigkeit entscheiden. Der nächste Absatz zeigt, dass solche Forderungen offene Türen bei ihren Adressat*innen einrennen.

3. Fairness als Voraussetzung und Ziel von Verfahren

Richard Berk und seine Kolleg*innen aus den Bereichen der Statistik, der Kriminologie und der Informatik sind, in der Diskussion um dieselben Rückfälligkeitsvorhersagealgorithmen, sogar auf fünf quantitative Fairnessbegriffe gekommen: Gleiche Genauigkeit („Overall accuracy equality"); Statistische Parität („Statistical parity"); Gleiche bedingte Gruppengenauigkeit („Conditional procedure accuracy equality"); Gleiche bedingte Vorhersagegenauigkeit („Conditional use accuracy equality"); Gleiches Fehlerverhältnis („Treatment equality") (Berk et al. 2018, S. 15f.).

Die Autor*innen fassen zwar alle fünf Begriffe algorithmischer Fairness im Konzept der „total fairness" zusammen, weisen jedoch unmittelbar darauf hin, dass diese praktisch unerreichbar sei, da sich gleiche Vorhersagegenauigkeit und gleiche Gruppengenauigkeit regelmäßig gegenseitig ausschließen. Generell seien Genauigkeit und Gerechtigkeit („accuracy and fairness") konfligierende Ziele. Damit beenden sie ihre Ausführungen aber nicht. Vielmehr beschäftigen sie sich gerade mit der Frage, was aus dieser mathematischen Schlussfolgerung gesellschaftlich folgt. Sie weisen explizit darauf hin: „in the end, it will fall to stakeholders – not criminologists, not statisticians and not computer scientists – to determine the tradeoffs. […] These are matters of values and law, and ultimately, the political process. They are not matters of science" (ebd., S. 35).

Sie fordern eine Stakeholder-Beteiligung zur Bearbeitung des Fairness-Problems, also gerade keine technische Lösung – allenfalls eine soziale, der sich dann wieder Technik anschließen könnte: „if there is a policy preference, it should be built into the algorithm" (ebd., S. 29). Sobald eine politische Präferenz benennbar wird, könne diese wiederum in Code übersetzt

werden. Nun lässt sich an dieser Stelle mit guten Gründen die offene Frage stellen, wie unterschiedliche soziale Vorstellungen von Fairness, Diskriminierung und „Normalität" in mathematische Modelle übersetzt werden können (Pöchhacker/Kacianka 2021, S. 7). Gleichwohl steht im Folgenden im Mittelpunkt, dass aus der Informatik selbst vernehmbar ist, dass wenn es die *eine* Fairness nicht gibt, es zu einer Norm- und Wertfrage wird, welche Art der Fairness in jeweiligen Fällen die gesellschaftlich gewünschte ist. Die Informatik erklärt es gerade nicht zu ihrem Geschäft, dies zu entscheiden und zu bestimmen, in welcher Weise wiederum hierüber zu entscheiden sei.

Spätestens hier also wird die Frage virulent, welche Verfahren gewährleisten, dass über Fairness-Fragen sachlich und sozial ausgewogen entschieden werden kann. Dies verweist unmittelbar auf das Feld der „procedural justice" (vgl. Bora/Epp 2000). In ihrer expliziten Anwendung dieses Konzepts auf das Feld der Künstlichen Intelligenz (KI) machen Frank Marcinkowski und Christopher Starke (2019) eher den Punkt, dass „sich die Machine-Learning-Literatur der Informatik vornehmlich mit Möglichkeiten und Grenzen der mathematischen Formalisierung sozialwissenschaftlicher Vorstellungen von gerechten Verteilungsnormen beschäftigt" habe (Marcinkowski/Starke 2019, S. 272). Um also zu fairen Algorithmen zu kommen, bedarf es offenbar fairer Verfahren. Während mit „fairen Algorithmen" wesentlich Diskriminierungsfreiheit gemeint ist, kommt „fairen Verfahren" die Bedeutung zu, eine breite Beteiligung über technische Expertise hinaus zu gewährleisten. Hierfür sind Informatik und Entwicklung aber nicht in der Bringschuld, vielmehr wird hier gerade auf Stakeholder-Beteiligung, den politischen Prozess und allgemeiner auf gesellschaftliche und politische Diskurse verwiesen.

Wer diese Debatte um algorithmische Fairness verfolgt, kann den Eindruck gewinnen, dass die Techniker*innen und Informatiker*innen nun hinreichend zugewartet hätten, ob Politik und/oder Gesellschaft den von ihnen zugespielten Ball annehmen. Zumindest hat derselbe Autor, Richard Berk, mit einer anderen Co-Autorin, Ayya A. Elzarka, Data Analyst bei *Google*, inzwischen ein weiteres Papier namens „Almost politically acceptable criminal justice risk assessment" vorgelegt. Erneut werden Rückfälligkeitsvorhersagealgorithmen aus Strafrechtsverfahren als Anwendungsfälle gewählt.

Als beinahe politisch akzeptabel – „almost politically acceptable" – werden Effekte algorithmenbasierter Entscheidungen oder Entscheidungshilfe hierin dann bezeichnet, wenn eine hinreichende Zahl an Stakeholdern zustimmen kann („*a sufficient number of stakeholders can agree*") (Berk/Elzarka 2020, S. 5; Herv. i.O.). Wieder wird darauf verwiesen, dass eine genauere Bestimmung jenseits der eigenen Expertise liege und man zum Gerechtig-

keitsbegriff besser bei John Rawls oder Amartya Sen nachfrage (vgl. ebd.). Beinahe politisch akzeptabel könnte eine algorithmische Bewertung z.B. dann sein, wenn der Algorithmus nur mit Daten der privilegiertesten Gruppe gefüttert und trainiert wird. Im Beispiel von eingesetzten Algorithmen im US-amerikanischen Strafrecht also würde dies bedeuten, nur Trainingsdaten von Weißen zu verwenden und somit alle später damit untersuchten Verdächtigen wie Weiße zu behandeln (vgl. Berk/Kuchibhotla 2020). Dies bearbeitet gewissermaßen den Standardeinwand der anglo-amerikanischen Debatte, dass historische Daten schon deswegen als gerechte oder auch nur realitätsgetreue ausfallen, weil diese Daten immer schon in systematisch benachteiligender Weise erfasst worden sind. „Overpolicing" etwa meint dann, dass African Americans schon dank häufigerer Kontrollen auch häufiger im Strafregister auftauchen müssen.

Explizit gegen Berk führt Rajiv Movva (2021) aus, dass die Frage nach fairen Verfahren vom eigentlichen Problem der Ungerechtigkeit ablenke. Ein fairer Algorithmus sei insofern ein Ding der Unmöglichkeit als das Justizsystem, aus dem alle verfügbaren Daten stammen, immer schon voreingenommen und dieser Bias nicht herauszurechnen sei. Eher müsse man umgekehrt fragen, ob Algorithmen nicht stattdessen dazu verwendet werden könnten, eine systematischere Analyse des richterlichen Handelns durchzuführen, um hierin implizite Muster zu explizieren (vgl. Movva 2021, S. 8f.). Darin ist schon wieder eine Annäherung an Berk zu erkennen, der an anderer Stelle ausführt, dass ein Vorteil von Code gegenüber dem menschlichen Gehirn sei, Risikoeinschätzungen nur in expliziter Form vornehmen zu können. Wenn die Software nicht proprietär ist, kann der Code studiert und ggf. korrigiert werden. Die meisten Stakeholder werden nicht die Fähigkeit oder Neigung haben, diese Aufgabe zu übernehmen, aber in vielen Universitäten und Unternehmen gäbe es Personen, die Hilfe anbieten könnten (vgl. Berk 2020, S. 231). Auch hierin ist eine Verfahrenskomponente zu erkennen.

Berk und Elzarka (2020) sind selbst nicht davon überzeugt, dass ihr Vorschlag einer beinahe politisch akzeptablen Risikobewertung alle Stakeholder überzeugte. Daher bedürfe es eines gerechten Verfahrens. Das Autorenduo gibt auch hierfür noch denkbare Beispiele: Beratungen eines Stadtrates oder anderer gesetzgebender Körperschaften, ein ständiger Ausschuss aus Vertretern von Interessengruppen, eine Kommission oder ein offizieller Beirat, in wieder anderen Fällen kann es einen Ad-hoc-Aufsichtsausschuss mit einer breiten Vertretung von Stakeholdern geben usw. In der Praxis gäbe es eine Fülle von Details auszuarbeiten, z.B. welche Interessengruppen bzw. Stakeholder teilnehmen können und welche Verfahrensregeln zu verabschie-

den wären, aber: „Further discussion would be a lengthy diversion and beyond the expertise of the authors“ (Berk/Elzarka 2020, S. 5).

Statt also zu fordern, der Technik-Expertise nicht die Deutungs- und Handlungshoheit zu überlassen, wäre offenbar eher zu fragen, warum Politik und Gesellschaft der Aufforderung der Informatiker*innen sowie der Entwickler*innen nicht folgen, über algorithmische Fairness oder politische Akzeptabilität ins Gespräch zu kommen.

4. Fazit: Kontrollarchitektur statt Verfahren?

Algorithmen umgibt die Aura des gleichzeitig Intransparenten wie Wirkmächtigen. Ihre Regulierung ist ein Spezialthema geblieben; politisch ist (bislang) mit entsprechenden Forderungen eine Wahl weder zu gewinnen noch zu verlieren. Dies trifft bei weitem nicht nur auf Algorithmen zu. Helmut Willke (2019) hat diesbezüglich das Modell der „reflexiven Repräsentativität“ vorgeschlagen. Komplexe Themen, die Bürger*innen, Parlamente und Regierungen überforderten, sollten durch Spezialsenate behandelt werden. Diese Expertisegremien müssten politisch verfasst beauftragt werden und könnten dann Fragen der Finanzregulierung, des Klimawandels, der Migration oder eben der Digitalisierung kompetenter bearbeiten (vgl. Willke 2019, S. 267). Unter den Bedingungen einer derart komplexen Gesellschaft ließen sich Fragen der Freiheit nur noch prozedural bestimmen (ebd., S. 259).

Gleichwohl kann die Suche nach angemessenen Verfahren durch eine Begutachtung der bestehenden de facto-Kontrollarchitektur zu algorithmischer Fairness ergänzt werden (vgl. Wischmeyer 2018). So wird sichtbar, dass es zwar womöglich keine Verfahren zur fairen Gestaltung fairer Algorithmen gibt, wohl aber zahlreiche mehr oder weniger private Kontrollinstanzen. Beispiele hierfür wären: *AlgorithmWatch*, *Open Sourced*, *ProPublica*, *The Algorithmic Justice League*, *The Information*, *The Markup* oder *The Protocol*. Diese offene Liste umfasst nur „zivilgesellschaftliche Watchdogs“ (Speth 2018). Zusätzlich ließe sich noch an Kollektive wie „White-Hat-Hacker“ oder gar an individuelle Datenschutzaktivist*innen denken, für die Max Schrems das vielleicht prominenteste und einflussreichste Beispiel wäre. Gleichwohl zeigt die von ihm gegründete *None of Your Business* (noyb) die Bedeutung organisierter Einflussnahme auf.[1] Gut möglich also, dass Algorithmen unter stärkerer Kontrolle und Beobachtung stehen als viele andere Regulierungsgegenstände, nur eben jenseits hoheitlicher Aufsicht.

1 Siehe https://noyb.eu/de [aufgesucht am 19.07.2021].

The Markup, gegründet übrigens von der ehemaligen *ProPublica*-Journalistin und Machine Bias-Autorin Julia Angwin, hat ein eigenes Verfahren kreiert. Das Citizen-Browser-Projekt ist ein von *The Markup* entwickelter Webbrowser.[2] Dieser überprüft die Algorithmen, mit denen Social Media-Plattformen bestimmen, welche Informationen sie ihren Usern zur Verfügung stellen, welche Nachrichten und Erzählungen verstärkt oder unterdrückt werden und welche Online-Gemeinschaften diese Nutzer*innen zum Beitritt ermutigen. Zunächst wird der Browser implementiert, um Daten von *Facebook* und *YouTube* zu sammeln.

Ein national repräsentatives Panel von 1.200 Personen wird dafür bezahlt, den angepassten Webbrowser auf ihren Desktops zu installieren, der es ihnen ermöglicht, Echtzeitdaten direkt von ihren *Facebook*- und *YouTube*-Konten mit *The Markup* zu teilen. Die von diesem Panel gesammelten Daten bilden statistisch valide Stichproben der amerikanischen Bevölkerung über Alter, *Race*, *Gender*, Geographie und politische Zugehörigkeit hinweg. Persönlich identifizierbare Informationen des Panels werden von *The Markup* entfernt.

Man kann dies durchaus als einen Versuch partizipativer Algorithmenkontrolle auffassen, keinesfalls aber als letztes Wort hierzu. Das Projekt leistet in vielerlei Hinsicht Pionierarbeit; Fairness allerdings wird hier geradezu klassisch über Repräsentativität aufgelöst.

Auch Berk (2020) beobachtet in den Vereinigten Staaten, Europa und andernorts, dass politische Institutionen zu reagieren beginnen. Es bildeten sich Bürgergruppen und viele NGOs engagierten sich. Mitunter werde sogar Technologie zur Gegenüberwachung entwickelt. Inzwischen seien etwa Brillen erhältlich, die das Licht einer Closed Circuit Television (CCTV)-Kamera streuen, um eine Gesichtserkennung zu verhindern (Berk 2020, S. 232).

Diese Art der Technikentwicklung wird von der TA typischerweise eher begleitet als zur Gegenkontrolle eingesetzt. Insofern scheint es angezeigt, dass sich die TA-Community nicht zuletzt als Verfahrensexpertin erkennbar einbringt in die Überlegungen zur algorithmischen Fairness. Die von Seiten der Informatik und der Technikentwicklung ins Auge gefassten Verfahren setzen, wie gezeigt, vor allem auf Stakeholder-Beteiligung. Solche Verfahren integrieren in erster Linie sachlich. „Partizipativer TA geht es vor allem darum, die Sach- und Sozialdimension in spezifischer Hinsicht zu verknüpfen und daraus Optionen für die Politikberatung zu schaffen" (Abels/Bora 2013: 114f.). Die Wahl eines Verfahrenstyps muss sich gegenüber drei Fragen als sinnvoll erweisen: *Wer* wird *wie* und *wozu* beteiligt? (vgl. ebd.) Wenn aus Algorithmenregulierung doch ein breiteres gesellschaftliches Thema werden

2 Siehe https://themarkup.org/series/citizen-browser [aufgesucht am 19.07.2021].

soll, würde der Blick in die „TA-Toolbox“ sich weniger auf Abstimmungen zwischen unterschiedlichen Interessenvertretungen (z.B. „Dialogverfahren“ oder „‚pTA‘ im engeren Sinne“; ebd.: S. 123ff.) und vielmehr auf Modelle zur Initiierung auch öffentlicher Debatten (z.B. „(erweiterte) Konsensuskonferenzen“, „Voting Conferences“; ebd.) richten. Bislang ist Algorithmenregulierung, wie dieser Beitrag gezeigt hat, ein Expertenthema – allerdings eines, auf das weit mehr Expert*innen einen Kontrollblick haben als mitunter angenommen wird.

Literatur

Abels, G.; Bora, A. (2013): Partizipative Technikfolgenabschätzung und -bewertung. In: Simonis, G. (Hg.): Konzepte und Verfahren der Technikfolgenabschätzung. Wiesbaden, S. 109–128

Aneesh, A. (2009): Global Labor: Algocratic Modes of Organization. In: Sociological Theory 27(4), S. 347–370

Angwin, J.; Larson, J.; Mattu, S.; Kirchner, L. (2016): Machine Bias. There is software that is used across the county to predict future criminals. And it is biased against blacks. ProPublica; https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [aufgesucht am 19.07.2021]

Berk, R.A. (2020): Artificial Intelligence, Predictive Policing, and Risk Assessment for Law Enforcement. In: Annual Review of Criminology 4(1), S. 209–237

Berk, R.; Elzarka, A.A. (2020): Almost politically acceptable criminal justice risk assessment. In: Criminology & Public Policy, S. 1–27

Berk, R.; Heidari, H.; Jabbari, S.; Kearns, M.; Roth, A. (2018): Fairness in Criminal Justice Risk Assessments: The State of the Art. In: Sociological Methods & Research 50(1), S. 3–44

Berk, R.; Kuchibhotla, A.K. (2020): Improving Fairness in Criminal Justice Algorithmic Risk Assessments Using Conformal Prediction Sets; https://arxiv.org/abs/2008.11664 [aufgesucht am 19.07.2021]

Bora, A.; Epp, A. (2000): Die imaginäre Einheit der Diskurse. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 52(1), S. 1–35

Chouldechova, A. (2017): Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments. In: Big Data 5(2), S. 153–163

Datenethikkommission der Bundesregierung (2019): Gutachten der Datenethikkommission. Bundesministerium des Innern, für Bau und Heimat. Berlin; https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.pdf [aufgesucht am 19.07.2021]

Fenton, N.E.; Neil, M. (2018): Criminally Incompetent Academic Misinterpretation of Criminal Data – and how the Media Pushed the Fake News; https://www.researchgate.net/publication/322603937 [aufgesucht am 19.07.2021]

Flores, A.W.; Bechtel, K.; Lowenkamp, C.T. (2016): False Positives, False Negatives, and False Analyses: A Rejoinder to Machine Bias: There's Software Used across the Country to Predict Future Criminals. And It's Biased against Blacks. In: Federal Probation 80(2), S. 38–46

Kleinberg, J.; Ludwig, J.; Mullainathan, S.; Sunstein, C.R. (2019): Discrimination in the Age of Algorithms; https://doi.org/10.2139/ssrn.3329669 [aufgesucht am 19.07.2021]

Kleinberg, J.; Mullainathan, S.; Raghavan, M. (2016): Inherent Trade-Offs in the Fair Determination of Risk Scores; https://arxiv.org/pdf/1609.05807 [aufgesucht am 19.07.2021]

König, P.D.; Krafft, T.D. (2021): Evaluating the evidence in algorithmic evidence-based decision-making: the case of US pretrial risk assessment tools. In: *Current Issues in Criminal Justice* 33(3), S. 359–381

Marcinkowski, F.; Starke, C. (2019): Wann ist Künstliche Intelligenz (un-)fair? Ein sozialwissenschaftliches Konzept von KI-Fairness. In: Hofmann, J.; Kersting, N.; Ritzi, C. (Hg.): Politik in der digitalen Gesellschaft. Zentrale Problemfelder und Forschungsperspektiven. Bielefeld, S. 269–288

Movva, R. (2021): Fairness Deconstructed: A Sociotechnical View of "Fair" Algorithms in Criminal Justice; https://arxiv.org/pdf/2106.13455 [aufgesucht am 19.07.2021]

Pöchhacker, N.; Kacianka, S. (2020): Algorithmic Accountability in Context. Socio-Technical Perspectives on Structural Causal Models. In: frontiers in Big Data 3, S. 1–9

Speth, R. (2018): Machtkontrolle durch Watchdogs. In: Forschungsjournal Soziale Bewegungen 31(3), S. 6–18

Willke, H. (2019): Komplexe Freiheit. Konfigurationsprobleme eines Menschenrechts in der globalisierten Moderne. Bielefeld

Wischmeyer, T. (2018): Regulierung intelligenter Systeme. In: Archiv des öffentlichen Rechts 143(1), S. 1–66

Zweig, K.A.; Krafft, T.D. (2018): Fairness und Qualität algorithmischer Entscheidungen. In: Mohabbat-Kar, R.; Thapa, B.E.P.; Parycek, P. (Hg.): (Un)berechenbar? Algorithmen und Automatisierung in Staat und Gesellschaft. Berlin, S. 204–227