

Was zu tun ist, damit Maschinen den Menschen dienen

Krüger, Julia; Lischka, Konrad

Veröffentlichungsversion / Published Version

Sammelwerksbeitrag / collection article

Empfohlene Zitierung / Suggested Citation:

Krüger, J., & Lischka, K. (2018). Was zu tun ist, damit Maschinen den Menschen dienen. In R. Mohabbat Kar, B. E. P. Thapa, & P. Parycek (Hrsg.), *(Un)berechenbar? Algorithmen und Automatisierung in Staat und Gesellschaft* (S. 440-470). Berlin: Fraunhofer-Institut für Offene Kommunikationssysteme FOKUS, Kompetenzzentrum Öffentliche IT (ÖFIT). <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-57617-2>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

<https://creativecommons.org/licenses/by/3.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:

<https://creativecommons.org/licenses/by/3.0>

Was zu tun ist, damit Maschinen den Menschen dienen

Julia Krüger & Konrad Lischka

freie Sozialwissenschaftlerin & Bertelsmann Stiftung, Gütersloh

Prozesse algorithmischer Entscheidungsfindung berühren in vielen Bereichen gesellschaftliche Teilhabe. Daraus ergeben sich Herausforderungen auf vier Feldern:

1. Gesellschaftliche Angemessenheit (Haben algorithmische Systeme sinnvolle Optimierungsziele?)
2. Überprüfen und erklären der Umsetzung (Verwirklichen algorithmische Systeme die sinnvollen Ziele?)
3. Schaffen von Diversität (Ist die Vielfalt der Systeme und Betreibermodelle groß genug, um gesellschaftlich angemessen zu sein?)
4. Übergreifende Rahmenbedingungen für teilhabeförderliche algorithmische Systeme (Genügen staatliche und individuelle Gestaltungskompetenzen dem gesellschaftlichen Anspruch?)

Der Beitrag stellt eine Auswahl von Lösungsideen in allen vier skizzierten Handlungsfeldern vor. Die Auswahl ist nicht umfassend und abschließend. Sie veranschaulicht aber auch in dieser Form, dass Akteure aus allen drei Sektoren Möglichkeiten haben, algorithmische Entscheidungsfindung für mehr gesellschaftliche Teilhabe zu gestalten. Es gibt viele Ideen für Maßnahmen und Methoden bzgl. der gesellschaftlichen Gestaltung, Intervention und Kontrolle algorithmischer Entscheidungsfindung. Keineswegs scheint der Mensch der Maschine ausgeliefert.

1. Worum es geht: Gesellschaftliche Anforderungen an algorithmische Entscheidungssysteme

Wie sehr algorithmische Entscheidungssysteme den Alltag durchdringen können, sieht man in New York. Dort bestimmen algorithmische Systeme, auf welche weiterführende Schule Kinder kommen,¹ wo die Polizei wie häufig Streife fährt und kontrolliert,² welche Lehrer Karriere machen,³ welche Gebäude vorrangig auf Brandschutz inspiziert werden⁴ und wer des Sozialleistungsbetrugs verdächtigt wird⁵.

Befürworter solcher Systeme führen eine Reihe von Vorteilen an, die sich grob in diese drei Bereiche gliedern lassen⁶:

- **Fairness und Konsistenz:** Algorithmenbasierte Prognosen arbeiten zuverlässig die vorgegebene Entscheidungslogik in jedem Einzelfall ab. Im Gegensatz zu Menschen ist Software nicht tagesformabhängig und wendet nicht in Einzelfällen willkürlich neue, unter Umständen ungeeignete Kriterien an. Gesellschaftlich nicht angemessene Kriterien lassen sich von vornherein ausschließen.
- **Umgang mit Komplexität:** Software kann größere Datenmengen analysieren als Menschen. So lassen sich Muster finden, anhand derer einige Probleme schneller, präziser oder günstiger gelöst werden können. Algorithmische Systeme können ihren Output

¹ Tullis, 2014

² Brennan Center for Justice, 2017

³ O'Neil, 2017

⁴ Heaton, 2015

⁵ Singer, 2015

⁶ vgl. Lischka & Klingel, 2017, S. 37f

günstig personalisieren und sie lassen sich neuen Umständen leichter anpassen als analoge Strukturen. Das in New York eingesetzte algorithmische System zur Schülerverteilung senkte zum Beispiel im ersten Jahr nach Einführung die Anzahl keiner weiterführenden Schule zugeteilter Schüler von 31.000 auf 3000.⁷

- Effizienz: Maschinen werten große Datenmengen in der Regel günstiger und schneller aus als Menschen das vergleichbare Pensum verarbeiten würden. Eine einmal entwickelte Entscheidungslogik eines Systems lässt sich günstig auf viele Fälle anwenden. In New York lobt die Feuerwehr zum Beispiel die Effizienz der zentralisierten, algorithmischen Auswertung von Gebäudedaten im Vergleich zum alten papierbasierten und auf 26 Standorte verteilten Verfahren.⁸

Der Einsatz algorithmischer Entscheidungssysteme garantiert nicht, dass diese Chancen tatsächlich verwirklicht werden. Auch das zeigt der Einsatz in New York. Die Herausforderungen für Teilhabechancen durch den Einsatz algorithmischer Systeme lassen sich grob in vier Felder unterteilen.

1.1. Gesellschaftliche Angemessenheit – haben algorithmische Systeme sinnvolle Optimierungsziele?

Der Rechnungshof in New York lobt das algorithmische System zur Schülerverteilung, weil es viel mehr Schüler den von ihnen präferierten Schulen zuordnet als das alte Verfahren. Zugleich zweifeln

⁷ New York City Independent Budget Office, 2016

⁸ Heaton, 2015

die Autoren des Berichts an, dass die Erfüllung individueller Wünsche wirklich das gesellschaftlich sinnvollste Optimierungsziel ist. Ihr Gegenargument: Das System teilt unterdurchschnittlich benotete Schüler meist unterdurchschnittlich bewerteten Schulen zu. Denn das entspricht den Präferenzen. Und doch benachteiligt diese Art der Verteilung systematisch Schüler aus ärmeren Vierteln, wo sich in der Stadt unterdurchschnittlich bewertete Schulen und unterdurchschnittlich benotete Schüler ballen.⁹

Hier geht es nicht um die Effizienz und Konsistenz des Systems, sondern um das Optimierungsziel: Soll die Verteilung individuelle Schulpräferenzen in möglichst vielen Fällen befriedigen? Oder sollen das System die Bildungschancen vom soziodemografischen Hintergrund entkoppeln? Beide Ziele sind vertretbar. Dass ein Prozess algorithmischer Entscheidungsfindung (im Folgenden als ADM abgekürzt nach dem im Englischen üblichen Begriff *algorithmic decision-making*)¹⁰ zuverlässig und nachvollziehbar arbeitet, sagt wenig über seinen gesellschaftlichen Sinn aus. Welches Ziel die Stadt und damit das von ihr beauftragte System verfolgen sollen, sollten in einem politischen Willensbildungsprozess möglichst viele Bürger und vor allem potenziell Betroffene mitbestimmen können. Hier geht es um die der technischen Umsetzung vorgelagerten Fragen, die nicht anhand technischer Kriterien zu beantworten sind. Die Gestaltung algorithmischer Systeme, die Teilhabechancen berühren, setzt so gut wie immer solche wertorientierten Zieldefinitionen voraus. Was einen guten Arbeitnehmer ausmacht, was eine relevante journalistische Nachricht auszeichnet, woran eine wichtige Freundschaft zu erkennen ist – auf solche Fragen gibt es keine eindeutig richtigen Antworten. Solche sozialen Konzepte müssen die Gestalter

⁹ New York City Independent Budget Office, 2016

¹⁰ Verwendet zum Beispiel in: USACM 2017; Ananny und Crawford 2016; Goodman und Flaxman 2016, Mittelstadt 2016a; Zarsky 2016.

von algorithmischen Systemen operationalisieren und messbar machen. Zur Entwicklung muss dabei auch gehören, die Ziele gesellschaftlich angemessen breit zu diskutieren. Das ist notwendig, um gesellschaftlicher Dynamik Raum zu geben. Sonst würde man im schlimmsten Fall durch algorithmische Systeme z. B. lediglich im Trainingsdatensatz geronnene gesellschaftliche Zustände der Vergangenheit fortschreiben.

Bei dem algorithmischen System zur Schülerverteilung in New York wirkt vermutlich ein ähnlicher Effekt der Status-quo-Reproduktion: Die Standorte überdurchschnittlich erfolgreicher Schulen und Schüler hängen mit der Verteilung von Reichtum und Bildung im Stadtgebiet zusammen. Diese Variable beeinflusst die Präferenzen der Schüler und das Verteilungsergebnis, unabhängig vom algorithmischen System zur Auswahl.

1.2. Umsetzung prüfen, erklären, falsifizieren: Verwirklichen algorithmische Systeme die sinnvollen Ziele?

Gut gemeint ist nicht gut gemacht: Algorithmische Systeme mit gesellschaftlich angemessenen Optimierungszielen können Teilhabe mindern, wenn es an der Umsetzung hapert. Um die Qualität eines algorithmischen Systems zu beurteilen, muss auch die Umsetzung im Einsatz untersucht werden.

Ein bekanntes Beispiel für solche Analysen ist die 2016 veröffentlichte Recherche der US-Rechercheorganisation Propublica zur Qualität algorithmischer Rückfallprognosen, die in vielen US-Bundesstaaten vor Gericht genutzt werden. Die Software war zu diesem Zeitpunkt seit Jahren im Einsatz, doch zuvor hatte niemand systematisch überprüft, welche Fehler bei den Prognosen auftreten. Kernergebnis der Propublica-Recherche: Die Art der Fehlprognosen

unterscheidet sich zwischen schwarzen und weißen Personen. Der Anteil Schwarzer mit hoher Rückfallprognose aber ohne Rückfall binnen zwei Jahren ist doppelt so hoch wie der Weißer.¹¹ Erst diese Rechercheergebnisse brachten die Diskussion über Fairnesskriterien der seit Jahren eingesetzten Systeme in Gang.

In New York hat die fehlende Nachvollziehbarkeit eines algorithmischen Systems zur Bewertung von Lehrern dazu geführt, dass ein Gericht den Einsatz dieser Software untersagte. Das System habe »willkürliche« und »unbeständige« (*»arbitrary and capricious«*) Ergebnisse geliefert, hieß es in der Urteilsbegründung.¹² Fehlende Überprüfbarkeit und Nachvollziehbarkeit sind auch die Hauptkritikpunkte an einem algorithmischen System zur Streifenplanung der New Yorker Polizei. Der Stadtrat James Vacca formuliert seine Bedenken so: Die Polizei habe ihm als Volksvertreter die Kriterien und Entscheidungslogik für die Einsatzplanung in der Bronx nie hinreichend erklären können.¹³

1.3. Diversität schaffen: Vielfältige Systeme und Betreibermodelle dienen der gesellschaftlichen Angemessenheit

Die gesellschaftliche Angemessenheit der Optimierungsziele und die Qualität der Umsetzung sind immer an einzelnen algorithmischen Systemen zu bewerten. Doch es gibt auch auf der darüber liegenden Ebene Handlungsbedarf: Eine große Vielfalt ist ein anzustrebender Zustand. Vielfalt in zweierlei Sinn:

- Vielfalt der Systeme: Unterschiedliche Optimierungsziele und Operationalisierungen in einem Einsatzgebiet.

¹¹ Angwin, Kirchner, Larson, & Mattu, 2016, S. 2

¹² Harris, 2016

¹³ Powles, 2017

- Vielfalt der Sektoren: Auftraggeber und Betreiber von ADM-Systemen aus dem öffentlichen, dem privatwirtschaftlich und dem zivilgesellschaftlichen Sektor

Vielfalt ist hier ein Wert an sich. Dafür sprechen vor allem diese drei Gründe:

- Skalierbarkeit: Die einmal entwickelte Entscheidungslogik eines ADM-Prozesses ist auf sehr viele Fälle anwendbar, ohne dass die Kosten für den Einsatz substantiell steigen. Das führt dazu, dass in einigen Lebensbereichen wenige algorithmische Systeme Verfahren dominieren können. Je größer die Reichweite eines algorithmischen Systems ist, desto schwieriger ist es für den Einzelnen, sich den Verfahren und ihren Folgen zu entziehen. Je geringer die Vielfalt algorithmischer Systeme in einem Einsatzbereich ist, desto härter treffen Fehler in der Umsetzung die Betroffenen.
- Gesellschaftliche Dynamik: Für soziale Phänomene oder Konzepte, wie zum Beispiel Nachrichtenrelevanz oder Eignung von Bewerbern, existieren viele richtige Operationalisierungen. Solche Konzepte unterliegen dem gesellschaftlichen Wandel. Je geringer die Vielfalt algorithmischer Systeme in einem Einsatzbereich ist, desto kleiner wird der Raum für gesellschaftliche Entwicklung und Vielfalt.
- Innovation: Wenn in einem Einsatzfeld unterschiedliche algorithmische Systeme im Einsatz sind, kann der Vergleich zwischen ihnen Erkenntnisse über Wirkung, Fehlerquellen und Alternativen befördern.

1.4. Übergreifende Rahmenbedingungen für teilhabeförderliche algorithmische Systeme

Handlungsbedarf auf den drei oben aufgeführten Feldern führt zu einem vierten: Es braucht kompetente Akteure, um den Rahmen für eine positive Entwicklung zu gestalten. Gesellschaftlicher Nutzen stellt sich nicht von alleine ein.

Gemeinwohlfördernde Gestaltung umfasst dabei einerseits Regulierung zur Abwehr und Nachsorge. Andererseits ist der Staat aber auch als Gestalter und Ermöglicher gefragt. Das Beispiel New York zeigt erste positive und negative Beispiele. Handlungsbedarf besteht hier offenkundig bei der staatlichen Gestaltungskompetenz.

2. Was man tun kann: Lösungsansätze

Es folgt eine Auswahl von Lösungsideen in allen vier skizzierten Handlungsfeldern. Die Auswahl ist nicht umfassend und abschließend. Sie veranschaulicht aber auch in dieser Form, dass Akteure aus allen drei Sektoren Möglichkeiten haben, algorithmische Entscheidungsfindung für mehr gesellschaftliche Teilhabe zu gestalten.

2.1. Zielsetzung algorithmischer Systemen auf gesellschaftliche Angemessenheit prüfen

Welche Optimierungsziele gesellschaftlich sinnvoll sind, lässt sich nicht präskriptiv für alle algorithmischen Systeme festlegen. Definieren und Priorisieren sind dynamische Prozesse. Und jedes neue algorithmische System mit Wirkung auf gesellschaftliche Teilhabe muss Anlass sein, diesen Prozess weiterzutreiben. Dafür ist eine breite gesellschaftliche Diskussion der Ziele und Operationalisierungen nötig. Wie können Einsatz und Entwicklung von ADM-Systemen gestaltet sein, um diese Diskussionen zu fördern, damit die

gesellschaftliche Angemessenheit im gesellschaftlichen Diskurs erörtert wird?

2.1.1. Transparenz von Zielen, Methoden, Ergebnissen: Transparenz gegenüber Betroffenen

Von einem algorithmischen System bewertete Menschen müssen wissen, dass sie bewertet wurden. Und sie müssen genug über Grundlage, Tenor und Folgen der Bewertung wissen, um mit diesen Informationen eine Evaluation und Korrektur anstrengen zu können.

Die europäische Datenschutz-Grundverordnung sieht im Kap. 3, Art. 13–15 in Verbindung mit Art. 22 Informationspflichten im Falle der Erhebung und Verarbeitung von Daten zum Zwecke der automatisierten Entscheidungsfindung vor. Diese umfassen aussagekräftige Informationen über die a) involvierte Logik, b) die Tragweite sowie c) die angestrebten Auswirkungen – zumindest, wenn die Entscheidung gegenüber der betroffenen Person rechtliche Wirkung entfaltet oder sie in ähnlicher Weise erheblich beeinträchtigt. Eine Reihe von Ausnahmen und Einschränkungen lässt an der Wirksamkeit der EU-DSGVO in diesem Punkt zumindest Zweifel zu, für eine ausführliche Diskussion ist hier nicht der Platz.

Als Merkmale einer wirksamen Transparenz beschreibt eine Studie zu den Menschenrechtsdimensionen automatisierter Entscheidungsverfahren im Auftrag des Europarats¹⁴ diese:

- Ziele der algorithmischen Entscheidungsfindung
- Variablen im Sinne der Modellierung

¹⁴ Council of Europe - Committee of experts on internet intermediaries, 2017

- Informationen über Methoden (Trainingsdaten, statistische Kennzahlen sowie die Menge und Art der automatisierten Entscheidungen zugrundeliegenden Daten)

Citron hebt darüber hinaus hervor, dass solche Auskunftspflichten grundlegend dafür sind, dass rechtsstaatliche Verfahren (*»due process«*) in prozeduraler und inhaltlicher Hinsicht gewahrt werden können.¹⁵

Die Forderungen bleiben allerdings oft recht allgemein: Gelten sie nur für Entscheidungssysteme, welche den individuellen Handlungsspielraum beeinflussen? Oder gelten sie auch für solche, welche die Wahrnehmung beeinflussen, wie zum Beispiel soziale Netzwerke (Google, Facebook, Twitter) oder Plattformen, die dem Online-Versandhandel (Amazon) oder der Verteilung anderer Güter oder Dienstleistungen dienen?¹⁶

Wissenschaftler des Oxford Internet Institute haben hier einen instruktiven Vorschlag gemacht, der als Counterfactual Explanations – eine Art Anwendungs- oder Nutzungserläuterung für das algorithmische Entscheidungssystem - debattiert wird. Sie soll Betroffenen eine Möglichkeit geben, algorithmische Entscheidungen zu verstehen und anzufechten, sowie eine Orientierung über Möglichkeiten, die algorithmische Entscheidung in der Zukunft zu verändern. Im Kern beschreibt eine Counterfactual Explanation die kleinsten nötige Änderungen im Input für das ADM-System, um einen anderen Output zu erzielen:

¹⁵ Citron, 2008, S. 1281ff

¹⁶ Otto, 2017, S. 18ff

»These counterfactual explanations describe the smallest change to the world that would obtain a desirable outcome, or to arrive at a 'close possible world'.«¹⁷

Im einfachsten Fall, etwa bei der Kreditbewilligung, könnte so eine Anwendungserläuterung eine Auskunft dazu geben, wie hoch das Jahreseinkommen eines Antragstellenden sein müsste, damit ein nicht bewilligter Kreditantrag doch bewilligt worden wäre. Interessant ist die Methode, wenn viele Variablen oder Variablensets zusammenkommen, wenn das gesamte Variablenset - eine Entscheidungsmatrix - vorliegt. Diese würde durch Analyse der einer konkreten Entscheidung zugrundeliegenden Variablen generell verschiedene Erklärungen für eine konkrete Entscheidung bieten können. Für das betroffene Individuum wäre voraussichtlich die Erklärung am hilfreichsten, welche die individuell nächstmögliche Veränderung fokussiert, so die Autoren.¹⁸ Es geht damit um eine algorithmisch generierte Erklärung algorithmischer Entscheidungsverfahren.

Neben der allgemeinen Auskunft zur Anwendung und Funktionsweise algorithmischer Entscheidungssysteme stellt die Kenntnis über erhobene und zum Zwecke einer maschinellen Entscheidung verarbeitete Daten eine der Forderungen dar. Diese soll einerseits Individuen die Kontrolle über die sie betreffenden Entscheidungen geben und andererseits der Allgemeinheit eine adäquate Funktionsweise algorithmischer Entscheidungssysteme sichern. Citron und Pasquale¹⁹ führen am Beispiel von Kredit-Scoring aus:

¹⁷ Wachter, Mittelstadt & Russell, 2017, S. 1

¹⁸ ebd.

¹⁹ Citron & Pascale, 2014

»An important question is the extent to which the public should have access to the data sets and logic of predictive credit-scoring systems. We believe that each data subject should have access to all data pertaining to the data subject. Ideally, the logics of predictive scoring systems should be open to public inspection as well. There is little evidence that the inability to keep such systems secret would diminish innovation.«²⁰

Die Autoren der »Future of Life«-Konferenz, die der Entwicklung gemeinwohlorientierter Künstlicher Intelligenz gewidmet ist, formulieren kurz und knapp:

»People should have the right to access, manage and control the data they generate, given AI systems' power to analyze and utilize that data.«²¹

Sie gehen damit noch einen Schritt weiter und fordern ebenfalls die Möglichkeit, persönliche Daten zu korrigieren.

2.1.2. Verpflichtende Folgenabschätzung: Transparenz gegenüber der Öffentlichkeit

Transparenz- und Auskunftspflichten sollen betroffenen Individuen einen Überblick über die Grundlagen einer sie betreffenden Entscheidung geben und darauf aufbauend die Möglichkeit einer Anfechtung und Veränderung einer konkreten Entscheidung. Wie aber sieht es aus mit dem Überblick für die Gesellschaft?

Folgenabschätzungen bzw. Verträglichkeitsprüfungen für algorithmische Entscheidungssysteme können eine Grundlage darstellen, auf der die Öffentlichkeit einen Überblick über eingesetzte Systeme

²⁰ Citron & Pasquale, 2014, S. 26

²¹ Future of Life Institute, 2017

erhalten kann. Auf dieser Grundlage könne dem Einsatz fehlerhafter, diskriminierender oder schädlicher Systeme Einhalt geboten werden, so der Informationswissenschaftler Ben Shneiderman: Eine konkrete Möglichkeit besteht etwa darin, Systembetreiber dazu zu verpflichten, eine Folgenabschätzung bereitzustellen, die Grundlage bietet für ein unabhängiges Zulassungsverfahren (*»independent oversight review«*).²²

Diese Vorschläge orientieren sich an den in der EU und den USA bekannten Umweltverträglichkeitsprüfungen bzw. -erklärungen, die bei Bauvorhaben Informationen zu betroffenen Stakeholdergruppen und potenziellen Folgen geben. Eine Algorithmenverträglichkeitsprüfung bzw. -erklärung (*»algorithm impact statement«*) könnte der allgemeinen Öffentlichkeit Überblick geben über

- Ziele eines algorithmischen Entscheidungssystems,
- Qualität des Daten-Inputs, sowie
- erwartete Ergebnisse.²³

Daran orientiert wären beispielsweise Abweichungen von einer intendierten Funktionsweise identifizierbar. In der deutschen Diskussion wird eine ähnliche Idee unter dem Schlagwort »Beipackzettel« diskutiert – ein Dokument, das »Einsatzgebiet, Modellannahmen und gesellschaftliche Nebenwirkungen« eines algorithmischen Systems benennt.²⁴

In eine ähnliche Richtung geht eine sogenannte Diskriminierungsprüfung bzw. -erklärung (*»discrimination impact assessment«*), wie sie der Rechtswissenschaftler Andrew Selbst für den Bereich vorhersa-

²² Shneiderman, 2016, S. 13539

²³ ebd.

²⁴ Zweig, 2016

gebasierter Polizeiarbeit (*»predictive policing«*) vorschlägt: Hier stehen Effektivität und potenzielle Diskriminierungseffekte algorithmischer Entscheidungssysteme im Mittelpunkt der Betrachtung. Der Vorschlag beinhaltet den Vergleich unterschiedlicher Algorithmen und Modelle im gleichen Anwendungsbereich. Er würde der Öffentlichkeit eine Möglichkeit für die Auswahl, Mitgestaltung oder Entwicklung algorithmischer Entscheidungssysteme eröffnen. Chancen und Risiken der Technologien könnten langfristig informiert debattiert werden. Zudem würde das Vertrauen in die Arbeit von Polizei- und Sicherheitsdiensten gestärkt.²⁵

2.1.3. Partizipation von Stakeholdern bei der Entwicklung und Anwendung automatisierter Systeme

Die Entwicklung und der Einsatz algorithmischer Entscheidungssysteme beinhalten von Beginn an Wertentscheidungen, die in das System-Design einfließen. Vorbilder für adäquate Verfahren der Stakeholder-Partizipation finden sich im Bereich der Bio- und Medizintechnik. Im Kontext von Genforschung wurden Lösungen für ethische Probleme auf verschiedenen gesellschaftlichen Ebenen gefunden. Die Einbindung von Patientenvertretungen in Organisationen, die mit der Entwicklung und Anwendung algorithmischer Entscheidungsfindung in der Medizin betraut sind, stellt eine Lösungsmöglichkeit dar, so Cohen et al..²⁶ Positive Erfahrungen liegen für Nutzung und Verwertung von Gendatenbanken vor, die ein Treuhänder (Individuum oder eine Gruppe von Betroffenen) betreibt.

²⁵ Selbst, 2016

²⁶ Cohen et al., 2014

2.1.4. Professionsethische Kodizes und Institutionen etablieren

Eine Reihe von Professionen haben professionsethische Prinzipien für Urteile über das Wohl von Menschen erarbeitet und Institutionen etabliert, die konkrete Fälle an diesen Prinzipien messen und beurteilen. In der Diskussion über die Bewertung gesellschaftlicher Angemessenheit algorithmischer Systeme empfehlen viele Experten Äquivalente zu Standards und Institutionen wie sie etwa bei medizinischer Forschung existieren.

Ob die Optimierungsziele gesellschaftlich angemessen sind, lässt sich nicht allein auf Grundlage der Gestaltung des Systems und der Datenauswahl diskutieren. Welche Folgen hat der Einsatz für den einzelnen Bewerteten? Welche Folgen hat der Einsatz absehbar auf kollektive Güter? Welche Alternativen bestehen? Diese Fragen können bei gesellschaftlich relevanten Systemen nicht allein die Data Scientists, Produktmanager, Implementierer und anderen an der Entwicklung Beteiligten beantworten. Hier braucht es viele Anknüpfungspunkte und Instrumente, um einen Diskurs in Gang zu bringen und zu ermöglichen. Dies ist ein Ansatz für eine Professionsethik: Wo algorithmische Systeme gesellschaftliche Teilhabe beeinflussen, müssen ihre Ziele, ihr Design und ihre Wirkung der gesellschaftlichen Kontrolle und Willensbildung unterliegen. Dies zu ermöglichen, ist die ethische Verantwortung der an der Entwicklung der Systeme Beteiligten im Sinne einer prozessbezogenen Professionsethik.

In diese Richtung gehen Vorschläge wie die »*Principles for Accountable Algorithms*« der Initiative FAT/ML.²⁷ Die Autoren postulieren fünf Prinzipien für die Gestaltung algorithmischer Systeme unter

²⁷ Fairness, Accountability, and Transparency in Machine Learning, 2016

dieser Maxime (im Folgenden aus dem englischen Original paraphrasiert):

- **Verantwortlichkeit:** Instanzen für Beschwerden und Berufung schaffen und öffentlich sichtbar machen.
- **Erklärbarkeit:** Sicherstellen, dass Entscheidungen Endanwendern und anderen Stakeholdern in nicht-technischer Sprache erklärt werden können.
- **Sorgfalt:** Quellen von Fehlern und Unsicherheit im System und den Datenquellen identifizieren, dokumentieren und protokollieren, sodass die Folgen verstanden und Linderungsmaßnahmen entwickelt werden können.
- **Überprüfbarkeit:** Interessierten Dritten das Testen, Verstehen und Evaluieren des Systems ermöglichen, z. B. durch geeignete APIs (Schnittstellen), Information, permissive Nutzungsbedingungen.
- **Fairness:** Sicherstellen, dass algorithmische Entscheidungen bei unterschiedlichen demographischen Faktoren (Geschlecht, Herkunft etc.) nicht systematisch ungerecht verschiedenen Output liefern.

2.2. Umsetzung von Zielen in Systemen prüfen, erklären, falsifizieren

Die Kenntnis der Ziele und erwarteten Resultate, der Methoden und der Anwendungsbereiche algorithmischer Entscheidungssysteme stellen die Grundvoraussetzung dar, um deren Angemessenheit zu gewährleisten. Doch damit ist natürlich weder gesichert, dass die intendierten Wirkungen erzielt werden, noch, dass nicht-intendierte Wirkungen ausgeschlossen werden können. Gerade im Bereich von selbst-lernenden Systemen gilt: Es können nicht nur Fehler oder

Bugs auftreten, sondern auch unerwartete Interaktionen von Systembestandteilen komplexer Systeme sowie den Einzelnen betreffende Fehler infolge zugrundeliegender Wahrscheinlichkeitsrechnung.

2.2.1. Etablierung zivilgesellschaftlicher Wächter: adäquate Ausstattung und Rechtssicherheit

Die Debatte über das COMPASS-System zur Risikoeinschätzung von Straftätern in den USA begann erst, nachdem das gemeinnützige US-Recherchebüro Propublica mit großem Aufwand Daten recherchiert, aufbereitet und ausgewertet hatte.²⁸ Es braucht nicht nur Ressourcen für die technische Analyse von Systemen, sondern auch für die Recherche potenzieller Fälle, das Sammeln und Einklagen von Daten und die Recherche tatsächlicher Einsatzpraktiken. Hierfür sind zivilgesellschaftliche Watchdogs nötig, wie das Beispiel Propublica in den Vereinigten Staaten zeigt. Inspiration könnten hier auch Nichtregierungsorganisationen aus der Umweltbewegung sein, die zum Teil als Beschwerdeinstanzen mit niedrigen Zugangshürden fungieren, Fälle aggregieren und Narrative entwickeln, die erst eine gesellschaftliche Debatte anstoßen. Forderungen nach Verantwortlichkeit gegenüber der Öffentlichkeit sind nicht zu verwirklichen, wenn nicht Mindestbedingungen für das Entstehen von Öffentlichkeit erfüllt sind. Nach allen Erfahrungen mit Fällen wie Compass, Centrelink und anderen Systemen gehören zu den Entstehungsbedingungen zivilgesellschaftliche Wächterorganisationen.

²⁸ Angwin, Kirchner, Larson & Mattu, 2016

2.2.2. Externe Analysen ermöglichen, Gesetzgebung anpassen

2015 erschien in Science eine einzigartige Studie zum Nutzungsverhalten von Facebook-Nutzern.²⁹ Die drei Autoren waren Facebook-Angestellte, die Datenbasis ihrer Forschung ist für niemanden außerhalb Facebook's zugänglich. Wissenschaftlicher wie Lazer warnen vor der Gefahr, dass »die einzigen, die Facebook erforschen können, Forscher bei Facebook sind«³⁰.

Die private Verfügungsgewalt über Algorithmen und Daten, die algorithmischen Entscheidungssystemen zugrunde liegen, können unter Umständen gesamtgesellschaftliche Schäden hervorrufen. Daher sind Wissenschaftsschranken für den Schutz solcher Systeme als Geschäftsgeheimnisse³¹ zu prüfen ebenso wie Transparenzvorschriften³².

2.2.3. Sicherung der Datenbasis: Datenauskunft und -korrektur, Herkunftsnachweise

Die Forderungen nach Datenauskunft und -korrektur schützen natürlich zunächst Individuen. Wie die Autoren des Future of Privacy Forum formulieren, kommt ein solches Recht auf Datenauskunft und -korrektur auch der Allgemeinheit zugute: Um negativen Effekten algorithmischer Entscheidungssysteme zu begegnen, brauche man sogenannte Data Methods Solutions - Methoden der Datenerhebung und Auswertung, welche die Aktualität, Korrektheit und Vollständigkeit von Daten absichern. Um ein Set adäquater Daten zu

²⁹ Bakshy, Messing und Adamic, 2015

³⁰ Lazer ,2015

³¹ Calo, 2017

³² Tutt, 2016

gewährleisten, das maschinellen Analyse- und Entscheidungsverfahren zugrunde liegt, bietet die Kontrolle von Individuen über sie betreffende Daten eine zuverlässige Option. Informationsrechte und Zustimmung von Individuen erhöhen aus dieser Perspektive die Zuverlässigkeit maschineller Entscheidungen³³

In der Literatur wird insbesondere über die Dokumentation der Herkunft von Daten etwa durch einen Lieferkettennachweis debattiert.³⁴

2.2.4. Widerspruchsverfahren institutionalisieren

Die Datenschutz-Grundverordnung sieht im Kontext algorithmischer Entscheidungsverfahren in Kap. 3, Art. 21-22 sowohl Widerspruchsrechte vor wie auch das Recht auf Erwirken des Eingreifens einer Person seitens des Verantwortlichen, auf Darlegung des eigenen Standpunktes und auf Anfechtung der Entscheidung.³⁵ Diese Rechte gelten auch unter Wissenschaftlern als zentral.³⁶ Ihre Umsetzung birgt Herausforderungen.

Generell gilt es, Widerspruchsrechte nicht nur für Individuen zu institutionalisieren. Auch die Möglichkeit der Einrichtung von Verbandsklagerechten sollte erörtert werden. Denn Wohlfahrtsverbände sind gesellschaftliche Institutionen, die einen direkten Überblick über eingesetzte Systeme maschineller Entscheidungen und ihre Implikationen haben, insbesondere in teilhaberelevanten Be-

³³ Future of Privacy Forum, 2017

³⁴ Association for Computing Machinery, 2017; Fairness, Accountability, and Transparency in Machine Learning, 2016

³⁵ Europäisches Parlament & Rat der Europäischen Union, 2016

³⁶ Citron, 2008

reichen. So werden sie auch von der Stadt New York zu Rate gezogen, um die von der Stadt in der Verwaltung eingesetzten Systeme zu klassifizieren und zu prüfen.³⁷

Als der Institutionalisierung von Widerspruchsverfahren dienlich gelten technische Designs, die Nachvollziehbarkeit ermöglichen³⁸ sowie Dokumentationspflichten für Betreiber algorithmischer Systeme, insbesondere zugrundeliegender Modelle, Algorithmen und Daten³⁹. Solche Dokumentationspflichten werden aktuell auch durch die Stadt New York geprüft, um die Angemessenheit maschinellen Verwaltungshandeln zu sichern.⁴⁰

2.3. Diversität schaffen

2.3.1. Frei zugängliche Trainingsdatensätze für Entwicklung und Training von Algorithmen

Für die enormen Fortschritte bei der Gesichts- und Bilderkennung durch algorithmische Systeme in den vergangenen Jahren ist neben neuer Hardware auch die breite Verfügbarkeit von Trainingsdaten verantwortlich. Im Jahr 2009 veröffentlichte ein Team um die Informatikerin Fei-Fei Li den Imagenet-Datensatz – eine Datenbank von damals 3,2 Millionen verschlagworteten Fotos. Inzwischen umfasst der Datensatz 13 Millionen Fotos, die Erkennungsrate der besten Software im jährlichen Imagenet-Wettbewerb ist von 71,8 Prozent im 2010 auf 97,3 Prozent 2017 gestiegen.⁴¹ Wertvolle Trainingsdaten sind heute bei weitem nicht per se frei zugänglich. Ein Beispiel sind

³⁷ Powles, 2017

³⁸ Kroll u. a., 2017

³⁹ Association for Computing Machinery, 2017; Shneiderman, 2016

⁴⁰ Powles, 2017

⁴¹ Gershgorn, 2017

Suchmaschinen und soziale Netzwerke – eine Anwendung algorithmischer Entscheidungsprozesse, mit denen die Mehrheit der deutschen Internetnutzer täglich konfrontiert ist. Die Strukturierung, Personalisierung und Bewertung von Inhalten erledigen bei sozialen Netzwerken und Suchmaschinen algorithmische Systeme, die als wesentliche Signale die Reaktionen der Nutzer auswerten.⁴² Diese Reaktionen kann kein anderer Anbieter auswerten, um eigene Empfehlungssysteme zu entwickeln. Die Konzentration der Nutzer bei wenigen Anbietern verbessert deren Datenlage und verschafft ihnen Vorteile gegenüber neuer Konkurrenz

Ein Vorschlag, um diesen Konzentrationstendenzen entgegenzuwirken: Öffentlich geförderte Forschung sollte produzierte Datensets der Allgemeinheit frei zugänglich machen. So ein Vorschlag des National Science and Technology Council in einem Bericht für den US-Präsidenten:

»Encouraging the sharing of AI datasets—especially for government-funded research—would likely stimulate innovative AI approaches and solutions. However, technologies are needed to ensure safe sharing of data, since data owners take on risk when sharing their data with the research community. Dataset development and sharing must also follow applicable laws and regulations, and be carried out in an ethical manner.«⁴³

2.3.2. Vielfältige Betreiber der Trainingsdatenallmende

Ein Geschäftsmodell für den Betrieb einer Trainingsdatenallmende zu entwickeln, ist eine Herausforderung. Gemeinwohl rechnet sich

⁴² Lischka & Stöcker, 2017, S. 15

⁴³ National Science and Technology Council, 2016, S. 31

nicht unbedingt nach unternehmerischer Buchführung. Deshalb braucht es auch andere Organisationsformen, eine Vielfalt von Betreibermodellen. Neben kommerziellen Akteuren kommen viele andere, gemeinwohlorientierte Organisationen in Betracht, zum Beispiel Hochschulen wie beim Imagenet-Datensatz. Cohen und andere schlagen als Inspiration Biobanken vor, die biologisches Material und zugordnete Daten etwa zur Krankengeschichte verwalten und der Forschung zugänglich machen. Solche Einrichtungen werden von Treuhändern im öffentlichen Auftrag betrieben wie die von Cohen beschriebene Michigan Department of Community Health Dried Blood Spot Specimen Bank.⁴⁴

2.3.3. Staat als aktiver Gestalter einer positiven Ordnung

Experten empfahlen der Obama-Regierung, dass die Exekutive ihre Macht bei der Beschaffung, Entwicklung und dem Einsatz von algorithmischen Systemen zum Gestalten einer positiven Ordnung nutzt. Drei Vorschläge:

- Offene Software fördern: *»To help support a continued high level of innovation in this area, the U.S. government can boost efforts in the development, support, and use of open AI technologies. Particularly beneficial would be open resources that use standardized or open formats and open standards for representing semantic information, including domain ontologies when available. Government may also encourage greater adoption of open AI resources by accelerating the use of open AI technologies within the government itself, and thus help to maintain a low barrier to entry for innovators. Whenever possible, government should contribute algorithms and software to open source projects.«⁴⁵*

⁴⁴ Cohen u. a., 2014, S. 1143

⁴⁵ National Science and Technology Council, 2016, S. 32

- Standards und Prozesse für den staatlichen Einsatz entwickeln:
»Agencies should work together to develop and share standards and best practices around the use of AI in government operations. Agencies should ensure that Federal employee training programs include relevant AI opportunities.«⁴⁶
- Kompetenz zur Entwicklung, Implementierung und Bewertung algorithmischer Systeme in neuen Agenturen konzentrieren:
»The Federal Government should explore ways to improve the capacity of key agencies to apply AI to their missions. For example, Federal agencies should explore the potential to create DARPA-like organizations to support high-risk, high-reward AI research and its application ...«⁴⁷

2.3.4. Standards durch staatliche Nachfrage setzen

Bei der staatlichen Beschaffung gehören Sozial- und Umweltstandards zum Teil zu den Vergabekriterien. Die Entwicklung und Einhaltung solcher Standards wird zum Teil staatlich gefördert. Zum Beispiel vom bundesweiten Netzwerk zur Fairen Beschaffung (mit Fokus auf Kommunen). Das Netzwerk wird vom Bundesentwicklungsministerium gefördert.

Die Idee, staatliche Nachfragemacht zur gemeinwohlorientierten Gestaltung zu nutzen, und die auf anderen Gebieten dabei etablierten Verfahren sollten Inspiration für Beschaffungsstandards für algorithmische Systeme sein.⁴⁸ Anforderungen an Erklärbarkeit, Angemessenheit und Diversität von Systemen könnten die Entwicklung entsprechender Praktiken und Werkzeuge fördern – zum Beispiel die Nutzung und Förderung offener Standards und Software,

⁴⁶ Felten u. a., 2016, S. 16

⁴⁷ Felten u. a., 2016, S. 16)(Felten u. a., 2016, S. 16

⁴⁸ Calo, 2017, S. 24

die Verfügbarkeit von Trainingsdaten, Monitoring, Erklärbarkeit, Entscheidungsforensik, Auszeichnung der Trainingsdaten und das Durchlaufen von Impact Assessments.

2.3.5. Förderung gemeinwohlorientierter Entwicklung

Forschungsförderung ist ein anderes Instrument, um durch staatliche Investitionen Technikentwicklung gemeinwohlförderlich zu gestalten. Calo sieht Handlungsbedarf bei Grundlagenforschung und Untersuchungen zur gesellschaftlichen Einbettung algorithmischer Systeme. Instrumente wie standardisierte Folgenabschätzungen oder professionsethische Kodizes müssen entwickelt und erprobt werden – das wären konkrete Ansatzpunkte für angewandte Forschung zur Einbettung von algorithmischen Systemen in gesellschaftliche Kontexte.

Die Förderung gemeinwohlorientierter algorithmischer Systeme sollte sich nicht allein auf etablierte institutionalisierte Forschung konzentrieren. Gemeinwohlorientierte Projekte unabhängiger Freiwilliger im Open-Source-Bereich oder aus zivilgesellschaftlichen Organisationen sollten gefördert werden. Als Inspiration kann der Prototype Fund dienen. Dieses von der Open Knowledge Foundation verantwortete und vom Bundesministerium für Bildung und Forschung finanzierte öffentliche Förderprogramm finanziert »gemeinnützige Software-Projekte in den Bereichen Civic Tech, Data Literacy und Datensicherheit«. Die geförderten Projekte sollen bis zur ersten Demoentwicklung gebracht werden – mit finanzieller Unterstützung und Coachingangeboten.⁴⁹

⁴⁹ Open Knowledge Foundation Deutschland, o. J.

2.4. Übergreifend: Kompetenzaufbau

2.4.1. Staatliche Gestaltungskompetenz stärken

Um der Komplexität algorithmischer Entscheidungen in der Gesellschaft gerecht zu werden, schlagen verschiedene Autoren vor, gezielt staatliche Kompetenzen aufzubauen bzw. entsprechende Kompetenzen bei existierenden staatlichen Institutionen zu stärken, und zwar in Abhängigkeit von unterschiedlichen Zielen.

Ein kontinuierliches Monitoring der technischen Entwicklung und Anwendung von ADM-Systemen sowie darauf aufbauende Empfehlungen stellen beispielsweise den Hintergrund für die Forderung nach der Einrichtung einer zentralen KI-Regulierungs- und/oder Beratungsinstitution dar.⁵⁰

Die meisten Autoren verorten eine solche zentrale Institution im Regelfall als Beratungsinstanz für Legislative, Exekutive und Judikative.⁵¹ Tutts Vorschläge gehen darüber hinaus: Er befürwortet die Einrichtung einer zentralen Behörde, die sowohl weiche Regulierung, wie etwa Transparenzvorschriften und Standardsetzung zum Gegenstand hat, als auch harte Regulierung: Zertifizierung und Zulassung von Systemen.

Auf der anderen Seite bedarf es insbesondere im Bereich hoheitlicher und öffentlicher Aufgaben auch des Ausbau von KI-Kompetenzen bei existierenden Institutionen, so Stone et al.⁵². Ziel ist das Verständnis für die Wechselwirkungen zwischen algorithmischer Entscheidungsfindung, Politikprogrammen und gesellschaftlichen Zielen.

⁵⁰ Tutt, 2016

⁵¹ Calo, 2017; Cave, 2017; Felten u. a., 2016; Geoff Mulgan, 2016

⁵² Stone et al., 2016

2.4.2. Individuelle Gestaltungskompetenz stärken

Stakeholder-Partizipation, Widerspruchsverfahren, Auskunftsrechte, zivilgesellschaftliche Wächter – solche Lösungsvorschläge eint, dass sie in der Umsetzung auf das Mitwirken von Betroffenen angewiesen sind. Von algorithmischen Entscheidungen betroffene Menschen müssen Auskunft verlangen, Informationen an Wächterorganisationen weiterleiten und Widerspruchsrechte in Anspruch nehmen. Das setzt ein Grundwissen darüber voraus, wo algorithmische Entscheidungen im Einsatz sind, welche Chancen und Risiken der Einsatz hat und wie man als (potenziell) Betroffener Einfluss auf Gestaltung und Einsatz nehmen kann.

Neben staatlicher Gestaltungskompetenz und professionsethischen Standards braucht es also auch Kompetenz in der Bevölkerung im Umgang mit algorithmischen Systemen. Die Royal Society definiert diese Kompetenz als *»a basic grounding in what machine learning is, and what it does, will be necessary in order to grasp, at a basic level, how our data is being used, and what this means for the information presented to us by machine learning systems.«*⁵³

Ein erster Schritt wäre es, den Kenntnisstand zu erfassen und dafür eine Methode zu entwickeln. Wie es um das Wissen über den Einsatz von algorithmischen Systemen und das Verständnis ihrer Funktionsweise in der Bevölkerung in Deutschland bestellt ist, wurde bislang nicht ermittelt. Zunächst wäre zu konkretisieren, wie sich in diesem Bereich Kompetenz äußert und wie sie zu erfassen wäre. In den Vereinigten Staaten wird diese Idee unter anderem unter der Bezeichnung *»algorithmic literacy«* diskutiert, allerdings fehlt bislang die Operationalisierung.⁵⁴

⁵³ Royal Society, 2017, S. 63

⁵⁴ Caplan, Reed, & Mateescu, 2016, S. 8

3. Fazit

Schon dieser unvollständige, knappe Überblick zeigt: Es gibt viele Ideen für Maßnahmen und Methoden zur gesellschaftlichen Gestaltung, Intervention bei und Kontrolle algorithmischer Entscheidungsfindung. Keineswegs scheint der Mensch der Maschine ausgeliefert. Allerdings gilt es nun, Chancen und Risiken im Einzelfall, d. h. unter Berücksichtigung des Anwendungsbereichs, der Komplexität und der Autonomie eines Systems, zu überprüfen und gegebenenfalls spezifische Handlungsoptionen zu entwickeln und zu erproben. Die Ideen existieren, es kommt nun auf Konkretisierung, Evaluation und Umsetzung an.

Quellen

- Angwin, J., Kirchner, L., Larson, J. & Mattu, S. (2016) »Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks«. *ProPublica*, 23. Mai 2016. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Association for Computing Machinery (2017). »Statement on Algorithmic Transparency and Accountability«, 12. Januar 2017. <http://s.fhg.de/eAU>
- Bakshy, E., Messing, S., & Adamic L. A. (2015). »Exposure to ideologically diverse news and opinion on Facebook«. *Science* 348, Nr. 6239 (2015), S. 1130–1132.
- Brennan Center for Justice (2017). »Brennan Center for Justice v. New York Police Department«, 19. Mai 2017. <http://s.fhg.de/t9k>
- Calo, R. (2017). »Artificial Intelligence Policy: A Roadmap«. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, 8. August 2017. <https://papers.ssrn.com/abstract=3015350>
- Caplan, R., Reed L. & Mateescu, A. (2016). »Who Controls the Public Sphere in an Era of Algorithms - Workshop Summary«. Data & Society Research Institute, 2016. <http://s.fhg.de/Eux>.

- Cave, S. (2017). »Written evidence - Leverhulme Centre for the Future of Intelligence«, 2017.
- Citron, D. K. (2008). »Technological due process«. *Wash. UL Rev.* 85 (2008), S. 1249. <http://s.fhg.de/M5H>
- Citron, D.K. & Pasquale, F. (2014). »The Scored Society: Due Process for Automated Predictions«. *Washington Law Review* 89, Nr. 1 (2014), S. 1–33
- Cohen, I. G., Amarasingham, R., Shah, A., Xie, B. & Lo, B. (2014). »The Legal And Ethical Concerns That Arise From Using Complex Predictive Analytics In Health Care«. *Health Affairs* 33, Nr. 7 (7. Januar 2014), S. 1139–1147.
- Council of Europe - Committee of experts on internet intermediaries (2017). »Study on the Human Rights Dimensions of Automated Data Processing Techniques (in Particular Algorithms) and Possible Regulatory Implications«. <http://s.fhg.de/vFW>
- Europäisches Parlament, und Rat der Europäischen Union (2016). Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung), Pub. L. No. Verordnung (EU) 2016/679 (2016). <http://s.fhg.de/DET>
- Fairness, Accountability, and Transparency in Machine Learning (2016). »Principles for Accountable Algorithms and a Social Impact Statement for Algorithms«, 2016. <http://s.fhg.de/DNC>
- Felten, E. (2016). »Preparing for the Future of Artificial Intelligence«, Oktober 2016. <http://s.fhg.de/cvB>
- Future of Life Institute (2017). »AI Principles«. Future of Life Institute, Januar 2017. <https://futureoflife.org/ai-principles/>
- Future of Privacy Forum (2017). »Unfairness by Algorithm: Distilling the Harms of Automated Decision-Making«. Gehalten auf der RightsCon, Brüssel, 30. März 2017
- Mulgan, G. (2016). »A machine intelligence commission for the UK: how to grow informed public trust and maximise the positive impact of smart machines«, 22. Februar 2016. <http://s.fhg.de/5ek>

- Gershgorn, D. (2017) »The data that transformed AI research—and possibly the world«. *Quartz*, 26. Juli 2017. <http://s.fhg.de/Eb5>
- Harris, E. A. (2016). »Court Vacates Long Island Teacher’s Evaluation Tied to Test Scores« *The New York Times*, 10. Mai 2016, Abschn. N.Y. / Region. <http://s.fhg.de/qAX>
- Heaton, B. (2015) »New York City Fights Fire with Data«. *Government Technology*, 15. Mai 2015. <http://s.fhg.de/rth>
- Zweig, K. A. (2016). »Medienethik: »Das Risiko liegt in der Intransparenz««. *Medienpolitik.net*, 11. April 2016. <http://s.fhg.de/gq9>
- Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G. & Yu, H. (2017) »Accountable Algorithms«. *University of Pennsylvania Law Review* 165, Nr. 3 (Februar 2017)
- Lazer, D. (2015). »The rise of the social algorithm«. *Science* 348, Nr. 6239 (6. Mai 2015).
- Lischka, K. & Klingel, A. (2017). »Wenn Maschinen Menschen bewerten« Bertelsmann Stiftung, 2017. <http://s.fhg.de/6Kq>
- Lischka, K. & Stöcker, C. (2017). »Digitale Öffentlichkeit«. *BStift - Bertelsmann Stiftung*, 2017.
- National Science and Technology Council (2016). »The National Artificial Intelligence Research and Development Strategic Plan«, Oktober 2016.
- New York City Independent Budget Office (2016). »A Look at New York City’s Public High School Choice Process« New York City Independent Budget Office, Oktober 2016. <http://s.fhg.de/ViL>
- O’Neil, C. (2017). »Don’t Grade Teachers With a Bad Algorithm«. *Bloomberg.com*, 15. Mai 2017. <http://s.fhg.de/ZLu>
- Open Knowledge Foundation Deutschland (2017). »Prototype Fund«. <https://okfn.de/projekte/prototypefund/>, abgerufen am 03.11.2017
- Otto, P. (2017). »Leben im Datenraum - Handlungsauftrag für eine gesellschaftlich sinnvolle Nutzung von Big Data«. In: *Perspektiven der digitalen Lebenswelt*, S. 9–36. *Verwaltungsressourcen und Verwaltungsstrukturen* 32. Nomos Verlag, Baden-Baden, 2017. <http://s.fhg.de/BXf>

Powles, J. (2017). »New York City’s Bold, Flawed Attempt to Make Algorithms Accountable«. *The New Yorker*, 21. Dezember 2017. <http://s.fhg.de/IVy>

Royal Society (2017). Machine learning: the power and promise of computers that learn by example. <http://s.fhg.de/68j>

Selbst, A. D. (2016). »Disparate Impact in Big Data Policing«. SSRN Scholarly Paper. Social Science Research Network, Rochester, NY, 5. August 2016. <http://papers.ssrn.com/abstract=2819182>.

Shneiderman, B. (2016). »Opinion: The Dangers of Faulty, Biased, or Malicious Algorithms Requires Independent Oversight«. *Proceedings of the National Academy of Sciences* 113, Nr. 48 (29. November 2016), S. 13538–13540.

Singer, N. (2015). »Bringing Big Data to the Fight Against Benefits Fraud«. *The New York Times*, 20. Februar 2015. <http://s.fhg.de/9vc>

Tullis, T. (2014). »How Game Theory Helped Improve New York City’s High School Application Process«. *The New York Times*, 5. Dezember 2014. <http://s.fhg.de/X8f>

Tutt, A. (2016). »An FDA for Algorithms«. SSRN Scholarly Paper. Social Science Research Network, Rochester, NY, 15. März 2016. <http://s.fhg.de/dpp>

Wachter, S., Mittelstadt, B. & Russell, C. (2017). »Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR«. SSRN Scholarly Paper. Social Science Research Network, Rochester, NY, 6. Oktober 2017. <https://papers.ssrn.com/abstract=3063289>

Web Foundation (2017). »Algorithmic Accountability: Applying the concept to different country contexts«, 28. Juli 2017. <http://s.fhg.de/6ey>

Anmerkung

Dieser Text basiert auf dem Arbeitspapier von Julia Krüger und Konrad Lischka: »Damit Maschinen den Menschen dienen«, Bertelsmann Stiftung, 2018.

Über die Autoren

Julia Krüger

Julia Krüger (Diplom Politikwissenschaft, Universität Potsdam) arbeitet als unabhängige Wissenschaftlerin an der Schnittstelle von Politik, Recht und Technik. Zuvor beschäftigte sie sich am Wissenschaftszentrum Berlin für Sozialforschung (WZB) und dem Think Tank Selbstregulierung Informationswirtschaft (SRIW) mit Internetpolitik, Datenschutz und Digitalisierung. Sie ist Autorin von netzpolitik.org und Fellow am Center for Internet & Human Rights (Europa-Universität Viadrina, Frankfurt an der Oder).

Konrad Lischka

Konrad Lischka ist Co-Leiter des »Projekts Ethik der Algorithmen« der Bertelsmann Stiftung. Zuvor stellvertretender Ressortleiter Netzwelt bei Spiegel Online, danach Wechsel in die Medien- und Netzpolitik als Referent Digitale Gesellschaft in der Staatskanzlei Nordrhein-Westfalen.