



Teilhabe, ausgerechnet

Wie algorithmische Prozesse
Teilhabechancen beeinflussen können

Teilhabe, ausgerechnet

Wie algorithmische Prozesse Teilhabechancen beeinflussen können

Kilian Vieth
Dr. Ben Wagner
im Auftrag der Bertelsmann Stiftung

Impressum

© Juni 2017 Bertelsmann Stiftung
Bertelsmann Stiftung
Carl-Bertelsmann-Straße 256
33311 Gütersloh
www.bertelsmann-stiftung.de

Verantwortlich

Konrad Lischka
Ralph Müller-Eiselt

Autoren

Kilian Vieth
Dr. Ben Wagner

Lizenz

Dieses Arbeitspapier ist unter der Creative-Commons-Lizenz [CC BY-SA 3.0 DE](https://creativecommons.org/licenses/by-sa/3.0/de/) (Namensnennung – Weitergabe unter gleichen Bedingungen) lizenziert. Sie dürfen das Material vervielfältigen und weiterverbreiten, solange Sie angemessene Urheber- und Rechteangaben machen. Sie müssen angeben, ob Änderungen vorgenommen wurden. Wenn Sie das Material verändern, dürfen Sie Ihre Beiträge nur unter derselben Lizenz wie das Original verbreiten.

Titelbild: Konrad Lischka, [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)

DOI 10.11586/2017027 <https://doi.org/10.11586/2017027>

Inhalt

1	Vorwort	7
2	Einleitung	8
3	Begriffliche Grundlagen.....	9
3.1	Teilhabe	9
3.2	Algorithmus.....	9
3.3	Algorithmische Entscheidungsfindung	10
3.4	Data Mining.....	10
3.5	Lernender Algorithmus	10
3.6	Modellierung	11
3.7	Soziale Einbettung von Algorithmen	11
3.8	Erklärbarkeit algorithmischer Entscheidungsfindung	12
4	Analysebereich: Computersysteme, die Menschen bewerten.....	13
4.1	Digitale Entscheidungssysteme.....	13
4.2	Bewertung von Menschen	13
5	Anwendungsszenarien algorithmischer Entscheidungsfindung	15
5.1	Arbeitswelt	15
5.2	Sicherheit.....	16
5.3	Meinungsbildung	17
5.4	Öffentliche Daseinsvorsorge	18
5.5	Marketing	19
5.6	Zwischenfazit: Welche Funktionen erfüllen Algorithmen?.....	20
6	Die potenzielle Wirkung algorithmischer Entscheidungsfindung auf Teilhabe berechnen	22
6.1	Methodischer Ansatz	22
6.2	Akteure	23

6.3	Soziale Einbettung.....	24
6.4	Konsequenzen.....	24
6.5	Zusammenfassung: So lässt sich Teilhabewirkungspotenzial bestimmen	26
6.6	Beispielrechnungen	27
7	Fazit und Ausblick.....	29
8	Literatur	30
9	Executive Summary.....	37
10	Über die Autoren.....	41

1 Vorwort

Was unterscheidet diese beiden Programme? Eine Software berechnet die Kosten einer KFZ-Haftpflichtversicherung basierend auf dem Fahrverhalten der Kunden. Eine andere Software sagt die Wahrscheinlichkeit von Wohnungseinbrüchen nach Straßenzügen vorher und hilft der Polizei bei der Streifenplanung.

Die intuitive Antwort könnte zum Beispiel lauten: Es erscheint mir bedrohlicher, das Fahrverhalten jedes Versicherten zu überwachen und mit einem entsprechenden Tarif zu sanktionieren, als Kriminalitätsprognosen für Straßenzüge abzugeben. Oder: Ich ziehe menschliche Entscheidungen vor und lehne beides ab. Egal wie die spontane Reaktion ausfällt: Menschen und ihr Verhalten werden heute in vielen Lebensbereichen von maschinellen Systemen, sogenanntem „algorithmic decision making“ (ADM), beobachtet und bewertet. Solche ADM-Prozesse haben Einfluss vor Gericht, bei der Vergabe von Krediten und Studienplätzen oder bei der Aufmerksamkeit, die Anrufer in einer Kundendienst-Hotline erfahren. Maschinelle Entscheidungssysteme dieser Art sind in anderen Ländern teilweise schon deutlich verbreiteter. Das große Einsatzspektrum zeigt aber, wie wichtig auch hierzulande eine informierte und vor allem sinnvoll strukturierte Debatte zu diesem Thema ist. Genau dazu soll dieses Arbeitspapier einen Beitrag leisten.

Ein entscheidendes Merkmal zur Differenzierung von ADM-Prozessen ist ihre mögliche gesellschaftliche Wirkung. Konkreter: Ihr Effekt auf Teilhabechancen. Welchen Einfluss haben ADM-Prozesse auf die gleichberechtigte Einbeziehung von Individuen und Organisationen in politische Entscheidungs- und Willensbildung sowie auf die faire Partizipation aller an sozialer, kultureller und wirtschaftlicher Entwicklung?

Die Autoren der vorliegenden Expertise, Kilian Vieth und Ben Wagner, stellen ein Instrument vor, mit dem sich dieses Wirkungspotenzial anhand sieben voneinander unabhängiger Merkmale bestimmen und auch tentativ quantifizieren lässt. Die Teilhaberelevanz kann durch die Beantwortung eines knappen Fragenkatalogs ermittelt werden, der anders als intuitive Einschätzungen konsistent auf unterschiedliche Fälle anwendbar ist. Es geht dabei nicht um die Güte der Verfahren an sich, sondern um eine relative Beurteilung ihrer potenziellen Wirkung auf Teilhabe, gleich ob positiv oder negativ. Je höher diese ausfällt, desto größer müssen die Anforderungen an einen ADM-Prozess sein – etwa beim Polizeieinsatz größer als bei der KFZ-Versicherung ([siehe Kapitel 6.6](#)).

Das dargestellte Instrument ist ein erster Vorschlag zur systematischen Auseinandersetzung mit Prozessen algorithmischer Entscheidungsfindung. Wir veröffentlichen es als Arbeitspapier unter einer freien Lizenz (CC BY-SA 3.0 DE), um einen Beitrag zu einem sich schnell entwickelnden Feld zu leisten, den auch andere aufgreifen können. Wir freuen uns über Erweiterungen, Anpassungen, die Anwendung der vorgeschlagenen Methodik auf konkrete Fallbeispiele ebenso wie über Anregungen und Kritik jeder Art.

Die Analyse von Kilian Vieth und Ben Wagner ist Teil einer Exploration zum Thema „Teilhabe in Zeiten von Algorithmen und Big Data“, in der sich die Bertelsmann Stiftung näher damit beschäftigt, wie sich Phänomene der digitalen Sphäre auf gesellschaftliche Teilhabe auswirken. Neben der Analyse internationaler Fallbeispiele von ADM-Prozessen ([Lischka und Klingel 2017](#)) werden weitere Arbeitspapiere zum Beispiel Fehlerquellen in Prozessen algorithmischer Entscheidungsfindung analysieren und sowie Wechselwirkungen zwischen algorithmischen Systemen und psychologischen Mechanismen in der digitalen Öffentlichkeit in der digitalen Sphäre untersuchen.



Ralph Müller-Eiselt
Senior Expert
Taskforce Digitalisierung
Bertelsmann Stiftung



Konrad Lischka
Project Manager
Taskforce Digitalisierung
Bertelsmann Stiftung

2 Einleitung

Was haben Bonitätsprüfungen, Grenzkontrollen und Suchmaschinen gemeinsam? Alle nutzen algorithmische Verfahren, um Entscheidungen vorzubereiten bzw. zu treffen. Das durch Algorithmen beeinflusste Urteil darüber, wer einen Kredit bekommt, wer einreisen darf und welche Treffer eine Suchanfrage liefert, ist zu einem zentralen Merkmal der digitalisierten Gesellschaft geworden.

Unser gesamter Alltag wird von und mit digitaler Technologie gestaltet und der Kernbestandteil all der Computer, die uns umgeben, sind die Algorithmen, die bestimmen, wie wir kommunizieren, arbeiten und uns bewegen. Die Digitalisierung generiert immer größere Datenmengen; neue Geschäftsmodelle und Plattformen entstehen, die alle in wesentlichen Bereichen durch Algorithmen gesteuert werden. Ohne Algorithmen lässt sich die digitale Gesellschaft nicht denken, wir brauchen sie, um unsere datenreiche Umgebung zu ordnen. „Wenn wir das gute Leben nicht auf einige wenige beschränken wollen, dann brauchen wir bessere Energieversorgung, bessere Mobilitätskonzepte und Ressourcenmanagement“ (Stalder 2017). Auch wenn sich die Entwicklung zahlreicher „smarter“ Produkte in manchen Fällen noch als hohle Versprechung entpuppt, ist doch eines klar: Ohne Algorithmen als Bestandteil sozialer Prozesse wird es nicht gehen, wenn wir drängende Probleme lösen und eine gerechte Teilhabe an der Gesellschaft ermöglichen wollen. Es geht um all die Prozesse, bei denen große Datensätze mithilfe von Software be- und verarbeitet werden, um Entscheidungen zu treffen oder Entscheidungsgrundlagen zu erstellen.

Dass Algorithmen immer öfter Entscheidungen mit uns, für uns und über uns fällen, wirft neue Fragen im Hinblick auf Teilhabe auf. Diese Expertise macht daher erste Vorschläge, wie die Teilhaberelevanz von Verfahren algorithmischer Entscheidungsfindung strukturiert und klassifiziert werden kann. Entscheidend ist dabei, sich von einem verbreiteten Technik-Determinismus zu lösen und die Gestaltbarkeit digitaler Technologie in den Mittelpunkt zu stellen. Technik ist formbar, deshalb gilt es im ersten Schritt zu klären, welche Ziele mit dem Einsatz algorithmischer Verfahren verfolgt werden. Nicht aus vermeintlichen technischen Zwängen heraus, sondern aus der Diskussion über wünschenswerte Ergebnisse müssen Rahmenbedingungen für Technologie entwickelt werden. In der digitalisierten Wissensgesellschaft wird die Gestaltung von Technologie zur elementaren Machtfrage. Denn wenn Wissen Macht ist, dann sind Algorithmen die Machtinstrumente unserer Zeit. Wie viel Beeinflussung durch Algorithmen ist für jeden Einzelnen und für die Gesellschaft als ganze verträglich und förderlich? Und wo müssen wir genauer hinsehen, um die Potenziale zu nutzen und die Risiken bestmöglich zu minimieren?

Diese Expertise bietet einen Überblick über das Gesamtphänomen von Prozessen algorithmischer Entscheidungsfindung, indem sie zentrale Begriffe erläutert (Kapitel 3), den relevanten Analysebereich solcher Prozesse abgrenzt (Kapitel 4) und typische Szenarien und Funktionen für den Einsatz von algorithmenbasierten Entscheidungsverfahren aufzeigt (Kapitel 5). Auf dieser Basis wird ein Vorschlag zur Klassifizierung von Prozessen algorithmischer Entscheidungsfindung entwickelt (Kapitel 6). Dieser soll helfen, die potenzielle Wirkung von algorithmischen Verfahren auf Teilhabe anhand weniger Kriterien zu bewerten und vergleichbar zu machen. Damit können zum Beispiel tiefer gehende Untersuchungen priorisiert und vorbereitet werden.

3 Begriffliche Grundlagen

3.1 Teilhabe

Im Rahmen dieser Studie umfasst der Begriff Teilhabe die gleichberechtigte Einbeziehung von Individuen und Organisationen in politische Entscheidungs- und Willensbildung sowie die faire Partizipation aller an sozialer, kultureller und wirtschaftlicher Entwicklung. Es geht also erstens um Teilhabe an demokratischen Prozessen – und damit um politische Gleichberechtigung – und zweitens um Teilhabe an Errungenschaften eines sozialen Gemeinwesens, „angefangen von guten Lebens- und Wohnverhältnissen, Sozial- und Gesundheitsschutz, ausreichenden und allgemein zugänglichen Bildungschancen und der Integration in den Arbeitsmarkt bis hin zu vielfältigen Freizeit- und Selbstverwirklichungsmöglichkeiten“ (Beirat Integration 2013).

Teilhabe in diesem Sinne hat zur Voraussetzung, dass die materiellen Mittel aller über einem Mindestniveau liegen, das allen Menschen die Mitwirkung am gesellschaftlichen Leben möglich macht. Die Gewährleistung von sozialer und politischer Teilhabe setzt also eine „Sockelgleichheit der sozialen Grundgüter“ (Meyer 2016) voraus. Elemente dieses Sockels werden zum Beispiel in der Allgemeinen Erklärung der Menschenrechte und im Internationalen Pakt über wirtschaftliche, soziale und kulturelle Rechte beschrieben (Bundesgesetzblatt 1966). Um chancengerechte Teilhabe in diesem Sinne zu ermöglichen, sind gezielte Investitionen in die Entwicklung individueller Fähigkeiten nötig (Bertelsmann Stiftung 2011: 31). Es liegt in der Verantwortung des Staates und des Gemeinwesens, kontinuierlich jedes Individuum zu befähigen, seine Chancen tatsächlich zu nutzen.

3.2 Algorithmus

Algorithmen sind mittlerweile Gegenstand diverser Debatten, dem Begriff werden jedoch je nach Kontext und Bezugsgruppe unterschiedliche Bedeutungen zugeschrieben. Die Informatik und Mathematik verwenden den Begriff anders, als es in der Öffentlichkeit oder im politischen Diskurs üblich ist. Um Debatten über „Algorithmen“ besser einordnen zu können, muss stets berücksichtigt werden, welche Senderin¹ sich an welches Publikum richtet. Insbesondere wenn es um die Analyse der gesellschaftlichen Auswirkungen von Algorithmen geht, muss zwischen der formellen Definition und einer allgemeinen Verwendung des Wortes unterschieden werden. In vielen Debatten über Algorithmen geht es nicht um einen eng definierten Algorithmus an sich, sondern vielmehr um die Rolle von Technologie in der Gesellschaft (Bucher 2016).

Der Begriff Algorithmus geht auf den persisch-arabischen Mathematiker Muhammad ibn Musa Al Chwarizmi zurück. Algorithmen sind nicht nur Software, sondern werden allgemein als programmierte Verfahren betrachtet, die einen bestimmten Input (einen Wert oder eine Gruppe an Werten) in einen gewünschten Output umwandeln (für eine klassische Lehrbuchdefinition siehe Cormen et al. 2001). Diese Verfahren verwenden bestimmte Berechnungen, um eine spezifische Aufgabe in genau definierten Schritten zu lösen (Gillespie 2014: 167). Algorithmen bestehen also aus einer genau definierten Folge von Schritten, um ein bestimmtes Ergebnis zu erzielen (Diakopoulos 2015: 400). Ein simple Aufgabe ist zum Beispiel: „Ordne die Zahlen dieser Zahlenfolge aufsteigend der Größe nach.“ Zur Lösung des Problems kommt dann eine Reihe von Sortieralgorithmen infrage. Ein Lösungsweg ist, der Reihe nach zwei benachbarte Zahlen zu vergleichen und sie zu vertauschen, falls sie in der falschen Reihenfolge liegen. Dieser Vorgang wird dann wiederholt, bis kein Zahlentausch mehr nötig ist. Dafür sind meistens mehrere Durchläufe notwendig.

¹ Die weibliche Form wird lediglich aus Gründen der Lesbarkeit gewählt. Sie ist als generisches Femininum zu verstehen.

Ein anschaulicher Vergleich für Algorithmen sind Backrezepte, denn auch sie bestehen aus einer Reihe von Schritten, die ausgeführt werden müssen, um ein konkretes Ziel zu erreichen. Es werden Daten eingegeben (die Zutaten), die dann Schritt für Schritt verarbeitet werden und schließlich ein vorab definiertes Ergebnis produzieren (z. B. einen Kuchen). Anders als beim Backen müssen für Computer allerdings alle Teilschritte in einer Art und Weise ausgedrückt werden, die für einen Computer direkt interpretierbar sind. Er kann nicht „eine Prise Salz“ hinzugeben, sondern nur eine genaue Menge, zum Beispiel „2 Gramm Salz“ (vgl. Zweig 2016).

3.3 Algorithmische Entscheidungsfindung

Algorithmische Entscheidungsfindung bezeichnet den Gesamtprozess von der Datenerfassung über die Datenanalyse bis hin zur Deutung und Interpretation der Ergebnisse und der Ableitung einer Entscheidung oder einer Entscheidungsempfehlung aus den Ergebnissen. Der Entscheidungsprozess ist festgelegt und kann beliebig viele Datenpunkte verarbeiten. Der Algorithmus fungiert dabei als (teil-)automatisiertes Hilfsmittel oder als ausschließlicher Entscheider. Algorithmische Entscheidungsfindung umfasst also mehrere Schritte und Unterprozesse, angefangen von der Entwicklung der Algorithmen selbst, über die Modellierung bis zur Interpretation des Outputs.

3.4 Data Mining

Data Mining bzw. Datenanalyse bedeutet, (große) Datenmengen nach Mustern zu durchsuchen. Die Art der Muster variiert, zum Beispiel kann nach Ereignissen gesucht werden, die oft zusammen auftreten. Oder es werden Daten automatisiert anhand konkreter Merkmale in vorab bestimmte Klassen (z. B. Gruppen von Menschen) einsortiert.

Die Mustererkennung beruht dabei meistens auf reiner Korrelation zwischen Datenpunkten und macht keine Aussage über Kausalzusammenhänge. Beim Einsatz von Mustererkennung ist also stets darauf zu achten, den entdeckten Zusammenhang (Korrelation) nicht mit einer Ursache-Wirkung-Beziehung zu verwechseln. Das menschliche Gehirn neigt jedoch stark dazu, Korrelationen als kausale Aussagen und sinnstiftende Geschichten zu interpretieren (Kahneman 2012).

Ein berühmtes Beispiel für erfolgreiches Data Mining in großen Datensätzen stammt von der US-amerikanischen Supermarktkette Target. Dort wurde herausgefunden, dass Schwangere oft bestimmte Produkte kaufen (Hill 2012). In ihrer Häufung sind die Einkäufe von Schwangeren so ungewöhnlich, dass die Firma sie dadurch zu einem bestimmten prozentualen Anteil identifizieren kann. Das entdeckte Muster ist hier also ein bestimmtes Konsumverhalten, das mit einer Schwangerschaft der einkaufenden Person korreliert.

3.5 Lernender Algorithmus

Als „maschinelles Lernen“ oder „lernende Algorithmen“ werden Verfahren bezeichnet, bei denen der Algorithmus selbst aus der Beobachtung vieler Fälle Regeln ableitet. Die Berechnung eines Ergebnisses besteht dabei aus zwei Phasen: Zuerst versucht der Algorithmus, Muster in Daten zu finden. Dazu wird er anhand von bekannten Datensätzen daraufhin trainiert, ein gewünschtes Ergebnis zu erkennen. Zum Beispiel kann ein Algorithmus aus der Beobachtung großer Mengen an Bildern „lernen“, auf diesen Bildern Gesichter zu erkennen. In den Trainingsdaten wird dafür angegeben, auf welchen Bildern Gesichter abgebildet sind und auf welchen nicht. Durch den wiederholten Vergleich zwischen Bildern mit und Bildern ohne Gesicht „lernt“ der Algorithmus Regeln, woran er ein Gesicht erkennen kann. Er sucht also nach Mustern in der Struktur des Bildes (z. B. Kanten und Schattierungen), die auf ein Gesicht hindeuten.

In der zweiten Phase werden diese Regeln dann auf neue, unbekannte Bilder angewendet. Auf Grundlage der erlernten Regeln kann der Algorithmus nun mit hoher Genauigkeit entscheiden, welche Bilder Gesichter abbilden und welche nicht. Die Eigenheit lernender Algorithmen ist also, dass die Entscheidungsregeln nicht direkt programmiert, sondern aus den Daten abgeleitet werden.

3.6 Modellierung

Wenn Algorithmen zur Beantwortung einer Frage eingesetzt werden, verarbeiten sie nie ein direktes Abbild der Realität, sondern ein eigens für dieses Verfahren entworfenes Modell der Realität. Bei der Entwicklung und Implementierung des Algorithmus gibt es immer einen Modellierungsfreiraum, je nachdem, in welche Teilschritte die Fragestellung zerlegt wird. Oft müssen Konzepte erst messbar gemacht werden, zum Beispiel Freundschaft oder die Relevanz einer Nachricht (Zweig 2017, im Erscheinen: 15). Dabei ist je nach Zielsetzung zu entscheiden, welche Merkmale für die Lösung relevant sind und welche vernachlässigt werden können. Jede Aufgabenstellung kann auf etliche verschiedene Arten in einem Algorithmus abgebildet werden. Wenn beispielsweise ein Navigationssystem eine Route von A nach B berechnen soll, muss zuvor entschieden werden, welches Verkehrsmittel genutzt und ob der kürzeste Weg (in Kilometern), der schnellste Weg (in Minuten) oder der günstigste Weg (z. B. ohne Mautstraßen) gewählt werden soll. Die Aufgabe „Zeige mir den Weg von A nach B an“ kann also auf unterschiedliche Art interpretiert und damit modelliert werden. Ein Modell ist immer eine Verkürzung und damit ein beschränktes oder abstrahiertes Abbild der Wirklichkeit. Die wichtigste Einflussgröße auf Teilhabe ist deshalb meist nicht die reine mathematische Konstruktion eines Algorithmus. Wichtiger sind hier in der Regel die Operationalisierung der Konzepte und die zu beantwortenden Fragen, die Implementierung eines passenden Algorithmus sowie die Erhebung geeigneter Daten.

3.7 Soziale Einbettung von Algorithmen

Oft werden die durch Algorithmen generierten Entscheidungen und Vorhersagen als neutral oder unabhängig angesehen. Doch tatsächlich sind Prozesse algorithmischer Entscheidungsfindung an vielen Punkten von Menschen beeinflusst: Menschen setzen die Ziele, entwerfen und implementieren die Prozesse, Menschen interpretieren die Ergebnisse und bestimmen entweder im Einzelfall über die Konsequenzen oder im Allgemeinen über die Bandbreite möglicher Konsequenzen. Nur weil ein algorithmisches Entscheidungssystem zum Einsatz kommt, ist der Gesamtprozess nicht weniger anfällig für Fehler und Befangenheit.

Prozesse algorithmischer Entscheidungsfindung werden oft eingesetzt, um menschliche Entscheidungsfindung vorzubereiten und zu unterstützen, oder auch, um sie ganz zu ersetzen. Dabei gehen zentrale Bestandteile menschlicher Entscheidungsprozesse verloren bzw. sie werden verlagert oder eingeschränkt. Zum Beispiel kann ein menschlicher Ermessensspielraum für Einzelfallentscheidungen nicht durch Algorithmen automatisiert werden (Spiekermann 2015).

Ohne die Interaktion mit Menschen bleiben viele der hier beschriebenen Algorithmen bedeutungslos. Deshalb darf Technologie nicht als neutrales Artefakt angesehen werden, denn ihre Ausgestaltung kann weitreichende politische Folgen haben (Denardis 2012). Um sich nicht von der Annahme der Objektivität in die Irre führen zu lassen, sind Algorithmen und ihre technologische Einbettung als soziale Konstrukte zu verstehen (Winner 1980). Algorithmen werden von Menschen erdacht und gestaltet und prägen in ihrer Anwendung wiederum menschliches Verhalten. Um die Auswirkungen von Algorithmen auf die Gesellschaft zu verstehen, gilt es, ihre soziale Konstruktion im Zusammenspiel zwischen Mensch und Technik zu betrachten (Brey 2005). Algorithmen sind also „technische Objekte, [die] ein immer höheres Niveau und einen höheren Grad an Handlungsträgerschaft erreichen“ (Rammert 2006: 28) und damit auch zu einem wichtigeren Teil der gesellschaftlichen Sozialstruktur werden.

Algorithmische Entscheidungsfindung basiert immer auf bestimmten Werten und Normen. Daher kann nie nur der Algorithmus „an sich“ untersucht werden, sondern es muss immer auch die Einbettung in einen sozialen Kontext berücksichtigt werden. Die Auswirkung von Algorithmen auf Teilhabe wird in vielen Fällen erst sichtbar, wenn der Algorithmus als soziales Konstrukt wahrgenommen wird. Auch wenn wir bewerten wollen, ob algorithmische Entscheidungsprozesse in konkreten Fällen „besser“ oder „effizienter“ funktionieren als menschliche Entscheidungsprozesse, kommt es auf die Interaktion zwischen Mensch und Technologie an. Die wachsende Bedeutung von Algorithmen erfordert also auch ein fundiertes Verständnis von sozialen Entscheidungsprozessen insgesamt.

3.8 Erklärbarkeit algorithmischer Entscheidungsfindung

Menschliche und algorithmische Entscheidungsprozesse sind grundsätzlich verschieden und verursachen unterschiedliche Arten von Fehlern. Das macht es nicht einfach, die Qualität und Güte beider Prozesse zu bewerten und miteinander zu vergleichen. Erschwert wird die Erklärbarkeit algorithmischer Entscheidungsfindung zudem durch drei Phänomene:

- Menschliche Entscheidungen bestimmen an vielen Stellen (z. B. bei der Messbarmachung) die Gestaltung der algorithmischen Systeme (vgl. Kapitel 3.7).
- Prozesse algorithmischer Entscheidungsfindung werten oft Daten aus, die aus menschlichem Verhalten entstehen (z. B. soziale Netzwerke das Nutzungsverhalten als Signal für die Relevanz von Beiträgen).
- Die Berechnungen von vielen typischen algorithmischen Systemen sind mittlerweile so komplex geworden, dass sie von Menschen nicht mehr nachvollzogen werden können. Vor allem bei lernenden Algorithmen (siehe Kapitel 3.5) bleiben die Muster und Logiken hinter den Entscheidungsprozessen verborgen. Bei derart komplexen Systemen können nicht nur Änderungen am Algorithmus, sondern auch an anderen Variablen große Wirkung auf das Verhalten des gesamten Systems haben. Selbst wenn formal ein Mensch die endgültige Entscheidung trifft, zum Beispiel über die Einladung zum Vorstellungsgespräch, nachdem ein Algorithmus die Bewerbungen vorsortiert hat, lässt das System nur einen begrenzten Spielraum offen. Es ist eher unwahrscheinlich, dass ein Mensch die Vorentscheidungen eines Algorithmus revidiert oder nur in Teilen übernimmt (Hannah-Moffat, Maurutto und Turnbull 2009).

Auch aus Mangel an Transparenz, Zeit oder Fachkenntnis ist es für die Nutzerinnen algorithmischer Entscheidungssysteme oft gar nicht möglich, Einzelfälle zu überprüfen. Zwischen einem teilautomatisierten und einem vollautomatisierten Prozess lässt sich in vielen Fällen nicht klar unterscheiden. Wenn ein Mensch die Entscheidung nicht mehr selbst mit Argumenten begründen kann, sondern sich auf den Output eines Computers verlässt, verschwimmt die Grenze zwischen menschlicher und algorithmischer Entscheidungsfindung.

4 Analysebereich: Computersysteme, die Menschen bewerten

Es gibt viele Einsatzbereiche für algorithmische Entscheidungsfindung. Für Einzelfälle lässt sich theoretisch fast immer ein teilhaberelevantes Szenario konstruieren. Doch wann wird Teilhabe grundsätzlich durch einen algorithmischen Entscheidungsprozess beeinflusst? Dazu werden hier vorläufige, pragmatische Kriterien für die Analyse der Teilhaberelevanz vorgeschlagen. Die Klassifizierung liefert kein absolutes Maß für Teilhaberelevanz, sondern eine Skala. Sie dient als ein erstes Instrument zum Vergleich der relativen Teilhaberelevanz von unterschiedlichen algorithmenbasierten Prozessen und Anwendungen. Der Analysebereich muss dafür auf die relevanten Phänomene mit vorhandenen oder absehbaren Auswirkungen auf Teilhabe eingegrenzt werden. Diese Teilmenge an generell teilhaberelevanten Anwendungsfällen kann dann einer eingehenden, qualitativen Analyse unterzogen werden. Algorithmische Entscheidungsprozesse sind für diese Expertise grundsätzlich relevant, wenn ein digitales Entscheidungssystem (Kapitel 1) zum Einsatz kommt, und wenn Menschen oder ihnen zugeschriebene Eigenschaften (Kapitel 2) dabei bewertet werden.

4.1 Digitale Entscheidungssysteme

Der erste Eingrenzungsschritt erfolgt nach technischen Gesichtspunkten: Die Expertise betrachtet nur digitale Entscheidungssysteme. Es gibt auch Algorithmen, die händisch angewendet werden. Da Algorithmen zunächst nur eindeutige Beschreibungen einer endlichen Serie von Aktionen sind, sind sie nicht zwangsläufig computergesteuert. Ein sehr aufwendiger Fragebogen, ein Hygieneleitfaden für Beschäftigte oder ein analoges Skript für eine Sachbearbeiterin können auch als Algorithmen verstanden werden. Vor allem in vielen formalisierten, bürokratischen Verfahren wird ein Input auf Grundlage einer definierten Schrittfolge in einen gewünschten Output umgewandelt. Analoge Algorithmen sind nicht neu, im Gegenteil prägen sie menschliches Zusammenleben bereits seit langer Zeit. Der erste Filter des Analysebereichs dieser Expertise lautet also: Ist das Entscheidungssystem digital?

Ein digitales Entscheidungssystem liegt vor, wenn der zugrunde liegende Algorithmus von einem Computer ausgeführt wird. Die Beschränkung auf digitale algorithmische Entscheidungssysteme bedeutet jedoch nicht, dass ihre soziale Einbettung ignoriert wird. Vielmehr ist dies eine pragmatische Fokussierung auf solche Fälle, die neuartige Fragestellungen in Bezug auf Teilhabe produzieren.

4.2 Bewertung von Menschen

Teilhaberelevant sind insbesondere solche Entscheidungsprozesse, die Menschen oder ihnen klar zugeordnete Eigenschaften und Erzeugnisse bewerten. Darunter fallen auch Systeme, die die Verbreitung und Produktion von Inhalten steuern, etwa in sozialen Medien. Sie können direkt oder indirekt auf Teilhabe wirken. Zum Beispiel wird Diskriminierung aufgrund persönlicher Merkmale oft erst durch den Blick auf gesamtgesellschaftliche Phänomene sichtbar. In der auf eine konkrete Person zugeschnittenen Nachrichtenauswahl, etwa dem individuellen Facebook-Newsfeed, sind für sich genommen kaum diskriminierende Strukturen zu erkennen, denn sie soll ja personalisiert sein. Erst in der Gesamtschau auf viele personalisierte Entscheidungen lässt sich gegebenenfalls Benachteiligung ausmachen. Ähnlich verhält es sich, wenn eine Software Bewerbungen vorsortiert und dabei Bewerberinnen aus bestimmten Stadtteilen überproportional oft nicht fürs Bewerbungsgespräch vorschlägt. Für den einzelnen nicht erkennbar, ob hier die Teilhabe einer bestimmten Gruppe an Menschen beeinträchtigt werden kann, wenn der betroffene Stadtteil etwa durch ein niedriges Einkommensniveau oder einen hohen Migrationsanteil geprägt ist.

Algorithmen, die in automatisierten Fertigungsprozessen, zum Beispiel in der Industrie oder in der Landwirtschaft, zum Einsatz kommen, werden durch den Filter „Bewertung von Menschen“ ausgeklammert. Auch die Aufzugssoftware, die mithilfe von Algorithmen automatisch entscheidet, welcher Aufzug zuerst in welches Stockwerk

fährt, soll in dieser Expertise nicht auf Teilhaberelevanz geprüft werden. Fälle wie die Abgasmanipulationen von Volkswagen können im Einzelfall zwar durchaus Teilhaberelevanz entfalten. Um den Analysebereich handhabbar einzuschränken, sollen sie in der vorliegenden Systematik aber nicht näher betrachtet werden.

Unser Klassifizierungsvorschlag analysiert somit Computersysteme, die Menschen bewerten. Diese Vorauswahlkriterien werden in Kapitel 6 wieder aufgegriffen. Für andere Zwecke oder Kontexte können engere Einschränkungen des Untersuchungsfelds sinnvoll sein. Den Analysebereich kann jede leicht anhand der verwendeten Teilhabedefinition einengen.

5 Anwendungsszenarien algorithmischer Entscheidungsfindung

Um die Klassifizierung algorithmischer Entscheidungsprozesse empirisch zu unterfüttern, werden im Folgenden Anwendungsszenarien aus verschiedenen gesellschaftlichen Bereichen erläutert. Diese Bestandsaufnahme erfolgt schlaglichtartig mit Blick auf die Relevanz für Teilhabe und erhebt keinen Anspruch auf Vollständigkeit. Im Fokus steht, die Vielfalt der Einsatzzwecke und betroffenen Lebensbereiche exemplarisch abzubilden, die hinsichtlich gesellschaftlicher Teilhabe wichtig sind (für eine ausführliche Analyse von Fallbeispielen siehe Lischka und Klingel 2017).

5.1 Arbeitswelt

5.1.1 Plattformarbeit: Ein Algorithmus als Chef?

In der Plattformökonomie werden in immer stärkerem Maße Algorithmen eingesetzt, um Arbeit zu verteilen, zu bewerten und zu verbessern (Agrawal et al. 2015; Bundesministerium für Arbeit und Soziales 2015; Choudary, Alstynne und Parker 2016). Algorithmen dienen hier vor allem zur Automatisierung und Effizienzsteigerung. Unter dem Begriff Plattformarbeit werden vielfältige neue Beschäftigungsmodelle und Managementprozesse zusammengefasst. Wesentlicher Bestandteil von Plattformarbeit ist es, dezentrale Auftragnehmerinnen mit ebenso dezentral verteilten Auftraggeberinnen zu verbinden. Neue technische Infrastrukturen ermöglichen neue Kommunikationswege, über die soziale Ressourcen geteilt und vielfältige Interessen zusammengebracht werden können (Hensel et al. 2016; Kocher und Hensel 2016). Die betroffenen Branchen und Arbeitsbereiche reichen von einfacher bis zu hoch qualifizierter Industriearbeit (z. B. Clickworker, TaskRabbit, Amazon Mechanical Turk), umfassen aber auch haushaltsnahe und persönliche Dienstleistungen (z. B. Uber, Lieferheld, careship.de); für einen ausführlichen Überblick siehe Leimeister, Durward und Zogaj (2016) sowie Schmidt (2016).

Über Software automatisieren und gestalten die Crowdwork-Unternehmen Arbeitsprozesse für Beschäftigte. Weisungen, Zugangs- und Auswahlentscheidungen bis hin zu Kontrollmechanismen erfolgen über Algorithmen. Außerdem werden meist spezifische Bewertungs- und Reputationssysteme installiert, deren Auswertung Einfluss auf zukünftige Auftragsvergaben hat. Lee et al. (2015) haben die Koordination von Arbeit über solche Intermediäre daher als „algorithmisches Management“ beschrieben. Das umfasst die Steuerung und Abstimmung zwischen den Betreiberunternehmen, Auftraggebern, Beschäftigten und Endkundinnen über eine meist zentrale Plattformsoftware. Die Plattform-Algorithmen übernehmen demnach klassische Managementfunktionen, wie zum Beispiel die Koordination von Arbeit, Optimierung von Arbeitsabläufen, Bewertung von Leistungen und Steuerung der Schichtplanung.

Die Koordination von Arbeit durch Algorithmen berührt mehrere Grundrechte, darunter das Recht auf Privatsphäre am Arbeitsplatz und das Recht auf Mitbestimmung und kollektive Interessenvertretung. Die dezentrale Aufteilung der Arbeit und die Abhängigkeit von den Kommunikationsstrukturen der Plattformen hat das Potenzial, das Schutzbedürfnis der Beschäftigten gegenüber den Plattformbetreiberinnen und Auftraggebern zu untergraben (Arthurs 2011; Freedland und Kountouris 2011).

Plattformarbeit lässt sich nicht in die klassische Dichotomie der Organisationsprinzipien „Hierarchie“ (abhängige Beschäftigung) und „Markt“ (Selbstständigkeit) eingliedern. Vielmehr steht sie für ein hybrides Organisationsprinzip, das wesentlich durch Algorithmen geprägt ist (Aneesh 2009). Die Arbeitsplattformen beziehen vielfältige Merkmale in die Bewertung der Beschäftigten mit ein. Es gibt auch spezifische Feedback-Mechanismen der Arbeitstätigkeit, die die Kunden in die Kontrolle einbeziehen und Ratingsysteme, die mitunter die klassische Weisung ersetzen (De Stefano 2015; Lingemann und Otte 2015). Die Gestaltung von Arbeitsbedingungen durch Plattformen wirft zudem Fragen über die Selbstbestimmung der Beschäftigten auf. Viele schätzen die erweiterte zeitliche und räumliche Flexibilität der Arbeit. Als kleiner Teil einer großen „Crowd“ werden Menschen, die auf Plattformen arbeiten, oft aber auch zu „humans as a service“ herabgestuft (Irani 2015). Die nahezu perfekte Austauschbarkeit entsteht unter anderem durch die globale Konkurrenz der Crowdworkerinnen und den geringen

Einarbeitungsaufwand für viele Tätigkeiten. Ohne auf Einzel- oder Kollektivinteressen eingehen zu müssen, definieren Algorithmen die Grenzen der Autonomie der Beschäftigten.

5.1.2 Bewerberauswahl: Wer wird zum Bewerbungsgespräch eingeladen?

In Großbritannien und den Vereinigten Staaten werden Umfragen zufolge 60 bis 70 Prozent der Bewerberinnen automatisierten Auswahlverfahren und Tests unterzogen (Lischka und Klingel 2017: 21). Die Aussagekraft der dort nachweislich genutzten Persönlichkeitstestverfahren über die Arbeitsleistung ist umstritten, die Methoden sind nicht transparent. Es gibt Hinweise, dass einige der Verfahren bestimmte Gruppen benachteiligen (etwa Menschen aus armen Wohngebieten oder mit Erkrankungen). Die zentrale Funktion der Algorithmen ist hier die Steigerung der Effizienz.

Die automatisierten Verfahren dienen der Vorauswahl von Bewerbungen. Ein Teil wird sofort auf Basis der Onlinetests abgelehnt, noch bevor ein Mensch die Bewerbungen sieht. Wie hoch der Anteil dieser Ablehnungen ist, kann der Arbeitgeberinnen bestimmen. Ein Testanbieter spricht von 30 Prozent automatisierter Ablehnungen (Weber und Dwoskin 2014).

Grundsätzlich besteht beim Einsatz solcher Systeme die Chance, Benachteiligungen abzubauen, wenn Annahmen in der Bewerbungssortierung transparent gemacht und überprüft werden. Für bestimmte Personengruppen können die automatisierten Verfahren in dieser Form aber den Zugang zum Arbeitsmarkt insgesamt erschweren oder gar vollständig blockieren. Eine bereits bestehende Form von Diskriminierung – zum Beispiel die Benachteiligung von Menschen mit „ausländisch“ klingenden Namen (vgl. Schneider, Yemane und Weinmann 2014) – droht so durch eine andere ersetzt zu werden. Da die Technik vor allem im Niedriglohnsektor zum Einsatz kommt, dürften Unternehmen wenig geneigt sein, in die Verbesserung und Prüfung der Systeme zu investieren. Denn: „Die Systeme müssen nicht alle Besten finden, sondern nur effizienter sein als das vorherige Auswahlssystem. Investitionen in die Kalibrierung der Systeme, die fortwährende Prüfung und Aktualisierung der Entscheidungslogiken und des Datenbestandes werden auf diesem Einsatzgebiet nie so hoch sein wie bei Stellen mit geringem Angebot an Arbeitskräften, hohem Bedarf und entsprechend hohen Gehältern“ (Lischka und Klingel 2017: 13).

5.2 Sicherheit

5.2.1 Polizei: Wo wird Streife gefahren?

Polizeibehörden setzen Algorithmen ein, um besonders kriminalitätsbelastete Orte zu prognostizieren. Im Kern geht es um eine bessere Antizipation durch die Nutzung von Algorithmen. Sie berechnen etwa Hotspots für bestimmte Delikte wie Wohnungseinbrüche. Die Ergebnisse werden genutzt, um Prioritäten beim Einsatz von Streifen zu setzen. Solche Systeme sind in Europa im regulären Einsatz, zum Beispiel in Zürich (Baumgartner 2015), Kent (Mohler et al. 2015) und Mailand (Mastrobuoni 2015). In Deutschland ist diese Form des sogenannten „Predictive Policing“ derzeit in 14 Pilotprojekten und Tests im Einsatz oder in der Entwicklung (Pilpul 2016). Zu den bekanntesten Analyseprogrammen mit diesem Ansatz zählen Predpol aus den USA und Precobs aus Deutschland.

Mit geographischen Mustern von Kriminalität arbeiten Behörden und Kriminologen seit dem 19. Jahrhundert, damals zuerst in London (Gluba 2014: 5). Lange bevor Software zum Einsatz kam, leiteten menschliche Analysten aus Statistiken ortsbezogene Kriminalitätsschwerpunkte ab. Zum Hintergrund: „Die Prognose von Hotspots für bestimmte Delikte basiert auf der Near-Repeat-Theorie. Dieser kriminologische Ansatz geht davon aus, dass bei Delikten wie Autodiebstahl, Wohnungs- und Autoeinbrüchen die Wahrscheinlichkeit für weitere Taten im örtlichen Umfeld nach einem Delikt steigt“ (Lischka und Klingel 2017: 13) Für Wohnungseinbrüche zeigen empirische Studien aus Großbritannien, den USA, den Niederlanden, Neuseeland und Australien statistisch signifikante Near-Repeat-Muster (Ferguson 2012: 19).

Precobs analysiert darüber hinaus wenige Parameter, wie Tatzeit, Beutetyp (z. B. Bargeld, Geldkassette), Gebäudetyp (z. B. Geschäfts- oder Wohngebäude), Vorgehen beim Einbruch (z. B. Hebeln, Drücken, Treten), um Muster von Serientätern zu erkennen (Brühl 2014). Die Vorhersage von Kriminalitätsschwerpunkten ist mithin nur ein Einsatzbereich von Predictive-Policing-Technologien. In Chicago werden Vorhersagetechnologien auch zur Identifizierung von einzelnen, potenziellen Straftätern genutzt (Davey 2016).

5.2.2 Justiz: Wer kommt auf Bewährung frei?

In vielen US-Bundesstaaten prognostiziert Software die Rückfallwahrscheinlichkeit von Straftäterinnen (Barry-Jester, Casselman und Goldstein 2015). Mehr als 60 Prognoseinstrumente sind auf dem Markt, viele kommen von privaten Unternehmen, darunter das weit verbreitete System COMPAS der Firma Northpointe. Die Softwareprodukte werden zur Antizipation von Verhalten und zur Klassifizierung von Personen eingesetzt.

Das COMPAS-System bewertet Straftäterinnen in 21 Kategorien, darunter auch das Rückfallrisiko bei Gewalttaten. Für alle Ausprägungen wird ein Wert auf einer Skala von 1 bis 10 ausgegeben. Bei einer Risikoprognose von 8 bis 10 sei die Rückfallwahrscheinlichkeit „hoch“, heißt es im Anwenderhandbuch (Northpointe 2015: 11). Die Scoring-Werte leitet das COMPAS-Verfahren aus Antworten auf 137 Fragen ab. „Welche Antworten bei der Berechnung wie gewichtet werden, ist nicht öffentlich“ (Lischka und Klingel 2017: 8).

Erste Untersuchungen zu den Prognoseverfahren deuten auf rassistische Diskriminierung hin: Der Anteil Schwarzer mit hoher Rückfallprognose, aber ohne Rückfall binnen zwei Jahren ist doppelt so hoch wie der Weißer (Angwin et al. 2016).

5.3 Meinungsbildung

5.3.1 Newsfeeds: Welche Inhalte werden präsentiert?

Intermediäre wie soziale Netzwerke und Suchmaschinen nutzen Prozesse algorithmischer Entscheidungsfindung, um für jeden Nutzer Inhalte personalisiert zusammenzustellen. Sie handeln dabei basierend auf Annahmen über Nutzerpräferenzen und machen Vorhersagen über Interessen der Nutzerinnen. Die dort eingesetzten Algorithmen erfüllen also die Kernfunktion der Individualisierung. Die Anbieter sammeln und bewerten vielfältige Nutzerdaten, um individualisierte Newsfeeds zu generieren. Der Strom an Bildern, Videos und Nachrichten zeigt nur einen Teil aller Inhalte, wodurch Social-Media-Plattformen mit großer Reichweite eine redaktionelle Funktion zukommt (Helberger und Trilling 2016). Newsfeed-Algorithmen wie Facebooks „EdgeRank“ selektieren und sortieren Inhalte, die für die Nutzer sichtbar sind. Darin liegt die Macht, Inhalte für die Nutzerinnen unbemerkt bevorzugen und benachteiligen zu können (York 2010).

Die algorithmische Auswahl von Inhalten betrifft die freie Meinungsäußerung unmittelbar. Darüber hinaus können Grundrechte und demokratische Grundprinzipien indirekt berührt sein. Facebook hat im Rahmen eines Experiments 2010 gezeigt, dass es mithilfe seines Newsfeeds Wahlverhalten beeinflussen kann (Bond et al. 2012). Dazu wurde am 2. November 2010, dem Tag der Mid-Term-Kongresswahlen in den USA, manchen Nutzerinnen auf Facebook angezeigt, wie viele ihrer Bekannten bereits gewählt hatten. Außerdem konnte man einen Button anklicken, um ebenfalls anzuzeigen, dass man gewählt hatte. Bei etwa zwei Prozent der Nutzerinnen, die die Wahlmeldung von ihren Freunden gesehen haben, wurde durch die Intervention die Wahrscheinlichkeit erhöht, ebenfalls auf „Ich habe gewählt“ zu klicken. Nicht alle davon sind jedoch tatsächlich auch wählen gegangen. In einem Abgleich zwischen den Facebook-Daten und der tatsächlichen Wahlbeteiligung zeigte sich nur eine Erhöhung von 0,4 Prozent. Die Kontrollgruppe wies keine Erhöhung der Wahlbeteiligung auf. Bei einem knappen Wahlausgang könnte ein solcher Effekt theoretisch groß genug sein, um über den Ausgang einer Wahl zu entscheiden, insbesondere in Mehrheitswahlsystemen (Tufekci 2014; Zittrain 2014). Eine gezielte Beeinflussung wäre schwer festzustellen und nachzuweisen, da sie erst in der Gesamtschau über viele Profile erkennbar würde. Auch für Suchergebnisse konnte experimentell ein Effekt auf die Wahlentscheidung unentschlossener Wählerinnen festgestellt werden (Epstein und Robertson 2015). Vor allem die obersten „Toptreffer“ in den Suchergebnissen hatten Einfluss auf die Meinungsbildung der Probanden.

5.3.2 Inhaltskontrollen: Welche Inhalte werden entfernt?

Der Einfluss von Intermediären liegt nicht nur darin, welche Inhalte sie anzeigen und priorisieren. Welche Inhalte nicht angezeigt oder gezielt entfernt werden, ist genauso teilhaberelevant. Viele Intermediäre, wie Facebook, Google und Microsoft, bearbeiten Meldungen von Inhalten und Löschanfragen mithilfe automatisierter oder teilautomatisierter Prozesse (Wagner 2016; Zhang, Stalla-Bourdillon und Gilbert 2016). Onlinedienste aller Art sind auf automatisierte Lösungen zum Filtern und Blockieren von Inhalten angewiesen, sei es zum Schutz vor Spam oder illegalen Inhalten, wie Beleidigungen oder Volksverhetzung.

Wenn Algorithmen Inhalte überprüfen und entfernen, ist wiederum die Meinungsfreiheit direkt betroffen. In vielen Fällen orientiert sich die Inhaltskontrolle nicht an den gesetzlichen Schranken, sondern an den allgemeinen Geschäftsbedingungen des jeweiligen Plattformbetreibers. Es entsteht ein Spannungsfeld zwischen dem, was gesetzlich erlaubt ist, und dem, was Intermediäre an Inhalten zulassen. Große Dienstanbieter stehen zunehmend unter Druck – vor allem durch Polizeibehörden –, Inhalte auch aufgrund vager, intransparenter Kriterien wie „terroristische Propaganda“ zu entfernen (Brown und Cowsls 2015; Europol 2015). Vollzugsbehörden verlagern die Kontrolle unerwünschter Inhalte hin zu den privaten Betreibern, was im Ergebnis nicht selten zulasten der Meinungsfreiheit geht.

5.4 Öffentliche Daseinsvorsorge

5.4.1 Jobvermittlung: Welche Jobmaßnahme ist die richtige?

Selbstverständlich nutzen auch öffentliche Verwaltungen in zunehmendem Maße automatisierte Verfahren (Eubanks 2013; Tufekci et al. 2015; van Haastert 2016). In Polen werden algorithmische Entscheidungssysteme zur Vermittlung von Arbeitsplätzen für Erwerbslose eingesetzt. Die verwendete Software unterteilt die Erwerbslosen in drei Kategorien, die mit drei verschiedenen staatlichen Arbeitsmarktprogrammen korrespondieren (Niklas 2017). Die Daten für die Bewertung der Erwerbslosen werden mithilfe einer Onlinebefragung gesammelt, auf deren Grundlage ein Scoring berechnet wird. Die Erstellung von Profilen durch die zuständige Behörde erfolgt seit 2014. Angaben zu Alter, Geschlecht, körperlichen oder geistigen Behinderungen und Familienstand fließen in das Entscheidungssystem ein. Forscherinnen der Panoptikon Stiftung untersuchten das Verfahren (Niklas, Sztandar-Sztanderska und Szymielewicz 2015). Sie betonen, dass die Nutzung von Algorithmen bei der Erbringung öffentlicher Dienstleistungen große Herausforderungen bei Nachvollziehbarkeit und klarer Verantwortlichkeit mit sich bringt. Da sich zum Beispiel die Auswahl von Qualifizierungsmaßnahmen sehr stark auf Teilhabechancen auswirkt, ist die Nutzung solcher Verfahren genau zu beobachten und kontinuierlich zu evaluieren.

5.4.2 Gesundheitsschutz: Welche Häuser zuerst sanieren?

Bleivergiftungen bei Kindern sind noch immer ein großes Problem in Chicago, weil wesentliche Teile des Wohnungsbestandes aus einer Zeit stammen, in der bleihaltige Farbe verwendet wurde. 2013 hatten zehn Prozent der Kinder unter sechs Jahren Bleikonzentrationen im Blut, die über dem von der US-Gesundheitsbehörde festgesetzten Grenzwert lagen (Hawthorne 2015). Die Stadt entwickelte deshalb mit der University of Chicago eine Software, die prognostizieren soll, in welchen Gebäuden und bei welchen Kindern das Risiko einer Bleivergiftung besonders hoch ist. Damit sollen frühe, zielgerichtete und daher günstige Interventionen ermöglicht werden. Das Risikoranking für Gebäude hilft, Schutzmaßnahmen für die Bevölkerung zu priorisieren (Potash et al. 2015). Zum Beispiel können dann Hochrisikogebäude schneller durch Inspektoren untersucht und bei Überschreiten der Grenzwerte renoviert werden.

Als Datenbasis stehen den Forschern zur Verfügung: 2,5 Millionen Bluttests mit Datum, Identität der Testpersonen, Alter und Wohnorte, zudem die Ergebnisse von 120.000 Hausinspektionen mit Datum- und Ortsangaben. Die bisherige Forschung zeigt, dass insbesondere Informationen über Alter, Inspektionsergebnisse und Zustand von Gebäuden die Prognosequalität auf Adressebene verbessern (a.a.O.: 2044). Das Verfahren ist ein

Beispiel dafür, wie die öffentliche Verwaltung von algorithmischer Entscheidungsfindung profitieren kann. Natürlich wäre es noch wirksamer, mit zusätzlichen Mitteln den Gesamtwohnungsbestand zu sanieren. Vor dem Hintergrund knapper Ressourcen stellt das Risikoranking vermutlich dennoch eine wesentliche Verbesserung für die öffentliche Gesundheitsfürsorge dar.

5.5 Marketing

5.5.1 Empfehlungssysteme: Wofür interessierst Du Dich als Nächstes?

Viele Onlinedienste verwenden Empfehlungssysteme, um einzelnen Nutzerinnen bestimmte Inhalte oder Produkte vorzuschlagen, zum Beispiel Bücher oder Videos. Mithilfe von Cookies und Technologien, wie dem sogenannten Browser-Fingerprinting sowie weiteren zur Verfügung stehenden Daten, etwa aus Suchanfragen, sozialen Netzwerken oder Smartphone-Apps, können sehr detaillierte Nutzerprofile erstellt werden (Tene und Polonetsky 2011). Auch ausgeloggte Nutzerinnen erhalten vermehrt personalisierte Darstellungen der Webseiten, sodass ein „Opt-out“ nicht ohne Weiteres möglich ist.

Die gesammelten Informationen werden mithilfe von Algorithmen analysiert, um personalisierte Empfehlungen auszusprechen. Personalisierung war in vielen Bereichen bisher ein Luxusgut, das bei hohem Kostenaufwand nur wenigen zugänglich war. Algorithmische Empfehlungssysteme können kostengünstige Personalisierung für alle ermöglichen. Dabei stellt sich die Frage, ob sie den Wünschen und Präferenzen der Nutzerinnen folgen und welche Wirkung sie zum Beispiel auf Meinungs- und Medienvielfalt haben.

Profilbildung birgt die Gefahr, dass die verwendeten Daten aus ihrem ursprünglichen Zusammenhang gerissen werden. Die Empfehlungssysteme konstruieren einen „digitalen Doppelgänger“ eines Menschen, der nicht deckungsgleich mit der natürlichen Person ist, die repräsentiert werden soll (Bauman et al. 2014). Außerdem können Filterblasen (Pariser 2011) entstehen, die dazu führen, dass Personen nur noch einseitigen Inhalten ausgesetzt sind. Die tatsächlichen Effekte solcher Empfehlungfilterblasen sind umstritten, dennoch gilt es genau hinzusehen, ob und wie sie zustande kommen (O’Callaghan et al. 2016). Die Anpassung der priorisierten Inhalte nach persönlichen Merkmalen, wie Geschlecht oder Einkommen, kann diskriminierende Effekte haben. Eine Studie ergab: Auf Rechnern, die mit einem als weiblich deklarierten Google-Werbeprofil verknüpft sind, spielt Google seltener Anzeigen für hoch bezahlte Stellenausschreibungen ab als auf Rechnern, die mit einem männlichen Profil verknüpft sind. Wie dieser Unterschied zustande kommt (z. B. durch gezielte Buchungen), konnten die Forscherinnen nicht herausfinden (Datta, Tschantz und Datta 2015).

5.5.2 Versicherungen: Wer zahlt welche Prämien?

Die drittgrößte Autoversicherung Großbritanniens, Admiral Group, gab im November 2016 bekannt, künftig die Facebook-Profile ihrer Kunden analysieren zu wollen. Mit den Daten sollte der Preis für die Versicherung von Fahranfängerinnen bestimmt werden (Stalder 2017). Da für sie noch keine Daten zum Fahrverhalten und keine Unfallhistorie existieren, wollte die Versicherung aus den Facebook-Daten Persönlichkeitseigenschaften identifizieren, die mit sicherem Autofahren in Verbindung stehen. Der eingesetzte Algorithmus sucht nach Korrelationen zwischen den Social-Media-Daten und den vorhandenen Daten über Schadensfälle (Ruddick 2016a). Die Beteiligung am Programm sollte freiwillig sein und Rabatte für die Teilnehmenden ermöglichen. Im letzten Moment verweigerte Facebook jedoch die Verwendung seiner Daten, stattdessen will die Versicherung nun über einen Fragebogen Daten von den Führerscheineulungen erheben (Ruddick 2016b). Die Individualisierung von Versicherungstarifen kann einerseits Kosten reduzieren, andererseits aber auch mit solidarischen Prinzipien kollidieren und einzelne Personengruppen stärker benachteiligen als andere.

Die Einwilligung in das Programm erfolgt bei vielen jungen Autofahrerinnen unter Umständen nicht freiwillig, wenn sie es sich nicht leisten können, auf den angebotenen Rabatt zu verzichten. Ähnliche Versicherungstarife

beziehen auch Daten über den Fahrstil mit ein. Der Telematik-Tarif „BonusDrive“ der Allianz² etwa richtet sich ebenfalls an junge Fahrer, die noch keine Unfallhistorie haben. Über eine App werden Daten zum Fahrverhalten an die Versicherung übermittelt. Wer Abstand hält und Tempolimits nicht überschreitet, soll Preisnachlässe erhalten. Nach welchen Kriterien die Fahrten im Einzelnen konkret bewertet werden und welchen Einfluss das auf die Preisgestaltung hat, wird nicht mitgeteilt.

5.6 Zwischenfazit: Welche Funktionen erfüllen Algorithmen?

Die Anwendungsszenarien zeigen, wie breit das Spektrum teilhaberelevanter Algorithmen ist. Im Folgenden werden die in den Fallbeschreibungen skizzierten Funktionen algorithmischer Entscheidungsfindungsprozesse kurz einzeln beschrieben. Die verwendete Systematik baut dabei auf den Ausarbeitungen von Gillespie (2014) und Diakopoulos (2016) auf. Damit werden die wichtigsten Aufgaben von Algorithmenbasierten Entscheidungssystemen zusammengefasst.

5.6.1 Automatisierung

Der zentrale Zweck von algorithmischen Entscheidungssystemen ist Automatisierung, das heißt zu handeln, ohne dass ein Mensch dabei interveniert (Winner 1978). Die Fähigkeit, automatisiert zu agieren bzw. zu reagieren, ermöglicht es Computersystemen, in immer mehr Situationen Menschen zu ersetzen. Automatisierte Systeme kommen sowohl für einfache als auch für komplexe Aufgaben zum Einsatz. Nicht immer ist die automatisierte Entscheidung eines Algorithmus für Menschen vorhersehbar oder erklärbar. Um automatisch verarbeitet zu werden, müssen die Inputdaten für den Algorithmus vorbereitet und angepasst werden. Automatisierung ist allerdings nicht mit Autonomie gleichzusetzen. Entscheidungssysteme sind nie vollständig autonom, sondern operieren immer innerhalb eines vorgegebenen Rahmens und eines sozialen Umfelds. Bei der Entwicklung und Implementierung werden stets Grenzen gesetzt, die die Funktionsweise und konkrete Aufgabe des Algorithmus vorgeben (Stalder 2017).

5.6.2 Effizienzsteigerung

Eng verknüpft mit der Automatisierung ist die Absicht, Prozesse effizienter zu gestalten. Effizienzsteigerungen resultieren aus einer Beschleunigung von Prozessen (technische Effizienz) und aus der verbesserten Skalierbarkeit algorithmischer Verfahren (ökonomische Effizienz). Letzteres bedeutet, dass die Grenzkosten oft sehr niedrig sind und mit relativ geringen Mitteln eine große Zahl an Fällen bearbeitet werden kann. Die Kostensenkung entsteht jedoch auch, weil menschliche Arbeit ersetzt wird und damit in verschiedenen Einsatzbereichen Personalkosten eingespart werden können. Das bedeutet zugleich, dass vielfältige, verteilte Entscheidungsinstanzen durch ein zentrales Entscheidungssystem ersetzt werden.

5.6.3 Priorisierung

Viele von Algorithmen beeinflusste Entscheidungen bewirken eine Priorisierung bzw. Hierarchisierung. Diese Urteile kommen im Prinzip „subjektiven Entscheidungen“ gleich, bei denen eine Bewertung vorgenommen wird, für die es keine offensichtlich richtige oder falsche Antwort gibt, kein klares „Ja“ oder „Nein“ (Pasquale 2015: 8). Dazu zählen Fragen wie „Sollte diese Beschäftigte eine Pause machen?“ oder „Welche Inhalte sollten am Wahltag ganz oben in einem Newsfeed stehen?“. Algorithmen definieren also mit, was relevant ist. Priorisierung als Filterung von Informationen gehört zum Alltag, ohne sie könnten wir Menschen nicht mit der Informationsflut umgehen, die tagtäglich auf uns einströmt. Algorithmen helfen uns, Zeit, Aufmerksamkeit, Geld und andere knappe Ressourcen zielgerichtet einzusetzen (Diakopoulos 2016: 57).

² <https://www.allianz.de/auto/kfz-versicherung/telematik-versicherung/> (abgerufen am 1.03.2017)

5.6.4 Klassifizierung

Viele algorithmische Entscheidungsverfahren ordnen Gegenstände, Personen, Begriffe o. Ä. bestimmten Klassen zu. Klassifizierung bedeutet Sortierung in vorab definierter Kategorien anhand typischer Merkmale. Wenn Menschen in Gruppen einsortiert werden, spricht der Soziologe David Lyon von „social sorting“ (Lyon 2003), also einer Art sozialen Klassifizierung. Die Einordnung in eine bestimmte Gruppe, zum Beispiel in die Kategorie „wird wahrscheinlich rückfällig“, kann dann wiederum allerlei Folgeentscheidungen nach sich ziehen. Klassifizierungen ziehen grundsätzlich „falsch positive“ oder „falsch negative“ Eingruppierungen nach sich. Ein Mensch, der als „wird nicht rückfällig“ klassifiziert wird, dann aber doch rückfällig wird, ist zum Beispiel falsch negativ klassifiziert.

5.6.5 Individualisierung

Wenn Systeme auf Einzelpersonen bzw. Einzelaccounts zugeschnitten werden, spricht man von Individualisierung. Die Anpassung an die Nutzerinnendaten erfolgt meist inkrementell, das heißt langsam durch das Hinzuziehen weiterer Daten über die Nutzerinnen im Zeitverlauf. Algorithmen ermöglichen Individualisierungen von Dienstleistungen und anderen Gütern in neuen Branchen und Bereichen. Nicht immer ist damit jedoch ein Mehr an Selbstbestimmung für die Nutzerinnen verbunden. Wenn Algorithmen etwa Nutzungsbedingungen oder Preise einzeln anpassen, kann es auch zu neuen Formen von Diskriminierung kommen.

5.6.6 Antizipation

Algorithmen werden auch eingesetzt, um Aussagen über die Zukunft zu machen. Die Vorhersage basiert auf statistischen Berechnungen. Es wird versucht, bestimmte Muster oder Ereignisse mit erhöhter Wahrscheinlichkeit vorherzusagen. Die Antizipation geht davon aus, dass eine bestimmte Beobachtung, etwa eine identifizierte Beziehung zwischen zwei oder mehreren Variablen, auch in der Zukunft eintreten wird. Die Prognosegenauigkeit erstreckt sich jedoch nur auf abstrakt definierte Gruppen von Fällen und kann keine präzisen Vorhersagen für Einzelfälle treffen.

6 Die potenzielle Wirkung algorithmischer Entscheidungsfindung auf Teilhabe berechnen

Je stärker der mögliche Effekt auf Teilhabe, desto genauer muss ein Verfahren geprüft werden. Das gilt auch im Positiven: Je größer die potenzielle Wirkung auf Teilhabe, desto höher kann auch das gesellschaftliche Verbesserungspotenzial ausfallen. Auch dann lohnt es sich, genauer hinzusehen. In jedem Fall gilt es, umso höhere Ansprüche an algorithmische Prozesse zu stellen und mögliche Folgen ausführlich zu reflektieren, je stärker diese Teilhabe berühren.

Die Bewertung der Teilhabewirkung, für die wir hier einen Strukturierungsvorschlag machen, bietet eine Grundlage für weitere Schritte, seien es eingehendere Analysen oder Maßnahmen zur Regulierung. Wenn eine hohe potenzielle Wirkung auf Teilhabe zu erwarten ist, muss konsequent dafür gesorgt werden, dass die positiven Effekte ausgeschöpft und negative Effekte, etwa Diskriminierung durch algorithmische Prozesse, verhindert werden. Gerade bei großer Teilhabewirkung von Algorithmen geht es deshalb oft darum, Vielfalt, fairen Wettbewerb und rechtsstaatliche Grundsätze abzusichern. Je höher der Anspruch an teilhaberelevante algorithmische Entscheidungsprozesse, desto umfangreicher sind die potenziellen Maßnahmen, die infrage kommen.

6.1 Methodischer Ansatz

Um die potenzielle Wirkung auf Teilhabe zu klassifizieren, kann es kein binäres Entscheidungskriterium geben. Das würde die vielen sozial konstruierten Logiken ignorieren, die in Prozessen algorithmischer Entscheidungsfindung eingebettet sind.

Eine Vielzahl von Kriterien kann Auswirkungen auf die Teilhaberelevanz haben. Die Auswahl und Gewichtung dieser Kriterien hängt dabei auch von der jeweiligen Teilhabedefinition ab. Es geht nicht um eine Bewertung guter oder schlechter Verfahren algorithmischer Entscheidungsfindung, sondern um eine relative Beurteilung ihres potenziellen Wirkungsgrads auf Teilhabe, gleich in welcher Richtung.

Die Kriterien gliedern sich in drei Gruppen: *Akteure*, *soziale Einbettung* und *Konsequenzen*. Sie orientieren sich also nicht an technologischen Gegebenheiten, sondern versuchen, politische und soziale Konstellationen zu erfassen.

Die Akteurskriterien untersuchen die effektive ökonomische und politische Macht hinter den Betreiberinnen der Entscheidungsprozesse. Bei den Prüfkriterien zur sozialen Einbettung geht es darum, die beabsichtigten und unbeabsichtigten gesellschaftlichen Wechselwirkungen zu bedenken. Schließlich werden die möglichen oder bereits real erkennbaren Konsequenzen in Bezug auf politische und soziale Grundrechte analysiert. Als inhaltliche Richtschnur für die Bewertung der Konsequenzen für Teilhabe ziehen wir hier rechtliche Normen heran. Als Basis dient der Anwendungsbereich des Allgemeinen Gleichstellungsgesetzes (§§ 1 und 2 AGG). Demnach ist Benachteiligung insbesondere dort unzulässig, wo es um den Zugang zur Erwerbsarbeit, die Arbeitsbedingungen, die Mitgliedschaft in Gewerkschaften und die Berufsausbildung geht. Auch den fairen Zugang zur Versorgung mit öffentlichen Gütern, wie Bildung, soziale Absicherung, Gesundheitsschutz oder Wohnraum, gilt es sicherzustellen.

Wie in Kapitel 4 bereits dargestellt, lässt sich prinzipiell für fast jeden Prozess algorithmischer Entscheidungsfindung ein (theoretisches) Fallbeispiel entwerfen, bei dem eine potenzielle Wirkung auf Teilhabe erkennbar wird. Daher schlagen wir vor, erst ab einer bestimmten Grundschwelle von Teilhabewirkung auszugehen. Mithilfe der nachfolgend skizzierten Kriterien kann die relative Teilhabewirkung überblicksartig erfasst und verglichen werden. Dies ermöglicht die Erstellung einer vereinfachten und handhabbaren Rangfolge von Algorithmen, um bei Bedarf weitere Schritte zu priorisieren.

Jedes der unten aufgeführten Prüfkriterien kann das *Teilhabewirkungspotenzial* (TWP) erhöhen (TWP = +1). Wenn ein Kriterium nicht zutrifft bzw. sich im Schnellverfahren nicht bewerten lässt, wird kein Teilhabewirkungspunkt vergeben (TWP_{neutral} = 0). Pro Kriterium können auch halbe Punkte (TWP = +0.5) vergeben werden, jedoch nicht mehr als zwei (TWP_{max} = +2). Das Gesamtspektrum reicht also von null bis zwei Punkten pro Kriterium. Je höher der Gesamtwert, desto größer die potenzielle Wirkung des Verfahrens auf Teilhabe.

Die Kriterien sind mit beispielartigen Testfragen hinterlegt. Die Beantwortung dieser Fragen fällt nicht immer leicht. Doch gerade weil die Anwendungsfälle so vielfältig und mitunter unübersichtlich sind, liegt hier der Mehrwert der abstrakten Klassifizierung: Sie reduziert Komplexität in einem von hoher Komplexität geprägten Themenfeld. Was kann Teilhabe einschränken oder stärken? Eine Organisation oder Person analysiert entlang der Prüfstruktur einen Prozess. Nach Anwendung der Kriterien auf den Einzelfall kann sie anhand des Ergebnisses einschätzen, welche Aspekte des Entscheidungsprozesses hervorstechen, und nach Anwendung auf mehrere Einzelfälle lassen verschiedene Verfahren miteinander vergleichen.

6.2 Akteure

A1 Wettbewerb: Wer ist der Betreiber des Algorithmus? Wie viel politische und ökonomische Macht hat er? Besteht effektiver Wettbewerb zwischen mehreren Betreibern (Oligopole, Monopole, Kartelle, marktbeherrschende Stellung, hoheitliche Macht)?

Beispiel Grenzkontrollen: Auch wenn private Firmen die Software für Visabewertungen produzieren und/oder betreiben, wird sie im Rahmen hoheitlicher Machtausübung angewendet. Der staatlichen Macht bei der Visavergabe kann sich niemand entziehen, insofern würde auch der private Betreiber mit staatlichem Auftrag als Monopol gewertet. Die reine Rechtsform des Betreibers ist also weniger relevant für die Wirkung auf Teilhabe.

- Ist der Wettbewerb zwischen mehreren Betreibern schwach, ist die Wirkung auf Teilhabe erhöht (TWP = +1).
- Wenn der Betreiber des Algorithmus ein Oligopolist oder Monopolist ist, dann ist jeder mögliche Eingriff in Teilhabe voraussichtlich besonders hoch (TWP = +2).

A2 Abhängigkeit: Ist es ein Problem, wenn der Prozess bzw. das produzierte Gut wegfällt oder der Zugang dazu verwehrt wird? Kann das Gut unmittelbar ersetzt werden? Gibt es alternative Verfahren/Angebote? Wie hoch sind die Kosten für den Umstieg auf ein alternatives Verfahren? Welche Organisationen oder Personengruppen sind strukturell abhängig von dem Verfahren? Erzeugen Aufbau und Funktionsweise des Systems eine informelle Abhängigkeit? Welche technischen oder sozialen Lock-in-Effekte gibt es?

Beispiel Crowdwork: Auf dem Markt für Crowdworking-Plattformen herrscht Wettbewerb. Mehrere Plattformanbieter konkurrieren sowohl um Auftraggeberinnen als auch Crowdworkerinnen. Eine Abhängigkeit der Beschäftigten könnte trotzdem entstehen, wenn die Auftragshistorie und Reputation an die spezifischen algorithmischen Bewertungsverfahren der Plattform gekoppelt sind. Ein ähnlicher Lock-in-Effekt kann etwa auch bei anderen Bewertungssystemen oder sozialen Medien auftreten.

- Wenn ein System Abhängigkeiten erzeugt, dann ist die potenzielle Wirkung auf Teilhabe erhöht (TWP = +1).
- Wenn die Abhängigkeit sehr groß ist, weil die Umstiegskosten prohibitiv hoch sind, dann ist jeder mögliche Eingriff in Teilhabe besonders stark (TWP = +2).

6.3 Soziale Einbettung

SE1 Selbstbestimmung: Inwiefern dient das Verfahren (nur) zur Vorbereitung einer Entscheidung? Bleibt die Nutzerin selbstbestimmt? Entscheidet das System (de facto oder offiziell) selbstständig? Wie viel Freiraum zur Änderung und Kontrolle einer algorithmischen Entscheidung hat die Anwenderin? Wie wirken sich Zeitdruck und operative Verwendung des Verfahrens auf die Selbstbestimmung der Anwender aus?

Beispiel Inhaltskontrolle: Die Gestaltung der Software zur Kontrolle von potenziell rechtswidrigen Inhalten kann großen Einfluss auf die Eigenständigkeit der Anwenderin haben. Durch Zeitdruck (z. B. hohe Taktung) und technische Einschränkung kann der Spielraum für menschliche Abwägungen de facto stark reduziert sein, auch wenn offiziell ein Mensch die finale Entscheidung trifft.

- Wenn die Anwenderin selbstbestimmt bleibt (SE1), dann wird dadurch der Eingriff in Teilhabe nicht beeinflusst (TWP=0).
- Wenn Entscheidungen nur formell von Menschen getroffen werden, de facto aber (weitgehend) automatisiert sind, dann ist die potenzielle Wirkung auf Teilhabe erhöht (TWP=+1 oder +2).

SE2 Anpassung: Wie passen sich Menschen an den Prozess an? Wie wirkt das Entscheidungssystem in der Praxis? Welche Wechselwirkungen zwischen Mensch und Computer verändern unter Umständen das Ergebnis? Welche effektive, soziale Autorität hat das algorithmische Verfahren? Welche (soziale) Bedeutung wird dem Prozess zugeschrieben? Hat er Diskurshoheit? Wie viel Macht steckt in den Ergebnissen des Prozesses?

Beispiel Trending Topics: Die sogenannten „Trending Topics“ auf Twitter werden nicht selten als soziale „Realität“ angesehen. Obwohl nicht bekannt ist, wie das Ranking zustande kommt, hat es Einfluss auf den Diskurs und unterschiedliche Akteure passen sich an die Logik des Trending-Topics-Algorithmus an.

- Wenn das Entscheidungsverfahren soziale Autorität besitzt, dann ist die potenzielle Wirkung auf Teilhabe erhöht (TWP = +1).
- Wenn Nutzerinnen ihr Verhalten an das Entscheidungssystem stark anpassen, dann ist die mögliche Wirkung auf Teilhabe besonders stark (TWP = +2).

6.4 Konsequenzen

K1 Reichweite: Wie viele Personen sind von dem Entscheidungsprozess betroffen (z. B. Nutzerzahlen)? Ist die Reichweite des Verfahrens bekannt bzw. begrenzt? Wie groß ist sie?

Beispiel Bewerberauswahl: Wenn nur ein Unternehmen eine Software zur Auswahl von Bewerberinnen einsetzt, ist die Reichweite relativ gering. Wird sie jedoch als ein zentrales Entscheidungssystem in vielen Unternehmen eingesetzt, steigt die Teilhabewirkung. Etwa wenn die Entscheidungslogik eines Systems in einer bestimmten Region oder Branche die Personalauswahl dominiert.

- Je größer die Reichweite des Entscheidungsprozesses, desto höher der potenzielle Eingriff in Teilhabe (TWP = +1 oder +2).

K2 Systemveränderung: Hebelt der Entscheidungsprozess solidarische Prinzipien aus? Ist eine Individualisierung systemverändernd? Hat das Verfahren eine potenzielle, eventuell nicht beabsichtigte Transformation des (sozialen) Wirkungsumfelds zur Folge?

Beispiel Krankenkasse: Wenn die Kosten für öffentliche Krankenversicherung durch Algorithmen mitentschieden und nach einer neuen Logik individualisiert werden, dann wird ein bisher auf Solidarität beruhendes System verändert. Anstatt alle entsprechend ihrer Leistungsfähigkeit einzahlen zu lassen, werden mindestens Teile des Beitrags auf Basis individueller Faktoren berechnet.

- Wenn der Prozess Prinzipien seines (sozialen) Wirkungsumfelds verändert, dann ist der Eingriff in Teilhabe potenziell erhöht (TWP = +1 oder +2).

K3 Diskriminierung: Wie stark könnten Menschen durch das Entscheidungsverfahren benachteiligt werden? Lässt sich für den Output ein diskriminierendes Muster, zum Beispiel aufgrund von Rasse, Herkunft, Geschlecht, Religion oder Weltanschauung, körperlicher Beeinträchtigung, Alter oder sexueller Identität feststellen bzw. erwarten?

Beispiel automatisierte Stellenanzeigen: Eine Untersuchung von Googles Werbeanzeigen für Jobs zeigte, dass als Männer kategorisierte Nutzer eher hoch bezahlte Stellen angezeigt bekommen als Frauen. Hier wirkt das Verfahren also auf die Teilhabe am Arbeitsmarkt.

- Wenn das Verfahren Menschen oder Gruppen strukturell benachteiligt oder benachteiligen könnte, dann ist die potenzielle Wirkung auf Teilhabe erhöht (TWP = +1 oder +2).

Tabelle 1: Kriterienübersicht zur Einschätzung potenzieller Wirkung auf Teilhabe (Quelle: eigene Darstellung)

Dimension	Kriterium	Erklärung
Akteure	A1 Wettbewerb	Wer ist der Betreiber des Algorithmus? Wie viel politische und ökonomische Macht hat er? Besteht effektiver Wettbewerb zwischen mehreren Betreibern (Oligopole, Monopole, Kartelle, marktbeherrschende Stellung, hoheitliche Macht)?
	A2 Abhängigkeit	Ist es ein Problem, wenn der Prozess bzw. das produzierte Gut wegfällt oder der Zugang dazu verwehrt wird? Kann das Gut unmittelbar ersetzt werden? Gibt es alternative Verfahren/Angebote? Wie hoch sind die Kosten für den Umstieg auf ein alternatives Verfahren?
Soziale Einbettung	SE1 Selbstbestimmung	Dient das Verfahren (nur) zur Vorbereitung einer Entscheidung? Bleibt die Nutzerin selbstbestimmt? Entscheidet das System (de facto oder offiziell) selbstständig? Wie viel Freiraum zur Änderung und Kontrolle einer algorithmischen Entscheidung hat die Anwenderin?
	SE2 Anpassung	Wie passen sich Menschen an den Prozess an? Wie wirkt das Entscheidungssystem in der Praxis? Welche Wechselwirkungen zwischen Menschen und Computer verändern unter Umständen das Ergebnis?
Konsequenzen	K1 Reichweite	Wie viele Personen sind von dem Entscheidungsprozess betroffen (z. B. Nutzerzahlen)? Ist die Reichweite des Verfahrens bekannt bzw. begrenzt? Wie groß ist sie?
	K2 Systemveränderung	Hebelt der Entscheidungsprozess solidarische Prinzipien aus? Ist eine Individualisierung systemverändernd?
	K3 Diskriminierung	Wie stark könnten Menschen durch das Entscheidungsverfahren benachteiligt werden? Lässt sich für den Output ein diskriminierendes Muster feststellen bzw. erwarten?

6.5 Zusammenfassung: So lässt sich Teilhabewirkungspotenzial bestimmen

Der Grad des Teilhabewirkungspotenzials (TWP) eines beliebigen neuen Prozesses algorithmischer Entscheidungsfindung lässt sich daher in der Gesamtschau wie folgt darstellen:

$$TWP = VA1 * VA2 * (A1 + A2 + SE1 + SE2 + K1 + K2 + K3)$$

VA1 und VA2 stehen für die Vorauswahl, die in Kapitel 4 ausgeführt wurde. Diese beiden Filterkriterien „Digitale Entscheidungssysteme“ (VA1) und „Bewertung von Menschen“ (VA2) können entweder den Wert 0 (nicht erfüllt) oder 1 (erfüllt) annehmen. Der Grad des Teilhabewirkungspotenzials (TWP) ergibt sich aus dem Produkt der beiden Vorauswahlkriterien und der Summe der Teilhabewirkungskriterien. Wenn einer der Faktoren den Wert 0 hat, schlägt dies sich dies unmittelbar auch im Gesamtergebnis (TWP = 0) nieder. Andernfalls entspricht die konkrete Höhe des TWP-Werts der Summe der Prüfkriterien (A1 + A2 + SE1 + SE2 + K1 + K2 + K3).

Das Gesamtergebnis (TWP) besagt, wie stark sich ein algorithmischer Entscheidungsprozess auf die Realisierung von Teilhabe auswirken kann. Die Skala reicht von 0 bis 14. Mit der Prüfstruktur lässt sich das Wirkungspotenzial unterschiedlicher Verfahren vergleichen. Diese relativen Aussagen lassen sich zum Beispiel nutzen, um Verfahren für die nähere Untersuchung zu priorisieren. Ein Fallbeispiel mit einem TWP von 11 ist dem mit einem TWP von 3 für eine tiefer gehende Analyse der Teilhabewirkung vorzuziehen. Das Instrument ist nicht für absolute Aussagen entwickelt. Es ermöglicht keine absoluten Aussagen, wie zum Beispiel: „Ab einem TWP von 6 sind Verfahren besonders teilhaberelevant.“

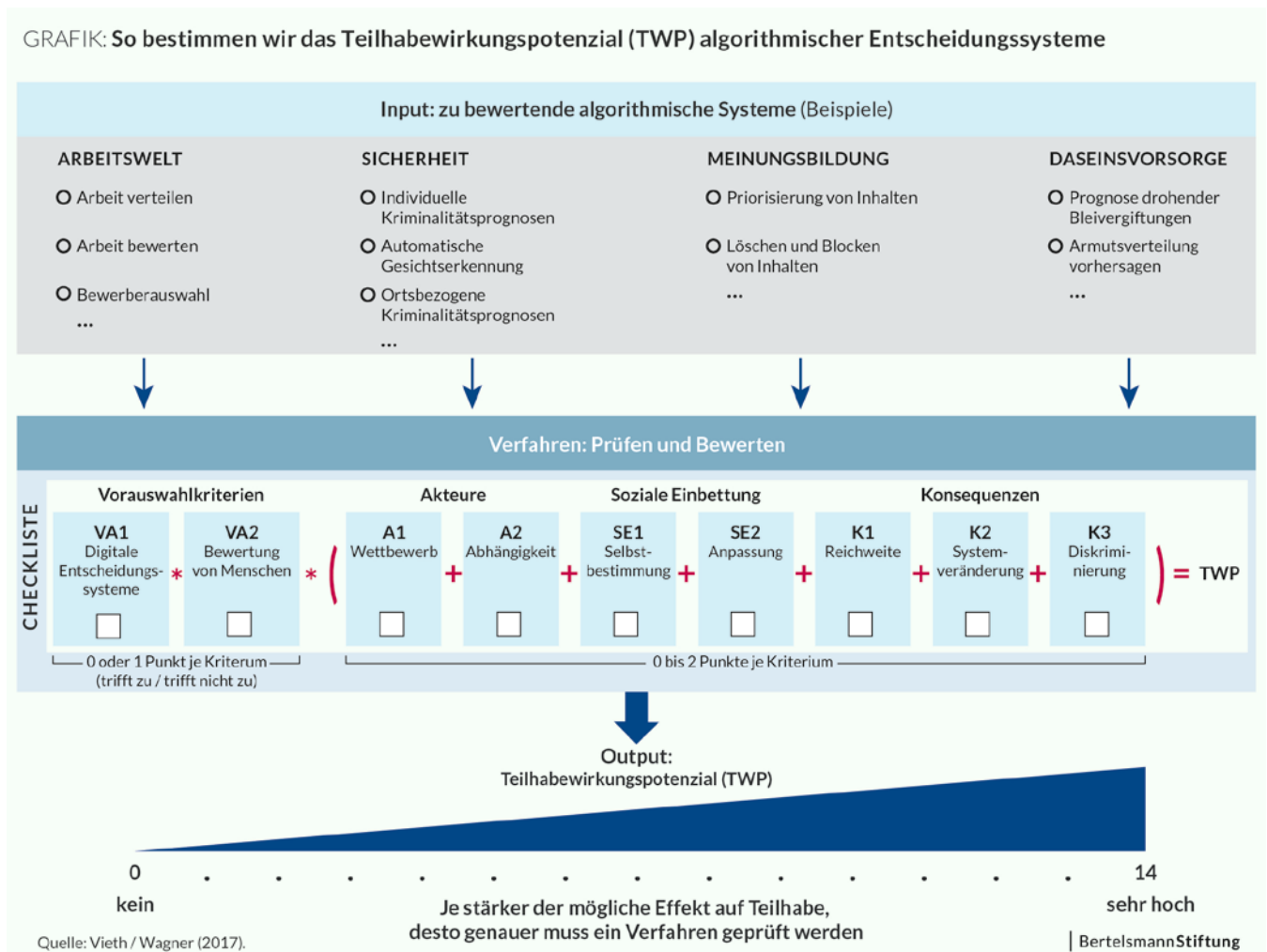


Abbildung 1: So lässt sich das Wirkungspotenzial von algorithmischen Prozessen auf Teilhabe berechnen

Aus den oben stehenden Kriterien lässt sich ein schnelles und relativ einfaches Prüfverfahren entwickeln, in dem Unternehmen und Prüferinnen auf zwei bis drei Seiten ihre algorithmischen Entscheidungsverfahren nach bestimmten Kriterien durchchecken.³ Derartige Kriterien könnten Akteurinnen nutzen, um Prozesse algorithmischer Entscheidungsfindung auf Teilhaberelevanz zu prüfen. Insofern dient das Verfahren als schneller und einfacher Einstieg für das Verständnis eines algorithmischen Entscheidungsprozesses. Eine gründliche Prüfung kann es nicht ersetzen. Es ist denkbar, dass Akteurinnen zur Analyse selbst Schwellenwerte der Teilhaberwirkung definieren. Dafür muss zunächst das zugrunde liegende Teilhaberverständnis transparent definiert und operationalisiert werden.

6.6 Beispielrechnungen

6.6.1 Teilhabewirkungspotenzial von Rechtschreibprüfung

Die Kontrolle von Grammatik und Orthographie in Textverarbeitungsprogrammen basiert auf einem digitalen Entscheidungssystem ($VA1 = 1$). Es wird automatisiert über die korrekte Schreibweise und Satzbau entschieden. Dabei kommt es jedoch nicht zur Bewertung von menschlichen Eigenschaften ($VA2 = 0$). Damit ergibt sich automatisch ein Teilhabewirkungspotenzial von 0, von einer weiteren Prüfung kann daher abgesehen werden.

$$TWP = 1 * 0 * (x)$$

$$TWP = 0$$

6.6.2 Teilhabewirkungspotenzial von Predictive Policing

Welches Teilhabewirkungspotenzial hat Predictive Policing (siehe Kapitel 4.2.1)? Es handelt sich um ein digitales Entscheidungssystem, das menschliche Eigenschaften bewertet ($VA1 = 1$; $VA2 = 1$). Auf Akteursebene lässt sich festhalten, dass die Polizei durch ihr Gewaltmonopol prinzipiell große Wirkung auf Teilhabe entfalten kann ($A1 = +2$). Auch die Abhängigkeit vom „Gut“ Sicherheit ist für alle Menschen hoch ($A2 = +2$). Die soziale Einbettung der Predictive Policing Software ist im Schnellverfahren nur schwer zu bewerten. Die Selbstbestimmtheit der Polizistinnen bleibt, soweit das bisher erkennbar ist, zumindest teilweise erhalten ($SE1 = 0,5$), doch ist soziale Autorität des Outputs infolge der Anpassung der Menschen an diese zu erwarten ($SE2 = +1$). Die Konsequenzen, die bisher auf wenige Pilotprojekte in Städten beschränkt sind, haben zunächst nur eine mittlere Reichweite ($K1 = +1$). Der Einsatz der Software hat das Potenzial, die Polizeiarbeit strukturell zu verändern ($K2 = +2$). Das Potenzial, Diskriminierung zu senken oder zu verfestigen, ist hoch ($K3 = +2$). In der Gesamtschau ergibt sich daraus:

$$TWP = 1 * 1 * (2 + 2 + 0,5 + 1 + 1 + 2 + 2)$$

$$TWP = 10,5$$

Mit 10,5 von maximal 14 Teilhabewirkungspunkten hat Predictive Policing eine sehr hohe zu erwartende Teilhabewirkung. In diesem Fall gilt es also, die Chancen und Risiken besonders aufmerksam und detailliert zu untersuchen.

6.6.3 Teilhabewirkungspotenzial von Plattformarbeit

Die Bewertung der Teilhaberelevanz von Plattformarbeit im Allgemeinen (siehe Kapitel 4.1.1) macht es schwer, die Akteurseigenschaften zu klassifizieren. Zunächst ist festzustellen, dass die Vorauswahlkriterien beide erfüllt sind ($VA1 = 1$; $VA2 = 1$). Die Betreiber sind bisher recht heterogen und es herrscht reger Wettbewerb zwischen ihnen ($A1 = 0$). Trotzdem können Abhängigkeiten durch plattformgebundene Reputations- und Auftragshistorien

³ Vergleiche zum Beispiel das „Quick Check Sheet“⁴³ des Danish Institute for Human Rights. Siehe https://hrca2.humanrightsbusiness.org/docs/file/HRCA%20Quick%20Check_English.pdf.

entstehen ($A2 = +1$). Die Souveränität der Nutzerinnen der Plattformsoftware wird beeinflusst ($SE1 = +1,5$), ebenso sind starke Anpassungen an das algorithmische Arbeitsumfeld zu erwarten ($SE2 = +2$). Die Reichweite ist branchenübergreifend im Schnellverfahren nicht eindeutig zu bewerten ($K1 = 0$). Systemverändernde Folgen sind aber wahrscheinlich. Zum Beispiel könnte sich die Plattformarbeit auf Teilhabebereiche wie Mitbestimmung auswirken ($K2 = +2$). Erste Diskriminierungsfolgen für Kundinnen sind bekannt, daher ist von einer erhöhten Wirkungspotenzial auszugehen ($K3 = 1$).

$$TWP = 1 * 1 * (0 + 1 + 1,5 + 2 + 0 + 2 + 1)$$

$$TWP = 7,5$$

Mit 7,5 von 14 Punkten ist die abstrakte Teilhabewirkung von Plattformarbeit in etwa im Mittelfeld der Skala. Eine detaillierte, branchenspezifische Untersuchung bietet sich an.

6.6.4 Teilhabewirkungspotenzial von Telematik-Versicherungen

Welches Teilhabewirkungspotenzial hat der Telematik-Tarif für die Autoversicherung (siehe Kapitel 4.5.2)? Auch hier sind die Vorauswahlkriterien für ein digitales Entscheidungssystem, das Menschen bewertet, klar erfüllt ($VA1 = 1$; $VA2 = 1$). Die Betreiberinnen stehen aktuell im Wettbewerb und es sind bisher keine starken Abhängigkeitseffekte zu erwarten ($A1 = 0$). Potenziell könnte es aber bei spezifischen Versicherungsunternehmen zu Lock-in-Effekten kommen, zum Beispiel wenn eine digitale Fahrhistorie nicht zu einem anderen Unternehmen übertragbar ist ($A2 = +0,5$). Die Selbstbestimmung der Kundinnen scheint zunächst nicht berührt ($SE1 = 0$). Anpassungseffekte, etwa ein vorsichtigerer Fahrstil, sind erwünscht und potenziell vorhanden ($SE2 = +1$). Die Reichweite ist noch stark begrenzt, da sich das Angebot nur auf Fahranfängerinnen bezieht und die Teilnahme prinzipiell freiwillig ist ($K1 = 0$). Der Telematik-Tarif kann sein Wirkungsumfeld potenziell verändern, da die individualisierte Datenauswertung zum Beispiel auch bei Unfällen hinzugezogen werden könnte ($K2 = +0,5$). Eine potenziell altersdiskriminierende Wirkung entsteht durch den Fokus auf Fahranfängerinnen ($K3 = +0,5$).

$$TWP=1 * 1 * (0 + 0,5 + 0 + 1 + 0 + 0,5 + 0,5)$$

$$TWP = 2,5$$

Nur 2,5 Teilhabewirkungspunkte könnten es erlauben, dieses Fallbeispiel aus einer genaueren Untersuchung auszuschließen.

7 Fazit und Ausblick

Zwei komplexe Themenfelder wie Teilhabe und algorithmische Entscheidungsfindung miteinander zu verknüpfen, ist herausfordernd, eröffnet aber insbesondere durch die Kombination qualitativer und quantitativer Ansätze spannende Perspektiven. Die Kriterien für unseren Klassifizierungsvorschlag stellen die Akteure, die Einbindung in den sozialen Kontext und die potenziellen Konsequenzen für Grundrechte in den Mittelpunkt. Natürlich sind auch andere Schwerpunkte denkbar. Doch für einen einfachen Einstieg in den Themenkomplex bietet eine Beschränkung wie bei unserer Auswahl einige Vorteile. Viele der Prüffragen sind mit relativ geringem Rechercheaufwand zu beantworten. Die Klassifizierung zeichnet sich durch ihre rasche Anwendbarkeit aus, zudem sind die Ergebnisse leicht vergleichbar. Die Kriterien sind auf Komplexitätsreduktion und Flexibilität ausgerichtet. Die grundlegende Struktur des Prüfverfahrens ist verhältnismäßig leicht anpassbar. Wenn wesentliche Faktoren fehlen, kann diese ohne großen Aufwand ergänzen. Außerdem eignen sich die vereinfachten Prüfkriterien gut für eine Visualisierung der Teilhaberelevanz, etwa in Form von Balkendiagrammen oder Scoring-Tabellen.

Auch ein Vergleich der verschiedenen Ergebnisse, wenn mehreren Personen das Teilhabewirkungspotenzial anhand der Kriterien bewerten und berechnen, könnte aufschlussreich sein: Wer beantwortet welche Fragen wie? Welche subjektiven Vorstellungen gibt es? Wie verändert sich das Scoring, wenn Menschen mit unterschiedlichen Erfahrungshintergründen die Kriterien beurteilen? Das Teilhabewirkungs-Scoring eignet sich damit auch als Methode für Workshops und als Diskussionsgrundlage für strategische Beratungen.

Die Debatte über Algorithmen und ihren Einfluss auf Teilhabe steht erst am Anfang. Die große Vielfalt an Anwendungsbereichen und Funktionen macht es schwer, eine kohärente Übersicht zu erstellen. Als Werkzeuge für die Entscheidungsfindung sind Algorithmen nicht per se gut oder schlecht. Es hängt von Zielsetzung und Umsetzung ab, ob ihr Wirkungspotenzial auf Teilhabechancen dazu genutzt wird, Chancengerechtigkeit und Grundrechte zu stärken und zu *besseren* Entscheidungsprozessen zu kommen.

Eine solche Verbesserung gegenüber menschlichen Entscheidungen ist oft möglich. In vielen Kontexten kann ein Algorithmus inkonsistenten und voreingenommenen menschlichen Entscheidungen überlegen sein (Kahneman 2012: 233; Meehl 1954). Doch die teilhabefördernde Zielsetzung erfolgt nicht durch die Technologie als solche, sondern muss gesellschaftlich ausgehandelt und gestaltet werden. Daher ist eine kritische Betrachtung und Reflexion teilhaberelevanter Algorithmen unabdingbar, sie kann nicht automatisiert werden.

Entscheidend ist dabei, Ziele explizit zu benennen. Denn wer das nicht tut, macht implizite Annahmen und läuft Gefahr, Teilhabe zu beschneiden. Annahmen und Erwartungen müssen geklärt werden, um positive Teilhabeeffekte zu erreichen. Hier können die in dieser Expertise entwickelten Klassifizierungskriterien einen Ansatz für Debatten über gesellschaftliche Ziele und politische Maßnahmen voranbringen. Gerade dann, wenn unterschiedliche Organisationen oder Personen einzelne Kriterien unterschiedlich bewerten. Denn erst im Dissens wird das Spektrum verschiedener Handlungsoptionen sichtbar.

8 Literatur

- Agrawal, Ajay, John Horton, Nicola Lacetera und Elizabeth Lyons (2015). „Digitization and the contract Labor Market: A Research Agenda“. *Economic Analysis of Digital Economy*. Hrsg. Avi Goldfarb, Shane M. Greenstein und Catherine E. Tucker. Chicago IL: University of Chicago Press. 219–250.
<http://www.nber.org/chapters/c12988.pdf> (Download 8.5.2017).
- Aneesh, Aneesh (2009). „Global Labor: Algoratic Modes of Organization“. *Sociological Theory* (27) 4. 347–370.
<https://doi.org/10.1111/j.1467-9558.2009.01352.x> (Download 8.5.2017).
- Angwin, Julia, Lauren Kirchner, Jeff Larson und Surya Mattu (2016). „Machine Bias: There’s Software Used Across the Country to Predict Future Criminals. And it’s Biased Against Blacks“. 23.5.2015.
<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> (Download 11.12.2016).
- Arthurs, Harry (2011). „Labour Law After Labour“. *The Idea of Labour Law*. Hrsg. Guy Davidov und Brian Langille. Oxford, New York: Oxford University Press. 13–29.
- Barry-Jester, Anna M., Ben Casselman und Donna Goldstein (2015). „The New Science of Sentencing“. *The Marshall Project*. 4.4.2015. <https://www.themarshallproject.org/2015/08/04/the-new-science-of-sentencing#.xXE6R5rD> (Download 24.4.2017).
- Bauman, Zygmunt, Didier Bigo, Paulo Esteves, Elspeth Guild, Vivienne Jabri, David Lyon und R. B. J. Walker (2014). „After Snowden: Rethinking the Impact of Surveillance“. *International Political Sociology* (8) 2. 121–144. <https://doi.org/10.1111/ips.12048> (Download 8.5.2017).
- Baumgartner, Fabian (2015). „Deutliche Zunahme im Jahr 2015: Wieder mehr Einbrüche in der Stadt Zürich“. *Neue Zürcher Zeitung* 27.10.2015. <http://www.nzz.ch/zuerich/stadt-zuerich/wieder-mehr-einbrueche-in-der-stadt-zuerich-1.18636278> (Download 24.4.2017).
- Beirat Integration (2013). „Soziale Teilhabe‘ Handlungsempfehlungen des Beirats der Integrationsbeauftragten“. Die Beauftragte der Bundesregierung für Migration, Flüchtlinge und Integration.
<http://www.bagiv.de/pdf/soziale-teilhabe-empfehlungen-beirat.pdf> (Download 8.5.2017).
- Bertelsmann Stiftung (2011). *Soziale Gerechtigkeit in der OECD – Wo steht Deutschland? Sustainable Governance Indicators 2011*. Gütersloh. http://news.sgi-network.org/uploads/tx_amsgistudies/SGI11_Social_Justice_DE.pdf (Download 8.5.2017).
- Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Carlow Marlow, Jaime E. Settle und James H. Fowler (2012). „A 61-million-person experiment in social influence and political mobilization“. *Nature* (489) 13.9.2012. 295–298. <https://doi.org/10.1038/nature11421> (Download 8.5.2017).

- Brey, Philip (2005). „Artifacts as social agents“. *Inside the Politics of Technology: Agency and Normativity in the Co-Production of Technology and Society*. Hrsg. Hans Harbers. Amsterdam: Amsterdam University Press. 61–84.
- Brown, Ian, und Josh Cowls (2015). *Check the Web – Assessing the Ethics and Politics of Policing the Internet*. Oxford: Oxford Internet Institute und VOX-Pol Network of Excellence. http://voxpol.eu/wp-content/uploads/2015/11/VOX-Pol_Ethics_Politics_PUBLISHED.pdf (Download 8.5.2017).
- Brühl, Jannis (2014). „Ermitteln mit ‚Predictive Policing‘-Algorithmen: Polizei-Software soll Verbrechen voraussagen“. *Süddeutsche Zeitung* 12.9.2014. <http://www.sueddeutsche.de/digital/ermitteln-mit-predictive-policing-algorithmen-polizei-software-soll-die-zukunft-voraussagen-1.2121942> (Download 24.4.2017).
- Bucher, Taina (2016). „The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms“. *Information, Communication & Society* (20) 1. 1–15. <https://doi.org/10.1080/1369118X.2016.1154086> (Download 8.5.2017).
- Bundesgesetzblatt (1966). „Internationaler Pakt über wirtschaftliche, soziale und kulturelle Rechte“. Bundesgesetzblatt (BGBl) 1976 II, 428. 19.12.1966. http://www.institut-fuer-menschenrechte.de/fileadmin/user_upload/PDF-Dateien/Pakte_Konventionen/ICESCR/icescr_de.pdf (Download 8.5.2017).
- Bundesministerium für Arbeit und Soziales (2015). *Arbeit 4.0 (Grünbuch)*. Berlin. https://www.bmas.de/Shared-Docs/Downloads/DE/PDF-Publikationen-DinA4/gruenbuch-arbeiten-vier-null.pdf?_blob=publicationFile (Download 8.5.2017).
- Choudary, Sangeel Paul, Marshall W. van Alstyne und Geoffrey G. Parker (2016). *Platform Revolution: How Networked Markets Are Transforming the Economy – And How to Make Them Work for You* (1. Auflage). New York NY: W. W. Norton & Company.
- Cormen, Thomas H., Clifford Stein, Charles E. Leiserson und Robert L. Rivest (und 2 mehr) (2001). *Introduction to Algorithms*. Cambridge MA: MIT Press.
- Datta, Amit, Michael Carl Tschantz und Anupam Datta (2015). „Automated Experiments on Ad Privacy Settings“. *Proceedings on Privacy Enhancing Technologies* 1. 92–112. <https://doi.org/10.1515/popets-2015-0007> (Download 8.5.2017).
- Davey, Monica (2016). „Chicago Police Try to Predict Who May Shoot or Be Shot“. *The New York Times* 23.5.2016. <http://www.nytimes.com/2016/05/24/us/armed-with-data-chicago-police-try-to-predict-who-may-shoot-or-be-shot.html> (Download 24.4.2017).

- De Stefano, Valerio (2015). „The Rise of the ‚Just-in-Time Workforce‘: On-Demand Work, Crowd Work and Labour Protection in the ‚Gig-Economy‘“. *Comparative Labor Law & Policy Journal* (37) 3. 471–504.
- Denardis, Laura (2012). „Hidden Levers of Internet Control“. *Information, Communication & Society* (September). 37–41.
- Diakopoulos, Nicholas (2015). „Algorithmic Accountability“. *Digital Journalism* (3) 3. 398–415.
<https://doi.org/10.1080/21670811.2014.976411> (Download 8.5.2017).
- Diakopoulos, Nicholas (2016). „Accountability in Algorithmic Decision Making“. *Communication of the ACM* (59) 2. 56–62. <https://doi.org/10.1145/2844110> (Download 8.5.2017).
- Epstein, Robert, und Ronald E. Robertson (2015). „The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections“. *Proceedings of the National Academy of Sciences* (112) 33. E4512–E4521. <https://doi.org/10.1073/pnas.1419828112> (Download 8.5.2017).
- Eubanks, Virginia (2013). „Caseworkers vs. Computers“. *Poptech* 11.12.2013. <https://virginiaeubanks.wordpress.com/2013/12/11/caseworkers-vs-computers/> (Download 8.5.2017).
- Europol (2015). „Europol's Internet Referral Unit to Combat Terrorist and Violent Extremist Propaganda“ (Press Release). 1.7.2015. <https://www.europol.europa.eu/newsroom/news/europol%E2%80%99s-internet-referral-unit-to-combat-terrorist-and-violent-extremist-propaganda> (Download 8.5.2017).
- Ferguson, Andrew Guthrie (2012). „Predictive Policing and Reasonable Suspicion“. *Emory Law Journal* 259. <https://papers.ssrn.com/abstract=2050001> (Download 24.4.2017).
- Freedland, Mark R., und Nicola Kountouris (2011). „The Legal Characterization of Personal Work Relations and the Idea of Labour Law“. *The Idea of Labour Law*. Hrsg. Guy Davidov und Brian Langille. Oxford, New York: Oxford University Press. 190–208.
- Gillespie, Tarleton (2014). „The Relevance of Algorithms“. *Essays on Communication, Materiality, and Society*. Hrsg. Tarleton Gillespie, Pablo Boczkowski und Kirsten Foot. Cambridge MA: MIT Press. 167–194.
- Gluba, Alexander (2014). *Predictive Policing – eine Bestandsaufnahme*. LKA Niedersachsen: Hannover.
https://netzpolitik.org/wp-upload/LKA_NRW_Predictive_Policing.pdf (Download 24.4.2017).
- Hannah-Moffat, Kelly, Paula Maurutto und Sarah Turnbull (2009). „Negotiated Risk: Actuarial Illusions and Discretion in Probation“. *Canadian Journal of Law and Society* (24) 3. 391–409.
<https://doi.org/10.1017/S0829320100010097> (Download 24.4.2017).
- Hawthorne, Michael (2015). „Could Chicago prevent childhood lead poisoning before it happens?“. *Chicago Tribune* 16.7.2015. <http://www.chicagotribune.com/news/ct-lead-poisoning-solutions-20150707-story.html> (Download 3.1.2017).

- Helberger, Natali, und Damian Trilling (2016). „Facebook is a news editor: the real issues to be concerned about“. 26.5.2016. <http://blogs.lse.ac.uk/mediapolicyproject/2016/05/26/facebook-is-a-news-editor-the-real-issues-to-be-concerned-about/> (Download 8.5.2017).
- Hensel, Isabell, Jochen Koch, Eva Kocher und Anna Schwarz (2016). „Crowdworking als Phänomen der Koordination digitaler Erwerbsarbeit – Eine inter-disziplinäre Perspektive“. *Industrielle Beziehungen* (23) 2. 162–186.
- Hill, Kashmir (2012). „How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did“. 16.2.2016. <http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/> (Download 8.5.2017).
- Irani, Lilly (2015). „Difference and Dependence among Digital Workers: The Case of Amazon Mechanical Turk“. *South Atlantic Quarterly* (114) 1. 225–234. <https://doi.org/10.1215/00382876-2831665> (Download 8.5.2017).
- Kahneman, Daniel (2012). *Thinking, Fast and Slow*. London u. a.: Penguin.
- Kocher, Eva, und Isabell Hensel (2016). „Herausforderungen des Arbeitsrechts durch digitale Plattformen – ein neuer Koordinationsmodus von Erwerbsarbeit“. *Neue Zeitschrift für Arbeitsrecht NZA* (33) 16. 984–990.
- Lee, Min Kyung, Dabiel Kusbit, Evan Metsky und Laura Dabbish (2015). „Working with Machines: The Impact of Algorithmic and Data-Driven Management on Human Workers“. ACM Press. 1603–1612). <https://doi.org/10.1145/2702123.2702548> (Download 8.5.2017).
- Leimeister, Jan Marco, David Durward und Shkodran Zogaj (2016). *Crowd Worker in Deutschland eine empirische Studie zum Arbeitsumfeld auf externen Crowdsourcing-Plattformen*. Düsseldorf: Hans-Böckler-Stiftung. <http://hdl.handle.net/10419/146407> (Download 8.5.2017).
- Lingemann, Stefan, und Jörn Christoph Otte (2015). „Arbeitsrechtliche Fragen der ‚economy on demand‘“. *Neue Zeitschrift für Arbeitsrecht NZA*. 1042–1047.
- Lischka, Konrad, und Anita Klingel (2017). *Wenn Maschinen Menschen bewerten*. Gütersloh: Bertelsmann Stiftung. <https://doi.org/10.11586/2017025> (Download 29.5.2017)
- Lyon, David (2003). *Surveillance as social sorting: privacy, risk, and digital discrimination*. New York NY: Routledge. <http://public.eblib.com/choice/publicfullrecord.aspx?p=240591> (Download 8.5.2017).
- Mastrobuoni, Giovanni (2015). „Crime is terribly revealing: Information technology and police productivity“. Unpublished Paper. http://cep.lse.ac.uk/conference_papers/01_10_2015/mastrobuoni.pdf (Download 24.4.2017).

- Meehl, Paul E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis MN: University of Minnesota Press. <https://doi.org/10.1037/11281-000> (Download 8.5.2017).
- Meyer, Thomas (2016). „Gleichheit – warum, von was und wie viel?“ *Neue Gesellschaft/Frankfurter Hefte* 11. 42–46.
- Mohler, George O., Martin B. Short, Sean Malinowski, Mark Johnson, George E. Tita, Andrea L. Bertozzi und P. Jeff Brantingham (2015). „Randomized controlled field trials of predictive policing“. *Journal of the American Statistical Association* (110) 512. 1399–1411. <https://doi.org/http://dx.doi.org/10.1080/01621459.2015.1077710> (Download 24.4.2017).
- Niklas, Jędrzej (2017). „Can an algorithm hurt? Polish experiences with profiling of the unemployed“. März 2017. <https://cihr.eu/can-an-algorithm-hurt/> (Download 8.5.2017).
- Niklas, Jędrzej, Karolina Sztandar-Sztanderska und Katarzyna Szymielewicz (2015). *Profiling the Unemployed in Poland: Social and Political Implications of Algorithmic Decision Making*. Warsaw: Panoptikon Foundation. <http://panoptikon.org/biblio/profiling-unemployed-poland-social-and-political-implications-algorithmic-decision-making> (Download 8.5.2017).
- Northpointe (2015). „Practitioners Guide to COMPAS Core“. <http://images.google.de/imgres> (Download 24.4.2017).
- O’Callaghan, Derek, Derek Green, Maura Conway, Joe Carthy und Pádraig Cunningham (2016). „Down the (White) Rabbit Hole: The Extreme Right and Online Recommender Systems“. http://doras.dcu.ie/20280/1/Down_the_Rabbit_Hole_DORAS_Version.pdf (Download 8.5.2017).
- Pariser, Eli (2011). *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. London. Penguin Publishing Group.
- Pasquale, Frank A. (2015). *Privacy, Autonomy, and Internet Platforms* (SSRN Scholarly Paper No. ID 2737413). Rochester NY: Social Science Research Network. <http://papers.ssrn.com/abstract=2737413> (Download 8.5.2017).
- Pilpul (2016). „Wo Predictive Policing eingesetzt wird“. Dezember 2016. <https://blog.pilpul.me/wo-predictive-policing-eingesetzt-wird/> (Download 24.4.2017).
- Potash, Eric, Joe Brew, Alexander Loewi, Subhanrata Majumdar, Andrew Reece, Joe Walsh, Eric Rozier, Emile Jorgenson, Read Mansour und Rayid Ghani (2015). „Predictive Modeling for Public Health: Preventing Childhood Lead Poisoning“. *KDD ’15 Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2039–2047. <https://doi.org/10.1145/2783258.2788629> (Download 24.4.2017).

- Rammert, Werner (2006). „Technik, Handeln und Sozialstruktur: Eine Einführung in die Soziologie der Technik“. *Technical University Technology Studies, Working Papers, TUTS-WP-3-2006*. Berlin.
- Ruddick, Graham (2016a). „Admiral to price car insurance based on Facebook posts“. 2.11.2016. <https://www.theguardian.com/technology/2016/nov/02/admiral-to-price-car-insurance-based-on-facebook-posts> (Download 8.5.2017).
- Ruddick, Graham (2016b). „Facebook forces Admiral to pull plan to price car insurance based on posts“. *The Guardian* 2.11.2016. <https://www.theguardian.com/money/2016/nov/02/facebook-admiral-car-insurance-privacy-data> (Download 8.5.2017).
- Schmidt, Florian A. (2016). *Arbeitsmärkte in der Plattform-ökonomie – Zur Funktionsweise und den Herausforderungen von Crowdwork und Gigwork*. Berlin: Friedrich-Ebert-Stiftung. <http://library.fes.de/pdf-files/wiso/12826.pdf> (Download 8.5.2017).
- Schneider, Jan, Ruta Yemane und Martin Weinmann (2014). *Diskriminierung am Ausbildungsmarkt Ausmaß, Ursachen und Handlungsperspektiven*. Berlin: Sachverständigenrat deutscher Stiftungen für Integration und Migration. https://www.svr-migration.de/wp-content/uploads/2014/11/SVR-FB_Diskriminierung-am-Ausbildungsmarkt.pdf (Download 8.5.2017).
- Spiekermann, Sarah (2015). *Ethical IT Innovation: A Value-Based System Design Approach*. Toronto: Apple Academic Press. <https://www.crcpress.com/Ethical-IT-Innovation-A-Value-Based-System-Design-Approach/Spiekermann/p/book/9781482226355> (Download 3.10.2016).
- Stalder, Felix (2017). „Algorithmen, die wir brauchen“. 16.1.2017. <https://netzpolitik.org/2017/algorithmen-die-wir-brauchen/> (Download 8.5.2017).
- Tene, Omer, und Jules Polonetsky (2011). *To Track or „Do Not Track“: Advancing Transparency and Individual Control in Online Behavioral Advertising* (SSRN Scholarly Paper No. ID 1920505). Rochester NY: Social Science Research Network. <https://papers.ssrn.com/abstract=1920505> (Download 8.5.2017).
- Tufekci, Zeynep (2014). „Engineering the public: Big data, surveillance and computational politics“. *First Monday* (19) 7. <http://firstmonday.org/ojs/index.php/fm/article/view/4901> (Download 8.5.2017).
- Tufekci, Zeynep, Jillian C. York, Ben Wagner und Frederike Kaltheuner (2015). „The Ethics of Algorithms: from radical content to self-driving cars“. Berlin: European University Viadrina. 1–18. <https://cihr.eu/publication-the-ethics-of-algorithms/> (Download 8.5.2017).
- van Haastert, Hugo (2016). „Government as a Platform: Public Values in the Age of Big Data“ (OII IPP 2016 Conference). Oxford Internet Institute.

- Wagner, Ben (2016). „Algorithmic regulation and the global default: Shifting norms in Internet technology“. *Nordic Journal of Applied Ethics* 22.3.2016. https://www.ntnu.no/ojs/index.php/etikk_i_praksis/article/view/1961 (Download 8.5.2017).
- Weber, Lauren, und Elizabeth Dwoskin (2014). „Are Workplace Personality Tests Fair?“. *Wall Street Journal* 30.9.2014. <http://www.wsj.com/articles/are-workplace-personality-tests-fair-1412044257> (Download 24.4.2017).
- Winner, Langdon (1978). *Autonomous Technology: Technics-out-of-Control as a Theme in Political Thought* (1st US-1st Printing edition). Cambridge MA u. a.: MIT Press.
- Winner, Langdon (1980). „Do Artifacts Have Politics?“. *Daedalus* (109) 1. 121–136.
- York, Jillian C. (2010). *Policing Content in the Quasi-Public Sphere* (Open Net Initiative Bulletin). Boston MA: Berkman Center, Harvard University. <https://opennet.net/policing-content-quasi-public-sphere> (Download 8.5.2017).
- Zhang, Pei, Sophie Stalla-Bourdillon und Lester Gilbert (2016). „A content-linking-context model for ‚notice-and-takedown‘ procedures“. Konferenz gehalten auf der WebSci '16 Proceedings of the 8th ACM Conference on Web Science. <http://eprints.soton.ac.uk/397678/> (Download 8.5.2017).
- Zittrain, Jonathan (2014). „Engineering an Election“. *Harvard Law Review Forum* (127) 8. 335–341.
- Zweig, Katharina Anna (2016). „2. Arbeitspapier: Überprüfbarkeit von Algorithmen“. 7.7.2016. <http://algorithm-watch.org/zweites-arbeitspapier-ueberpruefbarkeit-algorithmen/> (Download 24.4.2017).
- Zweig, Katharina Anna (2017, im Erscheinen). „Wo maschinelle Prognosen irren können“. Bertelsmann Stiftung.

9 Executive Summary

Algorithms are increasingly making decisions with us, for us and about us – thereby giving rise to new questions about participation. This expert opinion makes a number of initial suggestions for structuring and classifying the participation-related issues stemming from algorithmic decision-making (ADM). In today's digitized knowledge society, shaping technology is becoming a fundamental form of power. And if knowledge is power, then algorithms are becoming today's instruments of power. To what degree is it acceptable and desirable for algorithms to have an impact on the lives of individuals and on society as a whole? And which aspects of ADM must we consider more closely if we want to benefit from the potentials and minimize the risks to the greatest degree possible?

This expert opinion offers a general overview of ADM processes by explaining key concepts, delimiting the relevant area of analysis of ADM processes and presenting typical scenarios and functions relating to the use of these processes. With this as a basis, a system for classifying ADM processes is then proposed. The objective is to make it easier to evaluate and compare the potential impact of ADM processes on participation by using fewer criteria. This, in turn, will facilitate a prioritization and preparation of more in-depth research on the subject.

An evaluation of the impact on participation – a structure for which we are proposing here – is the point of departure for further steps, be they detailed analyses or regulatory measures. In those cases where ADM is expected to affect participation significantly, then a thoroughgoing effort must be made to ensure that positive impacts are exploited as much as possible, while negative impacts, such as ADM-driven discrimination, are prevented. When algorithms greatly influence participation, it is often necessary to counteract that influence by ensuring diversity, fair competition and principles of due process. The more prevalent the participation-relevant ADM processes are, the more comprehensive the potential responses must be.

Methodological approach

The potential impact on participation cannot be classified using binary decision-making criteria, since that would overlook the many socially relevant logics embedded in ADM processes.

Numerous criteria can affect the relevance to participation. The selection and weighting of these criteria, moreover, depend on how participation is defined. The point here is not to judge ADM processes as good or bad, but to evaluate their potential impact on participation in relative terms, regardless of the direction that might take.

The criteria can be divided into three groups: *actors*, *social embeddedness* and *consequences*. In other words, the groups do not reflect technological factors, but are an attempt to document political and social considerations.

The *actors*-related criteria examine the actual economic and political power of those supplying and/or operating the decision-making processes. The criteria used for *social embeddedness* reflect mutually reinforcing social interdependencies, both intended and unintended. Finally, the potential *consequences* – along with those that have already become apparent – are analyzed in terms of basic political and social rights. Existing legal norms are used as the guidelines for evaluating the impact on participation. The starting point is the German Antidiscrimination Act (§1 and 2 AGG) and its area of application. Discrimination is thus prohibited in the following areas in particular: recruitment, hiring, working conditions, membership in trade unions, and occupational training. Fair access to public goods such as education, social security, health care and housing must also be ensured.

We suggest assuming there is an impact on participation only after a certain threshold has been reached. Using the criteria sketched out below, the relative consequences for participation can be summarized and compared. This makes it possible to rank algorithms more simply and manageably in order to prioritize further steps, where needed.

Each of the criteria listed below can increase the *potential impact on participation* (Teilhabewirkungspotenzial, TWP; here, $TWP=+1$). If a criterion does not apply or cannot be evaluated as part of the quick-check process, no point is given for the potential impact ($TWP_{\text{neutral}}=0$). Half points can also be given for each criterion ($TWP=+0.5$), and the maximum is two points ($TWP_{\text{max}}=+2$). The values thus range from 0 to 2 points per criterion. The higher the overall value, the greater the process's potential impact on participation.

Test questions are used to evaluate the criteria, and answering them is not always easy. The various applications are highly diverse and, on occasion, obscure, which is precisely why the system of abstract classifications is beneficial: It reduces the complexity of a highly complex network of issues. What might limit or promote participation? To answer that question, a person or organization uses the evaluative structure to analyze one or more relevant processes. After applying the criteria to a specific case and considering the results, they can assess which aspects of the decision-making process are most prominent; they can also compare different processes with each other.

9.1 Actors

A1 Competition: Who is supplying/operating the algorithm? How much political and economic power do they have? Is there real competition between multiple suppliers/operators (oligopolies, monopolies, cartels, dominant market position, sovereign power)?

Example – Border controls: Even if private companies produce or supply software for granting visas, the software's use is an expression of sovereign power. No one can escape the state's power to grant visas, and to that extent a private supplier working on behalf of the state must also be seen as a monopoly. The actual organizational type or legal status of the supplier/operator is therefore less relevant for the impact on participation.

A2 Dependency: Is it a problem when the process or product is no longer available or access to it is denied? Can the product be readily replaced? Do substitute processes/products exist? What would it cost to switch to another process? Which organizations or groups are structurally dependent on the process? Do the system's construction and functioning result in an informal dependency? Which technical or social lock-in effects exist?

Example – Crowdwork: Robust competition exists in the market for crowdworking platforms. Multiple platform operators compete both for contract suppliers and crowdworkers. Participants could nonetheless become dependent if job histories and reputations become linked to the platform's specific algorithmic assessment processes. A similar lock-in effect can also arise with other assessment systems or with social media.

9.2 Social embeddedness

SE1 Self-determination: To what extent does the process serve (only) as a preliminary step for making a decision? Does the user retain any right to influence the process? Does the system decide (de facto or officially) on its own? How much freedom to change or manage an algorithmic decision does the user have? Do time pressure and the process's practical implementation affect the user's autonomy?

Example – Content control: The way software is designed for monitoring potentially illegal content can have a major impact on user autonomy. Time pressure (e.g. fast pace) and technical limitations can, de facto, greatly reduce the possibility of human input, even if a human will officially be the one to make the final decision.

SE2 Adaptation: How do people adapt to the algorithmic process? What impact does the decision-making system have on a practical level? Which interactions between humans and computers can potentially change the outcome? What is the process's actual power in social terms? Which (social) significance is ascribed to the process? Does it define and/or lead the discussion? How inherently influential are the process outcomes?

Example – Trending topics: The “trending topics” on Twitter are not infrequently seen as a social “reality.” Although it is not generally known how the ranking is created, it still affects the public discourse and different actors adjust their behavior to reflect the logic used by the “trending topics” algorithm.

9.3 Consequences

K1 Reach: How many people are affected by the decision-making process (e.g. user population)? Is the process’s reach known and/or can it be limited? How great is it?

Example – Candidate selection: If only one company uses a software application for choosing job candidates, then the application’s reach is relatively small. If, however, it becomes a core decision-making tool for many companies, its impact on participation increases accordingly, for example if the logic used by one system predominates when employees are hired in a particular region or industry.

K2 System change: Does the decision-making process undermine principles of solidarity? Would individualization change the system? Could the process result in an unintended transformation of the impacted (social) environment?

Example – Health insurance: If the costs for public health insurance are partially decided by algorithms and are subsequently individualized using a new logic, then a system which has been based on social solidarity will no longer be to some degree. Instead of having everyone pay only according to their ability to do so, at least part of each person’s contributions will be determined by individual factors.

K3 Discrimination: To what degree could people be disadvantaged by the decision-making process? Do the results exhibit a pattern of discrimination or might such a pattern be expected, e.g. based on race, ethnic or cultural background, gender, sexual identity, physical disability, age, religion or world view?

Example – Automated job listings: An analysis of Google’s employment advertisements revealed that users classified as men were shown more highly paid positions than were women. In this case, the process has an impact on job market participation.

9.4 Summary: Determining the potential impact on participation

The degree of the potential impact on participation (TWP) of a random new ADM process can thus be summarized as follows:

$$TWP = VA1 * VA2 * (A1 + A2 + SE1 + SE2 + K1 + K2 + K3)$$

For this expert opinion, ADM processes are fundamentally relevant when a digital decision-making system is used and when people or characteristics attributed to people are being evaluated by that system. These two filtering criteria – “digital decision-making systems” (VA1) and “evaluation of people” (VA2) – can have either a value of 0 (not true) or 1 (true). The degree of the potential impact on participation (TWP) is the product of both preliminary criteria and the sum of the impact criteria. The TWP value is the sum of the test criteria (A1 + A2 + SE1 + SE2 + K1 + K2 + K3). Expressing the extent to which an ADM process can affect participation, the overall result (TWP) can range from 0 to 14. This testing structure makes it possible to filter out processes which do affect participation but whose potential impact is limited. This could be the case when TWP is less than 3, for example, although this threshold value can be adjusted to reflect different definitions of participation.

The above criteria can be used to develop a fast and relatively simple testing procedure, allowing companies and evaluators to check their ADM processes using specific criteria in just two or three pages.⁴ Individuals and organizations can employ such criteria to evaluate ADM processes for their relevance to participation. To that extent, the process is a simple and quick introduction to understanding ADM processes. At the same time, it is not a substitute for a more comprehensive evaluation.

⁴ See the “Quick Check” developed by the Danish Institute for Human Rights:
https://hrca2.humanrightsbusiness.org/docs/file/HRCA%20Quick%20Check_English.pdf

10 Über die Autoren

Kilian Vieth

Kilian Vieth koordiniert das Projekt „European Surveillance Reform Initiative“ der Stiftung Neue Verantwortung in Berlin. Dort beschäftigt er sich mit Reformansätzen für eine demokratischere und effizientere Sicherheits- und Überwachungs politik in Europa. Seine Forschungsinteressen liegen im Bereich digitaler Menschenrechte, kritischer Sicherheitsforschung sowie politischer und sozialer Fragen algorithmischer Entscheidungsfindung. Zuvor war Kilian an der Forschungsstelle Internet & Menschenrechte an der Europa-Universität Viadrina in Frankfurt (Oder) in verschiedenen Forschungsprojekten tätig. Davor arbeitete er zwei Jahre im Bereich politische Kampagnen bei einer Kommunikationsberatung in Berlin. Kilian Vieth studierte Politikwissenschaft und European Affairs am Otto-Suhr-Institut der Freien Universität Berlin und an der Sciences Po Paris sowie Politik- und Verwaltungswissenschaft an der Zeppelin Universität Friedrichshafen.

Dr. Ben Wagner

Dr. Ben Wagner arbeitet an der Stiftung Wissenschaft und Politik (SWP) in Berlin. Zuvor war er Gründer und Leiter der Forschungsstelle Internet & Menschenrechte an der Europa-Universität Viadrina in Frankfurt (Oder). Seine Forschung beschäftigt sich mit digitalen Menschenrechten, algorithmischer Entscheidungsfindung und der Rolle des Internets in der Außenpolitik. Von 2013 bis 2014 war er Post-Doc an der Annenberg School for Communication, University of Pennsylvania. Er promovierte am Europäischen Hochschulinstitut in Florenz, Italien. Er ist als Berater für digitale Menschenrechte aktiv, u. a. für das Europäische Parlament, UNESCO, Europarat, Europäische Kommission und OSZE. Seine Forschung wurde u. a. in Politics, Telecommunications Policy, im Journal of Information Technology & Politics (JITP) und dem International Journal of Communications veröffentlicht.

Adresse | Kontakt

Bertelsmann Stiftung
Carl-Bertelsmann-Straße 256
33311 Gütersloh
Telefon +49 5241 81-81216

Konrad Lischka
Taskforce Digitalisierung
Telefon +49 5241 81-81216
konrad.lischka@bertelsmann-stiftung.de

www.bertelsmann-stiftung.de