Initiative D21

Kooperationspartner: KPMG

Unterarbeitsgruppe Algorithmen-Monitoring | Stand: 16. März 2019

DENKIMPULS DIGITALE ETHIK:

Bias in algorithmischen Systemen – Erläuterungen, Beispiele und Thesen

AUTOR_INNEN Corinna Balkow (Initiative D21 e.V.), Dr. Irina Eckardt (Initiative D21 e.V. / KPMG)

MITWIRKENDE Johann Jakob Häußermann (Center for Responsible Research and Innovation des Fraunhofer IAO), Lena-Sophie Müller (Initiative D21 e.V.), Ann Cathrin Riedel (LOAD e.V.), Dr. Nora Schultz (Geschäftsstelle Deutscher Ethikrat), Prof. Barbara Schwarze (Initiative D21 e.V. / Kompetenzzentrum Technik-Diversity-Chancengleichheit e.V.), Birgit Wintermann (Bertelsmann Stiftung), Mitarbeitende von KPMG AG und KPMG Law

- **Menschen bringen ihre eigenen soziokulturell geprägten Wahrnehmungen und Erfahrungen mit. Damit kann es keine komplexen Entscheidungen geben, die ohne Bias entstehen – weder analog, noch digital.**
- **Algorithmen agieren nicht unabhängig von den Menschen, die sie beauftragen, herstellen oder einsetzen – Pflichten und Verantwortung ergeben sich für alle am algorithmischen System beteiligten Menschen.**
- **Für den Umgang mit Bias in algorithmischen Systemen bedarf es keiner neuen Grundrechte.**

I. Einleitung

Kaum eine Entwicklung hat unser Leben in den letzten Jahren so verändert wie der Einsatz algorithmischer Systeme. Sie sollen uns das Leben einfacher machen, repetitive Arbeit abnehmen und bei der Entscheidungsfindung helfen. Allerdings wissen nur die Wenigsten, auf welcher Grundlage diese Systeme uns Ergebnisse als Entscheidungen präsentieren, welche Daten verarbeitet werden und was mit den Daten nach ihrer Nutzung passiert.

Der vorliegende Denkimpuls wurde im Rahmen der Unterarbeitsgruppe Algorithmen-Monitoring der Initiative D21 e.V. erarbeitet. Die vorgestellten Thesen für den Umgang mit Bias in Algorithmen sollen zu einer differenzierteren Debatte beitragen und eine breitere Diskussion initiieren. Wir haben Fragestellungen identifiziert und uns mit diesen aus sozioökonomischer, technologischer und ethisch-rechtlicher Sicht beschäftigt. Die Thesen beziehen sich auf den jeweiligen Umgang mit Bias und illustrieren diesen mit Beispielen. Es ist nicht das Ziel, mit diesem Papier abschließende Antworten auf zum Teil bereits drängende Fragen zu geben. Vielmehr soll eine Grundlage für einen nachhaltigeren Umgang mit dem Thema Bias in algorithmischen Systemen geschaffen werden.

Was sind Bias?

Mit Bias wird umgangssprachlich vieles beschrieben, von Vorurteilen über Verzerrungen in datengetriebenen Entscheidungssituationen bis hin zur Förderung oder Vernachlässigung bestimmter gesellschaftlicher Gruppen. Die Gründe hierfür können sowohl in bewussten wie auch

in unbewussten Herangehensweisen liegen.[1] Sie beruhen auf erlebten Erfahrungen oder eingeengten Informationen und Sichtweisen zu Personen oder Gruppen. So werden beispielsweise Datensätze spezifischer ethnischer Gruppen nur eingeschränkt zum Testen ausgewählt.[2] Oder es fehlen ausgewogene Datensätze, weil nicht bekannt ist, dass Krankheiten je nach Geschlecht und/oder Ethnie unterschiedliche Symptome und Auswirkungen haben können.[3] Beide Formen, bewusste und unbewusste Bias, wirken sich auf die Qualität der Ergebnisse algorithmischer Systeme aus. Statistische Qualitätsstandards aus der analogen Arbeit oder DIN-Normen zur Integration von Nutzerinnen und Nutzern bleiben auch für die Qualität zahlreicher digitaler Prozesse relevant. Sie sind nützlich für die Überführung in neue digitale Standards. Gleichwohl zeigen wissenschaftliche Studien, dass Chancengerechtigkeit in algorithmischen Systemen nicht ausschließlich durch mathematische Methoden bewirkt werden kann, da jeweils gleiche Daten häufig unterschiedlich interpretiert werden müssen.[4] Ein besonderer Effekt geht mit dem Vorhandensein der Bias einher: die sogenannte Bias-Blindheit. Sie beschreibt die Tendenz, dass sich die meisten Menschen für unbeeinflusst halten. Studien in Bereichen des Managements weisen beispielsweise darauf hin, dass Managerinnen und Manager – seien sie besonders oder weniger qualifiziert – sich nicht vorstellen können, wie stark sie selbst von einem Bias betroffen sind.[5]

Praktisch betrachtet ist die Welt, in der wir uns bewegen, – ob nun analog oder digital – sehr von unserer subjektiven Wahrnehmung geprägt. Diese Wahrnehmung basiert auf dem jeweiligen sozialen, kulturhistorischen und ökonomischen Hintergrund, welcher uns durch gesellschaftsspezifische Normen, Erziehung aber auch durch Medien und Kulturangebot vermittelt wird. Diese unterschiedlichen Einflüsse spiegeln sich in den Diskussionen und Entscheidungen der Gesellschaft bzw. des Individuums wider. Oft entscheiden Menschen, obwohl sie dafür zu viele oder zu wenige Informationen haben. Dabei überschätzen viele Menschen die Wichtigkeit der Informationen, die sie haben, oder tendieren dazu, nur Informationen zu vertrauen, welche ihre bisherige Meinung oder ihr bisheriges Wissen stützen. Zudem basiert die Entscheidung, welche sich aus dem Ergebnis ableiten lässt, oft nur auf dem Ergebnis selbst, aber nicht auf dem Umstand, wie es zu dem Ergebnis kam. Diese Einschränkungen werden als **kognitive Bias**[6] bezeichnet.

Durch die Nutzung digitaler Medien und anderer technologischer Lösungen stehen heutzutage immer mehr Daten zur Verfügung. Die Menschen geben nicht nur ihre Daten weiter, sondern mit ihnen auch ihre Ansichten, Meinungen und Vermutungen. Es muss daher von der Existenz von Bias in Daten ausgegangen werden. **Statistische Bias** beschreiben Verzerrungen bei der Verteilung von Datenpunkten, sowie systemische oder zufällige Fehler in der Datenerhebung. Dies führt dann zu fehlerhaften Ergebnissen in einer statistischen Untersuchung. So kann bspw. das Design eines Fragebogens die Ergebnisse hinreichend beeinflussen: Bei einer Skalenfrage – „Wie schätzen Sie auf einer Skala von 1-10 ... ein?“ – geben die meisten Befragten einen Wert in der Mitte an oder wählen einen Wert an den extremen Rändern, also 1 oder 10.[7]

Die notwendigen Annahmen, welche in Untersuchungen oder bei der Entwicklung von Lernalgorithmen, sowie von den Algorithmen zur Laufzeit angestellt werden müssen, um Beobachtungen verallgemeinern zu können, werden

1 Booth, Robert; Mohdin, Aamna (2018): Revealed: the stark evidence of everyday racial bias in Britain; online verfügbar unter: https://amp.theguardian.com/uk-news/2018/dec/02/revealed-the-stark-evidence-of-everyday-racial-bias-in-britain (letzter Abruf: 02.02.2019).

2 Cossins, Daniel (2018): Discriminating algorithms: 5 times AI showed prejudice; online verfügbar unter: https://www.newscientist.com/article/2166207-discriminating-algorithms-5-times-ai-showed-prejudice/ (letzter Abruf: 02.02.2019). .

3 European Society of Cardiology (2017): Sex in basic research: concepts in the cardiovascular field; online verfügbar unter https://www.dgesgm.de/images/pdf/Ventura-Clapier%20R%20Dworatzek%20E%20Seeland%20U%20et%20al%20Card%20Res%202017.pdf (letzter Abruf: 14.02.2019).

4 Friedler, Sorelle A.; Scheidegger, Carlos; Venkatasubramanian, Suresh; Choudhary, Sonam; Hamilton, Evan P.; Roth, Derek (2018): A comparative study of fairness-enhancing interventions in machine learning; online verfügbar unter: https://arxiv.org/abs/1802.04422 (letzter Abruf: 02.02.2019).

5 ProoV Team (2018): How Enterprises Overcome Digital Bias with International Collaboration; online verfügbar unter: https://proov.io/blog/enterprises-overcome-digital-bias-international-collaboration/ (letzter Abruf: 02.02.2019).

6 Wikipedia (2018): List of cognitive biases; online verfügbar unter: https://en.wikipedia.org/wiki/List_of_cognitive_biases (letzter Abruf: 02.02.2019).

7 Bogner, Kathrin; Landrock, Uta (2015): Antworttendenzen in standarisierten Umfragen; online verfügbar unter: https://www.gesis.org/fileadmin/upload/SDMwiki/Archiv/Antworttendenzen_Bogner_Landrock_11122014_1.0.pdf (letzter Abruf: 02.02.2019).

als **induktive Bias** bezeichnet. Sie stellen eine Grundlage vieler algorithmischer Systeme dar und bedürfen daher einer eigenen Betrachtung. Lernende algorithmische Systeme setzen auf Daten aus der Vergangenheit auf, diese beinhalten nicht automatisch heutige Zielvorstellungen. Wenn in der Vergangenheit eher Männer als Frauen eingestellt oder befördert wurden, dann muss dem algorithmischen System vorgegeben werden, ob das weiterhin gewollt ist oder eine unerwünschte Verzerrung darstellt. Jedoch ist eine Generalisierung ohne induktive Bias unmöglich. Beim induktiven Lernen wird aus Beispielen eine Funktion erlernt und schrittweise verallgemeinert. Die These dabei ist: Wenn sich eine lernende Funktion durch eine hinreichend große Beispielmenge gut einer Zielfunktion annähert, wird diese Funktion sich auch bei unbekannten Beispielen annähern.

Warum spielen Bias in algorithmischen Systemen eine besondere Rolle?

Bei einfachen Algorithmen (z. B. Taschenrechner) spielen Bias keine Rolle, wichtig werden sie erst, wenn Menschen von automatisierten Entscheidungen betroffen sind. Für betroffene Personen kann z. B. eine Suche im Internet wie ein einfacher Algorithmus wirken – man gibt etwas ein und der Algorithmus gibt etwas aus. Dabei handelt es sich im Hintergrund meist um datenbasierte algorithmische Systeme. Diese sind weltweit im Einsatz, z. B. bei Online-Flugbuchungen[8], bei Bewerbungsverfahren[9] oder in staatlichen Organisationen[10]. In Berichten über automatisierte Entscheidungssysteme gibt es viel Kritik an diskriminierenden Algorithmen.[11] Häufig wird dann mit Bias eine subjektive – gewollte oder nicht gewollte – Einflussnahme auf das algorithmische System beschrieben. In der Forschung wird noch zwischen automatisierten Entscheidungssystemen und unterstützenden Systemen unterschieden.[12] Wir werden der Übersichtlichkeit halber von algorithmischen Systemen sprechen, wenn wir über den gesamten Prozess reden und von Algorithmen, wenn es um ein eindeutiges Set von Anweisungen geht. Für die Diskussion ist es wichtig, zunächst die Komplexität eines algorithmischen Systems zu verdeutlichen.

Die folgende Illustration gibt eine Hilfestellung, um den Einfluss von Bias innerhalb des gesamten Lebenszyklus eines algorithmischen Systems zu verstehen. Durch die Berücksichtigung von verschiedenen Arten von Bias können bessere Standards geschaffen werden, welche den Umgang und speziell die Identifizierung von Bias ermöglichen. Die Grafik zeigt die unterschiedlichen Schritte in der Entwicklung eines solchen Systems, um die mögliche Einflussnahme unterschiedlicher Arten von Bias zu verdeutlichen. Vorurteile können Namen, Aussehen oder Fähigkeiten anderer Menschen betreffen, die als implizite Werturteile in die Arbeit am algorithmischen System einfließen. Daher gilt es, die Interessen der Menschen, welche die Algorithmen in Auftrag geben, planen, vorgeben, entwickeln, testen und einsetzen zu berücksichtigen. Zudem müssen die qualitative Expertise der Datenlieferanten sichergestellt sowie gesellschaftliche Erwartungen und Vorgaben beachtet werden.

8 Morris, Hugh (2018): Airlines are starting to price their seats based on your personal information – but is it legal?; online verfügbar unter: https://www.telegraph.co.uk/travel/news/dynamic-fare-pricing-airline-ticket-personalisation/ (letzter Abruf: 02.02.2019).

9 Buranyi, Stephen (2018): ‚Dehumanising, impenetrable, frustrating': the grim reality of job hunting in the age of AI; online verfübar unter: https://www.theguardian.com/inequality/2018/mar/04/dehumanising-impenetrable-frustrating-the-grim-reality-of-job-hunting-in-the-age-of-ai (letzter Abruf: 02.02.2019).

10 Stats NZ (2018): Algorithm assessment report; online verfügbar unter: https://www.data.govt.nz/assets/Uploads/Algorithm-Assessment-Report-Oct-2018.pdf (letzter Abruf: 02.02.2019).

11 Coffey, Helen (2018): Airlines face crack down on use of 'exploitative' algorithm that splits up families on flights; online verfügbar unter: https://www.independent.co.uk/travel/news-and-advice/airline-flights-pay-extra-to-sit-together-split-up-family-algorithm-minister-a8640771.html (letzter Abruf: 02.02.2019).

12 Gesellschaft für Informatik (2018): Technische und rechtliche Betrachtungen algorithmischer Entscheidungsverfahren. Studien und Gutachten im Auftrag des Sachverständigenrats für Verbraucherfragen; online verfügbar unter: http://www.svr-verbraucherfragen.de/wp-content/uploads/GI_Studie_Algorithmenregulierung.pdf (letzter Abruf: 14.02.2019).

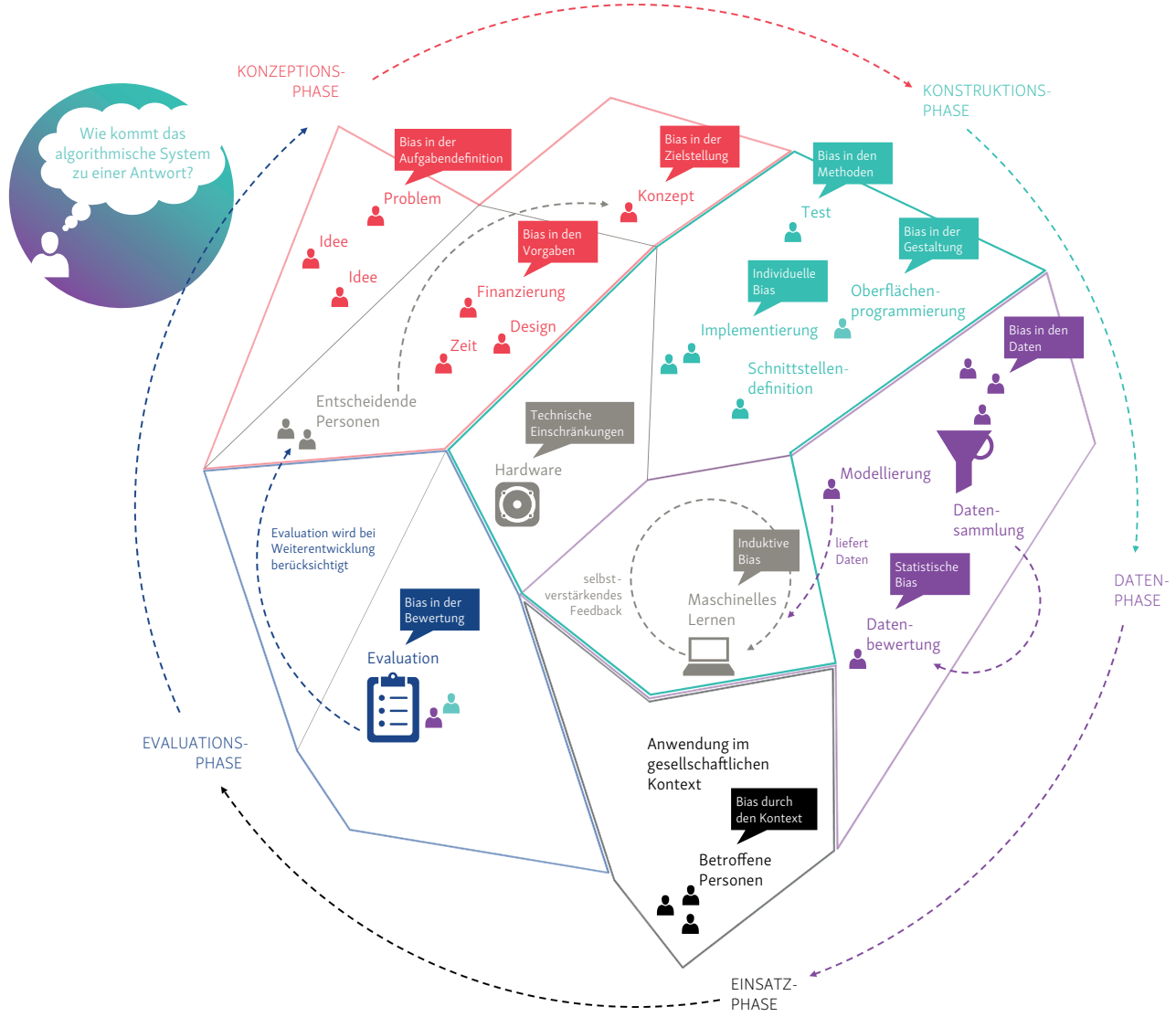

Abbildung 1: Verortung potenzieller Bias in einem algorithmischen System

- **Konzeptionsphase - Bereich des Auftrags:**
  Welche Probleme sollen durch einen Algorithmus gelöst werden? Welche Finanzierung / Zeit steht zur Verfügung? Welche Rahmenbedingungen wurden vorgegeben? Wie wird die Zielstellung definiert?
- **Konstruktionsphase - Bereich der Umsetzung:**
  Welche Ziele werden für die Algorithmen definiert? Welche Aufgabenstellung kann technisch wie umgesetzt werden? Welche Möglichkeiten werden programmiert? Sind Testphasen vorgesehen? Welche Hardware steht zur Verfügung?
- **Datenphase - Bereich der Datensammlung und -ordnung:**
  Welche Daten werden als relevant ausgewählt? Welche Datensätze werden priorisiert? Sind die Daten für das Ziel des Algorithmus geeignet und ausreichend? Besteht eine sinnvolle Auswahl von Trainings- und Testdaten? Liegen statistische Bias in den Daten vor?
- **Einsatzphase - Bereich der gesellschaftlichen Einbettung:**
  In welchem Kontext wird das algorithmische Entscheidungssystem eingesetzt? Wer verwendet die Empfehlungen für die eigenen Entscheidungen? Werden Auswirkungen auf gesellschaftliche Gruppen überprüft und getestet?
- **Evaluationsphase – Bereich der Bewertung und Verbesserung:**
  Wie wird Erfolg bewertet? Welche Möglichkeiten gibt es, Feedback auszuwerten? Wie werden Rückmeldungen berücksichtigt? Kommt es zu ethisch bedenklichen Wirkungen?

II. Technologische Perspektive auf Bias in algorithmischen Systemen

Die hohe Komplexität der Algorithmen, die mit dem maschinellen Lernen verbundenen ist, setzt voraus, dass die konzeptionelle Gestaltung eines Algorithmen-Monitorings nicht ohne technische Expertise passiert. Einem algorithmischen System wird das benötigte „Weltwissen" durch Menschen mit jeweils eigenen Auswahlkriterien vorgegeben. Es kann keine Entscheidungen über richtig oder falsch treffen, nur sichere oder unsichere Wahrscheinlichkeitswerte liefern. Jeder Input muss durch eine eigene Datenquelle integriert werden. Jeder Output wird von außen vorgegeben. Die Bewertung von daraufhin getroffenen Entscheidungen verbleibt bei den Entwicklerinnen und Entwicklern sowie den Personen, die das algorithmische System einsetzen.

Werden Korrekturen am algorithmischen System vorgenommen, müssen weitere Überprüfungen folgen, um sicherzustellen, dass das System fortlaufend funktioniert. Da sich in Unternehmen und Organisationen übliche Vorgehensweisen, angewandte Technologien, eingebundene Kooperationspartner und genutzte Recherche- und Informationsquellen häufig wiederholen, empfiehlt die Anti-Bias-Forschung auch, bewusst heterogene Sichtweisen einzubeziehen.[13]

These: Bias finden sich in verwendeten Daten.

Beschreibung: Auch bei einer sorgfältigen Auswahl von Datenquellen werden Bias in verwendeten Daten enthalten sein und durch die Nutzung ins algorithmische System einfließen. Es liegt in der Verantwortung von Personen, die algorithmische Systeme beauftragen, entwickeln oder einsetzen, die verwendeten Daten auf Bias zu untersuchen und ggf. geeignete Maßnahmen zum Umgang zu treffen.

Umgang: Manche Daten sind aufgrund von Bias für die intendierte Anwendung in algorithmischen Systemen nicht geeignet. Menschen, die algorithmische Systeme entwickeln, sollten im Laufe des Lebenszyklus eines Algorithmus Bias in den Daten identifizieren, mit neuen oder anderen Datenquellen testen, und ggf. Datenquellen anpassen oder austauschen. Die regelmäßige Überprüfung sollte fester Bestandteil des Entwicklungs- und Betriebszyklus werden.

Beispiele: Es gibt vielfältige technische Möglichkeiten zur Identifikation von Bias sowie potenzielle Gegenmaßnahmen. Ein Open-Source-Ansatz kann helfen, Transparenz über die eingesetzten technischen Methoden zu schaffen. Beispielsweise stellte die Universität Chicago ein „open source bias audit toolkit for machine learning developers, analysts, and policymakers" vor, mit dem Machine-Learning-Modelle auf Diskriminierungen und Bias untersucht werden können.[14] IBM stellte Lösungen im Rahmen des Projektes AIF360 vor. Diese beinhalten Beispiele für mögliche Anwendungen, den Umgang mit Bias, sowie Beispiele in konkreten Datensätzen.[15] Bias in Daten können zur strukturellen Benachteiligung von Gruppen führen, z. B. im Fall von arbeitssuchenden Frauen und Menschen mit Kindern.[16]

These: Bias werden von Menschen in die konkrete Ausgestaltung eines algorithmischen Systems eingebracht.

Beschreibung: Durch fehlerhafte Annahmen von Menschen, die das algorithmische System beauftragen, planen, gestalten, umsetzen, testen und einsetzen, bringen diese bewusst und unbewusst Bias in die Entscheidungen für die Formulierung des Auftrags, in die Auslegung des Systems, die Auswahl der Datensets, die Passung für den gesellschaftlichen Kontext und die erwarteten Ergebnisse ein. In ihren Rollen haben sie die Aufgabe, nicht nur die Funktionstüchtigkeit eines Systems zu bewerten, sondern auch die Existenz von Bias zu hinterfragen und mit bekannten typischen Fehlern abzugleichen.

13 ProoV Team (2018): How Enterprises Overcome Digital Bias with International Collaboration; online vefügbar unter: https://proov.io/blog/enterprises-overcome-digital-bias-international-collaboration/ (letzter Abruf: 02.02.2019).
14 Center for Data Science and Public Policy of The University of Chicago (2019): Aequitas; online verfügbar unter: https://dsapp.uchicago.edu/projects/aequitas/ (letzter Abruf: 14.02.2019).
15 NB Viewer Juypter (2019): Detecting and mitigating age bias on credit decision; online verfügbar unter: https://nbviewer.jupyter.org/github/IBM/AIF360/blob/master/examples/tutorial_credit_scoring.ipynb (letzter Abruf: 02.02.2019).
16 Wimmer, Barbara (2018): Computer sagt nein: Algorithmus gibt Frauen weniger Chancen beim AMS; online verfügbar unter: https://futurezone.at/netzpolitik/computer-sagt-nein-algorithmus-gibt-frauen-weniger-chancen-beim-ams/400345297 (letzter Abruf: 02.02.2019).

Umgang: Um Bias bei Menschen zu begegnen, ist es zuerst notwendig, diese in die Lage zu versetzen, Bias zu erkennen. Eine Quelle können Initiativen sein, die sich für größere Diversität rund um algorithmische Systeme einsetzen: Black in AI[17], Queer in AI[18], Women in Machine Learning[19], Lesbians Who Tech[20], Latinx in AI[21], Speakabled[22] oder AI for Deaf[23]. Wesentlich ist hierbei, sich nicht nur auf bestehende Initiativen zu beschränken, sondern auch zu untersuchen, ob für die eigene Anwendung oder den eigenen Markt weitere Personengruppen vor dem Hintergrund etwaiger Bias berücksichtigt werden sollten.

Beispiele: Bei der Ausgestaltung von algorithmischen Systemen können eine Vielzahl von möglichen fehlerhaften Annahmen, wie: Ein Produkt hat genau einen Preis. Es gibt genau zwei Geschlechter. Namen von Männern ändern sich nicht über die Zeit. Eine Sammlung solcher fehlerhaften Annahmen[24] kann bei der Identifizierung und Vermeidung dieser Bias eine wesentliche Rolle spielen. Auch die Verwendung von nicht vergleichbaren Datensätzen kann zu Bias führen, die das angestrebte Ziel des algorithmischen Systems gefährden. Ein Beispiel für eine Diskussion hierzu findet sich rund um die angestrebte Erkennung von Kriminellen und Nicht-Kriminellen.[25]

These: Menschen, die als „Objekt" in einem Algorithmus auftauchen, aber auch Menschen, die ein algorithmisches System aus eigenem Antrieb nutzen, können die Funktionsweise des algorithmischen Systems oft nur schwer nachvollziehen und so Bias nicht oder nur sehr eingeschränkt erkennen.

Beschreibung: Seitens der Nutzerinnen und Nutzer und der Zivilgesellschaft besteht Informationsbedarf gegenüber den Eignern von algorithmischen Systemen hinsichtlich der Frage, ob und wie Bias bestehen und für welche Aufgaben das algorithmische System eingesetzt wird. In vielen Fällen kann dies aufgrund der Vertraulichkeit der Daten oder proprietären Nutzung der Daten nicht direkt geschehen. Dann kann eine externe Prüfung der Algorithmen und der Daten auf Bias durch Dritte erfolgen.

Umgang: Transparenz und Nachvollziehbarkeit über potenzielle Bias und den Umgang damit in algorithmischen Systemen zu schaffen, fußt auf den Ergebnissen technischer Analysen. Diese sollten seitens der Entwickler und Eigner öffentlich gemacht werden. In vielen Fällen, in denen eine direkte Offenlegung der Daten oder des Algorithmus nicht möglich oder nicht erwünscht ist, können Analysen durch Dritte durchgeführt und deren Ergebnisse veröffentlicht werden. Dies ist dann möglich, ohne auf einzelne Daten oder Schritte im Algorithmus einzugehen. Hierbei ist es jedoch notwendig transparent zu machen, was im Rahmen der Prüfung untersucht wurde und welche Methoden zum Einsatz kamen.

17 Black in AI (2019). Online verfügbar unter: https://blackinai.github.io/#about (letzter Abruf: 02.02.2019).
18 Queer in AI (2019). Online verfügbar unter: https://queerai.github.io/QueerInAI/ (letzter Abruf: 02.02.2019).
19 Women in Machine Learning (2019). Online verfügbar unter: https://wimlworkshop.org/ (letzter Abruf: 02.02.2019).
20 Community of Queer Women in or around tech (2019). Online verfügbar unter: https://lesbianswhotech.org/about/ (letzter Abruf: 02.02.2019).
21 LatinX in AI (2019). Online verfügbar unter: https://www.latinxinai.org (letzter Abruf: 02.02.2019).
22 Speakabled. (2019): Liste von Menschen mit Behinderungen, die über Tech und Programmierung sprechen können; online verfügbar unter: https://www.speakabled.com/sprecherinnen/?members_search=Tech+und+Programmierung (letzter Abruf: 02.02.2019).
23 Rochester Institute of Technology (2019): National Technical Institute for the Deaf; online verfügbar unter: https://www.ntid.rit.edu/ (letzter Abruf: 02.02.2019).
24 Github (2019): Awesome falsehood; online verfügbar unter: https://github.com/kdeldycke/awesome-falsehood/blob/master/README.md (letzter Abruf: 02.02.2019).
25 Sullivan, Ben (2016): A New Program Judges If You're a Criminal From Your Facial Features; online verfügbar unter: https://motherboard.vice.com/en_us/article/d7ykmw/new-program-decides-criminality-from-facial-features (letzter Abruf: 02.02.2019).

Beispiele: Menschen werden beispielsweise als Objekt in einem algorithmischen System behandelt, wenn sie eine Bonitätsprüfung in der Kreditvergabe brauchen. Es bestehen seitens öffentlicher Organisationen[26] Bestrebungen, Transparenz über algorithmische Systeme und Bias zu schaffen. So wird in einem Projekt versucht, die Annahmen und Wirkungsweisen eines Bonitätsalgorithmus[27] nachzuvollziehen. Auch bei einer aktiven Nutzung von algorithmischen Systemen können Bias negative Effekte haben. So können Menschen momentan oft nur einer generellen Nutzung ihrer Daten für eine kostenlose Nutzung von Diensten zustimmen und selten einer eingeschränkten Nutzung, wie Ort, Dauer, oder Häufigkeit. Zurück bleibt ein Gefühl der Hilflosigkeit und die Schwierigkeit, ein digitales Bauchgefühl[28] zu entwickeln.

III. Sozioökonomische Perspektive auf Bias in algorithmischen Systemen

Übernehmen algorithmische Systeme Entscheidungen, die früher durch Menschen getroffen wurden, könnten deren Bias korrigiert werden. Die Gesellschaft wird das Ergebnis der Entscheidung kritisch hinterfragen und mit einer menschlichen Entscheidungsfindung vergleichen. Im Prozess zur Realisierung von algorithmischen Systemen müssen deshalb Festlegungen und Definitionen abgestimmt werden.

These: Durch die neue Konfrontation mit Bias in algorithmischen Systemen werden auch „analoge" Bias stärker auf den Prüfstand gestellt.

Beschreibung: Menschliche Entscheidungen basieren oft auf gesellschaftlich und kulturell anerkannten Standards, die Stereotypen beinhalten und damit Bias darstellen. Diese analogen Bias werden selten hinterfragt oder kritisiert, solange andere gesellschaftliche Gruppen durch das Festhalten an solchen profitieren.[29]

Bei der Einführung von algorithmischen Systemen stellen Menschen höhere Anforderungen und sind kritischer, was Bias anbetrifft. Dies liegt unter anderem darin begründet, dass Anweisungen in der analogen Welt oft mit einem gewissen Ermessensspielraum definiert werden und dieser Spielraum auch zumeist zugestanden wird. In der digitalen Welt hingegen, scheinen diese Anweisungen festgeschrieben und klar definiert.[30]

Umgang: Durch Aufklärung über Bias in algorithmischen Systemen sollen alle daran involvierten Akteure, aber auch gesellschaftliche Gruppen, dazu motiviert werden, eigene Gedanken zu dem Thema zu entwickeln. Die Auseinandersetzung mit potenziellen Bias in digitalen Systemen erfordert auch den Vergleich mit der bisherigen analogen Praxis. Die Aufklärung kann in der gesamten Bandbreite möglicher digitaler und analoger Informationsmedien, in Form von Vorträgen, Workshops oder Schulungen aber auch durch journalistische Aufbereitung im Rahmen der öffentlichen Meinungsbildung geschehen.

Hier bietet sich ein Zusammenwirken von Unternehmen, Wissenschafts- und Forschungseinrichtungen sowie öffentlichen und privaten Trägern an, insbesondere auch auf regionaler Ebene kleine und mittlere IT-Unternehmen. Menschen, die von diesem Wissen Gebrauch machen und sich über diese Erkenntnisse austauschen, können sich gegebenenfalls selbstständig im positiven

26 Algorithm Watch (2019): Mission Statement; online verfügbar unter: https://algorithmwatch.org/de/mission-statement/ (letzter Abruf: 02.02.2019).
27 Open Knowledge Foundation (2019): Get Involved: We crack the Schufa!; online verfügbar unter: https://okfn.de/blog/2018/02/open-schufa-english/ (letzter Abruf: 02.02.2019).
28 Müller, Lena-Sophie (2016): Das digitale Bauchgefühl. In: Friedrichsen M., Bisa PJ. (Hrsg.) Digitale Souveränität. Springer VS, Wiesbaden.
29 Hohlweg, Jelena; Salentin, Kurt (2014): Datenhandbuch ZuGleich. Zugehörigkeit & (Un-) Gleichwertigkeit IKG Technical Report Nr. 5, Version 1. Bielefeld; online verfügbar unter: https://pub.uni-bielefeld.de/download/2726338/2726339/Datenhandbuch_ZUGLEICH.v1.pdf (letzter Abruf: 02.02.2019).
30 Bullik, Andreas; Meiners, Kay (2018): Frau Zweig, was können Computer besser, und was Menschen?; online verfügbar unter: https://www.magazin-mitbestimmung.de/artikel/Frau+Zweig%2C+was+k%C3%B6nnen+Computer+besser%2C+und+was+Menschen%3F@6032 (letzter Abruf: 02.02.2019).

Sinne weiterentwickeln. Sie können Bias in algorithmischen Systemen aber auch in der analogen Welt in Zukunft während oder nach der Erstellung besser identifizieren und regeln.

Beispiele: Der Unterscheid in der Reichweite von Bias kann gut im Buchhandel verdeutlicht werden. Bei der Frage nach einer Buchempfehlung im lokalen Buchladen spielen bei der Empfehlung sowohl die Vorlieben der Angestellten, die derzeitige Bestsellerliste als auch die Auswahl an Büchern in dem Laden eine Rolle. Es ist weithin akzeptiert, dass die Empfehlungen von Fachangestellten auf deren Erfahrungen basieren. Wenn wir nun aber ein algorithmisches System mit diesen Erfahrungen trainieren und die Reichweite des Buchladens von lokal zu global erweitern, haben die Entscheidungen auf Grundlage der Erfahrungen ganz andere Tragweiten. Ähnlich zu diesem Fall ist es ebenfalls weithin akzeptiert, wie bzw. dass Schülerinnen und Schüler nach unterschiedlichen Auswahlkriterien auf spezifische Schulen respektive Klassen verteilt werden. Die Grundlagen für diese Entscheidungen werden häufig nicht bekannt gegeben und bleiben für die Ausgewählten wie auch deren Eltern weitestgehend intransparent. Wenn dies automatisiert und von einem algorithmischen System übernommen werden sollte, müssten die Kriterien für die Entscheidungen in den Auftrag einbezogen werden. Damit gäbe es unter der Voraussetzung, dass dies bei den kommunalen oder schulischen Auftraggebern auch gewollt wäre, mehr Chancen für eine Transparenz der Ergebnisse.

These: Keine Entscheidung ohne Bias.

Beschreibung: Soziokulturelle Erfahrungen, schulische Lernergebnisse sowie ökonomisch bedingte Lebensverhältnisse beeinflussen die menschliche Informationsaufnahme und -verarbeitung. Sie sind in der individuellen Interpretation von Information Grundlage menschlicher Entscheidungen. Ziel ist daher, die Reduzierung von Bias sowie der bewusste Umgang mit ihnen.

Qualitätsstandards und Qualitätskriterien aus der analogen wie der digitalen Welt können genutzt und angepasst werden, um Bias zu reduzieren. So können beispielsweise Berufsethiken, Qualitäts- und Antidiskriminierungsnormen, Vorgaben zur Nutzerintegration und gesetzliche Vorgaben zur Chancengleichheit in einen digitalen Anforderungskatalog für die jeweiligen Phasen des Entwicklungsprozesses aufgenommen werden. Folglich gibt es vielfältige Möglichkeiten, Bias in analogen sowie algorithmischen Systemen transparenter zu machen und zu reflektieren.

Umgang: Zu Beginn steht die Reflexion, inwieweit Bias in einem algorithmischen System vorhanden sind. Anhand von Gütekriterien[31] kann ein Urteilsvermögen bzw. eine Bewertung über bereits vorliegende algorithmische Systeme geschaffen werden. Werden Bias in einem algorithmischen System identifiziert, muss entschieden werden, wie mit dem System weiter vorgegangen werden kann. Als Ergänzung zu Normen könnten Checklisten für übergreifende Standards entwickelt werden.

Mögliche Punkte auf einer Checkliste könnten sein:

- **Wie schaffen Sie in Ihrem Unternehmen ein ausgeprägtes Verständnis für die Bedeutung von Bias im Zusammenhang mit algorithmischen Systemen?**
- **Wurden Mitarbeiterinnen und Mitarbeiter, die an der Entwicklung und Anwendung von algorithmischen Systemen beteiligt sind, zum Thema Bias geschult?**
- **Sind sie bei Konstruktion, Anwendung und Evaluation von algorithmischen Systemen aufgefordert zu reflektieren, wo Bias vorliegen könnten?**
- **Führt Ihr Unternehmen einen Daten- und Algorithmen-Katalog, in dem Details zur Datenherkunft und zu den verwendeten Modellen hinterlegt sind?**

Beispiele: In einer Studie des MIT im Bereich autonomes Fahren („Moral Machine!“) wurden 40 Millionen Nutzerinnen und Nutzer aus über 200 Ländern vor die Aufgabe gestellt, zu entscheiden, wen sie in einer Gefahrensituation im Verkehr retten würden. Auffällig war, dass es nur sehr wenige Unterschiede in den Ergebnissen gab, die sich auf das Alter der Teilnehmenden zurückführen ließen. Allerdings konnte man klare Cluster bezüglich der geografischen Regionen und der Kulturkreise erkennen. So war es möglich, Gruppen von Ländern zu definieren, die sich in östlich, westlich und südlich unterteilen ließen. Menschen aus dem südlichen Cluster votierten eher für die Rettung junger Menschen; Menschen aus dem östlichen Cluster hätten eher ältere Menschen gerettet. Für die Programmierung eines autonom fahrenden Fahrzeugs lässt sich nun fragen, ob man sich an diesen Ergebnissen orientiert kann.

31 iRights.lab (2019): Auf dem Weg zu Gütekriterien für den Algorithmeneinsatz; online verfügbar unter: https://irights-lab.de/auf-dem-weg-zu-guetekriterien-fuer-den-algorithmeneinsatz/ (letzter Abruf: 02.02.2019).

Auch ökonomische Interessen von Firmen können sich als Bias auswirken, welche in bestimmten automatisierten Entscheidungen auftreten. So ist es heutzutage bei vielen Fluggesellschaften üblich, dass Sitzplätze automatisch zugewiesen werden. Auf den ersten Blick scheint dies nur viel Zeit und Arbeit zu sparen. Allerdings legt eine Studie der britischen Luftfahrtbehörde CAA nahe, dass bei der vorgeblich zufälligen, automatisierten Vergabe von Sitzplätzen, Zusammenreisende gezielt auseinandergesetzt wurden, damit diese sich dann gegen Entgelt zusammenhängende Sitzplätze kauften.[32] Während dieses Vorgehen für den einzelnen Flug nur eine geringfügige Bedeutung hat, ergibt sich durch die Gesamtzahl der ca. 4,1 Milliarden Menschen, die 2017 weltweit durch Fluglinien befördert wurden, eine wesentlich größere Auswirkung.[33]

IV. Ethisch-rechtliche Perspektive auf Bias in algorithmischen Systemen

Aus ethischen und rechtlichen Erwägungen heraus stellt sich zunächst die Frage, was „unerwünschte" Bias ausmacht. Unerwünschte Bias können eine unrechtmäßige, ungewollte oder ungerechtfertigte Differenzierung und demnach Diskriminierung darstellen. Ein algorithmisches System, welches gemäß dieser Definition diskriminierend ist, sollte deswegen unzulässig sein. Aus ethisch-rechtlicher Perspektive besteht dann eine Notwendigkeit nach Lösungen zu suchen, die den durch Bias im algorithmischen System ausgelösten Diskriminierungen entgegenwirken.

Gesetze stellen einen ethischen Kompass einer Gesellschaft dar. In Übereinstimmung mit diesem Kompass muss eine ethische Leitlinie identifiziert werden, welche die gesellschaftlichen Anforderungen im Umgang mit algorithmischen Systemen berücksichtigt. Mit Hilfe dieser ethischen Leitlinie soll der bestehende gesetzliche Rahmen untersucht werden. Dort, wo sich bisher weiße Flecken finden, soll gegebenenfalls zusätzlicher Regelbedarf identifiziert werden.

These: Es bedarf nicht zwingend einer neuen gesetzlichen Regelung für algorithmische Systeme, sondern in erster Linie einer effektiveren Umsetzung des bestehenden Rechts.

Beschreibung: Für eine effektive Umsetzung kann von bestehenden Regelungen gelernt werden. Bereits nach Artikel 3 des Grundgesetzes darf niemand wegen seines Geschlechts, seiner Abstammung, seiner Rasse, seiner Sprache, seiner Heimat und Herkunft, seines Glaubens, seiner religiösen oder politischen Anschauungen benachteiligt oder bevorzugt werden. Niemand darf wegen seiner Behinderung benachteiligt werden. Somit ist ein algorithmisches System, welches durch einen der oben genannten Gründe eine Benachteiligung bedingt, unrechtmäßig. Eine systematische Bevorzugung (positive Diskriminierung) kann nur dann rechtmäßig und zulässig sein, wenn sie gewollt und gerechtfertigt ist. Es bedarf in Hinsicht auf algorithmische Systeme demnach lediglich einer Transformation in einfachgesetzliches Recht (durch Gesetze geregelt, nicht aber in der Verfassung verankert) mit verbindlichen Vorgaben. Im Einzelfall müsste dann überlegt werden, ob und welche Ausnahmetatbestände existieren.

32 flug-verspaetet.de (2019): Airlines setzen möglicherweise Familien gezielt auseinander; online verfügbar unter: https://www.flug-verspaetet.de/neuigkeiten/2018/11/29/airline-setzten-familien-auseinander (letzter Abruf: 02.02.2019).
33 DPA (2018): Weltweit mehr als 4 Milliarden Flugreisende; online verfügbar unter: https://www.handelsblatt.com/unternehmen/handel-konsumgueter/rekord-im-luftverkehr-weltweit-mehr-als-4-milliarden-flugreisende/20862546.html?ticket=ST-334738-6qoqA6ewIhEK-hkoIlgIC-ap1 (letzter Abruf: 02.02.2019).

Umgang: Die bestehenden rechtlichen Regelungen des Allgemeinen Gleichbehandlungsgesetzes (AGG) schützen Personen vor Diskriminierung. Wir empfehlen diese vor dem Hintergrund der technologischen Entwicklungen zu betrachten und auf allgemeingültige Regelungen zu algorithmischen Systemen zu übertragen. Das heißt, dass eine intensive Prüfung der Rechtslage stattfinden sollte. Insbesondere ist zu klären, wie Regelungen umgesetzt werden und wie gegebenenfalls notwendige Anpassungen oder Nachsteuerungen aussehen können. Zudem enthält das Bundesdatenschutzgesetz (BDSG) bereits eine erste Regelung zum sog. Scoring[34] sowie Vorgaben, die Bias durch Fokussierung auf bestimmte Daten vermeiden sollen. Wir empfehlen daher, diese Regelung auf die Übertragbarkeit auf andere Bereiche zu prüfen und dabei zu evaluieren, welche Bias tatsächlich unerwünscht sind und wie ein Schutz vor solchen Bias effektiv erreicht werden kann.

Beispiel: Bereits heute ist es unzulässig, Menschen, die sich auf einen Job bewerben, wegen ihrer Herkunft auszuschließen. Zulässig ist es indes, Vorgaben hinsichtlich der Sprachkenntnisse zu machen, wenn dies zum Beispiel eine Anforderung für einen Beruf ist (beispielswiese für die Tätigkeit des Übersetzens). Ein algorithmisches System dürfte also die Sprachqualität in einem Anschreiben auswerten, jedoch nicht von den Sprachkenntnissen auf die Herkunft schließen.

These: Der Umgang mit einem algorithmischen System soll davon abhängen, wie hoch das Diskriminierungs- und Schadenspotenzial ist.

Beschreibung: Eine Herausforderung besteht darin, das Diskriminierungs- und Schadenspotenzial zuverlässig zu ermitteln. Bei dessen Ermittlung soll insbesondere die Quantität der getroffenen Entscheidungen eine wesentliche Rolle spielen und ob davon Menschen direkt oder indirekt betroffen sind. Ebenso relevant ist die Abhängigkeit von dieser Entscheidung, beispielsweise wenn es keine Wechselmöglichkeit zu einem anderen Anbieter gibt.[35]

Wenn ein algorithmisches System ein geringes Diskriminierungs- und Schadenspotenzial besitzt, genügt eine Selbstregulierung. In diesen Fällen sollte jedoch ein Bias-Monitoring verpflichtend in das Qualitätsmanagement integriert werden. Dieses soll sowohl die Prognosequalität als auch das Ergebnis selbst zum Gegenstand einer eingehenden Betrachtung machen.

Umgang: Für den Fall, dass ein algorithmisches Entscheidungssystem ein hohes Diskriminierungs- und Schadenspotenzial besitzt, sollte eine externe, unabhängige Evaluation stattfinden. In sensiblen Bereichen, wo ein konstantes Risiko zur Diskriminierung besteht, sollte ein konstantes Monitoring eingeführt werden. Für dieses (externe) Monitoring sollte ein Verfahren gewählt werden, welches berechtigten Geheimhaltungsinteressen ausreichend Rechnung trägt (z. B. über In-Camera-Verfahren[36]).

Dadurch kann sowohl die Zulässigkeit bestimmter algorithmischer Systeme überprüft als auch deren Qualität näher beleuchtet werden. Dabei werden die korrekte Wahl und die Einbindung der Datengrundlage / Datenbasis untersucht. Das Algorithmen-Monitoring würde somit einen Mindeststandard für die Qualität algorithmischer Systeme darstellen. Der angelegte Mindeststandard könnte dabei auch abhängig vom Diskriminierungs- und Schadenspotenzial variieren.

Beispiele: Es sollte anhand des Diskriminierungs- und Schadenspotenzials festgelegt werden, durch wen und in welcher Regelmäßigkeit ein solches Monitoring erfolgt. Dabei besteht die Herausforderung sicherlich darin, dieses Diskriminierungs- und Schadenspotenzial zuverlässig zu ermitteln. Wenn ein algorithmisches System auf Grundlage vorherigen Nutzerverhaltens ein gleichartiges Produkt im Onlineshop zum Kauf vorschlägt (beispielsweise ein T-Shirt statt einer Jacke), ist das Diskriminierung und Schadenspotenzial relativ gering. Da von einem Eigeninteresse des Unternehmens an einer korrekten Prognose ausgegangen werden kann, genügt in solchen Fällen eine Selbstregulierung.

34 Nach § 31 Bundesdatenschutzgesetzes die „Verwendung eines Wahrscheinlichkeitswerts über ein bestimmtes zukünftiges Verhalten einer natürlichen Person zum Zweck der Entscheidung über die Begründung, Durchführung oder Beendigung eines Vertragsverhältnisses mit dieser Person".

35 Zweig, Katharina (2019): „Black Box Analysen zur Kontrolle von ADM-Systemen", Vortrag in der Enquete-Kommission Künstliche Intelligenz des Deutschen Bundestages, 14.01.2019; online verfügbar unter https://www.bundestag.de/dokumente/textarchiv/2019/kw03-pa-enquete-ki/585354 (letzter Abruf: 14.02.2019).

36 Beim In-Camera-Verfahren werden Unterlagen durch bei Verwaltungsgerichten eingerichtete „Fachsenate für In-Camera-Verfahren" überprüft. Die vorlegten Unterlagen werden weder der Öffentlichkeit noch den Beteiligten der Streitsache bekannt gegeben oder zugänglich gemacht, auch nicht dem Gericht der Hauptsache. Sie verbleiben im Fachsenat, also „in der Kammer". Im In-Camera-Verfahren wird

In Bereichen, in denen eine Diskriminierung bestimmter Gruppen anhand eines algorithmischen Systems erwartet werden kann, wie zum Beispiel im Bereich von Bewerbungen und Bewertungen von Menschen, sollte ein konstantes Monitoring eingeführt werden.

These: Sowohl in Fällen der Selbstregulierung als auch für ein externes Monitoring bedarf es eines Mindeststandards, gegen den das konkrete algorithmische System geprüft werden kann.

Beschreibung: Nur auf diese Weise kann sichergestellt werden, dass ein algorithmisches System dem Mindeststandard genügt. Bei mäßigem Schadens- und Diskriminierungspotenzial könnte eine freiwillige Selbstverpflichtung nach dem Prinzip „Comply or Explain" ausreichen.[37] Darüber hinaus sollte die Prüfung vertraulicher algorithmischer Systeme durch eine Prüfinstanz erfolgen. Wichtig ist hierbei, dass Transparenz darüber geschaffen wird, welche konkreten Prüfungshandlungen durchgeführt wurden.

Umgang: Ein Mindeststandard algorithmischer Systeme könnte durch eine Zertifizierung unabhängiger Instanzen realisiert werden. Diese Instanzen könnten beispielsweise die Datenbasis des Systems, die Modellierung von zugrundeliegenden Variablen und die Entscheidungslogik (auf Bias-Belastung) der Systeme überprüfen.[38] Eine zusätzliche Option stellt ein Schulungsnachweis für all diejenigen, die das algorithmische System in den verschiedenen Phasen begleiten, dar. Weiterhin könnte eine Prüfung der Repräsentativität der Daten – welche die Basis lernender Algorithmen bilden – zusammen mit einer Input-Output-Analyse Aufschluss über die Qualität der Resultate geben. Hierbei würden Unternehmen in einem jährlichen Bericht Auskunft über die Einhaltung selbst-auferlegter Compliance-Regelungen für den Umgang mit algorithmischen Systemen geben. Bei Nichteinhaltung würde eine Prüfung durch Externe erfolgen.

Beispiele: Von den gesetzlichen Regelungen zum Scoring, die einen Mindeststandard (ein „wissenschaftlich anerkanntes mathematisch-statistisches Verfahren") vorgeben, kann für das Monitoring von algorithmischen Systemen gelernt werden. Aufgrund der Komplexität der algorithmischen Verfahren sollten in diesem Kontext allerdings konkrete Mindeststandards entwickelt werden. Unzulässig ist es beispielsweise, ausschließlich vom Wohnort einer Person auf deren Kreditwürdigkeit zu schließen. Allerdings lässt das Gesetz zu, dass der Scorewert zu einem (sehr) hohen Anteil auf dem Wohnort beruht, solange sich wissenschaftlich belegen lässt, dass die Nutzung dieser Daten zu einer zutreffenden Aussage über die Kreditwürdigkeit führt.

festgestellt, ob die Behörde die Unterlagen zu Recht geheim halten darf. (Quelle: https://de.wikipedia.org/wiki/In-Camera-Verfahren, letzter Abruf 14.02.2019).

37 ICSA (2018): Comply or explain is vital for UK governance; online verfügbar unter: https://www.icsa.org.uk/knowledge/governance-and-compliance/features/comply-explain-uk-corporate-governance-code (letzter Abruf: 02.02.2019).

38 Krüger, Julia (2018): Wie der Mensch die Kontrolle über den Algorithmus behalten kann; online verfügbar unter: https://netzpolitik.org/2018/algorithmen-regulierung-im-kontext-aktueller-gesetzgebung/ (letzter Abruf: 02.02.2019).

V. Ausblick

Vielen Menschen sind die Begriffe aber auch die Auswirkungen rund um algorithmische Systeme noch nicht vertraut.[39] Dennoch werden immer mehr Lebens- und Arbeitsbereiche zunehmend durch diese Systeme geprägt. Forderungen nach informierter Einwilligung und digitaler Teilhabe können nur erfüllt werden, wenn den beteiligten Menschen die Auswirkungen bekannt und bewusst sind. Es liegt entsprechend in der Verantwortung von Fachleuten und Entscheidern, aufklärend tätig zu werden.

Die in diesem Denkimpuls vorgestellten Bias in menschlichen Entscheidungen und ihre Auswirkungen für algorithmische Systeme erfordern Maßnahmen in ethisch-rechtlichen, sozioökonomischen und technologischen Bereichen. Im Denkimpuls werden Vorschläge für den konkreten Umgang genannt, die es nun zu diskutieren gilt. Wir empfehlen, existierende algorithmische Systeme vor diesem Hintergrund zu evaluieren und damit gleichzeitig die Wirksamkeit der vorgeschlagenen Maßnahmen zu prüfen.

Neben der Schwerpunktbetrachtung zum Thema Bias wird die Unterarbeitsgruppe Algorithmen-Monitoring in weiteren Arbeitspapieren die Transparenz bzw. Nachvollziehbarkeit von algorithmischen Systemen und die Frage der Verantwortung bei (teil)automatisierten Entscheidungen bearbeiten.

39 Initiative D21 e.V. (2019): D21-Digital-Index 2018 / 2019, das jährliche Lagebild zur Digitalen Gesellschaft; online verfügbar auf https://initiatived21.de/publikationen/d21-digital-index-2018-2019/ (letzter Abruf: 14.02.2019).

Initiative D21

Kooperationspartner:

KPMG

Die Unterarbeitsgruppe Algorithmen-Monitoring

Algorithmen bergen ein immenses Potenzial, insbesondere kommt ihnen eine wachsenden Bedeutung bei technologischen Entwicklungen zu. Gleichzeitig entsteht eine zunehmende Komplexität und Intransparenz von algorithmischen Systemen. Dies bringt steigende Herausforderungen und verschiedene Fragestellungen mit sich. Vor diesem Hintergrund hat die Initiative D21 Anfang 2018 eine Unterarbeitsgruppe (UAG) der AG Ethik zur Bearbeitung von Fragestellungen rund um das Thema „Algorithmen-Monitoring“ gegründet.

Die UAG Algorithmen-Monitoring diskutiert die relevanten Fragestellungen mit Expertinnen und Experten aus drei Perspektiven: ethisch-rechtlich, sozioökonomisch und technologisch. Dabei bezieht sich die technologische Perspektive auf die praktische Umsetzbarkeit eines Algorithmen-Monitorings und setzt sich mit den Bedingungen, Problemen und Möglichkeiten auseinander. Die sozioökonomische Perspektive arbeitet heraus, welche sozialen und ökonomischen Chancen und Herausforderungen durch die Anwendung von algorithmischen Systemen entstehen und wie man den Herausforderungen gegebenenfalls entgegenwirken kann. Die ethisch-rechtliche Perspektive behandelt die Erschließung einer rechtlichen Grundlage, welche die Regulierung algorithmischer Systeme ethisch vertretbar sichert.

Das Ziel der UAG Algorithmen-Monitoring besteht darin, für die drei Schwerpunktthemen „Bias in algorithmischen Systemen“, „Transparenz bzw. Nachvollziehbarkeit“ und „Verantwortung im Umgang mit algorithmischen Systemen“ Thesen zu definieren und die Diskussionen zu Empfehlungen zusammenzufassen. Diese Empfehlungen sollen Vorschläge dazu enthalten, welche Regulierungen algorithmischer Systeme ethisch erforderlich sein könnten, wie sich diese gesellschaftlich und wirtschaftlich auswirken und wie sie technologisch umsetzbar wären.

ALGORITHMEN MONITORING

Impressum

Initiative D21 e.V.
Reinhardtstraße 38
10117 Berlin
www.InitiativeD21.de

Telefon: 030 5268722-50
kontakt@initiatived21.de

Download
initiatived21.de/publikationen/denkimpulse-zur-digitalen-ethik