Algorithmen, KI und soziale Diskriminierung

Ilona Horwath

1. Einleitung

> „Der Algorithmus diskriminiert nicht“ “New AI can guess whether you’re gay” “Apple Card Investigated After Gender Discrimination Complaints” “Google apologises for Photos app’s racist blunder” “Google Search thinks the most important female CEO is Barbie” „Microsoft Twitter-Nutzer machen Chatbot zur Rassistin“ „Alexa ist nicht mehr deine Schlampe“ “Why Is Silicon Valley So Awful to Women?” “Artificial Intelligence’s White Guy Problem” “UK ditches exam results generated by biased algorithm after student protests. Protesters chanted ‘Fuck the algorithm’ outside the country’s Department for Education” “After Google drama, Big Tech must fight against AI bias” “A thousand young, black men removed from Met gang violence prediction database” „Wenn Algorithmen Vorurteile haben“ “Design AI so that it’s fair”[1]

Dies ist nur eine kleine Auswahl an Schlagzeilen, die Fragen aufwerfen, inwiefern digitale Systeme, Daten und Künstliche Intelligenz (KI) Biases aufweisen und sexistisch oder rassistisch diskriminieren. Dass sie dies können, haben inzwischen hunderte Fälle gezeigt,[2] die auch zunehmend Gegenstand wissenschaftlicher Forschung und

1 Die Zeit online, 2018; The Guardian, 2017; The New York Times, 2019; BBC, 2015; The Verge 2015; Die Zeit Online, 2016; Die Zeit Online, 2018; The Atlantic, 2017; The New York Times, 2016; The Verge 2020, Financial Times 2021; The Guardian 2021; Süddeutsche Zeitung, 2016; Zou & Schiebinger 2018

2 Crawford 2021: 128

Kritik werden.[3] Quantität und Qualität dieser „Fehler“ weisen darauf hin, dass Sexismus, Rassismus und klassenbasierte Diskriminierung in die Architektur und Funktionsweise der technischen Systeme eingeschrieben ist. Technologien werden nicht im gesellschaftsfreien Raum entwickelt und angewandt, diskriminierungsrelevante Probleme sind in die Gesellschaft und in technische Infrastrukturen eingelagert, die sich wechselseitig beeinflussen. In diesem Beitrag soll der Frage nachgegangen werden, inwiefern datengetriebene und algorithmische Systeme die Un-/Gleichheit entlang der Kategorien gender, race and class befördern. Ein Schwerpunkt wird auf die Rolle von algorithmischen Systemen, KI und dafür erforderliche Daten gelegt. Auf der Grundlage des Konzeptes der *„gendered substructure“*[4] von Organisationen und Gesellschaften werden aktuelle Entwicklungen aufgegriffen und diskutiert. Den Abschluss bildet ein Ausblick auf Handlungsoptionen.

Um Diskriminierung durch technische Applikationen und Artefakte angemessen analysieren zu können, ist es erforderlich, sie im sozialen Kontext zu betrachten und als soziotechnische Systeme zu verstehen.[5] Es macht auch einen Unterschied, ob soziotechnische Systeme, wie Daten, Algorithmen und KI im Kontext naturwissenschaftlicher Phänomene entwickelt und angewendet werden, oder ob auch Menschen und soziale Institutionen eine (aktive) Rolle spielen.[6] In diesem Beitrag geht es um die Anwendung soziotechnischer Systeme auf Menschen und soziale Institutionen. Dabei ist immer zu beachten, dass soziotechnische Systeme in konkrete, intersektionale und soziohistorische Kontexte sowie soziale Beziehungen eingebettet sind.[7]

Digitale soziotechnische Systeme wie algorithmische Entscheidungssysteme oder KI versprechen Effektivität, Objektivität und Autonomie. Um die Systeme entwickeln und betreiben zu können, werden große Datenmengen benötigt. Entwicklung und Anwendung erfolgen Großteils in global aktiven, neoliberal-kapitalistischen Unternehmen, welche die Digitalisierung maßgeblich vorantreiben. Eine soziotechnische Perspektive fragt nach der co-construktiven Verwobenheit dieser und weiterer sozialer, und der technischen Dimensionen.[8] Dabei müssen auch die soziomateriellen Kontexte die-

3 U.a. Crawford 2021, D’Ignacio & Klein 2020, Campolo & Crawford 2020, Criado Perez 2019, Noble 2018, Eubanks 2017, O’Neill 2016, Barocas & Selbst 2016

4 Acker 1990

5 Draude et al. 2020

6 Campolo & Crawford 2020

7 Noble 2018: 13

8 Draude et al. 2020

ser Systeme einbezogen werden, um die (Re-)Produktion sozialer Ungleichheiten und Ausbeutungsverhältnisse erkennen zu können.[9]
Natürlich sollen auch Veränderungspotentiale zum Positiven – mehr soziale Gerechtigkeit – thematisiert werden. Allerdings werden aktuell in feministischer und kritischer Forschung[10] viele empirische Phänomene beschrieben, die hochproblematische Auswirkungen auf gender, race und class haben. Um diese in den Blick zu nehmen, greift der Beitrag das Konzept der „gendered organizations“[11] und die Metapher der „inequality regimes“[12] auf. Diese beziehen sich zwar primär auf „work organizations“,[13] können aber in einer breiteren analytischen Perspektive auf Prozesse gesellschaftlicher Organisation von Hierarchisierung, Stereotypisierung und Marginalisierung angewendet werden, aber auch Veränderungen und Abbau sozialer Ungleichheit erfassen.[14] Der Rückgriff auf gendered substructure und inequality regimes als Dimensionen der (Re-)Produktion sozialer Ungleichheit in Organisationen und Gesellschaften ermöglicht, Strukturen, Prozesse und Praktiken soziotechnischer Systeme, die die digitale Transformation grundlegend organisieren, in ihrer Wirkung auf soziale (Un-) Gleichheit zu reflektieren.

2. Daten, Algorithmen und KI

Algorithmen sind *„informatische Werkzeuge, um mathematische Probleme automatisiert zu lösen“*[15]. Sind Algorithmen in der Lage, sich selbst zu verbessern, d.h. ihre Regeln oder Parameter, dann spricht man auch von „lernenden Algorithmen“ oder Maschinellem Lernen.[16] Maschinelles Lernen stellt die Grundlage für KI dar.[17] Bei lernenden Systemen haben nicht nur Entwickler*innen großen Einfluss, sondern auch die (Trainings-)Daten, mit welchen die Systeme entwickelt und betrieben werden. Hier

9 Crawford 2021, Noble 2019, Zuboff 2018, vgl. auch D’Ignacio & Klein 2020, Eubanks 2017, O’Neill 2016
10 Vgl. z.B. Crawford 2021, D’Ignacio & Klein 2020, Noble 2019
11 Acker 1990
12 Acker 2006
13 Acker 2006
14 Acker 2011, 1990
15 Zweig 2018: 10
16 Sachverständigenkommission 2020: 21
17 Crawford 2021

werden große, diverse Datensätze benötigt, die aus verschiedenen Quellen, wie Social Media, Smartphones, Laptops usw. gewonnen werden. Die Verfügbarkeit und Qualität der Daten ist entscheidend, denn *„Training data (…) is the foundation on which contemporary machine learning systems are built.(…)These datasets shape the epistemic boundaries governing how AI operates and in that sense, create the limits of how AI can "see" the world."*[18]. KI als soziotechnische Systeme beinhalten soziale und technische Praktiken, Infrastruktur und Institutionen, Politik und Kultur und sie *„(…) both reflect and produce social relations and understanding of the world"*[19].

Algorithmen suchen in großen Datenmengen nach statistisch auffälligen Mustern („Data Mining"). Die Qualität der Daten spielt dabei eine wichtige Rolle. So können Algorithmen anhand von Trainingsdaten Diskriminierung lernen. Wenn zum Beispiel die Trainingsdaten für Recruitmentsoftware diskriminierende Personalentscheidungen aus der Vergangenheit beinhalten, lernt der Algorithmus, vergangene Diskriminierung in seinen Prognosen und Entscheidungen zu reproduzieren. Dies wird immer schwerer erkennbar, je komplexer Algorithmen werden. Was genau im Inneren eines solchen Systems passiert, ist bei komplexen soziotechnischen Systemen selbst für die Entwickler*innen nicht mehr nachvollziehbar. Dennoch geht mit Deep Learning Verfahren und algorithmengestützten Entscheidungen die Erwartung einher, dass diese „objektiver" und effizienter sind und damit weniger diskriminieren als menschliche Entscheidungen.[20]

Datengetriebene Systeme, Algorithmen und KI werden zunehmend in verschiedenen gesellschaftlichen Bereichen eingesetzt, um Prozesse „effektiver" zu gestalten und Entscheidungen zu unterstützen, z.B. Recruitment und Personalauswahl, Vergabe von Sozialleistungen, Vergabe von Studienplätzen, in der Kriminalitätsbekämpfung und im Gesundheitsbereich, sowohl in staatlichen Institutionen als auch in privaten Unternehmen.[21] Es ist daher wichtig zu fragen, wie sich in diesem Erwartungsfeld objektiver Entscheidungen und treffsicherer Prognosen Prozesse der Diskriminierung und Herstellung von sozialer Ungleichheit entlang der Dimensionen gender, race, and class ausmachen lassen, wie sie zustande kommen und welche Folgen sie zeitigen.

---

18 Crawford 2021: 98
19 Crawford 2021: 8
20 Campolo & Crawford 2020, Eubanks 2017
21 Vgl. AlgorithmWatch gGmbH und Bertelsmann Stiftung 2019

3. A Gendered Substructure of Data, Algorithms, and AI

> "All organizations have inequality regimes, defined as loosely interrelated practices, processes, actions, and meanings that result in and maintain class, gender, and racial inequalities within particular organizations (…)The ubiquity of inequality is obvious."[22]

Angesichts der historischen und gesellschaftlichen Verbreitung von Stereotypen, sozialer Ungleichheit und struktureller Diskriminierung wäre es geradezu vermessen zu erwarten, dass digitale Systeme frei davon sind. Damit ist aber gerade nicht gesagt, dass Diskriminierung auf rein sozialen Strukturen beruht. Vielmehr sind materielle und symbolische Wechselwirkungen zu analysieren, um die vermeintliche „Unausweichlichkeit" sozialer Ungleichheit hinterfragen und bekämpfen zu können und die Verantwortung beim Einsatz soziotechnischer Systeme greifbar machen und einfordern zu können. Mit dem Konzept der „gendered substructure" lassen sich komplexe Prozesse hierarchischer Vergeschlechtlichung erfassen, die in Organisationen und Gesellschaft oft unsichtbar ihre Wirkung entfalten und die Re-Produktion sozialer Ungleichheiten befördern. Geschlecht wirkt in diesen Prozessen intersektional, d.h. in Verwobenheit mit weiteren Kategorien sozialer Differenzierung wie Alter, Klasse und soziale Herkunft oder sexuelle Orientierung. Die Analyseperspektive der „gendered substructure" umfasst mehrere Dimensionen, die nachfolgend auf datengetriebene soziotechnische Systeme, Algorithmen und KI bezogen werden: a) Strukturen und die Konstruktion von Trennlinien entlang von gender, class und race; b) die Konstruktion von Symbolen und Bildern, die die Unterscheidungen und Trennungen zum Ausdruck bringen, reifizieren oder auch entkräften; c) Interaktionen, wobei im Kontext der Digitalisierung auch technisch vermittelte Interaktionen beachtet werden sollen; d) vergeschlechtlichten Komponenten individueller Identitäten, die durch a, b und c produziert werden mit dem Ergebnis von (c) Gender als fundamentalem Bestandteil der beständig ablaufenden Prozesse der Entwicklung und Konzeption sozialer Strukturen. Der Einsatz der soziotechnischen Systeme wird nicht unabhängig von den vorhandenen gesellschaftlichen Bedingungen vollzogen, im Gegenteil: die Systeme entwickeln sich in enger Verzahnung mit den gesellschaftlichen und organisationalen gendered sub-

22 Acker 2006: 443

structures und inequality regimes. Wie wirken Algorithmen und KI in diesen Feldern und Prozessen? Wie sind die Strukturen gelagert und wie die Konzepte und Erklärungen, die über die Strukturen hervorgebracht werden?

4. Strukturen und Trennlinien entlang von gender, race und class

Die erste Dimension der „gendered organization" bezeichnet Strukturen und die *„construction of divisions along the lines of gender, class and race"*[23], z.B. Arbeitsteilungen, Verhaltensnormen, physische Räume und Macht sowie die institutionalisierten Mittel, um diese Unterteilungen aufrecht zu erhalten, auch in den Strukturen von Arbeitsmarkt, Familie und Staat. Die Trennlinien verlaufen innerhalb und entlang der Entwicklung und Gestaltung soziotechnischer Systeme sowie der Klassifikationen und Bewertungen, die sie vornehmen. Beginnen wir mit den „work organizations". Ein Blick auf die vertikalen und horizontalen Unternehmensstrukturen der großen digitalen Player Facebook, Google, Amazon, Apple und Microsoft zeigt, sie sind vorrangig *„young, white, male"*[24]. Es ist bekannt, dass die Tech-Branche ein Diversity Problem hat (Noble 2018), das sich, wie das Beispiel Informatik und KI in den USA zeigt, tendenziell nicht bessert sondern verschärft.[25] Die neu entstehenden „Experten"-Disziplinen wie KI oder Data Science sind mathematisch-technischer Herkunft und speisen sich aus Feldern, in denen Frauen nach wie vor eine Minderheit darstellen[26] und die sich vielfach durch sexistische Arbeitskulturen auszeichnen.[27] Frauen und andere marginalisierte Gruppen haben damit nur eingeschränkten Zugang zur Gestaltungsmacht soziotechnischer Systeme und ihrer Anwendung. Dies führt dazu, dass ihre Lebensbedingungen und Interessen nicht ausreichend vertreten werden.

Diskriminierung ist dann nicht unbedingt intentional, sondern auch eine Folge des sogenannten *„privilege hazard"*[28]. Für privilegierte Gruppen sind Benachteiligungen auf der Basis von gender, race und class oft „unsichtbar", werden als „neutrale" oder legi-

23 Acker 1990: 146
24 D'Ignazio & Klein 2020: 183
25 D'Ignacio & Klein 2020, Crawford 2021
26 boyd & Crawford 2012
27 Sachverständigenkommission 2021; D'Ignacio & Klein 2020
28 D'Ignacio & Klein 2020: 28

time Unterschiede wahrgenommen[29] oder schlicht als *„other people's problems"* ignoriert.[30] Dies ist besonders problematisch auf Grund der Reichweite digitaler Systeme:

> „The problems of gender and racial bias in our information systems are complex, but some of their key causes are plain as day: the data that shape them, and the models designed to put those data to use, are created by small groups of people and then scaled up to users around the globe."[31]

Trennlinien verlaufen auch innerhalb von Datensätzen. Trotz der Unmenge an Daten, die aus allen Bereichen des sozialen Lebens über devices wie Smartphones, Laptops, Internet of Things usw. extrahiert werden, fehlen vielfach relevante und valide Daten über Frauen und Minderheiten.[32] So werben Unternehmen zum einen damit, einen *„pregnancy prediction score"* über ihre User*innen berechnen zu können, während es andererseits bis zum Maternal Deaths Act von 2018 keine systematische Datensammlung zu den bei schwarzen Frauen häufiger vorkommenden Schwangerschaftskomplikationen gab.[33] Datensätze bzw. „fehlende Daten" spiegeln immer gesellschaftliche Machtverhältnisse wieder: *„Power imbalances are everywhere in data science: in our data sets, in our data products, and in the environments that enable our work."*[34] Mit Blick auf Algorithmen und KI kommt Daten eine hohe Bedeutung zu.[35] Fehlen Daten bzw. Diversität in Datensätzen, kann es den Gebrauch von Technologien je nach Einsatzfeld unzugänglich, unbrauchbar oder gefährlich machen.[36] Die Konsequenzen sind je nach System unterschiedlich, sie können Trennlinien herbeiführen, die von einer Zutrittsverweigerung zum Fitnessstudio bis zur Einstufung als terroristische Gefahr reichen.[37]

---

29 Acker 2011, 2006
30 D'Ignacio & Klein 2020: 31
31 D'Ignacio & Klein 2020: 28
32 Criado Perez 2019
33 D'Ignazio & Klein 2020
34 D'Ignacio & Klein 2020: 201
35 Crawford 2021, D'Ignacio & Klein 2020, Zou & Schiebinger 2018
36 Criado Perrez 2019
37 Sachverständigenkommission 2021; D'Ignacio & Klein 2020, Mayer-Schöneberger & Cukier 2013

Dem *„gender data gap"* und der Problematik, dass nicht zählt, was nicht gezählt wird[38], steht das *„paradox of exposure"*[39] gegenüber: Sozial marginalisierte Gruppen, die von angemessener Repräsentation enorm profitieren würden, sind zugleich mit großen Risiken der (Miss-)Repräsentation in Datensätzen konfrontiert. Dies betrifft zum Beispiel LGBTQ Menschen, die in binären Geschlechternormen gemessen und klassifiziert, oder Menschen, die übermäßiger digitaler Überwachung ausgesetzt werden.[40] Die mit Klassifikation verbundenen Risiken der Hierarchisierung und Unterdrückung sind für bereits marginalisierte Gruppen wesentlich höher als für sozial etablierte und privilegierte Gruppen. Letztgenannte können sich eher Interventionen und Widerstand gegen Miss-Repräsentation und diskriminierende Effekte leisten.[41]

Inzwischen gibt es Ansätze, bias in Datensätzen und algorithmisch generierten Ergebnissen zu quantifizieren, zu qualifizieren und zu beheben.[42] Ein Versuch, fairere Datensätze zu generieren, wurde bei ImageNet vorgenommen: Diskriminierende und beleidigende Personenkategorien sollten aus der Datenbank entfernt werden, mit dem Ergebnis dass ca. 56% der Personenkategorien inklusive der entsprechenden Bilder bereinigt werden mussten[43] – was auf die Dimension der Problematik verweist. Eine kritische Reflexion des Kontextes, der Qualität und Validität ihrer Inhalte und der Grenzen von Datensätzen gehört aktuell nicht unbedingt zum „state of the art" von KI und Data Science – hier wird oft nur auf die Größe („Big Data") abgestellt.[44]

Diversität in Datensätzen und die Bereinigung von Verzerrungen sind wichtige Ansätze, reichen aber nicht aus, wenn keine Fragen nach den größeren Zusammenhängen von Unterdrückungsstrukturen[45] und struktureller Diskriminierung gestellt werden.[46] Diese Phänomene sind komplex, während Daten, die immer potentiell vielfache Bedeutungen, unlösbare Fragen und Widersprüche, komplexe Ambiguitäten beinhalten, bei der Klassifikation auf wenige, „eindeutige" Bedeutungen vereinfacht

38 D'Ignacio & Klein 2020: 34
39 D'Ignacio & Klein 2020: 105
40 D'Ignacio & Klein 2020, Sachverständigenkommission 2021
41 Noble 2018: 26
42 Rekabsatz et al. 2020, Draude 2020, D'Ignazio & Klein 2020, Bellamy et al. 2018
43 Crawford 2021: 141f.
44 D'Ignazio & Klein 2020, boyd & Crawford 2012
45 Crawford 2021, Noble 2018
46 D'Ignazio & Klein 2020: 55

und reduziert werden.[47] Obwohl Klassifikation nicht gleichbedeutend mit Unterdrückung ist,[48] ist die Gefahr der Stereotypisierung und Diskriminierung groß, wenn die sozialen Kontexte soziotechnischer Systeme ausgeklammert werden.[49] Die Praxis des Klassifizierens geht mit einer Zentralisierung von Macht einher, „*the power to decide which differences make a difference*“[50].

Marginalisierte Gruppen haben ein höheres Risiko, digital überwacht, stereotypisierend kategorisiert und algorithmisch diskriminiert zu werden. Auf schwarze Menschen trifft dies in einem Ausmaß zu, dass inzwischen von *„technological redlining“* gesprochen wird.[51] *„Redlining“* bezeichnet eine historische Praxis von Banken, in der Kreditvergaben für Hausbesitzer nicht nach der individuellen Kreditwürdigkeit, sondern nach der Demografie ihrer Nachbarschaft, insbesondere *„race and ethnicity“* vorgenommen wurden[52] – eine Praxis, die heute in Form von algorithmisch generierten credit scores fortgesetzt wird, die konsistent rassistische biases aufweisen.[53] Viele Algorithmen nehmen Bewertungen vor, die Handlungsrisiken einschätzen sollen, zum Beispiel das Risiko, Kreditraten zurückzuzahlen oder die Rückfallswahrscheinlichkeit für Straftaten. Diese Logik des „risk assessment“ ist dem militärischen und privatwirtschaftlichen Entwicklungskontext digitaler Systeme entwachsen,[54] sie ist anfällig für strukturelle Diskriminierung und Stereotypisierung. Zum Beispiel kam es 2019 zu Vorwürfen, dass Apples Algorithmus einen gender bias aufweist und Frauen systematisch ein deutlich niedrigeres Kreditlimit als Männern zuweist.[55]

Ein sehr bekanntes Beispiel ist die im US Justizsystem verbreitete kommerzielle Software COMPAS. Diese berechnet einen Risiko Score für die Rückfallswahrscheinlichkeit individueller Straftäter*innen. Wie sich herausstellte, wurde schwarzen Männern ein höherer, weißen Männern ein niedrigerer Rückfallswert zugeteilt, mit anderen Worten, das System diskriminiert rassistisch.[56] Kommerzielle Unternehmen fühlen

47 Crawford 2021: 143
48 Noble 2018
49 D’Ignazio & Klein 2020, Draude et al. 2020, Sachverständigenkommission 2021
50 Crawford 2021: 132
51 Noble 2018
52 D’Ignazio & Klein, 2020: 50
53 D’Ignazio & Klein 2020: 52
54 Crawford 2021
55 The New York Times, 2019
56 Crawford 2021; weniger bekannt ist, dass die Software auch sexistisch diskriminierend wirkt, vgl. D’Ignazio & Klein 2020: 55f.

sich für Diskriminierung nicht zuständig, „*vendors and contractors have little incentive to ensure that their systems aren't reinforcing historical harms or creating new ones*“[57]. Viele der soziotechnischen Systeme sind neoliberal-profitorientierte Techprodukte – die je nach geopolitischem Kontext unterschiedlich reguliert werden. Werden die soziotechnischen Systeme in staatlich regulierten Bereichen zur Entscheidungsfindung eingesetzt, entfallen dabei mitunter etablierte Mechanismen der Kontrolle und Verantwortlichkeit.[58]

„Predictive Policing“ ist ein Bereich, für den soziotechnische Systeme inzwischen vielerorts genutzt werden. Damit einher geht eine Gefahr des *„Racial Profiling“*, der verstärkten Kontrolle bestimmter Bevölkerungsgruppen und Räume, wie das Beispiel der „gang violence matrix“ der Londoner Metropolitan Police illustriert:[59] Diese Datenbank erfasst mutmaßliche Gang Mitglieder und Personen, die mit Gewalt in Gang Milieus in Berührung kommen. Eine Studie von Amnesty International zeigte, dass sie neben Vorstrafen auch Informationen aus sozialen Medien nutzt. 2017 umfasste die Datenbank 3806 Namen, davon 1500 Namen von Personen ohne Vorstrafen, von denen die meisten junge schwarze Männer waren – die Einstufung erfolgte u.a. nach Kriterien wie Musikpräferenzen. Die Polizei nutzt die Matrix für vorausschauende Polizeiarbeit und Prävention, die Namen werden auch mit anderen Behörden wie Schulen und Sozialeinrichtungen geteilt. Es sind Fälle bekannt, in denen die Weitergabe dieser Informationen die Entscheidungen von Behörden beeinflusst hat. Personen wurden über ihre Einstufung nicht informiert. Schließlich mussten in Zusammenhang mit Datenschutzverstößen tausend Namen aus der Gang Matrix entfernt werden.[60] Typisch an diesem Beispiel ist die Stereotypisierung, Überwachung und Kriminalisierung schwarzer, männlicher Identitäten, die nicht zuletzt das Risiko selbsterfüllender Prophezeiungen birgt, sie

> “(…) can lead to a feedback loop where those included in a criminal justice database are more likely to be surveilled and thus more likely to have more information about them included, which justifies further police scrutiny. (…)Inequity is not only deepened but tech-washed, justified by the systems that appear

57 Crawford 2021: 199
58 Crawford 2021: 199; Eubanks 2017
59 Kaltheuner & Obermüller 2018: 38ff.
60 The Guardian, 2021

> immune to error yet are, in fact, intensifying the problems of overpolicing and racially biased surveillance“.[61]

Ein ganz anderes Beispiel ist SAFElab:[62] in diesem Projekt wurden Twitter Daten genutzt um zu erforschen, wie Jugendliche in Chicago mit Gang-bezogener Gewalt umgehen. In einer Kooperation aus Forscher*innen, Sozialarbeiter*innen und vormals in Gangs involvierten jugendlichen Expert*innen wurden kontextuelle Social Media Analysen vorgenommen und ein Trainingsdatensatz als Grundlage für einen lernenden Klassifikationsalgorithmus entwickelt. Hier wurde der soziokulturelle Kontext genau analysiert und in die Entwicklung des Systems einbezogen, um Vorurteile und Diskriminierung zu vermeiden. Dieses Konzept setzt auf Wissenstransfer zwischen Forschenden und Betroffenen und auf die Ausbildung zivilgesellschaftlicher, sozialer Infrastrukturen als wesentlichem Bestandteil von soziotechnischen Systemen.
Schließlich soll noch ein für gendered organizations und inequality regimes zentrales Feld betrachtet werden, die „work organization“ im Schatten der großen Techunternehmen. Dass KI „effizient“ erscheint, hängt von viel „unsichtbarer“ menschlicher Arbeit,[63] vielfach unter prekären Bedingungen ab:

> „(…) the hard physical labor of mine workers, the repetitive factory labor on the assembly line, the cybernetic labor in the cognitive sweatshops of outsourced programmers. The poorly paid crowdsourced labor of Mechanical Turk workers, and the unpaid immaterial work of everyday users.“[64]

Die „unsichtbare Arbeit“ reicht von „Gratisarbeit“, die wir als User*innen soziotechnischer Systeme mit unseren täglichen Interaktionen verrichten (wir produzieren Daten und trainieren KI durch digitale Interaktionen), über neue crowdsourcing Plattformen wie Amazon Mechanical Turk bis zur Kinderarbeit beim Abbau von Rohstoffen. Digitale soziotechnische Systeme haben eine materielle Basis, deren Rohstoffe wie Kobalt für Lithiumbatterien in Smartphones und Laptops unter ausbeuterischen Bedingungen

61 Crawford 2021: 198
62 D’Ignazio & Klein 2020: 162f.
63 D’Ignazio & Klein 2020: 187ff.
64 Crawford 2020: 69

extrahiert werden.[65] Während Unternehmen ein vitales Interesse an der Unsichtbarkeit dieser Arbeits- und Ausbeutungsstrukturen haben, gibt es vermehrt Anstrengungen, diese sichtbar zu machen.[66] Die Trennlinie zwischen sichtbarer und „unsichtbarer" Arbeit verläuft entlang von gender, race and class: *„While the demographics of Silicon Valley tech workers remain steadily young, white and male, these global „ghost workers" are often older women of color, and always required to accept precarious labor conditions"*[67]. „Effektivität" bedeutet hier auch eine „Rationalisierung", in der bereits benachteiligte Gruppen das Nachsehen haben – so ging die Einführung automatisierter Systeme zur Versorgung mit Gesundheits- und Sozialhilfeleistungen in verschiedenen US-Bundesstaaten mit Jobverlusten erfahrener Sozialarbeiter*innen einher, ebenfalls zu einem erheblichen Teil schwarze Frauen.[68]

Was die Arbeitsmöglichkeiten von Frauen betrifft, wurden ambivalente Hoffnungen auf flexible Vereinbarkeitsmöglichkeiten von Erwerbs- und Reproduktionsarbeit und selbstbestimmte (Zu-)Verdienstmöglichkeiten in der neuen Plattformökonomie gesetzt. Digitale Plattformen zur Vermittlung von Arbeit existieren für die unterschiedlichsten Bereiche, z.B. Dienstleister wie Lieferando, Uber und nicht zuletzt digitale Dienstleistungen wie Softwaredesign, Bereinigung oder Kategorisierung von Daten.

Es hat sich gezeigt, dass algorithmische Systeme eine starke Überwachung und Kontrolle der Plattformarbeiter*innen ermöglichen. Zudem gibt es Hinweise, dass die für Plattformarbeitende und ihre Reputation zentrale Bewertung auch traditionell-stereotypen Mustern folgt, d.h. Frauen und schwarze Menschen weniger und schlechtere Bewertungen für ihre Arbeit erhalten. Auch entstehen neue Räume und Möglichkeiten für geschlechterbezogene Gewalt und sexuelle Belästigung am Arbeitsplatz, Einkommens- und Vergütungsunterschiede, fehlende Mitbestimmungsstrukturen und Interessensvertretungen.[69]

Algorithmengestützte Recruitmentverfahren sollen die Problematik von Vorurteilen bei Personalverantwortlichen lösen und den Aufwand von Auswahlprozessen reduzieren. Ob sich die Hoffnung erfüllen kann, hängt von mehreren Faktoren ab, u.a. der Modellierung der Lösung (z.B. Zielkriterien dafür, wer die bestgeeigneten sind), den verwen-

65 D'Ignacio & Klein 2020:184f., Crawford 2021
66 z.B. Crawford 2021, Crawford & Joler 2018
67 D'Ignacio & Klein 2020: 183
68 Eubanks 2017
69 Sachverständigenkommission 2021: 59-65

deten Trainingsdaten (z.B. was sind bisherige „gute Entscheidungen“, gibt es Diversität in den Datensätzen?). Je mehr Informationen der Algorithmus über eine Person hat, umso stärker kann er sie diskriminieren. Selbst wenn Variablen wie Geschlecht oder ethnischer Hintergrund explizit ausgeschlossen werden, können sie durch Metadaten oder korrelierende Variablen, z.B. Wohnort in segregierten Stadtteilen doch wieder als wichtiges Kriterium in Auswahl- und Entscheidungsprozesse einfließen.[70] Algorithmen und KI treffen also nicht automatisch vorurteilsfreie, objektive Entscheidungen – nicht zu diskriminieren ist wesentlich voraussetzungsvoller.

5. Konstruktion von Symbolen und Bildern

Die zweite Dimension der „gendered organizations“ bezeichnet die Konstruktion von Symbolen und Bildern, die die in der ersten Dimension manifesten Unterteilungen erklären, zum Ausdruck bringen, reifizieren oder manchmal diesen auch entgegenstehen, z.B. sprachlicher Ausdruck, Ideologien, Kulturen, Bekleidung, mediale Darstellungen, etc.. Hier bieten vor allem Social Media Plattformen breiten Raum für stereotype (Selbst-)Darstellungen, aber auch Möglichkeiten, zum Beispiel binäre Geschlechterstereotype aufzubrechen.[71] Schließlich sind Symbole wichtig, die die Daten, Algorithmen und KI, also die soziotechnischen Systeme selbst betreffen.

Im Jahr 2013 lancierte die UNO eine Kampagne gegen den weit verbreiteten Sexismus und die Diskriminierung von Frauen, die sich auch im Netz reproduziert, z.B. in der Funktion der Autovervollständigung der Google Suche. Eine Suchanfrage „Frauen sollten“ wurde vom Google-Algorithmus mit Aussagen wie „nicht studieren“, „keine Rechte haben“ oder „nicht arbeiten“ ergänzt.[72] Ergebnisse einer Bildersuche nach „black girls“ zeigten über Jahre hinweg vorrangig pornografische Darstellungen.[73] Bei der Jobsuche bekommen Frauen bei Google weniger Anzeigen in Zusammenhang mit gut bezahlten beruflichen Positionen angezeigt als Männer.[74] Auch weibliche Führungskräfte werden nicht angemessen repräsentiert: machen weibliche CEOs in den

70 Dewes 2015
71 Sachverständigenkommission 2021: 120
72 UN Women 2013
73 Noble 2018
74 Datta et al. 2015

USA etwa 27% aus, sind sie in Ergebnissen der Bildersuche nur mit 11% vertreten, zudem dominieren stereotype Darstellungen.[75] Eine weiße Braut wird von KI als solche erkannt, eine Braut aus Nordindien aber als Kostüm kategorisiert:[76] Eine automatische Beschlagwortung von Bildern weist Fotos von schwarzen Menschen die Bezeichnungen Affe und Gorilla zu.[77] Eine weitere Studie zeigte, dass die Google Suche nach Namen, welche afroamerikanisch klingen, häufiger mit Werbeanzeigen verknüpft sind, die einen Eintrag im Strafregister implizieren.[78] Der Chatbot Tay, der von Microsoft ins Netz gestellt wurde, um die Sprache junger Menschen zu lernen, eignete sich innerhalb von 24 Stunden derart extreme sexistische und rassistische Inhalte an, dass das selbstlernende Programm vom Netz genommen und viele seiner Tweets gelöscht werden mussten.[79]

Algorithmen lernen mit der Zeit, welche Einschaltungen die meisten Klicks bringen. Eine Perspektive, wonach Algorithmen „nur" Vorurteile von User*innen reproduzieren, ansonsten aber objektiv sind, greift allerdings zu kurz. Denn Algorithmen spiegeln auch die Werte, Normen und Strukturen der Werbepartner*innen, Kund*innen und algorithmischen Praktiken des Unternehmens. Suchergebnisse und Rankings werden zudem vielfach manipuliert, z.B. um Werbepartner*innen ganz oben auf Ergebnislisten zu platzieren oder verbotene Inhalte zu entfernen. Es gibt eine ganze „Schattenindustrie" zur „Optimierung" von Suchergebnissen.[80] Suchergebnisse sind daher „*deeply contextual and easily manipulated, rather than objective, consistent, and transparent, (…) they can be legitimated only in social, political, and historical context*"[81]. Dennoch *normalisieren* Suchergebnisse rassistische und sexistische Vorurteile:[82] Ergebnislisten werden als objektiv und glaubwürdig dargestellt und wahrgenommen.

Suchergebnisse und -vorschläge sind algorithmisch erzeugte Resultate mehrerer Faktoren inklusive der Popularität der eingegebenen Suchbegriffe", d.h., Häufigkeit, mit denen nach Begriffen gesucht wird, Ort und Zeitpunkt der Suche. Bei selbstlernenden Algorithmen sind die Faktoren nicht mehr vollständig nachvollziehbar. Aber selbst

75 Kay et al. 2015; die CEO Barbie findet sich ebenfalls unter den ersten Treffern, The Verge 2015
76 Zou & Schiebinger 2018
77 Noble 2018
78 Sweeney 2013
79 Zweig 2019
80 Noble 2018: 49
81 Noble 2018: 45
82 Noble 2018

explizite Eingriffe des Unternehmens – z.B. wann, wie und warum die pornografische Darstellung von „black girls“ im Laufe der Zeit doch noch geändert wurde, sind nicht transparent.[83] Google hat eine Monopolstellung, welche ins Internet gespeisten Informationen wann, wo und wie wieder auftauchen. Da staatliche bzw. öffentliche Strukturen zunehmend von kommerziellen, monopolisierten Organisationen abhängig sind, ist es wichtig, hier entgegenzuwirken.[84] Die Google Suchvervollständigung kann sich übrigens auch als selbsterfüllende Prophezeiung entfalten und ein bestimmtes Klickverhalten erst provozieren, z.B. mit angedeuteten Skandalen wie „Merkel schwanger“: *„dann kommt es zu einer selbsterfüllenden Prophezeiung: Was niemand suchte, wird zur beliebtesten Suchanfrage – einfach nur deshalb, weil es vorgeschlagen wird.“*[85]
Strukturelle und historisch gewachsene Diskriminierung schlägt sich in Suchanfragen, Nutzungsverhalten, Daten und schließlich den Algorithmen, Modellen und Ergebnissen nieder. Trennlinien wie problematische Repräsentationen von Frauen und anderen marginalisierten sozialen Gruppen, die in der Informationswelt geschaffen oder reproduziert werden, beeinflussen auch Wahrnehmungen, Verhalten und (Geschlechter-) Verhältnisse in der analogen sozialen Realität. So kann die exzessive Verbreitung von Stereotypen auf Social Media der Entwicklung diverser Rollenvorbilder entgegenstehen, z.B. indem sie die Berufswahl junger Menschen beeinflussen.[86]
Eines der wirkmächtigsten Symbole ist die Aura bzw. der Mythos der Objektivität[87] im Mainstreamdiskurs über Daten und insbesondere Big Data, Data Mining, algorithmische Verfahren und KI, der sich nicht zuletzt auf Grund der institutionellen Macht der großen Techunternehmen, z.B. ihrer exzessiven Publikationspolitik, hartnäckig hält.[88] Dabei hat die feministische Forschung bereits vor Jahrzehnten den *„view from nowhere“*, also die Vorstellung, dass sich die Produktion von Wissen unabhängig von konkreten Personen und ihrer sozialen Situiertheit vollzieht, als männlich geprägtes Ideal entlarvt, das die Tatsache verschleiert, dass Wissen immer in konkreten sozialen und politischen Kontexten entsteht und immer nur eine partielle Perspektive repräsentiert.[89] Für die öffentlichkeitswirksamen Objektivitätsdiskurse im Bereich der KI haben

83 Noble 2018
84 Noble 2018
85 Zweig 2019: 249
86 Sachverständigenkommission 2021
87 boyd & Crawford 2012
88 Zuboff 2018, Campolo & Crawford 2019
89 Haraway 1988

Campolo & Crawford (2020) in Anlehnung an Max Weber den Begriff „enchanted determinism“ eingeführt, „*claims about ‚superhuman‘ accuracy and insight, paired with the inability to fully explain how these results are produced“*. Dieser Diskurs produziert Technikoptimismus und „*shields the creators of these systems from accountability while its deterministic, calculative power intensifies social processes of classification and control“*[90]. Vermeintlich objektive Systeme wie algorithmische Verfahren, die große Intransparenzen aufweisen, bergen wie auch in der analogen Welt die Gefahr, in hohem Maß diskriminierungsanfällig zu sein, wenn sie ungeprüft eingesetzt werden.[91] Die Prüfung dieser Verfahren ist aus verschiedenen Gründen schwierig. Wenn Prüfungen vorgenommen werden, sind die Evaluationskriterien meist nicht ausreichend breit gesteckt: „*The focus on the ‚technical success‘ of a system in a discourse of enchanted determinism works to seal off its epistemological shortcomings and ethical problems”*.[92] Ein Problem ist demnach, dass die Systeme als technische und nicht als soziotechnische Systeme evaluiert werden. Eine in kleinem Rahmen gesteckte technische Funktionalität wird geprüft, nicht aber die Auswirkungen im sozialen System. Dies führt zu *„irrational rationalization“* [93]*:* Systeme, die außerordentlich detaillierte Informationen über die Lebensmuster der Menschen besitzen (die aus Smartphones, Laptops, Social Media, Sensornetzwerken etc. extrahiert werden), denen es aber an sozialem und historischem Kontext mangelt. Kontextualisierung stellt eine Voraussetzung für Validität und verantwortungsvollen Umgang mit den soziotechnischen Systemen dar.[94] Diese Problematik wird zunehmend durch Initiativen adressiert,[95] während kritische Data Science Ansätze wie *„auditing algorithms“* aufzeigen, wie Nutzen und Schaden soziotechnischer Systeme sozial ungleich verteilt sind.[96]

Interessant ist, dass etablierte (sozial-)wissenschaftliche Konzepte wie Klasse im Mainstream Diskurs um datengetriebene Systeme als nicht ausreichend/zu fehlerbehaftet und vereinfachend dargestellt werden, weil sie das individuelle Verhalten nicht angemessen berücksichtigen würden.[97] Digitale Daten, die wir im Alltag nebenbei pro-

90 Campolo & Crawford 2020: 1
91 Sachverständigenkommission 2021: 96
92 Campolo & Crawford 2020: 12
93 Campolo & Crawford 2020: 13
94 Sachverständigenkommission 2021, D’Ignazio & Klein 2020, Draude et al. 2020
95 Sachverständigenkommission 2021: 36
96 D’Ignazio & Klein 2020: 57; Eubanks 2017
97 Pentland 2013: 80

duzieren, werden demgegenüber als „honest signals“ verstanden, die „für sich selbst sprechen“ und für deren Analyse und Interpretation keine sozialwissenschaftlichen Kompetenzen mehr erforderlich seien. Es kommt also zu einer Verschiebung der Definitionsmacht von Sozialwissenschaften, die kritische Fragen über Datensätze, Validität und soziale Systeme stellen, in die mathematisch-naturwissenschaftlich geprägten Data Sciences.[98]

Daten sprechen aber nicht für sich selbst, sie müssen immer interpretiert werden.[99] Die Deutungsmacht verschiebt sich nicht nur, sie wird auch monopolisiert, dadurch dass primär kommerzielle Firmen die Kapazitäten aufweisen, Daten zu sammeln und datengetriebene Systeme zu entwickeln und einzusetzen. Das hat auch Auswirkungen darauf, wer Zugriff auf welche Ausschnitte dieser Daten hat, welche Fragen gestellt und wie sie beantwortet werden können; hier entstehen beträchtliche Asymmetrien.[100] Der ungleiche Zugang von Wissenschaftler*innen hat jedenfalls negative Auswirkungen auf die Objektivität, denn Ergebnisse können kaum noch validiert oder repliziert werden.

Grenzen von Datensätzen werden bei datengetriebenen Systemen meist nicht kritisch reflektiert. Aus dem Umstand, dass die Daten aus digitalen Interaktionen „nebenbei“ gewonnen werden (und nicht im Rahmen wissenschaftlicher Erhebungsformate), wird geschlossen, es handle sich um „natürliche“ Daten, obwohl z.B. aus social media gewonnene Daten hoch artifiziell sind. Die Nutzung user-generierter Daten tendiert dazu, mehr Wohlhabende, Gebildete, Weiße und Männer zu repräsentieren.[101] Außerdem sind „Aktionsformate“ auf soziotechnischen Plattformen vorformatiert, wie etwa liking, clicking, sharing, hyperlinks, hashtags oder auto-complete Funktionen. Digitale Plattformen unterwerfen soziale Interaktion bestimmten Logiken, z.B. sollen sie Vernetzung befördern, bieten viele Möglichkeiten zur Bewertung oder versuchen, uns zu möglichst vielen likes und shares zu bewegen. Aus dieser Perspektive zeichnen sich digitale Daten eher durch ihre „Künstlichkeit“ als durch ihre „Natürlichkeit“ aus.[102]

Algorithmen und KI leben von der Verfügbarkeit von Daten in großer Menge und Varietät. Vor diesem Hintergrund hat sich ein mächtiger Extraktionsimperativ entwickelt, der

98 boyd & Crawford 2012
99 D’ignazio & Klein 2020, Haraway 1988
100 boyd & Crawford 2012
101 boyd & Crawford 2012
102 Marres 2017: 52

sich metaphorisch im Ausdruck *„data, the new oil“* manifestiert: ein Glaube, dass die Welt aus Daten besteht und diese einfach extrahiert und verwendet werden dürfen.[103] Die *„Ideology of data extraction“*[104] verändert auch die Rolle alltäglicher User*innen für Technunternehmen: von Kund*innen zu Datenressourcen, die für den Profit von Werbekund*innen ausgebeutet werden.[105] Interaktionen von User*innen werden für die Datenextraktion instrumentalisiert „*This extractive system creates a profound asymmetry between who is collecting, storing, and analyzing data and whose data are collected, stored, and analyzed.“*[106]

Wieder sind marginalisierte Gruppen besonders von Datenextraktion und -verwendung ohne Konsens betroffen. So werden etwa Polizeifotos ohne Einverständnis (oder Wissen) der Betroffenen als Trainingsdaten für Gesichtserkennungssoftware verwendet.[107] IBM entwickelte einen experimentellen *„terrorist credit score“,* um ISIS Kämpfer*innen von geflüchteten Menschen zu unterscheiden – die Daten syrischer Geflüchteter wurden dafür ohne deren Wissen oder Konsens herangezogen.[108] Entgegen der Alltagsvorstellung, dass Menschen keine Probleme mit Datenextraktion haben, wenn sie im Gegenzug Gratisservices, Apps und Spiele nutzen können oder „data for good“ genutzt werden, gibt es tatsächlich sehr viel Widerstand gegen die Praktiken der Überwachung, Extraktion, Tracking, Targeting und unangemessener „Personalisierung“.[109]

Die Masse extrahierter Daten, Big Data, erlaubt genaue Verhaltensprognosen, die für Profitinteressen von Werbekund*innen genutzt werden.[110] In den dominanten Diskursen um data mining wird zwar eingeräumt, dass diese Praktiken Fragen nach Privacy/Datenschutz aufwerfen; diese werden jedoch als lösbar dargestellt angesichts des Versprechens, datengetriebene Systeme zur *„foundation of a healthier, more pros-*

103 boyd & Crawford 2012
104 Crawford 2021: 94; vgl. ausführlich: Zuboff 2019
105 Zuboff 2018
106 D'Ignazio & Klein 2020: 25
107 Crafword 2021. Das National Institute of Standards and Technology veranstaltet in Kooperation mit Intelligence Advanced Research Projects Activity sogar Wettbewerbe, in denen KI Forschende gegeneinander antreten, um unter Verwendung von Polizeifotos den „schnellsten und akkuratesten“ Gesichtserkennungsalgorithmus zu küren, Crawford 2021: 92.
108 Crawford 2021: 205
109 Zuboff 2018
110 Zuboff 2018

perous world" zu machen[111] und „data for the common good" zu nutzen. Im Gegensatz dazu stehen Ansätze wie „data for co-liberation" und Initiativen wie das „design justice network". Diese entwickeln Alternativen, um marginalisierte Gruppen in Designprozesse soziotechnischer Systeme aktiv einzubeziehen und soziale Gerechtigkeit als grundlegendes Designprinzip in der Praxis zu verankern.[112] Der Slogan „doing good with data" wird vor diesem Hintergrund kritisch herausgefordert:

> „‚doing good with data' requires being deeply attuned to the things that fall outside the dataset – and in particular how datasets, and the data science they enable, too often reflect the structures of power of the world they draw from. In a world defined by unequal power relations, which shape both social norms and laws about how data are used and how data science is applied, it remains imperative to consider who gets to do the 'good' and who, conversely, gets someone else's 'good' done to them."[113]

6. Interaktionen

Die dritte Dimension beschreibt Interaktionen zwischen Individuen, inklusive jener Muster, die Dominanz und Unterwerfung zwischen den Geschlechtern enaktieren, z.B. Unterbrechen, Gegenseitigkeit der Interaktion, Themenfestlegung in Diskussionen, Männer als Akteure und Frauen als emotionale Unterstützung.[114] Im Kontext der soziotechnischen Systeme Daten, Algorithmen und KI sollen nicht allein die unmittelbare Interaktion zwischen Individuen, sondern auch digital vermittelte Interaktionen zwischen Individuen und zwischen digitalen Anwendungen und User*innen betrachtet werden, wie z.B. virtuelle Assistent*innen oder Mikrotargeting zur Verhaltensbeeinflussung.

2017 berichtete der Australian über ein vertrauliches Dokument, in dem sich Facebook vor Werbekund*innen damit brüstete, über psychologische Einsichten von 6,4 Mio. junger Australier- und Neuseeländer*innen zu verfügen, die für Werbe- und Ver-

111 Pentland 2013
112 D'Ignazio & Klein 2020: 205f
113 D'Ignazio & Klein 2020: 47
114 Acker 1990

kaufszwecke manipuliert werden können.[115] Sprachassistentinnen sollen künftig Widerrede bei Beleidigungen und sexueller Belästigung leisten.[116] Seit 2017 wurde der Hashtag #MeToo millionenfach genutzt, um auf sexuelle Belästigung, Missbrauch und Übergriffe auf Frauen aufmerksam zu machen.

Wie angesprochen, sind bestimmte soziale Interaktionen in digitalen Umgebungen technisch vorkonstruiert. Hinzu kommen Verhaltenslogiken, die durch die Interessen der großen Tech-Unternehmen und ihrer Werbekund*innen Verhaltensformen in der digitalen Welt präformieren und vorantreiben. Technische Vorgaben und ökonomische Handlungslogiken sind digitalen Interaktionen eingeschrieben, oft mit dem Ziel, das Verhalten der User*innen zu manipulieren, z.B. mehr Zeit auf Plattformen zu verbringen, sich mit anderen zu vernetzen, jemanden „anzustubsen", zu liken, zu bewerten, Inhalte zu teilen und auf Werbebotschaften zu reagieren.[117] Zuboff spricht in diesem Zusammenhang von *„Aktionsvorteilen":* Sensoren, über die Daten erfasst werden, können gleichzeitig auch „Aktoren" sein, um ein bestimmtes Verhalten zu provozieren, „*Die wahre Macht liegt heute darin, Echtzeithandeln in der realen Welt zu modifizieren.*"[118]

Ziele und Operationen der *„automatisierten Verhaltensmodifikation"* werden von Unternehmen im Sinne ihrer Ertrags- und Wachstumsziele entworfen und kontrolliert auf Märkten, in denen User*innen nicht die Kund*innen, sondern die Verkaufsressource darstellen.[119] Ein Beispiel dafür ist das 2017 öffentlich gewordene Dokument von Facebook, das bei seinen Kund*innen damit wirbt, die emotionalen und psychologischen Schwächen seiner jugendlichen User*innen durch das Setzen von Werbeanreizen ausbeuten zu können.[120] Ein anderes Beispiel ist das Orchestrieren einer Situation aus der Ferne, wofür das Spiel Pokémon Go als *„Versuchsfeld für Fernstimulation im großen Stil"* verstanden werden kann: Um Verkaufsziele von Werbekund*innen zu erreichen, werden bestimmte Geschäfte oder Restaurants zu Anlaufstellen für Spieler*innen gemacht.[121]

---

115 Zuboff 2018: 348
116 Die Zeit Online 2018
117 Marres 2017
118 Zuboff 2018: 335
119 Zuboff 2018: 336
120 Zuboff 2018: 348
121 Zuboff 2018: 355f.

Viele dieser Interaktionmechanismen und Targeting-Methoden operieren unterhalb der Wahrnehmungsschwelle, sie umgehen also unser Bewusstsein.[122] Andererseits passen Menschen ihr Verhalten auch an die Logik algorithmischer Prozesse an, z.B. wann sie auf Facebook posten und wie Beiträge gestaltet werden, um maximale Aufmerksamkeit zu generieren; sie adaptieren ihr Verhalten gemäß der operativen Logik von Social Media Plattformen[123] eine Logik von likes, shares und rankings, die auch dazu führt, dass prominente Videos wie die „ice bucket challenge" wichtige andere Inhalte, z.B. Konflikte um rassistische Diskriminierung verdrängen.[124] Geschlechterspezifische Dimensionen der Verhaltensmodifikation sind noch kaum erforscht.

Die Interaktion mit virtuellen Assistent*innen wie SIRI oder ALEXA erfreut sich zunehmender Beliebtheit. Eine Analyse von Both (2014) zeigte auf, wie diese Assistentinnen vergeschlechtlichte Arbeitsteilung und Persönlichkeiten reproduzieren. Das Bestreben, „authentische" virtuelle Gefährt*innen zu entwickeln, führt zu Reifikation heteronormativer Geschlechterbeziehungen und nicht zu deren Dekonstruktion. Vorgeblich für alle User*innen entwickelt, repräsentieren sie männliche Berufs-, Lebens- und Erfahrungswelten, z.B. Konzentration auf Bedürfnisse von Geschäftsreisenden, Voraussetzung finanzieller Ressourcen etc.[125] Dem stehen Entwicklungen entgegen, so etwas wie „feministische KI" zu bauen – zumindest werden erste Schritte gesetzt, um negative und sexistische Vorurteile sowie sexuelle Belästigung und Gewalt gegen Frauen im Umgang mit diesen Assistent*innen nicht weiter zu bestärken.[126]

Sexualisierte und geschlechterbezogene Gewalt (hate speech, revenche porn…) ist in der digitalen Welt vielfältig und zahlreich vorhanden, was die Partizipation von Frauen be- bzw. auch verhindert.[127] Von Hate Speech betroffen sind v.a. Frauen, die öffentlich auftreten, z.B. sich beruflich exponieren, öffentlich engagieren oder mit geschlechterstereotypen Erwartungen brechen. Rassistische und sexistische Cybergewalt kann sich auch gegen Männer richten, v.a., wenn sie hegemonialen Männlichkeitsvorstellungen nicht entsprechen.[128] Feindseliger Sexismus und digitale Gewalt wirken als

122 Zuboff 2018
123 Bucher 2017, Marres 2017
124 Bucher 2017: 37
125 Both 2014
126 Die Zeit Online 2018
127 Sachverständigenkommission 2021: 124
128 Sachverständigenkommission 2021: 125f.

Platzanweiser und Mittel zur Stabilisierung tradierter Geschlechterrollen.[129] Diese Formen von Interaktion und Radikalisierung in sozialen Netzwerken diskriminieren vulnerable Gruppen, die sich dann auch aus Onlineumgebungen zurückziehen. Dies ist die Kehrseite des Umstandes, dass das Netz auch Möglichkeiten bietet, Marginalisierung und Gewalt entgegenzutreten, wie etwa der Erfolg der #MeToo Kampagne gezeigt hat.

7. Vergeschlechtlichte Komponenten der individuellen Identitäten

Die bisher besprochenen Dimensionen produzieren im Zusammenspiel vergeschlechtlichte Komponenten individueller Identitäten, z.B. die Wahl „geeigneter" Berufe, Sprachgebrauch, Bekleidung, und die Repräsentation des Selbst als vergeschlechtlicher Teil der Organisationen.[130] Wie bereits angesprochen, trifft dies auf jede Menge user*innengenerierten Content und Social Media Plattformen zu. An dieser Stelle soll jedoch das Phänomen der Prognose sozialer Identitäten aufgegriffen werden.
„Deep learning systems are at their most deterministic when they are applied to ascribe identity or other social characteristics from a set of inputs understood as signals"[131], z.B. die Vorhersage der sexuellen Orientierung anhand von Gesichtern.[132] Die „Gay Faces Study" von Kosinski & Wang erregte 2017 weltweites Aufsehen. Die Forscher hatten eine Software entwickelt, die vermeintlich sexuelle Orientierung von Gesichtern ablesen kann. Kurz darauf veröffentlichte der Psychologe Todorov gemeinsam mit KI Forschern von Google eine Reihe von Gegenstudien, die nachwiesen, dass Gesichtserkennungstechnologien stark auf Kriterien wie Gesichtsausdruck, Kopfhaltung etc. reagieren. Die Studien zeigten, dass jene Unterschiede, die Kosinski und Wang als Hinweise auf biologischen Determinismus interpretiert hatten, durch Differenzen in Pflegeroutinen, Selbstdarstellung und Lebensweise – also soziokulturelle Unterschiede und „doing gender" erklärt werden konnten (Kaltenheuner & Obermüller 2018).
In der kritischen KI Forschung werden Studien zur Prognose sozialer Identitäten und ihre Auswirkungen als hochproblematisch eingestuft: „*When these techniques*

129 Sachverständigenkommission 2021: 128, vgl. auch Rudman & Glick 2008
130 Acker 1990
131 Campolo & Crawford 2020: 10
132 Kosinski & Wang 2018

are introduced in social domains, they have the potential to intensify hierarchies and differences while closing them off to political debate, visibility, or accountability.“[133]
Die Schwierigkeit mit derartigen Klassifikationen besteht darin, dass Konzepte wie „Identität“, „gender“, „sexuelle Orientierung“ oder „Kriminalität“, die hochgradig von sozialem Kontext abhängige und relationale Phänomene darstellen, „vereindeutigt“ werden, d.h. auf wenige Bedeutungsdimensionen reduziert, deren Validität oft mehr als fragwürdig ist. Trotzdem wird die Praxis in Mainstreamdiskursen der KI kaum hinterfragt: *„the ideas that race and gender can be automatically detectable in machine learning is treated as an assumed fact and rarely questioned by the technical disciplines, despite the profound political problems this presents*.“[134]
Die Diskriminierungsanfälligkeit zeigt sich besonders in Überwachungstechnologien, *„where software code and a false sense of objectivity come together to contain and control the lives of Black People, and of other people of color“*[135]. Gesichtserkennung und biometrische Verfahren erkennen weiße Männer am besten, Frauen mit dunkler Haut am schlechtesten.[136] Überwachungstechnologische Systeme werden im staatlichen wie privaten Bereich beworben und eingesetzt. Falsche Identifikation kann, wie oben angesprochen, von Unannehmlichkeiten bis zu bedrohlichen Szenarien reichen. Obwohl seit Jahrzehnten im Einsatz, führten erst kritische Untersuchungen in den letzten Jahren dazu, dass Verzerrungen bei der Erkennung bestimmter Personengruppen, v.a. sexistische, rassistische und altersdiskriminierende Verzerrungen, getestet und adressiert werden.[137] Auch hier wirkt der „privilege hazard“: „*prototypical whiteness as well as proto-typical maleness, youth, and able-bodidness inscribes racial categories into surveillance technologies like facial recognition (…) and pervades machine learning*.“[138]
Die Beispiele machen deutlich, dass soziotechnische Systeme bestehende Diskriminierung automatisieren können, dies bestimmte Menschengruppen stärker betrifft und sie häufiger mit den Auswirkungen konfrontiert sind, während Hersteller- und Anwender*innen behaupten, dass Prognosen von Faktoren wie Hautfarbe oder Ge-

133 Campolo & Crawford 2020: 12
134 Crawford 2021: 144
135 Benjamin, zit. nach D’Ignazio & Klein 2020: 55
136 D’Ignacio & Klein 2020
137 Sachverständigenkommission 2021: 32
138 Campolo & Crawford 2020:14f.

schlecht unabhängig sind. Viele der soziotechnischen Systeme sind durch Geschäftsgeheimnisse geschützt, was es schwer macht, programmierten Vorurteilen und Diskriminierungen entgegenzuwirken. Die Gefahr in Zusammenhang mit KI ist nicht, dass die Systeme intelligenter als Menschen werden könnten, sondern dass sie Sexismus, Rassismus und andere Formen der Diskriminierung in der digitalen Infrastruktur der Gesellschaften festschreiben:[139] „*Sexism, racism and other forms of discrimination are built into the machine-learning algorithms that underlie the technology behind many ‚intelligent' systems that shape how we are categorized and advertised to.*"[140]

Als letztes Beispiel soll noch die Klassifikation und Prognose sozialer Identitäten zum Zwecke der Personalisierung und des targeted advertising angesprochen werden. Algorithmen folgen User*innen und ihren digitalen Interaktionen („tracking") mit dem Ziel, Profile über ihre Identitäten und Aktivitäten anzulegen, die dann für die Auswahl „relevanter Inhalte" und zielgenaue Werbebotschaften genutzt werden. Wie in der analogen Werbung auch, ist digitale Werbung und die Praxis des „algorithmic profiling" gespickt mit Stereotypen, z.B. demografische Klassifikationen mit „*stereotypical assumptions about what the typical middle aged women is like*"[141], welche Bedürfnisse und Vorlieben sie hat und welche Produkte sie am wahrscheinlichsten kaufen wird (und wann). Auch wenn User*innen Stereotype erkennen und Werbebotschaften als unpassend, irritierend oder ärgerlich empfinden können, wirken sie dennoch darauf ein, welche digitalen Inhalte wir überhaupt angeboten bekommen, z.B. die konsequente Bewerbung von Jojo Moyes Romanen, aber keinerlei feministische Literatur. Es ist schwierig, algorithmischer Kategorisierung zu entkommen oder diese zu modifizieren[142] – und das beeinflusst und begrenzt Erfahrungs-, Entfaltungs- und Entwicklungsmöglichkeiten.[143] „*The question is not just whether the categories and classifications that algorithms rely on match our sense of self, but to what extent we come to see and identify ourselves through ‚the eyes' of the algorithm?*"[144]

139 D'Ignacio & Klein 2020: 29
140 Crawford 2016
141 Buchner 2017: 34
142 Bucher 2017
143 Zuboff 2018
144 Buchner 2017: 34f.

8. Diskussion und Ausblick

Vorurteile und Diskriminierung sind historisch gewachsen und in gesellschaftliche Strukturen eingegossen. Strukturelle Diskriminierung manifestiert sich in unterschiedlichen sozialen und ökonomischen Lebens- und Entwicklungschancen, z.B. Bildungssystem, Arbeitsmarkt, Gesundheit. Es sind diese Bereiche, in denen Lebens- und Entwicklungschancen zusehends von digitalen soziotechnischen Systemen beeinflusst werden. Die historisch gewachsene Diskriminierung geht in Daten, Algorithmen und digitale Systeme ein, es kommt zu gravierenden Ungleichheitseffekten, die nicht von unmittelbar diskriminierenden Absichten und Handlungen von Akteur*innen ausgehen (müssen). Dennoch entfalten sie Wirkung, oft „unsichtbar", bis aufgezeigt werden kann, dass bestimmte Technologien für bestimmte Personengruppen nicht funktionieren, benachteiligend wirken und die Diskriminierung soziale Folgen hat.[145] Nun sind viele dieser Probleme nicht neu – neu ist aber die enorme (zeitliche, geografische und mikropolitische) Reichweite der digitalen soziotechnischen Systeme und ihr Potential zur Automatisierung sozialer Ungleichheit. Sie sind jedenfalls mit erheblichen Diskriminierungsrisiken verbunden.[146] Auch gut trainierte und „funktionierende" Software kann zu diskriminierenden Entscheidungen kommen, wenn sie Vorurteile in vergangenen Entscheidungen aus einem Datensatz gelernt hat, die Zielkriterien diskriminierend sind oder das System auf ein Feld angewendet wird, in dem strukturelle Diskriminierung wirkt.

Data Mining ermöglicht es, intendierte Diskriminierung – unter dem Deckmantel der Objektivität – zu verschleiern, z.B. durch den Ausschluss geschützter Merkmale; und erschwert es, nicht intendierte Diskriminierung aufzuspüren, nicht zuletzt auf Grund der Komplexität der Systeme, die Entwickler*innen oft selbst nicht mehr ganz nachvollziehen können.[147] Es sind konservative Systeme, die Muster aus Daten (gesammelte Erfahrungen aus der Vergangenheit) in gegenwärtige Entscheidungen einfließen lassen und in die Zukunft projizieren.[148] Aber selbstlernende Algorithmen projizieren nicht nur die Vergangenheit in die Zukunft, sie individualisieren auch treffsicher strukturelle Diskriminierung. So werden aggregierte Daten verwendet, um auf Basis der

145 Zweig 2019, Eubanks 2017
146 Sachverständigenkommission 2021: 91
147 Barocas & Selbst 2016
148 O'Neill 2016

Zugehörigkeit zu sozialen Gruppen Prognosen für Individuen zu erstellen, die deren Leben maßgeblich beeinflussen.[149] Zweig spricht in diesem Zusammenhang von algorithmisch legitimierten Vorurteilen.[150] Die in diesem Beitrag ausgeführten Beispiele machen deutlich, dass algorithmische Diskriminierung bereits marginalisierte Menschengruppen stärker betrifft und sie häufiger mit den Auswirkungen konfrontiert sind.[151]

Breite interdisziplinäre Kooperationen sind erforderlich, um soziotechnische Systeme kontextualisieren, bias bearbeiten und gerechtere Systeme entwickeln zu können.[152] Dafür braucht es Theorien und Konzepte der kritischen Sozialwissenschaften, der Sozialen Ungleichheitsforschung, der Feministischen Forschung und Erkenntnistheorien, der Critical Race Studies, der Critical Data Studies sowie das Einbeziehen von Aktivist*innen, und von Betroffenen in die Entwicklung der soziotechnischen Systeme[153] – Expertisen, die über digitale Technologie weit hinaus gehen, um Geschlechtergerechtigkeit, und soziale Gerechtigkeit aktiv als Zielsetzung in die Entwicklung digitaler soziotechnischer Systeme und in die Rahmenbedingungen für ihren Einsatz einzubringen.[154]

> „The histories of classification show us that the most harmful forms of human categorization – from the Apartheid system to the pathologization of homosexuality – did not simply fade away under the light of scientific research and ethical critique. Rather, change also required political organizing, sustained protest, and public campaigning over many years.“[155]

Technikunternehmen sehen die Verantwortung gerne ausschließlich auf Seiten der Gesellschaft. Wird bei soziotechnischen Systemen soziale Gerechtigkeit aber aktiv berücksichtigt, d.h. bei Planung, Gestaltung, Entwicklung, Einsatz und Evaluation digitaler Systeme angestrebt und umgesetzt,[156] dann können gesellschaftliche Diskrimi-

149 D'Ignazio & Klein 2020: 55
150 Zweig 2019: 229
151 Draude et al. 2020
152 Draude et al. 2020
153 Sachverständigenkommission 2021: 37; D'ignacio & Klein 2020, Draude et al. 2020
154 Sachverständigenkommission 2021
155 Crawford 2021: 149
156 Sachverständigenkommission 2021

nierungen mit technischen Mitteln auch sichtbar gemacht werden und Möglichkeiten, diese auszuräumen, erschlossen werden. Datengetriebene Technologien können soziale Probleme nicht lösen, aber sie können zu Lösungen beitragen: *„Counting and measuring do not always have to be tools of oppression. We can also use them to hold power accountable, to reclaim overlooked histories, and to build collecitivity and solidarity.“*[157]

Gefördert von der Deutschen Forschungsgemeinschaft (DFG, German Research Foundation): TRR 318/1 2021 – 438445824

Bibliographie

Acker, Joan (2011): “Theorizing Gender, Race, and Class in Organizations”. Jeanes, E.L.; Knights, D.; Martin, P.Y. (Eds.): Handbook of Gender, Work & Organization. Chichester: Wiley, 65-80.

Acker, Joan (2006): “Inequality Regimes. Gender, Class, and Race in Organizations”. In: *Gender & Society* Vol. 20/4, 441-464.

Acker, Joan (1990): “Hierarchies, Jobs, Bodies: A Theory of Gendered Organizations”. In: *Gender and Society*, Vol.4/2, 139-158.

AlgorithmWatch gGmbH und Bertelsmann Stiftung (2019): “Automating Society. Taking Stock of Automated Decision-Making in the EU”. In: https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/automating-society (03.09.2021).

The Atlantic (2017): “Why Is Silicon Valley So Awful to Women?”. In: https://www.theatlantic.com/magazine/archive/2017/04/why-is-silicon-valley-so-awful-to-women/517788/ (03.09.2021).

Barocas, Solon/Selbst, Andrew D. (2016): „Big data‘s disparate impact”. In: *California Law Review* 104, 671-732.

BBC (2015): “Google apologises for Photos app’s racist blunder”. In:https://www.bbc.com/news/technology-33347866 (03.09.2021).

157 D’Ignazio & Klein 2020: 123

Bellamy, Rachel K. E./Dey, Kuntal/Hind, Michael/Hoffman, Samuel C./Houde, Stephanie/Kannan, Kalapriya/Lohia, Pranay/Martino, Jacquelyn/Mehta, Sameep/Mojsilovic, Aleksandra/Nagar, Seema/Ramamurthy/Karthikeyan Natesan/Richards, John/Saha, Diptikalyan/Sattigeri, Prasanna/Singh, Moninder/Varshney, Kush R./Zhang, Yunfeng (2018): *AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias.* arXiv:1810.01943.

Both, Göde (2014) "Multidimensional Gendering Processes at the Human-Computer-Interface: The Case of Siri. Gender-UseIT: HCI, Usability und UX unter Gendergesichtspunkten". Marsden, Nicole/Kempf, Ute (Hg.) *Technik-Diversity-Chancengleichheit*, Berlin, München, Boston: De Gruyter Oldenbourg, 107-112.

boyd, danah/Crawford, Kate (2012): "Critical questions for big data. Provocations for a cultural, technological, and scholarly phenomenon". In: *Information, Communication & Society*, Vol. 15/5, 662–679.

Bucher, Taina (2017): "The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms". In: *Information, Communication & Society* Vol. 20/1: 30-44.

Campolo, Alexander/Crawford, Kate (2020): "Enchanted Determinism: Power without Responsibility in Artificial Intelligence". In: *Engaging Science, Technology, and Society* 6, 1-19.

Crawford, Kate (2021): *Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven & London: Yale University Press.

Crawford, Kate (2016): "Artificial Intelligence's White Guy Problem". In: The New York Times, https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html (16.08.2021).

Crawford, Kate/Joler, Vladan (2018): "Anatomy of an AI System". In: https://anatomyof.ai/ (03.09.2021).

Criado Perez, Caroline (2019): *Invisible Women. Exposing data bias in a world designed for men*. London: Penguin Random House.

Datta, Amit/Datta, Anupam/Tschantz, Michael Carl (2015): "Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination". In: *Proceedings on Privacy Enhancing Technologies*, 1, 92–112.

Dewes, Andreas (2015): "Say hi to your new boss: How algorithms might soon control our lives. Discrimination and ethics in the data-driven society". In: https://media.ccc.de/v/32c3-7482-say_hi_to_your_new_boss_how_algorithms_might_soon_control_our_lives#t=1021 (03.09.2021).

D'Ignacio, Catherine/Klein, Lauren F. (2020): *Data Feminism*. Cambridge, Massachusets: The MIT Press.

Draude, Claude/Klumbyte, Goda/Lücking, Phillip/Treusch, Pat (2020): "Situated algorithms: a sociotechnical systemic approach to bias". In: *Online Information Review* Vol. 44/2, 325-342.

Eubanks, Virginie (2017): *Automating Inequality. How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.

Financial Times (2021): "After Google drama, Big Tech must fight against AI bias. 'Silicon Valley's problems with gender and racial imbalance are endemic and likely to last for years'". In: https://www.ft.com/content/ef0c61ab-240d-42b1-af3c-aca2e4896bd2 (03.09.2021).

The Guardian (2021): "A thousand young, black men removed from Met gang violence prediction database. Exclusive: Sadiq Khan review into 'discriminatory' police matrix found 38% on list posed little or no risk." In: https://www.theguardian.com/uk-news/2021/feb/03/a-thousand-young-black-men-removed-from-met-gang-violence-prediction-database (03.09.2021).

The Guardian (2017): "New AI can guess whether you're gay or straight from a photograph". In: https://www.theguardian.com/technology/2017/sep/07/new-artificial-intelligence-can-tell-whether-youre-gay-or-straight-from-a-photograph (03.09.2021).

Haraway, Donna (1988): "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective". In: *Feminist Studies* Vol. 14/3, 575-599.

Kaltheuner, Frederike/Obermüller, Nele (2018): *Datengerechtigkeit*. Berlin: Nicolai.

Die Zeit Online (2018): "Alexa ist nicht mehr deine Schlampe." In: https://www.zeit.de/digital/internet/2018-01/sprachassistenten-alexa-sexismus-feminismus-sprachsteuerung-ki (25.08.2021).

Lischka, K./ Klingel, A. (2017): „Wenn Maschinen Menschen bewerten: Internationale Fallbeispiele für Prozesse algorithmischer Entscheidungsfindung". In: Bertelsmann Stiftung, https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/wenn-maschinen-menschen-bewerten (03.09.2021).

Lupton, Deborah (2015): *Digital Sociology*. London and New York: Routledge.

Marres, Noortje (2017): *Digital Sociology. The Reinvention of Social Research*. Cambridge/Malden: polity Press.

Mayer-Schönberger, Viktor/Cukier, Kenneth (2013): *Big Data. Die Revolution, die unser Leben verändern wird*. München: Redline.

The New York Times (2019): “Apple Card Investigated After Gender Discrimination Complaints”. In: https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html (03.09.2021).

The New York Times (2016): “Artificial Intelligence’s White Guy Problem”. In: https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html (03.09.2021).

Noble, Safiya Umoja (2018): *Algorithms of Oppression. How search engines reinforce Racism*. New York: New York University Press.

O’Neill, Catherine (2016): *Weapons of Math Destructions. How Big Data Increases Inequality and Threatens Democracy*. UK/USA/Canada/Ireland/Australia/India/New Zealand/South Africa: Penguin Books.

Pentland, Alex (2013): “The Data-Driven Society”. In: *Scientific American* 309, 80–83.

Sachverständigenkommission für den Dritten Gleichstellungsbericht der Bundesregierung (2021): „Digitalisierung geschlechtergerecht gestalten. Gutachten für den Dritten Gleichstellungsbericht der Bundesregierung“. In: https://www.dritter-gleichstellungsbericht.de/de/topic/73.gutachten.html (03.09.2021).

Süddeutsche Zeitung (2016): „Wenn Algorithmen Vorurteile haben.“ In: https://www.sueddeutsche.de/digital/diskriminierung-wenn-algorithmen-vorurteile-haben-1.2806403 (03.09.2021).

Sweeney, Latanya (2013): “Discrimination in Online Ad Delivery. Google ads, black names and white names, racial discrimination, and click advertising”. In: *ACM Queue*, 11/3, 1–19.

UN Women (2013): “UN Women ad series reveals widespread sexism.” In: https://www.unwomen.org/en/news/stories/2013/10/women-should-ads (03.09.2021).

Rekabsaz, Navid/West, Robert/Henderson, James/Hanbury, Allan (2021): “Measuring Societal Biases from Text Corpora with Smoothed First-Order Co-occurrence”. In: *ICWSM* 2021, 549-560.

Kosinski, Michal/Wang, Yilun (2018): “Deep Neural Networks Are More Accurate Than Humans at Detecting Sexual Orientation From Facial Images.” In: *Journal of Personality and Social Psychology*, Vol. 114/2, 246–257.

The Verge (2020): “UK ditches exam results generated by biased algorithm after student protests”. In: https://www.theverge.com/2020/8/17/21372045/uk-a-level-results-algorithm-biased-coronavirus-covid-19-pandemic-university-applications?utm_campaign=theverge&utm_content=chorus&utm_medium=social&utm_source=twitter (03.09.2021).

The Verge (2015): “Google Search thinks the most important female CEO is Barbie”. In: https://www.theverge.com/tldr/2015/4/9/8378745/i-see-white-people (20.08.2021).

Die Zeit online (2018): „Der Algorithmus diskriminiert nicht“. In: https://www.zeit.de/arbeit/2018-01/roboter-recruiting-bewerbungsgespraech-computer-tim-weitzel-wirtschaftsinformatiker (03.09.2021).

Die Zeit Online (2016): „Twitter-Nutzer machen Chatbot zur Rassistin“. In: https://www.zeit.de/digital/internet/2016-03/microsoft-tay-chatbot-twitter-rassistisch (03.09.2021).

Zou, James/Schiebinger, Londa (2018): “Design AI so that it’s fair”. In: *Nature* Vol. 559, 324-326.

Zuboff, Shoshana (2018): *Das Zeitalter des Überwachungskapitalismus*. Frankfurt/New York: Campus.

Zweig, Katharina (2019): *Ein Algorithmus hat kein Taktgefühl. Wo künstliche Intelligenz sich irrt, warum uns das betrifft und was wir dagegen tun können*. München: Heyne.