Bianca Prietl

Algorithmische Entscheidungssysteme *revisited*: Wie Maschinen gesellschaftliche Herrschaftsverhältnisse reproduzieren können

Zusammenfassung

Der vorliegende Beitrag unternimmt eine konzeptionelle Analyse des Zusammenspiels von algorithmischen Entscheidungssystemen mit gesellschaftlichen Herrschaftsverhältnissen aus der Perspektive feministischer Technik- und Rationalitätskritik. Hierzu fasst er diese automatisierten Entscheidungstechnologien als herrschaftsförmige Erkenntnistechnologien, die spezifische Weltzugänge, Erkenntnismöglichkeiten und (politische) Gestaltungsoptionen begründen, und verfolgt die These, dass letztere ein Schlüssel sind um zu verstehen, wie algorithmische Entscheidungssysteme gesellschaftliche Herrschaftsverhältnisse reproduzieren. Konkret rekonstruiert er (1) die Ein- und Ausschlusseffekte eines datenfundamentalistischen Zugriffs auf Welt, (2) die ungleichheitskonservierende Tendenz eines post-explanativen Erkenntnispragmatismus und (3) die *Politics* von *Data Science*.*

1. Einleitung

2015 publizierten drei Informatiker*innen die Ergebnisse ihrer experimentellen Untersuchung von Googles Ad Posting-Algorithmus, derzufolge Werbeschaltungen für Führungskräfte trainings signifikant öfter Nutzer*innen angezeigt werden, deren Profil sie als Männer ausweist, als solchen, deren Profil sie als Frauen ausweist (Datta / Tschantz / Datta 2015). Letzten Herbst sorgte Amazon mit einer vergleichbaren Nachricht für Negativschlagzeilen: Ein zur Automatisierung von Personalrekrutierungen entwickelter, selbstlernender Algorithmus musste nach heftiger Kritik eingestellt werden, weil er Männer bevorzugte (Der Standard 2018). 2016 veröffentlichte die investigativ-journalistische Plattform ProRepublica ihre Analyse des Einsatzes datenbasierter Risikobewertungssysteme im US-amerikani-

* Für Ermutigung und wertvolles Feedback danke ich Paola Lopez sowie den anonymen Gutachter*innen.

schen Strafvollzug, derzufolge Afroamerikaner*innen systematisch ein höheres Rückfallrisiko attestiert wird als sog. weißen Amerikaner*innen (Angwin et al. 2016). Im selben Jahr berichtete die ZEIT online (2016) unter dem vielsagenden Titel »objektiv rassistisch« von einem datenbasiert selbstlernenden Algorithmus, der als Juror in einem fiktiven Schönheitswettbewerb überdurchschnittlich viele »kaukasisch« aussehende Frauen unter die Schönsten wählte. Vor diesem Hintergrund interessiert sich der vorliegende Beitrag für das Zusammenspiel von algorithmischen Entscheidungssystemen mit gesellschaftlichen Herrschaftsverhältnissen, wie sie entlang von Geschlechter-, Klassen-, Ethnizitäts- und anderen ungleichheitsproduzierenden sozialen Grenzen verlaufen. Dabei verfolgt er die These, dass wir es hier weniger mit zufälligen Fehlern, Verzerrungen oder *Bia-ses* zu tun haben, als mit algorithmischen Diskriminierungen, also einem Phänomen sozialer Ungleichheit.¹

Im Zuge der fortschreitenden Digitalisierung und Datafizierung der Gegenwartsgesellschaft gewinnen algorithmische Entscheidungssysteme in immer mehr gesellschaftlichen Bereichen an Bedeutung für die Produktion von Wissen und die Herbeiführung von Entscheidungen (Kitchin 2017; Prietl/Houben 2018). Besondere Aufmerksamkeit gebührt dabei dem Einsatz datenbasiert (selbst-)lernender Algorithmen, stehen sie doch im Verdacht, ein neues Paradigma der gesellschaftlichen Wissensproduktion (Kitchin 2014) bzw. einen neuen Governance-Modus (Danaher 2016) zu begründen. In Deutschland finden sich (selbst-)lernende Algorithmen gegenwärtig hauptsächlich in »sogenannten »Entscheidungsunterstützungssystemen« (Zweig 2018, 12), im Rahmen derer sie Expert*innenentscheidungen beratend informieren. Besonders brisant – und deshalb auch hier interessant – ist ihr Einsatz für die Produktion von Wissen über das Soziale und damit ihre Bedeutung für die Verteilung von gesellschaftlichen Ressourcen, sozialen Partizipationsmöglichkeiten und Lebenschancen.²

Technikoptimistische Stimmen verbinden mit algorithmischen Entscheidungssystemen die Hoffnung auf mehr, besseres und prädiktives Wissen (Mayer-Schönberger/Cukier 2013). Im Kontrast zu menschlichen Entscheidungen gelten automatisierte Entscheidungstechnologien als

¹ Weitere Fälle algorithmischer Diskriminierung: O’Neil (2018 [2016]), Eubanks (2018), »Data Harm Record« des Data Justice Lab der Cardiff University unter <https://datajusticelab.org/data-harm-record/> (1.7.2019).

² Nicht im Zentrum stehen hingegen algorithmische Entscheidungssysteme, die etwa im Kontext der Industrie 4.0 wartungsbedürftige Maschinenteile in Produktionsanlagen identifizieren und unter Optimierung der Produktionsabläufe (selbsttätig) einer Wartung zuführen.

effizienter, leistungsfähiger und beinahe unbegrenzt skalierbar, aber auch als objektiver, da vorgeblich frei von Vorurteilen oder kognitiven Verzerrungen (Zweig 2018, 15). Die gesellschaftliche Herstellung und höchst voraussetzungsvolle Einbettung und Nutzung von Technik, wie sie etwa die *Science and Technology Studies* betonen, geraten dabei rasch in den Hintergrund. Werden Fälle algorithmischer Diskriminierung publik, wird zwar schon einmal die Frage aufgeworfen, ob diese neuen Technologien sexistisch oder rassistisch sein können; fundierte Auseinandersetzungen mit den *Politics* dieser Artefakte (Paulitz 2005), d. h. den ihnen inhärenten Machtrelationen, stehen jedoch ebenso erst am Anfang wie die Frage, wie diese in und durch Technik materialisiert und verstetigt werden.

Der aktuelle Forschungsstand lässt sich grob in zwei Diskussionsstränge gliedern: Eine wachsende Zahl von Arbeiten dokumentiert und problematisiert die ungleichheitsproduzierenden *Effekte* algorithmischer Wissensproduktion und Entscheidungsfindung (Mau 2017; Eubanks 2018; Redden 2018). Während sie damit instruktive Hinweise auf deren Herrschaftsförmigkeit geben, bleiben die Technologien in ihrer *technischen Verfasstheit* weitgehend unbeleuchtet (Büchner 2018, 345), sodass unklar bleibt, *wie* algorithmische Entscheidungssysteme zu Produzent*innen sozialer Ungleichheit werden. Demgegenüber stehen Beiträge, die in den Datenwissenschaften bzw. den sie speisenden Feldern der Informatik, Mathematik oder Statistik angesiedelt sind und detailliertere Einsichten in die *informationstechnischen Funktionsweisen* algorithmischer Wissensproduktion und Entscheidungsfindung erlauben. Aufgrund zumeist fehlender Transparenz oder Dokumentation (als Ausnahme siehe Lopez 2019 zum Klassifizierungsalgorithmus des österreichischen Arbeitsmarktservices) handelt es sich hierbei entweder um konzeptionelle Analysen potentieller Bias-Quellen (Danks/London 2017; Zweig 2018; Suresh/Guttat 2019) oder um sog. *black box*-Experimente, bei denen ein algorithmisches Entscheidungssystem mehr oder weniger automatisiert, systematisch mit verschiedenen Inputdaten »gefüttert« wird, um anhand des jeweiligen Outputs Rückschlüsse auf seine Funktionsweisen zu ziehen (Sweeney 2013; Datta/Tschantz/Datta 2015; Ensign et al. 2018). Um zu verstehen, wie algorithmische Entscheidungssysteme und gesellschaftliche Herrschaftsverhältnisse zusammenspielen und warum die vermeintlich zufälligen *Biases* existierende soziale Ungleichheiten reflektieren, bedarf es (1) eines noch weitgehend ausstehenden Zusammendenkens beider Diskussionsstränge und (2) einer Betrachtung von automatisierten Entscheidungstechnologien als soziotechnisches Phänomen, d. h. als gleichermaßen technisch wie sozial (erste Beiträge liegen hier vor von Nobel 2018; Prietl 2019; Hagendorff 2020).

Überlegungen zu den Verquickungen von Technik, Macht und Wissen, wie sie u. a. den *Feminist Science and Technology Studies* zu verdanken sind, können ein solches Vorhaben informieren. Nachfolgend wird deshalb zunächst im Anschluss an Donna Haraway eine Perspektive auf algorithmische Entscheidungssysteme als *herrschaftsförmige Erkenntnistechnologien* entwickelt. Vor diesem Hintergrund wird eine informationstechnisch informierte, konzeptionelle Fehleranalyse automatisierter Entscheidungstechnologien vorgestellt und im Sinne einer *feministisch inspirierten Technik- und Rationalitätskritik* weitergedacht. Herausgearbeitet werden derart (1) die Ein- und Ausschlusseffekte eines datenfundamentalistischen Zugangs zur Welt, (2) die ungleichheitskonservierende Tendenz eines post-explanativen Erkenntnispragmatismus und (3) die *Politics* von *Data Science*. Das Fazit stellt weiterführende Überlegungen zur Diskussion.

2. Algorithmische Entscheidungssysteme als herrschaftsförmige Erkenntnistechnologien

Algorithmen bezeichnen »computer-programmed step-by-step instructions for taking a given set of inputs and producing an output« (Danaher 2016, 245), um ein mathematisches (Optimierungs-)Problem automatisiert zu lösen. Algorithmische Entscheidungssysteme bestehen in der Regel aus (mindestens) zwei Algorithmen: Der erste Algorithmus »lernt« auf Basis sog. Trainingsdatensätze, wie Fälle in der Vergangenheit kategorisiert wurden, indem er Muster in diesen (Groß-)Datensätzen identifiziert und daraus ein Regelwerk für die Kategorisierung ableitet, anhand dessen dann in der Zukunft neue Fälle kategorisiert werden können.³ Die Anwendung des derart »gelernten« Regelwerks auf neue Input-Daten erfolgt dann durch einen zweiten Algorithmus (Zweig 2018, 11 ff.). Wenn die Entwicklung und Einbettung von algorithmischen Entscheidungssystemen als eine »lange Kette von Verantwortlichkeiten« (Zweig 2018, 17) vorgestellt wird, in deren Verlauf »leicht Hunderte Personen [involviert sind]« (Zweig 2018, 20), wird deutlich, dass diese technischen Artefakte nicht losgelöst von den gesellschaftlichen Kontexten ihrer Herstellung und

³ Optimiert werden soll dabei i. d. R. die *Accuracy* oder Treffsicherheit, mit der (Input-) Daten klassifiziert werden. Daraus folgt zum einen, dass die identifizierten Regeln weniger mit inhaltlich-qualitativen Erklärungen zu tun haben, als damit, eine bestmögliche Klassifikation zu erzielen; zum anderen bedeutet dies, dass das algorithmische Entscheidungssystem darauf »trainiert« wird, die im Trainingsdatensatz angelegten Unterscheidungen auch selbst vorzunehmen (Lopez 2019, 7, 9).

Nutzung mit ihren je spezifischen Macht- und Herrschaftsverhältnissen verstanden werden können.

Um algorithmische Entscheidungssysteme als herrschaftsförmige sozio-technische Phänomene in den Blick zu nehmen, sind Donna Haraways Arbeiten geradezu prädestiniert. Bereits in den 1980er-Jahren betonte sie hinsichtlich der aufstrebenden Informations- und Kommunikationstechnologien (IKT) das unauflösliche Zusammenspiel von Technikgestaltung und Wissensproduktion mit patriarchalen, kolonialen und kapitalistischen Herrschaftsverhältnissen (2004 [1985]). Mit Blick auf ihre berühmte (Denk-)Figur der* Cyborg, einem »cybernetic organism, a hybrid of machine and organism« (2004 [1985], 7), problematisiert sie etwa: »The main trouble with cyborgs, of course, is that they are the illegitimate offspring of militarism and patriarchal capitalism, not to mention state socialism.« (2004 [1985], 10).⁴

IKT und andere *Technosciences*, die jenseits etablierter Grenzen von »reiner« und angewandter Forschung, Natur- und Technikwissenschaften, aber auch Wissenschaft, Ökonomie und Politik operieren, repräsentieren, so Haraway, nicht nur aufstrebende Technologien, sondern etablierten auch eine neue Wissensordnung und begründeten damit spezifische Weltzugänge, Erkenntnismöglichkeiten und politische Gestaltungsoptionen (Haraway 1995; Weber 2017, 350 ff).

Ähnlich anderen Akteur-Netzwerk-Theorien verwehrt sich Haraway dabei einer Unterscheidung von menschlichen und technowissenschaftlichen Entitäten und betont vielmehr deren unauflösliche Verwobenheit in semiotisch-materiellen Apparaten der Wissensproduktion. Dabei seien Technologien der Visualisierung und Vermessung entscheidend an den Prozessen der Konstruktion von Daten als spätere Fakten beteiligt, indem sie etwa prästrukturieren, was mit ihrer Hilfe wie gesehen, gemessen und damit gewusst – bzw. nicht gewusst – werden kann (Haraway 2017 [1995], 399). Wahrheitsansprüche seien deshalb stets das kontingente Ergebnis lokaler soziotechnischer – nämlich semiotisch-materieller – Konstruktions- und Aushandlungsprozesse, bei denen menschliche ebenso wie nicht-menschliche Aktant*innen handlungstragend beteiligt sind. Vor diesem Hintergrund gilt Haraway *jeder* Wahrheitsanspruch als »situierter« und in diesem Sinne als partikular. Objektivität will sie folglich nicht län-

⁴ Damit kann Haraway als eine Vordenkerin von Betrachtungen sozialer Ungleichheit angesehen werden, wie sie aktuell unter dem Stichwort »Intersektionalität« verhandelt werden und darauf abheben, dass sich kontextspezifisch je verschiedene Herrschaftsverhältnisse, wie Klassismen, Sexismen oder Rassismen, überkreuzen, wechselseitig verstärken oder auch abschwächen (u. a. Yuval-Davis 2006; Bührmann 2009).

ger an einen vorgeblich neutralen Blick »von nirgendwo« geknüpft sehen, sondern an die Sichtbarmachung der (eigenen) Erkenntnisposition, die es erst ermögliche, für Wahrheitsansprüche zur Verantwortung gezogen zu werden (2017 [1995], 389).

Mit Haraway lassen sich algorithmische Entscheidungssysteme also als in gesellschaftlichen Herrschaftsverhältnissen situierte, semiotisch-materielle Apparate der Wissensproduktion, bestehend aus technischen und menschlichen Aktant*innen, verstehen. Als herrschaftsförmige Erkenntnistechnologien bringen sie Wissensbestände hervor, die gleichsam weder neutral noch objektiv sind, und sind dementsprechend selbst hochgradig politisch. Vorgeschlagen wird damit ein »shift away from algorithms as *the* agential and consequential entity« (Neyland / Möllers 2016, 3; H.i. O.) hin zu ihrer Betrachtung als stets in komplexe soziotechnische Zusammenhänge eingebunden (hierzu pionierhaft auch Paulitz 2005). Eine derart perspektivierte, feministische Technik- und Rationalitätskritik interessiert sich sowohl dafür, welche Weltzugänge, Erkenntnismöglichkeiten und Gestaltungsoptionen algorithmischen Entscheidungssystemen innewohnen als auch dafür, wie diese mit gesellschaftlichen Herrschaftsverhältnisse verflochten sind.⁵

3. Algorithmische Entscheidungssysteme: Diskriminierung qua Design?

Die von der Informatikerin und Mitbegründerin der Initiative *Algorithm Watch*, Katharina Zweig, vorgelegte konzeptionelle Fehleranalyse algorithmischer Entscheidungssysteme ist nicht nur beispielhaft für vergleichbare informationstechnisch informierte Synopsen möglicher Bias-Quellen entlang des Entwicklungs- und Nutzungsprozesses automatisierter Entscheidungstechnologien (Danks / London 2017; Suresh / Gutttag 2019); die von der Bertelsmann Stiftung geförderte Arbeit tritt auch dezidiert dazu an, eine ethische Diskussion über Algorithmen im deutschsprachigen Raum zu eröffnen. Konkret beschreibt Zweig (2008) die Entwicklung und Einbettung von algorithmischen Entscheidungssystemen in praktische Nutzungskontexte als einen fünfteiligen Prozess. In jeder Phase können »Fehler« gemacht werden, die zu diskriminierenden Entscheidungen führen. Schon der Titel, *Wo Maschinen irren können*, aber auch einleitende Bemerkungen, wonach »Algorithmen diskriminierungsfrei entscheiden [können]« (Zweig

⁵ Im Sinne Haraways geht damit weder eine Dämonisierung von Technik einher, noch die Reklamation einer »unschuldigen« Perspektive für feministische Positionen.

2018, 15), »[s]olange sie keinen Input als Trainingsdatensatz bekommen, der eine Diskriminierung beinhaltet, und solange eine Diskriminierung nicht explizit in den Programmcode implementiert wird« (ebd.), deuten an, dass dabei von einer grundsätzlichen Neutralität und Objektivität von Technik ausgegangen wird, die es durch Behebung der menschengemachten Fehler zu bewahren oder wieder herzustellen gilt. Damit bleiben Zweigs Ausführungen tendenziell der Hoffnung verhaftet, dass objektive und neutrale Ergebnisse möglich sind, wenn denn erst alle Fehler behoben wären.

3.1 »Wo Fehler passieren können«

In Phase 1 der Entwicklung algorithmischer Entscheidungssysteme, Algorithmen-Design und Implementierung, wird ein soziales Problem formal-analytisch als ein mathematisches (Optimierungs-)Problem formuliert, wobei der Algorithmus das Verhältnis von In- und Output beschreibt. Laut Zweig können in dieser Phase vor allem »handwerkliche Fehler« (2018, 21) auftreten, deren Identifikation und Behebung umso wahrscheinlicher ist, je mehr Informatiker*innen Zugang zum Code haben.

In der entscheidenden Phase 2, Methodenwahl, wird in einem iterativen Vorgehen über die Art des verwendeten Lernalgorithmus ebenso entschieden wie über die Operationalisierung des (sozialen) Problems und den Trainingsdatensatz (Zweig 2018, 18 ff.). Hier ist nicht nur höchst relevant, welche Daten der Trainingsdatensatz beinhaltet, anhand dessen der Algorithmus sein Regelwerk »lernt«, sondern auch wie ein (soziales) Phänomen messbar gemacht wird (Zweig 2018, 22 ff.). So wird vermutet, dass der sich in einem Schönheitswettbewerb rassistisch gebärende Algorithmus in der Lernphase kaum Gesichter von *People of Color* zu sehen bekam, weshalb diese nicht in das von ihm identifizierte Muster von Schönheit passten. Denkt man an androzentrische Anerkennungs- und Zitationszirkel in der Wissenschaft (Rossiter 2003), wird außerdem augenscheinlich, dass es einen erheblichen Unterschied macht, ob die Professorabilität von Wissenschaftler*innen anhand der Anzahl all ihrer Publikationen bemessen wird oder – wie bei dem vermehrt auch in Berufungsverfahren zu Rate gezogenen H-Index (Hirsch 2005) – anhand der Anzahl n an Publikationen, die alle mindestens n -Mal zitiert worden sind.

In Phase 3 kommt es zur eigentlichen Konstruktion des Entscheidungssystems, indem der maschinelle Lernalgorithmus – oftmals auch mehrere – in einem *trial-and-error*-Verfahren mit den Trainingsdaten zusammengeführt und solange modifiziert wird, bis die anvisierte *Accuracy* bzw. Treffsicherheit erreicht wird. Damit hängt die Konstruktion des Entschei-

dungssystems erheblich davon ab, welche Fehlerrate bei der Klassifikation von Daten als akzeptabel gesetzt wird und wie diese überhaupt berechnet wird (Zweig 2018, 19, 24f.). So weist das von ProRepublica als rassistisch identifizierte Risikobewertungssystem im US-amerikanischen Strafvollzug zwar eine hohe Gesamttrefferquote auf, jedoch eine deutlich schlechtere Treffersicherheit bei der Klassifikation von afroamerikanischen im Vergleich zu »weißen« Amerikaner*innen (Gummadi 2019).

In Phase 4 erfolgt die Einbettung des algorithmischen Entscheidungssystems in den Nutzungskontext, indem das algorithmische Entscheidungssystem mit neuen Daten gespeist wird, die bewertet werden bzw. auf deren Basis Vorhersagen über das zukünftige Verhalten von Personen getroffen werden sollen (Zweig 2018, 19). Hier könne es zu Fehlinterpretationen kommen, wenn den Anwender*innen etwa nicht klar ist, dass das algorithmische Entscheidungssystem keine Vorhersage über das tatsächliche Verhalten einer Person macht, sondern nur die gruppenbasierte Wahrscheinlichkeit für ein bestimmtes Verhalten angibt. Auch könne Unkenntnis über die Fehlerrate des algorithmischen Entscheidungssystems dazu führen, dass dessen Entscheidungen allzu leicht unhinterfragt bleiben. Problematisch kann aber auch der unsachgemäße oder manipulative Umgang mit dem System werden, wie dies etwa der Fall des selbstlernenden Chatbots *Tay* zeigt, der von Nutzer*innen mit rassistischen Äußerungen »gefüttert« wurde und schon nach kurzer Zeit mit rechtsradikalen Statements für Aufregung sorgte (Zweig 2018, 25f.).

Schließlich geht es in Phase 5, Re-Evaluation, darum, das Funktionieren des algorithmischen Entscheidungssystems im Nutzungskontext zu bewerten und diese Bewertung als Feedback in das System zurückzuspielen, um es (kontinuierlich) weiterzuentwickeln (Zweig 2018, 20). Dieser Prozess kann durch selbstverstärkende Feedbackschleifen oder Feedbackasymmetrien systematisch gestört werden. Im Rahmen von *Predictive Policing* etwa wird die Einsatzplanung der Polizei durch datenanalytische Prognosetechniken informiert, mit der Konsequenz, dass vermehrte Streifen an manchen Orten auch zu mehr Feststellungen von Straftaten in diesen Gebieten führen, während Straftaten in weniger unter polizeilicher Beobachtung stehenden Gebieten vermehrt unerfasst bleiben, sodass »police are repeatedly sent back to the same neighborhoods« (Ensign et al. 2018, 1). Auch können im Fall von Hafterleichterungen nur jene Fehlentscheidungen überhaupt als solche erkannt werden, in denen jemand entlassen, aber rückfällig geworden ist; jemand, der oder die nicht rückfällig geworden wäre, aber gar nicht erst entlassen wurde, kann dies per definitionem niemals unter Beweis stellen.

Die Ausführungen von Zweig geben instruktive Hinweise auf die *sozio-technische Kontingenzt* algorithmischer Entscheidungssysteme. Wenngleich soziale Faktoren durchaus als relevante Stellschrauben in den Blick kommen und somit Einblick in die Konstruktions- und Aushandlungsprozesse gegeben wird, die diesen Technologien und ihren Modi der Wissensproduktion und Entscheidungsfindung zugrunde liegen, bleiben die Ausführungen doch *technikzentriert* und verlieren die *gesellschaftlichen Voraussetzungen* dieser Konstruktions- und Aushandlungsprozesse tendenziell ebenso aus dem Blick wie die hierin angelegten Herrschaftsverhältnisse. Hier kann eine sozialwissenschaftlich informierte, feministische Technik- und Rationalitätskritik den Faden aufgreifen und weiterspinnen.

3.2 Algorithmische Entscheidungssysteme in technik- und rationalitätskritischer Perspektive

Nachfolgend werden – ohne Anspruch auf Vollständigkeit – drei Facetten der herrschaftsförmigen Verflechtung von algorithmischen Entscheidungssystemen mit sozialen Ungleichheitsverhältnissen analytisch rekonstruiert.

3.2.1 Ein- und Ausschlusseffekte eines datenfundamentalistischen Weltzugangs

Versuche der Wissensproduktion und Entscheidungsfindung mittels algorithmischer Entscheidungssysteme basieren auf der (stillschweigenden) Annahme, »that life can be captured and modeled by data or even fully transformed into it« (Thatcher 2014, 1768) – eine erkenntnistheoretische Prämisse, die Kate Crawford (2013) kritisch als »Datenfundamentalismus« bezeichnet. Mit ihrem datenfundamentalistischen Zugang zur (sozialen) Welt gerieren sich algorithmische Entscheidungssysteme als Wiedergängerinnen streng positivistischer Erkenntnismodelle und erneuern »the primacy of inductive reasoning in the form of a technology-based empiricism«, wie Fulvio Mazzocchi (2014, 1250) problematisiert. Gleichzeitig ist auch in informationstechnischen und datenwissenschaftlichen Kreisen weitestgehend unstrittig, dass Daten entgegen der alltagsweltlichen Repräsentationsthese niemals einfach »gegeben« und neutral, sondern immer schon »zubereitet, verarbeitet und »cooked« (Gitelman2013; Busch 2017; Mützel/Saner/Unternährer 2018) sind, wovon nicht zuletzt die datenwissenschaftlichen Auseinandersetzungen mit (Trainings-)Datensätzen als primärer Quelle von *Biases* zeugen. Weitaus weniger reflektiert wird jedoch, dass schon der datafizierende Weltzugriff selbst kein neutraler ist.

Wissenschafts- und erkenntnistheoretische Arbeiten argumentieren etwa, dass der Ruf nach mathematischer Repräsentation ein mechanistisches

Menschenbild favorisiert, das auf vereinzelte codifizierbare Komponenten menschlichen Handelns und Seins fokussiert und nicht in der Lage ist, deren Komplexität, Variabilität oder auch Disparität zu erfassen (Hara-way 2017 [1995]; Barnes / Wilson 2014, 10). Historische Erfahrungen mit Quantifizierungsbestrebungen in der Humangeographie und Sozialphysik lehren zweitens, dass quantifizierende Erkenntniszugänge tendenziell blind für Machtverhältnisse und Interessenkonflikte sind und auch kulturell-symbolische Phänomene nur schwer erfassen können (Barnes / Wilson 2014, 10; Mazzocchi 2014).⁶ In ihrer Geschlechteranalyse der in der Informatik vorherrschenden »Epistemologie des Formalen« demonstriert Corinna Bath (2009, 101 ff.) drittens, wie die hierin angelegte Fokussierung auf Abstraktion, Modellierung und Standardisierung zu einer vereinheitlichenden und normierend / normalisierenden Abbildung der Wirklichkeit führt, die in der Tendenz Aspekte der Realität jenseits der (andro- und eurozentrischen) Norm ausblendet.

Nicht nur die für algorithmische Entscheidungssysteme so essentiellen Daten(sätze) sind also niemals neutral und objektiv, sondern auch der datenfundamentalistische Weltzugang *selbst* zeitigt durch die in diesem Zugriff auf das Soziale begründeten (Un-)Sichtbarkeits-effekte herrschaftsförmige Ein- und Ausschlüsse. Diesen wohnt die Tendenz inne, Leben jenseits der herrschenden Norm(en) zu marginalisieren, gesellschaftliche Macht- und Ungleichheitsverhältnisse ebenso wie symbolisch-kulturelle Phänomene auszublenden und ein mechanistisches Menschenbild zu profilieren, das der Komplexität sozialen Lebens nur bedingt gerecht wird. Wenn eine von der australischen Regierung eingesetzte Software zur Prüfung von Sozialleistungsansprüchen überdurchschnittlich viele Personen mit »ungewöhnlichen« Erwerbsverläufen fälschlicherweise zu Rückzahlungen aufruft, dann liegt dies also nicht nur daran, dass Algorithmen grundsätzlich keine Ermessensspielräume abbilden können und stets dieselben Regeln zur Anwendung bringen (Zweig 2018, 5, 15); hier zeigt sich vielmehr auch, wie gerade ohnehin schon marginalisierten Menschen wie Arbeits- und Wohnungssuchende, Menschen mit Migrationshintergrund oder auch Frauen vermehrt von den herrschaftsförmigen Ein- und Ausschlüsseffekten eines datenfundamentalistischen Weltzugriffs betroffen sind, weil ihre Leben eher von den typisierten Normallebensläufen und

⁶ Im Kontext der Profilierung von Big Data-Methoden im Human Resource Management beobachtet Renyi Hong (2016) etwa aktuell, dass dem Ruf nach »harten Zahlen« ein »latenter Sexismus« innewohnt, da durch die Fokussierung auf Zahlen die nicht gleichermaßen quantifizierbare emotionale Arbeit – von zumeist Frauen – tendenziell aus dem Blick gerät.

-umständen der »weißen« Mittelschicht(smänner) abweichen, und damit öfter nicht (korrekt) erfasst und angemessen beurteilt werden.

3.2.2 Ungleichheitskonservierende Tendenz eines post-explanativen Erkenntnispragmatismus

Im Zentrum algorithmischer Entscheidungssysteme stehen maschinelle Lernalgorithmen, die im Kern darauf ausgelegt sind, aus Mustern, die in zwangsläufig immer schon aus der Vergangenheit stammenden Daten gefunden werden, Annäherungen über die Zukunft abzuleiten. Erkenntnistheoretisch ist damit ein zweifacher Wandel verbunden: Zum einen geht es nicht länger darum, ein Phänomen zu verstehen oder zu erklären – also die klassischen *Warum*- oder *Wie*-Fragen zu stellen; stattdessen wird Erkenntnisgewinnung auf die Produktion von probabilistischen Vorhersagen und damit auf das Bearbeiten der *Was*- oder *Wer*-Fragen reduziert (Bowker 2014; Andrejevic 2015). Zum anderen implementiert die Fokussierung auf prädikatives Entscheidungswissen ein »Regime der Antizipation« (Adams / Murphy / Clarke 2009), in dem Vorhersagen über die Zukunft, die aus der Vergangenheit abgeleitet werden, Aktionen in der Gegenwart verlangen. Im Kontext dieses von Mark Andrejevic (2014, 1675) entsprechend als post-explanativer Erkenntnispragmatismus bezeichneten Erkenntnismodells wird Korrelation zunehmend gewichtiger und bedeutsamer als Kausalität (Bowker 2014, 1796) und Wahrheitsansprüche werden daran gemessen, ob sie »actionable« sind, also Handlungen anleiten können.

Geoffrey Bowker (2014) argumentiert, dass dieses Korrelationsprimat zwar grundsätzlich die Chance berge, Zusammenhänge zu ergründen, die jenseits a priori gesetzter Stereotype existieren, wenn etwa das Kaufverhalten von Personen analysiert wird, ohne vorab Geschlecht oder Alter als unabhängige Variable zu setzen und als entsprechend bedeutsam zu unterstellen. Der hierin angelegte Verzicht darauf, die identifizierten Zusammenhänge erklärend zu verstehen, lässt diesen Vorteil jedoch rasch in sein Gegenteil umschlagen, wie ein vom Arbeitsmarktservice Österreich (AMS) entwickelter Algorithmus zur Klassifikation von Arbeitssuchenden zeigt (Lopez 2019): Dieser sortiert Arbeitssuchende entsprechend ihrer prognostizierten Chance auf Arbeitsmarktwiedereingliederung in drei Gruppen, denen je unterschiedliche Unterstützungsleistungen zustehen. Grundlage der Klassifikation sind vom Algorithmus identifizierte Muster in der über mehrere Jahre aufgebauten Datenbank des AMS, wonach etwa Ältere, Frauen, Mütter und Migrant*innen schlechtere Chancen haben, eine neue Stelle zu finden – was angesichts existierender Ungleichheiten am Arbeitsmarkt wenig überrascht. Diese *sozialen* Strukturmuster infor-

mieren nun aber die Klassifikation von gegenwärtigen wie zukünftigen Arbeitssuchenden, mit der Konsequenz, dass Personen höheren Alters oder weiblichen Geschlechts nicht länger Anspruch auf bestimmte Ressourcen haben. Damit drohen nicht nur für die Vergangenheit identifizierte soziale Ungleichheitsstrukturen schicksalsgleich in die Zukunft einzelner Menschen fortgeschrieben zu werden; das algorithmische Entscheidungssystem implementiert auch einen soziotechnischen *self-fulfilling prophecy*-Apparat.

Algorithmische Entscheidungssysteme entfalten also nur allzu rasch eine konservierende Tendenz (Bowker 2014, 1797), wenn soziale Muster und Strukturen, die sich fast zwangsläufig in Big *Social Data* finden, als positivistischer Ausdruck von Wahrheit *verkannt* werden. Der algorithmischen Entscheidungssystemen zugrundeliegende post-explanative Erkenntnispragmatismus birgt deshalb in Kombination mit dem gleichermaßen in diesem Kontext propagierten Antizipationsregime die Gefahr, existierende Herrschaftsverhältnisse zu affirmieren und damit Möglichkeiten der Veränderung und des Wandels zu unterminieren. Die erhoffte Unabhängigkeit von menschlichen Einflüssen droht damit anstatt zu mehr Objektivität und Neutralität, zur unausweichlichen Reproduktion sozialer Ungleichheitsstrukturen qua algorithmischer Erkenntnisteknologie zu führen.

3.2.3 Die Politics von Data Science

Sowohl die für algorithmische Entscheidungssysteme essentiellen Datensätze als auch die informationstechnischen Infrastrukturen für ihre Erhebung, Verarbeitung und Analyse liegen weitestgehend entweder in den Händen von Internet- und Datenkonzernen oder aber von staatlichen Nachrichtendiensten und Verteidigungseinrichtungen (Lyon 2014). Wie schon technische Innovationen vor ihnen sind damit auch algorithmische Entscheidungssysteme primär in militärischen und kapitalistischen Entstehungs- und Verfügungskontexten situiert. Zudem dominiert in diesem Kontext eine »virtual class« (Barbrook / Cameron 1996) aus überwiegend »weißen«, jungen, gut ausgebildeten und sozioökonomisch privilegierten Männern, wodurch die primär für die Entwicklung von automatisierten Entscheidungstechnologien verantwortlich zeichnenden Personen gleich auf mehreren Achsen der gesellschaftlichen Hierarchisierung die Position der Herrschenden einnehmen.⁷

⁷ Von den mitunter offen sexistischen Arbeitskulturen in der New Economy zeugte zuletzt ein von einem Google Mitarbeiter veröffentlichtes anti-feministisches Manifest (Bovensiepen 2017).

Obschon als »sexiest job of the 21st century« (Davenport / Patil 2012) ausgelobt, ist das Betätigungsfeld *Data Science* noch kaum institutionalisiert und Ausbildungs- und Studienformate werden gerade erst etabliert (Zweig 2018, 17 ff.). Datenwissenschaftler*innen werden deshalb bislang v. a. unter Absolvent*innen von Informatik-, Mathematik- oder Physikstudiengängen rekrutiert. Damit verfügen sie nicht nur selten über fachliche Expertise für die Gegenstände ihrer technischen Verarbeitung (Bartlett et al. 2018); ihre Ausbildung umfasst für gewöhnlich auch keine Inhalte, die sie dazu in die Lage versetzen, die ethischen Implikationen und sozialen Folgen ihrer Arbeit zu reflektieren (Zweig 2018, 30). Hinzu kommt, dass unter ihnen eine Weltanschauung zu dominieren scheint, die liberale und meritokratische Ideale als Grundlage wahrer Gerechtigkeit beschwört⁸ und in Information(szugang) den Schlüssel zur Lösung aller – auch sozialer – Probleme erkennt (Dickel / Schrape 2015; Nachtwey / Seidl 2017). Am Beispiel eines digitalen Registers zur automatisierten Zuteilung von Sozialwohnungen in Los Angeles zeigt sich hingegen (Eubanks 2018, 125), wie die Reduktion eines hochgradig durch Klassismen, aber auch Sexismen, Rassismen und Bodyismen strukturierten Verteilungskonfliktes auf ein Informationsproblem nicht nur die zugrundeliegenden sozialen Ungleichheitsstrukturen unberührt lässt, sondern diese sogar verdeckt. Denn es ist keineswegs die schlechte Informationslage von Wohnungssuchenden, sondern schlicht der Mangel an leistbarem Wohnraum, der so viele Menschen auf der Straße leben lässt.

In der von Datenwissenschaft und -wissenschaftler*innen propagierten Fokussierung auf »rein sachliche« Probleme und deren technische Lösung artikuliert sich also nicht nur die Machtvergessenheit von Akteur*innen, die sich ihrer eigenen Position als Herrschende nicht gewahr sind; sie verstellt auch tendenziell den Blick auf wohl nicht nur in diesem Fall notwendige politische Auseinandersetzungen mit Machtasymmetrien, Interessenskonflikte und Ungleichheitsstrukturen.

⁸ Barbrook und Cameron (1996) beschreiben die im Silicon Valley verbreitete Weltanschauung des Internet- und Informationszeitalters als »bizarre« und »ambivalente« Kombination von »hippie anarchism« (56) und »anti-corporatism« (52) auf der einen und »economic liberalism« (56), »entrepreneurial zeal of the yuppies« (45) und »laissez faire ideology« (52) auf der anderen Seite.

4. Fazit

Dieser Beitrag argumentiert, dass die sog. *Biases* automatisierter Entscheidungstechnologien als algorithmische Diskriminierung und damit als Phänomen sozialer Ungleichheit zu verstehen sind. Werden automatisierte Entscheidungstechnologien als herrschaftsförmige soziotechnische Erkenntnisteknologien in den Blick genommen, kann analysiert werden, wie diese ob der ihnen zugrundeliegenden Weltzugänge, Erkenntnisprämissen und Gestaltungsoptionen Herrschaftseffekte entfalten, die in der Tendenz bestehende soziale Ungleichheiten reproduzieren.

Dies wird dadurch verstärkt, dass den Ergebnissen algorithmischer Entscheidungssysteme nur schwer zu widersprechen ist. Einerseits zeichnen sich diese durch eine gewisse Undurchsichtigkeit bzw. Undurchdringlichkeit aus (Burrell 2016): Wo sie nicht von Geheimhaltungsklauseln geschützt werden, bleiben sie dem Großteil der Bevölkerung aufgrund fehlender fachlicher Expertise intransparent. Aber auch Fachleute können die Kluft zwischen den semantischen Logiken menschlichen Verstehens und der mathematisch-informationstechnischen Logik von Algorithmen mit ihrer extremen Skalierung nicht einfach überbrücken. Andererseits kann den Ergebnissen algorithmischer Entscheidungssysteme ob der symbolischen Macht und Autorität von (numerischen) Daten nur unter Verweis auf alternative Daten oder mittels einer fundierten Kritik der vorliegenden Daten widersprochen werden, wie Bettina Heintz (2010, 172) schon für numerische Zahlen erklärte. Beides verlangt Ressourcen bzw. Einblick in die Methoden und Entstehungsbedingungen automatisierter Entscheidungen, über die nur wenige verfügen.

Folgt man der hier vorgelegten Argumentation, so gilt es, die grundlegende Herrschaftsförmigkeit von algorithmischen Entscheidungssystemen anzuerkennen und von der vielfach noch immer hochgehaltenen Hoffnung Abschied zu nehmen, wonach automatisierte Entscheidungstechnologien objektiver oder neutraler seien als Menschen. Zukünftig bedarf es sowohl empirischer Untersuchungen, die den vorliegenden Einsichten Befunde darüber zur Seite stellen, wie sich das Zusammenspiel von Technik, Wissen und Macht in der Praxis konkret entfaltet, als auch Anstrengungen, automatisierte Entscheidungstechnologien zu entwickeln, die nicht nur möglichst akkurat die existierenden Strukturen sozialer Ungleichheit abbilden, sondern auf ihren Abbau hinwirken. Gesellschaftspolitisch ist außerdem eine möglichst breite Debatte darüber zu fordern, wo derartige Erkenntnisteknologien eingesetzt werden sollen bzw. ob es Bereiche ge-

ben soll, die nicht in die »Hand« von Maschinen gelegt werden (zur Regulierung von Algorithmen siehe Leighton et al. 2017).

Literatur

- Adams, Vincanne / Murphy, Michelle / Clarke, Adele E. (2009): Anticipation: Technoscience, life, affect, temporality. In: *Subjectivity* 28, 246–265.
- Andrejevic, Mark (2014): The Big Data Divide. In: *International Journal of Communication* 8, 1673–1689.
- Angwin, Julia / Larson, Jeff / Surya, Mattu / Kirchner, Lauren / Parris, Terry Jr. (2016): Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks. In: *ProPublica* (23.5.2016). <https://www.propublica.org/article/machine-bias-riskassess-ments-in-criminal-sentencing> (30.12.2018).
- Barbrook, Richard / Cameron, Andy (1996): The Californian Ideology. In: *Science as Culture* 26, 44–72.
- Barnes, Trevor J. / Wilson, Matthew W. (2014): Big Data, social physics, and spatial analysis: The early years. In: *Big Data & Society*. DOI: 10.1177/2053951714535365.
- Bartlett, Andrew / Lewis, Jamie / Reyes-Galindo, Luis / Stephens, Neil (2018): The locus of legitimate interpretation in Big Data sciences. In: *Big Data & Society*. DOI: 10.1177/2053951718768831.
- Bath, Corinna (2009): De-Gendering informatischer Artefakte: Grundlagen einer kritisch feministischen Technikgestaltung. Bremen.
- Bovensiepen, Nina (2017): Die weißen Männer von Google. In: *Süddeutsche Zeitung* (9.8.2017). <https://www.sueddeutsche.de/karriere/gleichberechtigung-die-weissen-maenner-von-google-1.3620497> (6.1.2019).
- Bowker, Geoffrey C. (2014): The Theory / Data Thing. In: *International Journal of Communication* 8, 1795–1799.
- Burrell, Jenna (2016): How the machine »thinks«: Understanding opacity in machine learning algorithms. In: *Big Data & Society*. DOI: 10.1177/2053951715622512.
- Busch, Lawrence (2017): A Dozen Ways to Get Lost in Translation: Inherent Challenges in Large-Scale Data Sets. In: *International Journal of Communication* 8, 1727–1744.
- Büchner, Stefanie (2018): Zum Verhältnis von Digitalisierung und Organisation. In: *Zeitschrift für Soziologie* 47, 332–348.
- Bührmann, Andrea (2009): Intersectionality – ein Forschungsfeld auf dem Weg zum Paradigma? In: *GENDER* 2, 28–44.
- Crawford, Kate (2013): The Hidden Bias in Big Data. In: *Harvard Business Review* (1.4.2013).
- Danaher, John (2016): The Treat of Algocracy: Reality, Resistance and Accomodation. In: *Philosophy & Technology* 29, 245–268.
- Danks, David / London, Alex J. (2017): Algorithmic Bias in Autonomous Systems. In: *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, 4691–4697.
- Datta, Amit / Tschantz, Michael Carl / Datta, Anupam (2015): Automated Experiments on Ad Privacy Settings. A Tale of Opacity, Choice, and Discrimination. In: *Proceedings on Privacy Enhancing Technologies*, 92–112.

- Davenport, Thomas H. / Patil, D.J. (2012): Data Scientist: The Sexiest Job of the 21st Century. In: Harvard Business Review (Okt. 2012). <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century> (3.7.2019).
- Der Standard (2018): Amazon streicht KI-Rekrutierungstool wegen Frauenfeindlichkeit (10.10.2018). www.derstandard.de/story/2000089096622/amazon-streicht-ki-rekrutierungstool-wegen-frauenfeindlichkeit (30.1.2019).
- Dickel, Sascha / Schrape, Jan-Felix (2015): Dezentralisierung, Demokratisierung, Emanzipation. Zur Architektur des digitalen Technikutopismus. In: Leviathan 43, 442–463.
- Ensign, Danielle / Friedler, Sorelle A. / Neville, Scott / Scheidegger, Carlos / Venkatasubramanian, Suresh (2018): Runaway Feedback Loops in Predictive Policing. In: Proceedings of Machine Learning Research 81, 1–12.
- Eubanks, Virginia (2017): Automating Inequality. How High-Tech Tools Profile, Police, and Punish the Poor. New York.
- Gitelman, Lisa (Hrsg., 2013): Raw Data is an Oxymoron. Cambridge.
- Gummadi, Krishna P. (2019): Foundations for Fair Algorithmic Decision Making. Keynote auf der transdisziplinären Tagung »Bias in AI and Neuroscience«, 17.–19. Juni 2019, Radboud University Nijmegen.
- Hagendorff, Thilo (2020): Maschinelles Lernen und Diskriminierung: Probleme und Lösungsansätze. In: Musik, Christoph / Bogner, Alexander (Hrsg.): Digitalization and Society. Wiesbaden, 53–66.
- Haraway, Donna (2004 [1985]): A Manifesto For Cyborgs: Science, Technology, and Socialist Feminism in the 1980s. In: dies. (Hrsg.): The Haraway Reader. New York, London, 7–45.
- Haraway, Donna (2017 [1995]): Situiertes Wissen. Die Wissenschaftsfrage im Feminismus und das Privileg einer partialen Perspektive. In: Bauer, Susanne / Heinemann, Torsten / Lemke, Thomas (Hrsg.): Science and Technology Studies. Frankfurt/Main, 369–403.
- Haraway, Donna (1995): Die Neuerfindung der Natur. Primaten, Cyborgs und Frauen. Frankfurt/Main.
- Heintz, Bettina (2010): Numerische Differenz. Überlegungen zu einer Soziologie des (quantitativen) Vergleichs. In: Zeitschrift für Soziologie 39, 162–181.
- Hirsch, Jorge E. (2005): An index to quantify an individual's scientific research output. In: Proceedings of the National Academy of Sciences 102, 16569–16572.
- Hong, Renyi (2016): Soft skills and hard numbers: Gender discourse in human resources. In: Big Data & Society. DOI: 10.1177/2053951716674237.
- Kitchin, Rob (2017): Thinking critically about and researching algorithms. In: Information, Communication & Society 20, 14–29.
- Kitchin, Rob (2014): Big Data, new epistemologies and paradigm shifts. In: Big Data & Society. DOI: 10.1177/2053951714528481.
- Leighton, Andrews / Benbouzid, Bilel / Brice, Jeremy / Bygrave, Lee A. / Demortain, David / Griffiths, Alex / Lodge, Martin / Mennicken, Andrea / Yeung, Karen (2017): Algorithmic Regulation. LSE Discussion Paper 85, London.
- Lopez, Paola (2019): Reinforcing Intersectional Inequality via the AMS Algorithm in Austria. In: Proceedings of the 18th Annual IAS–STS Conference on Critical Issues in Science and Technology Studies, i.E.
- Lyon, David (2014): Surveillance, Snowden, and Big Data: Capacities, consequences, critique. In: Big Data & Society. DOI: doi.org/10.1177/2053951-714541861.

- Mau, Steffen (2017): *Das metrische Wir. Über die Quantifizierung des Sozialen*. Berlin.
- Mayer-Schönberger, Viktor / Cukier, Kenneth (2013): *Big Data. Die Revolution, die unser Leben verändern wird*. München.
- Mazzocchi, Fulvio (2014): Could Big Data be the end of theory in science? In: *EMBO reports* 16, 1250–1255.
- Mützel, Sophie / Saner, Philippe / Unternährer, Markus (2018): *Schöne Daten! Konstruktion und Verarbeitung von digitalen Daten*. In: Houben, Daniel / Prietl, Bianca (Hrsg.): *Datengesellschaft. Einsichten in die Datafizierung des Sozialen*. Bielefeld, 111–132.
- Nachtwey, Oliver / Seidl, Timo (2017): *Die Ethik der Solution und der Geist des digitalen Kapitalismus*. IFS Working Paper 11, Frankfurt/Main.
- Neyland, Daniel / Möllers, Norma (2016): Algorithmic IF ... THEN rules and the conditions and consequences of power. In: *Information, Communication & Society*. DOI: 10.1080/1369118X.2016.1156141.
- Nobel, Safiya U. (2018): *Algorithms of Oppression. How Search Engines Reinforce Racism*. New York.
- O'Neil, Cathy (2018 [2016]): *Angriffe der Algorithmen*. Bonn.
- Paulitz, Tanja (2005): *Netzsubjektivität/en. Konstruktionen von Vernetzungen als Technologien des sozialen Selbst*. Münster.
- Prietl, Bianca (2019): BIG DATA: Inequality by Design? In: *Proceedings of the 2nd Weizenbaum Conference on Digital Inequality*. doi.org/10.34669/wi.cp/2.11.
- Prietl, Bianca / Houben, Daniel (2018): *Einführung. Soziologische Perspektiven auf die Datafizierung der Gesellschaft*. In: Houben, Daniel / Prietl, Bianca (Hrsg.): *Datengesellschaft. Einsichten in die Datafizierung des Sozialen*. Bielefeld, 7–32.
- Redden, Joanna (2018): *The Harm That Data Do*. In: *Scientific American* 319, 5, 76.
- Rossiter, Margaret W. (2003): *Der Matilda-Effekt in der Wissenschaft*. In: Wobbe, Theresa (Hrsg.): *Zwischen Vorderbühne und Hinterbühne. Beiträge zum Wandel der Geschlechterbeziehungen in der Wissenschaft vom 17. Jahrhundert bis zur Gegenwart*. Bielefeld, 191–210.
- Singer, Mona (2005): *Geteilte Wahrheit. Feministische Epistemologie, Wissenssoziologie und Cultural Studies*. Wien.
- Suresh, Harini / Gutttag, John V. (2019): *A Framework for Understanding Unintended Consequences of Machine Learning*. In: *ArXiv page*. <https://arxiv.org/abs/1901.10002> (3.7.2019).
- Sweeney, Latanya (2013): *Discrimination in Online Ad Delivery*. <https://dataprivacylab.org/projects/onlineads/1071-1.pdf> (3.7.2019).
- Thatcher, Jim (2014): *Living on Fumes: Digital Footprints, Data Fumes, and the Limitations of Spatial Big Data*. In: *International Journal of Communication* 8, 1765–1783.
- Weber, Jutta (2017): *Einführung*. In: Bauer, Susanne / Heinemann, Torsten / Lemke, Thomas (Hrsg.): *Science and Technology Studies*. Frankfurt/Main, 339–368.
- Winner, Langdon (1980): *Do Artifacts Have Politics?* In: *Daedalus* 109, 121–136.
- Yuval-Davis, Nira (2006): *Intersectionality and Feminist Politics*. In: *European Journal of Women's Studies* 13, 193–209.
- Zeit Online (2016): *Objektiv rassistisch. Beauty-AI*. (9.9.2016) <https://www.zeit.de/digital/2016-09/beauty-ai-kuenstliche-intelligenz-schoenheitswettbewerb-rassismus-algorithmen> (6.3.2019).
- Zweig, Katharina (2018): *Wo Maschinen irren können. Verantwortlichkeiten und Fehlerquellen in Prozessen algorithmischer Entscheidungsfindung*. Gütersloh.