

Über die epistemische Macht lernender Maschinen

Abstract

Dieser Text untersucht, was es für den Horizont menschlicher Erfahrungen und Handlungsräume bedeutet, wenn auf maschinellern Lernen beruhende Systeme in gesellschaftlich relevanten Bereichen als Akteur:innen auftauchen. Gesellschaftlich relevante Bereiche sind zumindest potentiell auch Bereiche einer politischen Gestaltungsmöglichkeit. Die Hauptthese dieses Artikels ist, dass diese Möglichkeit in Gefahr ist, wenn in solchen Bereichen menschliches Handeln durch maschinelle Analysen und Weichenstellungen ersetzt wird. Anhand des von der US-amerikanischen Soziologin Ruha Benjamin untersuchten Beispiels KI-basierter Recruiting-Verfahren wird zunächst ein phänomenologisches Verständnis des Problems erarbeitet, wobei Unerbittlichkeit und statistische Diskriminierung eine besondere Rolle spielen und die Beobachtung gemacht wird, dass an die Stelle eines handelnden ein herstellendes Verhältnis zur Welt treten kann. In einem zweiten Schritt wird die Bedeutung von Verstehen und Wissen für individuelle und gesellschaftliche Handlungsmöglichkeiten in den Mittelpunkt gestellt und deshalb untersucht, inwiefern wir es mit einem epistemischen Problem zu tun haben. Mit Kristie Dotson wird der Einsatz auf maschinellern Lernen beruhender Systeme in gesellschaftlich relevanten Bereichen drittens als epistemische Unterdrückung besonderer Schwere (>dritter Ordnung<) betrachtet, da hierdurch von hegemonialen bzw. zur Norm gemachten Sichtweisen abweichende Perspektiven auf kaum mehr zugängliche Weise aus der Produktion von Wissen ausgeschlossen werden können.

This text explores how the emergence of machine learning (ML) systems as actors in socially relevant domains alter the horizon of human experience and their space of action. Socially relevant domains are political in character, at least potentially. The main thesis of this paper will be that the introduction of ML-systems in socially relevant domains threatens the potential to act politically. First, using the example of AI-based recruiting processes studied by the U.S. sociologist Ruha Benjamin, a phenomenological understanding of the problem will be developed. Second, the importance of understanding and knowledge acquiring for individual and societal possibilities of action is brought into focus, showing to which extent we are dealing with an epistemic problem. ML-based systems exercise epistemic power in various ways, as they have interpretive power over data, they restrict access to knowledge in a specific way, and they prevent people from actively participating in generating that knowledge. Most importantly, they reconfigure realities and lead people to acquire socially relevant knowledge based purely on behavioral mass observation. This may result in a shift of competencies, at worst disclosing the possibility to act politically within the given socially relevant domains. Third, the use of ML-systems in socially relevant areas is then discussed with Kristie Dotson as a form of epistemic suppression of particular severity (>third order<). The usage of ML-systems can exclude perspectives that deviate from hegemonic or normative views within the production of knowledge in a way that is hardly accessible anymore.

Um was es in diesem Text geht. Und um was nicht.

Dieser Text stellt den Versuch dar, zu verstehen, was es für den Horizont möglicher menschlicher Erfahrungen und Handlungsräume bedeutet, wenn auf maschinellern Lernen beruhende Maschinen oder Systeme in gesellschaftlich relevanten Bereichen als Akteur:innen auftauchen. In diesem Satz sind einige klärensvalue Begriffe und Einschränkungen enthalten.

Mit einem auf maschinellern Lernen basierenden System (ML-System) meine ich hier ein softwarebasiertes System, das nicht im hergebrachten Sinne nach von Menschen formulierten Verfahrensvorschriften (Algorithmen) arbeitet, sondern dessen spätere Verfahren auf einem maschinell entwickelten, komplexen Modell der Trainingsdaten beruhen. In einer Studie der Fraunhofer-Allianz Big Data und Künstliche Intelligenz von 2018 klingt das so:

»Maschinelles Lernen bezweckt die Generierung von ›Wissen‹ aus ›Erfahrung‹, indem Lernalgorithmen aus Beispielen ein komplexes Modell entwickeln. Das Modell, und damit die automatisch erworbene Wissensrepräsentation, kann anschließend auf neue, potenziell unbekannte Daten derselben Art angewendet werden. Immer wenn Prozesse zu kompliziert sind, um sie analytisch zu beschreiben, aber genügend viele Beispieldaten – etwa Sensordaten, Bilder oder Texte – verfügbar sind, bietet sich Maschinelles Lernen an. Mit den gelernten Modellen können Vorhersagen getroffen oder Empfehlungen und Entscheidungen generiert werden – ganz ohne im Vorhinein festgelegte Regeln oder Berechnungsvorschriften.«¹

Mit der Bezeichnung Data Science verweist das akademische wie professionelle Feld, das sich heute mit maschinellern Lernen befasst, zudem auf einen wissenschaftlichen Anspruch, mit dem »mehr Wissen aus den Daten extrahiert werden soll, als diesen prima facie anzusehen ist«.²

Philosophisch interessant ist das allemal. Maschinelles Lernen setzt zu wesentlichen Teilen auf Korrelationen und induktives Schließen. Die ›Beispiele‹ oder Trainingsdaten sind digitale Repräsentationen der ›Gegenstän-

1 Inga Döbel, u.a.: *Maschinelles Lernen. Eine Analyse zu Kompetenzen, Forschung und Anwendung*, 2018, https://www.bigdata-ai.fraunhofer.de/content/dam/bigdata/de/documents/Publicationen/Fraunhofer_Studie_ML_201809.pdf (aufgerufen: 15.7.2022).

2 Dominik Meinecke: *Datenschutz und Data Science. Herausforderungen und Handlungsoptionen im Beschäftigtendatenschutz*, Baden-Baden 2022, S. 23.

de«, über die das System etwas lernen soll (in einem sehr einfachen Fall zum Beispiel die Pixel eines digitalen Fotos von einem Auto/Schaf/Menschen etc.). Ein einzelner Datensatz kann dabei sehr komplex sein und aus vielen Merkmalsdimensionen bestehen. Entwickler:innen solcher Systeme können nur indirekt – dabei aber mitunter auch unwillentlich – Einfluss auf deren spätere Funktionsweise nehmen, zum Beispiel durch Auswahl, Strukturierung³ oder Anordnung⁴ der Trainingsdaten, durch die Vorgabe des Lernalgorithmus oder Berechnungsmodells⁵ oder durch die Implementierung von Feedback-Schleifen. Der Trainingsprozess wird so lange fortgeführt, bis Ergebnisse als ausreichend präzise erachtet werden. Anschließend gilt das System als einsatzbereit. Ein häufig diskutiertes Problem in diesem Zusammenhang ist, dass die so entwickelten internen Modelle und Verfahren des Systems eine *Blackbox* darstellen und auf Grund ihrer Komplexität selbst für Entwickler:innen und Fachleute weitgehend undurchsichtig bleiben.⁶ So wünschenswert transparentere Systeme auch wären: das, was hier kritisiert werden soll, würde damit nicht behoben. Die Nutzung induktiver Schlüsse und das Zustandekommen der Repräsentation eines Menschen als Datensatz stellen – auch wenn oder gerade weil sie methodisch nicht neu sind – schon genügend Ansatzpunkte für eine kritische Betrachtung der genannten Verfahren dar. Doch dazu später.

In diesem Beitrag geht es um maschinelles Lernen und darauf basierende Systeme auch nur insofern, als sie in gesellschaftlich relevanten Bereichen zum Einsatz gebracht werden. Als gesellschaftlich relevant sollen dabei Situationen und Bereiche gelten, in denen über Menschen geurteilt und damit über deren Möglichkeit zur Teilhabe an sozialen, kulturellen, ökologischen, ökonomischen und politischen Ressourcen entschieden wird. Wer

-
- 3 Maschinelles Lernen kann auf strukturierten oder auf unstrukturierten Trainingsdaten erfolgen. Strukturierte Daten sind solche, die in einem vorgegebenen Format oder Schema vorliegen, z.B. wenn in einem Text bestimmte Abschnitte maschinenlesbar als <title> oder <author> gekennzeichnet sind.
 - 4 Dass die Reihenfolge der Eingabe der Trainingsdaten den von der Maschine erlernten Algorithmus beeinflussen kann, wurde erst kürzlich im Rahmen eines Angriffsversuchs belegt. Siehe Cory Doctorow: »Attacking Machine Learning Training by Re-ordering Data«, 26.05.22, <https://doctorow.medium.com/attacking-machine-learning-training-by-re-ordering-data-c59f7ec0f18e> (aufgerufen: 14.6.2022).
 - 5 Der Methodenkanon des maschinellen Lernens umfasst eine große Anzahl vorgefertigter Lern- oder Berechnungsverfahren, die heute von klassischen statistischen Verfahren bis hin zu tiefen neuronalen Netzen reichen und deren Auswahl vom Anwendungsfall und von der Erfahrung der damit betrauten Personen abhängt.
 - 6 Vgl. Jenna Burrell: »How the Machine ›Thinks‹. Understanding Opacity in Machine Learning Algorithms«, in: *Big Data & Society* 3 (2016), Heft 1, S. 1–12.

über den Zugang zu den Ressourcen entscheidet, nach welchen Maßgaben eine solche Entscheidung erfolgen sollte und tatsächlich erfolgt, wie die möglichen Wege zur Teilhabe gestaltet sind, und was die Teilhabe an den jeweiligen Ressourcen bedeutet, welche Ein- und Ausschlusskriterien dabei wirksam werden und wie diese miteinander verwoben sind – all dies sind Fragen, die immer wieder gestellt und neu verhandelt und beantwortet werden können, zumindest potenziell. Gesellschaftlich relevante Bereiche haben also eine kontingente Geschichte mit auslotbaren Freiheitsgraden, sie sind Bereiche einer möglichen politischen Gestaltung. Bis jetzt. Die Hauptthese dieser Arbeit wird sein, dass diese Möglichkeit in Gefahr ist, und dass dem Wesen dieser Gefahr nur entgegengewirkt werden kann, indem sie mitsamt ihrer Verwurzelung in den Tiefen hegemonialer Wissenspraxen der westlichen Moderne in Frage gestellt wird.

Dies soll exemplarisch anhand eines konkreten Falls gezeigt werden. Die Regelung des Zugangs zur Arbeitswelt in Recruitingverfahren wird von mir als ein im obigen Sinne gesellschaftlich sogar besonders relevanter Bereich aufgefasst. Eine Bestätigung dieser Einschätzung findet sich in der gerade (2022) erst erschienenen Dissertation des Juristen Dominik Meinecke:

»Im Beschäftigungskontext sind Datenschutz und der Einsatz von Algorithmen besonders sensible Gebiete, nicht zuletzt aufgrund der Vielzahl der Betroffenen. Konsequenterweise ist daher die Wertung der EU-Kommission, die Anwendungen zur Entscheidung über Einstellungen, Beförderungen und Kündigungen in ihrem Vorschlag für ein KI-Gesetz grundsätzlich als Hochrisiko-Systeme einstufen möchte.«⁷

Dennoch werden im Personalmanagement zahlreicher Unternehmen bereits heute auf maschinellem Lernen basierende Systeme eingesetzt.⁸ Wo Implementierung und Ausgabe des Algorithmus menschliches Urteilen und Handeln *ersetzen*, werde ich sagen, dass das auf maschinellem Lernen beruhende System zum:zur Akteur:in in einem gesellschaftlich relevanten Bereich wird. Was es für einen Unterschied macht, ob in gesellschaftlich relevanten Bereichen Menschen oder Maschinen Entscheidungen treffen, soll im Verlauf der Arbeit deutlich werden.

7 Meinecke: *Datenschutz und Data Science*, S. 25.

8 Je nach Anbieter ist zum Beispiel von HR Analytics, People Analytics oder Workforce Analytics die Rede. Bei Meinecke finden sich relativ aktuelle Zahlen zur Nutzung derartiger Software in Europa. Vgl. ebd., S. 54f.

Wenn Menschen Maschinen Handlungshoheit übertragen, dann hoffen sie dabei (meistens), dass diese nicht nach dem Zufallsprinzip, sondern so agieren, dass vorliegende Informationen auf rationale Weise in Aktionen umgesetzt werden. Insofern sie als wissensbasierte Systeme angesprochen und eingesetzt werden, sollen sie hier entsprechend als epistemische Ressourcen und als Teil epistemischer Systeme angesehen werden. Diese Auffassung wird durch den Begriff Data Science nur bestärkt. Wiederum gilt: ich rede hier nicht einer Neudefinition von Begriffen aus dem Bereich menschlicher Kompetenzen das Wort, sondern umgekehrt: im gesellschaftlich relevanten Bereich messe ich die Maschine an dem, was sie ersetzt – und werde darlegen, dass dabei viel auf der Strecke bleibt, worum emanzipatorische Bewegungen lange gekämpft haben.

Unerbittlichkeit und Verweigerung des Dialogs

Es ist das Kennzeichen derjenigen auf maschinellern Lernen (ML) beruhenden Systeme, von denen hier die Rede sein soll, dass sie es zu erlauben scheinen, Aussagen über Menschen zu treffen, ohne sie direkt befragen oder zu Wort kommen lassen zu müssen. Aussagen in diesem Sinne können entweder Klassifikationen (also die Einordnung von Einzelnen in vorgegebene Kategorien oder Gruppen) oder mit einem Wahrscheinlichkeitsindex versehene Prognosen über künftiges Verhalten (künftige *Performance*) sein.⁹ Grundlage dieser Ergebnisse ist der Vergleich des vorliegenden Datensatzes mit bereits vorhandenen Datensätzen mit denselben Merkmalsstrukturen und bekannten Ergebnissen. Die so erfolgenden Klassifikationen oder Prognosen führen, je nach Einsatzgebiet, zu einer Weichenstellung in umgebenden Prozessen, können also z.B. entscheidend werden dafür, ob jemand ein Jobangebot erhält oder nicht, welche Informationen jemand in einer App zu sehen bekommt, ob jemand einen Kredit bekommt oder, wie ebenfalls in den USA bereits praktiziert, ob er oder sie vorzeitig aus der Haft entlassen wird.¹⁰

9 Meinecke unterscheidet drei Arten möglicher Analysen: »Die deskriptive (>beschreibende<) Analyse beschreibt einen Ist-Zustand, die prädikative (>vorhersagende<) Analyse liefert die Voraussage eines künftigen Zustands bzw. einer Entwicklung und die präskriptive (>vorschreibende<) Analyse bietet Lösungen für die vorhergesagten Entwicklungen an« (ebd., S. 55).

10 Juristisch erfüllen solche Anwendungen in der Regel die in der europäischen Datenschutzgrundverordnung (DSGVO) gegebene Definition des *Profiling*s. Vgl. ebd., S. 81.

Rein ML-basierte Systeme können nicht anders als den in die Daten eingeschriebenen sozialen Ist-Zustand fortzuschreiben und tendieren so dazu, soziale Ungleichheiten zu reproduzieren. Die dahinterliegenden Mechanismen wurden von anderen bereits ausführlich beschrieben und kommentiert.¹¹ Über die daraus folgenden Ergebnisverzerrungen (rassistische, genderbasierte uvm.) wurde ebenfalls schon viel geschrieben und diskutiert.¹² In der Folge – und in Anerkennung des Problems – wurden Ansätze vorgestellt, die es erlauben, ingenieurmäßig in diese Mechanismen einzugreifen, wodurch eine Überwindung existierender Benachteiligungen durch eben diese Systeme technisch ermöglicht werden soll.¹³

Derartige Vorschläge berühren das eigentliche Problem meiner Einschätzung nach jedoch nur an der Oberfläche. Den Einsatz ML-basierter Systeme in bestimmten gesellschaftlich relevanten Prozessen unter dem Aspekt epistemischer Unterdrückung zu betrachten, erlaubt dagegen eine weiterführende Kritik.¹⁴

Die wesentliche Anregung zu diesem Text verdanke ich der Lektüre von Kristie Dotsons Aufsatz »Conceptualizing Epistemic Oppression« von 2014,¹⁵ auf den ich später noch genauer eingehen werde. Ursprünglich aufgerüttelt hat mich Ruha Benjamins soziologische Studie *Race After Technology*, die sich vornimmt, »die wirkmächtigen Vorannahmen und Werte zu decodieren, die in die materielle und digitale Architektur unserer Welt eingeschrieben sind«.¹⁶ Dieser habe ich die Schilderungen von Betroffenen und zahlreiche Hinweise auf ML-Systeme in Recruitingverfahren entnommen.

11 Vgl. z.B. Cathy O'Neil: *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York 2016. Auch in Alma Kolleck und Carsten Orwat: *Mögliche Diskriminierung durch algorithmische Entscheidungssysteme und maschinelles Lernen – ein Überblick*, Berlin 2020. Oder in Kate Crawford: *Atlas of AI*, New Haven, CT. 2021.

12 Z.B. Ruha Benjamin: *Race After Technology: Abolitionist Tools for the New Jim Code*, Cambridge 2019.

13 In der Forschung diskutiert als »algorithmic fairness« oder »fair AI«.

14 Aus technikphilosophischer Sicht könnte es sich auch als lohnend erweisen, auf diesen Fall bezogen spezielle Erscheinungsformen der Medialität von Technik herauszuarbeiten, was hier aber nicht geleistet werden kann.

15 Kristie Dotson: »Conceptualizing Epistemic Oppression«, in: *Social Epistemology: A Journal of Knowledge, Culture and Policy* 28 (2014), Heft 2, S. 115–138.

16 Ruha Benjamin: *Race After Technology: Abolitionist Tools for the New Jim Code*, Hoboken, NJ. 2019, S. 12, eigene Übersetzung.

ML-basierte Technologien werden in Bewerbungsverfahren bereits vielfach eingesetzt.¹⁷ Ein Anbieter, der in Benjamins Fallstudien vorkommt, ist die Firma MYA,¹⁸ ein in den USA ansässiges Unternehmen, das sich auf die digitale Durchführung von Recruitings spezialisiert hat. Der Preis der Software mache sich für Unternehmen, so die Marketingbehauptung, schnell durch Einsparungen von Personal im Personalmanagement bezahlt. Die Software selbst besteht aus verschiedenen Modulen, zu denen auch ein Chatbot gehört, der erste Recruitinggespräche mit potenziellen Bewerber:innen führt. Auch Videoanalysen sind möglich. Andere Anbieter, wie HireVue¹⁹ bieten daneben auch automatisierte Assessments, Terminvergaben, Auswahlverfahren und affektbasierte Persönlichkeitsanalysen an.²⁰ Die Systeme werden ständig ausgebaut und kommen weltweit immer häufiger zum Einsatz: zu den bekannten Kunden allein von MYA gehören global operierende Leiharbeits- und Recruiting-Firmen wie Adecco, Deloitte, Randstad und Hays, aber auch Großkonzerne wie L'Oréal. MYA wurde im Mai 2021 von der Stepstone GmbH, Betreiberin einer etablierten Online Recruiting Plattform, übernommen.

Oft wird argumentiert, dass der Technologieeinsatz in Bewerbungsverfahren, wenn überhaupt, so lediglich ohnehin bestehende Diskriminierungen fortschreibe und daher in sich keine Veränderung darstelle. Das ist jedoch nicht richtig. Benjamin zeigt sehr eindrücklich, dass durch das Ausschalten des Faktors Mensch soziale Aufstiegschancen für Angehörige historisch benachteiligter Gruppen komplett eliminiert werden könnten. Denn sobald das Aussortieren der Bewerbungsunterlagen überall nach derselben Logik – weil durch dieselbe Software – erfolgt, entfällt die Möglichkeit, irgendwo vielleicht doch einmal einer anders oder weniger vorurteilsbelasteten Einstellungspolitik zu begegnen: Jede weitere Bewerbung ist damit von vornherein zwecklos. Der Computerwissenschaftler

17 Eine Auflistung der Bereiche, in denen auf ML basierende Systeme im Recruiting zur Anwendung kommen können, von der Strukturierung des Bewerber:innenfelds, über Persönlichkeits- und andere Tests, Sprach- und Videoanalysen bis hin zu vollautomatisierten Einstellungsentscheidungen findet sich bei Meinecke: *Datenschutz und Data Science*, S. 57–67.

18 <https://www.stepstone.com/en/mya/>.

19 <https://www.hirevue.com>.

20 Mit KI-basierter Affekterkennung und deren wissenschaftlich höchst umstrittenen Grundannahmen befasst sich Kate Crawford sehr ausführlich. HireVue wird von ihr explizit als Anbieter genannt, der diese für Bewerbungsverfahren kommerzialisiert. Vgl. Crawford: *Atlas of AI*, S. 151–180.

Arvind Narayanan aus Princeton stellt dem die potenzielle Vielfalt menschlicher Vorurteile gegenüber:

»Human decision makers might be biased, but at least there is a ›diversity‹ of biases. Imagine a future where every employer uses automated resume screening algorithms that all use the same heuristics, and job seekers who do not pass those checks get rejected everywhere«. ²¹

Andererseits wird zugunsten des Technologieeinsatzes wie oben angerissen gelegentlich argumentiert, dass mit seiner Hilfe Diskriminierungen effektiv, schneller und flächendeckender ausgeschaltet werden könnten als in menschenbesetzten Prozessen. Das mag zunächst nachvollziehbar scheinen, ist aber ein Fehlschluss, denn Diskriminierung (zunächst im wertneutralen Sinne) ist ja genau Aufgabe und Verfahrensweise dieser Systeme, so dass eigentlich alle dafür herangezogenen Merkmale auf ihre Relevanz und die Fairness gegenüber den darin nachteilig bewerteten Menschen untersucht und diskutiert werden müssten. ^{22,23} Kolleck und Orwat diskutieren diesen Punkt im Absatz »Statistische Ungleichbehandlung und statistische Diskriminierung«. ²⁴ Unter statistischer Diskriminierung verstehen sie eine »Form der technikbasierten Diskriminierung, die sich im Kontext komplexer algorithmischer Systeme – insbesondere bei lernenden Systemen – häufig zeigt [...] Sie stützt sich auf Ersatzinformationen und stellt eine Sonderform der sozialen Diskriminierung dar«. ²⁵ »Ersatzinformationen« ²⁶ bezeichnen dabei Merkmale, die als Hinweise auf andere Merkmale verstanden werden. Als Beispiel nennen Kolleck und Orwat das Alter, das herangezogen wird,

21 Arvind Narayanan, Princeton University, zitiert nach Benjamin: *Race After Technology*, S. 133.

22 Diese sind allerdings nicht immer herauszufinden, vor allem dann nicht, wenn es sich um ein System handelt, das auf unstrukturierten Daten operiert, dem also Kategorien wie Alter, Geschlecht etc. nicht vorgegeben werden – wie bei Bild- und Spracherkennungsalgorithmen.

23 Meinecke erörtert darüber hinaus die Frage, inwiefern an die aus den Daten gewonnenen Informationen auch der sich aus dem Datenschutzrecht ergebende »Grundsatz der Richtigkeit, bestehend aus sachlicher Richtigkeit und Aktualität« als Maßstab anzulegen sei. Die juristische Antwort hängt stark davon ab, ob diese Informationen als personenbezogene Daten gewertet werden können, wofür es gute Argumente gibt. Vgl. Meinecke: *Datenschutz und Data Science*, S. 130ff.

24 Kolleck: *Mögliche Diskriminierung durch algorithmische Entscheidungssysteme und maschinelles Lernen*, S. 37.

25 Ebd., S. 34.

26 Ebd.

um in gesellschaftlich anerkannter Weise auf Verantwortungsbewusstsein im Umgang mit alkoholischen Getränken schließen zu dürfen:

»Die Altersdiskriminierung wird meist nicht als problematisch wahrgenommen, solange sie ein gesellschaftliches legitimes Ziel verfolgt und der statistische Zusammenhang zwischen der Ersatzinformation Alter und der Zielinformation Verantwortungsbewusstsein/Sinnesstärke evident erscheint, selbst wenn sie auf den Einzelfall nicht zutreffen mag.«²⁷

Ersatzinformationen werden statistischen Korrelationen entlehnt und »auf den Einzelnen übertragen [...], ohne dass der/dem Einzelnen die Möglichkeit gegeben wird, diese Annahme zu widerlegen.«²⁸ Maschinelles Lernen automatisiert die Beurteilung des Einzelnen anhand von Ersatzinformationen, wobei zu betonen ist, dass die Wahl derselben hier weder von Sozialwissenschaftler:innen vorgenommen wird noch demokratisch legitimiert ist. Statistische Diskriminierung in automatisierten Bewerbungsverfahren bedeutet, dass Bewerber:innen *prinzipiell* nicht nach ihrem tatsächlichen Verhalten, ihren individuellen Fähigkeiten, Eigenschaften, Motiven und Zielen, sondern nur noch anhand von Ersatzinformationen beurteilt werden.²⁹

Kolleck und Orwat weisen zwar darauf hin, dass hierin aus juristischer Sicht grundsätzlich eine Unvereinbarkeit mit der Einzelfallgerechtigkeit, eine Verletzung des Rechts auf Selbstdarstellung und der zugrundeliegenden Garantie der freien Entfaltung der Persönlichkeit bestehen kann sowie ein Verstoß gegenüber des in der Datenschutzgrundverordnung festgelegten Rechts einer betroffenen Person, »nicht einer ausschließlich auf einer automatisierten Verarbeitung beruhenden Entscheidung unterworfen zu werden, die ihr gegenüber rechtliche Wirkung entfaltet oder sie in ähnlicher Weise erheblich beeinträchtigt.«³⁰ Trotzdem stellen sie unter »Handlungsoptionen«³¹ im Umgang mit *Algorithmischen Entscheidungssystemen* deren Anwendung in gesellschaftlich relevanten Bereichen nicht grundsätzlich in Frage, sondern diskutieren Maßnahmen zur Herstellung von Trans-

27 Ebd., S. 35.

28 Ebd., S. 36.

29 Selbst wenn wir davon ausgehen müssen, dass manche Unternehmen Bewerber:innen ohnehin einfach als Leistungserbringer:innen »objektivieren« und nicht an Persönlichkeiten interessiert sind, dürfte das, was Kolleck und Orwat als statistische Diskriminierung bezeichnen, nicht einmal im Sinne der Auftraggeber:innen sein.

30 Ebd., S. 42.

31 Ebd., S. 61–69.

parenz, Kontroll- und Evaluierungsmöglichkeiten sowie die Möglichkeiten juristischer Regulierung zur »Reduktion von Diskriminierungsrisiken«.³² Gegen die statistische Diskriminierung empfehlen sie lediglich eine »Öffnung der bislang besonders geschützten Merkmale«³³, ohne jedoch zu erklären, wie eine entsprechende Anpassung der Gesetze aussehen könnte.

Wenn es um ungerechtfertigte Benachteiligungen durch automatisierte Personalentscheidungen geht wird gerne eingeworfen, dass auch Menschen in Personalabteilungen vorurteilsbehaftet und Bewerber:innen gegenüber nicht rechenschaftspflichtig seien. Auf den ersten Punkt lässt sich mit Narayanan antworten, dass bei einer Vielfalt von Gegenübern Diskriminierung lokalisiert und lokal umgangen werden kann, während sie bei einem einheitlichen und automatisch agierenden Gegenüber globalen Charakter bekommt und – das wird später noch wichtig – nicht auf individueller, sondern nur noch auf grundsätzlicher Ebene angegangen werden kann. Diesen Punkt möchte ich vorerst schon einmal als die »Unerbittlichkeit« von Automaten bezeichnen. Der zweite Teil des Einwands wird dadurch entkräftet, dass die Frage der Verantwortung und der Problembhebung bei Diskriminierungsvorwürfen (sofern sie sich überhaupt nachweisen lassen, was in beiden Fällen schwer ist) beim Einsatz von auf ML basierenden Entscheidungssystemen noch weitgehend ungeklärt ist.³⁴

Im Erleben der Jobsuchenden erkennt die Soziologin Benjamin einen weiteren Unterschied zu herkömmlichen Bewerbungsverfahren:

»[A]pplicants are frustrated not only by the lack of human contact, but also because they have no idea of how they are evaluated and why they are repeatedly rejected. One job seeker described questioning every small movement and micro-expression and feeling a heightened sense of worthlessness because the company couldn't even assign a person for a few minutes. The whole thing is becoming less human.«³⁵

Sogar eine körperliche Verunsicherung wird erzeugt, sobald im automatisierten Bewerbungsverfahren neben den schriftlichen Unterlagen auch Videomaterial ausgewertet wird. Die Unsicherheit bezüglich der Funktionsweise der Maschine verstärkt die Frustration, und die Bewerber:innen sehen kaum eine Möglichkeit, zu erfahren, was sie tun können, um ihre

32 Ebd., S. 62.

33 Ebd., S. 67.

34 Dies wird in der Forschung unter »accountability of AI« diskutiert.

35 Benjamin: *Race After Technology*, S. 132.

Chancen in der Zukunft zu erhöhen. Prinzipiell kann – wie oben gesehen – jedes, auch ein selbst noch nie wahrgenommenes Detail des eigenen Erscheinens in der Welt für das maschinelle Gegenüber (oder für die stochastischen Verfahren dahinter) ein Merkmal sein, das zum Ausschluss führt. Vor dieser Ungewissheit entwickelt sich das Gefühl, nichts richtig machen zu können und nichts wert zu sein, nicht einmal den kurzen Einsatz einer menschlichen Arbeitskraft.

Ein als wesentlich wahrgenommener Unterschied zwischen Mensch und Maschine als Gegenüber scheint also tatsächlich die wenigstens prinzipiell vorhandene Möglichkeit zu sein, Erklärungen zu finden. Dass dieser Unterschied zu ethischen Einwänden gegen den Einsatz solcher Systeme führen kann, ist in der Fachwelt auch längst angekommen, was die relativ neue Forschungsrichtung Explainable AI (etwa: erklärungs-fähige Künstliche Intelligenz) bezeugt. Dies ist nicht nur für juristische Belange wichtig,³⁶ sondern, wie Benjamin uns lehrt, besonders auch im Hinblick auf die psychosozialen Auswirkungen auf die individuell Betroffenen und die Gesellschaft als Ganzes. Das von Benjamin angeführte Fallbeispiel lehrt uns, den Blick darauf zu richten, dass der Einsatz solcher Dienste und Produkte sich auf die Möglichkeit zur Entwicklung persönlicher Handlungsfähigkeit in der Welt auswirkt und macht damit einen besonderen Aspekt der Medialität von Technik sichtbar.

Handlungsfähigkeit beruht auf einem Verständnis der Welt, einem Verständnis möglicher Zusammenhänge zwischen dem, was wir tun und dem, was dann passiert. Nur so ist zielorientiertes Handeln überhaupt möglich. Soziales Lernen findet nur mit und vor anderen statt. In Bewerbungsverfahren eröffnet ein qualifiziertes Feedback die Möglichkeit, Ursachen des Scheiterns zu erkennen und unser Verhalten so anzupassen, dass die Erfolgsaussichten beim nächsten Versuch größer werden und kann so dabei helfen, Chancen zur Teilhabe zu erhöhen.

Freilich werden auch in herkömmlichen, durchgängig von Menschen betreuten Bewerbungsverfahren oft Absagen ohne Erklärung verschickt oder Bewerbungsschreiben gar nicht erst beantwortet. Dennoch besteht gegenüber einem automatisierten Aussortierungsverfahren ein wesentlicher Unterschied. Denn um sich die Ablehnung durch einen Menschen zu erklären, haben wir immer mindestens zwei Möglichkeiten: einerseits können

36 Dies erläutert z.B. Joanna Bryson: *Six Kinds of Explanation for AI (one is useless)*, 19.09.2019. <https://joanna-bryson.blogspot.com/2019/09/six-kinds-of-explanation-for-ai-one-is.html> (aufgerufen: 11.12.21).

wir direkt nachfragen (und haben dabei gelegentlich auch Erfolg), andererseits haben wir die Möglichkeit des vorgestellten Dialogs, d.h. wir können versuchen, uns in den:die Andere:n hineinzusetzen und uns durch dessen:deren (vorgestellte) Augen zu sehen. Wie treffsicher wir dabei sind, können wir beeinflussen, indem wir uns zum Beispiel über gängige Entscheidungskriterien der Personalabteilungen informieren. Dies jedoch ist unmöglich, sobald wir es mit Maschinen zu tun haben, deren Entscheidung auf einer Auswertung von Ersatzinformationen und Gewichtungungen beruht, die die Bewerber:innen nicht beeinflussen können und die nicht einmal deren Entwickler:innen vollständig kennen. Mit dem Einsatz von Maschinen entfällt nicht nur wie oben beschrieben die Möglichkeit, als individuelle Person in Erscheinung zu treten, sondern gleichzeitig auch die des vorgestellten Dialogs und damit der Selbstreflexion im Lichte eines Gegenübers. Und was die Möglichkeit der direkten Nachfrage angeht: ein Mensch muss zwar nicht antworten, aber die Maschine kann es nichtmal – zumindest nicht im Sinne eines qualifizierten Feedbacks.

Es wird also etwas für das In-der-Welt-Sein ganz Wesentliches verändert, wenn Menschen ihr Auftreten im Blick auf die Kriterien ML-basierter Systeme optimieren müssen, weil diese in gesellschaftlich relevanten Prozessen die Stelle von Menschen einnehmen. Diese Veränderung besteht wesentlich darin, dass die Möglichkeit entfällt, Handlungsfähigkeit durch Kommunikation innerhalb eines menschlichen Bezugsgewebes herzustellen.³⁷

Es verwundert daher nicht, dass die Betroffenen nach anderen Wegen suchen, sich ein Wissen über die Ursachen des Ausschlusses anzueignen und ihre Handlungsfähigkeit wiederherzustellen.

»Some job seekers are already developing ways to subvert the system by trading answers to employer's tests and by creating fake applications as informal audits of their own. In fact, one HR employee for a major company recommends slipping the words ›Oxford‹ or ›Cambridge‹ into a CV in invisible white text, to pass the automated screening.«³⁸

Menschen entwickeln also Strategien und finden Tricks, mit denen sie auch im kommunikationsfeindlichen Umfeld ihren Zielen näherkommen

37 Petra Gehring beschreibt in ihrem Buch *Über die Körperkraft der Sprache* anhand des Mythos der Echo sehr drastisch, inwiefern der Ausschluss von der Möglichkeit, zu kommunizieren Menschen seelisch wie körperlich Gewalt antun und sie letztlich vernichten kann. Vgl. Petra Gehring: *Über die Körperkraft der Sprache*, Frankfurt am Main 2019.

38 Benjamin: *Race After Technology*, S. 133.

können. Manche davon können durch den Dialog mit Expert:innen in Erfahrung gebracht werden. Als Beispiel hierfür nennt Benjamin den in Bewerber:innenkreisen kursierenden Tipp einer Personalerin, die Worte ›Cambridge‹ oder ›Oxford‹ irgendwo in der Bewerbung unterzubringen, und zwar in weißer Schrift, so dass diese zwar maschinenlesbar, aber für menschliche Augen nicht als Betrugsversuch erkennbar sind. Alternativ dienen individuelle oder kollektiv organisierte Testläufe dazu, die Funktionsweise des selektierenden Systems zu erkennen, um einem befürchteten Ausschluss zu entgehen. Das maschinelle Gegenüber kann also auf experimentelle Art zu einer Art des Dialogs gezwungen werden, weil die Software sich zwar als Blackbox präsentiert, aber nicht umhinkann, auf einen gegebenen Input einen Output zu geben. Und auf diesen Input-Output-Paaren können nun die Nutzer:innen ihren eigenen (menschlichen oder ebenfalls softwarebasierten) Algorithmus zur Mustererkennung laufen lassen. Aus solchen Erkenntnissen ergeben sich neue Handlungsmöglichkeiten.

Derartige Coping-Strategien zeigen nicht nur die Bedeutung des Verstehens für die Herstellung von Handlungsfähigkeit, sondern auch, inwiefern sich der Dialog mit Menschen von dem mit Maschinen unterscheidet. An die Stelle eines in vernünftige Handlungen umsetzbaren Abgleichs von Selbst- und Fremdwahrnehmung tritt als einzig verbleibender Ausweg die trial-and-error-getriebene technische Manipulation des fremden Systems. In Anlehnung an Hannah Arendt³⁹ könnte der Gegensatz auch wie folgt beschrieben werden: Ein Erscheinungsraum im menschlichen Bezugsgewebe, in dem das ›Wer‹ sichtbar werden könnte, ist entfallen, die Bewerber:innen werden nurmehr als ein ›Was‹ gesehen, und an die Stelle ihres handelnden tritt ein herstellendes Verhältnis zur Welt.

Was damit verloren geht, ist auch ein Raum, in dem Politisierung möglich ist. Politisierung möchte ich in diesem Zusammenhang mit Tobias Matzner verstehen als »Handlungen, die dazu führen, dass grundlegende Überzeugungen im Weltbild einer Gruppe von Menschen ›verflüssigt‹ werden.«⁴⁰ Das wird gleich noch wichtig, wenn es um die Resilienz epistemischer Systeme geht.

39 Vgl. Hannah Arendt: *Vita Activa oder Vom tätigen Leben*, München 2020.

40 Tobias Matzner: *Vita Variabilis. Handelnde und ihre Welt nach Hannah Arendt und Ludwig Wittgenstein*, Würzburg 2013, S. 180.

An dieser Stelle möchte ich kurz zusammenfassen, was bis hierher deutlich geworden sein sollte:

- Eine Betrachtung der gesellschaftlichen Auswirkungen des Einsatzes von ML-basierten Entscheidungssystemen greift zu kurz, wenn sie nur die Fortschreibung bereits bestehender Diskriminierungsebenen in den Blick nimmt. Auf dieser Ebene besteht ein grundsätzlicher Unterschied zwischen Menschen und Maschinen nur im Ermessensspielraum der ersten und in der *Unerbittlichkeit* letzterer.⁴¹
- Maschinelles Lernen auf unstrukturierten Daten beruht systematisch auf statistischer Diskriminierung – nur ist diese (anders als zum Beispiel bei der Gesetzgebung in Sachen Jugendschutz) bislang meist den Augen der Öffentlichkeit entzogen.
- Was damit generell unterdrückt und aus dem Bereich des (von einem ML-basierten System) Wissbaren ausgeschlossen wird, sind vom Mainstream abweichende Kulturen, soziale Zusammenhänge und individuelle Lebensentwürfe.
- Durch den Wegfall zwischenmenschlicher Erscheinungsräume erleben sich die Betroffenen als wertlos. Sie verlieren eine Möglichkeit, als ›Wer‹ in Erscheinung zu treten und werden nur noch als ›Was‹ behandelt. Wie mit Benjamin gesehen, kann dies als sehr verunsichernd und schmerzhaft empfunden werden.
- Um in den Bereichen, in denen Menschen durch Maschinen ersetzt wurden, ›handlungsfähig‹ zu bleiben, müssen die Betroffenen ihrerseits auch in ein herstellendes Verhältnis zur Welt treten.

All dies wurde anhand der soziologischen Studie Benjamins und der von ihr diskutierten Beispiele aus Bewerbungsverfahren illustriert, um ein phänomenologisches Verständnis des Problems zu ermöglichen. Die so herausdestillierten Erkenntnisse und Fragen sind ein Vorschlag und können im besten Fall dazu dienen, den Einsatz ML-basierter Entscheidungssysteme auch in anderen gesellschaftlich relevanten Bereichen besser zu beurteilen.

41 Dieser Aspekt könnte noch ausgeweitet werden, wenn wir Hannah Arendts Ausführungen zum Verzeihen und zum Versprechen dagegenhalten, was an dieser Stelle jedoch nicht in ausreichender Dichte getan werden kann.

Die Unterdrückung des Individuellen und des Zwischenmenschlichen als epistemisches Problem

Bis hierhin habe ich dargelegt, wodurch eine Unterdrückung des Individuellen und des Zwischenmenschlichen entsteht, wenn in bestimmten Bereichen bislang von Menschen wahrgenommene Tätigkeiten und Entscheidungen an ML-basierte Systeme übergeben werden, und welche Auswirkungen das auf die Betroffenen hat.

Im Folgenden soll das Ganze nun als epistemisches Problem gefasst werden. Dabei geht es mir um den Punkt, dass durch den Einsatz solcher Systeme nicht Einzelne oder bestimmte Gruppen, sondern Menschen in ihrer Eigenschaft als individuelle, veränderungsfähige, schöpferische, wissende und handelnde Subjekte generell ausgeschlossen werden aus dem Zustandekommen von Wissen, das dennoch zur Grundlage für gesellschaftlich relevante Entscheidungen wird. Dieser Ausschluss betrifft Menschen gleichermaßen als Wissende über sich selbst (die in ihrem Handeln nämlich gerade Auskunft geben könnten über das was fehlt, nämlich ihre individuellen Fähigkeiten und Potentiale) wie auch als Produzent:innen von Wissen (nämlich darüber, welche Kriterien denn nun gesellschaftlich wünschenswert wären für eine bestimmte Art von Entscheidungen, und welche nicht).

Mit Kristie Dotson möchte ich epistemische Unterdrückung hier begreifen als dauerhaften, gewissermaßen institutionalisierten Ausschluss eines Subjekts und dessen Perspektive aus der Produktion von Wissen. Sie schreibt:

»Epistemic oppression refers to persistent epistemic exclusion that hinders one's contribution to knowledge production«. ⁴²

In ihrem 2014 erschienenen Aufsatz »Conceptualizing Epistemic Oppression« schlägt sie vor, drei Arten epistemischer Unterdrückung zu unterscheiden. Damit möchte sie dem Einwand begegnen, dass epistemische Unterdrückung immer auf Formen sozialer und historischer Unterdrückung zu reduzieren ist. Ihre These lautet, dass es eine Art epistemischer Unterdrückung gibt, bei der das nicht der Fall ist, und dass sich die einen von den anderen Fällen am deutlichsten anhand der Überlegung unterscheiden lassen, was anzugreifen wäre, um sie zu überwinden.

42 Dotson: »Conceptualizing Epistemic Oppression«, in: *Social Epistemology: A Journal of Knowledge, Culture and Policy* 28, S. 1.

Dotsons Analyse baut auf feministischen Ansätzen auf, die drei wesentlich soziale Aspekte von Wissen sichtbar gemacht haben: 1) die schon durch unsere Leiblichkeit bedingte Situiertheit als Wissende, die entscheidend dafür ist, *was* wir wissen oder erfahren können, 2) die wechselseitige Abhängigkeit von gemeinsam geteilten Ressourcen oder Schemata wie beispielsweise einer gemeinsamen Sprache, die das *Wie* unseres Wissens bestimmt, und schließlich 3) die Resilienz epistemischer Systeme, verstanden als die geteilten Annahmen und akzeptierten Mittel, die in einer gegebenen Gemeinschaft für die kollektive Wissensproduktion zur Verfügung stehen.⁴³

Auf diesen Ansätzen aufbauend, unterscheidet Dotson verschiedene Arten epistemischer Unterdrückung anhand der Frage, durch welche Prozesse sie jeweils erkannt und überwunden werden könnten. Sie denkt also – und das ist wichtig – von der Veränderbarkeit her. Wo Unterdrückung ist, will sie wissen, wie sie überwunden werden kann.

Dotson spricht von epistemischer Unterdrückung erster Ordnung, wenn nicht die institutionalisierten Grundlagen der Wissensproduktion selbst in Frage gestellt werden müssen, sondern lediglich ihre *ineffiziente Anwendung*. Sie könnte zum Beispiel in einer Gesellschaft entstehen, die eine prinzipielle Gleichberechtigung aller Perspektiven zwar vorsieht, aber nicht lebt. Die Ignoranz gegenüber bestimmten Perspektiven (also der Situiertheit) kann in einem solchen Kontext von Betroffenen mit verfügbaren Mitteln wirkungsvoll thematisiert und in der Folge behoben werden, ohne Mittel oder Grundlagen anzutasten.

Von epistemischer Unterdrückung zweiter Ordnung spricht Dotson, wenn die gesellschaftlich geteilten Ressourcen und Mittel der Wissensproduktion *ungenügend* sind. Sie entsteht durch die wechselseitige Abhängigkeit von diesen Mitteln und kann zum Beispiel bedeuten, dass bestimmte Erfahrungen – oder Erfahrungen bestimmter Gruppen – keinen Platz in einer gemeinsamen Sprache haben und nicht sichtbar gemacht werden können. Für einen Großteil der Bevölkerung, diejenigen nämlich, deren Wissen und Praktiken gesellschaftlich anerkannt sind, ist kein Problem wahrnehmbar. Voraussetzung für eine Veränderung hier ist, so Dotson, dass Angehörige privilegierter Gruppen sich darauf einlassen, Begriffe, Schemata oder andere geteilte Ressourcen im Licht der Beiträge Benachteiligter zu überarbeiten.

43 Vgl. ebd., S. 6f.

Epistemische Unterdrückung erster und zweiter Ordnung sind nach Dotson eng mit historisch-sozialen Machtunterschieden und Unterdrückungsmechanismen verwoben und lassen sich auf diese reduzieren.⁴⁴

Darüber hinaus bestehe jedoch eine dritte Form epistemischer Unterdrückung. Nach Dotson handelt es sich um epistemische Unterdrückung dritter Ordnung, wenn das epistemische System – gefasst als die Gesamtheit epistemischer Ressourcen, Praxen und Grundannahmen⁴⁵ – sich als *inadäquat* erweist, im Sinne von unfähig, zu erkennen, warum sich bestimmte Dinge darin nicht wissen lassen. Eine Kritik der eigenen Grundlagen ist hier nicht möglich, obwohl nur sie dazu führen könnte, epistemische Mängel zu beseitigen. Wer sich epistemischer Unterdrückung dritter Ordnung widersetzen will, bekommt es, so Dotson, nicht in erster Linie oder nicht nur mit historisch-sozial gewachsenen Machtstrukturen zu tun, sondern mit der immanenten Resilienz epistemischer Systeme. Während die beiden vorgenannten Formen epistemischer Unterdrückung sich innerhalb eines epistemischen Systems erkennen und adressieren lassen, ist eine Kritik epistemischer Unterdrückung dritter Ordnung nach Dotson nur von außerhalb möglich. Wer sie vorbringt, wird oftmals ausgelacht oder für verrückt erklärt. Das erinnert an Tobias Matzners mit Wittgenstein und Arendt beantwortete Frage nach der Veränderbarkeit von etablierten gemeinsamen Handlungsweisen und Überzeugungen. In *Vita Variabilis* fasst er diese zunächst unter Wittgensteins Begriff des »Weltbilds«, das über die Praxis erworben wird und als Hintergrund fungiert, auf welchem zwischen wahr und falsch unterschieden werde, ohne seinerseits begründbar zu sein und ohne, dass der/die Einzelne sich selbst davon überzeugt hat. Ein solches Weltbild ist auch nach Matzner nur von einem ganz bestimmten Außen, vom »Rande der Gesellschaft«⁴⁶ aus kritisierbar, sei es, weil – wie er mit Arendts Verweis auf die Figur des Pariah erläutert – dessen Fehler von dort aus besonders gut sichtbar sind, oder weil sich – wie sich an den möglichen Schicksalen dieser Figur erweist – automatisch an den Rand begibt, wer

44 Vgl. ebd., S. 12, 15.

45 Oder wie es bei Dotson heißt: »Epistemological systems, here, refer to our overall epistemological life ways. It includes operative, instituted social imaginaries, habits of cognition, attitudes towards knowers and/or any relevant sensibilities that encourage or hinder the production of knowledge. An epistemological system is a holistic concept that refers to all the conditions for the possibility of knowledge production and possession.« (Ebd., S. 7.).

46 Matzner: *Vita Variabilis*, S. xx.

es wagt, die gemeinsamen Grundlagen in Frage zu stellen. »It is the stuff ›culture clashes‹ are made of«,⁴⁷ schreibt Dotson.

Kritik von außen kann, aber muss für sie nicht mit einer historisch unterprivilegierten Position zusammenfallen: »Those gaps can affect everyone, but not necessarily equally.«⁴⁸ Bei epistemischer Unterdrückung dritter Ordnung liegt für Dotson eine Form epistemischer Unterdrückung vor, die nicht auf historisch-soziale Machtstrukturen reduzierbar ist. Diese Konzeptualisierung epistemischer Unterdrückung möchte ich im Folgenden auf die Betrachtung technischer Systeme anwenden, die mit Data Science Methoden arbeiten und in der Automatisierung von Bewerbungsverfahren zum Einsatz kommen.

Epistemische Unterdrückung durch ein ML-basiertes System?

Suchen wir nun nach den epistemischen Dimensionen in unserem Beispiel der ML-basierten Recruitingverfahren, so lassen sich drei verschiedene Ebenen betrachten:

1. Die Ebene der einzelnen Bewerber:innen, auf der eine Diskriminierung anhand bestimmter Merkmale stattfindet,
2. die allgemeinere Ebene der einzelnen Bewerber:innen, auf der ein Ausschluss als Wissende über sich selbst stattfindet, und die
3. Ebene des Kollektivs, das sich auf das erzeugte Wissen verlässt.⁴⁹

Mit dem oben eingeführten Framework lassen sich diese Ebenen getrennt unterschiedlich bewerten.

1. Wer von einem ML-basierten System auf Grund bestimmter Merkmale aussortiert wird, deren Zusammenstellung und Belegung er oder sie nicht willentlich beeinflussen kann, wird zunächst nicht von der Wissensproduktion, sondern zum Beispiel von der Teilhabe an anderen

47 Ebd., S. 17.

48 Ebd., S. 16.

49 Hier ist nicht nur an die Unternehmen zu denken, die sich erhoffen, den oder die optimale Bewerber:in zu finden, sondern genauso an die Gesellschaft, deren Mitglieder lernen, was eine Bewerbung erfolgreich macht und dabei gerne übersieht, dass die Vorbedingungen eines möglichen Erfolgs durch die statistischen Diskriminierungen eines ML-basierten Systems definiert werden können – denn die Merkmalsträger:innen, die hier aussortiert werden, bekommen nicht einmal die Chance, vielleicht in Zukunft Erfolg zu haben.

gesellschaftlichen Möglichkeiten (hier der Erwerbsarbeit in einem bestimmten Beruf) ausgeschlossen. Wenn jenseits der Diskriminierung überhaupt ein epistemischer Ausschluss in Betracht kommt, dann meines Erachtens nur indirekt und insofern, als die gesellschaftliche Position, die ihm: ihr verwehrt ist, zur künftigen Wissensproduktion beiträgt, sei es aktiv (zum Beispiel wenn es um eine akademische Position geht) oder sei es passiv (indem das folgende Schicksal der Person auf dem Erwerbsmarkt in die Suche nach neuen Ersatzinformationen und Merkmalskorrelationen einfließt oder – wie eine Automatisierung von Self-fulfilling Prophecies – alte bestätigt). Es kommt auf die im Einzelnen durch das System wirksam herangezogenen Merkmale an, ob sich diese Ebene auf bereits bestehende historisch-soziale Machtunterschiede reduzieren lässt – auch wenn viele Studien bereits gezeigt haben, dass und warum das häufig so ist (s.o.). Doch sei das mal dahingestellt: wie ließe sich diese Form von epistemischen Ausschlüssen eliminieren? Kommen Ansätze, die auf Transparenz, Erklärbarkeit und Überwachung derartiger Systeme abzielen, nicht immer schon zu spät? Wie wir gesehen haben, liegt die Besonderheit der Maschine hier vor allem in dem, was ich *Unerbittlichkeit* genannt habe. Aus diesem Grund scheint der wirksamste Hebel hier im in der Datenschutzgrundverordnung verbrieften Recht der Betroffenen zu liegen, nicht ausschließlich durch Maschinen beurteilt zu werden. Wer dieses Recht in Anspruch nimmt (vorausgesetzt, dass es für sie oder ihn überhaupt gilt), macht sich damit sicherlich nicht überall beliebt, stellt damit aber nicht mehr als die effiziente Anwendung der Mittel in Frage und kann dies im gegebenen Rahmen (eben durch die rechtliche Regelung) zumindest potenziell auch tun. Auf dieser Ebene sehe ich daher, wenn überhaupt eine epistemische Unterdrückung, dann maximal eine erster Ordnung.

2. Auf der allgemeineren Ebene der einzelnen Bewerber:innen, auf der ein Ausschluss als Wissende über sich selbst stattfindet, stellt sich die Lage schon anders dar. Auch hier ist eine Form der *Unerbittlichkeit* maschineller Systeme zu finden, die allerdings eher auf die Unmöglichkeit zielt, überhaupt noch (in diesem Bereich) als ›Wer‹ wahrgenommen zu werden und als handelnde Person erscheinen zu können. Die Betroffenen sind auf in der Vergangenheit hinterlassene und nur in Form digitaler Repräsentationen verfügbare Datenspuren reduziert und festgeschrieben und damit auch um das beraubt, was Arendt als Möglichkeit des Neuanfangs und einen wesentlichen Teil des Menschseins beschreibt. Die

Verweigerung des Dialogs, die Eliminierung des zwischenmenschlichen Erscheinungsraums, und sei es auch nur in diesem kleinen betrachteten Ausschnitt der Welt,⁵⁰ führt damit in der Tat zu einer Unterdrückungssituation: unterdrückt wird das individuelle Person-Sein als solches (in einer gegebenen Situation). Die zur Anwendung kommenden epistemischen Verfahren sind *ungenügend* in einem ganz fundamentalen Sinn. Der vorliegende Text zeigt, dass es Mittel gibt, dies zu benennen. Methoden, mit denen diese Form der epistemischen Unterdrückung vermieden werden könnten, sind mit verfügbaren Ressourcen der Wissensproduktion vorstellbar. Sind diese jedoch für alle Betroffenen (alle betroffenen Gruppen) gleichermaßen zugänglich? Sicherlich nicht. Es erweist sich im Gegenteil als ein stark mit historisch-sozial gewachsenen Strukturen verbundenes Privileg, Zugang zu den benötigten Informationen und Verhandlungsmedien zu haben. Aus den genannten Gründen könnte diese Ebene heute einer epistemischen Unterdrückung erster *oder* zweiter Ordnung entsprechen, so deutlich scheint mir die Unterscheidung an dieser Stelle nicht.

3. Interessant ist nun, wie sich der Fall auf der Ebene des Kollektivs, das sich auf das erzeugte Wissen verlässt, darstellt. Wie selbstverständlich ist es uns bereits geworden, die Entscheidungen ML-basierter Systeme hinzunehmen? Wie viele Menschen – und welche Teile der Gesellschaft – verlassen sich in ihrem täglichen Tun bereits darauf, dass sich valide Aussagen und Prognosen über das Verhalten einzelner Menschen aus der Analyse von Massendaten ableiten lassen?⁵¹ Wie sehr haben ›wir‹ uns bereits mit statistischen Diskriminierungen abgefunden, wenn wir sie nur noch verstehen (erklärt haben) wollen, ihre Anwendung aber nicht grundsätzlich in Frage stellen?

Im Rückblick zeigt sich zwar, dass die Grenzen zwischen den von Dotson vorgeschlagenen Kategorien nicht trennscharf, sondern fließend sind: die zwischen Diskriminierung und Unterdrückung, zwischen historisch-sozial und epistemisch, die zwischen geteilten Ressourcen und resilienten Grundlagen bei der Entstehung von Wissen. Dennoch können sie helfen, das

50 ..., der allerdings nur ein breiteres Phänomen illustrieren soll....

51 Dass, wie und warum derartige Verfahren trotz fragwürdiger Grundlagen vor allem in Managementkreisen nicht erst durch Data Science salonfähig wurden beschreibt Karl-Heinz Brodbeck in seiner philosophischen Kritik der Wirtschaftswissenschaften. Vgl. Karl-Heinz Brodbeck: *Die fragwürdigen Grundlagen der Ökonomie. Eine philosophische Kritik der modernen Wirtschaftswissenschaften*, Darmstadt 2009.

Problem des Einsatzes ML-basierter Systeme zur Beurteilung von Menschen in gesellschaftlich relevanten Bereichen und das Zusammenspiel von historisch-sozialer und epistemischer Unterdrückung in diesem Teilbereich besser zu verstehen.

Zusammenfassung und Ausblick

ML-basierte Systeme üben auf verschiedene Weise epistemische Macht aus, denn

1. sie haben eine Deutungsmacht über die Daten – maschinelles Lernen beruht darauf, dass der Algorithmus selber die Regeln (beim nicht-überwachten Lernen auch die Merkmale) entdeckt, mit denen sich sein Output optimieren lässt,
2. sie beschränken den Zugang zu dem von Ihnen genutzten Wissen und hindern Menschen, gegebenenfalls mit Ausnahme von Technologie-Expert:innen, daran, sich aktiv am Zustandekommen dieses Wissens zu beteiligen,
3. sie konfigurieren Wirklichkeiten neu und bringen Menschen ebenfalls dazu, sozial relevantes Wissen aufgrund reiner Verhaltensbeobachtung zu erwerben. Kommunikative Aspekte des Wissenserwerbs können so andere Formen annehmen oder teilweise entfallen. Damit entsteht eine Verlagerung von Kompetenzen, schlimmstenfalls in eine Richtung, in der der Faktor Mensch zunehmend verschwindet und die Unterscheidung zwischen Korrelationen und Handlungsgründen sowie die Bedeutung menschlicher Handlungsfreiheit aus den Augen gerät.

Bei Anwendungen, die in der Lage zu sein behaupten, aus Massendaten eine für sozial relevante Aktionen ausreichende Wissensgrundlage zu generieren, und die diese Aktionen dann auch durchführen (Aussortierung eines Anteils der eingegangenen Bewerbungen) ist maximale Vorsicht geboten. Zum einen sind die Eingangsdaten hinsichtlich ihrer Auswahl und die Ergebnisse hinsichtlich ihrer Interpretation daraufhin zu hinterfragen, ob sie persönliche und gesellschaftliche Verhältnisse überhaupt adäquat abbilden. Unabhängig davon wird die so erzeugte Sicht auf die Welt und der dahinter liegende Status quo perpetuiert, sobald ein solches System so eingesetzt wird, dass seine Ausgaben auch über künftige Zugänge zu gesellschaftlichen Räumen entscheiden. Vor allem aber werden durch ihren Einsatz soziale und politische Handlungs- und Erfahrungsräume beschnitten.

Um dies deutlich zu machen hilft es, den Einsatz ML-basierter Systeme als Akteur:innen in gesellschaftlich relevanten Bereichen unter dem Blickwinkel epistemischer Unterdrückung zu betrachten. Künftige Arbeiten könnten diesen Ansatz weiterverfolgen und nach Besonderheiten suchen, die mit der automatisierten Produktion von Wissen in weiteren Anwendungsbereichen zu tun haben. Im Fall der ML-basierten Bewerber:innen-vorauswahl liegt eine solche Besonderheit darin, dass sie die Betroffenen durch Ähnlichkeitsvergleiche mit in der Vergangenheit positiv bewerteten Fällen unerbittlich aussiebt, ohne ihnen die Möglichkeit zu geben, mit ihren individuellen Qualitäten in Erscheinung zu treten. Dabei kann auch für die einstellenden Unternehmen ein Zugriff auf potenziell neues Wissen verloren gehen. Gleichzeitig wird der gesellschaftlich relevante Vorgang, durch den der Zugang zur Erwerbsarbeit geregelt wird, dem Raum des Politischen und damit einer wenigstens potenziell existierenden Verhandbarkeit grundsätzlich und unhintergebar entzogen.

Eine Auseinandersetzung mit den Spezifika *automatisierter* Wissensproduktion kann andererseits dazu beitragen, Theorien epistemischer Ungerechtigkeit und Unterdrückung aktuell zu halten und zu bereichern. Denn es reicht nicht, sich diesem Themenfeld unter rein intersubjektiven Gesichtspunkten zu nähern. Technik – zum Beispiel in Form von ML-basierten Systemen und Data Science – hat im Bereich der Wissensproduktion teils in ermöglichender, teils in verhindernder Weise, auf jeden Fall jedoch in relevantem Maße längst Einzug gehalten.

Während die geläufigere Rede von epistemischer Ungerechtigkeit den Blick eher auf spezifische Gruppen von Ausgeschlossenen lenkt und sich auf historisch-soziale Ungleichheiten reduzieren lässt (was es nicht weniger schlimm macht), sollte nicht vergessen werden, dass der Einsatz neuer Technologien auf alle Beteiligten zurückwirkt. So lässt sich auf einer tieferen Ebene beschreiben, was gemeint sein könnte, wenn Ruha Benjamin mahnt: »Technology that might be working just fine for some of us (now) could harm or exclude others and [...] even when the stakes seem trivial, a visionary ethos requires looking down the road to where things might be headed. *We're next.*«⁵²

Kristie Dotsons Überlegungen zu epistemischer Unterdrückung haben geholfen, begreifbar zu machen, inwiefern die hier betrachteten Technologien Teile epistemischer Systeme sind und auf diese zurückwirken, vor

52 Benjamin: *Race After Technology*, S. 174.

allem aber, warum sie die epistemische Unterdrückung des Individuellen und des Sozialen verfestigen und gleichzeitig der Politisierung entziehen.

Wir alle werden erleben, wohin sich eine Gesellschaft entwickelt, die einen Teil ihrer gesellschaftlichen Gestaltungsmöglichkeiten an Technologien abtritt. Über welche Mittel wir dann verfügen, um das zu erkennen und zu beeinflussen, liegt heute in unserer Hand.

