Zeitschrift für Praktische Philosophie
Band 9, Heft 1, 2022, S. 279–304
www.praktische-philosophie.org
https://doi.org/10.22613/zfpp/9.1.11

Epistemische Ungerechtigkeiten in und durch Algorithmen – ein Überblick[1]

Epistemic Injustices in Algorithms – An Overview

Nadja El Kassar, Berlin

Zusammenfassung: Die Erkenntnis, dass Algorithmen diskriminieren, benachteiligen und ausschließen, ist mittlerweile weit verbreitet und anerkannt. Programme zur Verwendung im *predictive policing*, zur Berechnung von Rückfälligkeitswahrscheinlichkeiten bei Straftäter:innen oder zur automatischen Gesichtserkennung diskriminieren vor allem gegen nicht-Weiße Menschen. Im Zuge dieser Erkenntnis wird auch vereinzelt die Verbindung zu epistemischer Ungerechtigkeit hergestellt, wobei die meisten Beiträge die Verbindungen zwischen Algorithmen und epistemischer Ungerechtigkeit nicht im Detail analysieren. Dieser Artikel unternimmt einen Versuch, diese Lücke in der Literatur zu verkleinern. Dabei umreiße ich zunächst das Feld der Algorithmen, um so die Entitäten, die epistemisch ungerecht sein könnten, klar zu fassen und zu unterscheiden. Dann erläutere ich eine Auswahl von epistemischen Ungerechtigkeiten, die für die Analyse von ungerechten Algorithmen relevant sind. Schließlich führe ich epistemische Ungerechtigkeiten und Algorithmen zusammen und zeige anhand von drei Beispielen – automatische Geschlechtserkennung, Googles Suchmaschinen-Algorithmus, PredPol (*predictive policing)* – dass Algorithmen an testimonialer Ungerechtigkeit und hermeneutischer Ungerechtigkeit beteiligt sind. Sie tragen so zudem auf verschiedenen Ebenen zur Exklusion von marginalisierten Gruppen bei.

Schlagwörter: Epistemische Ungerechtigkeit, Exklusion, Algorithmen, Predictive Policing, Automatische Geschlechtserkennung

1 Ich danke Natalie Ashton, Josh Habgood-Coote, Hilkje Hänel und zwei anonymen Gutachter:innen dieser Zeitschrift für wertvolle Anmerkungen zu den Überlegungen in diesem Aufsatz.

Abstract: It is widely acknowledged that algorithms discriminate against people, disadvantage them, and exclude them. Predictive policing, recidivism predictions, automatic facial recognition discriminate against non-Whites in particular. These insights have been mentioned in conjunction with epistemic injustice, but most contributions do not discuss the relation between algorithms and epistemic injustices in detail. This article attempts to reduce the gap in the literature. I start by outlining what algorithms are in order to capture those objects that can be epistemically unjust. Then I introduce varieties of epistemic injustice that are relevant for analyzing epistemically unjust algorithms. On that basis I examine three algorithms – automatic gender recognition, Google's search engine algorithm, PredPol (predictive policing) – with respect to epistemic injustice. I argue that these algorithms contribute to varieties of testimonial and hermeneutical injustice and are part of different levels of exclusion of marginalized groups.

Keywords: Epistemic injustice, exclusion, algorithms, predictive policing, automatic gender recognition

Die Erkenntnis, dass Algorithmen oft diskriminieren, benachteiligen, ausschließen, ist mittlerweile weit verbreitet und auch weit anerkannt (z. B. Eubanks 2017, O'Neil 2016, Orwart 2019, Zweig 2017). Insbesondere Programme zur Verwendung im *predictive policing*[2], zur Berechnung von Rückfälligkeitswahrscheinlichkeiten bei Straftäter:innen oder zur automatischen Gesichtserkennung diskriminieren vor allem gegen nicht-Weiße Menschen. Im Zuge dieser Einsicht wird auch vereinzelt die Verbindung zu epistemischer Ungerechtigkeit hergestellt, z. B. dass digitale Umgebungen und digitale Identitäten epistemisch ungerecht sind (Origgi und Tirana 2017, Scotto 2020), aber die meisten Beiträge analysieren die Verbindungen zwischen Algorithmen und epistemischer Ungerechtigkeit nicht im Detail. Dieser Artikel unternimmt einen Versuch, diese Lücke in der Literatur zu verkleinern. Ich werde verschiedene Hinsichten, in denen Algorithmen epistemisch ungerecht sind und zu epistemischer Ungerechtigkeit beitragen, unterscheiden. Dadurch entsteht ein einleitendes, sortiertes Bild der Rolle von Algorithmen in der Debatte um epistemische Ungerechtigkeit, das einen weiteren Schritt zur Übertragung der Konzeption von epistemischer Ungerechtigkeit in die digitale Welt darstellt. Es wird sich dabei unter anderem zeigen, dass Algorithmen Manifestationen von epistemischer Ungerechtigkeit sein können.

2 Ich lasse diesen Ausdruck unübersetzt, weil er auch im deutschen Diskurs oft unübersetzt verwendet wird.

Eine wichtige Einschränkung möchte ich bereits am Anfang hervorheben: Die Analyse soll nicht Verantwortungsfragen diskutieren, vielmehr soll Material entwickelt werden, dass für eine dezidiert moralische Diskussion und Verantwortungsfragen relevant ist.

In meiner Analyse gehe ich wie folgt vor. Ich umreiße zunächst das Feld der Algorithmen, um so die Entitäten, die epistemisch ungerecht sein können, klar zu fassen und zu unterscheiden (Abschnitt 1). Dann erläutere ich eine Auswahl von epistemischen Ungerechtigkeiten, die für die Analyse von ungerechten Algorithmen relevant sind (Abschnitt 2). Schließlich führe ich epistemische Ungerechtigkeiten und Algorithmen zusammen und zeige anhand von drei Beispielen – automatische Geschlechtserkennung, PredPol (*predictive policing*), Googles Suchmaschinen-Algorithmus – dass Algorithmen an testimonialer Ungerechtigkeit und hermeneutischer Ungerechtigkeit beteiligt sind. Sie tragen so zudem auf verschiedenen Ebenen zur Exklusion von marginalisierten Gruppen bei (Abschnitt 3). In einem Ausblick sammele ich die Erkenntnisse und verweise kurz auf die Möglichkeit, dass Algorithmen auch im Widerstand gegen epistemische Ungerechtigkeit verwendet werden können (Abschnitt 4).

1. Um welche Algorithmen geht es?

Im Grunde genommen – und aus nicht-technischer Perspektive definiert – sind Algorithmen (mathematische) Handlungsvorschriften, die vorgeben, wie ein bestimmtes Ziel unter bestimmten Bedingungen zu erreichen ist.[3] Algorithmen ähneln insofern Kochrezepten, die die Abfolge von Handlungen vorschreiben, die eine Person vornehmen muss, um ein bestimmtes Gericht zuzubereiten. Anders als Kochrezepte sind Algorithmen aber nicht nur auf ein Gericht beschränkt, sondern sind allgemeiner auf Typen von Problemen bezogen (Lenzen 2019, 42). Lenzen spricht daher von Algorithmen als einer „[A]utomatisierung … eine[s] Problemlösungsprozesses“ (Lenzen 2019, 43). Die Kernaufgaben von Algorithmen sind dabei (1) das Priorisieren oder Sortieren – eine geordnete Liste erstellen, (2) das Klassifizieren – etwas einer Kategorie zuordnen, (3) das Assoziieren – Verbindung herstellen, und (4) das Filtern – Wichtiges hervorheben (Fry 2018, 8–9).

Zusätzlich zu den Kernfunktionen von Algorithmen und den verschiedenen Manifestationen dieser Aufgaben in diversen Programmen müssen

3 Siehe Hill (2016) für eine nicht formale Definition von Algorithmen, die dennoch technisch informiert ist.

Algorithmen ganz grundsätzlich in einer weiteren Hinsicht unterschieden werden: Algorithmen können in *prima facie* unproblematischen Kontexten oder in problematischen Kontexten, d. h. *high-stakes* bzw. Hochrisiko-Kontexten eingesetzt werden. *Prima facie* unproblematische Kontexte sind beispielsweise die Verwendung von Algorithmen bei der Erkennung von Müll in Gewässern (Wolf et al. 2020). Problematische Kontexte sind diejenigen Kontexte, in denen es „um etwas geht", in denen ein hohes Risiko gegeben ist. Das meint vor allem Kontexte, in denen Menschenleben von den Entscheidungen, die auf der Grundlage der Algorithmen oder von Algorithmen getroffen wurden, direkt betroffen sind. Solche Kontexte sind beispielsweise bei der Verwendung von Algorithmen in der Rechtsprechung, der Kreditvergabe oder Sozialhilfevergabe zu finden. Aufgrund dieser zumindest potenziell problematischen Einsatzgebiete sind Algorithmen und auch ihre Produktion normativ und (zumindest teilweise) politisch.[4]

Die Analyse von epistemischen Ungerechtigkeiten ist besonders bei diesen *high-stakes* oder auch Entscheidungen-treffenden (*decision making*) Algorithmen essenziell. Ich möchte mich jedoch nicht nur auf Entscheidungen treffende Algorithmen beschränken, vielmehr soll es um Algorithmen gehen, „bei denen es um etwas geht", Algorithmen, die Auswirkungen auf Menschen(leben) haben. Mit dieser Umgrenzung gehört auch Googles Suchalgorithmus zum Untersuchungsgegenstand, denn er trifft nicht Entscheidungen, aber bei den Suchergebnissen kann es um etwas gehen. Mehr zu den ausgewählten Algorithmen am Ende dieses Abschnitts.

Für alle Algorithmen gilt zudem, dass sie epistemisch sind. Denn Algorithmen tragen sowohl in unproblematischen als auch in high-stakes Kontexten zur Wissensproduktion bei. Sie produzieren Wissensbeiträge, in unproblematischen Kontexten etwa Wissen über Plastikabfälle in Flüssen, aber in Hochrisiko-Kontexten auch Ergebnisse über die Kreditwürdigkeit von Personen, die dann als Wissen behandelt werden. Da Algorithmen immer von Entwickler:innen, Programmierer:innen, der kooperierenden Techniker:innen sowie den Auftraggeber:innen produziert sind, sind diese Personen zumindest indirekt beteiligt an der Wissensproduktion durch die Algorithmen. Insofern die Algorithmen normativ und politisch relevant sind, sind auch die Entwickler:innen etc. mit impliziert. Inwiefern sie verantwortlich sind für die Algorithmen ist eine Frage, die ich nicht behandeln werde, da sie zu weit

4 Sie fügen sich also in Langdon Winners größere These ein: „Artifacts have politics" (Winner 1980).

führen würde. Festzuhalten ist, dass ihre Verantwortung für die Algorithmen über die Verantwortung von anderen epistemischen Subjekten hinaus geht, da sie in ihrer Produktion und an ihrem Einsatz direkt beteiligt sind.

Ihre Involviertheit lässt sich wie folgt weiter spezifizieren: Für alle Algorithmen ist es die Aufgabe von Programmierer:innen und kooperierenden Techniker:innen ist, das gesuchte Ziel und die damit verbundenen Teilziele in eine mathematische Sprache zu übersetzen und somit erfassbar für die mathematischen Handlungsempfehlungen zu machen (Zweig 2019, 60). Dazu gehört auch die *Modellierung* der spezifischen Aufgabe. Schauen wir uns den Fall einer Dating-App an. Allgemein gesprochen soll diese App Menschen in Kontakt bringen. Dazu muss sie auswählen, welche Menschen sie in Kontakt bringt. Welche Parameter und Kriterien dazu verwendet werden, hängt von den Auftraggeber:innen, Entwickler:innen, Techniker:innen ab. Müssen die Personen Fragen gleich beantworten? Müssen sie sich auf Fotos gegenseitig auswählen, um in Kontakt gebracht zu werden? In allgemeinen Algorithmus-Typen gesprochen, sollen die Algorithmen hinter der Dating-App Menschen gemäß den gewählten Kriterien assoziieren. Für die Ausführung der Anweisungen ist dann noch Input in Form von Daten notwendig, mit denen der geschriebene Algorithmus ablaufen kann. Allerdings können die in der App verwendeten Algorithmen auch aus Datenanalysen entwickelt werden. Dann durchsucht ein erster Algorithmus Daten nach Mustern, und auf der Grundlage der entdeckten Muster wird dann ein zweiter Algorithmus entwickelt, der die Muster auf weitere Daten anwendet und quasi ‚nur noch' entscheidet (dieser Prozess ist das sogenannte „maschinelle Lernen").[5]

Diese basalen Unterscheidungen sind auch für die Hauptfrage des Artikels nach epistemischer Ungerechtigkeit in und durch Algorithmen wichtig, denn sie zeigen auf, dass Ungerechtigkeiten (hier dezidiert nicht nur epistemisch) in Algorithmen an mehreren Stellen eindringen und vorkommen können. Zudem können verschiedene Instanzen und Prozesse zu Ungerechtig-

5 Wie der erste Algorithmus die Muster findet, kann wiederum ganz unterschiedlich sein. Es können beispielsweise Entscheidungsbäume und andere klassische mathematische Verfahren verwendet werden (Lenzen 2019, 131–133), Aber auch das *deep learning* (eine Art des maschinellen Lernens) kann verwendet werden, bei der künstliche neuronale Netze, die mehrere Schichten haben, Gewichtungen berechnen und damit ein statistisches Modell des operationalisierten Gegenstands erstellen (vgl. Mitchell 2019, 70–80 für eine genauere Beschreibung).

keiten führen. Und diese Ungerechtigkeiten können dann auch epistemischer Art sein, wie ich im Verlauf an drei Beispielen zeigen werde. Beispielsweise könnte die Dating-App nur mit Daten von Weißen, heterosexuellen Männern mit einem bildungsbürgerlichen Hintergrund trainiert worden sein, und dadurch Dating-Entscheidungen etwa von nicht-Weißen Personen nicht abbilden können – ein Defizit, das auch epistemisch relevant ist.

Einen realen problematischen Fall von einem Algorithmus, der aus Daten von tatsächlichen Entscheidungen von Menschen lernt, und auf der Grundlage einen Algorithmus ermöglicht, der ungerechte Entscheidungen trifft, liefert Amazons mittlerweile aufgegebener Bewerbungsalgorithmus. Dieser Algorithmus sollte auf der Grundlage von (a) alten Bewerbungen an Amazon, (b) den tatsächlich erfolgten Anstellungen und (c) weiteren mit den Anstellungen verbundenen Daten lernen, welche Bewerbungen gleich aussortiert werden und welche Bewerbungen gute Chancen für eine Anstellung bei Amazon haben. Aus dieser Datenmenge schloss der entstandene Algorithmus unter anderem, dass Bewerbungen mit dem Wort „women's", z. B. „women's chess club captain" schlechter bewertet werden *sollten* (Dastin 2018). Denn die Bewerbungen, die dieses Wort enthielten, waren in der Datenmenge, mit der der Algorithmus trainiert wurde, eher erfolglose Bewerbungen gewesen. Der Algorithmus findet ja schlicht geteilte Eigenschaften, die bisher erfolglosen und erfolgreichen Bewerbungen zukommen, und bestimmt daraus Eigenschaften, die anwendbar sind, z. B. dass das Wort „women's" in einer erfolglosen Bewerbung vorkommt. Wenn man auf dieser Grundlage, ohne externe Kontrolle, einen Entscheidungsalgorithmus schriebe, dann würde dieser zweite Algorithmus einfach Bewerbungen mit dem Wort „women's" schlechter bewerten, weil das so vorgeben ist.[6] Man würde dabei allerdings übersehen, dass in den Bewerbungen, aus denen der Algorithmus lernen sollte, Frauen auch ungerechtfertigterweise abgelehnt wurden, und der Indikator „women's" als ein Ausschlusskriterium für Bewerbungen, somit nur alte ausschließenden Muster wiederholt und nun in den Entscheidungen materialisiert. Die Leistung des Algorithmus kann nicht sinnvoll ohne soziale, strukturelle und politische Überlegungen bewertet werden. Das Programm wurde von Amazon mittlerweile eingestellt, weil es die Bewerbungen nicht gerecht bewerten würde.

6 Natürlich kommen noch andere Kennzeichen dazu, aber ich habe mich hier aus Gründen der Einfachheit auf ein Kriterium beschränkt.

Für viele Algorithmen besteht das zusätzliche Problem, dass sie nicht transparent sind. Je nach Akteurin stellt sich die Intransparenz unterschiedlich dar. Die Personen, die durch die Algorithmen bewertet werden, kennen die Handlungsschritte und die zugrundeliegenden Daten nicht, z. B. bei Entscheidungen bezüglich der Kreditwürdigkeit einer Person. Aber auch die Personen, die die Algorithmen verwenden, könnten die Handlungsschritte und Daten nicht kennen – so etwa wenn man mit Googles Suchmaschine nach einem Stichwort sucht und Ergebnisse präsentiert bekommt. Diese Intransparenz kann zwei grundlegend verschiedene Ursachen haben. Einerseits sind manche Algorithmen nicht transparent, weil sie das Ergebnis von *deep learning*[7] und damit nicht mehr vollständig verständlich oder zugänglich für Menschen sind. Erklärungen, wie die Algorithmen arbeiten, sind dann meist nur *posthoc* und erklären das Ergebnis, nicht aber die eigentlichen Prozesse, die zu dem Ergebnis führen (vgl. Rudin 2019).[8] Andererseits liegt in einer Vielzahl von Fällen die Intransparenz darin begründet, dass die spezifischen Algorithmen Eigentum einer Firma sind und dadurch geschützt sind. Der Algorithmus, der in Googles Suchmaschine verwendet wird, und tiktoks Empfehlungsalgorithmus sind in erster Linie undurchsichtig, weil die Firmen das Betriebsgeheimnis nicht verraten wollen.[9] Diese Unterscheidungen sind für die unternommene Analyse von epistemischen Ungerechtigkeiten wichtig, weil die Intransparenz die epistemische Ungerechtigkeit in einigen Fällen ko-konstituiert. Wir werden die Rolle der Intransparenz in der Analyse sogenannter Automatischer Geschlechtserkennungsprogramme und Programmen zur Vorhersage von Straftaten sehen.

Ich werde in diesem Artikel drei Algorithmen im Hinblick auf enthaltene epistemische Ungerechtigkeiten diskutieren. Das erste Beispiel sind Algorithmen für automatische Geschlechtserkennung, die bisher vor allem für personalisierte Werbung verwendet werden (Gebru 2020). Das zweite Beispiel ist der Algorithmus, der Googles Suchmaschine ausmacht. Und das dritte Beispiel ist „PredPol“, ein kommerzielles Modell, das verwendet wird,

7 Siehe Fußnote 5.

8 Es gibt Möglichkeiten auch *deep learning* Algorithmen einsehbar zu machen, aber sie werden nicht weit angewendet (vgl. Rudin 2019, Chen et al. 2019).

9 Diese komplexen Algorithmen können natürlich auch aufgrund ihrer Funktionsweise, etwa durch die Verwendung von *deep learning*, im Detail für Menschen unzugänglich, d.i. unverständlich, sein, aber das muss nicht der Fall sein

um *Predictive Policing* zu betreiben, d. h. vorherzusagen, wo Verbrechen geschehen werden.

2. Arten von epistemischer Ungerechtigkeit

Grundsätzlich finden sich epistemische Ungerechtigkeiten in strukturell ungerechten Gesellschaften, in denen dominante Gruppen die epistemische Akteurschaft von Mitgliedern nicht-dominanter, marginalisierter Gruppen beeinträchtigen. Die Ungerechtigkeit ist epistemisch, weil sie das Subjekt als epistemisches Subjekt, also als Wissende:n, betrifft (Fricker 2007, 20). Weil aber die Fähigkeit zu Wissen ein elementarer Bestandteil unseres Personenstatus ist, betrifft epistemische Ungerechtigkeit auch die Person als Ganze. Wie Fricker ausführt, „wronging someone as a giver of knowledge – by perpetrating testimonial injustice – amounts to wronging that person as a knower, as a reasoner, and thus as a human being" (2007, 44). Die paradigmatischen Formen epistemischer Ungerechtigkeit sind die von Fricker eingeführte testimoniale Ungerechtigkeit und hermeneutische Ungerechtigkeit (Fricker 2007). Testimoniale Ungerechtigkeit findet sich, wenn das Zeugnis von Sprecher:innen deswegen nicht angenommen oder geglaubt wird, weil die Hörer:innen ein identitätsbezogenes Vorurteil gegenüber den Sprecher:innen haben (Fricker 2007, 28). Die Sprecher:innen sind glaubwürdige Zeugnisgebende, aber sie werden aufgrund der identitätsbezogenen Vorurteile nicht als solche behandelt. Testimoniale Ungerechtigkeit tritt beispielsweise dann auf, wenn in einem akademischen Gespräch eine Aussage nicht als Wissensbekundung ernst genommen wird, weil die Sprecherin eine Frau ist und die Zuhörer:innen aufgrund von Vorurteilen nicht anerkennen, dass die Frau eine wissende Person ist, die Wissen weitergeben kann.

Hermeneutische Ungerechtigkeit basiert ebenfalls auf identitätsbezogenen Vorurteilen. In diesem Fall beschränken diese Vorurteile allerdings die begriffliche Ausstattung einer Gemeinschaft; vereinfacht gesagt, die Gemeinschaft besitzt nicht die passenden Ausdrücke, die für das Verständnis und Selbstverständnis von marginalisierten Gruppen notwendig wäre. Frickers einschlägiges Beispiel beschäftigt sich mit dem Ausdruck der sexuellen Belästigung, der erst in den 1960er Jahren geschaffen werden musste, um die Erfahrungen von Frauen zu erfassen und sprachlich auszudrücken. Weil die Mitglieder der dominanten Gruppe diese belästigenden Praktiken als normal oder sogar legitim angesehen haben, haben sie für diese Erlebnisse keinen Ausdruck in ihrem Begriffsrepertoire entwickelt. Die benachteiligten Gruppen, die aufgrund von hermeneutischer Marginalisierung nicht im sel-

ben Maße wie die dominante Gruppe an der Bildung von Begriffen beteiligt sind, machen Erfahrungen, die von den Begriffen der dominanten Gruppe nicht erfasst werden.

Gaile Pohlhaus hat diese hermeneutische Ungerechtigkeit ergänzt durch eine bestimmte Art von Unwissenheit, die durch hermeneutische Ungerechtigkeit entsteht, nämlich die sogenannte *vorsätzliche hermeneutische Ignoranz* [„willful hermeneutical ignorance", (Pohlhaus 2012, 716)]. Bei dieser Ignoranz weigern sich Mitglieder dominanter Gruppen Begriffe anzuerkennen, die marginalisierte Gruppen in den Diskurs einbringen, um ihre Erfahrungen zu beschreiben (Pohlhaus 2012, 715f.). Dadurch ignorieren dominante Subjekte Teile der Welt, die für marginalisierte Gruppen signifikant sind, oder missverstehen diese Teile, denn ihnen fehlen die Begriffe, um diese Bereiche angemessen zu erfassen (ebd.). Dieses Missverstehen und Ignorieren von marginalisierten Lebensrealitäten in der geteilten Welt wirkt sich auf die epistemische Funktionalität der gesamten Gemeinschaft aus, denn so werden nicht-dominante Beiträge aus dem allgemeinen Diskurs ausgeschlossen. Die Weitergabe von Wissen und Überzeugungen aus marginalisierten Gruppen ist gestört, weil die dominante Gruppe die Mittel ablehnt, die notwendig sind, um dieses Wissen zu verstehen und aufnehmen zu können. Wie Rebecca Mason hervorhebt, können zweierlei Arten von Unwissen aus diesen Dysfunktionen entstehen: (a) Die Mitglieder von marginalisierten Gruppen wissen nicht, dass sie unterdrückt und marginalisiert werden, weil die begrifflichen Ressourcen fehlen, um diese Unterdrückung zu erfassen; und (b) die dominante Gruppe ist unwissend – vielleicht sogar ignorant – darüber, dass sie marginalisierte Gruppen ausschließen (Mason 2011).[10]

Die Effekte von testimonialer und hermeneutischer Ungerechtigkeit zeigen sich auch auf der psychologischen Ebene. *Testimonial smothering*, ein Begriff, den Kristie Dotson eingeführt hat, beschreibt das Schweigen einer marginalisierten Person, die sich bewusst entscheidet, nicht zu sprechen, weil sie auf Grundlage vorheriger Erfahrungen weiß, dass die Hörer:innen ihr nicht zuhören werden (Dotson 2011, 244).[11] *Testimonial quieting*, ein

10 Unwissenheit ist im Gegensatz zur Ignoranz ein neutraler kognitiver Zustand. Eine unwissende Person weiß eine wahre Aussage nicht oder hat eine falsche Überzeugung. Ignoranz ist eine Form von Unwissenheit, bei der die betreffende Person auch nicht wissen will. Die Unterscheidung lässt sich auch auf Gruppen anwenden. Siehe auch El Kassar (i.E.).

11 Ich lasse *testimonial smothering* und auch *testimonial quieting* unübersetzt, da die Übersetzungen eher Beschreibungen liefern würden.

weiterer Begriff von Dotson, ist eine Variante von testimonialer Ungerechtigkeit, bei der das Zeugnis der Sprecher:innen aufgrund von identitätsbezogenen Vorurteilen überhaupt nicht angehört wird (Dotson 2011, 242). Während es bei testimonialer Ungerechtigkeit um ein Glaubwürdigkeitsdefizit im Hinblick auf eine bestimmte Aussage von einer bestimmten Person geht, geht es bei *testimonial quieting* darum, dass die Aussage einer Person nicht gehört wird; es ist, als ob die Person gar nicht gesprochen hätte. *Testimonial smothering* und *quieting* greifen ebenso wie andere Formen epistemischer Ungerechtigkeit bzw. Gewalt in die epistemische Handlungsfähigkeit eines Subjekts ein. Sie liefern hierbei im Vergleich zu hermeneutischer Ungerechtigkeit aber vor allem Erklärungen, wie die betroffenen Subjekte aus epistemischen Gemeinschaften ausgeschlossen werden und somit nicht zur Wissensproduktion beitragen können.

Dotson unterscheidet zudem drei Ebenen von epistemischer Exklusion, die die folgende Diskussion von epistemischer Ungerechtigkeit in und durch Algorithmen erweitert, indem sie einerseits erklärt, warum die Ungerechtigkeiten schwer zu erkennen sind und andererseits Unterscheidungen einführt, die für die Reaktion auf die epistemische Ungerechtigkeit sinnvoll sind. In diesem Artikel geht es noch nicht um Lösungsschritte, aber die Analyse der Problemlage sollte nichtsdestotrotz einen Anschluss hier vorbereiten. Die Exklusion erster Ordnung „results from the incompetent functioning of some aspect of shared resources with respect to some goal or value" (Dotson 2014, 123). Dazu gehört beispielweise testimoniale Ungerechtigkeit, bei der nicht das ganze System des Zeugnisgebens und Zeugnisannehmens defekt ist, sondern nur ein Aspekt in der Verwendung des Systems; nämlich die Annahme von Zeugnis ohne ungerechtfertigte identitätsbezogene Vorurteile. Bei Exklusion zweiter Ordnung gibt es zusätzlich auch einen Defekt im System, denn die epistemischen Ressourcen funktionieren nicht für alle Teilnehmer:innen gleich gut (Dotson 2014, 126–129). Ein Beispiel für so eine Exklusion zweiter Ordnung ist die hermeneutische Ungerechtigkeit, bei der für bestimmte Mitglieder die Ressourcen fehlen, um ihre Lebensrealität auszudrücken. Auch das *testimonial smothering* fällt unter diese Exklusion zweiter Ordnung. Bei Exklusion dritter Ordnung funktioniert das epistemische System wie es soll – allerdings nur aus der Sicht der dominanten Gruppe. Es ist als System an sich nicht in der Lage eine bestimmte epistemische Aufgabe zu erfüllen, weil es unangemessene, dominante epistemische Ressourcen verwendet und so andere Mitglieder nicht-dominanter Gruppen systematisch ausschließt (Dotson 2014, 129–131). Hierzu gehört *contributo-*

ry injustice, bei dem die Hörenden sich weigern, die passenden begrifflichen Mittel zu verwenden, um die Beiträge der marginalisierten Gruppe als epistemisch wertvolle Beiträge aufnehmen und somit überhaupt verstehen zu können (Dotson 2012, 31–32).

Es gibt noch weitere Arten von epistemischer Ungerechtigkeit (vgl. Pohlhaus 2017, Bratu und Hänel 2021) und konzeptionelle Details, aber diese knappe Übersicht reicht aus, um die epistemische Ungerechtigkeit von Algorithmen in einem ersten Versuch zu erfassen. Im Kontext dieses Artikels wird also auf zwei Typen von epistemischer Ungerechtigkeit fokussiert, die sich in Bezug auf Algorithmen finden lassen könnten:

a) Testimoniale Ungerechtigkeit, *testimonial quieting*, *testimonial smothering*, die alle das Subjekt als Mitglied einer sozialen Gruppen als Wissende und als Zeugnis gebende Person betreffen.
b) Hermeneutische Ungerechtigkeit, welche zu vorsätzlicher hermeneutischer Ignoranz und zu Unwissenheit sowohl bei Mitgliedern von marginalisierten als auch dominanten Gruppen führen kann. Wie oben erklärt, sind hierbei vor allem die begrifflichen epistemischen Ressourcen der Subjekte und Gemeinschaften betroffen. Die daraus folgenden epistemischen Dysfunktionen finden sich also in der gesamten Gesellschaft, aber haben nicht die gleichen Konsequenzen für alle Personen.

3. Epistemische Ungerechtigkeiten in und durch Algorithmen

Wenn wir uns nun epistemische Ungerechtigkeit im Hinblick auf spezifische Algorithmen anschauen, werde ich mich, wie angekündigt, auf Algorithmen beschränken, die *high stakes* sind bzw. bei denen es um etwas geht. Die drei diskutierten Algorithmen sind: (1) Geschlechtserkennungssoftware, (2) Googles Suchmaschinen-Algorithmus, (3) PredPol.

3.1. Automatische Geschlechtserkennungsprogramme (Automatic Gender Recognition, AGR)

Automatische Geschlechtserkennungsprogramme (*Automatic Gender Recognition*, AGR) werden derzeit u. a. für personalisierte Werbung eingesetzt, aber ihr möglicher Verwendungsrahmen ist weitaus größer und alle großen Tech Unternehmen arbeiten an solchen Erkennungsprogrammen.[12]

12 Zu den Anwendungen von Automatic Gender Recognition siehe den Überblick von Vincent (2021).

AGR baut auf Algorithmen auf und fällt damit auch in die für diesen Artikel relevante Gegenstandsgruppe.[13] Die meisten technischen Details der AGR Programme sind für unsere Zwecke nicht wichtig, wichtig ist aber, dass die bestehenden Programme ausschließlich mit einer binären Geschlechtsverteilung arbeiten, d. h. sie können nur „Mann" oder „Frau" erkennen und haben nicht-binäre Identitäten gar nicht als Optionen in ihrem Setup. Des Weiteren können sie nur trans* Personen zuordnen, die Geschlecht in einer cisnormativen[14] Weise manifestieren, also Standardkonzeptionen von männlichem und weiblichem Aussehen entsprechen (Scheuerman et al. 2019, 144:15). Die Erkennungsprogramme beschränken sich dabei etwa auf Gesichtszüge und Lippen- und Augenmakeup (Muthukamar et al. 2019). Buomlawini und Gebru (2018) zeigen zudem für drei kommerzielle Geschlechtserkennungsprogramme, dass sie bei der Erkennung von Schwarzen Frauen die höchste Fehlerquote haben.

Durch diese Design-Entscheidungen, die von binären Geschlechterverteilungen ausgehen, wird die Existenz von nicht binären trans* Personen ausgeschlossen und verunmöglicht (vgl. auch Scheuerman et al. 2019, 144:22). Sie können nur als männlich oder weiblich erkannt werden, nicht als nicht-binär oder trans*. Auch bei einfachen Sicherheitsscannern an Flughäfen werden nicht-binäre und trans* Personen als auffällig gekennzeichnet und behandelt, denn der Scanner markiert sie und ihre Körper als Sicherheitsrisiko, da der gescannte Körper mit dem Geschlecht, das die Sicherheitsmitarbeitenden manuell eingeben müssen, nicht übereinstimmt. Costanza-Chock beschreibt eindrücklich, wie es für eine nicht-binäre trans* weibliche Person ist diesen Prozess zu durchlaufen.

> … [M]y heartbeat speeds up slightly as I near the end of the line, because I know that I'm almost certainly about to experience an embarrassing, uncomfortable, and perhaps humiliating search by a Transportation Security Administration (TSA) officer, after my body is flagged as anomalous by the millimeter wave scanner. I know that this is almost certainly about to happen because of the particular sociotechnical configuration of gender normativity (cis-normativity, or the assumption that all people have a gender identity that is consistent with the sex

13 Für einen Überblick über Unterschiede in Funktion und Leistung von Gesichtserkennung und Labelling siehe die Analyse von Scheuerman et al. (2019).

14 Cis-Normativität ist die Annahme, dass alle Personen die Gender Identität haben, die ihnen bei ihrer Geburt zugewiesen wurde.

> they were assigned at birth) that has been built into the scanner, through the combination of user interface (UI) design, scanning technology, binary-gendered body-shape data constructs, and risk detection algorithms, as well as the socialization, training, and experience of the TSA agents. (Costanza-Chock 2020)

Diese Prozesse wiederholen sich auch bei Algorithmen, die nur mit binären Geschlechterkategorien arbeiten.[15] Und die entstehenden Erfahrungen von nicht-binären Personen sind nicht einfach nur unangenehm. Vielmehr erfahren sie durch die Konstruktion der Technologien, Scanner und AGR, epistemische Ungerechtigkeit. Darin dass ihre Existenzweise nicht erkannt und anerkannt werden *kann*, weil die Technologien – u. a. die Algorithmen – keine passenden Begriffe für ihre Existenz erhalten haben und somit auch nicht anwenden können, liegt die Grundlage für testimoniale und hermeneutische Ungerechtigkeit. Ich konzentriere mich weiter auf AGR, aber die Ungerechtigkeiten finden sich auch in anderen Programmen, die gleiche Designmerkmale in Bezug auf Geschlecht haben.

Testimoniale Ungerechtigkeit manifestiert sich bei der AGR auf zwei Ebenen. Auf der einen Ebene als testimonial quieting, und zwar wenn Wissen von nicht-binären und trans* Personen beim Design der AGR nicht berücksichtigt wird. Eine weitere Ebene betrifft die Erkennung des Geschlechts durch die AGR. Dabei legt die Person, die erkannt werden soll, dem Programm gegenüber ein Zeugnis in Bezug auf ihr Geschlecht ab. – Dieses Zeugnis ist auch körperlicher Art, siehe dazu weiterführend Medina und Henning (2021). – Das AGR Programm kann dann das Zeugnis richtig oder falsch verstehen, aufnehmen oder ablehnen. Für nicht-binäre Personen können AGR Programme, die auf binären Geschlechterkonzeptionen aufbauen, niemals das Zeugnis annehmen oder richtig verstehen, weil sie gar nicht die Begriffe dazu einprogrammiert haben. Für trans* Personen schränkt sie den Interpretationsrahmen ein und zwingt ihnen eine binäre Interpretation auf. Der Versuch das eigene Geschlecht auszudrücken, wird systematisch verzerrt.[16]

15 Mar Hicks (2019) zeigt wie die zunehmende Digitalisierung ab den 1950ern es trans* Personen in England erschwerte, ihr Geschlecht offiziell zu ändern. Auch da wurden dann binäre Kategorien verwendet, die keinen Raum für Alternativen ließen. Das Problem lässt sich also historisch untersuchen.

16 Die Situation kann auch mit Hilfe von Catalas Begriff der *hermeneutical domination* beschrieben werden, weil das Zeugnis der nicht-binären und trans* Personen angehört, aber nicht umgesetzt wird (2015).

Testimoniale Ungerechtigkeit und *quieting* werden in der Tätigkeit des Algorithmus also materialisiert und verstärkt. Infolge der Fehlklassifikationen können falsch klassifizierte nicht-binäre und trans* Personen auch vermehrt Selbstzweifel erleben (Scheuerman et al. 2019, 144:22), ebenfalls ein anerkannter Effekt von testimonialer Ungerechtigkeit.[17] Die Blackbox AGR lässt sie nicht erkennen, warum sie falsch klassifiziert wurden, und vor allem zeigt es nicht auf, dass die Fehlklassifikation nicht an ihnen oder ihrer Gender Präsentation liegt, sondern schlicht am Design des AGR Programms (Scheuerman et al. 2019, 144:22). Dazu bräuchten sie Wissen von Aktivist:innen, etwa dem *Design Justice Network* (vgl. Costanza-Chock 2020) oder Analysen wie die von Scheuerman et al., die die Funktionsweise des AGR Programms erläutern.

Bei der Entwicklung der Programme zeigt sich zudem auch hermeneutische Ungerechtigkeit in Form vorsätzlicher hermeneutischer Ignoranz (Pohlhaus 2012). Sie zeigt sich bei den Designer:innen, die die AGR Programme so strukturiert haben. Wer in den späten 2010ern AGR Programme schreibt, die auf einer binären Geschlechtsverteilung basieren, ignoriert Begriffe und Konzeptionen, die trans* und nicht-binäre Personen selbst verwenden und die in der Öffentlichkeit durchaus zugänglich sind.

AGR Programme schließen somit nicht-binäre und trans* Personen aus. Einerseits handelt es sich um eine Exklusion zweiter Ordnung, denn das System AGR funktioniert nicht für alle Teilnehmer:innen gleich gut. Nicht-binäre und trans* Personen werden in ein binäres Geschlechterschema gepresst, das sie nicht erfasst. Und auch eine Exklusion dritter Ordnung liegt vor, da die AGR Programme nur aus der cisnormativen Perspektive fehlerfrei funktioniert und beispielsweise trans* Personen weniger gut klassifiziert als nicht-trans Personen. Man muss allerdings aus dem System heraustreten, um die Struktur kritisch analysieren zu können (Dotson 2014).

Dezidiert epistemisch ist die Ungerechtigkeit, weil nicht-binäre und trans* Personen zur sozialen Konstruktion von Geschlecht sowie Konzeptionen von Geschlecht beitragen aber durch die AGR Programme in ihrer Eigenschaft als Wissende missachtet und missverstanden werden. Zudem bildet das durch die AGR manifestierte Vokabulare ihre Geschlechter und die Geschlechtsmanifestationen nicht ab. Auch diese Mängel im Begriffssystem sind epistemischer Natur.

17 Die betroffenen Subjekte beschreiben dann auch oft Fälle von Gender Dysphorie als Ergebnis der Erfahrungen mit der AGR.

3.2. Googles Suchmaschine – Was man sucht und was man findet

Der Algorithmus hinter Googles Suchmaschine wird oft kritisiert, weil er Minderheiten benachteiligt, etwa durch sexistische und rassistische Suchergebnisse (um nur einige Problemfelder zu nennen).[18] Betroffen sind dabei sowohl Algorithmen hinter der sogenannten „autocomplete"[19] Funktion, die angefangene Suchfragen mit Vorschlägen ergänzt, als auch die Reihenfolge der Suchergebnisse selbst.[20] Bei Suchanfragen, die mit den Worten „Frauen sollten…" beginnen, waren noch 2013 die *autocomplete*-Vorschläge sexistisch oder frauenverachtend; z. B. „Frauen sollten nicht studieren", „Frauen sollten keine Hosen tragen", „Frauen sollten keine Rechte haben" (Weber 2015). Ähnliche Beispiele gibt es für rassistische *autocomplete*-Vorschläge; so berichtet Safiya Umoja Noble, dass bei einer Google-Suche für den Suchanfang „Black girls" ausschließlich pornografische Ausdrücke vorgeschlagen wurden – Noble selbst suchte eigentlich nach Tipps für Freizeit-Aktivitäten mit ihren Nichten (Noble 2018). Auch die Suchergebnisse selbst sind problematisch, denn die Sortierung relevanter Ergebnisse in Bezug auf *race* bevorzugt rassistische, *white supremacist*[21] Ergebnisse. Dylan Roof, der 2015 in South Carolina neun Afroamerikaner:innen bei einer Bibelstunde in einer Kirche erschoss, gab vor seiner Tat die Suchbegriffe „black on white crime" bei Googles Suchmaschine ein und erhielt dafür primär Ergebnisse aus der

18 Auch *Google Photos* verwendet(e) rassistische Algorithmen, z. B. wurde ein Schwarzes Paar auf einem Foto als Gorillas gelabelt (u. a. Weber 2016, Gebru 2020). Im Jahr 2018 war die einzige Lösung für das Problem, die Kategorie „Gorilla" aus den verfügbaren Kategorien zu entfernen. Google Photos kann dadurch keine Gorillas auf Fotos erkennen (Simonite 2018).

19 Ich verwende den englischen Ausdruck, da er auch in deutschsprachigen Diskussionen teilweise unübersetzt verwendet wird. Eine Alternative ist „Autovervollständigung".

20 Google erklärt, dass die Reihenfolge der Suchergebnisse und auch die automatischen Vorschläge aus den Suchanfragen der Nutzer:innen entstehen und nicht durch den Algorithmus. Da diese Antwort umstritten ist (vgl. Noble 2018) und das Thema epistemischer Ungerechtigkeiten in dem Such-Algorithmus in einem ersten Schritt auch unabhängig von dieser konkreten Verantwortungsfrage untersucht werden kann, klammere ich diese Problematik in dem vorliegenden Text aus.

21 Ich lasse auch diesen Ausdruck unübersetzt, da er als technischer Terminus behandelt werden kann.

white supremacy Bewegung. Laut Roofs Angaben waren diese Ergebnisse der Beginn seiner Radikalisierung, die zu dem Amoklauf führte (Hersher 2017).

Diese Mängel von Googles Suchmaschinen-Algorithmus sind epistemisch relevant, weil die Suchmaschine essentielle epistemische Funktionen erfüllt; u. a. liefert sie Informationen oder wird bei offenen Fragen jeglicher Art zu Rate gezogen. So wie die Suchmaschine derzeit diese Aufgaben übernimmt, trägt sie zu epistemischen Ungerechtigkeiten bei. Grundlegend ist festhalten, dass wie bei AGR der Suchalgorithmus und die Ergebnisse bestehende epistemische Ungerechtigkeiten in Kommunikation und Wissensproduktion materialisieren. Die Perspektiven und Beiträge, die Google einsetzt und präsentiert, sind überwiegend aus einer dominanten – weißen, männlichen – Perspektive verfasst und lassen marginalisierte Perspektiven nicht zur Sprache kommen. Diese Mängel sind folgenreich, denn die Google Suchmaschine bestimmt als „surrogate expert" (Simpson 2012) oder auch als „artificial testifier" (Gunn und Lynch 2018) den Inhalt von Überzeugungen mit und trägt zur Qualität des epistemischen Systems bei.

Darin stecken zwei verbundene Facetten von epistemischer Ungerechtigkeit. Als „artificial testifier" wird Googles Suchmaschine zu sehr geglaubt, ihr kommt ein Glaubwürdigkeitsüberschuss zu, und wie Medina betont geht dieser Überschuss auf Kosten anderer Akteur:innen, die weniger Glaubwürdigkeit erhalten (Medina 2013 , 62–64). Die Gruppen, die nicht validiert werden, indem sie zu einem Suchergebnis auf einer der ersten Seiten gemacht werden, verlieren dadurch sogar Glaubwürdigkeit. Hier behandele ich Googles Suchalgorithmus wie eine (künstliche) epistemische Akteur:in, die in das komplexe Netz von epistemischen Ungerechtigkeiten selbst eingebunden ist und von epistemischen Ungerechtigkeiten profitiert. Die zweite Facette betrifft marginalisierte Gruppen und ihre Beiträge. Wenn die Beiträge dieser Gruppen oder Beiträge, die Perspektiven von marginalisierten Gruppen berücksichtigen, erst auf den späteren Ergebnisseiten erscheinen, dann können marginalisierte Gruppen nicht so gut zum Überzeugungssystem beitragen, das Google durch seine Ergebnisliste formt, ihr Zeugnis wird geringer bewertet oder gar ausgeschlossen – eine Form von testimonialer Ungerechtigkeit. Ihr Wissen wird nicht weitergegeben oder abgewertet. Zudem können diese Suchergebnisse es marginalisierten Gruppen erschweren, ihre eigenen Perspektiven wiederfinden. So können die Suchvorschläge zusätzlich den begrifflichen Rahmen so einschränken, dass die Begriffe, die für ein Mitglied einer marginalisierten Gruppe notwendig sind, ausgeschlossen

und verdrängt werden. Dabei handelt es sich dann um Beiträge zu hermeneutischer Ungerechtigkeit, da diese epistemische Ressource Begriffe und Wissen, die für marginalisierte Erfahrungen notwendig sind, nicht verfügbar macht.

Unabhängig davon wie die Intentionen und Überzeugungen von dem Kollektiv „Google" oder der Dachgesellschaft „Alphabet" bestimmt werden, liegt hier eine Exklusion erster Ordnung und eine Exklusion dritter Ordnung vor. Eine Exklusion erster Ordnung erfolgt durch Googles Suchalgorithmus, weil ein Aspekt dieser "shared epistemic resource[]" (Dotson 2014, 123) nicht korrekt funktioniert, beispielsweise der Algorithmus im Hinblick auf *autocomplete* Vorschläge für *race*-sensitive Begriffe oder die Sortierung der Ergebnisse. Diese Mängel würden nicht den ganzen Suchalgorithmus im Hinblick auf das Ziel, Wissen im Internet durchsuchbar, findbar und zugänglich zu machen, betreffen. Eine Exklusion dritter Ordnung liegt vor, weil der Algorithmus nicht die epistemische Aufgabe erfüllt, Wissen fair und allgemein zu verbreiten und zugänglich zu machen, sondern die dominante Perspektive bevorzugt. Die Exklusion dritter Ordnung erklärt auch, warum es so schwierig ist, die epistemischen Ungerechtigkeiten und den Ausschluss zu erkennen: Man muss aus Googles Leistungen und Produkten heraustreten, um die Mängel zu erkennen.

3.3. PredPol – Der vermeintliche Versuch, Verbrechen zu reduzieren

PredPol ist ein Programm, das in den USA verwendet wird, um die Zahl von Verbrechen zu reduzieren.[22] PredPol erstellt auf der Grundlage von vergangenen Arten von Straftaten, dem Ort der Straftat und der Uhrzeit der Straftat Vorhersagen, wo es in der nächsten Zeit zu einer Straftat kommen wird, damit die Polizei ihre Patrouillen an dieses Gebiet anpassen kann (Lum und Isaac 2016). Weitere verwandte Versuche Straftaten zu reduzieren verwenden geografische Daten eines Ortes und die dortigen Verhaftungszahlen

22 PredPol selbst wird überwiegend in den USA verwendet, aber es gibt andere Programme, mit einer ähnlichen Funktion, die auch in Europa verwendet werden. PRECOBS (Pre-Crime-Observation-System) ist beispielsweise ein deutsches Programm, das u. a. in Zürich eingesetzt wird, um die Zahl von Einbrüchen zu reduzieren (vgl. Leese 2018). Ich konzentriere mich hier auf PredPol, weil andere Modelle möglicherweise anders funktionieren, und weil für PredPol derzeit genauere Analysen existieren. Siehe bspw. Kayser-Bril (2020) für Kritik an PRECOBS und weiteren in der Schweiz verwendeten Programmen.

dazu, um eine sogenannte „heat list" zu erstellen, mit Personen, die Straftaten begehen könnten. In Chicago etwa sucht die Polizei dann Personen auf, die auf der Liste stehen, und warnt sie davor Straftaten zu begehen (Lum und Isaac 2016).

Das mag zunächst nach einer guten Idee klingen, allerdings gelingt es PredPol nicht zukünftige Straftaten vorherzusagen, sondern nur zukünftige Polizeieinsätze und das liegt an dem Daten-Input, mit dem PredPol arbeitet. Die Daten sind vergangenen Polizeieinsätzen entnommen, so dass sie beispielsweise nur tatsächliche Festnahmen wegen Drogenbesitz enthalten und nicht alle Verbrechen, die mit Drogen zu tun haben. Lum und Isaac zeigen durch einen Vergleich zwischen PredPol-Ergebnissen und Daten aus der „National Survey on Drug Use and Health" (NSDUH) aus dem Jahr 2011 für einen Teil von Oakland, dass Drogen weiter verbreitet sind als nur an den Orten, an denen Polizeieinsätze stattfinden. Somit zeigt PredPol nur einen kleinen Ausschnitt an Orten, an denen tatsächlich Drogen zu finden sind (Lum und Isaac 2016). Die Orte, an denen die Einsätze gehäuft stattfinden und die von PredPol angezeigt werden, sind Orte, an denen vor allem nicht-Weiße leben. Dementsprechend, so Lum und Isaac, sind Schwarze Bürger:innen von Oakland etwa doppelt so häufig wie Weiße von *Predictive Policing* betroffen. Bürger:innen, die nicht Schwarz oder Weiß sind, sind etwa anderthalb Mal so häufig wie Weiße von *Predictive Policing* betroffen. Der Drogenkonsum in Oakland ist laut der Daten der „National Survey on Drug Use and Health" jedoch zwischen Weißen, Schwarzen, und Anderen gleich verteilt und rechtfertigt damit die faktische Verteilung der Einsätze nicht (Lum und Isaac 2016, 18). PredPol ist damit nicht das „scientific, evidence-based and race-neutral" (Lum und Isaac 2016, 18) Programm, das es vorgibt zu sein. PredPol diskriminiert, weil es auf Grundlage von Daten funktioniert, die die rassistischen Vorurteile spiegeln, die bei der Bekämpfung von Drogendelikten häufig zum Tragen kommen. Dabei verstärkt PredPol die Verzerrungen sogar und sorgt so für eine Feedbackschleife, da durch die erhöhten Einsätze in den Gebieten auch mehr Verbrechen in den Gebieten registriert werden als in den Gebieten, die nicht markiert wurden durch PredPol, was wiederum zu einer größeren Anzahl an Einsätzen führt.

Es ist nicht nur der Datensatz, der zu Fehlern führt, sondern die Kombination von Datensatz und der Funktionsweise des Algorithmus sowie der Praxis, in der er zur Anwendung kommt. Diese Komponenten dürfen nicht getrennt betrachtet werden, denn der Vorhersage-Algorithmus existiert in einem Kontext, in dem Menschen, Städte und Institutionen Algorithmen zur

Prävention verwenden. Daher muss die Analyse über den scheinbar klar umgrenzten Algorithmus hinausgehen und auch Anwendungskontexte, Datenproduktion, Datenverwendung, etc. einbeziehen.

Ein Beispiel für diese komplexen, übergreifenden Zusammenhänge bietet der folgende Fall aus Chicago, der nicht PredPol betrifft, aber dennoch aussagekräftig für die Probleme in und durch Algorithmen ist. Lum und Isaac berichten beispielsweise von einem jungen Mann in Chicago, den die Polizei ermahnte keine weiteren Straftaten zu begehen, dabei hatte er noch nie eine Straftat begangen, er hatte kein Strafregister. Er wurde ausgewählt als eine Person, die bald eine Straftat begehen würde, auf einer , wie oben bereits erwähnt, sogenannten „Heat list" (Lum und Isaac 2016, 15). Hier zeigt sich besonders deutlich das in Abschnitt 2 erwähnte Problem der Intransparenz: Es ist unklar, was für ein Algorithmus, mit was für Daten verwendet wird, um diese „Heat List" zu erstellen. Vielleicht ist es ja gar kein Algorithmus, sondern nur eine bestimmte Tabelle? Der betroffene junge Mann und in der Regel auch die Öffentlichkeit hat keinen Zugang zu diesen Details.

Was ist bei diesen Algorithmen – inklusive der Daten, mit denen sie arbeiten – nun der Platz von epistemischen Ungerechtigkeiten? Man findet (1) testimoniale Ungerechtigkeit und *testimonial quieting* sowie (2) hermeneutische Ungerechtigkeit und Ignoranz. Die testimoniale Ungerechtigkeit zeigt sich im Glaubwürdigkeitsüberschuss, der dem Algorithmus, den Polizeidaten und den Entscheidungen der Polizist:innen, die die Polizeidaten konstituieren, zukommt. Die Daten präsentieren ein scheinbar objektives Dokument, aber sie verzerren die Realität, da benachteiligte Orte mehr geprüft werden und dann auch dort mehr Straftaten aufgezeichnet werden, die wiederum in die Daten eingefügt werden.

Auch hier entsteht durch einen Glaubwürdigkeitsüberschuss kontrastiv ein Glaubwürdigkeitsmangel bei den anderen Personen, denen als Folge des einseitigen Überschusses umgekehrt nicht geglaubt wird (vgl. Medina 2013, 62–64). Ihnen wird epistemische Autorität abgesprochen und sie werden nicht als epistemische Subjekte behandelt, sondern vielmehr zum Schweigen gebracht. Im Fall von PredPol wird den Daten der Polizei und den Berechnungen des Algorithmus mehr geglaubt, sie werden als objektiv angenommen. Im Zuge dessen werden die Personen, die in den betroffenen Gebieten wohnen und Minderheiten sind, zu epistemischen Objekten über die geurteilt wird, die aber nicht selbst zu diesen Urteilen beitragen können bzw. dürfen.

Diese testimoniale Ungerechtigkeit ist eng verbunden mit hermeneutischer Ungerechtigkeit, denn sie schlägt sich auch in den Begriffen und Vor-

stellungen der Gesellschaft wieder. Sie schreiben „biased social imaginaries" (Medina 2013, 68) fort, in denen nicht-Weiße viel häufiger in Drogendelikten involviert sind.[23] Diese Bilder aus Biases und Feedbackschleifen werden unbewusst weiter in der Polizeiarbeit genährt. Es zeigen sich so hermeneutische Ungerechtigkeiten im Bezug auf das Fremdbild und möglicherweise das Selbstbild von nicht-Weißen Subjekten. Natürlich ist PredPol nicht ursächlich und einzeln erklärend für die hermeneutische Ungerechtigkeit, aber das Programm ist ein essentieller Bestandteil der Ungerechtigkeiten.

Trotz aller Kritik werden PredPol und ähnliche Ansätze weiterverwendet. Dies können wir wiederum mit Pohlhaus' Begriff der vorsätzlichen hermeneutischen Ignoranz erklären (2012, 715). Hermeneutisch ignorante Subjekte ignorieren epistemische Beiträge von marginalisierten Gruppen und sind so nicht in der Lage die geteilte Welt zu verstehen. Im Fall von PredPol ignorieren sie die alternativen Beschreibungen und Konstruktionen, die marginalisierte Gruppen liefern, und halten an ihrer eigenen Konzeption und Evaluation fest. Sie sind sich nicht bewusst, dass sie marginalisierte Gruppen aus dem Gespräch über Strafprävention systematisch ausschließen. Diese hermeneutische Ungerechtigkeit kann auch Mitgliedern marginalisierter Gruppen die Erkenntnis versperren, dass sie benachteiligt werden und im Diskurs nicht als vollwertige Beteiligte teilnehmen können (Mason 2011). Sie kommen bei der Arbeit mit PredPol ja nur als Täter:innen und Datensätze vor.

Man könnte an dieser Stelle einwenden, dass es übertrieben ist, zu behaupten, dass PredPol epistemisch ungerecht ist. Schließlich spiegelten die Daten nur die Praxis der Polizeiarbeit und somit sind allenfalls diese epistemisch ungerecht, nicht aber PredPol. Doch dieser Einwand fußt auf einer individualistischen Konzeption von epistemischen Ungerechtigkeiten und übersieht materielle und strukturelle Manifestationen von epistemischen Ungerechtigkeiten. PredPol ist ein Teil der Strukturen, die epistemische Ungerechtigkeit fortschreiben und erhalten. Und epistemische Ungerechtigkeit wird nie nur von einem Individuum oder einer Gruppe von Individuen ausgeübt, vielmehr ist sie inhärent auch kollektiv und strukturell (vgl. Anderson 2012) und ko-konstituiert durch Algorithmen wie PredPol.

Die Exklusion dritter Ordnung in Dotsons Terminologie ist hier besonders salient, denn der Algorithmus funktioniert nur aus der dominanten

23 Catalas Begriff der hermeneutischen Dominanz würde die Analyse auch an dieser Stelle sinnvoll vertiefen (2015).

Perspektive richtig. Seine Mängel erklären sich daraus, dass dominante epistemische Ressourcen verwendet werden und andere Mitglieder systematisch ausgeschlossen werden. Wiederum erklärt die Exklusion dritter Ordnung, warum es aus dominanter Perspektive *prima facie* schwierig ist, epistemische Ungerechtigkeiten in PredPol zu erkennen und sie zu korrigieren. Denn man muss über das verwendete epistemische System hinausgehen, um PredPol kritisch analysieren zu können. Man muss etwa über die Ergebnisse der Polizeiarbeit hinausgehen, diese Arbeit hinterfragen, und damit auch anderen Input einbeziehen, beispielsweise den Input von denjenigen, die durch unangemessene Polizeieinsätze oder Präventionsversuche besonders betroffen sind. Nur so kann man erkennen, dass PredPol Polizeieinsätze voraussagt und nicht Drogenverbrechen.

4. Abschluss und Ausblick

Die vorliegende Analyse hat gezeigt, dass Manifestationen von epistemischer Ungerechtigkeit in den drei untersuchten Algorithmen zu finden ist. Besonders häufig sind tatsächlich Formen von testimonialer und hermeneutischer Ungerechtigkeit sowie vorsätzliche hermeneutische Ignoranz. Ein größerer Untersuchungsrahmen von epistemischer Ungerechtigkeit, der beispielsweise auch epistemische Ausbeutung beinhaltet, würde weitere Manifestationen zeigen, bspw. im Ausschluss der essentiellen Beiträge Schwarzer Personen bei der Entwicklung von sozialen Medien (McIlwain 2020, Nelson 2021). Aus Platzgründen muss diese Erweiterung auf eine andere Gelegenheit verschoben werden. In diesem Zusammenhang sollte dann auch die befreiende und widerstandausübende Funktion von Algorithmen untersucht werden, denn ironischerweise können Algorithmen auch Mittel im Kampf gegen epistemische Ungerechtigkeit sein. Das Projekt „On the books“[24] der *University of North Carolina* verwendet beispielsweise Algorithmen und maschinelles Lernen, um rassistische Passagen in Rechtstexten aus der Zeit zwischen Bürgerkrieg und Bürgerrechtsbewegung zu identifizieren.[25]

Die vorliegende Analyse legt nahe, dass epistemische Ungerechtigkeit in Bezug auf Algorithmen sich vor allem an zwei Orten findet: (a) In den Daten, die von epistemischer Ungerechtigkeit durchzogen sind, beispielsweise, weil das Zeugnis von marginalisierten Gruppen gar nicht abgebildet ist. (b)

24 Siehe auch die Website: https://onthebooks.lib.unc.edu/.

25 Siehe für einen weiteren Zugang auch (Velkova und Kaun 2019).

In den Algorithmen, die mit Konzeptualisierungen arbeiten, die marginalisierte Gruppen ausschließen, und damit Fälle von testimonialer und hermeneutischer Ungerechtigkeit sind, wie bei den diskutierten *Automatic Gender Recognition* Programmen.

Wichtig ist, dass *alle* Algorithmen (nicht nur *high-stakes* Algorithmen) Fehler in diesen Feldern haben können. "Garbage in, Garbage out" ist ein Sprichwort unter Programmierenden, das alle Algorithmen betrifft, denn für alle – egal ob klassisch oder auf Basis von maschinellem Lernen – gilt, dass mangelhafter Dateninput auch mangelhafte Ergebnisse produziert. Und auch falsche Konzeptualisierungen in den Algorithmen oder fehlende Inklusion von diversen Perspektiven verschlechtern die Qualität jedes Algorithmus'. Aber bei Algorithmen, in denen es um sozial und persönlich signifikante Themen, wie Polizeieinsätze oder Identität, geht, haben diese Probleme besonders einschneidende Auswirkungen. Und es sind auch diese Algorithmen, bei denen epistemische Ungerechtigkeit besonders virulent ist, weil sie mit ihrem Bezug auf Identität und einem Ausschluss von marginalisierten Gruppen die zentralen Elemente von epistemischer Ungerechtigkeit tangieren. Das heißt nicht, dass diese Art von Algorithmen immer und *per definitionem* epistemisch ungerecht sind, aber es heißt, dass bei ihnen eine besondere Gefahr für epistemische Ungerechtigkeit besteht, die von Entwickler:innen, Anwender:innen, Forscher:innen und der Gesellschaft immer geprüft und verhindert werden sollte.[26]

Literatur

Anderson, Elizabeth. 2012. „Epistemic Justice as a Virtue of Social Institutions". *Social Epistemology* 26 (2): 163–73. https://doi.org/10.1080/02691728.2011.652211.

Berenstain, Nora. 2016. „Epistemic Exploitation". *Ergo, an Open Access Journal of Philosophy* 3 (20201214). https://doi.org/10.3998/ergo.12405314.0003.022.

Bratu, Christine, und Hilkje Haenel. 2021. „Varieties of Hermeneutical Injustice: A Blueprint". *Moral Philosophy and Politics* 8 (2): 331–50. https://doi.org/10.1515/mopp-2020-0007.

Buolamwini, Joy, und Timnit Gebru. 2018. „Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification". *Proceedings of Machine Learning Research* 81: 77–91.

26 Feministische Zugänge zu Technologie können hier wichtigen informierten Input liefern, für Beispiele siehe den Sammelband von netzforma* e.V. Verein für feministische Netzpolitik (2020) oder Loh und Coeckelbergh (2019).

Catala, Amandine. 2015. „Democracy, Trust, and Epistemic Justice“. *The Monist* 98 (4): 424–40. https://doi.org/10.1093/monist/onv022.

Costanza-Chock, Sasha. 2020. „Introduction: #TravelingWhileTrans, Design Justice, and Escape from the Matrix of Domination“. In *Design Justice*. https://design-justice.pubpub.org/pub/ap8rgw5e/release/1.

Dastin, Jeffrey. 2018. „Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women“. *Reuters*, 10. Oktober 2018, Abschn. Retail. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G.

Dotson, Kristie. 2011. „Tracking Epistemic Violence, Tracking Practices of Silencing“. *Hypatia* 26 (2): 236–57. https://doi.org/10.1111/j.1527-2001.2011.01177.x.

———. 2012. „A Cautionary Tale: On Limiting Epistemic Oppression“. *Frontiers: A Journal of Women Studies* 33 (1): 24–47. https://doi.org/10.5250/fronjwomestud.33.1.0024.

———. 2014. „Conceptualizing Epistemic Oppression“. *Social Epistemology* 28 (2): 115–38. https://doi.org/10.1080/02691728.2013.782585.

El Kassar, Nadja. i.E. „Wissen, Unwissenheit und Ignoranz: Corona als Chance und Herausforderung für den epistemologischen Diskurs“. In *Wissensproduktion und Wissenstransfer in Zeiten der Pandemie*, herausgegeben von Pedro Schmechtig und Rico Hauswald. Karl Alber.

Eubanks, Virginia. 2017. *Automating inequality: how high-tech tools profile, police, and punish the poor*. New York, NY: St. Martin’s Press.

Fricker, Miranda. 2007. *Epistemic injustice: power and the ethics of knowing*. Oxford ; New York: Oxford University Press.

Fry, Hannah. 2018. *Hello World: How to Be Human in the Age of the Machine*. New York, London: W. W. Norton & Company.

Gebru, Timnit. 2020. „Race and Gender“. In *The Oxford Handbook of Ethics of AI*, herausgegeben von Markus D. Dubber, Frank Pasquale, und Sunit Das, 251–69. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190067397.013.16.

Gunn, Hanna Kiri, und Michael P. Lynch. 2018. „Googling“. In *The Routledge Handbook of Applied Epistemology*, herausgegeben von David Coady und James Chase, 1. Aufl., 41–53. Routledge. https://doi.org/10.4324/9781315679099.

Medina, José, und Tempest Henning. 2021. „My Body as a Witness: Bodily Testimony and Epistemic Injustice“. In *Applied Epistemology*, herausgegeben von Jennifer Lackey, 171–90. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198833659.003.0008.

Hersher, Rebecca. 2017. „What Happened When Dylann Roof Asked Google For Information About Race?“ *NPR*, 10. Januar 2017, Abschn. America. https://www.npr.org/sections/thetwo-way/2017/01/10/508363607/what-happened-when-dylann-roof-asked-google-for-information-about-race.

Hicks, Mar. 2019. „Hacking the Cis-tem“. *IEEE Annals of the History of Computing* 41 (1): 20–33. https://doi.org/10.1109/MAHC.2019.2897667.

Hill, Robin K. 2016. „What an Algorithm Is“. *Philosophy & Technology* 29 (1): 35–59. https://doi.org/10.1007/s13347-014-0184-5.

Introna, Lucas D., und Helen Nissenbaum. 2000. „Shaping the Web: Why the Politics of Search Engines Matters“. *The Information Society* 16 (3): 169–85. https://doi.org/10.1080/01972240050133634.

Kayser-Bril, Nicolas. 2020. „Swiss Police Automated Crime Predictions but Has Little to Show for It“. *AlgorithmWatch* (blog). 22. Juli 2020. https://algorithmwatch.org/en/swiss-predictive-policing.

Leese, Matthias. 2018. „Predictive Policing in der Schweiz: Chancen, Herausforderungen, Risiken“. *Bulletin zur schweizerischen Sicherheitspolitik*, 57–71.

Lenzen, Manuela. 2018. *Künstliche Intelligenz: was sie kann & was uns erwartet*. Originalausgabe. C.H.Beck Paperback 6302. München: C.H. Beck.

Loh, Janina, und Mark Coeckelbergh, Hrsg. 2019. *Feminist Philosophy of Technology*. Techno:Phil – Aktuelle Herausforderungen der Technikphilosophie. J.B. Metzler. https://doi.org/10.1007/978-3-476-04967-4.

Lum, Kristian, und William Isaac. 2016. „To predict and serve?“ *Significance* 13 (5): 14–19. https://doi.org/10.1111/j.1740-9713.2016.00960.x.

Mason, Rebecca. 2011. „Two Kinds of Unknowing“. *Hypatia* 26 (2): 294–307. https://doi.org/10.1111/j.1527-2001.2011.01175.x.

Medina, José. 2013. *The epistemology of resistance: gender and racial oppression, epistemic injustice, and resistant imaginations*. Oxford ; New York: Oxford University Press.

Mitchell, Melanie. 2019. *Artificial intelligence: a guide for thinking humans*. New York: Farrar, Straus and Giroux.

Muthukumar, Vidya, Tejaswini Pedapati, Nalini Ratha, Prasanna Sattigeri, Chai-Wah Wu, Brian Kingsbury, Abhishek Kumar, Samuel Thomas, Aleksandra Mojsilovic, und Kush R. Varshney. 2018. „Understanding Unequal Gender Classification Accuracy from Face Images“. *arXiv:1812.00099 [cs, stat]*, November. http://arxiv.org/abs/1812.00099.

netzforma* e.V. Verein für feministische Netzpolitik. 2020. *Wenn KI, dann feministisch. Impulse aus Wissenschaft und Aktivismus*. Berlin.

Noble, Safiya Umoja. 2018. *Algorithms of oppression: how search engines reinforce racism*. New York: New York University Press.

O’Neil, Cathy. 2016. *Weapons of math destruction: how big data increases inequality and threatens democracy*. First edition. New York: Crown.

Origgi, Gloria, und Serena Ciranna. 2017. „Epistemic Injustice. The Case of Digital Environments“. In *The Routledge Handbook of Epistemic Injustice*, herausgegeben von Ian James Kidd, Gaile Pohlhaus Jr., und José Medina, 303–12. London ; New York: Routledge, Taylor & Francis Group. https://doi.org/10.4324/9781315212043.ch29.

Orwat, Carsten. 2019. *Diskriminierungsrisiken durch Verwendung von Algorithmen: eine Studie, erstellt mit einer Zuwendung der Antidiskriminierungsstelle des Bundes*. 1. Auflage. Baden-Baden: Nomos.

Pohlhaus Jr., Gaile. 2012. „Relational Knowing and Epistemic Injustice: Toward a Theory of Willful Hermeneutical Ignorance“. *Hypatia* 27 (4): 715–35. https://doi.org/10.1111/j.1527-2001.2011.01222.x.

———. 2017. „Varieties of epistemic injustice“. In *The Routledge handbook of epistemic injustice*, herausgegeben von Ian James Kidd, Gaile Pohlhaus Jr., und José Medina, 13–26. Routledge handbooks in philosophy. London ; New York: Routledge, Taylor & Francis Group.

Rudin, Cynthia. 2019. „Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead“. *Nature Machine Intelligence* 1 (Mai): 206–15.

Scheuerman, Morgan Klaus, Jacob M. Paul, und Jed R. Brubaker. 2019. „How Computers See Gender: An Evaluation of Gender Classification in Commercial Facial Analysis Services“. *Proceedings of the ACM on Human-Computer Interaction* 3 (CSCW): 1–33. https://doi.org/10.1145/3359246.

Scotto, Silvia Carolina. 2020. „Digital Identities and Epistemic Injustices“. *HUMANA.MENTE Journal of Philosophical Studies* 13 (37): 151–80.

Segev, Elad. 2010. *Google and the digital divide: the bias of online knowledge*. Chandos internet series. Oxford: Chandos Pub.

Simonite, Tom. 2018. „When It Comes to Gorillas, Google Photos Remains Blind“. *Wired*, 11. Januar 2018. https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/.

Simpson, Thomas W. 2012. „Evaluating Google as an Epistemic Tool“. *Metaphilosophy* 43 (4): 426–45. https://doi.org/10.1111/j.1467-9973.2012.01759.x.

Velkova, Julia, und Anne Kaun. 2021. „Algorithmic resistance: media practices and the politics of repair“. *Information, Communication & Society* 24 (4): 523–40. https://doi.org/10.1080/1369118X.2019.1657162.

Villani, Cédric. 2018. „For a meaningful artificial intelligence: Towards a French and European strategy.“ https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf.

Vincent, James. 2021. „Automatic Gender Recognition Tech Is Dangerous, Say Campaigners: It’s Time to Ban It“. The Verge. 14. April 2021. https://www.theverge.com/2021/4/14/22381370/automatic-gender-recognition-sexual-orientation-facial-ai-analysis-ban-campaign.

Weber, Sara. 2016. „Wenn Algorithmen Vorurteile haben“. Süddeutsche.de. 15.Januar 2016. https://www.sueddeutsche.de/digital/diskriminierung-wenn-algorithmen-vorurteile-haben-1.2806403.

Winner, Langdon. 1980. „Do Artifacts Have Politics?“ *Daedalus* 109 (1): 121–36.

Wolf, Mattis, Katelijn van den Berg, Shungudzemwoyo P. Garaba, Nina Gnann, Klaus Sattler, Frederic Stahl, und Oliver Zielinski. 2020. „Machine Learning for Aquatic Plastic Litter Detection, Classification and Quantification (APLASTIC-Q)“. *Environmental Research Letters* 15 (11): 114042. https://doi.org/10.1088/1748-9326/abbd01.

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