



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Butler, Gates und Cyborg

Gender und Künstliche Intelligenz in posthumanen Zeiten “

verfasst von / submitted by

Katharina Klappheck B.A.

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Master of Arts (MA)

Wien, 2021/ Vienna 2021

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

UA 066 824

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Masterstudium Politikwissenschaft

Betreut von / Supervisor:

Univ.-Prof. Dr. Barbara Prainsack

Inhaltsverzeichnis

0.) Disclaimer	4
I.) Einleitung	4
A.) Relevanzbegründung	7
B.) Forschungsfrage	12
C.) Gliederung.....	12
D.) Forschungsstand.....	13
1.) Fallstudien und exemplarische Illustrationen schadhafter KI Systeme.....	14
a.) Intersektionale Fallstudien.....	14
b.) Gender Queere Studien mit exemplarischen Charakter.....	15
2.) Studien mit sozialwissenschaftlicher Perspektive.....	17
3.) Neue theoretische Perspektiven.....	18
E.) Weiße Flecken der Forschung.....	20
F.) Vorstellung des Untersuchungsgegenstands	22
1.) Beschreibung.....	22
2.) Fallauswahl	23
a.) Artefakt oder Studie.....	23
b.) Software als Eigentum.....	24
c.) Die Frage des ethischen Samples.....	24
d.) Begründung der vorliegenden Studienwahl.....	25
G.) Definitionen grundlegender Begriffe.....	25
1.) Künstliche Intelligenz (KI).....	26
a.) Alte und neue KI.....	27
b.) Big Data und/oder KI.....	29
c.) Algorithmen/Software oder worüber wir sprechen.....	29
d.) KI und die Sozialwissenschaften.....	31
2.) Queer.....	32
II.) Theorieteil.....	33
A.) Einordnung der Theoriematrix.....	34
1.) Posthumanismus als Zeitdiagnose	34
2.) Technowissenschaft als Paradigma der KI Forschung.....	36
B.) Geschlechtertheoretischer Ausgangspunkt der Arbeit.....	40
1.) Geschlecht und Technologie (eine Verhältnisbestimmung).....	41
2.) Gender als Technik.....	42
a.) Geschlecht als Norm.....	42
b.) Geschlecht als automatische Konstruktion.....	44
c.) Geschlecht als digitale Kategorie.....	46
d.) Gender, KI und Macht.....	47
C.)Theorie Matrix.....	49
III.) Hauptteil.....	50
A.) Zielstellung der Studie.....	50
1.) Definition und Vorstellung der Zielstellung.....	51
a.) Hinführung zur Zielstellung des Gendershades Projekts.....	52
b.) Definition der Zielsetzung (FAT Algorithmen).....	55
i.) AI Ethics	56
ii.) Fairness.....	57
iii.) Accountability.....	58
iv.) Transparency.....	61
2.) Emperie.....	63

a.) Bias als konkrete Genderregulierung.....	63
i.) Gender Data Gap.....	64
ii.) Privilege Hazard.....	66
iii.) Queerfeindlichkeit und Genderclassifyern.....	68
b.) Genderperformance im Gendershades Projekt.....	69
c.) Von der Gender Performance zu Posthumanistischer Performance.....	72
i.) Politischer Cut.....	77
ii.) Gender als Messstandpunkt.....	78
d.) Protokategoriales Genderkonzept im Gendershades Projekt.....	79
3.) Polittheoretische Analyse des Gendershades Projekts	82
a.) Posthumanistische Performativität und das Subjekt.....	83
b.) Posthumanistische Performativität und Macht.....	85
c.) Posthumanistische Performativität und Handlungsmacht.....	88
4.) Zwischenfazit.....	89
B.) Technische Funktionsweise.....	89
1.) Gendertheoretische Definition von (KI-)Technologie	90
a.) Herleitung.....	91
i.) Gender als Norm/ Gender als Automat?.....	91
ii.) Race und Technologie.....	92
b.) Definition.....	93
i.) ...Technologie.....	93
ii.) Gender als.....	95
c.) politikwissenschaftliche Implikationen/ Spezifizierung.....	97
i.) Macht.....	97
ii.) Politische Ebene.....	99
2.) Technische Grundlagen der Genderclassifier.....	100
a.) Automatische Mustererkennung.....	100
i.) Data Mining.....	101
ii.) Statistik als theoretische Grundlage.....	101
b.) Gesichtserkennungsalgorithmen.....	102
c.) Benchmarks.....	103
3.) Gender als kategorialer Standard.....	104
a.) Definition des Konzeptes Kategorie	105
i.) Herleitung (Praxis der Klassifikation).....	106
ii.) Begriffsbestimmung.....	106
iii.) Ressourcen, Grenzen und das Internet.....	108
iv.) Gender als Standard.....	108
v.) Exklusionsmechanismen.....	110
b.) Empirie.....	111
i.) Gender als Kategorie des PPB.....	111
ii.) Sortierung.....	113
iii.) Normierung.....	113
iv.) Standardisierung als Wissenschaftlichkeit.....	115
α.) Der reliable Genderstandard.....	116
β.) Gender als valider Standard	117
c.) Polittheoretische Analyse.....	118
4.) Gender als agentielles Protokoll.....	124
a.) Definition.....	125
i.) Theoretische Herleitung.....	125
ii.) Protokollfeminismus (Begriffsbestimmung).....	127

iii.) Der Apparat (Begriffsbestimmung).....	132
b.) Empirie.....	135
i.) Das Label als Genderprotokoll.....	135
ii.) Protokoll als „how to do benchmarks“.....	135
iii.) protokollarische Relation.....	136
iv.) Das Benchmark als agentieller Apparat.....	138
c.) Polittheoretische Analyse.....	140
i.) Technopolitics.....	141
ii.) Operationalisierung und Agency.....	141
iii.) Mesopolitik.....	142
d.) Queeres Benchmark als feministisches Protokoll (Polittheoretische Intervention).....	145
i.) How to queer AI.....	146
5.) Zwischenfazit.....	152
C.) Arbeitsteilung und Gender im Gender Shades Projekt.....	153
1.) Definition.....	155
a.) Arbeitsteilung und Themengebiete.....	157
i.) Hardware und Arbeitsteilung.....	157
ii.) Prekäre und elitäre Arbeit mit und an KI.....	158
iii.) Revolution und das Verschwinden der (weißen, westlichen) Arbeit.....	160
2.) Empirie.....	162
a.) Die besondere Stellung Schwarzer KI Wissenschaftlerinnen.....	162
b.) Elitäre Wissensarbeit.....	163
c.) Der Umgang mit Daten.....	164
3.) Polittheoretische Analyse.....	165
a.) Der globale Süden als Vorbild	165
b.) Frau sein und Arbeitsteilung.....	166
4.) Zwischenfazit.....	167
IV.) Fazit.....	167
A.) Vorstellung der Forschungsergebnisse.....	168
1.) Antwort zur Forschungsfrage 1.....	168
2.) Antwort zur Forschungsfrage 2.....	172
B.) Methodenkritik.....	176
1.) Geografische Einschränkungen.....	176
2.) Theoretische Homogenität.....	177
3.) Interessenkonflikt.....	178
4.) Ein generelles Verbot.....	178
C.) Ausblick.....	179
1.) Das Fortbestehen der Genderclassifier.....	180
a.) Intersektionale Inklusion.....	180
b.) Alternative Datengewinnung.....	181
2.) Das Verschwinden der Genderclassifier	183
D.) Abschluss.....	185
V.) Literaturverzeichnis	186
VI Anhang.....	201
A.) Zusammenfassung.....	201
B.) Abstract.....	201

0.) Disclaimer

Diese Arbeit enthält Rechtschreibfehler und fehlerhafte Grammatik. Dies ist meiner Behinderung zu zuschreiben. Diese Arbeit wurde Freiwilligen zum Lektorat vorgelegt. Ein kostenpflichtiges Lektorat habe ich abgelehnt, da dieses nicht für alle Studierende vorgesehen ist und ich aufgrund meiner Behinderung keine zusätzlichen Kosten sowie Arbeitszeit investieren möchte, im Sinne einer Gleichbehandlung von Studierenden mit und ohne Behinderung.

1.) Einleitung

20% 18% 2,5 % 1,4%

Diese Zahlen beschreiben eine multiple Krise der Künstlichen Intelligenz (KI) auf ökonomischer und akademischer Ebene:

20% verdienen LatinX-Programmierer:innen im Schnitt weniger als weiße männliche Kollegen in den USA. (vgl. West et. al. 2019, S. 13).

20% der Professor:innen im Feld KI sind Frauen (vgl. ebd. S. 10).

18% der Autor:innen der 21 bedeutendsten Konferenzen zu diesem Fachbereich sind Frauen (vgl. ebd.).

2,5% der Festangestellten bei dem Techkonzern Google sind Schwarz (vgl. ebd. S.11).

1,4% der Aktivist:innen, die in Vereinigungen organisiert zur gemeinwohlorientierten Nutzung Künstlicher Intelligenz sind, sind Schwarz (vgl. Nkonde 2019, S.30).

Dies sind Spuren einer tiefgreifenden Problematik der Künstlichen Intelligenz.

So bezeichnet die ehemalige Leiterin der Abteilung für ethische KI des Konzerns Google, Timnit Gebru, diese Homogenität der KI-Schaffenden als „Diversity Crisis“ (vgl. Snow 2018). Diese Krise wird durch unzureichende Daten, vermutlich sogar noch verstärkt, bzw. das Ausmaß verschleiert.

Die wenigen Daten, die sich auf Konzerne (und dabei exklusiv nur auf US-amerikanische Unternehmen) beziehen, wie im vorgelegten Beispiel, sind verzerrt. Google beispielsweise hat für die oben genannten Zahlen, lediglich 80 % Prozent der Angestellten einbezogen und hat nur Abteilungen berücksichtigt, die nicht nur aus weißen Männern bestehen. Andere Konzerne, wie etwa Tesla oder Netflix, veröffentlichen keine Zahlen zur Diversität innerhalb ihrer Firmen (vgl. West et. al. 2019, S.14; Evans, Rangarajan 2017). Dabei bleiben ganze Personengruppen völlig ausgeschlossen von diesen Erhebung zur Geschlechterparität. Personen, die sich weder dem weiblichen noch männlichen Geschlecht zuordnen, werden durch die hier zitierten Zahlen nicht erfasst. Fraglich ist ferner, ob es sich bei den Frauen oder Männern, die diese Zahlen repräsentieren, um Cis-Personen oder Trans-Personen handelt (vgl. West 2019, S.4). Gleichzeitig sind intersektionale Daten, also Zahlen die beispielsweise schwarze Frauen abbilden würden, kaum vorhanden, lediglich der Software Hersteller GitHub veröffentlichte explizite Zahlen zu Geschlecht und Race¹, die intersektional verbunden waren (vgl. Evans, Rangarajan 2017).

Gleichzeitig bezieht sich diese skizzierte Krise nicht nur auf die Arbeitsteilung innerhalb der Herstellung von KI. Die Exklusion von marginalisierten Gruppen lässt sich weiterhin auch innerhalb bereits bestehender Anwendungen beobachten, in Form von rassistischen und sexistischen Technologien:

“2009: LA-based blogger Joz Wang (2009) wrote that the new Nikon S360 camera she’d bought her mom for Mother’s Day kept asking her family members “Did someone blink?”, even though their eyes were always open. Safiya Umoja Noble (2012) observed that her searches for the terms “ Black girls” and “Latina girls” in Google’s search engine turned up blatantly sexist and pornographic results.

2015:Jacky Alciné (2015) logged in to Google Photos to find a new set of albums tagged by the company’s facial recognition software, including one that categorized him and his friend as “Gorillas.”

2016:Microsoft’s AI chatbot Tay learned to use racist and misogynist language in a single day (Vincent, 2016), and

¹ Ich nutzte hier bewusst nicht die deutsche Übersetzung des Wortes. Erstens aufgrund des National sozialistischen Kontextes und der damit verbundenen mörderischen Ideologie der „Rassen Lehre“. Anders als die englische theoretische Arbeit an und mit dem Begriff, lässt sich die deutsche Definition nicht von dem Schrecken der NS Herrschaft lösen (vgl. Knapp, 2005, S. 73).

Zweitens nutzte ich Race hierbei, ausgehen einer voranging US amerikanischen Begriffstradition, in Anlehnung an die geografische Verortung meines Forschungsgegenstand, einer US amerikanischen Studie und der von den Autor:innen verwendeten theoretischen Tradition der Intersektionalität in Anlehnung an die Juristin Kimberle Crenshaw (vgl. Buolamwini, Gebru 2018, S. 3; Crenshaw 1991). Race verweist dabei auf die Herrschaftsverhältnisse, rassistischer Gesellschaft (vgl. Hall 2018, S.32-33). Gleichzeitig spiegelt dieser Begriff,aufgrund der Arbeit Schwarzer, Indigener und Wissenschaftler:innen of Color, die Kämpfe einer kritischen Umdeutung wieder, nach der der Begriff Race als ein Ausgangspunkt Kollektiver Selbstbeschreibung und politischer Kämpfe für BIPOCs funktionieren kann (vgl. Coleman 2009).

ProPublica published an investigative article reporting on the racially discriminatory risk assessment system COMPAS, which they found was twice as likely to falsely label Black defendants as future criminals than white defendants. The algorithm was used in decision-making at every stage in the criminal justice system, from setting bond amounts to evaluating parole (Angwin et al. 2016).“ (Myers West 2020, S. 2)

Des weiteren entwickelte Amazon im Jahr 2018 eine Software, die, anhand von eingereichten Bewerbungen, Kandidat:innen für den weiteren Bewerbungsprozess heraussuchen sollte. Dieses Programm wurde mittels Daten-Sets konzipiert, welche überwiegend männliche, erfolgreiche Bewerbungen enthielten, so dass es in der Anwendung später Männer bevorzugte (vgl. D'Ignazio, Klein 2020, S. 28).

Dies ist nur eine kleine Auswahl an empirischen Beispielen und setzt sich nicht nur im US amerikanischen Kontext fort.

So haben die Künstler Adam Harvey und Jules LaPlace aufgezeigt, dass ein Daten-Set der Stanford University aus dem Jahr 2014 unter anderem von der chinesischen Regierung dazu gebraucht wurde, Gesichtserkennungssoftware zu entwickeln, welche zur Überwachung und Unterdrücken der muslimischen Minderheit der Uiguren diene. Das Daten-Set enthielt Bewegtbilder von Student:innen, die durch Überwachungskameras auf dem Campus aufgenommen wurden. Die Bilder entstanden ohne die Zustimmung der Beteiligten (vgl. Weinberg, 2020 S. 2; Harvey, LaPlace 2021)

Die Soziolog:in Anna Hoffman spricht in diesem Zusammenhang von „Data Violence“, einem Konzept, welches den Ausschluss marginalisierter Gruppen durch KI-Systeme beschreibt. Dabei würden in Anlehnung an staatliche Bürokratisierungs- und Sortierungsprozesse von Staatsbürger:innen, Menschen mittels des staatlichen Gewaltmonopols in spezifische Identitäten gruppiert, etwa durch die Geschlechtszuteilung in der Geburtsurkunde, und so erst als Bürger:innen konstituiert. Bürger:innen sind dabei Machtstrukturen ausgesetzt, wie sexistischer Diskriminierung aufgrund der Zuordnung Frau (vgl. Hoffmann 2020, S. 3). „Data Violence“ würde diese staatliche Gewaltanwendung in Anbetracht der verstärkten Nutzung kategorial arbeitender künstlicher Intelligenz von einem öffentlichen/staatlichen Sektor auch in einen privaten technischen Sektor überführen (vgl. ebd.). Die Technik² ist hier die entscheidende Differenz zur prädigitalen bzw. intelligent vernetzten Gesellschaft. KI ist jetzt der bürokratische Akt, der Menschen zu Bürger:innen, User:innen und Konsu-

2 Im weiteren Verlauf verwende ich Technik und Technologie synonym.

Prinzipiell ist jedoch Technologie, epistemologisch betrachtet die Lehre der Techniken. Technik bezieht sich dabei auf spezifische Fertigkeiten und auch Artefakte. Im englischen jedoch wird Technologie wesentlich weiter gefasst und umfasst prinzipiell auch die deutsche Definition von Technik (vgl. Coleman 2009, S.178) Aufgrund meiner verwendeten Literatur, die vorrangig englischsprachige Literatur und des Ansatz der soziotechnischen Artefakte, dem der Technikbegriff auch trennscharf zugrunde gelegt worden ist, von den von mir zitierten Wissenschaftler*innen, werde ich beide Worte synonym verwenden.

menten macht (vgl. ebd. S.5)

Die oben beschriebene Krise wird für Hoffman in diesem Kontext jedoch nicht durch eine Nachjustierung der sie verursachenden Technologie gelöst, die beispielsweise Schwarze Menschen auf öffentlichen Plätzen besser erkennt, sondern verweist auf ein Problem, welches sich aus sozialer und technischer Verwobenheit ergibt.

So schreibt Hoffmann, daran anschließend, dass Fragen der Machtverteilung und der Legitimation staatlichen und privatwirtschaftlichen Handelns nicht eine Frage technischer oder sozialer Lösung sei sondern eine Verbindung beider Ebenen:

„Indicative of a larger body of limited and atomistic response to problems as structural violence and data discrimination [...], inclusion discourses work as a convenient cover for problems that are [...] not separate from, but intrinsic to technologies predicated on surveillance, social sorting, and optimisation“ (ebd. S. 2)

KI wird somit in Hoffmanns Analyse ein soziotechnisches Artefakt, das einerseits eine gesellschaftliche und ordnungspolitische Funktion für verschiedene Akteur:innen (private Unternehmen, Staaten) übernimmt und andererseits, wie ich bereits beschrieben habe, immer noch ein spezifisches Artefakt des 21. Jahrhunderts ist, welches sich auf informationstechnologischer Ebene von Computerprogrammen der 70iger Jahre unterscheidet. Soziotechnisch meint hierbei folglich diese nicht voneinander zu trennende Verbindung sozialer Einflussnahme durch bestimmte Technologien.

Zweitens verweist das Konzept der „Data Violence“ in diesem Zusammenhang auf die Notwendigkeit, politikwissenschaftlicher Analyse, da es sich um eine Frage staatlicher Macht, vermittelt durch Künstliche Intelligenz, handelt.

Dabei sind diskriminierende KI-Anwendungen und die Exklusion von marginalisierten Gruppen bei deren Produktion keinesfalls zwei getrennte Phänomene oder Krisen sondern zusammenhängende Phänomene. Es handelt sich um eine multiple Krise, die sowohl die Zugänge zur Arbeit an Künstlicher Intelligenz als auch die Auswirkungen betrifft, wobei die Bedeutung von KI als Infrastruktur heutiger Gesellschaften immer weiter zunimmt (West 2020, S.7-8). Ich möchte im Folgendem deshalb die Relevanz einer intersektionalen Studie in Hinblick auf Genderqueere Menschen und ihrer Exklusion durch und mit KI genauer herausarbeiten.

A.) Relevanzbegründung

Technologien, die durch staatliche Behörden und Unternehmen eingesetzt werden, stellen, so meine

weiteren Ausführungen, eine konkrete Gefahr für Transpersonen und vor allem Transpersonen of Color dar. Dabei, so werde ich zeigen, ist die Transfeindlichkeit eine paradigmatische Bedingung für die von mir untersuchten KI-Systeme. West et. al. schreiben dazu in ihrer groß angelegten, mehrjährigen Studie zur Diskriminierung bei der Produktion von KI aufgrund der Kategorien Gender und Race, dass die, von mir vorgestellte, multiple Krise, ein strukturelles Problem sei. Es handelt sich hierbei keinesfalls um eine Reihe von Einzelfällen, d.h. zufälligen firmenpolitischen Entscheidungen hinsichtlich einflussreicher Positionen und/oder einzelnen defekten KI-Anwendungen, die sexistischen Output generieren, sondern um eine Verstärkung bereits existierender Ungleichheitsverhältnisse entlang von Race und Gender. KI, so die Autor:innen, ist ein gesamt gesellschaftspolitisches, und somit politikwissenschaftlich relevantes, Problemfeld des 21. Jahrhunderts.

„From this perspective, locating individual biases within a given technical system—and attempting to fix them by tweaking the system—becomes an exercise in futility. Only by examining discrimination through the lens of its social logics (who it benefits, who it harms, and how) can we see the workings of these systems in the context of existing power relationships. [...] it is not just that AI systems need to be fixed when they misrecognize faces or amplify stereotypes. It is that they can perpetuate existing forms of structural inequality even when working as intended.“ (West et. al. 2019, S.10)

Gerade für demokratische Staaten stellt diese Technologie daher auch eine Legitimationsproblem dar. So werden auf der einen Seite immer stärker KI-Anwendungen, wie Gesichtserkennung in den öffentlichen Raum integriert, sei es als Hilfsmittel polizeilicher Erkennungsdienste oder aber als Zutrittsregularien für Firmengebäude. Gleichzeitig werden diese Technologien von einer sehr homogenen Gruppe von weißen, zumeist US-amerikanischen Männern entworfen (vgl. Ensmenger 2010). Hierbei sind Menschen, die von den negativen Auswirkungen dieser Anwendungen betroffen sind, marginalisierte Gruppen, wie FLINTA-Personen und/oder BIPOCs. Anders formuliert, KI wird von weißen Männern gemacht, die negativen Folgen tragen Andere. Dabei werden wichtige sozio-technische Entscheidungen, wie der Anwendungsradius solcher Technologien oder aber welche Akteur:innen diese überhaupt nutzen dürfen, nicht als demokratische Entscheidung in einem Parlament getroffen, sondern entweder als Anschaffungspunkt im staatlichen Finanzhaushalt bzw. der jeweiligen Behörde oder aber als Firmeneigentum gelabelt und rechtlich abgesichert, so dass eine öffentliche oder akademische Debatte über deren genauen Funktionsweisen und über mögliche Veränderungen nicht stattfinden kann (vgl. Barabas 2020, S.95, Coleman 2009, S.84-89, Schuklenk 2020; Ochigame 2019).

Ich möchte in diesem Zusammenhang, noch einmal dezidiert, auf die Wichtigkeit und Relevanz der

politikwissenschaftlichen Analyse dieser Technologie eingehen, da dieser Punkt zentral für die Ausgangslage der vorliegenden Studie ist. Die multiple Krise von KI dreht sich nicht mehr um Fragen der technischen Ausbesserung von spezifischen Computeranwendungen, etwa universitäre Hochleistungsrechenprobleme. Die Krise weist zwei, für politikwissenschaftliche Analysen bedeutende Spezifika auf. Erstens beruht sie auf einer Technologie, die ein Teil der Infrastruktur kapitalistisch vernetzter Gesellschaften ist. So findet sich Künstliche Intelligenz in Form von digital vernetzten Smart-Cities im Städtebau wieder, Lieferketten werden durch Künstliche Intelligenz für eine Vielzahl an Produkten optimiert und gesteuert, wie beispielsweise für Autos. Das akademische Publizieren ist ohne Künstliche Intelligenz in Form von Forschungsdatenbanken, Netzwerken wie Research Gate, Suchmaschinen und vielen verwandten Anwendungen nicht mehr denkbar (vgl. Halpern et. al. 2017, S. 108—110). Zweitens sind strukturelle Machtverhältnisse entlang von Kategorien wie Gender und Race, gestaltgebend für KI. Ich möchte diesen zweiten Punkt mit einer Analyse eines empirischen Falls genauer ausleuchten.

Dieser konkrete Fall ist im Bereich der automatischen Gesichtserkennung angesiedelt. Es ist ein breites Spektrum bestimmter Technologien. Dabei nehmen entweder Sensoren, wie Kameras, Bilder von Menschen auf und eine KI verarbeitet diese dann, indem sie markante, unveränderlich Marker des gezeigten Gesichts abstrahiert, oder aber, indem ein Foto in das zu bearbeitende Daten-Set einer KI eingespeist wird. Gesichtsanalysesoftware unterscheidet dabei grob zwei Funktionen:

- (a) Das Erkennen (Recognition). Dabei sollen diese Anwendungen, entweder Gesichter verifizieren, also Eintritt gewähren zu Gebäuden oder aber auch das Handy entsperren können, bzw. das Identifizieren von Gesichtern, d.h. anhand vorhandener Daten feststellen, ob eine Person mit Bildern einer Datenbank, beispielsweise dem Passregister von Staaten identisch ist.
- (b) Das Klassifizieren, also anhand eines Fotos dem gezeigten Gesicht beispielsweise eine Genderkategorie zuordnen (vgl. Castelvecci 2020).

Diese Technologien ist bereits jetzt ein wichtiges politikwissenschaftliches Analysefeld, da sie innerhalb hoheitsstaatlicher Aufgaben, wie dem Grenzschutz oder der polizeilichen Überwachung von Menschen, zum Einsatz kommen. So werden Gesichtsidentifikationstechnologien bereits in London und den USA von den dortigen Polizeibehörden eingesetzt (vgl. Richardson et. al. 2019; Fussey, Murray 2019).

Diese KI-Anwendungen werden dabei vorrangig durch das Feld der Surveillance Studies untersucht. Gesichtserkennungstechnologie ist hierbei vordergründig ein besonders prägnantes Beispiel

für die Verbindung von KI und staatlicher exekutiver Macht.

Dabei zählen diese Technologien zur einer neuen Art der „Surveillance“, da sie zwar auf Menschen angewandt wird, jedoch nicht zwingend von Menschen durchgeführt werden muss (vgl. Marx 2016, S.20-21). Zugleich geht diese neue Art der Überwachung mit einer veränderten politischen Situation einher, nicht mehr ausschließlich die exklusiv staatliche, souveräne Handlung, etwa durch die Polizei, ist als gesellschaftspolitischer Hintergrund von Bedeutung für die politikwissenschaftliche Analyse, sondern die Verbindung staatlicher Effektivität und wirtschaftlicher Produktivität. Die Kommunikationswissenschaftlerin Kelly Gates beschreibt dieses auf neue Technik basierende Verhältnis von Staat und Konzern als

„response to a set of conflicting demands of both the state and the business system to individualize and to classify, to include and to exclude, to protect and to punish, to monitor and define parameters, and to otherwise govern populations in the face of their radical destabilization under the wrenching neoliberal reforms instituted in the United States and across the globe during the latter part of the twentieth and early twenty-first centuries.“ (Gates 2011, S.28).

Dabei wird auch in diesem Zitat deutlich, dass eine Analyse des politischen Rahmens in dem Technik entsteht von entscheidender Bedeutung für das Verständnis der Funktionsweisen von KI ist. Die Anforderungen der Staaten und der Unternehmen lassen klassifizierende und sortierende Technologien wie KI möglich werden. Diese spezifischen Bedürfnisse eröffnen somit einen Raum der Kontingenz, der auf Korrelation zwischen Funktionsweise und gesellschaftlichen Erfordernisse hindeutet (vgl. ebd.)

In dem Versuch diese Surveillance Studies auf Geschlechter, politische Bedürfnisse und Faktoren zu beziehen entwirft die Soziolog:in Nadine Dannenberg das Konzept der Queer Surveillance Studies. Hierbei liegt der Forschungsfokus vorrangig auf Fragen der Auswirkungen auf marginalisierte Gruppen in Hinblick auf Geschlecht und Sexualität. Dabei kritisiert Danneberg, dass zwar Einzelfallstudien zu Fragen von diskriminierenden Überwachungspraxen durchgeführt würden, diese aber nicht intersektional konzipiert würden. Diese Studien untersuchten lediglich anhand einer Ungleichheitskategorie wie Race oder Gender oder Sexualität, auf welche Weise Gesichtserkennungstechnologie Menschen aufgrund dieser soziodemografischen Marker diskriminiert und ihnen Schaden zufügt (vgl. Danneberg 2019, S. 32-33).

Im US-amerikanischen Kontext deckte dabei eine kürzlich erschienene Studie auf, dass die Polizeidirektion von New Orleans Vorhersagetechnologie einsetzt, die ermitteln soll, ob und wann Menschen straffällig werden, wobei sie mit rassistischen und cissexistischen Daten unterfüttert wurde.

Ich möchte an dieser Stelle eine längere Passage zitieren:

The report on the Department of Justice investigation of the New Orleans Police Department revealed that officers improperly targeted and arrested transgender residents, sometimes fabricating evidence of a crime as well as exploiting archaic and biased statutes like “crimes against nature [by solicitation]”—a statute that criminalizes sexual conduct that is considered morally unacceptable and requires registration as a sex offender.[...] The Department of Justice found that in addition to these practices being discriminatory, they also raised significant concerns about the “efficient and effective use of resources to ensure public safety” since individuals convicted of “crimes against nature” made up approximately 40% of the jurisdiction’s sex offender registry and the police department was charged with monitoring all registrants’ compliance.[...] Moreover, since 80% of those registrants were also Black, (Richarson et. al. 2019, S.42)

Diese Gefahren sind so allgegenwärtig, dass sich sogar das US-amerikanische Justizministerium gezwungen sah, diese Vorhersagetechnologien als sexistisch und rassistisch einzuschätzen. Um es noch einmal in aller Deutlichkeit zu betonen, diese Daten sind nicht nur diskriminierend, sie bedeuten für die Personen oftmals reale Gewalt, Stigmatisierung und eine ständige Überwachung durch Software und Polizeibeamte (vgl. Carpenter, Marshall, 2018, S.9).

Es lässt sich somit festhalten, dass KI in einem immer stärkeren Maße nicht mehr nur in einem Unternehmerischen Kontext aufzufinden ist sondern für staatliche Aufgaben genutzt wird und auch das Verständnis bezüglich staatliches Handelns, sich weg von einem exekutiven Souveränitäts- hin zu einem Managementbegriffes, der staatlichen Macht, in der Politikwissenschaftlichen Forschung, entwickelt (vgl. Gates 2011, S.28). Andererseits sprechen Wissenschaftler:innen von einer Krise dieser Technologie auf verschiedenen Ebenen, von der Herstellung bis zum Einsatz von KI-Systemen, hinsichtlich der Exklusion und Betroffenheit von marginalisierten Gruppen. Dabei bleibt jedoch die Gruppe Gnder queere Personen, sowohl weißer als auch Schwarzer und of Color Menschen gänzlich unbeleuchtet, trotz ihres extrem hohen Risikos und tatsächlicher Erfahrung mit schädlicher KI in Kontakt zu kommen, wie ich es am Beispiel der Überwachungstechnologien des Polizeidepartments von New Orleans illustriert habe. Die KI Diversitätskrise ist damit auch eine Herrschaftsfrage hinsichtlich der Wirkmächtigkeit von KI in heutigen digitalen Gesellschaften. Gleichzeitig ist es jedoch aufgrund zumeist privatrechtlicher Beschränkungen, wie der Deklaration von KI Anwendungen als Firmeneigentum, sehr schwierig, diese Technologien einer politikwissenschaftlichen Analyse, hinsichtlich Teilnahmevoraussetzungen im Designprozessen, Funktionsabläufen, zugrunde gelegten Daten und ähnlichen, zu bekommen. Dennoch braucht es Studien die empirische Fälle analysieren, um Handlungsempfehlungen abgeben zu können hinsichtlich , demokratische Benutzungsvoraussetzungen und Designentscheidungen, um die oben beschriebene Krise zu meistern.

B.) Forschungsfrage

Ich möchte deshalb, ausgehend von dem Mangel an Theoriearbeit und Fallanalyse und der gleichzeitigen Relevanz dieser im Umgang mit der KI Krise, die ich im weiteren Verlauf noch näher erläutern werde, eine konkrete Gesichtserkennungsanwendung untersuchen und sie politikwissenschaftlich interpretieren, um zumindest einen Ansatzpunkt für eine theoretische Auseinandersetzung mit dieser Technologie zu ermöglichen, die sich aufgrund der Besonderheit eines KI-Systems ergibt. Dabei möchte ich vorrangig auf Gender als ein Konzept eingehen, welches nicht nur Binär zu operationalisieren und zu analysieren ist. Deshalb werde ich mich auf Gender als eine Funktion des Sozialen beziehen, um einerseits Genderbinaritäten, wie sie bereits verwendet werden, mit einzubeziehen und andererseits einen Ausblick gegen zu können, wie eine trans- und queerinklusive Forschung und vielleicht auch Gestaltung von KI im Allgemeinen möglich ist.

Aufgrund der fachlichen Nähe der Technik zur Disziplin der Politikwissenschaft möchte ich mich dabei auf Gesichtserkennung spezialisieren. Meine Forschungsfrage wird sich aufgrund dessen an einem Fall orientieren.

Sie lautet:

F1: Wie ist Geschlecht in Gesichtserkennungssoftware eingelassen, am Beispiel des Gendershades Projektes?

Und ferner:

F2: Wie lassen sich die Ergebnisse politikwissenschaftlich interpretieren?

Ich möchte im folgenden nun die Gliederung der Forschungsarbeit vorstellen und erläutern, wie ich diese Fragen beantworten werde.

C.) Gliederung

Die Folgende Arbeit unterteilt sich in fünf Hauptabschnitte. Ich werde zuerst einen Theoretischen Rahmen vorstellen der die Arbeit anleiten wird.

Hier werde ich zuerst die spezifische theoretische Prämisse erläutern, die meiner Geschlechter theoretischen Überlegungen zugrunde liegt, dem Posthumanismus. Danach werde ich darauf aufbauen ein Wissenschaftsverständnis einführen, welches das Fundament für meine Analyse der zugrunde

gelegten Technologie ist. Dann werde ich meinen Geschlechter theoretischen Rahmen erläutern, den ich für die Arbeit und somit die Bearbeitung der Forschungsfrage verwende.

Dann werde ich ausgehen von der, eingangs beschriebenen, Problemskizze, der multiplen Krise der KI Produktion, den Hauptteil in der Kapitel unterteilen: Erstens die Analyse der Zielsetzung des Gendershades Projektes, als Antwort von Buolamwini und Gebru auf die Diversitätskrise der KI. In einem zweiten Teil analysiere ich die Technischen Funktionen des Gendershades Projektes hinsichtlich der Funktion von Gender und in einem dritten Teil werde ich die Arbeitsteilung im Gendershades Projektes hinsichtlich Geschlecht untersuchen. Dabei sind alle drei Teile identisch aufgebaut. Ich werde zuerst eine Definition des zu untersuchenden Problemfelds im Gendershades Projekt geben und meiner spezifisch für den jeweiligen Bereich verwendete theoretischen Begriffe. Dann werde ich meine Theoretischen Annahmen auf das Gendershades Projekt anwenden. Diese Teile tragen den Titel Empirie. In einem dritten Schritte werde ich jeweils eine dezidiert politikwissenschaftliche Analyse vornehmen, die sich voranging auf Macht und Herrschaftsverhältnisse im Gendershades Projekt bezieht. Alle drei Teile werden von mir danach jeweils mit einem Zwischenfazit versehen. I

n einem fünften Teil der Arbeit werde ich ein Abschließendes Fazit abgeben einschließlich Methodenkritik. Dabei werde ich zuerst meine beiden Forschungsfragen beantworten und dann eine Reflexion Der Arbeit vornehmen. Abschließend gebe ich einen Ausblick für später Forschung hinsichtlich Genderclassifyer allgemein, unter zwei Prämissen, (a) diese Technologie besteht weiterhin und (b) sie ist nicht mehr existent etwa durch Verbote.

Zuvor möchte ich jedoch meine Einleitung beenden, in dem ich zuerst den aktuellen Forschungsstand hinsichtlich KI und der skizzierten Krise gebe. Danach werde ich weiße Flecken der Forschung skizzieren, die meine Forschungsfragen versuchen zu verkleinern. Ich werde daran Anschließend meine Fallauswahl begründen und wichtige Begriffe für die restliche Arbeit klären.

D.) Forschungsstand

Im Folgenden werde ich einen Einblick in den aktuellen Forschungsstand zur multiplen Krise von KI mit besonderem Schwerpunkt auf die Gesichtsanalyse geben. Ich werde zuerst die wenigen bereits vorhandenen Fallstudien zu schadhaften KI-Anwendungen aufzählen und dann, in einem zweiten Schritt, einerseits die sozialwissenschaftlichen Analysen, die diese Studien aufgreifen, andererseits eine breitere Analyse der gesellschaftlichen Auswirkungen dieser Technologien beschreiben. Diese Literatur werde ich anhand zweier unterschiedlicher theoretischer Zugänge behandeln. Eine Vorgehensweise ist es dabei, die bestehenden sozialwissenschaftlichen Theorien zu sozialer Ungleichheit

auf neue Technologien anzuwenden, wie es etwa Chathy O'Neil in ihrem Werk „Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy“ (2017) tut. Der zweite methodische Zugang wie ihn etwa die Medienwissenschaftler:in Wendy Chun vertritt, geht hingegen, vermittelt durch neue intelligente Software, von sich verändernden Herrschaftspraxen aus, so dass diese Ansätze neue politische Theorien durch spezifische Technologien, wie Künstliche Intelligenz, versprechen.

Dies ist vorrangig eine von mir gewählte analytische Trennung, um einen Überblick über die Vielzahl an Literatur in diesem Feld zu gewähren.

Ich werde daran anschließend eine Forschungslücke skizzieren.

1.) Fallstudien und exemplarische Illustrationen schadhafter KI Systeme

Fallstudien sind eine Grundvoraussetzung dafür, die oben beschriebene Krise analysieren zu können, Fallstudien, die sich mit konkreten KI-Systemen beschäftigen liefern die empirische Grundlage für darauf aufbauende systematische und kritische (sozialwissenschaftliche) Arbeiten. Aufgrund jedoch der sehr geringen Anzahl an tatsächlich empirisch durchgeführten Studien dieser Art, habe ich mich entschieden in den Forschungsstand auch Beispiele aufzuführen die beispielcharakter haben, um die Problematiken die ich Eingangs erwähnt habe zu verdeutlichen. Im folgendem werde ich deshalb Fallstudien und Beispiele vorstellen, die einerseits intersektionale diskriminierende KI-Systeme analysieren und andererseits Studien, die sich mit Gender und Gesichtsanalysetechnologien beschäftigen, die Geschlecht nicht nur als cis-geschlechtliche und binäre Variable fassen.

Die Artikel zu Gesichtserkennungstechnologie im speziellen und gleichzeitig ein Großteil der zitierten Beispiele, die auch ich eingangs aufgegriffen habe, stammt aus Zeitungen und Blog-Beiträgen und weniger aus wissenschaftlichen Fallstudien zu, beispielsweise, rassistischen Bilderkennungsprogrammen.

a.) Intersektionale Fallstudien

Trotz der Aktualität und Präsenz des Themas diskriminierender Gesichtserkennungstechnologie, ist die wissenschaftliche Erforschung sozialer Ungleichheit durch und innerhalb dieser Anwendungen entlang intersektionaler, verschränkter Kategorien wie Race und Gender relativ gering.

Nach einer Online-Literaturrecherche, bei der die Tagungsbeiträge großer HCI-Konferenzen (Human Computer Interaction) zwecks kritischer Auseinandersetzung mit der Verbindung von Nutzer:innen und Gesichtserkennungstechnologien einbezogen wurden, und Sichtung von Sekundärliteratur über kritische Studien zu ungerechter KI habe ich drei intersektionale Studien gefunden,

die sich mit den ungerechten Auswirkungen von KI entlang der Achse Gender und Race beschäftigen. Einmal das Gendershades-Projekt von Joy Buolamwini und Timnit Gebur (2018), dann eine Studie von Raji und Star et. al (2020) sowie die Begleit- und Evaluationsstudie von Raji und Buolamwini (2019).

Dabei beziehen sich die zwei letztgenannten Studien auf das Gendershades-Projekt, welches eine Auditstudie ist, die als erste wissenschaftliche Arbeit erforscht hat inwiefern Gesichtserkennungssoftware entlang der Achse Race und Gender performt. Gebur und Buolamwini fanden dabei heraus, dass die Analyse von schwarzen, weiblichen Gesichtern, im Gegensatz zu einer von weißen, männlichen, Fehlerraten von bis zu 25 % aufwies, was das Klassifizieren nach dem Marker Gender anbelangte

Die anderen beiden Studien untersuchten zusätzlich und unter Mitwirkung von Buolamwini und Gebur die Akkuranz der von Amazon vertriebenen Gesichtsanalysesoftware hinsichtlich der Kategorie Race und Gender . Dabei, so die Autor:innen, „erkennen“ alle untersuchten Gesichtsanalyseanwendungen schwarze, weibliche Gesichter signifikant schlechter als weiße, männliche (vgl. Buolamwini, Gebur, 2018, S. 8-9; Raji und Star et. al. 2020, S. 147. Raji, Buolamwini 2019, S.6). Dabei setzten diese Studien gleichzeitig auf ein binäres Geschlechterverständnis welches Trans-Personen und genderqueere Personen nicht mit einschließt (vgl. ebd.).

b.) Gender Queere Studien mit exemplarischen Charakter

Bezüglich trans- und queer-inklusive Studien, also Arbeiten die sich mit Geschlecht jenseits von Mann/ Frau auseinandersetzen, ist das Feld, im Hinblick auf Gesichtserkennungstechnologie, ebenfalls sehr begrenzt. Dabei zeigen diese nun vorgestellten Studien exemplarisch anhand von qualitativen Interviews und Literaturstudien auf in wie fern KI Systeme Genderqueeren Menschen schadet (vgl. Hamidi et. al. 2018, S. 3). Ich habe, anhand des beschriebenen Suchradius, zwei Studien auffindig machen können, die sich mit der Inklusion von Trans-Personen und Gesichtserkennungssoftware beschäftigen. Dabei waren die Studien, im Gegensatz zu den intersektionalen Arbeiten von Buolamwini et. al., nicht auf Design oder Evaluation von Gesichtserkennung angelegt. Eine Studie von Hamidi et. al. ist eine qualitative Erhebung von 12 Trans- und Queer-Perspektiven auf Genderclassifayer³. Alle Teilnehmer:innen, gaben an, dass für sie Genderclassifayer eine Bedrohung ihrer Sicherheit darstellten bzw. darstellen könnten, wenn diese verstärkt im öffentlichen Raum, beispielsweise als Zugangstechnologie dienen sollten. Diese Ängste äußerten die Teilnehmer:innen unabhän-

3 Eine Technologie, die mittels künstlicher Intelligenz, anhand von zumeist Bildern des Gesichts eine Gender-Kategorie diesem Bild zuordnet (vgl Keyes 2018, S. 2018, S. 88:4). Diese Technik wird im weiteren Verlauf der Arbeit noch ausführlicher dargestellt.

gig von der Idee, dass sich diese Technologie mittels technischer Modifikationen verbessern, also akkurater auch für nicht binäre oder Trans-Personen arbeiten könnte (vgl. ebd. S.8). Die zweite Studie ist eine Arbeit von Os Keyes. Hier untersucht Keyes mittels einer quantitativen Literaturanalyse über 400 Arbeiten auf transinklusive/queer inklusive Ansätze innerhalb der Forschung zu eben jenen Genderclassifyern. Keyes fand kein einziges Paper, welches eine Perspektive enthielt, die genderqueere Menschen mit in Design-Überlegungen einschloss oder diese Personen als Nutzer:innen oder Betroffene dieser Technologie einbezog (vgl. Keyes 2018, S. 88:12).

Fallstudien und/oder exemplarische Illustrationen, die sich mit schadhafter KI auseinandersetzen und dabei sowohl intersektional als auch queer- und transinklusiv arbeiten, habe ich nicht ausfindig machen können. Zusammenfassend möchte ich festhalten, dass das Feld der kritischen Analyse von Gesichtsanalyseanwendungen sehr klein ist. Ich habe lediglich vier Studien mit den genannten Parametern gefunden, wobei letztendlich nur das Gendershades Projekt eine wirkliche empirische Fallstudie darstellt. Gleichzeitig steigt die Anzahl der kritischen Analysen dieser Technologien und mit ihr die Forderungen nach mehr und vertiefenden Arbeiten zu und an konkreten empirischen Beispielen, um entweder spezifische Formen der rassistischen und sexistischen Ausbeutung zu identifizieren, oder aber um genauere theoretische Konzepte zu erarbeiten, die wiederum zu einem besseren Verständnis der bereits aufgeführten multiplen Krise der KI und, im Fall meiner Arbeit, der Gesichtserkennungsanalyse durch KI beitragen (vgl. Benjamin 2019, S.67-69, Hester 2018, S.70).

Die Notwendigkeit von Fallstudien spiegelt sich letztendlich auch in den zitierten Quellen wieder, so werden die wenigen, bereits vorhandenen Studien, wie das Gendershades-Projekt, in Sekundärstudien vielfältig und breit zitiert, andererseits werden vorrangig empirische Beispiele aufgezeigt, die aus Blogs und Zeitungsartikeln stammen, wie auch das eingangs zitierte Beispiel aus solchen Quellen stammt (vgl. West 2020, Ananny, Crawford 2018, Weinberg 2020, Hoffmann 2019). Dabei stellt der Zugang zu KI-Systemen als Untersuchungsgegenstand wissenschaftlicher Forschung selbst ein Problem dar, da ein Großteil der Anwendungen, die eingesetzt werden, wie etwa automatische Gesichtserkennungssoftware, Firmeneigentum und damit zumeist durch das Patentrecht geschützt ist, so dass eine genaue Einsicht in die Funktionsweisen dieser Technologie, durch urheberrechtliche Bestimmungen begrenzt ist (vgl. Bucher 2018, S. 41).

Kritische Studien zu KI-Systemen und ihren Auswirkungen sind daher dringend notwendig einerseits aus dem Mangel an Fallstudien, andererseits aus der Notwendigkeit, diese Systeme politikwissenschaftlich zu analysieren, um diese Technologie in bestehende Machtstrukturen einzubeziehen und wiederum ihre Wirkung auf Gesellschaften zu analysieren.

Sozialwissenschaftliche Analysen zu Gesichtserkennungstechnologie sind sehr zahlreich ich möchte deshalb eine analytische Vorstellung vornehmen, die verstärkt auf unterschiedliche Analyseansätze eingeht statt empirische Vollständigkeit zu beanspruchen. Dabei unterteile ich die aufgeführten Ansätze anhand ihrer Verwendung sozialwissenschaftlicher Konzepte zur Analyse exkludierender KI-Systeme. Aufgrund der weniger spezifischen Fallarbeit beinhalten diese Studien, jedoch nicht nur Gesichtserkennungssoftware als empirische Referenzen.

Ich unterscheide diese Studien in Arbeiten, die durch sozialwissenschaftliche Konzepte KI untersuchen, und solche, die diese Technologien nutzen, um neue theoretische sozialwissenschaftliche Konzepte zu entwickeln. Dabei stehen theoretische Überlegungen zu Macht und Ungleichheit im Fokus dieser Studien

2.) Studien mit sozialwissenschaftlicher Perspektive

Studien, die sich aus gesellschaftswissenschaftlicher Perspektive mit einer Vielzahl von diskriminierenden Technologien und Arbeitsteilung innerhalb der KI-Produktion beschäftigen, sind zahlreicher als die eben beschriebenen Einzelfallstudien. Die Analyse schädlicher KI-Auswirkungen ist dabei keinesfalls neu. Bereits in den 1980iger und 1990iger Jahren beschäftigten sich feministische Anthropolog:innen und Informationswissenschaftler:innen aus den USA, aber auch aus Nordeuropa, mit sexistischen Arbeitsverhältnissen und Geschlechterdiskriminierung als implizites Wissen in KI-Systemen (vgl. Adam 1998, Forsythe 2001, Suchman 1987, Hayles 1999). Diese Studien beinhalten jedoch keinen Ansatz der sich kritisch mit Race als Ungleichheitskategorie auseinandersetzt oder das Weißsein der führenden KI-Elite thematisierte. Eine Ausnahme bildete Donna Haraways Ansatz des Cyborgs, einer theoretischen Figur, die die Verbindung des Menschen mit tierischen und technischen Komponenten zur Jahrtausendwende untersuchte und auf die Verbindung internationaler Produktionsketten zur Herstellung von Computertechnologie und Software hinwies (vgl. Haraway 1995a, S.54-59).

In den letzten Jahren jedoch hat sich eine zunehmende Zahl an Publikationen mit dem Thema der intersektionalen Diskriminierung durch und mit KI beschäftigt, die verstärkt auf die heutige Bedeutung dieser Technologie für die Gesellschaften des 21. Jahrhunderts eingehen (vgl. Suchman 2019, S. 56).

Als Standardwerke gelten dabei Safiya Umoja Nobles „Algorithms of Oppression: How Search Engines Reinforce Racism“ (2018), Cathy O'Neils „Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy“ (2017) und Virginia Eubanks „Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor“ (2019).

Diese Werke behandeln dabei anhand einer Vielzahl an empirischen Beispielen, wie bestehende und historisch verwurzelte soziale Ungleichheiten, wie etwa rassistische Diskriminierung in den USA, durch intelligente Software, wie automatisierte Kreditvergabesysteme, aktuell weiterhin ausgeübt und implementiert wird, durch Institutionen, wie Banken und die staatliche Wohlfahrt (vgl. O'Neil 2017, 141-5; Eubanks 2019, S. 181- 184).

Die Autor:innen gehen dabei, auch im Vergleich zu älteren Studien auf die Verbindung von Race und Gender ein. So analysiert Safiya Umoja Noble in ihrer Studie die sexualisierte Darstellung schwarzer Frauen und Mädchen durch Googles Suchergebnisliste (vgl. Noble 2018, S.99-103). Den Studien ist dabei gemein, dass sie die bestehenden Ungleichheitsverhältnisse fortgesetzt sehen durch KI. Die Arbeiten thematisieren, dabei weniger neue Formen von Macht, die durch diese Technologien entstehen könnten. KI ist hier eher ein weiteres Untersuchungsfeld für intersektionale Wissenschaft bezüglich von Machtverhältnissen, und das in einem stark US-amerikanisch geprägten Umfeld. Alle drei Studien behandeln die USA als geographischen Rahmen. So kritisiert der Soziologe Clapperton Chakanetsa Mavhunga die Fixierung westlicher Forschung zu intelligenter Software auf den globalen Norden als Akteur:in und einzigen geographischen Ort, in dem diese Technologie stattfindet. Vielmehr würde sich eine Erweiterung des Fokus auf beispielsweise Afrika als Kontinent der Erfindung alternativer und kritischer Technologien anbieten. Wobei auch das koloniale Erbe westlicher Wissenschaften beleuchtet werden könnte (vgl. Mavhunga 2017, S.8). Allerdings gibt es bereits jetzt wissenschaftliche Arbeiten zu den Auswirkungen von KI als Technologie auf den globalen Süden, die zum einen auf die Ungleichstrukturen eingehen, die auf die Entstehung von KI als Produkt hinweisen, in dem sie beispielsweise analysieren, woher die für heutige Anwendungen gebrauchten Daten in Form von Bildern, Gesundheitsdaten oder anderen personenbezogenen Fakten kommen und wie sie erhoben werden (vgl. Arora 2016). Zum anderen erarbeiten Wissenschaftler:innen jedoch auch Narrative, die sich mit dem globalen Süden als Akteur:in für Hochtechnologie auseinandersetzen (vgl. Pype 2017; Okune 2020; Roussel, Stolfi 2020, S.11-12).

3.) Neue theoretische Perspektiven

Dem gegenüber stehen Theoretiker:innen, die sich verstärkt mit neuen Konzepten von Herrschaft und politischer Handlungsmacht auseinandersetzen, die sich aus der Anwendung von künstlicher Intelligenz und der damit verbundenen Infrastruktur, wie dem Internet, ergeben. Exemplarisch hierfür stehen die Medientheoretiker:innen Wendy Chun mit ihren Arbeiten zur materialistischen und ideologiekritischen Auseinandersetzung mit Software und dem Internet als Topos einer neuen, nicht menschlichen Idee von Freiheit sowie Eugene Thacker und Alexander Galloway, die mit ihrer Ana-

lyse von intelligenten Netzwerken und Protokollen versuchen, das Konzept einer politischen Theorie des 21. Jahrhunderts mittels Netzwerktechnologie zu beschreiben (vgl. Chun 2009, 2013, Thacker, Galloway 2013).

Hierbei entwickelt Chun in den frühen 2000ern entlang der Architektur des Internets eine politische Theorie von Freiheit und Kontrolle, die sich nicht mehr auf klassische Untersuchungsräume der politischen Theorie bezieht, wie das Gefängnis oder die Schule, als Ort der staatlichen Disziplinierung, sondern sie sieht das Internet als (materiellen und theoretischen) Ort einer Kontrollmacht, die Freiheit jenseits einer konkreten Person verspricht und sich auf die Verteilung von Information durch neue Technologien wie intelligente Suchfunktionen bezieht. So schreibt Chun:

„The current configuration of fibre-optic networks challenges disciplinary and regulatory power. Telecommunications monopolies, rules, and regulations have been and continue to be revised, many regulatory techniques have been rendered ineffectual, and many new, more invasive techniques are being introduced.“ (Chun 2009, S. 15).

Hier wird noch einmal deutlich, dass Technik und Macht in Chuns Analyse nicht zeitlich versetzt entstehen, sondern direkt miteinander verbunden. In eine ähnliche Richtung geht dabei auch das Konzept der Politikwissenschaftlerin Barbara Prainsack, für die Firmen wie Amazon oder Google und deren Big-Data-Anwendungen eine neuen Souverän kreieren, den iLeviathan. In Anlehnung an den Vertragstheoretiker Thomas Hobbes garantiert dieser iLeviathan aber nicht mehr den Schutz des Lebens sondern,

„the possibility to purchase or exchange services and goods faster and more conveniently, or to communicate with others across the globe in real time. Increasingly, the iLeviathan also demands that people trade privacy and freedom from surveillance for access to services provided by public authorities [...]“ (Prainsack 2019, S.2).

Die Wissenschaftler:in skizziert hier, im Gegensatz zu Chun, Arbeiten, die jetzige intelligente Anwendungen durchführen. Diese ermöglichen, mittels großer Mengen Daten gesellschaftlich relevante Prognosen zu treffen, wie beispielsweise medizinische Diagnosen (vgl. Ho, Teo 2020, S.2-3)

Galloway und Thacker hingegen entwickeln eine politische Ontologie des Netzwerks, das sich auch auf das Internet als Ausgangspunkt neuer politischer Möglichkeiten, und vorrangig auf die Notwendigkeit neuer politischer Theorie zu Erfassung dieser Möglichkeiten bezieht. Dabei erklären die Autor:innen, dass Netzwerke in Form von Protokollen eine Form der Macht ausüben, die sich nicht mehr auf staatliche Handlungen bezieht oder auf einzelne Konzerne als Akteur:innen, sondern auf einen Verbund menschlicher/nichtmenschlicher Akteur:innen und verschiedener Institutionen, vermittelt durch Technologien wie dem Internet (vgl. Thacker, Galloway 2014, S. 305).

Dabei verweisen Thacker und Galloway explizit auf die Notwendigkeit, weiterer analytischer Arbeit zur Entwicklung politischer Zugänge für die kritische Forschung zu KI (vgl. ebd. S.290). Dabei würden sich technische Bedingungen als politische erweisen und das Politische als technische Funktionsweise wieder gegeben werden. Ihre politische Ontologie von Netzwerken begründet sich dabei für beide in der konkreten Praxis von Informationsnetzwerken (vgl. ebd. 307-308).

Dies lässt mich zur oben erwähnten Forschungslücke überleiten, in die ich daran anschließend meine Forschungsfrage einbetten werde.

E.) Weiße Flecken der Forschung

Der weiße Fleck, die ich dabei versuche zu verkleinern, besteht vor allem in der fehlenden Bandbreite tieferer Analysen bezüglich konkreter Design-Praxen von Gesichtserkennungstechnologie oder, anders formuliert, im Fehlen von Fallstudien, wie ich sie skizziert habe.

Dabei konstatiert eine Vielzahl an Literatur den Mangel an spezifischer theoretischer Arbeit, die Exklusion marginalisierter Gruppen und damit verbundene Machtstrukturen innerhalb von Gesellschaft analysiert (vgl. Bath et. al 2013, S.23; Couldry, Mejias 2019, S.87; Marino 2020, S. 227).

Um noch einmal West et. al. zu zitieren, bedarf es vermehrt Forschung, die politikwissenschaftlich analysieren kann, wie KI konstruiert wird:

„Future work could examine the politics of system design itself, and study AI systems in situated realities. Such work could ask why a system was designed in a particular way, how it was constructed, and whose interests shaped the metrics by which its success or failure is assessed. Rather than solely focusing on improving existing datasets or individual algorithms, future work could also more thoroughly account for how societal discrimination surfaces in data provenance, examining the history and process of dataset construction, and considering how cultural norms and stereotypes were numerated and represented at the time of data creation.“ (West et. al. 2019, S.16)

Ein besonders eklatante Leerstelle stellt dabei die Frage nach der Auswirkung und Konstruktion für genderqueere und Transpersonen dar, da Gender bisher in der Forschung zur multiplen Krise als männlich/weiblich codiert worden ist. Transpersonen finden zwar in manchen Studien Beachtung, wenn es um empirische Beispiele der Exklusion von Transpersonen geht, wie etwa die Gesichtserkennungssoftware des Chauffeurdienstes Uber, die Mitarbeiter:innen, die sich als Trans outeten, keinen Zutritt zur Plattform ließ. Aber genauere Forschung zu Einschreibpraxen binärer Gender-Konzepte und der Exklusion genderqueerer und Transpersonen fehlen fast vollständig (vgl. Myers West 2020, S. 2).

Gleichzeitig beklagen Forscher:innen, wie Ruha Benjamin, dass bisherige theoretische Arbeiten diese Verbindungen von Technik und Macht nicht ausreichend behandeln, da bisherige Konzepte der politischen Theorie die Besonderheit künstlicher Intelligenz als konkreten empirischen Fall nicht fassen können (vgl. Benjamin 2019, S.192-3). Doch ist dieser weiße Fleck der Forschung tatsächlich ein Problem; denn diese KI-Anwendung ist nicht unumstritten, wenn es um die Einbeziehung marginalisierter Gruppen geht.

So argumentiert beispielsweise die Soziologin Anna Hoffmann, dass eine rein technische Inklusion von marginalisierten Personen innerhalb der technischen Anwendung, beispielsweise genauere Gesichtserkennungssoftware, die auch Schwarze Frauen „erkennt“, die soziale bzw. gesellschaftliche Dimension der Exklusion dieser Gruppen verkennt. Nach Hoffmann ist die Technologie selbst ein Ergebnis rassistisch sexistischer Strukturen und reflektiert gesellschaftliche Verhältnisse. So führen, wie in dem zitierten Beispiel, Evaluationen der von US-amerikanischer Polizei eingesetzten Gesichtserkennungssoftware hinsichtlich genauerer Erfassung von Schwarzen Frauen, nicht zu einer kritischen Auseinandersetzung mit den Strukturen, die es erst nötig erscheinen lassen, diese Technologie anzuwenden. Eine rassistische Inhaftierungspraxis, die auf jahrhundertealtem Erbe der Sklaverei beruht, kann nicht kritisch analysiert werden, indem neueste Polizeitechniken „inklusiver“ gestaltet werden. Die Aktivist:in und Fotograf:in Zoé Samudzi fasst diese Gefahr wie folgt zusammen:

„it is not social progress to make black people equally visible to software that will inevitably be further weaponized against us. We are considered criminal and more surveillable by orders of magnitude; whatever claim to a right to privacy that we may have is diminished by a state that believes that we must always be watched and seen.“ (Samudzi 2019)

Die Untersuchung von Machtstrukturen und Exklusion marginalisierter Gruppen durch und mit Technologie, ist somit immer in einem forschungsethischen Graubereich, der, solange diese Technologien existieren, nicht aufgelöst werden kann. Die von mir beschriebene Lücke in der Forschung, ist somit nach dieser Argumentation ein ambivalenter Forschungsstand. Einerseits wird automatische Gesichtserkennung angewendet und das auf höchst bedenkliche Weise, nachgewiesen rassistisch und sexistisch in ihren Auswirkungen, entwickelt durch eine exklusiv cis-männliche weiße Gruppe und ist dabei eine Technologie, die immer stärker in gesellschaftliche Strukturen eingebunden ist (vgl. Campalo, Crawford 2020, S.14-15).

Daher möchte ich deutlich machen, dass ich mich im weiteren Verlauf mit der Funktion von Gender befasse, um eine Studie zu erstellen, die sich mit Genderbinarität und deren Auswirkungen auf die Gestaltung von KI auseinandersetzt. Dabei werde ich auf die Exklusion und Gefahren dieser Exklu-

sion hinweisen. Grundsätzliche ist eine theoretisch begründete Ablehnung dieser Technologien durchaus mit dieser Arbeit vereinbar, da ich mich hier nicht für einer weiteren Nutzung ausspreche, sondern die Funktionsweise von Gender analysiere.

Bevor ich jedoch meinen theoretischen Rahmen erläutere werde ich im folgendem Teil meine Fallauswahl vorstellen und Begründen und daran anschließend zwei für diese Arbeit wesentliche Begriffe erläutern: Künstliche Intelligenz und Queer.

F.) Vorstellung des Untersuchungsgegenstands

Als Untersuchungsgegenstand dient mir für diese Arbeit die Auditstudie des MIT Media Labs „Gender Shades“. Hierbei beziehe ich mich auf die beiden daraus entstandenen Forschungsberichte(vgl. Buolamwini, Gebru 2018; Buolamwini 2017).

1.) Beschreibung

Das Gender Shades Projekt ist ein Audit-Projekt zur Geschlechterkennung der 3 größten Anbieter für Genderclassifier (Microsoft, IBM, Face++).

Die Autor:innen untersuchten hierbei, inwiefern die automatische Gesichtserkennung der dieser Unternehmen die Beispielbilder richtig, also das angegebene Gender der jeweiligen Personen korrekt, feststellt. Das Sample umfasste 1270 Porträtaufnahmen von 1270 verschiedenen Personen. Es wurde, um geschlechtliche binäre Parität und Schwarze Menschen und People of Color sichtbar zu machen, ein neues Datensample im Projekt erstellt. Dieses Sample setzt sich aus 6 Parlamenten zusammen. Drei europäische und drei afrikanische Parlamente wurden aufgrund der ausgeglichenen Verteilung der Marker Race und Gender gewählt (s. Abbildung 2). Die Autor:innen konnten nachweisen, dass die Genderclassifier „darker females“ mit einer Fehlerquote von bis zu 34,7 % missgenderten. „lighter-skinned males“ wurden hingegen mit einer Fehlerquote nur von 0,8% missgendert.

Set	n	F	M	Darker	Lighter	DF	DM	LF	LM
All Subjects	1270	44.6%	55.4%	46.4%	53.6%	21.3%	25.0%	23.3%	30.3%
Africa	661	43.9%	56.1%	86.2%	13.8%	39.8%	46.4%	4.1%	9.7%
<i>South Africa</i>	437	41.4%	58.6%	79.2%	20.8%	35.2%	43.9%	6.2%	14.6%
<i>Senegal</i>	149	43.0%	57.0%	100.0%	0.0%	43.0%	57.0%	0.0%	0.0%
<i>Rwanda</i>	75	60.0%	40.0%	100.0%	0.0%	60.0%	40.0%	0.0%	0.0%
Europe	609	45.5%	54.5%	3.1%	96.9%	1.3%	1.8%	44.2%	52.7%
<i>Sweden</i>	349	46.7%	53.3%	4.9%	95.1%	2.0%	2.9%	44.7%	50.4%
<i>Finland</i>	197	42.6%	57.4%	1.0%	99.0%	0.5%	0.5%	42.1%	56.9%
<i>Iceland</i>	63	47.6%	52.4%	0.0%	100.0%	0.0%	0.0%	47.6%	52.4%

Table 2: Pilot Parliaments Benchmark decomposition by the total number of female subjects denoted as F, total number of male subjects (M), total number of darker and lighter subjects, as well as female darker/lighter (DF/LF) and male darker/lighter subjects (DM/LM). The group compositions are shown for all unique subjects, Africa, Europe and the countries in our dataset located in each of these continents.

(Abbildung 1, Buolamwini, Gebru 2018, S. 7).

2.) Fallauswahl

Ich möchte meine Fallauswahl im Folgenden anhand dreier Punkte erläutern. Zuerst werde ich erläutern, warum ich mich für eine Studie und nicht für ein technisches Artefakt, also die Software selbst entschieden habe. Hierbei werden allgemeine forschungsethische Aspekte eine Rolle spielen. Danach werde ich die spezifischen ethischen Überlegungen für die Studie Gender Shades erläutern und schließlich noch eine methodische Begründung anschließen. Aufgrund der politischen Brisanz des Themas Gesichts- und Geschlechtserkennung habe ich mich entschieden diese Begründung ausführlich zu halten.

a.) Artefakt oder Studie

Die Frage, was Geistes- und Sozialwissenschaftler:innen untersuchen, wenn sie Algorithmen und Software untersuchen, ist methodisch gesehen eine weitreichende und viel diskutierte Frage (vgl. Vertsi, Ribes et al. 2019). Dabei sind die Herangehensweisen vielfältig, von ethnografischen Zugängen über literaturwissenschaftliche Interpretationsansätze der Softwarecodes (vgl. Marino 2020). Diese Methoden ermöglichen es, einige Aspekte von Software zu analysieren und andere in den Hintergrund treten zu lassen, wie etwa materialistische Studien von Software, die sich auf Produktionsweisen beziehen (vgl. Chun 2013). Das Analysieren einer Gesichtserkennungssoftware aus einer geisteswissenschaftlichen Perspektive kann methodisch abgesichert werden. Dennoch möchte ich in dieser Arbeit auf eine wissenschaftliche Studie zurückgreifen die kein Algorithmus ist. Was sind die Gründe hierfür?

b.) Software als Eigentum

Im Jahr 2020 stellen Gesichtserkennungssoftwaretools Firmeneigentum dar. Sie sind patentiert und gehören Firmen, wie beispielsweise IBM oder Microsoft. Das macht es, von einem Transparenzstandpunkt aus, schwer nachvollziehbar, wie diese Geschlechtserkennungsprogramme arbeiten. So sind technische Details fast nicht einzusehen. Zwar nutzen Wissenschaftler:innen diese Programme, jedoch verfügen sie nicht über die Informationen, die diesen zugrunde liegen, wie etwa Trainingsdatensets, die verifizierten Parameter, die Gesichtserkennung ermöglichen (beispielsweise welche Punkte im Gesicht wissenschaftlich relevant sind) (vgl. Keyes 2018).

c.) Die Frage des ethischen Samples

Des Weiteren werden Gesichtserkennungsprogramme oftmals durch die Sample des NIST „The National Institute of Standards and Technology“ getestet. Eine US-amerikanische Behörde, ähnlich der DIN, welche Testsets zur Verfügung stellt, an welchen die Algorithmen getestet werden können. Performen die Tools überdurchschnittlich, gibt es ähnlich der Stiftung Warentest, Auszeichnungen, die den Marktwert eines Produktes fördern. Os Keyes, Nikki Stevens und Jacqueline Wernimont zeigten in einer neueren Studie, dass Bilder aus diesen Datensätzen von Strafgefangenen oder Kindern in Kinderpornografie stammen. Eine Einwilligung dieser Menschen zur Veröffentlichung lag nicht vor (vgl. et. Al. 2019). Um diese Praktiken nicht weiter zu fördern, möchte ich auf eine Dokumentation der Auditstudie zurückgreifen, da die technischen Details mathematisch beschrieben und die verwendeten Daten von den Wissenschaftler:innen angegeben wurden.

Ich möchte ferner in meiner Untersuchung nach Funktionsweisen von Geschlecht fragen bzw. diese offenlegen und analysieren. Jedoch sind die untersuchten Genderclassifier und die zu analysierende Auditstudie auf dem reduktionistischen Bild einer Geschlechterbinarität aufgebaut. Diese Binarität ist in sämtlichen untersuchten Classifiern der Beispielstudie, aber auch in der Forschung zu Gendererkennung maßgeblich (vgl. Keyes 2018). Eine Untersuchung der Gendererkennungssoftware würde neue Testbilder und Geschlechtserkennung erfordern und somit wieder Transpersonen oder genderqueere Menschen verunsichtbaren, was an dieser Stelle nicht sinnvoll oder zielführend ist. Zumal diese Arbeit an anderer Stelle, wie bereits zitiert, geschehen ist.

Ich werde im weiteren Verlauf dieser Arbeit diese Binarität analysieren und sie in meine Überlegungen einbeziehen, jedoch werde ich sie, durch das Verweigern der Benutzung solcher Technologie, nicht verstetigen.

Ich habe mich deswegen entschieden, eine bereits bestehende Auditstudie zu Geschlechtserkennung

als Fallbeispiel auszuwählen, um den bereits benannten forschungsethischen Bedenken entgegenzuwirken und dennoch dieses, für die Politikwissenschaft wichtige, Thema fachlich aufbereiten zu können.

d.) Begründung der vorliegenden Studienwahl

Ich habe die Studie Gender Shades ausgewählt, da sie ein Datenset verwendet, das sich auf die Idee der repräsentativen Demokratie beruft. So wurden Parlamentarier:innen ausgewählt anhand einer paritätischen Verteilung von Gender und race. Die verwendeten Fotos sind offiziell und mit Einverständnis der Betroffenen, als Personen des öffentlichen und demokratischen Lebens der ausgewählten Länder, eingefügt. Die Fotos stammen aus den Parlamentsarchiven (vgl. Buolamwini, Gebru 2018)

Ferner habe ich diese Arbeit ausgewählt, da sie die drei Marktführer auf dem Gebiet der automatischen Gendererkennung behandelt und die von den Unternehmen bereitgestellten Algorithmen nutzt. Mittels dieses Projektes kann ich somit auf eine breite technische Basis verweisen und die Funktion der Gendererkennung breiter fassen, als es die Begrenzung auf ein Unternehmen erlaubt. Des Weiteren ermöglicht die auf Fairness angelegte Arbeit von Joy Buolamwini und Tinit Gebru eine Analyse von Geschlecht, die sich zuvorderst nur mit der Kategorie an sich beschäftigt und nicht mit daraus resultierenden Verwendungszwecken, wie beispielsweise Marketing oder militärische Zwecke (vgl. ebd.).

Drittens habe ich die Studie aufgrund meiner verwendeten Theorie gewählt. Der feministische Rahmen meiner Arbeit gibt mir einen Fokus auf geschlechtersensible Themen in einem intersektionalen Kontext vor. Eine zentrale Kategorie in dieser Arbeit ist Race. Aufgrund des Einsatzfelds, in dem ein von vielen Forscher:innen aufgezeigtes Racial Bias vorherrscht, habe ich mich für diese Auditstudie entschieden, da sie diese beiden Kategorien Race und Geschlecht methodisch miteinander verknüpft (vgl. Benjamin 2019). Außerdem möchte ich hier explizit auf das bereits vorhandene Wissen, rund um intersektionale Forschung zu KI, andere Wissenschaftsdisziplinen aufbauen. Da meine Theorien einen transdisziplinären Rahmen aufspannen, muss auch der Untersuchungsgegenstand dieser Arbeitsweise genügen.

Ich werde abschließen im Folgenden die Schlüsselkonzepte meiner Arbeit vorstellen: Künstliche Intelligenz und Queer.

G.) Definitionen grundlegender Begriffe

Bevor ich meinen theoretischen Ansatz für diese Studie erläutere, möchte ich zwei Begriffe, die

zentral für diese Arbeit sind, einführen: „Künstliche Intelligenz“ und „queer“.

1.) Künstliche Intelligenz (KI)

Ich untersuche die Funktion von Gender innerhalb von Genderclassiferyer und einer Evaluationsmethode dieser Technologie, dem Benchmark, bzw. dem von Buolamwini und Gebru entwickelten Daten-Set zur Kontrolle der Genderclassiferyer (vgl. Buolamwini, Gebru 2018, S. 4). Dabei beruht die Technologie der Genderclassiferyer, die eine Form der automatischen Gesichtserkennung darstellen⁴, auf neueren Ansätzen der künstlichen Intelligenz, dem maschinellen Lernen (vgl. Hamidi et. al. 2018, S.2).

Ich möchte im Folgenden deshalb einführend definieren, was künstliche Intelligenz bedeutet. Ich werde ferner erläutern, warum ich für meine Studie Genderclassiferyer als intelligente Software verstehe und weniger das Konzept Algorithmus als theoretisches Konzept verwende. Im zweiten Kapitel dieser Arbeit wird noch genauer auf die technischen Besonderheiten der, dem Gender Shades Projekt zugrunde liegenden, Genderclassiferyer eingegangen. An dieser Stelle geht es mir um eine Kontextualisierung und Darlegung meiner im weiteren Verlauf der Arbeit verwendeten Begriffe.

Künstliche Intelligenz ist ein transdisziplinäres(, universitäres Forschungsfeld, welches ein Teilgebiet der Informatik darstellt und Methodender Kognitionswissenschaft, System- und Informationstechnologie oder der Robotik vereint (vgl. Elish, boyd 2018, S.60). Die Ursprünge dieser Fachrichtung gehen auf einen Workshop des Dartmouth College 1956 zurück, dieser Workshop gilt als die erste offizielle Nennung von künstlicher Intelligenz als Forschungsgegenstand (Campalo, Crawford 2020, S.2). Künstliche Intelligenz kann dabei als Idee verstanden werden, maschinelle Intelligenz zu gestalten, die menschenähnlich handelt (vgl. Adam 1998, S.26). Diese zweite und grundsätzliche Definition, die Marvin Minsky einer der Gründungsväter der Disziplin aufstellte, ist bis heute gebräuchlich und dennoch umstritten aufgrund ihrer begrifflichen Unschärfe. In Ihrem Standardwerk zur Einführung in die künstliche Intelligenz beschreiben die Informatiker Stuart Russell und Peter Norvig deshalb 4 begriffliche Nahrungsversuche. Künstliche Intelligenz spaltet sich demnach in definitorischer Sicht in „menschliches Denken“ „menschliches Handeln“, „rationales Handeln“ „rationales Denken“ (vgl. Russell, Norvig 2016 S.2). Dennoch sind auch diese Unterscheidungsversuche nicht sonderlich trennscharf und wirken oft tautologisch, so ist eine Definition für den KI-Begriff bezogen auf rationales Handeln als Grundprämisse „AI ... is concerned with intelligent behavior in

4 Besonders im Fall des Gendershades-Projektes. Hier stellen Genderclassiferyer eine Unterart der Gesichtsanalyse dar (vgl. Buolamwini, Gebru 2018, S2-3). Genderclassiferyer tauchen in der Forschung jedoch als Verhaltensprognosesoftware auf, um beispielsweise Verhalten von Nutzer:innen auf Social-Media-Plattformen zu kategorisieren, indem Nutzer:innen automatisch von Genderclassiferyern aufgrund ihres Nutzer:innenverhaltens als männlich oder weiblich klassifiziert werden (vgl. Bives, Haimson, 2016, S.3)

artifacts“ (Nilsson, 1998 [nach Russell, Norvig 2016, S.2).

Ich konzentriere mich daher auf die Funktionsweisen von künstlicher Intelligenz.

Dabei werden alte und neue KI-Ansätze unterschieden (vgl. Adam 1998, S.28-32).

a.) Alte und neue KI

Als alte KI-Ansätze gelten sogenannte symbolische Ansätze. Hierbei werden fest definierte Probleme durch determinierte Algorithmen gelöst.

„Both GPS [General Problem Solving] and the later Soar system are based upon a highly constrained problem solving situation, with an artificial and formally defined problem domain, and with only a rather limited amount of empirical data.“ (ebd., S. 27)

Das bedeutet, KI arbeitet nach formalisierten, hierarchisierten Arbeitsschritten. Algorithmen bedeuten hier mathematische Problemlösefunktionen. Dabei gehen symbolische Ansätze von der Prämisse aus, dass sich Probleme und Lösungswege durch mathematische Ansätze formalisieren lassen. Diese Formalisierung ist die Darstellung von Problemen und Lösungswegen in Form mathematischer Symbole. Intelligenz wird in diesem Konzept zur Fähigkeit, diese Symbole zielführend zu manipulieren (vgl. Misselhorn 2018, S.19).

Neue KI-Ansätze fallen unter Konzepte des maschinellen Lernens und des Deep Learning. Dabei bezieht sich die Neuartigkeit dieser Form von Künstlicher Intelligenz nicht auf die Entwicklung der Methode, die in etwa zeitgleich zu den alten KI-Ansätzen auftrat, sondern in ihrer zunehmenden Verwendung ab den frühen 00er Jahren (vgl. Sejnowski, 2018, S. 2; 32-34). Diese beruhen auf der Prämisse, maschinelle Intelligenz solle sich nicht an Symbolverarbeitung abarbeiten, sondern sich am menschlichen Gehirn auf einer neuronalen Ebene, in etwa der Übertragungsebene von Informationen im menschlichen Gehirn, orientieren. Vereinfacht ausgedrückt orientierten sich symbolverarbeitende Ansätze stärker an der Repräsentation von zuvor menschlich definierten Inhalten, die neueren Ansätze orientieren sich nicht an konkreten Inhalten sondern an Fragen der Verarbeitung von Information. Diese Funktionsäquivalenz wird durch künstliche neuronale Netzwerke hergestellt, welche an menschliche Neuronen erinnern sollen⁵ (vgl. Elish, boyd 2017, S.61).

Das Ziel dieser Ansätze ist aus heutiger Sicht die Lösung sogenannter „real world Probleme“. Diese sind weniger mathematisch definiert, sondern angelehnt an nicht formalisierte Probleme, wie beispielsweise die eingangs beschriebene Gesichtserkennung an Flughäfen (vgl. Campolo, Crawford 2020, S. 3). Neue Ansätze von KI beruhen dabei immernoch auf formalisierten und hierarchisierten

5 Diese Überlegung künstlicher Neuronen gab es schon während der Fünfziger Jahre wurde jedoch zugunsten der Idee der symbolverarbeitenden Ansätze vernachlässigt und erst in den 80iger Jahren wieder prominent (vgl. Sejnowski, 2018, S.32-34)

Problemlösestrategien, jedoch sind diese Ansätze darauf ausgelegt, Muster durch Abstraktion aufzuspüren, ohne ein vorher menschlich definiertes Konzept dieser Muster in ihren Lösungsalgorithmen implementiert zu haben.

Der Kommunikationswissenschaftler Andreas Sudmann fasst das maschinelle Lernen, welches diesem KI-Ansatz zugrunde liegt, dabei wie folgt zusammen.

„Der maschinelle Lernprozesses läuft dabei typischerweise wie folgt ab: Gegeben ist ein Input (x) sowie ein Output (y). Ziel ist es, für irgendeinen Input x (z.B. Bilder bzw. Pixels) den dazu passenden Output y (z.B. den Inhalt eines Bildes) möglichst genau vorherzusagen. Dafür wird das maschinelle Lernsystem im Rahmen einer sogenannten Lernphase auf Grundlage riesiger Mengen von Beispieldaten (anhand gelernter Übereinstimmungen von x, y) so lange trainiert, bis das System, auf der Basis der Lernerfahrung aus dem Training, in der Lage ist, auch neue, unbekannte Inputdaten zutreffend zu generalisieren.“ (vgl. Sudmann 2018, S.10)

Dabei lassen sich jedoch zwei Arten des Lernens unterscheiden, das überwachte, wie in dem Zitat dargestellte, und das unüberwachte. Das unüberwachte Lernen betrifft demgegenüber Inputdaten, die keinen festgeschriebenen Output haben. Dies bezieht sich auf sogenannte „unstrukturierte“ Daten, die vorher beispielsweise nicht kategorisiert worden sind. Am Beispiel des Genderclassifiers würde dies beispielsweise bedeuten, dass die eingepflegten Bilder anhand derer diese Anwendungen „trainieren“ soll nicht mit dem Label Mann/Frau versehen sind.

Die Soziolog:innen Kate Crawford und Alexander Campolo machen dabei auf die Besonderheit dieser Ansätze aufmerksam, in dem sie den Output dieser Anwendungen als empirisch verwertbar definieren und fassen diese Definition wie folgt zusammen:

„The network’s outputs can be tested in real-world contexts, usually in the form of classifications or predictions, such as deciding whether an email should be classified as spam. This “empirical” orientation differentiates deep learning from predecessors that relied on an initial, a priori specification of a task“ (Crawford, Campalo, 2020, S.6).

Diese Ausrichtung ist auch innerhalb des Gender Shades Projektes festzustellen, so beschrieben Gebru und Buolamwini die Genderclassifier einerseits als kommerzielle, also bezahlpflichtige Dienste der jeweiligen Unternehmen und nehmen innerhalb ihres Problemaufrisses und der Relevanzbegründung der Studie auf die Anwendungen außerhalb universitärer Kontexte am Beispiel des Strafvollzugsystems Bezug (vgl. Buolamwini, Gebru 2018, S.1-2).

Dabei beruhen diese Ansätze des maschinellen Lernens, wie beschrieben, auf großen Mengen an Daten, um in „der echten Welt“ angewendet zu werden (vgl. Crawford, Campalo, 2020, S.6).

Künstliche Intelligenz, verstanden als maschinelles Lernen, verweist dabei auf einen weiteren wichtigen Begriff, nämlich auf „Big Data“ (vgl. Alpaydin, 2016, S.17-18).

b.) Big Data und/oder KI

Big Data ist ein weiter Begriff, der, wie KI, wenig trennscharf ist. Ich möchte ihn hier wie folgt verstehen als die Entwicklung eines großen exponentiellen Anwachsens von digitalen Daten, durch die Massenverbreitung von Hardware in Form von Computern und internetfähigen Smartphones, so wie immer neuen, datengenerierenden Anwendungen, wie Social-Media-Plattformen (vgl. ebd. S.9-11). Dabei sind diese digital generierten Daten durch „3 Vs“ - „volume, velocity and variety“ - gekennzeichnet (Elish, boyd 2018, S.59). Das meint, dass diese Art von Daten nicht nur durch ihr Volumen ein Novum innerhalb der Informationstechnologie darstellen, sondern auch durch ihre Qualität. Daten können heute, in Form sozialer Daten, komplette Biografien von jeder, Social-Media nutzenden Person abbilden und aufgrund ihrer digitalen Form mit anderen Daten verglichen und vernetzt werden, und sie sind multimedial. Das bedeutet, wir haben es mit einer Vielzahl an verschiedenen Datenmodi zu tun, Bilddateien, Videos, Text oder Sprachaufzeichnungen. Gleichzeitig können sie in Echtzeit oder mit minimaler Verzögerung erstellt und durch entsprechende Computerprogramme verarbeitet werden (vgl. Alpaydin, 2016 S. 15, Kitchin 2014, S. 2).

Big Data ist dabei ein in den frühen 2000ern geschaffener Neologismus, welcher vorrangig auf eine wirtschaftliche Nutzung dieser neu geschaffenen Daten referiert, in Form von Aufbewahrungsdiensten oder Analyse- und Vorhersageangeboten (vgl. van Dijk 2014, S.199).

Big Data ist nach meiner Definition damit eine Grundvoraussetzung für das maschinelle Lernen, welches die von mir zu untersuchende Technologie ist.

Ich werde in meiner Arbeit daher beide Begriffe synonym verwenden.

Dennoch sind Big Data und künstliche Intelligenz nicht dasselbe, sondern beziehen sich auf unterschiedliche wissenschaftliche und ökonomische Ursprünge (vgl. Elish, boyd 2017 S. 58).

c.) Algorithmen/Software oder worüber wir sprechen

Ein vielgenutzter Begriff innerhalb des Diskurses um KI, sowohl aus der informationstheoretischen als auch aus der sozialwissenschaftlichen Perspektive, ist dabei der Algorithmus (vgl. Amoore, Piotukh et al. 2015, Galloway 2006, Kitchin 2016, Kushner 2013, Striphos 2015,). So wird der Algorithmus in Studien zu KI oftmals als Synonym gebraucht (vgl. Jatón im Erscheinen, Kitchin 2017, S. 14-15)

Algorithmus ist aus mathematischer Sicht eine schrittweise, hierarchische Lösungsoperation, nach

dem Muster „Wenn X, dann Y“. Mit Bezug auf den Computer als Umwelt lässt sich dieser wie folgt definieren „An algorithm was understood to be a set of defined steps that followed the correct order will computally process input (instructions and/ or data) to produce a desired outcome“ (Kitchin 2017, [nach Bucher 2019], S.21).

Jedoch verweist der heutige (sozial) wissenschaftliche Diskurs über Algorithmen als Bestandteil von KI, auf weit mehr als nur eine mathematische Rechenoperation. Algorithmen sind zur Gesellschaftsdiagnose geworden (vgl. Kushner 2013, S.1243)

Sie beruhen dabei, nach der inzwischen einschlägig geworden Definition des Philosophen Andrew Goffey, weniger auf konkreten Ausformungen, etwa durch Programmiersprachen, sondern sie sind Allgemeinplätze der Informationsordnung (vgl. Goffey 2008, S. 15). Hierbei ist es unerheblich, ob beispielsweise Musterkennungsprogramme in der Sprache Python oder Pearl programmiert sind. Der Algorithmus, also die Problemlösungsstrategie dahinter, kann insgesamt auch in beiden Sprachen formuliert sein, der Algorithmus entspricht dabei einem transformierbaren Inhalt.

„An algorithm is an abstraction, having an autonomous existence independent of what computer scientists like to refer to as ‘implementation details,’ that is, its embodiment in a particular programming language for a particular machine architecture“ (ebd.).

Dabei verweisen Sozialwissenschaftler:innen immer wieder auf den paradigmatischen Charakter, den Algorithmen als theoretisches Konzept dadurch erhalten. So steht der Algorithmus immer auch für einen analytischen Allgemeinplatz sozialwissenschaftlicher Theorien über Computertechnologie, Künstliche Intelligenz, Big Data, neue Formen der Arbeit und eine Vielzahl weiterer Themen (vgl. Ames 2018, S. 1-2).

Ich möchte deshalb für meine Arbeit den Begriff Software, statt Algorithmus, zur Beschreibung von KI nutzen, um auf die spezifischen, historischen und materiellen Kontexte meines Forschungsgegenstands eingehen zu können. Ich werde deshalb von intelligenter Software schreiben und nur für die Vermittlung technischer Referenzen Algorithmus als Begriff nutzen.

Dabei beziehe ich mich auf das Konzept der Medienwissenschaftlerin Wendy Chun von Software als Ding.

„Software as a thing reconceptualized society, bodies and memories in ways that both compromise and extend the subject, the user. Importantly, software as a thing cannot be reduced to software as a commodity: software as „thing“ is a return to older definitions of thing as a „gathering“, as pertaining to anything related to „man““ (Chun 2009, S.6)

Software, so Chun, verweist dabei auf eine vergeschlechtlichte Ursprungsgeschichte. Zur Anfangszeit des Digitalcomputers war Software eine Bezeichnung für eine Arbeitsleistung, die vorrangig

von FLINTA Personen ausgeführt wurde (vgl. ebd. S. 5). Diese Leistung wurde in Lohnarbeitszeit verrechnet, pro Rechenoperation wurde, zumeist in militärischen und universitären Einrichtungen in den USA und England, bezahlt (vgl. ebd. 29). Software wandelte sich im Laufe der Jahrzehnte und ist heute eine automatisierte Funktion um Rechenoperationen auszuführen, die das Funktionieren des Computers und darauf basierend des Internets ermöglichen (vgl. ebd. S.33). Software fällt dabei eine Vermittlungsfunktion zu zwischen Benutzer:in und Netzwerk (zumeist das Internet). Dabei sind einzelne Computeruser:innen weniger auf ihr individuelles Wissen hinsichtlich der konkreten Funktion des jeweiligen Endgerätes und der genutzten App angewiesen, sondern auf die Software, die die dahinter stehenden Vernetzungen ermöglicht (vgl. ebd. S.18)

Software wirkt dabei performativ. Erst durch das Benutzen eben jener Anwendung entstehen Nutzer:innen und Programmierer:innen. Software als Ding verweist dabei auf den nie abgeschlossenen Charakter der Software (vgl. ebd.)

Aufgrund dieser materialistischen und performativen Ebene erlaubt Software, als Definitionsbasis für automatische Gesichtserkennungsanwendungen, eine Analyse der verborgenen Arbeitsteilung und der Hierarchien, die durch benutzen und benutzt werden von und durch diese Technologien, entstehen. Diese theoretische Perspektive ermöglicht es mir, im weiteren Verlauf der Arbeit die gewählten Theorien, die ich im anschließenden Theoriekapitel vorstelle, auf meinen Untersuchungsgegenstand anzuwenden. Algorithmus als Begriff ist für meine theoretische Arbeit zu allgemein und zu abstrakt, als dass das Konzept die spezifischen technischen und sozialen Wirkweisen einbeziehen könnte, da ich mich auf einen empirischen Fall konzentriere und eben nicht auf eine allgemeine Analyse des Prinzips Genderclassifier.

d.) KI und die Sozialwissenschaften

Innerhalb der letzten Jahren etablierte sich eine sozialwissenschaftliche Begleitforschung zu diesen neuen Ansätzen der künstlichen Intelligenz, die sich vermehrt mit der kritischen Auseinandersetzung zum Ist-Stand des maschinellen Lernens befasst und mit der Frage, ob künstliche Intelligenz überhaupt ein angebrachter Begriff ist für die eben vorgestellten Techniken oder nicht zu einer Verschleierung von Machtstrukturen und der eigentlichen Funktionsweise dieser Technologien führt (vgl. Wajcman 2019, Elish, boyd 2017). Ich möchte diese Diskussion nicht führen, sondern mich an das Argument der Informationstheoretiker:in Alison Adam halten. Die Wissenschaftlerin erläuterte bereits 1999, dass Kritik an dem Begriff der künstlichen Intelligenz und dessen implizite Versprechungen menschenähnlicher Intelligenz und tatsächlicher Praxis berechtigt ist, jedoch künstliche Intelligenz bereits eingesetzt werde und diese untersucht werden müsse.

„Shifting the focus from the philosophical to the cultural, in this way, means that it does become important to ask how AI systems are used to represent knowledge, what kind of knowledge and whose knowledge they contain.“ (Adam 1999, S. 407)

Ausgehend von dieser Argumentation und dem momentan fortschreitenden Gebrauch künstlicher Intelligenz innerhalb jetziger Gesellschaften auf einer globalen Ebene, konzentriere ich mich in dieser Arbeit auf die Einsatzorte und geplanten Anwendungen von Genderclassifyern und weniger auf die kritische Analyse der Lücke zwischen Praxis und zukünftiger Verwendung bzw. der Fundamentalkritik an dem Konzept von KI. Dabei möchte ich Adams Argument weiter spinnen und herausarbeiten, warum es nicht mehr ein Wechsel von einer philosophischen zur kulturellen Analyse ist, sondern zu einer politischen.

Deshalb werde ich Genderclassifier als existierende und funktionierende Technologie behandeln und mich weniger auf die kritische Auseinandersetzung der zukünftigen Nutzbarkeit fokussieren.

Genderclassifier werden bereits jetzt kommerziell angeboten und erprobt (vgl. Raji, Buolamwini 2019). Sie bieten dabei Raum für Spekulationen über die zukünftigen Anwendungen, da sie in einem frühen Stadium der Entwicklung sind (Keyes 2018, S.12-13). Dennoch werden sie bereits heute getestet und an Menschen ausprobiert (vgl. Hamidi et. al. 2018, S.2-3). Deshalb werde ich Genderclassifier als existierende und funktionierende Technologie behandeln und mich weniger auf die kritische Auseinandersetzung der zukünftigen Nutzbarkeit fokussieren.

Zusammenfassend referiere ich, wenn ich von Gendererkennung schreibe, auf das Konzept der Software und nicht das des Algorithmus, dabei ausgehend von sozialwissenschaftlichen Theorien. Im Verlaufe dieser Arbeit werde ich auch von Algorithmen schreiben, da dieses Konzept auch innerhalb der Definitionen von Gesichtserkennungssoftware und Genderclassifier auftaucht. Dies geschieht jedoch als Referenzangabe und entspricht nicht meiner Definition. Ferner werde ich KI und Big Data synonym verwenden, da beide Konzepte in meiner Studie auf automatische Gesichtserkennung und maschinelles Lernen referieren.

2.) Queer

Ein zweites wichtiges Konzept, welches für meine Arbeit unerlässlich ist, ist der Terminus „queer“. Im Laufe der Arbeit werde ich mich nämlich auf das Konzept Genderqueer beziehen. Ich möchte dieses deshalb kurz erläutern. Dabei handelt es sich um eine Arbeitsdefinition, die eine Inklusion von Transmenschen und Menschen die sich weder dem männlichen noch dem weiblichen Geschlecht zuordnen. Queer ist ein sehr breites theoretisches Konzept, welches sich aus politischen Kämpfen entwickelt hat. Queer ist eine Bezeichnung innerhalb von LGBTQIA-Gemeinschaften, welche sich auch innerhalb des akademischen Diskurses rund um Judith Butlers Werk „Das Unbe-

hangen der Geschlechter“ entwickelt hat (vgl. Muñoz 2009, S.19-23, Funk 2018, S.93). Ich möchte nicht alle Facetten oder Kritiken an diesem Begriff hier ausführen, sondern lediglich eine begriffliche Klärung vornehmen, die sich für meine Studie anbietet

Ich orientiere mich hierbei, bezogen auf meinen Forschungsgegenstand, auf Queer Software Studies, ein akademisches Feld, welches sich mit der Verbindung von Software und Gender auseinandersetzt. Nach den Theoretiker:innen Daniel G. Cockayne und Lizzie Richardson erfüllen Ansätze von Queer Software Studies dabei bestimmte Funktionen.

„Queer approaches are invested in conceptualising and (therefore) challenging both social and digital code(s) – or the norm – to show how they constrain normativity but also how forms of intimate life can transgress, disrupt, and distribute what is normal.“ (Cockayne, Richardson 2017 a, S. 1643)

Queer verweist dabei auf das Konzept der Gendernorm, welches ich in der theoretischen Konzeptualisierung aufbereiten werde. Das Konzept von queer bezieht sich, nach den Autor:innen, einerseits auf eine Ebene der sexuellen Orientierung (vgl. ebd.). Andererseits ermöglicht der theoretische Begriff von Queerness, die Auseinandersetzung mit cisnormativen und binären Vorstellungen also angeborener, unveränderlicher und binär ausgestellter Geschlechtskategorien (vgl. Cockayne, Richardson 2017 b, S.1587). Damit ermöglicht die Referenz auf Queer Software Studies die Einbeziehung von alternativen Genderidentitäten als kritischen Ausgangspunkt der Analyse von Software, die nicht nur in eine Studie über Cismänner und -frauen münden müssen. Gleichzeitig verweist queer auf die normative Wirkweise von Gender auch innerhalb der Nutzung und Entstehung von Software, die ich spezifisch für das Gender Shades Projekt untersuche.

Nachdem ich diese beiden wichtigen Begriffe eingeführt habe werde ich zuerst meine Fallauswahl vorstellen und begründen und danach meinen Theoretischen Rahmen vortellen mit welchen ich meine Forschungsfrage beantworten werde.

II.) Theorieteil

Ich werde in diesem Theorieteil zuerst mein grundlegendes Verständnis zum Zusammenhang von Technologie und Gesellschaftstheorie darlegen, da es die theoretische Grundlage meiner spezifischen Überlegungen zum theoretischen Verhältnis von Technologie und Geschlecht ist. Dieses werde ich in einem zweiten Schritt genauer konzeptualisieren.

A.) Einordnung der Theoriematrix

1.) Posthumanismus als Zeitdiagnose

Die Frage der Verbindung von Macht und Technik ist prinzipiell keine sonderlich neue Frage. Schon bei Karl Marx lassen sich Überlegungen hinsichtlich der sozialen und politischen Rolle der Technologie für kapitalistische Gesellschaften finden (vgl. Meyer 2018, S.28; Castoriades 1990, S.584-587).

Theoretische Ansätze, die sich mit Fragen von Sexismus und Rassismus in Bezug auf Technologien beschäftigen sind seit den 70er Jahren präsent im wissenschaftlichen Diskurs, etwa in der Arbeit von Shulamith Firestone zum emanzipatorischen Potential reproduktiver Techniken wie der Empfängnisverhütung oder in den Beiträgen des Combahee River Collective, einer Gruppe Schwarzer lesbischer Theoretiker:innen und Aktivist:innen, mit ihren Werken zur Frage der Zugänge für Schwarze Frauen zur Gesundheitsversorgung (vgl. Firestone 2015 [1970]; Combahee River Collective 1997).

Dabei gleichen viele damals vorgetragene Ideen zur Technikgestaltung jenseits patriarchaler und sexistischer Unterdrückung aktuellen Debatten um Fragen von KI und Fairness, wie ich sie im Problemaufriss dargestellt habe. So schreibt die Soziolog:in Judy Wrayman bezüglich früherer feministischer Auseinandersetzungen mit Technik und Gender etwa „It [feminist writing] was based on an empiricist view of science and technology as fundamentally (gender) neutral. Sexism and androcentrism were understood as social biases capable of correction by stricter adherence to the methodological norms of scientific inquiry. The problem was framed in terms of uses and abuses to which science and technology has been put by men.“ (Wrayman 2004, S.16-17).

Diese strikte Binarität der Geschlechter ist in heutigen, kritisch feministischen Debatten um KI jedoch weniger stark ausgeprägt und wird zumeist mit einem Verweis auf Gender-Pluralität versehen (vgl. Keyes 2018).

Dennoch unterscheiden sich die von mir skizzierte Problemlage und die historischen Auseinandersetzung von Gender und Technik, und zwar in Form einer paradigmatischen gesellschaftstheoretischen⁶ Verschiebung. Ich beziehe mich hierbei auf die Idee des posthumanen Zeitalters (vgl. Hayles 1999). Anders als der von mir angeführte algorithmic turn, wie ihn Alexander Galloway beschrieb (vgl. Galloway 2004), handelt es sich bei der Idee, dass Gesellschaften im 21. Jahrhundert posthuman seien, nicht nur um die Frage, wie diese mittels neuer (algorithmischer) Technik funktionieren bzw. wie sich ihre Funktionsweise der technischen

⁶ Ich schreibe hier bewusst von gesellschaftswissenschaftlicher Verschiebung, da sich die folgenden zitierten Theorien nicht auf eine Fachdisziplin beschränken und gerade die transdisziplinäre Verortung als Prämisse posthumanistischer Theoretiker:innen gilt (vgl. Loh 2019, S. 150-155; Braidotti 2019, S. 150).

Funktionsweise smarter Technologie annähert. Viel grundsätzlicher beschreibt die Idee des Posthumanen eine neue Art des gesellschaftlichen und der Subjektposition (vgl. Braidotti 2014, S.8-9). Mittels synthetischer Biologie⁷, Gentechnologie und anderen Lebenswissenschaften werden Organismen im Labor geschaffen. Die Naturhaftigkeit des Menschen wird mittels immer ausgefeilterer Möglichkeiten zur Genveränderung brüchiger denn je. Die Patentierung der DNA lässt Fragen der Vermarktbarkeit des menschlichen Körpers virulent werden und omnipräsente Computertechnologie in Form tragbarer Smartphones und Heimcomputer stellen dabei eine materielle Infrastruktur dar (vgl. Braidotti 2019; Thacker 2004; Cooper 2008; Waldby, Mitchell 2006). Dabei ist diese gesellschaftswissenschaftliche Gegenwartsdiagnose rasant zu einer Prämisse kritischer wissenschaftlicher Auseinandersetzung mit Gesellschaft und Technik des 21. Jahrhunderts aufgestiegen (vgl. Braidotti 2014, S.11). So schrieb das Künstler:innen Kollektiv Critical Art Ensemble (CAE) 1998 noch von einem potentiellen posthumanen Zeitalter, abgeleitet aus dem scheinbaren Heraufziehen virtuellen Bewusstsein⁸ (vgl. CAE 1998, S. 22). Bereits ein Jahr später jedoch erklärte die Literaturwissenschaftlerin Cathrin Hayles das Posthumane zur theoretischen Beschreibung des neuen, nahenden Jahrtausend. Diese Zeit sei von der Untrennbarkeit der Menschen und ihrer Hochtechnologien, wie DNA-Sequenzierung gekennzeichnet (vgl. Hayles 1999, S.37-40). Diese Gesellschaftsanalyse referiert dabei auf eine posthumanistische Theorie die das neue Subjekt und seine Umwelt/Gesellschaft versucht zu analysieren.

Eine; für die politikwissenschaftliche Analyse moderner KI-Systeme, wichtige Prämisse ist dabei die Idee der politischen ökonomischen Bedingungen für die posthumane Gesellschaft. Wissenschaftler:innen, wie Donna Haraway oder die bereits zitierte Judy Wajcman, verweisen in ihrer Analyse posthumaner Gesellschaften auf die kapitalistischen Bedingungen von Technologie. Technik, so die Autor:innen, ist dabei Ausdruck steigender Bedürfnisse, nicht der einzelnen Menschen, sondern eines globalen kapitalistischen Systems (vgl. Wajcman 2004, S.39, Haraway 1995a, S.54-59).

KI als Technologie nimmt innerhalb posthumaner Gesellschaften eine besondere Funktion(en) ein. Nach Hayles ist KI Ausdruck einer tiefen Verunsicherung darüber, was Menschsein bedeuten kann. Das aufklärerische Ideal eines vernünftig handelnden Subjektes, welches konstitutiv für den demokratischen Nationalstaat ist, wird imitiert durch eine Software, die in Sekundenbruchteilen mensch-

⁷ Synthetische Biologie ist ein Teilbereich der Biologie. Zu diesen zählen beispielsweise Computergenerierter Biologische Modulation oder die Bioinformatik. "We are now no longer in the "wet lab" but in a "dry lab," doing "biology in silico." And, we have moved from the field of biochips to the field of bioinformatics. If the encoding process carried patterns of relationships across material substrates, then the receding process will extend the functionality of that encoded or translated data in ways that are specific to its medium." (vgl. Thacker 2004, S.19)

⁸ einer Technologie, die das Hochladen des menschlichen Geistes in das Internet oder eines beliebigen digitalen Speichermediums ermöglichen soll (vgl. Kurzweil 2005, S. 138-140).

liches Verhalten analysiert und bewertet (vgl. Hayles 1999, S.286-288). Des Weiteren wird auch die Frage eines kapitalistisch, menschlichen Subjektes in Frage gestellt; denn in digitalen Zeiten sind keinesfalls Menschen die einzigen Teilnehmer:innen an den Märkten, dazu gehören auch automatisierter Warenverkehr oder der Zugang zu bestimmten Märkten durch Authentifizierungsanwendungen und Datenschutzvereinbarung und ähnliche Anwendungen (vgl. Braidotti 2014, S.94-97)

Dabei fasst Braidotti diese spezielle Stellung des Menschen innerhalb dieser technologisierten Märkte wie folgt:

„In any Case, having to demonstrate one's humanity in order to access goods and service seems to be the interperatave(???) of a „new“ economy, centred on the algorithmic culture of computational networks, not good old Man/Homo/Anthrophos - the human. In our information age the boundaries between anthropomorphic humans and quasi-human technological substitutes have been radically displaced“ (Braidotti 2019, S. 14).

An dieser Stelle möchte ich hinzufügen, dass die Entstehung der menschenähnlichen Technologie gleichfalls eine Grenzverschiebung markiert, die neuer theoretischer Konzepte bedarf, um überhaupt gesellschaftlich und politisch bearbeitbar zu werden.

So verweist Hayles, daran anschließend, auf die Frage der kritischen Auseinandersetzung mit KI. Nach der Wissenschaftler:in ist die Frage, wer im Zeitalter intelligenter Maschinen wie handeln kann, drängend. Dennoch ist die Rückkehr zu einem menschlich autonomen Wesen mit Handlungsmacht, der:die dem technischen Versagen oder etwa der Krise der KI entgegnet, nur ein theoretisches Auflebenlassen patriarchaler Machtphantasien, die den Mann als autonomen Heilsbringer und Herrscher über die Maschine instituiert (vgl. Hayles 1999, S. 238-239). Die Medienkünstlerin Beth Coleman zeigt in ihrer Analyse von Race und Wissenschaft ferner die Einschreibung eines weißen kolonialen Denkens, welches sich hinter der Idee eines menschlichen Beherrschens von Maschinen verbirgt. Die Vorstellung von Herrschaft aufgrund technischer bzw. kognitiver Überlegenheit ist eine Grundvoraussetzung für die Deprivilegierung und Unterdrückung von BIPoCs und die Kolonisation durch die westliche Gesellschaft. Ausgehend von einem aufklärerischen Gedanken des 18. Jahrhunderts, welcher weiße Männer als intellektuell überlegene gesellschaftliche Gruppe einführte, sei der Gedanke der Beherrschung der Maschine aufgrund rationaler Überlegenheit keine wissenschaftliche Möglichkeit der Kritik an KI (vgl. Coleman 2009, S.190). Wie also theoretisch arbeiten mit KI?

2.) Technowissenschaft als Paradigma der KI Forschung

Ich möchte den Ansatz der Technowissenschaft hier als eine mögliche Antwort vorstellen und als analytische Grundlage für meine theoretische Arbeit einführen.

Die technowissenschaftliche Idee beruht auf der vorgestellten gesellschaftstheoretischen Prämisse, dass sich die Menschheit in einem posthumanen Zeitalter befindet. Dabei ist Technologie eine

konstitutive Voraussetzung für global vernetzte kapitalistische Gesellschaften. Zentral ist dabei Donna Haraways berühmte Figur des Cyborgs als Protosubjekt dieser Zeit. Der Cyborg, wie ihn Haraway versteht, ist dabei eine Referenz für das Verschmelzen und Kolabieren der Grenzen zwischen menschlich/animalisch und maschinell (vgl. Haraway 1995a, S.36-39) .

Doch hinter diesen figurativen Ideen steht auch ein Verständnis der neuen Wissenschaft, der Techno-Science. Das beschreibt eine Art der Technikgestaltung, die sich durch Laborpraxen, kapitalintensive Forschung und Herstellung und die starke Anbindung an die Universität als Ort des Designs auszeichnet. Technowissenschaften grenzen sich somit von sozial anthropologischen Überlegungen ab, die den Menschen prinzipiell als Technik nutzendes Wesen sehen, indem sie auf die spezifische Produktion der Technik im posthumane Zeitalter hinweisen (vgl. Coleman 2009, S.199-200; Reichle 2009, S.11). Dieses Wissenschaftsverständnis beruht dabei auf vier Grundprinzipen: (a) Technik ist relational, (b) Technik ist performativ, (c) Technik ist posthuman und (d) Technik ist politisch.

Durch die Co-Konstitution von Gesellschaft und Technik durch sich selbst, sind beide als relationale Strukturen zu verstehen, die sich nur durch die Einbeziehung des jeweils anderen Bereichs erklären lassen. Anders formuliert, Gesellschaft lässt sich nur durch ihre Technologie verstehen und Technologie nur durch die Gesellschaft in der sie angewendet wird. (Wajcman 2004, S,39)

Diese Verbindung ist darauf aufbauend auch immer performativ, sie wird immer gemacht. Sie ist kein Fakt welcher sich unabhängig von dem jeweiligen Kontext erklärt. Hayles weist in diesem Zusammenhang auf ?auf die enge Beziehung zwischen der Performativität von Software und Gender hin. Denn Software, so die Wissenschaftler:in wird wie Gender immer wieder performt (vgl. Hayles 1999, S.274).

Drittens beruht diese posthumane Technologie immer auf das Zusammenwirken menschlicher und nicht menschlicher Akteur:innen, gerade in Bezug auf KI wird dies deutlich. Denn das Ziel von KI ist es ja gerade, eine Anwendung zu schaffen, die ohne menschliche Hilfestellung in verschiedenen Bereichen handeln kann (Russel, Norvig 2016). Ich werde diese Prämisse im ersten Kapitel weiter ausformulieren.

Viertens ergibt sich aus diesen drei Prämissen, technowissenschaftlich arbeiten verweist auf den irreduziblen Kern des Politischen innerhalb von Technologie. Dabei folge ich der Idee des Politischen nach Lee Edelmans queerem Verständnis von Politik als Sichtbarmachung des „Außerhalb“: Außerhalb der normativen Ordnung, außerhalb eines binären Geschlechterverständnisses. Politik beginnt somit immer in einem Sichtbarwerden der Grenzen des Diskurses und lässt Optionen erscheinen (vgl. Edelman 2004, S. 3-4). Ich möchte mit der

theoretischen Konzeptualisierung von Geschlecht/Technologie auf die Möglichkeit von Handlung und Veränderbarkeit hinweisen, die eine politikwissenschaftliche Analyse erst erlauben. Denn würde es sich bei diesem Verhältnis um einen Determinismus handeln, könnten politikwissenschaftliche Analysen nicht greifen. Dafür muss ich jedoch erläutern, inwiefern das Verhältnis von Geschlecht/Technik kontingent und somit gesellschaftstheoretisch verhandelbar ist. Das dahinterliegende Argument lautet wie folgt, wenn Technologie performativ ist, bedeutet das immer, dass Technik nicht einfach in Form fertiger Artefakte und alternativloser Innovationsprozesse geschieht, sondern in komplexen sozialen Settings gegründet ist, die sowohl menschliche als auch nichtmenschliche Akteur:innen beinhalten (vgl. Haraway 1995a, S. 48-49). Dabei ist diese Verantwortung nicht als individualistische Herausforderung zu verstehen, der sich nur einzelne Forscher:innen stellen sollten, vielmehr handelt es sich um Prozesse der verteilten und kollektiven Verantwortung, die sich mit gesellschaftlichen Machtstrukturen auseinandersetzt, indem sie Technikgestaltung und technowissenschaftliche Analyse als situierte Praxis begreift. Eine Art der Wissenschaft, die auf ein partielles Wissen setzt, das sich durch seine Rückbestimmung und Offenlegung der Verortung des:der Wissenschaftler:in widerspiegelt. Gender sowie Race sind Begrenzungen und Orte dieser theoretischen Grenzziehungspraxen. Gender so Haraway „ist ein Feld strukturierter und strukturierender Differenz“ (Haraway 1995b, S.88). Diese Differenz beinhaltet immer Macht, da sie ein Innen und Außen markiert, eine Grenze zieht und somit, um wiederum auf Edelman zurückzukommen, politisch ist. Gender wird somit zu einem wichtigen Analyse-Tool und lässt mich überleiten zu meiner Konzeptionalisierung von Gender für die anstehende Analyse und Beantwortung der Forschungsfrage.

Davor möchte ich noch einmal zusammenfassen, welches grundlegende Verhältnis ich für meine Arbeit hinsichtlich KI und Gesellschaft verwende und auf welche theoretischen Grundlagen ich mich hierbei stütze.

Bis jetzt habe ich eine theoretische Basis geschaffen, welche das Verhältnis von Technik und Gesellschaft(stheorie) erläutert. KI ist theoretisch in meiner Arbeit eine relationale, posthumane, politische Kategorie der Techno-Science, die sich aus einer posthumanen Zeitdiagnose ableiten lässt. Ich habe somit das Verhältnis von Technologie und Gesellschaft spezifiziert in Anbetracht der weitreichenden Tradition sozialwissenschaftlicher Forschung zu dieser Beziehung. Ich habe ferner erläutert, weshalb KI in der spezifisch technowissenschaftlichen Sichtweise eine besondere Stellung einnimmt.

Im Folgenden werde ich genauer auf die Verbindung von Macht und Grenzziehung durch und mittels der Kategorie Gender eingehen, die Haraway beschreibt (vgl. ebd.).

Davor möchte ich jedoch diese Grenzziehung als spezifische Methode näher erläutern. Ich beziehe

mich auf das Konzept der „Diffraction“ der Physiker:in Karen Barad, welches sich wiederum auf Haraways Überlegungen zu Interferenz bezieht (vgl. Barad 2013, S. 57).

Barad baut ihr Model der Diffraktion als Methode auf dem Widerspruch zu herkömmlichen Ideen der Science and Technology Studies (STS) auf. Diese, so die Wissenschaftler:in, beruhen ihrerseits auf sozialwissenschaftlichen Theorien der Reflexion und Repräsentation, als Mittel der kritischen Analyse von Gesellschaft und Technik (vgl. Barad 2007, S. 88).

Diese klassischen STS-Ansätze wiesen jedoch zwei fundamentale Schwächen auf, die einer feministischen posthumanistischen Analyse entgegenstünden.

Einerseits würden klassische STS zwar von der Prämisse ausgehen, dass Technologie eine performative Praxis ist, wie ich sie bereits als Prämisse für die technowissenschaftliche Interpretation von Technologie eingeführt habe. Dabei würden es diese Ansätze jedoch vermeiden, soziale (Ungleichheits-)Faktoren mit in diesem Prozess einzubeziehen. Gender und Race, so Barad, würden so als vorgängige Entitäten verstanden, die nicht in den Entstehungsprozess von Technologie mit einfließen (vgl. ebd. 87). Nach Barad ist dies eine logische Ableitung aus der zweiten methodischen Problematik der Repräsentation. Nach Barad verweisen die STS-Ansätze auf eine kritische sozialwissenschaftliche Reflexion technischer Artefakte, diese jedoch würden lediglich soziale Verhältnisse in Technologie repräsentiert sehen. Dabei hätte die Repräsentation an sich keine Auswirkungen auf die Objekte, die sie wiedergibt. Technologie wäre somit bloß ein Spiegel sozialer Ungleichheit und nicht selbst ein spezifischer Ort dieser, der eigene Herrschaftsverhältnisse produziert (vgl. ebd.). Nach Barad verkennt diese theoretische Perspektive jedoch in letzter Konsequenz ihre performativen sozialkonstruktivistische Prämissen. Für Barad muss sich eine Gesellschaftstheorie vielmehr einer Methodeder Interferenz zuwenden, um diesem Ansatz gerecht zu werden (Meißner 2013, S. 180). Wobei Interferenz und Diffraktion von Barad als Methode synonym verwendet werden:

„Eine diffraktive Methodologie [ist] in erster Linie eine kritische Praxis, um einen Unterschied in der Welt zu machen. Sie ist ein Bemühen, zu verstehen, welche Differenzen von Bedeutung sind, wie sie etwas ausmachen und für wen.“ (Barad 2013, S. 57)

Damit gliedert sich die Interferenz als grundlegender methodischer Gedanke in die von Myers-West et. al. entworfene Forderung nach mehr kritischer Forschung zu KI als Ort gesellschaftlicher Machtprozesse ein. So beschreiben Meyers West et. al. den Bedarf an zukünftiger Forschung zu KI und Machtstrukturen als Fragen nach der politischen Beschaffenheit eben jener Anwendungen in folgender Form „Such Work could ask why a system was designed in a particular way, how it was constructed, and whose interest shaped the metrics by which its success or failure“ (West et. al. 2019, S.16). Die Interferenz als Methode ermöglicht es, solche Fragen zu bearbeiten, damit sie KI

als Phänomen erfassen kann, welches eine konkretes Zusammenspiel menschlicher und nichtmenschlicher Akteur:innen erfasst. Dabei ist das konkrete methodische Vorgehen nach Barad durch das sogenannte „reading through“ gekennzeichnet einem „transdisciplinary engagement attend to the fact that boundary production between disciplines is itself a material discursive practice“ (ebd. S.90).

Die Philosophin Janina Loh beschreibt in ihrer vergleichenden Arbeit zu posthumanistischen Theorien, den gemeinsamen Ansatz von Barad und Haraway hinsichtlich einer Wissenschaftsdisziplinen überwindenden transdisziplinären Arbeitsweise als ein Eingeständnis, dass naturwissenschaftliche und soziale (Forschungs-)Fragen konstitutiv miteinander verbunden sind (vgl. Loh 2019, S.152-153).

Ich möchte in einem nächsten Schritt näher erläutern, inwiefern diese Verbindung Gender als spezifischen Forschungsgegenstand konturieren kann.

Dafür werde ich zwei Linien skizzieren, die diese Verbindung untersuchbar werden lassen, nämlich Geschlecht UND Technik und Geschlecht ALS Technik. Ich werde dafür erstens Geschlecht und Technik als theoretischen Rahmen erläutern und erklären, warum ich die von Landström entwickelte queere Perspektive auf Technologie und Geschlecht als Einstieg wähle.

In einem zweiten Schritte möchte ich „Geschlecht als Technologie“ als weiterführenden Rahmen der Masterarbeit und der zu entwickelnden Perspektive erläutern. Dabei werden drei Subkonzepte von mir eingeführt: einmal nach Judith Butler Geschlecht als Norm zum anderen, von Jack Halberstam, Geschlecht als automatische Konstruktion, sowie Geschlecht als Protokategorie in einem digitalen intersektionalen Kontext nach Cheney-Lippold. In einem weiterführenden Teil werde ich die Verschiebung von Gender und Technologie zu Gender als Technologie aus einer machttheoretischen Perspektive unterfüttern, um eine politikwissenschaftliche Analyse der Einlassung von Geschlecht in Gesichtserkennungssoftware vornehmen zu können. Dafür werde ich die Konzeption von Macht in Anschluss an Haraways „Informatik der Herrschaft“ erläutern.

Dennoch bleibt anzuführen, dass die von mir gewählten Ansätze lediglich eine partielle Antwort liefern können und andere theoretische Entscheidungen andere Perspektiven eröffnen.

B.) Geschlechtertheoretischer Ausgangspunkt der Arbeit

Ich möchte hier darauf verweisen, dass ich mich speziell auf die Beziehung von Geschlecht und Technik beziehe und von daher mein Verständnis von Gender als theoretischen Rahmen entwickle. Ich kann hier in keinsten Weise ein umfassenden Einblick in die geistes- und sozialwissenschaftlichen Debatten um Geschlecht wieder geben. Ich fokussiere mich hierbei besonders auf die Beziehungen von Geschlecht und Technik. Mein genaues Verständnis der zu

untersuchenden Software bzw. des Auditprogrammes werde ich im Abschnitt Untersuchungsgegenstand darlegen und definieren. Ich werde im Folgenden das Verhältnis Geschlecht/Technologie erläutern, welches ich für die Arbeit verwende.

1.) Geschlecht und Technologie (eine Verhältnisbestimmung)

Gender stellt, wie beschrieben, einen wichtigen soziodemografischen Parameter im Bereich der Gesichtserkennung dar, so dass Geschlechtserkennung einen eignen, abgetreten Bereich innerhalb diese Forschungsfeldes darstellt.

Andererseits ist Geschlecht und Technologie ein breites und institutionalisiertes Forschungsfeld welches verschiedenen Stränge beinhaltet (vgl. Faulkner 2001). Dabei sind das Unterscheidungsmerkmal und die Zuordnung entlang der Frage des Verhältnisses von Geschlecht und Technologie organisiert. Ausgehend von Judy Waycmans Arbeiten zur Geschichte der feministischen Forschung zu Technologie und Wendy Faulkners Untergliederung der feministischen STS, möchte ich drei dieser Forschungsstränge vorstellen (vgl. Wacyman 2004, S.11-23, Faulkner 2001, S.79)

Frauen in Technik: Hierbei handelt es sich vorrangig um Fragen der Repräsentation, warum, beispielsweise, gibt es wenig Frauen in technischen Bereichen (vgl. Faulkner 2001, 79).

Frauen und Technik: Hierbei geht es um Forschungsinteressen, die sich um Fragen der spezifischen Nutzung von Frauen und Technologie drehen (vgl. ebd.)

Gender und Technik: Die Frage ist in diesem Fall, wie Technologie gegendert ist (vgl. ebd.)

Ich werde mich für meinen theoretischen Rahmen auf die Forschungsrichtung Gender und Technologie stützen, da ich meine Forschungsfrage in diesem Bereich angesiedelt habe. Dafür möchte ich als Ausgangspunkt den Ansatz der Queer Science and Technologie Studies von Landström verwenden. Dieser Ansatz, in dem Technologie performativ von Landstöröm gefasst wird, spiegelt die von mir eingeführte Idee der Techno Science und deren Technikverständnis wieder, insofern Technologie performativ von Landstöröm gefasst wird. Andererseits wird Gender dabei im Produktionsprozess verortet und als spezifischer Faktor der Entsehungsgeschichte verstanden. Dabei geht die:der Autor:in von einem funktionalen Verständnis von Geschlechtin Verbindung mit Technologie aus. Geschlecht ist in diesem Sinne keine stabile Kategorie der Subjekte, sondern eine Wirkungsweise. Hierbei verweist Landström auf einen konstruktivistischen Reflexionsmoment:

„Technology is, in this view, created and used in a changing, socio-cultural system, and gender, as a feature of this system, is also shaped by technology.” (Landström, 2012, S. 9).

Für diese Art der Analyse wird Geschlecht von Landtsröm als, an einem an Deleuze orientierten

Konzept, Assamblage verstanden. Dies meint die durch Macht strukturierte Verbindung verschiedener Komponenten menschlicher und nichtmenschlicher Akteur:innen (vgl. ebd. 16). Empirisch nützlich wird dieses Konzept eines sich konstruierenden Genders, so Landström, durch die Überlegung von Geschlecht als Oberfläche. Hierbei ist Gender nicht etwas den Personen Innerliches sondern eine Funktion die situativ bedient und verkörpert wird. Gender wird somit, nach Landström zur verorteten Subjektivierungsweise. Um diese Funktion jedoch spezifischer zu fassen, möchte ich mich von hier an auf „Gender als“ fokussieren, um genauer zu erläutern, wie Gender in Bezug auf Technologie arbeitet.

2.) Gender als Technik

a.) *Geschlecht als Norm*

Landström selbst leitet das Verständnis von Gender von Butlers Ideen zu Geschlecht ab und fokussiert sich hierbei besonders auf den performativen Aspekt von Gender, also das ständige reartikulieren von Geschlechternormen, im verbalen aber auch physischen Sinne (vgl. ebd., S. 17). Ich hingegen möchte mich mehr auf den Normcharakter von Geschlecht in der Butlerschen Theoriekonzeption beziehen. Hierbei geht es mir um Geschlecht als Etwas. Ich möchte mich von dem konkreten Subjekt als Forschungsgegenstand abgrenzen, um in eine theoriegestützte Analyse des noch zu behandelnden Auditprogrammes „Gender Shades“ zu kommen.

Die Norm als theoretisches Konzept ermöglicht diesen Analyseschritt. Butler leitet die Konzeption von Gender als Norm, von Foucaults Begriff der Norm ab. Sie ist dabei stets eine Art und Weise der Hervorbringung von Regeln und somit kein konkretes juristisches Gesetz (vgl. Butler 2011, S. 86). Dabei ist die Norm „ein Maß und ein Mittel, um einen Standard hervorzubringen“ (vgl. ebd. 87). Die Norm funktioniert als Moment der Individuation und der Vergleichbarkeit, indem sie diskrete Einheiten hervorbringt (vgl. ebd. 88). Foucault selbst bezieht diese Funktionsweise auf moderne Regierungen bzw. die Norm als neue Form der Macht. Sie stellt ein Feld der sozialen Macht dar und ermöglicht Intelligibilität, also die Sichtbarmachung von Subjekten bzw. deren Ausschluss aus dem sozialen Feld. Sichtbarmachung meint hierbei auch die abgesicherte Existenz von Subjekten (vgl. Foucault 1990, 24)

Dabei werden mittels normalisierender Institutionen Entitäten geschaffen; wie der Schüler, der Kranke oder der Delinquent. Solche Prozesse produzieren eine standardisierte Form des Subjektes, mit Funktionen und erwartbarem Verhalten innerhalb der gesetzten Rahmen. Ein Schüler geht in den Unterricht, schwänzt er diesen, widersetzt er sich in einem ersten Blick, kann dies jedoch immer nur in Bezug auf das System Schule. Die Abweichung wird zu einer Situation der Artikulation

der regulatorischen Macht, indem diese sanktioniert wird und die Norm, das im Unterricht erscheinen, als unsanktionierte Verhaltensweise bestätigt wird (vgl. ebd. 44).

Absichtlich habe ich im letzten Abschnitt das männliche Gender verwendet, im Gegensatz zu Foucault sieht Butler nämlich in Geschlecht eine spezifische Norm, die eigene Regulierungsmechanismen hervorbringt. Geschlecht wird somit zu einem neuen Ort von Macht ohne aber die regulierende Macht per se zu sein (vgl. Butler 2011, S.72). Für meine Analyse ergibt sich daraus eine erste systematische Einteilung. So schreibt Butler das sich Gender als Norm immer auf eine über den jeweiligen Fall hinausgehende Subjektivierungsweise beruft, also nicht die jeweilige Handlung sei, jedoch, dass auch immer die konkreten Regulierungen konstitutiv für das Bestehen der Norm sei (vgl. ebd. 71). Die Frage nach konkreteren Regulierungen, wie beispielsweise die Frage, welche Menschen dürfen wen heiraten, ist Bestandteil der tieferliegenden Frage, wie werden diese Menschen überhaupt erst zu Personen die heiraten dürfen. Somit möchte ich meine Analyse im nachfolgenden Teil auch mit den von den Autor:innen vorgelegten Fragen der Geschlechterdiskriminierung beginnen, um empirische Fälle der konkreten Regulierung von Gender durch automatische Gesichtserkennung aufzuzeigen zu können, und dann auf die Frage der Herstellung von Standards von Geschlecht als Norm eingehen zu können.

Dabei beschreibt Butler Geschlecht als Norm in der westlichen, kapitalistischen Gesellschaft als die Absicherung der geschlechtlichen Binarität. Für Butler schafft Gender als Norm eine Intelligibilität für das Männliche und das Weibliche. Damit sei nicht die normative Vorstellung von Weiblichkeit oder Männlichkeit sondern die Binarität selbst gemeint (vgl. ebd. S. 74). Geschlecht als Norm reguliert nicht mehr einzelne Bereiche, wie Gesetze gegen sexuelle Belästigung am Arbeitsplatz, sondern ist vielmehr „die Produktion von Parametern der Personalität, dass heißt die Herstellung von Personen in Übereinstimmung mit abstrakten Normen, welche die einzelnen Menschenleben zugleich bedingen und übersteigen und auch zerbrechen (vgl. ebd. S. 96). Gender als Norm ermöglicht also einen binären Standard, der Menschen erst zu vergeschlechtlichten Wesen werden lässt. Geschlecht ist somit nicht Bestandteil des Subjektes per se, sondern muss immer wieder erzeugt werden. So wird in diesem Konzept auch die Abweichung von der Norm selbst zu einem stabilisierenden Beispiel, da es eben eine Abweichung darstellt und aus dem Rahmen der sozialen Akzeptanz fällt bzw. nicht intelligibel ist (vgl. ebd. 91).

Jedoch sieht Butler in ihrem Aufsatz die Möglichkeit des Widerstandes, nämlich Gender zu queeren als eine Art der Infragestellung der zugrunde liegenden Binarität. Diesen Punkt möchte ich in einem späteren Kapitel der Arbeit aufgreifen. Ich werde diskutieren, inwieweit Gender als Technologie, als emanzipatorischer Ansatz geschlechtergerechter Politik verstanden werden kann oder aber, als Kategorie der Differenz, emanzipatorische Politik in Bezug auf Geschlechtererkennung untermi-

niert.

In einem nächsten Schritt möchte ich die Idee der Standardisierung, wie sie Butler beschreibt, mittels Jack Halberstams Idee der Automation von Gender verbinden, um den Ansatz Geschlecht als Technologie, wie ich ihn entwickeln möchte, spezifisch auf meinen Untersuchungsfall anwenden zu können.

b.) Geschlecht als automatische Konstruktion

Nach Butler ist „Gender [...] der Apparat, durch den die Produktion und Normalisierung des Männlichen und Weiblichen vonstatten geht“ (vgl. ebd. 74). Dabei führt Butler jedoch nicht aus, inwieweit diese Konzeption metaphorisch oder aber als produktive Funktion von Existenzweisen zu verstehen ist. Er bleibt untertheoretisiert in den Ausführungen. Es ist unklar ob Apparat hierbei meint, ob Geschlecht im Sinne einer Maschine funktioniert oder die Wortwahl ungenau getroffen wurde; denn der weitere Text bietet keinen Anhaltspunkt für die Fragen: Ist Geschlecht eine Technologie? Wie kann diese Funktion theoretisch beschrieben und somit für Forschung nutzbar gemacht werden? Interessanterweise ist auch der Aufsatz von Landström hier nicht weiterführend, da Landström diese Frage ausspart.

Um diese Lücke zu füllen möchte ich Jack Halberstams Aufsatz „Automating Gender: Postmodern Feminism in the Age of Intelligent Machine“ besprechen und sein Konzept von Gender erläutern. Jack Halberstam bettet sein Konzept in eine materialistische Gegenwartsanalyse ein, die auch heute noch Bestand hat. Für ihn leben wir in einem Zeitalter der Massentechnologie. Diese schreibt sich mit der Verbreitung des „Homecomputers“ materiel in westlichen Industriegesellschaften ein (vgl. Halberstam 1991, S.441). Die Aufgabe der feministischen Theorie könne deshalb nicht mehr die Ablehnung oder Befürwortung von Technologie sein, sondern die Frage der verschiedenen Konfiguration von Technik (vgl. ebd.). Das Neuartige des späten 20. Jahrhunderts sind demnach für Halberstam nicht einfach nur neue Technologien, sondern die hybriden Formen von Subjektivität, die diese neue Ära bevölkern (vgl. ebd. S.440). Ausgehend von Donna Haraways Figur des Cyborgs als Wesen zwischen Maschine und Mensch, fokussiert sich Halberstam auf die explizit vergeschlechtlichte Version dieser Figur: die Female Cyborg. Diese sei eine verkörperte Grenzverwischung von Mensch/Maschine/Tier und somit auch eine neuartige Figuration von Gender, jenseits des weißen Cismannes (vgl. ebd.). Für meine Arbeit ist hierbei jedoch nicht die:der Cyborg als konkrete Subjektform von Interesse sondern Gender als Automationsprozess. Denn Halberstam verknüpft Gender hierbei mit einer technischen Funktionsweise. Anders als Butlers auf die (juridische) Norm bezogenes Geschlechtskonzept vertritt Halberstam in Anlehnung an Haraway eine technologische Sichtweise. Nicht das Gesetz sondern der Computer wird zum Paradigma für Gender:

“Gender, we might argue, like computer intelligence, is learned, imitative behavior that can be processed so well that it comes to look natural. [...]. In other words, gender, like intelligence, has a technology” (vgl. Halberstam 1991, 443)

Hier verbindet Halberstam Geschlecht noch mit Technologie als etwas, das besessen werden kann, beziehungsweise als Attribut von Geschlecht. Später werde ich mittels der von mir gewählten theoretischen Ansätze zeigen, dass die Effekthaftigkeit von Geschlecht in Bezug auf Gender Recognition Software, Geschlecht als Technologie konstituiert.

Die Halberstamsche Spezifizierung der Wirkungsweise von Geschlecht bleibt jedoch nicht bezogen auf eine rein semantische Konstruktion, sondern bindet die Materialität von Technologie und Gender ein. Der vergeschlechtlichte Körper wird nicht aufgeteilt in Geist/Code und Körper/Hardware, sondern ist eine durch machtvollen Kategorien geformte Akteur:in. In der Theorie ist die Female Cyborg nicht der:die utopisch/dystopisch aufkommende Andere im Cyberspace, sondern ist Subjekt innerhalb des Informationskapitalismus. Geschlecht und Ausbeutung werden auch hier zusammen gedacht. Vergeschlechtlichte Arbeitsteilung ist für Halberstam nicht mehr die exklusive Verknüpfung mit der industriellen Revolution, sondern mit neuen Technologien und fließenden Machtstrukturen (vgl. ebd. 457). Dabei ist Gender, um es noch einmal deutlich zu machen, keine versteckte essentialistische Existenzweise moderner Menschen, sondern „an electronic text that shifts and changes in dialogue with users and programs“ (vgl. ebd. 458). Bezogen auf meine Forschung lässt sich hier auch die Verbindung von Materialität und digitaler Technik nachvollziehen, Geschlecht und automatische Geschlechtererkennung lassen anhand des Bildes einer Person, die Frage nach Körper und Geist verschwimmen. Das Bild eines Menschen wird zum Ausdruck seines/ihrer Genders, jedoch innerhalb einer textbasierten maschinellen Umgebung.

Kann Gender also als personelle Kategorie verstanden werden? Handelt es sich um ein universell humanes Merkmal? Sind Norm/Apparat automatisch generiert, und, wenn ja, welche Möglichkeiten des Politischen ergeben sich daraus?

Um diese Fragen anhand des Untersuchungsgegenstandes beantworten zu können, möchte ich noch Geschlecht in einer dritten Funktion erläutern, nämlich die intersektionale Kategorisierung von Ge-

schlecht.

c.) *Geschlecht als digitale Kategorie*

Ausgehend von meinem Forschungsgegenstand ist Geschlecht immer verbunden mit anderen sozio-demografischen und politischen Kategorien, wie in der zu untersuchenden Studie „Gender Shades“ Race: So wird Geschlecht von den Forscher:innen in einer interskettionalen Matrix untersucht. Um dieser theoretisch gerecht zu werden, möchte ich im Folgendem Geschlecht als Subjektkategorie einführen und auf die Verknüpfungen mit Race hinweisen. Dabei verwende ich die theoretischen Konzepte von John Cheney-Lippold⁹. Dieser aktualisiert noch einmal Halberstams Zeitdiagnose und fokussiert sich knapp 30 Jahre später auf Big Data Technologie als paradigmatische gesellschaftliche Struktur (vgl. Cheney-Lippold 2017, 88-89). Dabei sei das Einteilen in abgrenzbare, vergleichbare teileeine Form der sozialen Ordnung (vgl. ebd. 53). Der Autor unterscheidet zwischen digitalen, aufgeteilten Identitäten und dem analogen Gegenpart. Dies erläutert der Autor für Gender. So gebe es eine analoge politische Identität Gender und eine algorithmische Kategorie Geschlecht, beispielsweise durch Suchmaschinen, wie Google, zwecks Marketing generiert (vgl. ebd. 64). Geschlecht wird in dieser algorithmischen Umgebung zu einer Aggregation statistischer Wahrscheinlichkeiten, die sich in Korrelationen und nicht in Kausalitäten ausdrücken. Paradigmatisch dafür sind die Kaufempfehlungen großer Internetversandhändler:innen. Cheney-Lippold verweist in diesem Zusammenhang auf eine Neue Form von Intersektionalität: die „protocategorical“ Intersektionalität. In Anlehnung an die Wissenschaftler:innen Kimberlé Crenshaw und Leslie McCall bezieht Cheney-Lippold sich auf die konstitutive Verwobenheit verschiedener kategorialer Zuschreibungen. Ein Fokus liegt dabei, in Anlehnung an das Erbe der Intersektionalität als schwarze feministische Theorie, auf Race und Gender. Dabei entwirft er, ausgehend von McCalls Unterscheidung verschiedener Intersektionalitätskonzepte, seine Überlegungen zu einer auf Algorithmen bezogenen intersektionalen Theorie.

„The protocategorical is how [...] the algorithmic world makes the logic of intersectionality useful [...]. But this computational perspective reads intersectional patterns without context and thus erases the “complexities” of discourse—and ultimately our lived experience. [...] The possibility for a protocategorical intersectional politics comes into the picture only after the clustered category is

⁹ John Cheney-Lippold ist durch antisemitisches Veralten aufgefallen (vgl. <https://www.bbc.com/news/world-us-canada-45592553> zuletzt abgerufen am 12.03.2021). Ich habe keiner seiner theoretischen Überlegungen bezüglich Israel übernommen. Dennoch habe ich seine Überlegungen zur Intersektionalität und Digitalen Infrastrukturen einbezogen, da ich sie für grundlegend halte. Dennoch ist diese Lösung natürlich nicht zufrieden stellend. An dieser Stelle möchte ich deshalb auf die Publikation von Steffen Klävers „Decolonizing Auschwitz?: Komparativ-postkoloniale Ansätze in der Holocaustforschung“ zur systematischen Untersuchung dieses Spannungsfeldes.

created, not before. It is a politics that is useful for whoever is doing the algorithmic processing, not for those who are living the output of that processing.” (ebd. 81; 84).

Geschlecht wird hierbei weniger als Technik denn als statistische Referenz begriffen. Der Algorithmus, hier von Cheney-Lippold als KI verstanden, extrahiert aus einer großen Menge an Daten Muster. Diese Muster sind, in Cheney-Lippolds Theorie, Kategorien. Diese sind jedoch, anders als bei Butler, nicht produktive Norm sondern entpolitisierte Datencluster zur Bedeutungsgewinnung in einer Informationsökonomie. Für meine Forschung ist in diesem Zusammenhang vor allem das schon in der Analyse der algorithmischen Arbeitsweise angelegte intersektionale Netz an Identitätskategorien. Cheney-Lippold weist in seiner Analyse auf die neuen technischen Möglichkeiten der Verkettung von Kategorien hin. Es stellt keine Ressourcen bindende Hürde dar, im Zeitalter informationsverarbeitender Tools, große Mengen an demographische Daten zu verbinden und in ihnen Muster zu erkennen. Gleichzeitig, so Cheney-Lippold, werden jedoch spezifische Verkettungen von Kategorien, ihres politischen Kontextes beraubt, in dem diese als Platzhalter der statistischen Korrelation dienen.

Ich werde in meiner Analyse teils auf die theoretische Spannung zwischen dieser Konzeption algorithmischer Intersektionalität und dem politischen Ansatz von Buolamwini in der zu untersuchenden Studie eingehen.

Vorher möchte ich jedoch in einem letzten Schritt meine zugrunde liegende Machtkonzeption erläutern, um die angeführten theoretischen Anschlusspunkte meiner ausgewählten wissenschaftlichen Ansätze nachvollziehbar zu gestalten. Denn eine machtheoretische Perspektive ermöglicht mir die Verbindung der bis jetzt noch eher rudimentären, aufeinander bezogenen theoretischen Ausführungen.

d.) Gender, KI und Macht

Die vorgestellten Genderkonzeptionen bedienen sich zur machtanalytischen Rahmung um eine Foucaultsche Perspektive von Macht als Norm, die sich mit Standardisierung von Subjekten beschäftigt und dadurch mit ihrer Schaffung *durch* Standardisierung (vgl. Butler 2011, S. 92-93). Hierbei jedoch fehlt die theoretische Bezogenheit von Macht auf den Vorgang der automatischen Gesichtserkennung. Wenn Geschlecht, wie Butler schreibt, einem spezifischen Regulierungsapparat eingeschrieben ist, wo genau liegt dann der fachtheoretische Unterschied zwischen der Schule als Institution und neuronalen Netzwerken? Wenn Cheney-Lippold von einer Macht spricht, die sich erst auf Individuen auswirkt, nachdem die digitale Kartierung ihres Geschlechtes abgeschlossen ist, bleibt die Frage, ob sich Macht auch jenseits analoger Subjekte vollziehen kann, die sich nicht auf

den Menschen, ohne technologische Verwobenheit bezieht.

Um diese Problematik fassen zu können, möchte ich Donna Haraways Konzept der Informatik der Herrschaft einführen. Wie die bisherigen Ansätze bezieht auch Haraway ihre Theorie auf eine neue, paradigmatische gesellschaftliche Umwälzung, die sich durch und mittels Technik vollzieht (vgl. Haraway 1995a, S.48). Dabei ist Mirkotechnik(Computerchiptechnologie) die Infrastruktur der Gesellschaft im 21. Jahrhundert. Für Haraway gibt es keine technische Lösung und keinen technischen Determinismus, der auf die soziale Welt angewandt wird, vielmehr ist diese zur Technologie oder genauer formuliert zum Schaltkreis geworden (vgl. ebd. S.53). Die Informatik der Herrschaft drückt sich hierbei nicht mehr in Formen des Gesetzes oder der Hierarchie aus, sondern in „Design Strategien, Constraints, Durchsatzraten, Systemlogiken oder Kostenverminderung“ (ebd. S.49). Haraway verweist dabei auf eine neue Subjektposition, die sich nicht mehr am Menschen fest macht, sondern an verschwimmenden Grenzen von Maschinen/Mensch und Tier, dem schon besprochenen Cyborg (vgl. ebd.). Dabei ist der:die Cyborg in Haraways Sinne gerade nicht durch die Machtkonzeption Foucaults geprägt sondern der Simulation dieser. Der Cyborg verweise nicht auf eine produktive, subjektivierende Macht, da es keine Subjekte im eigentlichen Sinne mehr gäbe. Haraway weist somit über die von Cheney-Lippold eingeführte Unterscheidung zwischen Subjekt und digitalem Doppelgänger hinaus. Beides fällt in Haraways Analyse zusammen und bildet eine sich wieder und wieder reproduzierbare Subjektivität, die sich jedoch nicht an einem Individuen festmachen lässt. Des Ursprungs durch unendliche Wiederholungen beraubt wird das postfoucaultsche Subjekt zum Simulakrum, der Kopie ohne Original (vgl. ebd. S.50). Der Mensch wird zu einer Figuration des neuen universalen Problems der Übersetzbarkeit.

„Wie jede andere Komponente und jedes andere Subsystem auch, müssen menschliche Lebewesen in einer Systemarchitektur verortet werden, deren grundlegende Operationsweisen probabilistisch und statistisch sind. Kein Objekt, Raum oder Körper ist mehr heilig und unberührbar. Jede beliebige Komponente kann mit jeder anderen verschaltet werden, wenn eine passende Norm oder ein passender Kode konstruiert werden kann, um Signale in einer gemeinsamen Sprache auszutauschen“ (ebd. S.50)

Diese Übersetzbarkeit funktioniert mit den Methoden des maschinellen Lernens, mittels Wahrscheinlichkeit, Übersetzbarkeit etc.. Haraway betrachtet Geschlecht dabei selbst als in andere Technologien zu überführendes Moment moderner Vorstellung von Individuen. Gender selbst wird somit zu einer Frage der politischen Gestaltung und bekommt, anders als bei Butler, einen Herrschafts-/Unterdrückungsmoment zugeschrieben, der Geschlecht als Kategorie des Widerstandes und der emanzipatorischen Zukunftsgestaltung disqualifiziert (vgl. ebd. S.53). Gleichwohl ist diese Kategorie, wie auch Class und Race, für Haraway wichtiger Startpunkt der Möglichkeit einer kritischen Auseinandersetzung mit Gesellschaft und Welt (vgl. ebd. S.35). Diese beiden zukunftsorien-

tierten Geschlechterdimensionen (von Butler und Haraway) möchte ich später in der Arbeit miteinander vergleichen.

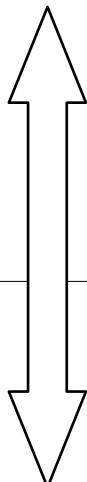
Des Weiteren wird bei Haraway die intersektionale Verbindung von Geschlecht, Race und Klasse sichtbar. Denn die Übersetzungen für Menschen im Zeitalter der Informationstechnologie geschehen auf unterschiedliche Arten und Weisen, nicht abhängig von individuellen liberalen Leistungsversprechen, sondern kategorialer Positionierung im globalen kapitalistischen Informationsnetzwerk. Haraways sozialstatistisch feministische Perspektive lässt in diesem machtheoretischen Kontext Arbeit als wichtigen Faktor der Stabilisierung sichtbar werden.

„Die Haushaltsökonomie als internationale Organisationsstruktur des Kapitalismus wurde durch die neuen Technologien ermöglicht (nicht etwa verursacht). Die Folgen der neuen Technologien bekommen Frauen sowohl im Verlust des (männlichen) Familieneinkommens (falls sie jemals Zugang zu diesem weißen Privileg besaßen) als auch an ihren eigenen Arbeitsplätzen, deren Kapitalintensität ständig steigt, wie z.B. Büroarbeit oder Krankenpflege, zu spüren.“ (vgl. ebd. S.55)

Hiermit habe ich meinen machtheoretischen Ansatz dargelegt und möchte im Folgenden noch einmal zusammenfassend meine Matrix zur Untersuchung der Studie „Gender Shades“ darstellen.

C.)Theorie Matrix

Ich möchte mit aufbauend auf dem Ansatz von Geschlecht und Technik auf ein aus den Theorien von Geschlecht als Norm/automatische Konstruktion/Protokategorie. Ich möchte auf dem Ansatz von Geschlecht und Technik, dabei mit aufbauend auf den Theorien von Geschlecht als Norm/automatische Konstruktion/Protokategorie ein allgemeineres Verständnis von Gender als Technologie entwickeln, um die Forschungsfrage in einem funktionalistischen Sinne beantworten zu können. Ich unterlege diese theoretische Begriffsarbeit mit einem Machtkonzept von Haraway, um einer politikwissenschaftlichen Analyse gerecht werden zu können. Die Anwendung auf die Studie „Gender Shades“ erfolgt dabei einerseits aus einem aus dem Artefakt hergeleiteten Forschungsinteresse und andererseits aus einer Einordnung eben jenes Forschungsgegenstands als symptomatische Ausformung automatischer Geschlechtererkennung (s. Abbildung 2)

Macht als Informatik der Herrschaft • Kapitalistische Verhältnisse Grundlage technischer Gesellschaft • Mikroelektronik ist die Struktur von Gesellschaft • Keine Subjekte mehr sondern Vernetzte Cyborgs • Macht ist die Vernetzung von allem mit allem • Intersektionale Knotenpunkte der Machtvollen Vernetzung	Gender als Technik • zu entwickelnde Perspektive	
	Gender als Protokategorie • Statistische Referenz von Digitalem Gender ◦ Ist Informationscluster: entsteht aus Daten • Politische Identität Gender • Beides verknüpft mit anderen Kategorien	
	Gender als automatische Konstruktion • Protokoll für menschliches Verhalten und Analogie zu Künstlicher Intelligenz	

	Gender als Norm • Standard zur Einteilung von Menschen in Binäre Kategorien • Norm des Intelligiblen im sozialen Feld	
	Gender Und Technik: • Gender und Technik konstruieren sich gegenseitig • Gender stellt eine soziale Oberfläche dar, kein Kernelement von Subjekten	

(Abb.2. Theoretische Matrix)

III.) Hauptteil

A.) Zielstellung der Studie

Ich werde im Folgenden die Zielsetzung und die Notwendigkeit der vorgestellten Studie Gender Shades analysieren. Dieser Schritt folgt dabei zweierlei Überlegungen. Einerseits arbeite ich entlang der Gliederung die ausgewählten Untersuchungsgegenstände (vgl. Buolamwini, Gebru 2018, Buolamwini 2017). Zweitens berücksichtige ich die politischen Arbeit Buolamwinis, welche sich um das Projekt Gender Shades entwickelt hat und schließlich in der Gründung der Algorithmic Justice League durch die Wissenschaftlerin mündete (vgl. D'Ignazio, Klein 2020, S. 32). Die Grenzverschiebung von praktischer Politik und politischer Analyse der Auditstudie möchte ich so verdeutlichen. Meine theoretische Einführung in die Verbindung von Technologie und Gender bzw. Race, Gender und Technologie wird in diesem Abschnitt der Arbeit, zu einem Moment gelebter Technopolitik. Ich untersuche nicht nur einen technischen oder politischen Zweck bzw. eine technologische Antwort auf ein Problem der statistischen Repräsentation. Vielmehr analysiere ich eine Artikulation der politischen Gestaltbarkeit technologischer Artefakte, wie der automatischen Geschlechtererkennung. Ich werde so eine Teilantwort auf die Einschreibung von Geschlecht im Gender Shades Projektes liefern.

Dies geschieht wie folgt: Ich werde zuerst die genaue Zielsetzung Buolamwinis mit der Auditstudie Gender Shades vorstellen. Dies wird der theoretischen Orientierung dienen.

In der anschließenden Analyse der Zielsetzung werde ich, ausgehend von meinem theoretischen Rahmen, die Idee von Judith Butler der Genderregulierung auf die Zielstellung des Auditprogrammes anwenden und diese, mittels Karen Barads Idee der posthumanistischen Performativität, erweitern. Um den intersektionalen Charakter des Gender Shades Projektes zu untersuchen, werde ich das Konzept der Protokategorie von Cheney-Lippold auf diese Studie anwenden und diskutieren.

In einem letzten Teil werde ich diese Ergebnisse ausgehend von meinen machttheoretischen Überlegungen diskutieren, um noch einmal die Frage der politikwissenschaftlichen Bedeutung zu behandeln.

1.) Definition und Vorstellung der Zielstellung

Das Auditprogramm verfolgt zwei verbundene Ziele. Erstens untersucht Buolamwini die drei gängigen Genderclassifyer auf dem Markt (IBM, Face++, Microsoft) auf intersektionale Verbringungen von Race und Gender (vgl. Buolamwini, Gebru 2018, S.1). Zweitens entwickeln die Autor:innen ein eigenes Sample, welches als Maßstab für die Genderclassifyer dient (vgl. ebd.). Damit reagieren sie auf eine strukturelle Problematik statistischer Erfassungen im Allgemeinen und der automatischen Gesichtserkennung im Speziellen. Caroline Criado-Perez beschreibt dies, zugespitzt, über auf die Frage der datentechnischen Repräsentation von marginalisierten Menschen, als tiefgreifendes Problem heutiger digitaler Gesellschaften.

„Es gibt eine Leerstelle in den wissenschaftlichen Daten in Bezug auf Frauen im Allgemeinen [...] doch über Schwarze Frauen, behinderte Frauen oder Frauen aus der Arbeiterschicht gibt es praktisch keinerlei wissenschaftliche Daten. Nicht nur weil sie gar nicht erst erhoben werden, sondern auch weil, sie nicht von über Männer erhobenen Daten getrennt sind.“ (vgl. ebd. 2020, S. 15).

Buolamwini geht jedoch in ihrer Bedarfsanalyse und Zielstellung des Gender Shades Projekts weiter und erklärt, dass es auf der einen Seite einen massiven Gap zwischen der Repräsentation von marginalisierten Gruppen, insbesondere Schwarzer Menschen gibt, und gleichzeitig diese Menschen nicht einfach „nur“ aus der Betrachtung oder dem Designprozess verschwinden, sondern wiederum proportional stärker von diesen Devices betroffen sind. Sie unterlegt diese These mit dem Beispiel der US-Strafvollzugsbehörde. Diese nutzt Gesichtserkennungssoftware zur Ermittlung von Straftäter:innen. Schwarze Menschen werden häufiger von der Polizei angehalten und somit öfter einer Gesichtserkennungsanwendung unterzogen als weiße Menschen. Gleichzeitig sind Schwarze Menschen und Menschen of Color deutlich unterrepräsentiert in den verwendeten Datensets zum trainieren dieser Anwendungen. Dies führt zu erhöhter Gefahr einer Fehlidentifikation (vgl. Buolamwini, Gebru 2018). Das Auditprogramm, welches sich also auf die Korrektheit von automatischer Gesichtserkennung entlang der soziodemographischen Kategorien Geschlecht und Race orientiert, hat das soziotechnische¹⁰ Ziel einer Überprüfung der Akkuratheit der Ergebnisse von Genderclassifyern, aufgrund der politischen Wirkmächtigkeit bereits bestehender Anwendungen (vgl. Buolamwini 2017, S.14).

10 Ich leite diesen Ansatz von meinem technowissenschaftlichen Grundverständnis ab, in dem digitale Technologien die Grundvoraussetzung der Gesellschaften im 21. Jahrhundert darstellen. Somit sind diese immer gesellschaftlich bedingt:

„Der soziotechnische Ansatz betrachtet und beurteilt automatisierte Prozesse und datengestützte Entscheidungen von Menschen immer in ihrem gesellschaftlichen Kontext. Dazu zählen politische Rahmenbedingungen, Interessen und Einflussmöglichkeiten staatlicher Institutionen, Unternehmen und anderer Organisationen sowie Wertvorstellungen der beteiligten Akteur*innen; zudem schließt er soziale, ökonomische und ökologische Folgen jenseits des geplanten Einsatzes der jeweiligen Technologien ein.“ (Sachverständigenkommission für den Dritten Gleichstellungsbericht der Bundesregierung, 2021, S. 22)

Catherine D'Ignazio und Lauren F. Klein heben diese Bedingtheit des Projektes besonders hervor und charakterisieren Buolamwinis Arbeit bzw. ihr Sein als Wissenschaftlerin als „model for the data scientist as public intellectual – who yes, works on technical audits and fixes, but also works on cultural, legal and political efforts too.“(vgl. ebd. 2020, S. 32).

Buolamwini beschreibt ihr politisches/technisches Ziel als Transparenz sensibler Algorithmen. Dabei meint Transparenz, die Offenlegung erhöhter Risiken für Diskriminierung, also unverhältnismäßiger Outputs basierend auf soziodemografischen Faktoren wie Race und Gender und deren Verteilungen auf Gesellschaften. Dies würde zum Beispiel für die Nutzung von Gesichtserkennungssoftware der Polizei bedeuten: offen legen der verwendeten Datensets, um einen Abgleich der soziodemographischen Merkmale wie Race und Gender im Datenset mit der empirischen Verteilung jener Marker im Gebiet, in dem jene Software eingesetzt wird, zu ermöglichen.

Buolamwini folgt hier auf den ersten Blick bereits etablierten normativen Überlegungen zu KI und deren gesellschaftlichen Folgen, nämlich Debatten um Verantwortung (accountability), Transparenz (Transparency) und Fairness. Diese Konzepte sind bereits jetzt Schauplätze wissenschaftlicher Debatten und durch eigene Journals, Tagungen und Professuren ein aufstrebendes Forschungsfeld, den Forschungen rund um Bias (vgl. Buolamwini 2017, S. 14). Das Alleinstellungsmerkmal der Forschung von Buolamwini ist jedoch die Verkettung verschiedener Diskriminierungen innerhalb der automatischen Gesichtserkennung (vgl. ebd. 19).

a.) Hinführung zur Zielstellung des Gendershades Projekts

Künstliche Intelligenz ist mittlerweile Teil einer globalen sozialen Infrastruktur, von Pflegebots, Kreditvergabesystemen und des militärischen Einsatzes in Form von Drohnen. Dabei zeigt die wissenschaftliche Auseinandersetzung mit den Folgen und Auswirkungen dieser Technik, dass KI bestehende soziale Ungleichheiten verstärk und diskriminierend agiert (vgl. Myers West 2020 S.2, O'Neil 2016, S. 193-ff.)

Gesichtserkennungsanalyse durch Künstliche Intelligenz ist dabei eine Unterkategorie der Analyse sozialer Ungleichheit durch und mit KI. Automatische Gesichtsanalyse ist ein Teilgebiet der Künstlichen Intelligenz, in welchem, mittels maschinellen Lernens, Gesichter entweder bekannten, in Datenbanken abgespeicherten Fotos zugeordnet werden sollen, oder es handelt sich um automatische Erkennung. Hier soll das Programm erkennen, ob sich ein menschliches Gesicht auf dem Foto befindet (vgl. Buolamwini, Gebru 2018, S.2). Es zeigte sich, dass diese

Gesichtserkennungstools oftmals Gesichter von Schwarzen Menschen oder Menschen of Color nicht erkannten oder missklassifizierten (vgl. Benjamin 2018, S. 50-51). In diesem Kontext bewegt sich das Gender Shades Projekt. Ausgehend von empirischen Fällen der Diskriminierung von BIPOC untersuchten Buolamwini und Gebru, ob und inwiefern 3 kommerzielle Genderclassiifer, Gender an und auf Gesichtern erkannten, die Schwarz und of Color waren (vgl. Buolamwini, Gebru 2018). Das Problem, dass das Gender Shades projekt behandeln sollte, ist der Fakt, dass Gesichtserkennungssoftware, Schwarze Menschen und Menschen of Color systematisch missklassifiziert und unkenntlich macht, wie im Falle der Gesichtserkennungsanwendung von Google, die Schwarze Menschen als Gorillas diffamierte (vgl. West 2020, S.2). Dabei verweisen Buolamwini und Gebru auf die Gefahr der Missklassifizierung, aufgrund des Einsatzes von KI in immer sensibleren gesellschaftlichen Bereichen, unter anderem dem US-amerikanischen Strafvollzugssystem. Gesichtserkennungssoftware wird bereits heute innerhalb der Polizeibehörden eingesetzt und soll Straftäter:innen ermitteln, mittels Abgleich der behördlichen Datenbanken (Browne 2015, 127-8, Barabas 2020). Dabei, konstatiert die Soziologin Simone Browne, seien diese Technologien so stark von einer weißen Hegemonie bestimmt, dass diese das Fundament biometrischer Technologien, wie der automatischen Gesichtserkennung, sei.

„What Gordon insightfully calls the “notion of white prototypicality” is the enabling condition of the structured violence of “the dialectics of recognition.” [...] This prototypical whiteness is one facet of the cultural and technological logic that informs many instances of the practices of biometrics and the visual economy of recognition and verification that accompanies these practices“ (Browne 2015, S. 110).

Nach Browne sind die eben beschriebenen KI-Systeme nicht etwa die Ausnahme, sondern sie deuten auf das rassifizierte „Funktionieren“ einer Technologie, die weiße Hegemonie absichert.

Buolamwini und Gebru weisen vor diesem Hintergrund auf die Gefahr für Schwarze Menschen hin, von solchen Tools nicht nur falsch erkannt zu werden und somit potentiell dem ungerechtfertigten Eingreifen der US-amerikanischen Behörden ausgeliefert zu sein, sondern auch auf die besondere Gefahr für Schwarze Menschen speziell hinsichtlich der Strafvollzugsbehörde. Die Bürgerrechtsaktivistin Michelle Alexander erklärt in ihrem Werk „The New Jim Crow“, dass die USA 5% der Weltbevölkerung stellen und 25% aller Gefängnisinsass:innen. 2006 saß einer von 14 männlichen Schwarzen im Gefängnis und einer von 106 weißen Männern (vgl. Alexander 2017, S.145).

„Das heutige Kontrollsystem sperrt einen großen Prozentsatz der Afroamerikaner aus der Gesellschaft und der

Arbeitswelt aus [...] Wie Jim Crow (und die Sklaverei) stellt die Masseninhaftierung ein eng gestricktes Netz aus Maßnahmen dar, die zusammen dafür sorgen, dass sich eine weitgehend durch Rasse definierte Gruppe nicht aus ihrem untergeordneten Status befreien kann.“ (ebd. 35).

Dieses eng gestrickte Netzwerk wird durch Gesichtserkennungstechnologie weiter verdichtet, indem beispielsweise das „racial profiling“, also die Kontrollen, die von Polizeibehörden aufgrund rassifizierter Zuschreibung von Verbrechen durchgeführt werden, und die Daten, die so entstehen, wiederum zu scheinbar abgesicherten erneuten Durchsuchungen führen. Ein Zirkelschluss, der datentechnisch abgesichert ist. Der Jurist Rodrigo Ochigame benannte diese Praxis als „fundamental misattribution of agency“ (Barabas 2020, S. 104). So würden rassistische Inhaftierungspraxen zu rassistischen Daten führen, die wiederum Gesichtserkennungstechnologie begründen und formen, die wiederum zu mehr Überwachung Schwarzer Gemeinschaften und schließlich zu mehr Festnahmen führen (vgl. ebd.),

Dabei stellen Missklassifikationen für Schwarze Menschen und Menschen of Color ein zusätzliches Sicherheitsrisiko dar, in diesen Kreislauf mit einbezogen zu werden (vgl. Buolamwini, Gebru 2018, S.1).

Buolamwini und Gebru reagieren mit ihrer Studie damit auf ein System der rassifizierten Unterdrückung, welches sich auch auf Technologien der Gesichtserkennung als Hilfsmittel beruft.

Diese Problematik der Missklassifikation und des Nichterkennens versuchen Buolamwini und Gebru mit einem sogenannten Benchmark zu begegnen (vgl. ebd.). Benchmarks sind innerhalb der Arbeit zu Künstlicher Intelligenz Evaluationsmethoden. Sie sind feste Standardwerkzeuge, mit denen verschiedene KI-Programme getestet werden. Innerhalb des Feldes der Gesichtserkennung sind das Datensets, bestehend aus einer bestimmten Menge an verschiedenen Bildern. Gesichtserkennungsanwendungen müssen dann, je nach Funktion, diese Bilder Personen zuordnen oder, wie im Fall des Gender Shades Projektes, Geschlechter zuordnen, da es Genderclassifier untersucht (vgl. ebd. S.2-3). Dies sind Anwendungen der automatischen Gesichtsanalyse, die anhand eines Fotos einer Person, deren Geschlecht bestimmen sollen (vgl. Hamindi et al. 2018, S.2).

Genderclassifier arbeiten mit einer binären Ordnung, die Bilder als männlich oder weiblich kategorisiert (vgl. Keyes 2018, S. 88:4). Der Benchmark ist im Falle des Gender Shades Projektes ein Datenset, bestehend aus 1270 Fotos von 1270 Parlamentarier:innen, pro Foto eine Person. Die Parlamentarier:innen wurden aufgrund der Zugänglichkeit der Bilder gewählt. Dabei stammten diese Bilder aus jeweils 6 Länderparlamenten, die aufgrund ihrer ethnischen Zusammensetzung

gewählt wurden: drei europäische und drei afrikanische Parlamente¹¹. Diese Parlamente wurden aufgrund ihrer ausgeglichenen binären Genderparität von den Wissenschaftler:innen gewählt (vgl. Buolamwini, Gebru 2018, S.7). Diese Fotos werden dann durch Buolamwini und Gebru kategorisiert, einmal entlang von Gender als entweder männlich oder weiblich, und zweitens aufgrund der Hautfarbe entlang des „Fitzpatrick Skintype“, einer Skala zur Bestimmung der Hautfarbe in 6 Kategorien (vgl. ebd.). Die Genderclassifier werden dann auf dieses Datenset angewendet und ihre Trefferquote, bzw. die Anzahl an Übereinstimmungen mit den vorher festgelegten Genderlabeln, und ihre Fehlerquote werden dann miteinander verglichen.

Somit konnten Buolamwini und Gebru aufzeigen, dass Genderclassifier nicht nur Frauen deutlich schlechter „erkannten“, ihnen somit kein korrektes Genderlabel zuordnen konnten, und dass zweitens Genderclassifier statistisch deutlich schlechter Gender zuordnen konnten, wenn es sich um „darker subjects“ handelte. Dies kulminierte in der höchsten Fehlerquote für alle drei Genderclassifier bei Fotos von „darker female subjects“ (Buolamwini, Gebru 2018, ebd.).

b.) Definition der Zielsetzung (FAT Algorithmen)

Das zweite Ziel dieser Studie ist es, einen inklusiven Ansatz für Datensets zu beschreiben. Dabei meint Inklusion die Aufnahme möglichst diverser Subgruppen (Gruppen, die aufgrund mehrerer soziodemographischer Marker wie Race und Gender) in Datensets für Training und Anwendung, um eine diskriminierende Gesichtserkennungssoftware zu vermeiden. Denn, wie das Gender Shades Projekt auch aufzeigte, waren bereits verwendete Benchmarks zur Evaluation von Genderclassifiern fast vollständig aus weißen Gesichtern und zumeist aus weißen männlichen Gesichtern erstellt worden. Diskriminierung, so die Annahme von Gebru und Buolamwini, beruhe dabei auf impliziten Vorurteilen. Datensets, homogene Designgruppen sowie das Ausschließen von spezifischen Kontexten bei der Entwicklung von solchen Programmen, tragen zu solchen rassistischen Ergebnissen bei, nicht einzelne bewusste Entscheidungen (vgl. Buolamwini 2017; Hoffmann 2019, S.903).

Die Zielsetzung lässt sich dabei einem größeren Wissenschaftszweig zuordnen, der Ethik rund um Künstliche Intelligenz, genauer gesagt dem sogenannten FAT-Ansatz (Fairness, Accountability, Transparency), welcher auch von Buolamwini explizit als Hintergrund ihrer Forschung genannt wird (vgl. Buolamwini 2017, S. 44).

In einem folgendem Schritt möchte ich diese Grundlage genauer beschreiben und ihre Kritikpunkte herausarbeiten. In einem zweiten Schritt möchte ich dann alternative Ansätze kritischer Forschung zu Diskriminierung und KI darlegen, die das Forschungsziel von Buolamwini und Gebru vor der

11 Diese Länder sind: Südafrika, Senegal, Ruanda, Schweden, Finnland, Island (vgl. Buolamwini, Gebru 2018, S: 7)

von ihnen aufgeworfenen Problemstellung, rassistischer und vergeschlechtlicher Unterdrückung, weitreichender fassen, als es der FAT-Ansatz vermag. Ich werde dabei speziell und in Anbetracht meiner Forschungsfrage auf Gender als Ungleichheitskategorie eingehen.

i.) AI Ethics

In den letzten 9 Jahren hat sich, innerhalb der wissenschaftlichen und wirtschaftlichen Gemeinschaft rund um Künstliche Intelligenz, das Wissen um das potentielle Risiko der Diskriminierung durch KI konsolidiert, in Form von Debatten um Ethik für KI (Hutchison, Mitchell 2019, S.49). Dies verstärkte sich noch einmal in den letzten 5 Jahren. Jobin et al. zeigten, in ihrer Studie zur Kartierung von KI-Ethik, dass 80 % der gefertigten Richtlinien für KI ab 2016 entstanden. 25 % davon stammen aus den USA (vgl. Jobin et al. 2019, S.391). Der Medienwissenschaftler Mark Andrejevic spricht deshalb auch von einem „ethical turn“ innerhalb der Debatte um KI (vgl. Andrejevic 2020, S.565). Dabei ist die Definition für diese Ethik wenig grenzscharf und wird unter Ethics of AI, Ethical AI, Data Ethic und ähnlichen Begriffen verhandelt (vgl. Powers, Ganascia, 2020, S.29). Die Philosophin Janina Loh, die mit ihrem Vorschlag zu Roboterethik einen Versuch unternimmt, Grenzen innerhalb dieses Feldes zu ziehen, definiert dabei Ethik wie folgt als Beschäftigung „einerseits mit den menschlichen Sitten, Bräuchen und Gewohnheiten, andererseits mit dem guten Leben, nimmt darüber hinaus die wissenschaftliche Reflexion menschlicher Praxis vor und beurteilt die Kriterien guten und schlechten Handels“ (vgl. Loh 2019, S.10).

Loh erklärt darüber hinaus, dass Maschinenethik hier eine Form der Ethik sei, die sich auch mit nicht menschlichen Akteurinnen befasse (vgl. ebd. 11).

Ethik für KI kann sich dabei am ehesten mit ihren spezifischen Themenfeldern von verwandten Ethiken abheben, ist jedoch oft konzeptionell stark an andere Bereichsethiken, wie der Medizinethik¹², angelehnt (vgl. Mittelstadt 2019). Ethik um KI befasst sich mit Themen, wie der Implementierung ethischer Grundsätze in der Technik, metaethischer Betrachtungen über den Gebrauch von KI oder empirische Studien über diskriminierende Auswirkungen von Technologie, wie unter anderem Buolamwini und Gebru's Gender Shades Projekt (vgl. Hagendorff 2020, S.99)

12 Medizinethik: „Principlism’ emerged from medicine as a theoretical moral framework joining traditional ethical standards with the requirements of practitioners, research ethics committees, and medical institutions for practical ethical decision-making.11 Principlism proposes four core principles that require specification and balancing in different decision-making contexts. Whereas principlism in medical ethics provides a common language to identify and conceptualise ethical challenges, and provides guidance for setting health policy and clinical decision-making.“ (Mittelstadt 2019, S. 2)

Jobin et al. beziehen Ethik dabei auf eine politische Funktion, die der Soft Laws. KI-Ethik habe dabei die Funktion, führenden Akteur:innen, sowohl staatlichen als auch privaten, als Richtlinien zu dienen. KI-Ethik hat keinen rechtsverbindlichen Charakter (vgl. Jobin et. al. 2019, S.389).

Als konkrete Ausformungen gilt hier der FAT-Ansatz. Ethik befasst sich dabei mit den Themengebieten der Fairness, Accountability und der Transparenz.

ii.) Fairness

Das Konzept der Fairness stellt hierbei eine Antwortmöglichkeit auf diskriminierende technologische Ursachen dar. Dabei ist Fairness eine informationstechnologische Zugriffsmöglichkeit, bei der das Problem und die Lösung in den zugrunde liegenden Programmen gesucht wird. Algorithmen, Datensets oder Ähnliches werden dabei als schadhaft gekennzeichnet und mittels statistischer Messungen, mathematischer Problemdefinitionen und Begrenzung sozialer Probleme auf die jeweils untersuchte Technik bezogen (vgl. Hoffmann 2019, S. 901-2).

Nach der Studie von Ben Hutchinson und Margaret Mitchell zur Entwicklung von Fairness als Kriterium für Künstliche Intelligenz, gibt es dabei vier Kriterien: individuelle, gruppenbezogene nicht vergleichende, gruppenvergleichende und correlative Fairness (vgl. Hutchinson, Mitchell 2019, S.52)

Buolamwini definiert Fairness demgegenüber wie folgt auch als Fairness, die einzelne Gruppen vergleicht (Männer/Frauen Schwarz/Weiß und dies auch intersektional also kategorieübergreifend). “For this thesis, fairness is defined as having comparable classification accuracy rates across intersectional subgroups” (Buolamwini 2017, S.20).

Vor diesem Hintergrund verweist Sarah Myers West in ihrer Kritik des Fairness-Diskurses innerhalb smarter Software auf die Gleichsetzung des Fairnesskonzepts mit Computertechnologie. So gebe es zwar einen transdisziplinären Austausch über Fairness innerhalb von AI, wie ihn Buolamwini in ihrem Papier anführt, in dem sie beispielsweise auf das Konzept der Intersektionalität der Juristin KimberléCrenshaw verweist, diese blieben jedoch weitestgehend untertheoretisiert in Anbetracht der Lösungskonzeption als akkurate Technologien, bzw. akkuratere Genderclassifier. Anders formuliert wird durch die mathematisch, informationstechnologische Definition des Problems der diskriminierenden Algorithmen, eine informationstechnologische Lösung notwendig. Ich möchte in der anschließenden Diskussion auf den Zwiespalt eingehen, der sich für das Gender Shades Projekt daraus ergibt. So rekurren Buolamwini und Gebru, wie aufgezeigt, auf ein breites die gesamte Gesellschaft betreffendes (spezifisch: die US-amerikanische Gesellschaft betreffendes) Problem mit einer technischen Lösungskonzeption.

So kritisiert Anna Lauren Hoffmann in ihrer Auseinandersetzung mit der Fairnesskonzeption von

Buolamwini und Gebru und deren Ziel antirassistische Politik zu betreiben: „However, both fall short in terms of addressing the institutionalization of (and liberation from) social hierarchy. For example, Buolamwini and Gebru’s (2018) stated aim of ‘increasing phenotypic and demographic representation and algorithmic evaluation’ (p.12) does not address justice issues arise from the institutional context within which facial recognition is employed a limitation Buolamwini [...] and Gebru [...] have emphasized elsewhere.” (Hoffman 2019, S. 907).

Ich möchte in der anschließenden Diskussion diese, von Hoffman aufgeworfene Zweiteilung von technischer Evaluation einerseits und politischer Kontextualisierung andererseits durch Gebru und Buolamwini, in Frage stellen und argumentieren, dass das Gender Shades Projekt seinerseits als solche Verbindung durchaus gewertet werden kann bzw. auf die Politiken des Gender Shades Projekts eingehen. Zuerst möchte ich jedoch noch die beiden anderen Aspekte der von Buolamwini und Gebru verfolgten Ziele erläutern Transparency und Accountability.

iii.) Accountability

Accountability erfreut sich in wissenschaftliche Diskursen derzeit nicht nur in Bezug auf KI sondern auch allgemeiner ausgedrückt im politikwissenschaftlichen Diskurs großer Beliebtheit, ist jedoch zumeist wenig definiert bzw. steht oftmals für andere breitere Ansätze politikwissenschaftlichen Arbeitens, wie Gerechtigkeit oder Demokratie (vgl. Bovens et al. 2014). Der Accountability-Forscher Mark Bovens kritisiert dabei eine nicht vorhandene Kohärenz der Begriffsdefinition, die neben der Fülle an Funktionen, die dieses Konzept je nach Studie einnimmt, eine grundlegende Problematik für die wissenschaftliche Arbeit darstellt (vgl. Bovens 2010, S. 947). Die Accountability-Debatte sei zwar ausgeprägt und fachlich differenziert, jedoch arbeiten die Wissenschaftler:innen größtenteils mit Eigendefinitionen, was vergleichende Arbeiten erschweren würde (vgl. ebd.). Ich möchte daher im folgenden den Ansatz, den Buolamwini und Gebru wählen, kurz erläutern und einbetten in bestehende Forschung zu Accountability und KI, beschränke mich aber, aufgrund meiner Forschungsfrage, vorrangig auf politikwissenschaftliche und feministische Ansätze.

Accountability ist, im Gegensatz zu Fairness oder Ethik, ein viel genutztes Konzept der wissenschaftlichen Diskussion im Hinblick auf KI-Produktion, da es weniger auf gemeinsame Werte verweist, sondern auf die Tatsache, dass es Akteur:innen geben muss, die für die Systeme, gleich in welcher Form, verantwortlich seien (vgl. Kroll 2020, S.12). Wie ich bereits im vorangegangenen Abschnitt beschrieben habe, bezieht sich das Konzept der Accountability oftmals auf Transparenz als eine Grundvoraussetzung. Durch das Sichtbarwerden fehlerhafter Techniken/Abläufe/homogener Datensets können Betroffene und/oder Entscheidungsträger:innen Verantwortung ableiten (vgl. ebd.

S. 11).

Ich folge dabei der Definition von Accountability von Joshua Kroll, der in seiner Studie zu KI und Accountability den Versuch unternimmt, dies Konzept hinsichtlich verschiedener Fachdisziplinen zu systematisieren. Accountability nach Kroll sei

„with respect to some entity – a relationship that involves reporting of information to that entity and in exchange receiving praise, disapproval, or consequences when appropriate. Successfully demanding accountability around an entity, person, system, or artefact requires establishing both ends of this relationship: who or what answers to whom or to what?“ (ebd. S. 4).

Dabei würden, nach Kroll, die einzelnen Disziplinen Verantwortung in unterschiedlichen Bereichen rund um KI verorten, so würden sich Politikwissenschaftler:innen verstärkt auf die Ebene der Mensch-Maschinen-Interaktionen und abstrakte Normen berufen, wie der Menschenwürde oder dem Gleichbehandlungsgrundsatz, wo hingegen Designer:innen verstärkt auf den Produktionsprozess oder auf die KI an sich abstellen würden, wenn es um den Ort der Verantwortungsübernahme geht (vgl. ebd. S.4). Daraus entwickelt Kroll ein Stufenmodell für Verantwortlichkeit innerhalb von KI-Systemen bzw. ein Model welches Abstraktionsgrade der Verantwortungsübernahme erlaubt.

„instrumentally, we may demand that the score is correct if and only if it adequately rates risk or rates risk in an equal way across demographic groups, holding the system’s performance to an objective and mathematical standard of correctness; at a systems level, we might hold the score to a standard of defensibility in litigation or another oversight mechanism; at a societal level, we might ask whether the distribution of risk elucidated by the score is the correct and morally appropriate distribution, a standard of fidelity to normative goals“ (ebd. S. 5).

Für Kroll beinhaltet dabei die Ebene des „oversight“ Möglichkeiten der Policy-Implementierung und eine Abstraktionsebene höher würden Fragen formativer Ordnungen und Gerechtigkeit verhandelt werden können.

Ähnlich multidimensional erklärt auch Buolamwini die Schwierigkeit der konkreten Implementierung von Accountability, im Hinblick auf Entwicklung von Genderclassifern als ein Produkt, welches sich aus vielen Arbeitsschritten zusammensetzt.

Buolamwini geht es dabei um eine komplexe Zeitlichkeit, die sowohl Output als auch den Designprozess im Hinblick auf Accountability parallel erfasst.

„Further still, if an unfair or harmful decision is made by an automated system, the question of accountability remains. Deciding who is accountable for an algorithmic decision is complicated by the distributed nature in which algorithms are combined and developed.“ (vgl. Buolamwini 2017, S.46)“

„Accountability must also start at the conceptualization stage when a construct or target variable is defined. When supervised com-

puter vision models are used to assess constructs like gender, the models have been trained to learn the visual displays of the construct“ (vgl. ebd. S. 100)

Das Konzept ist hierbei definitorisch untertheoretisiert und spiegelt die eingangs von mir beschriebene begriffliche Weite der Debatte rund um Accountability innerhalb der akademischen Literatur zu diesem Thema wieder (Bovens et al. 2010, S. 12). Hierbei wird das technische Tool zwar von Buolamwini als handelnde Akteur:in erkannt, jedoch verbleibt die Analyse auf einer oberflächlichen Betrachtung der Dualität von Mensch und Maschine, die beispielsweise auch Kroll in seinem aktuellen Versuch der begrifflichen Systematisierung von Accountability beibehält. Denn es verschränken sich Strukturen und Akteur:innen multidimensional, beispielsweise Artefakt (Gender-classifier) mit Prozessen (Design der Algorithmen, Design des Datensets im Gender Shades Projektes) und Akteur:innen (Betroffene, Entwickler:innen, Datenset). Jedoch werden diese Verschränkungen wiederum nicht weiter im Gender Shades Projekt verhandelt, was beide Autor:innen in einem kritischen Artikel 2020 bestätigen (vgl. Buolamwini, Gebru 2018, S. 12, Raji et al. 2020). In diesem schlagen die Autor:innen einen mehrstufigen Prozess der Kontrolle von Algorithmen vor, der sich aus verschiedenen Modulen zusammensetzt, wie der Evaluation des Nutzens der Technik, Checklisten vor dem eigentlich Design-Prozess und Governing-Methoden während der Entwicklung (vgl. Raji et. al. 2020 S. 38).

Diese breiter aufgestellten Ansätze für mehr Verantwortlichkeit innerhalb von KI-Systemen sind keinesfalls neu. Bereits 1996 schrieb die Wissenschaftler:in Helen Nissenbaum einen Artikel, der vor der Gefahr warnte, dass im Zeitalter ubiquitärer Computernutzung¹³. Verantwortlichkeiten schwer herzustellen und zu sichern seien, und zwar aufgrund der verteilten Herstellung von Software, Softwarefehlern (BUGs), alleiniger Schuldzuweisungen an den Computer und der Entkopplung von Haftung und Eigentumsrechten. So würden Firmen zwar die Eigentumsrechte von Software in Form von Patenten und Firmeneigentum reklamieren, jedoch Nutzer:innen dieser Programme, die Verantwortung für den Gebrauch zuschreiben (vgl. Nissenbaum 1996, S. 36). Sie schlug deshalb eine Reihe tiefgreifender Maßnahmen zur Sicherstellung von Verantwortbarkeit vor, die in der Erstellung von Richtlinien zur einfacheren Gestaltung von Software lägen (vgl. ebd. S. 37).

Die Frage der individuellen Verantwortung erschwert es, nach Buolamwini, an diesem Punkt zusätzlich, klarer zu formulieren, wie Lösungen auszusehen haben. Ein Problem, welches in der Forschungsliteratur als das Problem der vielen Hände beschrieben wird. Verantwortung ist dabei

13 Sociologically, ubiquitous computing may mean the decline of the computer addict. In the 1910's and 1920's many people "hacked" on crystal sets to take advantage of the new high tech world of radio. Now crystal-and-cat's whisker receivers are rare, because radios are ubiquitous. In addition, embodied virtuality will bring computers to the presidents of industries and countries for nearly the first time. Computer access will penetrate all groups in society. (Weiser 1991, S.104)

schwer zu adressieren, da sich moderne Softwareentwicklung nicht mehr auf eine Programmierer:in oder ein festes Team beschränkt, sondern aus einer Vielzahl an Akteur:innen, die KI produzieren (vgl. Kroll 2020, S. 1-4).

Entgegen dieser Maßnahmen möchte ich einen dritten Weg vorschlagen, der sich aus meinem gewählten Theorieverständnis ergibt; den Weg der Techno Science, die ich bereits vorgestellt habe als grundlegende theoretische Basis meiner Arbeit. Alle bis hierhin vorgestellten Ansätze gingen dabei von klaren disziplinären Trennlinien und der Unterscheidbarkeit von menschlichen und nicht menschlichen Akteur:innen innerhalb von KI-Systemen aus.

Der von mir gewählte Ansatz der posthumanistischen Performativität, nach Barad, wird in der folgenden analytischen Betrachtung noch einmal einen anderen theoretischen Ausgangspunkt bieten, jenseits von FAT-Ansätzen, zur Produktion gerechterer Genderclassifier.

Ich möchte zuerst jedoch noch einmal zusammenfassen, weshalb FAT-Ansätze unzureichende theoretische Rahmen sind, um erstens meine Forschungsfrage zu beantworten, aber allgemeiner warum sie unzureichend sind, auf die multiple Krise von KI einzugehen.

iv.) Transparency

Transparenz ist in dem Projekt Gender Shades zusammen mit Accountability ein Baustein für faire Algorithmen. Dabei definiert sie Transparenz wie folgt: „transparency as providing information on the demographic and phenotypic composition of training and benchmark datasets.“ (Buolamwini 2017, S. 14). Buolamwini schränkt Transparenz als Zielbestandteil stark ein. So macht sie zwar deutlich, dass Transparenz ein breites, umfassendes politisches Feld, dies jedoch nicht Ziel der Studie ist (vgl. ebd.)

Dabei wird Transparenz von zwei Grundbedingungen heutiger Software limitiert, einerseits durch die, schon von mir beschriebene, Patentierung von gewerblicher Gesichtserkennungssoftware. Durch das Patent an den untersuchten Systemen fallen wichtige Teilbereiche der Software unter das jeweilige Firmengeheimnis, wie beispielsweise der Source Code (vgl. Buolamwini 2017, S. 46). Des Weiteren gäbe es keinen Automatismus, der einen Algorithmus nur durch bloßes Aufheben des Patentschutzes einsehbar machen würde, ob eine weiterführende Interpretation möglich sei, so Buolamwini sei fraglich (vgl. ebd.). Die Systeme selbst werden immer komplexer und das Nachvollzie-

hen jeder einzelnen Operation von Geschlechtserkennungssoftware sei unmöglich. Die darauf aufbauenden Systemarchitekturen können den Entwickler:innen nicht mehr kausal zugeordnet werden (vgl. ebd.)

Die Medientheoretikerin Wendy Chun führt dieses Phänomen weiter aus und beschreibt, dass Computer und die dahinterliegende Software selbst immer transparenter werden, aber sich die Opaktizität dieser Tools immer weiter steigert. Der Ruf nach Transparenz so Chun ist ein Impuls der Kompensation und Verschleierung der Unmöglichkeit einer Transparenz, im Sinne der Beherrschbarkeit von Algorithmen bzw. Software¹⁴. Dabei ist das Paradoxon des immer durchsichtiger werdenden Computers mit seinen spezifischen Programmen produktiv und der Technologie inhärent.

„The Linking of [...] know ability with what is unknown, makes it a powerful fetish that offers its programmers and users alike a sense of empowerment, of sovereign subjectivity, that covers over -barely- a sense of profound ignorance.“ (Chun 2013, S. 18).

Chun lässt dieser kritischen Beobachtung um Transparenz jedoch ein differenziertes Fazit folgen. So sieht sie die Entwicklung der Software zu immer komplexeren Gebilden als umkehrbar und auch in vielerlei Hinsicht als Vorteil. Eine Rückkehr zur Schreibmaschine befürwortet sie nicht, stattdessen öffnet sie noch einmal die Wirkweise von Transparenz nicht nur als Wissen über, sondern als politische Strategie. Denn Transparenz könnte auch eine Form der Transformation sein. Als Beispiel verwendet Chun Projekte des MIT Medialabs, welche Daten, die sonst Staaten und Konzerne vorbehalten wären für Bürger:innen öffnen. Öffnen meint hier einerseits Sichtbarmachen, also Transparenz, aber auch offen (open) im Sinne des Konzepts der Open Data¹⁵.

„They give users the collective power to transform these database and these debates over what counts as evidence“ (vgl. ebd. S. 93).

Diese Konnotation von Transparenz, und der damit einhergehenden Potentiale von Subjektivierung, Politisierung und Handlungsmacht, wird, in Anlehnung an die sich anschließende vertiefende Analyse von Gender, mittels Butler eine Verbindung zwischen Buolamwinis Zielsetzung und queertheoretischen Überlegungen liefern. Die sich einerseits auf Buolamwinis Ziel stützt und andererseits das Konzept Transparenz vor diesem Hintergrund erweitert.

14 Chun spricht hier von Software als spezifischem theoretischem Begriff und spezifischer Anwendungsform von Algorithmen. Algorithmen in der sozialwissenschaftlichen Forschung und der Data Science meint demgegenüber, dass der Algorithmus das allgemeine Prinzip hinter einer jeder Software ist (vgl. Bucher 2018, S.30).

15 „any kind of data and content - sonnets to statistics, gene to geodata - that can be freely used reused and redistributed.“ (Open Data Foundation [nach Bath 2013, S. 85]). Die Frage der Freiheit ist jedoch umstritten, gerade im Hinblick auf Bezahlung der jeweiligen Programmierer:innen (vgl. Coleman 2009, S. 38-40)

2.) Emperie

Das Ziel des Gender Shades Projekt ist vorrangig eine Evaluierung bestehender Gesichtserkennungsanwendung und die Aufdeckung und Behebung sexistischer/rassistischer Bias in der Software (vgl. Buolamwini, Gebru 2018, S. 1). Dieses Ziel möchte ich unter der theoretischen Annahmen der Genderregulierung von Butler genauer analysieren.

Augenscheinlich ist, dass, ausgehend von der Evaluationsidee des Projektes Gender Shades, Gender von den Autor:innen als Binarität gefasst wird. Ich halte in Bezug auf meine Forschungsfrage somit fest: Geschlecht ist als Binarität in das Gender Shades Programm als Evaluationsfaktor eingelassen.

„In this work we use the sex labels of „male“ and „female“ to define gender classes since the evaluated benchmarks and classification systems use these binary labels“ (Ebd. S3).

Dennoch wird diese Binarität von Buolamwini auch als artifiziell dekonstruiert, indem die Wissenschaftlerin auf die komplexe Realität von Gender in verschiedenen nationalen Kontexten aufmerksam macht, welche sich nicht in einer Aufteilung in Mann und Frau auflöst (vgl. Buolamwini 2017, S. 100). Ich möchte in einem ersten Schritt bei dieser Binarität bleiben und diese als konkrete Genderregulierung fassen. In einem zweiten möchte ich spezifischer auf den scheinbaren Bruch von Buolamwinis queerem Genderverständnis und der binären Designentscheidung eingehen. Ich werde darlegen, dass die Klassifizierung von Gender als female/male auf eine symptomatische Verengung von Geschlecht als Merkmal des Subjektes zurückzuführen ist. Hierbei möchte ich, um bei Butler zu bleiben, aufzeigen, inwiefern die Verwendung von binären Genderkategorien in der Auditstudie eine hybride Übergangsform von „konkreter Genderregulierung“ zu „Gender als Norm“ ist. Ich werde dabei diese Hybridisierung auch als Verschränkung von Subjekt, Forschungsgegenstand und Forschungsprozess analysieren. Ich greife dafür, wie bereits angekündigt, auf das Konzept der posthumanistischen Performativität von Karen Barad zurück.

a.) Bias als konkrete Genderregulierung

In einem ersten Schritt werde ich das Konzept der konkreten Genderregulierung einführen, ich werde anhand dieses Konzeptes eine theoretische Perspektive eröffnen, die es mir erlaubt, diskriminierende Software gendersensibel zu analysieren.

Butler geht in ihrer Konzeption von Gender von einer machtvollen Unterteilung der Geschlechter aus, die sich in konkreten Regulierungen und empirischen Instanzen zeigen (vgl. Butler 2011, S.17).

Welche Regulierungsmechanismen könnten das im Bereich des Gender Shades Projektes sein?

Ich möchte im Folgenden zwei konkrete Konzepte der binären und darauf aufbauend sexistischen Genderkategorisierung im Bereich der datengetriebenen Gesichtserkennung aufzeigen. Diese sind

im Gegensatz zu den vorangehend diskutierten FAT-Ansätzen, empirisch belegte und genderspezifische, theoretische Ansätze.

Dabei werde ich den Gender Data Gap vorstellen, ein Konzept der Journalistin Caroline Criado-Perez (vgl. Criado-Perez, 2020, S. 11-12). Dieses Konzept ermöglicht mir die theoretische Einordnung des Gender Shades Projektes in einen übergeordneten gesellschaftlichen Zusammenhang, der den Ausschluss von marginalisierten Gruppen durch Unter- oder Nichtrepräsentation mittels Daten beschreibt (vgl. ebd. S.41). Dann werde ich den Begriff Privileg Hazard, ein Konzept der Datenwissenschaftler:innen Catherine D'Ignazio und Lauren Klein, dass sich vor allem auf die Homogenität der Entwickler:innenteams in der KI-Produktion und deren Folgen beschäftigt. In einem dritten Schritt werde ich auf die Exklusion von genderqueeren Menschen eingehen, um aufzuzeigen, inwiefern sexistische Strukturen nicht nur Cis-Frauen betreffen und wie sich Genderregulierung auch als Regulierung von Zweigeschlechtlichkeit in den Kontext der KI-Produktion einschreibt.

i.) Gender Data Gap

Nach Butler lassen uns Genderregulierungsinstanzen sichtbar werden in gesellschaftlichen Feldern, wenn wir gewisse normalisierte Positionen einnehmen, wie die Rolle der Ehefrau oder des Ehemanns bei einer Ehezeremonie (vgl. Butler 2011, S.84-85). Diese geschlechtsspezifische Sichtbarkeit lässt sich, so werde ich es zeigen, aber auch im Feld der Big-Data-Applikationen ablesen. Caroline Criado-Perez prägte dafür den Begriff des Gender Data Gap:

„Es ist eine geschlechtsbezogene Lücke in den wissenschaftlichen Daten, eine Gender Data Gap [...] Neu ist aber der Kontext, in dem Frauen weiterhin „das Andere“ bleiben – eine Welt, die immer stärker auf Daten basiert und immer stärker von Daten beherrscht wird.“ (Criado-Perez 2020, 11-12).

Das Fehlen von Frauen in verschiedensten Datensets (klinische Studien; Bilderkennungsanwendungen, Bewerbungsalgorithmen) ist politisch konnotiert, da es vermeidbar wäre, es stellt technisch keine Unmöglichkeit dar, Fotos von weißen und Schwarzen Frauen einzubeziehen (vgl. Buolamwini, Gebu 2018). Sichtbarkeit wird so, nach Butler, zu einer Form der Macht des Intelligiblen (vgl. Butler 1991, S. 72). In diesem Fall drückt sie sich durch Datensamples aus; denn, wie bereits beschrieben, leiten sich aus den darauf aufbauenden Algorithmen konkrete neue Handlungen ab. Dabei bilden Gesichtserkennungssoftware und Gender-Erkennungssoftware ein für die Politikwissenschaft spezifisches Feld der Macht des Intelligiblen, in dem diese Technologien Fragen der Staatsbürger:innenschaft und Grenzpolitik berühren (vgl. Buolamwini 2017, S. 29). So erläutert der Queertheoretiker Tobi Beauchamp in seiner Arbeit zu Körperscannern an US-amerikanischen Flug-

häfen,

„that such imaging techniques are rooted in racializing and gendering processes that produce and distinguish the categories of citizen and terrorist, normative and deviant. These images circulate in political contexts that necessarily cast certain bodies as violent or as threateningly anomalous in an effort to stabilize the notion of healthy, safe bodies“ (Beauchamp 2019, S. 76.)

Er verweist hier auf den Mechanismus der Normalisierung, der einerseits den cis-geschlechtlichen Körper als Norm etabliert und somit genderqueere Körper extrem sichtbar werden lässt, als Gefahr und Antithese zur Staatsbürger:innenschaft (Butler).

Das Gender Shades Projekt ist aus diesem Grund auch von Buolamwini und Gebru in ein, wie bereits erwähnt, breiteres politisches Vorhaben eingebettet worden. Die Wissenschaftler:innen setzten sich dabei mit den Unterdrückungsmechanismen auseinander, die sich aus der Abwesenheit von Schwarzen Frauen in den Datensets der untersuchten Genderclassifier von IBM, Microsoft und Face++ ergeben (vgl. Buolamwini 2017, S. 17). Die konkrete Unsichtbarmachung von Schwarzen Frauen und Women of Colour, durch die wiederholte Entscheidung gegen die Aufnahme von Bildern dieser marginalisierten Personen in die Datenbanken, erzeugt eine prekäre Position innerhalb der immer stärker zunehmenden digitalen Gesellschaft. Fehlende Daten von marginalisierten Frauen führen so zu politischen und in der Lebenswelt relevanten Folgen, in verschiedensten Bereichen heutiger Gesellschaften. Fehlende Digitale Sichtbarkeit kann beispielsweise zu fehlender medizinischer Betreuung führen. So ist beispielsweise die Müttersterblichkeit unter Schwarzen Frauen in den USA deutlich höher als die weißer Frauen, dennoch fehlen genaue Daten, die die Gründe für diesen Misstand offen legen (vgl. Buolamwini, Gebru 2018, S. 2, Klein, D'Ignazio 2020, S.22). Dies führt zu einer ambivalenten Situation, die einerseits in Hypersichtbarkeit und somit konstante Überwachung marginalisierter Gruppen, vor allem BIPoCs und genderqueere Personen, führt, andererseits aber zu fehlenden Daten und somit fehlender (digitaler) Sichtbarkeit dieser Personengruppen (vgl. Beauchamp 2016, S.104- 105, Criado-Perez 2020, S.92-95). Der Gender Data Gap erklärt dabei eine asymmetrische Verteilung von Sichtbarkeit.

Einerseits ist Sichtbarkeit als Repräsentation in Datensets eine Art der Produktion von Rechten und Privilegien in verschiedenen Formen wie die Berücksichtigung von Bedürfnissen, beispielsweise entlang der Kategorie Gender oder aber individueller Zugänge für repräsentierte Menschengruppen, indem sie Zugang zu bestimmten Leistungen erhalten (vgl. Criado-Perez 2020, S. 30) Andererseits kann diese, in Form von Überwachung durch eine übergeordnete Instanz, wie beispielsweise staatlicher Strukturen, zu Repressionen und Gewalt führen (vgl. Browne 2015, S. 38-39).

Der Gender Data Gap zeigt somit auch die Privilegien, die mit erhöhter Intelligibilität einhergehen.

Diese beschreibt Butler als den produktiven Aspekt von Macht, da Macht nicht nur verhindert sondern auch ermöglicht (vgl. Butler 2011, S.148-9). Der Gender Data Gap verweist damit auf die vergeschlechtlichte Machtdimension, die sich durch Privilegien des Sichtbarwerdens innerhalb von Datensets, in einer immer stärker auf Daten beruhenden Gesellschaft ergeben. Diese Sichtbarkeit erlaubt dabei Privilegien, die sich in Form naturalisierter Subjektpositionen ausdrücken, wie beispielsweise eine männlich weiße Designer:innen-Norm für die Gestaltung öffentlicher Plätze (vgl. Criado-Perez 2018, S. 80-85). Simone Browne spricht vor dem Hintergrund der Gesichtsanalyse deshalb auch von einer

„structured violence of ‘the dialectics of recognition.’ This prototypical whiteness is one facet of the cultural and technological logic that informs many instances of the practices of biometrics and the visual economy of recognition and verification that accompanies these practices.“ (vgl. Browne 2015, S.110).

Nach Browne ist diese Dialektik Ausdruck der Ungleichverteilung von Macht und Möglichkeit der Einflussnahme auf diese Technologien. Weißsein wird so zu einer grundlegenden Bedingung von Überwachungstechnologie (vgl. ebd.). Ich möchte diese „Prämisse“ des Technikdesigns im folgenden näher und spezifisch innerhalb der KI-Produktion anhand des „Privilege Hazard“ erklären (D'Ignazio, Klein 2020, S. 28).

ii.) Privilege Hazard

Im Kontext der Data Science weisen D'Ignazio und Klein diese soziale Macht der Sichtbarkeit als „Privilege Hazard“ aus (vgl. D'Ignazio, Klein 2020, S. 28).

„the phenomenon that makes those who occupy the most privileged positions among us - those with good education, respected credentials, and professional accolades - so poorly equipped to recognize instances of oppression in the world.“ (ebd.)

Hierbei ist wichtig zu betonen, dass es sich um eine Definition handelt, die subjektive Intention nicht als die Ursache begreift. Vielmehr handelt es sich nach D'Ignazio und Klein um Prozesse, die sich in konkreten Situationen zeigen, aber auf eine dahinterliegende gesellschaftliche Machtstruktur, wie Sexismus und Rassismus, verweisen (vgl. ebd.). Dabei sei eine hegemoniale weiße Männlichkeit prägend für die empirischen Konstellationen, sprich für die konkreten Gruppen und nicht (Cis-)Männer. Kulturelle Narrative, vermittelt durch Popkultur, erzeugen dabei Vorstellungen von dem Ideal eines Computerwissenschaftlers, der weiß und männlich ist und den Zugang für eben jene Menschen sehr offen und durchlässig hält (vgl. Ensmenger 2015, S. 62-64; Eglash 2002).

Dieser Prozess lässt sich durch die enorme Gestaltungsmacht einer relativ kleinen weißen männlichen Gruppe von Datenwissenschaftlern anhand einer Vielzahl an empirischen Studien nachzeichnen (vgl. Ensmenger 2010, Misa et. al. 2010, Margolis, Fisher 2003, Webb 2020). Die Entscheidung für oder gegen gendersensible Datensamples würde von dieser Gruppe abhängen. So machen innerhalb der KI-Forschungsabteilung von Facebook lediglich 15 Prozent Frauen, bei Google 10 Prozent, 2,5% der Programier:innen bei Google¹⁶ sind Schwarz (vgl. D'Ignazio, Klein 2020, S. 27, West et al. 2019, S.5). Buolamwini fand demgegenüber heraus, dass 78 Prozent der Gesichter innerhalb der untersuchten Samples der großen kommerziellen Gesichtserkennungssoftware Männern zu zuordnen sind (vgl. D'Ignazio, Klein 2020 S. 27; 29). Diese Entscheidungen würde jedoch eben nicht aus einer intendierten Handlung heraus entspringen, sondern aus der Ignoranz der eigenen Positionalität, mit Butler könnte Mensch hinzufügen: und der ständig wiederholten Verunsichtbarung von FLINT-Personen und BIPoC (vgl. ebd. 28). So verweisen D'Ignazio und Klein auf die Doppeldeutigkeit des Phänomens „Privilege Hazard“, das sich in konkreten empirischen Beispielen, wie den zitierten Geschlechterverhältnissen zeigt, jedoch als allgemeiner Wesenzugs dieses Feldes aufgefasst wird.

„The privilege hazard occurs at the level of the individual [...] but it is much more harmful in aggregate [...] So it matters deeply that data science and artificial intelligence are dominated by elite white men because it means there is a collective privilege hazard so great that it would be a profound surprise if they could actually identify instance of bias priori to unleashing them onto the world.“ (ebd. S.29).

Dieses hier von den Wissenschaftler:innen angerissene Aggregieren würde ich in diesem Kontext komplexer einordnen und stärker noch unter dem Gesichtspunkt der Normierung nach Butler fassen. Ich werde im weiteren Verlauf erläutern, dass nicht einfach nur das Vorhandensein vieler weißer Männer zu einer „männlichen“ Norm in der Geschlechtererkennung wird. Vielmehr möchte ich das Aggregieren selbst analysieren, um die Antwort auf meine Forschungsfrage spezifischer zu fassen, genauer gesagt die Binarität von Gender als Prozess zu dekonstruieren. Bis hierhin habe ich einerseits die ambivalente Sichtbarkeit durch die und innerhalb der KI-Produktionsverhältnisse erläutert und mich auf den „Privilege Hazard“ als empirische Referenz für eine hegemoniale weiße Zweigeschlechtlichkeit bezogen. Ich möchte in einem letzten Schritt der Hervorhebung konkreter Genderregulierung auf die Exklusion von genderqueeren Menschen innerhalb der von Gebru und Buolamwini untersuchten Genderclassifier verweisen, um dann auf meine theoretischen Überlegungen hinsichtlich Barads posthumaner Performativität zu kommen.

16 Ich habe bereits in der Einleitung auf die Problematik dieser Zahlen, bezüglich ihrer Akkuranz, hingewiesen.

iii.) Queerfeindlichkeit und Genderclassifyern

Ich möchte an dieser Stelle exemplarisch aufzeigen, inwiefern auch Transpersonen und genderqueere Menschen in Genderclassifyer eingebunden sind. Denn nach Butler funktionieren Geschlechter außerhalb der Mann/Frau-Binarität, als Grenzmarkierungen.

„Weicht man von der Gender-Norm ab, bringt man gleichzeitig ein Beispiel für eine Abweichung hervor. Diese Abweichungen können regulatorische Mächte [...] sehr schnell nutzen, um ihren eigenen, fortwährenden regulatorischen Eifer zu stützen.“ (Butler 1991, S.91)

Os Keyes arbeitet in einer Audit Literaturstudie zu Genderclassifyern eben jene regulatorische Funktion von Gender heraus (vgl. Keyes 2018, S.88:7). So stellt Keyes, anders als Buolamwini, nicht Binarität als Grundlage der Kategorie Gender in den Mittelpunkt, sondern Binarität als transgender und genderqueeres exklusives Moment. Die Verwendung von Gender in dieser Form produziert ihre eigene Wahrheit

„if a HCI (Human Computer Interaction) researcher's method only allows them to gather data for a binary model of gender, they are going to end up with research that presumes gender is a binary.“ (ebd. S.88:12).

Genderbinarität wird in Keyes Analyse zu einer Arbeitsgrundlage, die die Norm innerhalb des Feldes der Geschlechtererkennung darstellt und durch Wiederholungen festigt. Sie stellt eine statistische Operationalisierungsgrundlage dar, nach der Genderclassifyer arbeiten können, nämlich, dass Menschen sich immer nach male/female einordnen lassen. Geschlecht wird normalisiert und somit für diese Anwendungen standardisiert. Nach Keyes sei diese Vorstellung bei den Wissenschaftler:innen, die sich mit dieser Technologie befassen, so selbstverständlich, dass sie Geschlecht nicht näher definieren, sondern Genderbinarität, als Prämisse ihrer Arbeit begreifen (vgl. ebd. S. 88:3, Schlesinger et al. 2017, S. 5419), Genderqueere Menschen werden so zu einer marginalisierten und immer wieder ausgeschlossenen Gruppe „gemacht“. So fanden beispielsweise Schlesinger et al. heraus, dass es lediglich 3 Paper auf HCI¹⁷ Konferenzen zu dem Thema Genderqueer und Transgender gegeben hat.

Diese von Keyes vertretene Idee der Performativität der Genderbinarität als Forschungsgrundlage für die Produktion dieser Technologie, ist forschungsethisch für mich relevant. Diese Feststellung erkennt an, dass einerseits Geschlecht in diesem Forschungsfeld als strikt trennbare und binäre Kategorie verhandelt wird, aber auch eine Entscheidung gegen genderqueere Menschen und Transpersonen beinhaltet. In einem nächsten Schritt möchte ich diesen Standpunkt als Prämisse für eine ge-

¹⁷ HCI meint Human Computer Interaction. Es ist ein Subfeld der Computerwissenschaften und behandelt voranging Technologien, wie Genderclassifyer (vgl. Keyes 2018, S.88:2).

nauere Analyse der Performativität von Gender in Buolamwinis Studie nutzen. Ich werde so die eben erwähnten konkreten Instanzen der Genderregulierung in Butlers Konzeption als Normierungsprozesse zusammenführen.

b.) Genderperformance im Gendershades Projekt

Keyes Kritik an der Binären Struktur von Gender in der Forschung zu Genderclassifier kann im meinem Theorierahmen als Konkretisierung von Butlers Konzept, der Performativität von Geschlecht, gelten. Ich möchte dies noch einmal ausformulieren.

Im Hinblick auf Gender als Norm ist nach Butler die ständige Wiederholung der Stabilisierungsmechanismus, der die Zweigeschlechtlichkeit als Notwendigkeit und naturalisierten Fakt erscheinen lässt (vgl. Butler 1991, S. 205). Für die von mir untersuchte Studie kann hier die Evaluierung (die Zielsetzung des Gender Shades Projektes) als performativer Akt der Geschlechterbinarität gelten. Also das immer währende Anwenden der Genderclassifier, die Geschlecht binär labeln und die Kategorisierung der Bilder nach Geschlecht in Gebru und Buolamwinis Benchmark (vgl. Buolamwini, Gebru 2018, S. 6).

Dabei ist die Performativität auf zwei Ebenen verteilt. Erstens wird durch das binäre neue Datenset des Gender Shades Projektes wieder Zweigeschlechtlichkeit als Prüfstein eingeführt, also als Norm etabliert. Fotos von Abgeordneten werden von den Forscher:innen als männlich/weiblich etikettiert und später als Prüfstein für die Akkuranz der Genderclassifier genutzt (vgl. Buolamwini, Gebru 2018, S. 5). Zweitens wird durch das Prüfen, also das verwenden der Classifier, Binarität wiederholt und scheinbar verfeinert. Das explizite Ziel ist es ja, Männer und Frauen akkurater zu „erkennen“, so dass die untersuchten Classifier in Zukunft mit höherer Wahrscheinlichkeit Männer und Frauen unterscheiden können. Die Norm der Binarität wird hierbei also zweiteilig performativ in die Genderclassifier eingeschrieben, durch verschiedenen Akteur:innen und auf verschiedenen Ebenen.

So weist die Idee des Gesetzes in Butlers Konzeption von Geschlecht ein Funktionsäquivalent zu der Einschreibung von Gender bei Buolamwini auf.

„Unweigerlich „produziert“ die Rechtsgewalt, was sie (nur) zu repräsentieren vorgibt. Demnach muß es der Politik um diese Doppelfunktion der Macht gehen, nämlich um die juristische und die produktive [...] Es genügt also nicht, zu untersuchen wie Frauen in Sprache vollständiger repräsentiert werden können. Die feministische Kritik muß auch begreifen, wie die Kategorie „Frau(en)“, das Subjekt des Feminismus, gerade durch jene Machtstrukturenhervorgebracht und eingeschränkt wird.“ (vgl. Butler 1991, S. 16).

Für meine Analyse des Gender Shades Projektes bedeutet dies zweierlei. Einerseits müssen die von

Buolamwini und Gebru verwendeten binären Kategorien von mir ebenfalls in die kritische Analyse der Funktion von Geschlecht mit einbezogen werden, da sie diese, um Butler zu zitieren, produzieren. Zweitens muss ich, um die Binarität als artifiziell und somit als gesellschaftlich bzw. in dem Fall des Gender Shades Projekts als soziotechnisch konstruiert, zu dekonstruieren, erläutern wie die Kategorien female/male konstruiert werden, sowohl durch Buolamwini und Gebru als auch durch die untersuchten Genderclassifier.

Schematisch lässt sich mein Interpretationsvorhaben des Gender Shades Projektes alstheoretische Spiralbewegung fassen. Mein Startpunkt ist dabei der Gender Data Gap als Problemstellung und Ausgangsvorraussetzung des Gender Shades Projektes. Buolamwini und Gebru, so meine Lesart der zugrunde gelegten Studien, entwerfen Datensets, die sowohl hinsichtlich Gender als auch Race durch die Wissenschaftler:innen ausgeglichen konzipiert wurden, so dass das PPB einer der wenigen Benchmarks für die Evaluation von Genderclassifier ist, das auch Schwarze Menschen und vor allem Schwarze Frauen in statistisch signifikanter Weise repräsentiert, nämlich zu ungefähr 25% innerhalb des Datensets. In einem dritten Schritt testen Buolamwini und Gebru drei verschiedenen Genderclassifier, die das Datenset, bzw. jedes Bild im Datenset, als Mann oder Frau kategorisieren. In der nächsten Phase müssten neue Datensets für das Training der Genderclassifier erstellt werden, um Bilder akkurater einordnen zu können. Dieser Prozess würde so von neuem beginnen. Diese „Verbesserung“ ist jedoch Spekulation bzw. trifft nicht zu, wie Buolamwini und Gebru in einem selbstkritischen Artikel offen legen (Raji et. al. 2020, S. 6). Dennoch möchte ich bei dem Bild der Spirale bleiben, da unabhängig von der Evaluation, immer neue Bilder nach Geschlechtern sortiert werden, so lange Genderclassifier angeboten werden, sei es als kommerzielle Software, wie bei den von Buolamwini und Gebri untersuchten Anwendungen, oder durch wissenschaftliche Bemühungen, diese Technologie zu verfeinern bzw. anwendungsreif zu gestalten. Die Spirale drückt somit in diesem Fall die zeitliche Reartikulation von Gender durch die algorithmische Geschlechtererkennung aus, bzw. die immer wieder neu zu produzierende Zweigeschlechtlichkeit. Anders formuliert, Genderclassifier produzieren aufgrund ihrer Funktionslogik immer wieder eine Zweigeschlechtlichkeit. So formalisiert möchte ich auf zwei Grundvoraussetzungen für dieses Verständnis der Performativität von Geschlecht nach Butler aufmerksam machen, die die vollständige Anwendung auf das Projekt von Buolamwini und Gebru verunmöglichen: Erstens das menschliche Subjekt als Prämisse von Geschlechterperformativität und zweitens, darauf aufbauend, dieses als analytischen Fixpunkt dieser Theorie. Dabei werde ich zeigen, dass Buolamwini, genau wie Butler, am Subjekt als Locus politischer Intervention festhält, aber die spezifische technische Wirkweise künstlicher Intelligenz diesem Verständnis Grenzen setzt. Diese Grenzen möchte ich im Anschluss mittels des Konzeptes der posthumanistischen Performativität sichtbar machen und analysieren, um aufzeigen

zu können wie eine emanzipatorische Umsetzung des Gender Shades Projektes hinsichtlich queerer Geschlechtlichkeit gelingen könnte. Ich werde in meinem zweiten Kapitel auf die Grenzen eines inklusiven Ansatzes eingehen, aufgrund der bereits vorgetragenen Probleme hinsichtlich der staatlichen Überwachung marginalisierter Gruppen.

Doch zurück zu Butlers Fokus auf das Subjekt und auf Performativität. Butlers Theorie ist eine Theorie der Subjektivierung. Genauer formuliert weist, nach Butler, die Kategorie Geschlecht per se auf eine Subjektivierungsstrategie hin, welche das Subjekt als Mensch fasst und die Macht als soziale Macht (vgl. Butler 1991, S.30-31). Die Verwendung von Geschlecht erzeugt Subjekte.

Buolamwini setzt ebenfalls auf menschliche Subjekte, sowohl als Verantwortliche für die Produktion von Genderclassifier und als Betroffene dieser Technologie (vgl. Buolamwini 2017). Darauf aufbauend begreift Buolamwini ihr normatives Eingreifen in bestehende sexistisch, rassistische Praxen als menschengemacht und auch von Menschen wandelbar. Dies lässt sich durch die Wahl der Ethik und aus der Spezifikation der Verantwortung als ethischen Wert ableiten, welche beide auf den Menschen als Akteur:in verweisen (vgl. ebd. S. 17-18).

Zweitens benötigt diese intendierte menschliche Komponente in Butlers Theorie einen fixierten Standpunkt der Beobachtung bzw. Analyse. Die Subjektivierung ist das immer wiederkehrende Moment der Genderperformance, und alle Handlungen, Gesetze und Normen führen an diesen Punkt. Das Subjekt wiederum ist der empirische Referent der andauernden Normierung. Ähnlich eines Zuges in Bewegung, der nur als bewegter Zug wahrgenommen werden kann, wenn Mensch einen festen Referenzpunkt hat, braucht Butlers Idee der Performativität einen Fixpunkt, an Hand dessen sich die Performativität von Gender ablesen lassen kann. Im meinem Schema braucht es die maschinelle und menschliche Einteilung/Unterteilung von Männern und Frauen, um wiederum die Kategorien, die die Genderclassifier verwenden, stabil zu halten. Buolamwinis Kategorisierung von „female“ und „male“ in der algorithmischen Praxis und ihr Verweis auf andere gegenderte Realitäten sind mit meiner theoretischen Präzisierung des fixen Referenzpunktes, so nicht nur folgenlose, Antwort auf bereits existierende methodische Zwänge der kommerziellen Genderclassifier, sondern Teil der analytischen Notwendigkeit ein menschliches Subjekt zu erhalten (vgl. ebd. S. 82). So erkennt sie einerseits die Beweglichkeit bzw. die Performativität von Gender an. Andererseits demonstriert Buolamwini die von Butler herausgearbeitete, funktionale, normierende Wirkung von Geschlecht, indem Buolamwini und Gebru mit dem Gender Shades Projekt wiederum den Akt der zweigeschlechtlichen Einsortierung ausführen. Diese Wirkweise lässt zwar andere Geschlechteridentitäten theoretisch im Gender Shades Projekt sichtbar werden, wie die angeführten Beispiele von Buolamwini, jedoch bleiben diese von der Produktion von Geschlecht durch die Genderclassifier ausgeschlossen und können somit auch nicht als Kategorien dieser Technologien

aufgenommen werden.

Vergeschlechtlichte Subjekte jenseits von cis-geschlechtlicher Binarität werden andererseits zu Fixpunkten, anhand derer sich die Performativität messen lässt, in diesem Fall die Verstetigung einer cis-geschlechtlichen Diskriminierung von Schwarzen Frauen innerhalb der automatischen Geschlechtererkennung. Wie Butler, in Anlehnung an Foucault, schreibt, sind diese außenstehenden Personen konstitutiv für Gender als Norm, da sie eine Grenze markieren anhand derer „Normalität“ zu Tage tritt und Männer und Frauen in Gesellschaft eintreten könne (vgl. Butler 1991, S. 17).

An dieser Stelle möchte ich Butlers Theorie für die Analyse des Gender Shades Projektes erweitern, da es, so werde ich darlegen, die spezifische Konfiguration von Menschen, Geschlecht und Künstlicher Intelligenz innerhalb des Gender Shades Projektes für meine Analyse erforderlich macht, Performativität des Geschlechts, von ihrem exklusiv menschlichen Kern zu lösen, die ich herausgearbeitet habe.

c.) Von der Gender Performance zu Posthumanistischer Performance

Um die Performativität von Geschlecht in einem automatischen Prozess, wie dem automatischen Genderclassifier zu fassen, werde ich in diesem Teil Karen Barads Erweiterung von Butlers Genderkonzeption einführen, dafür werde ich zuerst auf die Besonderheit meines Analysegegenstands eingehen. Ich werde erläutern warum posthumanistische Performativität für die Untersuchung von Gender in KI eine produktive Perspektive bietet.

Ich beginne mit der Spezifizierung von KI als Feld der Ges Corinna Bath, die für ihre Fallanalyse des „semantic Webs“¹⁸ den Begriff des „Hybriden Objektes“ als Ausgangspunkt kritischer Geschlechtertheorie aufgreift. Geschlecht so Bath würde in den komplexen neuen Infrastrukturen der Software Entwicklung in kollektive Praxis umgesetzt und Wissen als konstruiert und veränderbar durch die Entwickler:innen verstanden, eine Fokussierung auf Machtverhältnisse sei notwendig, doch können Wissenschaftler:innen nicht mehr von einem einzelnen männlichen, weißen Subjekts als Ursache von schadhaften KI-Systemen ausgehen (vgl. Bath 2013, S. 93). Bath verweist dabei auf die Schwierigkeiten geschlechtertheoretischer Analysen dieser neuen technischen Artefakte

„Gerade in den maschinell erstellten Ontologien sind Ein- und Ausschlüsse aus epistemologischer wie geschlechtertheoretischer Perspektive schwer einzuschätzen, da wir es hier mit einer techno-sozialen bzw. posthumanen Wissensmodellierung zu tun haben.“ (ebd.)

18 Webinhalte, die implementiert werden, die maschinell lesbar sind, um so intelligenter Software zu ermöglichen, nutzer:innenspezifisch Wissen anzubieten (vgl. Bath 2013, S. 69-70)

Um vergeschlechtlichte Exklusionen zu erfassen und analysieren zu können, braucht es eine neue methodische Ausgangslage. Bath schlägt dafür die Idee des Hybridobjektes, als Prämisse vor. Dieses sei ein Grenzobjekt welches stets zwischen informatischer Formel und realen Handlungen oszilliere (vgl. Bath 2013, S. 98). Dabei würde nicht nur die Trennung zwischen materiell und geistig brüchig sondern auch die Grenze zwischen

„Deskription und Konstruktion. Ein Hybridobjekt nieder zu schreiben wird nicht nur als Deskription, sondern genauso als (Auf)bauen und damit Konstruktion interpretiert. (Siefkes et al. 1998, S.3 [nach Bath 2013, S.99]).

Diese Trennungslinien, die das Hybridobjekt einrahmen, sind dabei immer politische Grenzen, sie bestimmen den Handlungsraum der Software. Auf das Gender Shades Projekt bezogen, habe ich bei der Zielvorstellung auf die politische Dimension der brüchigen Grenzen von Deskription und Konstruktion hingewiesen. Die Frage der Verantwortung hängt beispielsweise für Buolamwini maßgeblich von der Trennung menschlicher und maschineller Handlung ab, um Verantwortlichkeit als Mechanismus etablieren bzw. fordern zu können. Gleichzeitig verwischt diese Grenze bei automatisierten Prozessen wie den Genderclassifyern (vgl. Buolamwini 2017, S.46).

Ausgehend von dieser Prämisse schlägt Bath Barads agentiellen Realismus vor, um die Grenzziehung als politischen Prozess untersuchen zu können (vgl. ebd. S.100). Ich führe deshalb diesen theoretischen Ansatz ein. Ich habe die grundlegenden Unterschiede in der nachfolgenden Tabelle thesenhaft zusammengefasst (s. Abbildung 3):

Konzept/ Funktionen	Butler Gender Performativität	Barad posthumanistische Performativität
Prämisse	„grundlegende Kategorien des Geschlechts, der Geschlechtsidentität und des Begehrens als Effekte einer spezifischen Machtform zu enthüllen“ (Butler 1991, S. 9)	„In traditional humanist accounts, intelligibility requires an intellectual agent [...] and intellection is framed as a specifically human capacity. But in my agential realist account, intelligibility is an ontological of the world in its ongoing articulation. It is not a human-depend characteristic.“ (S.149)
Apparat	Gender ist ein Apparat und die Norm. Binarität wird durch Gender eingeführt in das soziale Feld (vgl. 84) Sie [die Norm] erlegt dem Sozialen ein Gitter der Lesbarkeit auf“ (S.73) Da es sich um einen Machapparat handelt, könnte Gender auch dazu	Apparate sind Praxen der Grenzziehung, wie beispielsweise Natur/Kultur; Sozial/Wissenschaft; Mann/Frau. Sie entstehen aus anderen Praxen der Grenzziehung (vgl. 148) Apparate produzieren Phänomene (170). Diese Phänomene lassen sich nicht

	beitragen, die Norm abzuwandeln (vgl. 74)	auf Menschen und soziale Strukturen wie Macht zurückführen (vgl. 171)
Akteur:innen	Subjekte, die Effekte der Gendernorm sind (vgl. 1991, S.213) Verworfenen Subjekte, die durch ihre nicht binäre Geschlechtlichkeit die Produktion der Norm verdeutlichen (S. 92) Subjekte und verworfene Subjekte stehen in konstitutiver Beziehung. Es kann Subjekte nur in Hinblick auf Gesellschaft geben (vgl. 245) Normen als produktive Macht, die gesellschaftliche Strukturen hervorbringen, in dem sie Identifikation und Begehren als Ressource bereithalten (vgl. 1991, S. 104) • Institutionen • Verfahrensweisen • Diskurs (vgl. 1991, S. 9)	Menschliche/nicht menschliche Akteur:innen, in spezifischen Beziehungen. Diese heißen Intraaktionen. Erst durch diese Beziehungen entstehen Grenzen zwischen Mensch/nicht Mensch. 171) Akteur:innen, auch technischer Natur, wie beispielsweise Labore und Computer (vgl. 171) Agency ist hier keine binäre Eigenschaft von Subjekten (haben oder nicht haben). Sie ist graduell im Apparat verteilt (vgl. ebd.)

(Abb. 3. eigene Abbildung, Theorievergleich Butler, Barad)

Barads Theorie des agentiellen Realismus beruht auf dem Konzept der Intraaktion. Dieses begründet sich auf der Annahme, dass Phänomene der Schlüssel zu Weltbetrachtung sind. Phänomene, wie die oben beschriebenen Hybridobjekte, entstehen in einem komplexes Gebilde aus Handlungen. Sie sind keine stabilen Einheiten, vielmehr stabilisieren sie sich durch eine ständige Wiederholung. Durch Barads Fokus auf Phänomene wird die menschliche Subjektivität als Prämisse von Handlung verworfen, vielmehr setzt Barad auf das Zusammenspiel nicht menschlicher und menschlicher Akteur:innen. Dies lässt sich mit der Butlerschen Idee der Genderperformance vergleichen, nach der Geschlecht immer wieder, in Form von Rechtsvorschriften, Familienbeziehungen, medizinischen Diagnosen und vielen externen Faktoren, neu definiert wird. Barad selbst leitet diese Einsicht jedoch aus einem technischen Kontext ab, dem Doppelspaltexperiment. In diesem wird Licht durch einen Doppelspalt (in Form einer Platte mit zwei Öffnungen) auf einen dahinterliegenden Bildschirm geworfen. Das so entstehende Muster ist abhängig von der experimentellen Betrachtungsweise, je nach Fokus erscheint ein Wellen- oder Teilchenmuster (vgl. Barad 2013, S.36). Barad führt hier aus, dass Subjekt/Objekt und Messapparat sich erst durch das Experiment konstituieren und das in einem kontingenten Verhältnis, je nach Ausrichtung des Messstandpunktes verschiebt sich das Objekt bzw. wird durch die Verschiebung auch der Subjektstatus verschoben. So erkenne ich als Forscher:in Wellen- oder Teilchenmuster, abhängig davon, ob ich, theoriegeleitet, Licht als

wellenförmig begreife oder aber als aus Teilchen bestehend (vgl. Barad 2007, S. 104-105). Barad verweist in ihrem Konzept des agentuellen Realismus auf eine politische Dimension dieser Idee, der posthumanistischen Performativität, die sich über die erwähnten physikalischen Experimente erstreckt und deswegen furchtbar für meine Analyse werden lassen.

„I propose a posthumanist performative approach to understanding technoscientific and other naturalcultural practices [...]. The move toward performative alternatives to representationalism shifts the focus from questions of correspondence between descriptions of and reality [...] to matters of practices, doings, and action.“ (Barad 2007, S. 135)

Mit der jeweiligen Anschauungsweise verschiebt sich wiederum, bezogen auf meine Analyse, die Betrachtung spezifischer Ungleichheitsverhältnisse, wie des binären Geschlechter-Bias in den analysierten Genderclassifern. Die von Gebru und Buolamwini beibehaltende Zweigeschlechtlichkeit ist so nicht Ausdruck einer neutralen Beschreibung der Funktionsweise der Genderclassiifer. Vielmehr führt mich die posthumanistische Perspektive von Barad zu der Frage: Welche Mechanismen führen zu diesem Phänomen (der Genderbinarität), und zwar nicht im Butlerschen Sinne von Gender als (menschliche) Norm sondern zu Gender als Apparat (vgl. Bath 2013, S. 102- 103)?

Hierbei würde ich Barads Kritik an Butler, nach der Butlers Genderkonzeption als zu stark auf Sprache und weniger auf Materie ausgerichtet ist, nicht stärken wollen. Nach Barad würde Butler in ihrer Performativitätskonzeption zu wenig auf die eigene Agency der Materie an sich eingehen und stärker an der Idee der Sprache als Medium der Konstruktion von Körpern verhaftet bleiben (vgl. Barad 2007, S.213-214). Ich würde mit Blick auf meine untersuchte Studie eher den, auch von Bath beschriebenen, Hybridisierungseffekt und den sich wandelnden theoretisch Standpunkt unterstreichen, also den Fokus eher auf die Besonderheit informatischer Artefakte, sowohl theoretische Ansätze als auch Praxen zu sein, die Aufgaben, wie automatische Geschlechtererkennung erfüllen, legen (vgl. Bath 2013, S.100).

In meiner schematischen Darstellung habe ich Butlers Konzept der Geschlechterperformativität als Spiralbewegung beschrieben. Hierbei dient die Bindung von Gender an Subjekte einem stabilen Standpunkt, der eine feministische Analyse der Geschlechternormierung als binär zulässt (vgl. Butler 1991, S. 17). Diese Verbindung eröffnet somit theoretisch einen Kausalzusammenhang zwischen Norm und Subjekt, der performativ und nicht naturalistisch begründet ist. Es gibt ein binäres Geschlechtersystem, nicht weil es „natürlich“ ist, sondern weil es eine sich wiederholende politische, rechtliche Praxis ist, die genderqueere Menschen aus dem Feld der gesellschaftlichen Sichtbarkeit drängt, beispielsweise in Form der rechtlichen Regulation sexueller Belästigung am Arbeitsplatz, welche von einer heterosexuellen, zweigeschlechtlichen Beziehung zwischen übergeordneten Männern und untergeordneten Frauen ausgeht (vgl. Butler 2017, S. 95). Wie aber Corinna Bath und

Buolamwini selbst angemerkt haben, ist diese exklusive Verbindung von Norm und Subjekt in Anbetracht der Hybridisierung als Forschungsperspektive und Policy-Ausgangspunkt wenig zielführend (vgl. Bath 2013, S.96, Buolamwini 2017, S.44). Im Feld der KI, spezifisch der Genderclassifier, kann diese Beziehungen erstens auch auf maschinellen Prozessen beruhen. Gender wird hier durch automatische Genderclassifier zugeteilt und nicht durch ein Richter:innenspruch. Andererseits verweist die zweigeschlechtliche Norm der Genderclassifier nicht mehr zwangsläufig auf Menschen als direkt Betroffene, sondern durch ihre mittelbare Beziehung in Form der Fotos der Menschen, die für das Benchmark des Gender Shades Projektes von Buolamwini und Gebru ausgesucht wurden, auch auf technische Artefakte, wie digitale Bilddateien (vgl. Buolamwini, Gebru 2018, S. 4).

Mittels Barads posthumanistischer Performativität kann ich Butlers Verbindung von Subjekt, Norm und Geschlecht aufgeben und trotzdem mit Gender als Norm weiterarbeiten, um die von Gebru und Buolamwini beschriebene und benutzte Geschlechterbinarität zu untersuchen. Nach Barads Idee ist beispielsweise die Genderbinarität ein Effekt der Genderclassifier. Dieser bezieht sich spezifisch auf KI als Ausgangspunkt der Zweigeschlechtlichkeit. Die Evaluation Buolaminis und Gebrus wäre ein Apparat, der einen Einschnitt in diesen Prozess vornehmen würde, um seinerseits eine Veränderung hervorzurufen, welche gleichzeitig durch ein binär gegendertes Datenset und die Anerkennung von Geschlechterdiversität bestimmt wären, jedoch mit unterschiedlich verteilten Handlungskompetenzen (vgl. Buolamwini 2017, S.100).

Die Evaluation ist folglich ein spezifischer Moment der Sichtbarmachung und der Herstellung hybrider instabiler Artefakte. Die Intelligibilität erfolgt dabei durch menschliches Handeln in Form der Kategorisierung der im Datenset verwendeten Fotos durch die Wissenschaftler:in und durch das maschinelle Zuordnen von Genderkategorien durch die untersuchten Genderclassifier, sowie durch das aggregierte „Wissen“ der Genderclassifier hinsichtlich ihrer sich ständig wiederholenden Praxis des Kategorisierens von Fotos (vgl. Buolamwini, Gebru 2018, S.5-8). Dieser Zusammenschluss, den Barad als Apparat bezeichnet, ist jedoch nicht nur durch das bloße Vorhandensein von Mensch und Genderclassifier gekennzeichnet, sondern selbst als spezifische Funktion charakterisiert.

„Apparatuses are [...] open-ended practices involving specific intra-action of humans and non-humans, where the differential constitutions of humans and non humans, designates particular phenomena that are themselves implicated in the dynamics of intra-activity” (Barad 2003, S. 171).

Corinna Bath hat diese Prozesshaftigkeit noch einmal genauer auf die nicht kausalen Zusammenhänge algorithmischer Artefakte zugespitzt. Es kann, nach Bath, im Bereich intelligenter Software nicht mehr von einer Kette diskreter Ursachen (Akteur:innen) für eine bestimmte Wirkung ausge-

gangen werden. Vielmehr sei eine theoretische Analyse notwendig die bestimmte Zusammenarbeit von Artefakt (KI) und Mensch beschreibt und für die Fallanalyse somit sichtbar macht (vgl. Bath 2013, S. 94).

Bezogen auf Genderclassifier lässt sich schwer bis unmöglich ermitteln, welche Bilder von Personen in welcher Weise zu den Mustern der Genderclassifier geführt haben, nach denen diese Geschlecht ermitteln. Vielmehr analysieren Buolamwini und Gebru die Effekte ihres Datensets und den Algorithmen, ohne eine kausale Wirkweise zu kennen oder anzunehmen. Die Wissenschaftler:innen erklären gleichzeitig, dass diese Muster sozialen Ungleichheitsstrukturen folgen, in dem sie Männer genauer erkennen als Frauen und weiße Gesichter akkurater in Mann/Frau unterteilen als Schwarze Gesichter (vgl. Buolamwini 2018, S. 4-ff.).

Anders formuliert stellt Geschlecht so keinen Ursache-Wirkungszusammenhang her. Ein Mann ist somit nicht zwangsläufig einer, weil er sich selbst oder so begreift und gesellschaftlich so agiert, sondern weil Software dem Bild dieses Menschen, eine Kategorie zuweist, die durch Auditstudien wiederum überprüft werden (vgl. Barad 2013, S.55-57). Dabei lassen sich die „Gründe für diese Einordnung jedoch nicht rückwirkend feststellen. weiße Männlichkeit wird so mittels posthumanistischer Performativität hergestellt, die sich in statistischer Akkuranz ablesen lässt, anhand der Auditstudienresultate des Gender Shades Projektes (vgl. Buolamwini 2017, S. 92) . Das Labeln selbst wird dabei zur machtvollen Operation mit politischen Auswirkungen, in Form von Ausschlüssen genderqueerer Menschen und Diskriminierung von Frauen/Frauen of Color und Männern of Color.

Ich möchte im Folgenden noch genauer auf die Frage der politischen Auswirkungen einer posthumanistischer Performativität von Gender eingehen, auf welche Barad mit dem Bezug auf die politische Wirkmächtigkeit der Diffraction nach Haraway hinweist; denn so Haraway

„Unlike reflections, diffractions do not displace the same elsewhere, in more or less distorted form [...] Diffraction is a narrative, graphic, psychological, spiritual and political technology for making consequential meanings.“(Haraway 1997, S.273).

Was diese Bedeutungen auszeichnet in Bezug auf das Gender Shades Projektes möchte ich nun erläutern.

i.) Politischer Cut

Die Geschlechterbinarität, die sich in der Studie Gendershades zeigt, verstehe ich, wie bereits beschrieben, als posthumanistisch performativ. Ich erweitere so meine theoretische Perspektive auf Gender als Norm, hinsichtlich KI als handelndes und hybrides Objekt (vgl. Bath 2013, 93). Ein wichtiger, und für eine politikwissenschaftliche Analyse der Auditstudie zentraler Punkt ist dabei

die auf Macht und gesellschaftliche Verortung beruhende Konzeption von Barads Performativität (Barad 2007, S.245-246).

Ich möchte im Folgenden auf diese politische Dimension eingehen und mich wieder auf das breitere Ziel des Gender Shades Programms fokussieren. Habe ich bis jetzt die Funktionsweisen einer Genderbinarität ausgearbeitet, so bleibt der eigentliche Impetus von Buolamwinis Studie nämlich die Sichtbarmachung und Bearbeitung einer intersektionalen Verschränkung von Geschlecht und Race, in meiner Arbeit bislang unbearbeitet (vgl. Buolamwini, Gebur 2018, S. 1-2).

Ich möchte Anhand zweier Punkte erläutern weshalb die posthumanistische Performativität, nach Barad, als intersektionale Verschränkung innerhalb der Zielsetzung des Gender Shades Projektes verstanden werden kann und auf politikwissenschaftlich relevante Aspekte dieser Interpretation eingehen.

Erstens erläutere ich die Verschränkung von Gender und Race, indem ich Gender als fixen methodischen Ausgangspunkt des Gender Shades Projektes deute. Diese Interpretation geht aus meiner Verwendung der posthumanistischen Genderperformativität hervor. Zweites werde ich, anhand des bereits eingeführten Konzeptes der Protokategorie von Cheney-Lippold als theoretische Verbindung zwischen Gender als Norm und Gender als Technologie in meiner Theoriematrix, die Perspektivverschiebung von rein digitalen zu performativ digitalen Auswirkungen¹⁹ der Geschlechterbinarität des Gender Shades Projekts erörtern.

ii.) Gender als Messstandpunkt

Ich habe bis hierhin erläutert, weshalb sich Gender theoretisch nicht auf ein rein menschliches Subjekt beziehen kann, indem ich, in Abgrenzung zu Butler, zu verdeutlichen versucht habe, dass auch der Genderclassifier einen Referenzpunkt darstellt. Ich möchte in diesem Schritt jedoch auf die Funktionsäquivalenzen von Gender eingehen, indem ich den Nutzen einer stabilen Genderkategorie für KI am Beispiel der Genderclassifier erläutere. Dieser Schritt ist notwendig, damit ich das politische Potential der posthumanistischen Performativität von Gender aufzeigen kann, die sich mit Informationstechnologie als Untersuchungsfeld auseinandersetzt (vgl. Barad).

Barad begreift die posthumanistische Performativität als Werden und Verbinden von Mensch, nicht Mensch und Messapparat. Diese Verbindung wird durch eine bestimmte Grenzziehung sichtbar. Ich deute diese Grenzziehung auch als Ordnen eines Prozesses, da Barad Foucaults Konzept der Normierung übernimmt und auf nicht menschliche Akteur:innen überträgt (vgl. ebd. S. 199). Apparate ermöglichen es Phänomene wahrzunehmen, da sie durch Grenzziehungen Sinn stiften.

¹⁹ Performativ digital meint hierbei, dass die durch die Genderclassifier und die Benchmark produzierten Geschlechterkonstruktionen, anders als in Cheney-Lippolds Ansatz, nicht getrennt von den beteiligten Personen betrachtet werden können (vgl. Cheney-Lippold 2017, S.63-67)

„Apparatuses enact agential cuts that produce determinate boundaries and properties of ‘entities’ within phenomena, where ‘phenomena’ are the ontological inseparability of agentially intra-action“ (Barad 2003, S.148)

Dabei werde ich sowohl Gender als auch die Genderclassifier als Apparate bezeichnen, die sich gegenseitig konstruieren. Beide, so meine Prämisse, bedingen sich wechselseitig. Einerseits benötigen Genderclassifier Gender als Aufgabe, sonst wären es keine Genderclassifier, andererseits lassen Genderclassifier ein Konzept von Gender wirklich werden, welches binär, maschinenlesbar und interpretierbar ist.

Ausgehend von der stabilisierenden Kategorie der Genderbinarität ist es Buolamwini und Gebru hierbei möglich, in einem komplexen Prozess des Labelns der verwendeten Bilddateien mit der Fitzpatrick-Skala, Race als zweite intersektionale Kategorie einzuführen, und den untersuchten Genderclassifiern rassistische Erkennungspraxen nachzuweisen, was einerseits eine antirassistische Politik Buolamwinis ermöglicht, gleichzeitig jedoch auch auf die Schwierigkeit der Verbindung von Bild, Hautfarbe und demografischen Konzepten wie Race hinweist.

„Though previous studies have used nationality as a proxy for ethnicity, ethnicity and race distinctions are unstable, overlapping, and defined differently in nations around the world. The rules for membership evolve overtime.“ (Buolamwini 2017, S.99)

Die Genderbinarität wird hierbei zu einer Art Bruchlinie, die eine intersektionale Evaluation der Genderclassifier ermöglicht. Gender ist der fixierte theoretische Standpunkt, der das Analysieren von rassistischen automatischen Analysen ermöglicht, da die verwendete Binarität mit der Funktion der untersuchten Genderclassifier übereinstimmt (vgl. Gebru, Buolamwini 2018, S. 5-6). Ich möchte nun zu dem bereits angekündigten zweiten Punkt kommen, die intersektionale Verschränkung der automatisch zugeordneten Kategorien Gender und Race.

d.) Protokategoriales Genderkonzept im Gendershades Projekt

Ich beziehe mich an dieser Stelle auf die bereits erwähnte protokategoriale Intersektionalität. Dabei wird die Verschränkung soziodemografischer Kategorien, nach Cheney-Lippold, zu einem Fundament des Internet und, damit verbunden, intelligenter Such- und Kategorisierungssoftware, wie den automatischen Warenvorschlägen großer Techkonzerne wie Google (vgl. Cheney-Lippold 2017, S.57-58). Der Autor verweist dabei auf die Besonderheit des digitalen Raumes und entwickelt ein Konzept, welches diese widerspiegeln soll, die protokategoriale Intersektionalität.

Anders als beispielsweise Butler oder Barad macht der Wissenschaftler in seiner Studie auf die kommerzielle Notwendigkeit genauer und individualisierter Marketinganalysen als Mittel digitaler Wertschöpfung aufmerksam. Die Prämisse der Unternehmen sei, dabei möglichst differenziert,

jede:n Nutzer:in, ausgehend von soziodemografischen Markern wie Alter, Gender oder Race zu identifizieren und mit Produkten oder Dienstleistungen auszustatten. Auch staatliche Akteur:innen würden diese neue intersektionale Akkuranz nutzen, um, wie bereits beschrieben, Menschen zu überwachen, die eine vermeintliche oder tatsächliche Bedrohung darstellen würden (vgl. ebd., S.80).

Dabei bezieht Cheney-Lippold sich auf die machtvolle Wirkung der Kategorien. Diese ist bei dem Wissenschaftler immer zweiteilig. So folgt er einerseits dem Konzept von Butler, Kategorien als Norm zu sehen, die auf ein menschliches Subjekt referieren, andererseits gäbe es immer auch eine digitale Variante der Kategorie, diese sei jedoch immer getrennt vom Subjekt und der Handlungsmacht des Subjekts. Dies sei eine Trennung zwischen digitalen, durch kapitalisierte/staatlich kontrollierte Kategorien und einer „social and political identity that articulates various bodies, performance, and self identifications“ (vgl. ebd. S.73).

Diese theoretische Grenzziehung unterscheidet sich grundlegend von dem von mir verwendeten Konzept der posthumanistischen Performativität von Gender. Zwar erkennt Cheney-Lippold an, dass mit Algorithmen neue technische Akteur:innen in das Feld der kapitalistischen Produktion kommen und neue Ansätze von Agency und Politik notwendig werden lassen, gleichzeitig führt er eine unbewegliche Binarität ein zwischen politischer und unpolitischer Kategorie (vgl. ebd. S. 58). Gender wie Race wird in Cheney-Lippolds Sicht zu einer klar trennbaren Typologie entlang der Achse digital/sozial (vgl. ebd.)

Dabei stehen emanzipatorische gesellschaftliche Kämpfe auf der einen Seite, fest gebunden an den Subjekten vor dem Rechner und auf der anderen Seite steht ein Marketing oder nationale Sicherheitsfragen betreffendes Datencluster, die einer strikten asymmetrischen Befehlshierarchie folgen: Auf der einen Seite der Staat/Google auf der anderen Seite die passiven Empfänger:innen des Outputs.

„The possibility for a protocategorical intersectional politics comes into the picture only after the clustered category is created, not before. It is a politics that is useful for whoever is doing the algorithmic processing, not for those who are living the output of that processing.“ (ebd. 84).

Diese ontologisch fixierte Binarität Cheney-Lippolds, die in der menschlichen Politik liege, führt zu einer anderen Betrachtungsweise und einer anderen politischen Herangehensweise als sie eine posthumanistische Performativität erlauben.

Wie ist das zu verstehen?

Mein Punkt hierbei ist, dass sich, wenn wir davon ausgehen, dass Subjekt, Algorithmus und

Datenset in einem Prozess miteinander verbunden sind und immer wieder konstruktiv durch Grenzziehungen markiert aber nicht letztendlich definiert werden, nach Barad auch Agency für Algorithmen ergibt, die sich nicht in strikten Kausalketten auflösen lassen, wie oben beschrieben (vgl. Barad 2003, S.135-137). Damit berühre ich hier aber noch einen fundamentaleren Punkt, und zwar die Frage der Rolle der Genderclassifier und ihrer Beziehung zu gesellschaftlichen Unterdrückungsformen. Folge ich Cheney-Lippold in seiner Argumentation, wäre die einzige Möglichkeit politischen Handelns beschränkt auf analoge politische Vereinigungen und die verwendeten Künstlichen Intelligenzen wären transparente (Macht durchlässige, aber nicht mit Macht handelnde) Artefakte, wartend auf die Bedienung herrschender Klassen. Demgegenüber machen Buolamwini und Gebru jedoch deutlich, dass einerseits Genderclassifier sowohl einen negativen Einfluss auf das Leben Schwarzer Menschen und insbesondere Schwarzer Frauen haben, dieser jedoch nicht zwangsläufig die Exklusion ihrer selbst aus einem politischen Forschungsprojekt zur Folge hat. Vielmehr bringen sich Buolamwini und Gebru ja gerade aus dem Anspruch der Veränderung heraus mit dem Gender Shades Projekt in die digitale Sphäre der Protokategorien ein, um Cheney-Lippold zu paraphrasieren. Ruha Benjamin verweist in diesem Zusammenhang auf das grundlegende Problem der systemischen Unterdrückung. So seien Race und Rassismus Technologien, die Welten ordnen und sichtbar machen (vgl. Benjamin 2019, S. 53). Es gibt demnach keine Zweiteilung der politischen und digitalen Kategorie. KI wird in diesem Zusammenhang ein Apparat, der Race verdeutlicht, der Einfluss auf das Leben von Menschen hat, aber nicht mit Menschen identisch ist.

„Robots exemplify how race is a form of technology itself, as the algorithm judgement of Beauty AI extend well beyond adjudicating attractiveness and into questions of health, intelligence, criminality, employment, and many other fields, in which innovative techniques give rise to new flanged form of racial discrimination.” (ebd. S52).

Das Gender Shades Projekt kann so einerseits als Selbstermächtigung gesehen werden, das sich jenen simplizistischen Zuschreibungen von Handlungsmacht und Unterdrückung widersetzt, die Cheney-Lippold anwendet.

Andererseits wird mittels meiner Abhandlung der posthumanistischen Performativität sichtbar, dass der prozesshafte Charakter des neuen Datensets von Buolamwini und Gebru ambivalente Situation herrührt, die Race als Komponente des Evaluierungsprozesses der drei untersuchten Genderclassifier in Erscheinung treten lassen. Auf der anderen Seite wird Gender als Binarität stabilisiert und ermöglicht es Buolamwini und Gebru, aufzuzeigen, inwiefern Genderclassifier

weiße Männer mit fast 99% Wahrscheinlichkeit als solche erkennen aber nur 70% Schwarzer Frauen (vgl. Buolamwini, Gebru 2018, S. 9)

Ferner ist die Evaluierung, das Ziel des Projektes, nicht nur bloße technische Datenakkumulation, sondern ein politischer Akt, der einen Unterschied macht in der Welt, auf widersprüchliche Art und Weise. Durch das Gender Shades Projekt kann ein Genderclassifier auch beispielsweise auf Schwarze Menschen angewandt werden, und diese könnten in Zukunft weniger mangelhaften Anwendungen ausgesetzt sein, was wiederum potentiell Leben retten könnte. Gleichzeitig bleibt, wie ich bereits geschildert habe, die überproportionale Überwachung Schwarzer Menschen²⁰ (vgl. Raji et al. 2020; Browne 2015). Gleichzeitig bleiben auch Monopolstellungen der Unternehmen oder Ausbeutungsverhältnisse hinsichtlich der Arbeitsverhältnisse der KI-Produktion durch die Auditstudie von Buolamwini und Gebru unangetastet (vgl. Raji et al. 2020), S.6). Posthumanistische Performativität, als theoretisches Konzept, ist somit nicht einer teleologischen Idee des Fortschritts verhaftet. Nur weil soziotechnische Prozesse in den Mittelpunkt der Analyse neuer Informationstechnologien, wie in meiner Studie einfließen, heißt das nicht, dass dies einen absoluten ethischen, oder in meinem Fall politischen Wert darstellt. Wie Donna Haraway es formulierte, gibt es keine unschuldigen Positionen (Haraway 1995b, S.84-85)

Im nächsten Abschnitt werde ich detaillierter auf die hier angerissenen ambivalenten Momente der Politik hinter Buolamwinis Zielsetzung eingehen.

3.) Polittheoretische Analyse des Gendershades Projekts

In diesem abschließenden Teil der Untersuchung zur Zielsetzung des Gender Shades Projekts möchte ich, wie bereits erwähnt, eine Analyse folgen lassen, die sich mit der Machtdynamik innerhalb und im Umfeld des Projektes beschäftigt.

Dies werde ich mittels zweier Punkte nachvollziehen.

Erstens möchte ich, ausgehend von der posthumanistischen Performativität, noch einmal einer kritischen Verortung der prozesshaften Zusammenwirkung von Mensch/Maschine nachgehen. Zweitens werde ich mit dem bereits eingeführten Machtbegriff von Haraway erläutern, wieweit Buolamwinis Handeln aus einer soziotechnischen Perspektive als machtvoll gelten kann, und welche Problematiken sich damit aber auch ergeben, bezogen auf die herausgearbeitete Binarität und die spezifische Wirkmächtigkeit der Evaluation Buolamwinis hinsichtlich der Genderclassifier. Ich komme so auf die anfangs aufgestellte These zurück, dass gerade die intersektionale Verkettung und die spezifische Funktionsweise der Genderrecognition zu einem neuen Denken über die Frage der Ausgestaltung von Politik führt.

²⁰ Ausgehend von der fast ausschließlich US-amerikanischen Literatur möchte ich hier darauf verweisen, dass es sich um eine lokal begrenzte Aussage meinerseits handelt.

a.) Posthumanistische Performativität und das Subjekt

Durch die von mir herausgearbeitete posthumanistische Komponente, also der aufgeteilten Agency von Algorithmus, Datenset und Mensch, bleibt die Frage bestehen, wie diese Aufteilung vor dem Hintergrund einer machtkritischen Analyse zu bewerten ist bzw. was die Besonderheit einer Perspektive der posthumanistischen Performativität ist.

Die posthumanistische Theoretikerin Rosi Braidotti merkt vor diesem Hintergrund an, dass ein wichtiges Differenzierungsmerkmal kritischer posthumanistischer Theorien, wie ich eine verwende, das Moment politischer Arbeit und Interpretation ist. Die Differenz sieht Braidotti gegenüber Actor Network Theory (ANT), von der sie Posthumanismus abgrenzt. Diese Theorieschule befasst sich, aufbauend auf ihren Gründer Bruno Latour, mit den Überlegungen zu symmetrischen Handlungskompetenzen, sowohl menschlicher als auch nicht menschlicher Akteur:innen. Braidotti verweist vor diesem Hintergrund auf die wichtige Stellung des Subjektes als Ausgangspunkt der posthumanistische Theorie bzw. kritischen posthumanistischen Theorie. Die Vermeidung der Gleichsetzung von Subjekt und Objekt in ihren Handlungs- und Verantwortungsspielräumen ermöglicht erst feministische und antirassistische Analysearbeit, da das vernetzte Subjekt als Mittelpunkt der Theorie verstanden wird, ohne jedoch die Einbettung in maschinelle, biochemische und andere nicht menschliche Zusammenhänge auszuschließen, bzw. die Objekte nicht als ebenbürtige Partner:innen zu betrachten.

„By flattening out different degrees of power as potentia, Latour relinquishes the subject function altogether. If there are no subjects, in fact, then there is no need for genders, class, race and age-oriented analyses of power relationsrelations(???) [...] This is the point where I disagree: no amount of claim equality between human and non-human actors [...] can compensate for the Lack(???lack???) of an epistemology that does justice to the power structures of contemporary subjects“ (Braidotti 2019, S.57).

Mit dieser Konturierung möchte ich wieder auf Buolamwinis, von Fairness und Ethik geprägtes Ziel der Vermeidung der Repression Schwarzer Menschen und Menschen of Color und besonders Schwarzer Frauen kommen. Denn für die Wissenschaftler:in steht außer Frage, dass wir durch und mit Technik vergesellschaftet werden und diese Technik wiederum in Gesellschaft geschaffen und performt wird (vgl. Buolamwini 2017, S. 14). Subjekte stehen hier im Vordergrund der Forschung nicht als passive Empfänger:innen der Technik oder Partner:innen von Genderclassifyern. Buolamwini geht es ja gerade nicht um eine neue Spezifizierung der Genderclassifier sondern um eine Evaluierung dieser, zum Schutz marginalisierter Gruppen. Diese Gleichzeitigkeit der Politik für Subjekte und auch würdevoller Subjekte aus einer eigenen politischen Positionierung heraus mittels Künstlicher Intelligenz ist aus meiner Untersuchung heraus als posthumanistisch zu deuten. Die Zusammenhänge der menschlichen und nicht menschlichen Akteur:innen werden entlang von

Machtstrukturen sichtbar. Um noch einmal auf das von Benjamin angeführte Beispiel der „giggling programmer“ zurückzukommen, die eine spezifische Position innerhalb des Phänomen Genderclassifier haben. So handelt es sich aus einer posthumanistischen Perspektive heraus, wie ich sie aufgeworfen habe, um eine Akteur:innengruppe, die aufgrund gesellschaftlicher Privilegien einen höheren Grad an intendiertem Einfluss hinsichtlich der Technikgestaltung hat, als beispielsweise Schwarze Menschen bei denen Genderclassifier proportional öfter angewendet werden, bezogen auf beispielsweise Polizeikontrollen (vgl. Buolamwini, Gebru 2018, S. 2). Einzelne Akteursgruppen haben so einen höheren Grad an Gestaltungsmöglichkeiten und somit an Einfluss. Dennoch ergibt sich aus der Verwebung und Gestaltung der Genderclassifier und deren Ausführung eine Agency die spezieller Handlungen erlauben, die diesen Einfluss im Buolamwinischen Sinne fairer verteilt. Diese beziehen sich nicht nur auf analoge Politikinstrumente wie der Quotierung von Datenwissenschaftler:innen bei Google.

Indiesem Sinne kann das Projekt Gender Shades als politischer Ausdruck posthumanistischer Subjektwerdung verstanden werden. Das Feld der politischen Aktionen erweitert sich in dem Moment der Widersetzung gegen den bisherigen Status von Genderclassifiern, beispielsweise auf das Erstellen des Gender Shades Datensets, mit Auswirkungen, die einer Gruppe von Subjekten helfen. Wenn D'Ignazio und Klein Buolamwini als „public data scientist“ beschreiben, weil sie ihre Forschung in politische Kontexte einbettet, würde ich dies folglich umkehren und Buolamwini als posthumanistische Datenwissenschaftlerin lesen, die politisch codiert. Nicht das Design wird durch politische Rahmung zum Tool von Emanzipation, sondern die politischen Bestrebungen sind mit dem Datenset verbunden. Das Aussuchen der Bilder, die Fragestellung und das in die Öffentlichkeit tragen sind teil des posthumanistischen Projektes, nicht als trennbare Elemente sondern als ununterscheidbare Phänomene. Das Subjekt erhält somit eine Stellung, die sich auf eine technische Bedingtheit stützt. Gleichzeitig wird Gender zu einem Prozess der sich auf eine komplexe prozesshafte Entität des posthumanistischen Subjekts bezieht. Diese Beziehung ist keinesfalls vorherbestimmt oder aber zukünftige Notwendigkeit. Sie entsteht durch das Handeln Buolamwinis, rassistischen und sexistischen Strukturen, maschinellern Handeln, Daten und großen Unternehmen, um nur einige Faktoren zu nennen.

Ich möchte im abschließenden Teil erläutern, warum die Genderkonzeption nicht nur auf den emanzipatorischen Charakter des Projektes verweist, sondern auch auf Macht, im Sinne eines Hierarchisierungsprozesses, und auf Fragen der Verantwortlichkeit und Ethik, wie ich sie am Anfang des Kapitels vorgestellt habe. Denn die Verbindungen sind nicht alle von gleicher Art und Form, sondern vielmehr von unterschiedlichen Positionierungen geprägt. Das ist der politische Impetus kritischer posthumanistischer Analysearbeit.

b.) Posthumanistische Performativität und Macht

Gender wie ich es vor dem Hintergrund Butlers und Barads analysiert habe, bezieht sich auf eine Machtform. Butler beruft sich wie Barad hierbei auf Foucault .

Ich habe bereits angeführt, dass mir diese Konzeption der Macht als reine Subjektivierung in Anbetracht der Genderclassifier als Untersuchungsgegenstand zu kurz greift. Im Folgenden möchte ich dies jedoch anhand der Rückbesinnung auf Buolamwini noch einmal ausformulieren.

Wie ich im vorherigen Teil beschrieben habe, ermöglicht es die posthumanistische Performativität die Idee des Gender Shades Programmes als umfassendes soziotechnisches Phänomen zu beschreiben. Technik ist nach meiner theoretischen Ausarbeitung per se ein politischer Prozess und ist eben nicht erst als fertiges Produkt eine Ursache politischer und rechtlicher Regulierung (vgl. Coleman 2009, Chun 2009).

Der Gendertheoretiker Paul Preciado beschreibt dabei diesen Wechsel einer Foucaultschen Disziplinarmacht hin zu einer technikzentrierten Macht der Übersetzung, Vermischung und Einsickerung in Bezug auf Geschlecht. Denn

„Es geht nicht mehr um die Bestrafung individueller sexueller Übertretung oder das Überwachen und Korrigieren von Verwirrungen, nicht um schärfere Gesetze oder verinnerlichte Disziplin, sondern um die Modifikation der Körper“ (Preciado 2016, S.211).

Ich möchte deshalb auf das von Haraway entworfene Konzept der Informatik der Herrschaft zurückkommen, um die Prozesshaftigkeit des Gender Shades Projektes genauer beschreiben zu können, indem ich bestimmte Grundvoraussetzungen der Genderclassifier als politische Praxis einführe. Anders als Preciado konzentriere ich mich jedoch weniger auf die Körpermodifikation als auf Bildverarbeitung, und zwar aufgrund meines Forschungsgegenstandes. Wie Haraway in ihrer Definition zur neuen Form der Herrschaft ausführt, ist die Praxis der Macht geprägt durch das Paradigma der Informationstechnologie und somit ist das Codieren und das skalierbar machen von Subjekten, Objekten, Cyborgs und technischen Artefakten souveräner Akt (vgl. Haraway 1995a, S.49-50).

Vor dem Hintergrund dieser Machtdefinition wird Buolamwinis und Gebrus Erstellung des Datensets und des Labelprozess eine gewichtige Komponente ihrer politischen Arbeit. Sie macht Geschlecht und Race sichtbar, oder mit Butlers Worten intelligibel, für Programme, die namenlose (subjektlose) Gesichter scannen (vgl. Butler 2011, S.209-211). Doch anders als Cheney-Lippold widersetzen sich die Wissenschaftler:innen hier der Rolle der passiven Empfänger:in von Macht, sie greifen in die Reproduktion posthumanistischer Genderapparate, wie Genderclassifier, ein (vgl. Cheney-Lippold 2017, S.84).Wie Buolamwini in einer kritischen Evaluierung ihrer eigenen Arbeit

anmerkt, ist ihnen dabei die Grenze ihrer Tätigkeit bewusst und die niemals endende Arbeit in einem weißen, digitalen Kapitalismus. Gleichzeitig nutzen sie diese Mechanismen aus, um in den Diskurs einzugreifen (vgl. Buolamwini 2017, S. 102). Die Ziele von einer Ethik der Genderclassifier, die Schwarze Menschen zumindest in diesem Bereich des Lebens vor Gewalt und Tod schützen soll, wird nicht mit Menschenrechten erlangt sondern über Daten, Skalen und der Verwendung von Algorithmen.

Dabei verweisen Buolamwini und Gebru auf den von Simone Browne beschriebenen Konflikt der Überwachung Schwarzer Menschen. Die Technologie der Genderclassifier, müsste prinzipiell abgeschafft werden, um die hierarchisierte und rassistische, allgegenwärtige Ausspähung und die damit verbundenen Strafvollzugsmaßnahmen zu unterbinden. Dennoch ist diese Technologie weiterhin verfügbar (vgl. Browne 2015, S.109-115, Buolamini, Gebru 2018, S. 1).

Daraus ergibt sich eine neue Verantwortung für Forschung zu Genderclassifyern und zu den verwendeten Kategorien. So hat Haraway neben dem Konzept der Informatik der Herrschaft auch erklärt, dass Wissensproduktion nicht mehr als unschuldige Arbeit verstanden werden kann. In komplexer Art und Weise seien das Handeln und die damit einhergehende Veränderung aufgrund der Einbettung in rassistische, kapitalistische, sexistische Verhältnisse immer ein Oszillieren zwischen dem Eingeständnis der eigenen Einbettung und Mitwirkung in diesen Verhältnissen und andererseits dem Begehren einer besseren Zukunft, anders ausgedrückt, eine Politik im Angesicht der Apokalypse, sei es eine Pandemie, der Klimawandel oder anstehende Präsidentschaftswahlen (vgl. Haraway 1995a, S. 36). Damit müssen jedoch Fragen der Verantwortlichkeit auch in Hinblick auf die Verwendung der binären Label Mann/Frau und deren Festschreibungen in KI gestellt werden, und zwar in einer praktisch politischen Art und Weise.

Diese Spannung wird deutlich, wenn Mensch sich unterschiedliche kritische Positionen bezüglich Genderclassifyern anschaut. Für Os Keyes, einer weißen Transforscher:in, ist der Genderclassifier, aufgrund der Fremdbestimmung von Gender und der damit immer wieder verbundenen Gefahr des Misgendering, abzulehnen. Dabei spricht Keyes vor allem die Frage der Nutzen dieser Tools an.

„But whether or not AGR (Automatic Gender Recognition) can be made to work technically trans inclusive way does not answer the question of whether that is meaningful. [...] the technologies fundamentally premised on the idea that gender is something assigned. Yet to be trans [...] means to stand as a testament to the idea that it is self-knowledge[...] that has primacy in defining gender. (Keyes 2018, S. 88:13).

Buolamwini wiederum weist auf die Notwendigkeit eines Genderclassifier hin, welcher Schwarze Menschen mit einschließt, und nimmt dafür die trans-exklusiven Formen in Kauf. Es ist ein Abwägen zwischen sich im Moment ausschließenden Zielen. Buolamwini insistiert auf bereits

verwendete Techniken, die nicht, oder nur unter größtem politischen Aufwand zu entwerfen seien. Technik wird hierbei zu einem eng verzahnten Teilstück innerhalb eines rassistisch agierenden Staates (vgl. Buolamwini 2017, S. 100-102).

Gleichzeitig weist Buolamwini auf die Schwere der Entscheidung der binär verwendeten Label hin, kann diese aufgrund der Einbettung in bestehende Genderclassifiers jedoch nicht herausnehmen, bzw. entscheidet sich für diese Technik als bereits bestehende reale bedrohende Funktion. Andererseits macht auch Keyes auf die eigene Verortung als weiße Person aufmerksam und weist auf die Frage der Ausführbarkeit und Notwendigkeit von Genderclassifiern gerade auch für besonders bedrohte marginalisierte Personengruppen, wie Trans Women of Colour oder Schwarzen Personen hin.

Als selbst weiße Wissenschaftler:in ist es in diesem Kontext eine verantwortungsvolle Aufgabe, diese Positionen gegeneinander abzuwägen, und die Frage, die sich mir stellt, ist, ob politikwissenschaftliche Analysen nicht auch an ihre Grenzen stoßen. Denn die letztendliche Gewichtung, ob ich eine binäre Auditstudie als sinnvolles Tool erachte, ist vor dem Hintergrund der Trans-Exklusion, aber gleichzeitig der Verbesserung eines Systems, um Schwarze Cis-Menschen und Menschen of Color zu schützen, für mich unentscheidbar. Ich werde niemals von rassistisch motivierten Gewalttaten betroffen sein und Race als Kategorie bietet mir ein Privileg, dass Buolamwini beispielsweise bewusst thematisiert und dass ihr verwehrt ist. Sie wurde nicht erkannt von Genderclassifiern. Die Zielsetzung Buolamwinis ist vor dem Hintergrund des systemischen Rassismus in unserer Gesellschaft für Menschen of Color überlebenswichtig. Es geht hier um ein Datenset, dass bestimmte Menschen schützen kann.

Dennoch bleibt die Frage, was mit Menschen passiert, die sich auf der intersektionalen Linie von Gender und Race an marginalisierten Punkten befinden, insbesondere Trans-Frauen of Color, die noch einmal signifikant stärker von Gewalt, Polizeigewalt und Exklusionsmechanismen, wie Geschlechterseparation betroffen sind (vgl. Carpenter, Marshal 2018). Auch hier möchte ich auf Haraways Idee der nicht unschuldigen Position zurückkommen. Denn

„unschuldige „Identitätspolitik“ und Epistemologien [sind] unmögliche Strategien für eine klare Sicht von den Standpunkten der Unterworfenen aus sind. Man kann nicht Zelle oder Molekül „sein“ - oder Frau, kolonisierte Person [...] Ebenso wenig kann man den eigenen Standpunkt an einen anderen Ort verlegen, ohne dafür verantwortlich zu sein.“ (Haraway 1995a, S. 85)

Wissenschaftliches Eingreifen verändert Gesellschaften. Diese Veränderungen sind ambivalent, wie ich gezeigt habe. Diese Ambivalenz lässt sich nicht auflösen oder korrigieren, vielmehr ist sie Teil der gesellschaftlichen Struktur im 21. Jahrhundert, die durch eine Hypervernetztheit gekennzeichnet

ist (vgl. Chun 2013, S.92-93). Das bedeutet nicht, dass politisches Handeln verunmöglicht wird. Das Gender Shades Projekt beweist ja genau einen Aktivismus in Zeiten maschineller Entscheidungen und Effekte, mit nachvollziehbaren Handlungsabläufen. Es bedeutet jedoch, dass ich als Wissenschaftler:in offenlegen muss, welche Position ich einnehme, um die Grenzen der eigenen Arbeit sichtbar zu machen und bearbeitbar (vgl. Haraway 1995a, S. 79).

Gleichzeitig können Wissen, und damit verbunden Politiken nicht mehr in kausale Wirkungsketten gebracht und verstetigt werden, wer also ist Adressat:in politischer, smarter Intervention?

c.) Posthumanistische Performativität und Handlungsmacht

In einem letzten Schritt möchte ich noch einmal zu den anfangs aufgeworfenen Zielstellungen von Buolamwini kommen, der fairen Algorithmen, Accountability und Transparenz. Denn mit der verbundenen Unterlegung von Buolamwinis Studie mit Haraways Konzept der Herrschaft habe ich gezeigt, dass diese Auditstudie machtvoll in eine politische Praxis der Genderclassifier kritisch, aber nicht unschuldig sondern ambivalent eingreift.

Und dass diese, gerade aufgrund der Codierung von soziodemografische Daten in Maschinensprache, geschieht. Dieses Codieren zeugt von einer weiteren Folge der Verbundenheit. Drückt sich in Buolamwinis Fragen der Verantwortung genau jene Grenzverschiebungen aus die gerade ihre machtvolle Intervention charakterisieren.

Algorithmische Systeme beruhen dabei auf einer Distribution von Handlungen und Ressourcen. Sie bestehen aus Bildern, Strukturkategorien, Informatiker:innen, Fotograf:innen und Nutzer:innen sowie Firmen und Staaten (vgl. Bucher 2018, S.20-30). All diese Komponenten sind eingebunden und eingeschrieben, entgegen einer Foucaultschen Disziplinarmacht die sich von außen in Körper einschreibt und diese formt. Wie die Institution Schule Schüler:innen formt, beruhen die hier untersuchten Künstlichen Intelligenzen auf das funktionale Zusammenwirken unzählbarer und unkontrollierbarer Komponenten, die sich in einen Funktionszusammenhang, dem Genderclassifier, einfügen. Dieses Einfügen ist durch gesellschaftliche Machtstrukturen bestimmt, die selbst aufeinander bezogen sind. Dies zeigt sich beispielsweise in den für Benchmarks sehr populären sehr weißen und männlichen Daten Sets²¹ mit denen Genderclassifier evaluiert werden (vgl. Benjamin 2019, S. 64).

Gleichzeitig sind diese Strukturen selbst Teil der Technik und zwar nicht als äußeres Element, sondern codierter Bestandteil dieser, in Form der zweigeschlechtlich arbeitenden Genderclassifier oder den statistischen Fehlerraten dieser Technologien hinsichtlich Schwarzer und weißer Gesichter.

21 Ein Beispiel welches auch Buolamwini und Gebru heranziehen ist das IJB-A Datenset der US Regierung mit einem Anteil von 20% an Schwarzen Gesichtern (vgl. Buolamwini, Gebru 2018, S. 7)

Dieses sind Orte der Machtausübung. Wie Butler deutlich gemacht hat, schafft soziale Macht eine Sichtbarkeit für Subjekte, und ich würde anfügen, für Technik, die Leben und Politik ermöglicht (vgl. Butler 2011, S.84).

Bezogen auf das Gender Shades Projekt bleibt die Frage der Verantwortung zwar nicht unsichtbar, aber die Lösungen bzw. Antworten auf diese Frage werden ohne genaue Analysen und Zuschreibungen unsichtbar. Dies ist nach meiner vorgestellten Analyse die Rolle einer kritischen politikwissenschaftlichen Analyse, eine genaue Untersuchung der einzelnen Handlungsträger:innen, die in neuer Art und Weise handeln und auch mit normativen Entscheidungen und Abwägungen konfrontiert sind (vgl. Benjamin 2019, S.66).

4.) Zwischenfazit

Nach dieser Analyse der Zielsetzung von Buolamwini lässt sich Folgendes für die Beantwortung der Forschungsfrage festhalten. Erstens ist Gender binär in das Gender Shades Projekt eingelassen. Diese Binarität ist performativ, wird also ständig durch neue Daten, die den Genderclassifyern zugeführt werden, wiederholt und somit gefestigt. Diese Binarität verfestigt sich jedoch nicht nur im Datenset sondern wird auch in einem größeren Kontext sichtbar, z.B. beim Gender Data Gap. Die Performativität ist von mir als posthumanistische spezifiziert worden. Gender ist somit nicht an ein menschliches Subjekt kausal geknüpft, sondern durch multiple datengetriebene Aktionen sowie politische Interventionen in einem Spannungsverhältnis zu einem posthumanistischen Subjekt.

Diese performative Binarität wirkt sich dem gegenüber auf politische Handlungsspielräume aus. Codieren wird zu einer Politik, die über zurückverfolgbare und adressierte menschliche Agency hinausragt. Dieses Hinausragen wird deutlich in der Zielsetzung von Buolamwini, wenn sie Verantwortung und Transparenz als zu erreichende Normen aufführt, gleichzeitig aber auch auf die Unmöglichkeit jene Normen vollständig zu erfüllen. Dabei habe ich parallel herausgearbeitet, dass die Intelligibilität marginalisierter Gruppen, die Buolamwini mittels des entworfenen Datensets herstellt, über Transparenz und Verantwortung hinausweist und die machtvollen Interventionen mittels Informationstechnologie sichtbar werden.

Ich werde im zweiten Teil der Arbeit genauer auf einen Aspekt dieser Sichtbarmachung eingehen und mich genauer mit Formen der Arbeit und der kapitalistischen Produktion des Datensets von Buolamwini beschäftigen. Der dritte Teil beleuchtet noch einmal genauer die eingeschriebenen Machtverhältnisse, die durch die Nutzung von Gender als algorithmische Erkennungsstrategie dient.

B.) Technische Funktionsweise

In diesem Teil der Arbeit möchte ich den Zusammenhang von Gender und der technischen Wirkweise des Auditdatensets und der untersuchten Genderclassifier von Buolamwini und Gebru untersuchen (ebd. 2018).

Die Frage „Wie funktioniert das Gendershades Projekt?“ werde ich mit meiner Forschungsfrage „Wie ist Gender in das Gendershades Projekt eingelassen?“ verbinden.

Dafür gliedert sich das Kapitel in vier Teile.

Zuerst werde ich meine analytische Perspektive erläutern, das Konzept Gender als Technologie. Dabei werde ich die Gründe für die Wahl dieser Perspektive erklären und diese definieren. Daraus leite ich Thesen ab, die dann die anderen anschließenden Abschnitte strukturieren

In einem zweiten Teil werde ich die technischen Funktionsweisen von Genderclassifyern und dem Pilot Parliaments Benchmark (PPB) erklären. Hierfür werde ich automatische Gesichtserkennungsapplikationen und maschinelles Lernen, sowie die Funktion und Bedeutung des Benchmarks erklären.

In einem dritten Schritt werden Empirie und die zugrundeliegenden theoretischen Überlegungen von mir zusammengeführt. Hier analysiere ich, wie Gender als spezifische Technologie im Gendershades Projekt wirkt. Ich werde hier zwei Konzepte herausarbeiten, einmal Gender als kategorialer Standard, basieren auf der Arbeit der Soziolog:innen Geoffrey C. Bowker und Susann Light Star und einmal Gender als agentieles Protokoll, als Zusammenführung des von der Genderforscherin Michelle Murphy verwendeten Protokollfeminismus und Karen Barad Begriffes des Apparats (vgl. Barad 2007, Bowker, Star 2000, Murphy 2012)

Diese Konzepte werde ich in einer dreiteiligen Analyse einführen, die eine Definition, die Verankerung im Gendershades Projekt und politische Implikationen dieser Konzepte beinhaltet.

Der vierte Teil stellt, darauf aufbauend, ein Zwischenfazit dar, der eine Teil Antwort auf die Forschungsfrage beinhaltet.

1.) Gendertheoretische Definition von (KI-)Technologie

Ich möchte hier keine Technikfolgenabschätzung als politikwissenschaftliche Analyse durchführen, welche ein gängiges Feld der Policy Forschung, wäre, sich aber vorrangig den Dualismus Technik und Gesellschaft beibehält. Im Bereich der Mensch-Maschine-Interaktion, als Oberbegriff einer Technikfolgeabschätzung, die sich mit intelligenter Software auseinandersetzt, die auch beispielsweise das Einsetzen von Gesichtserkennung beinhaltet, werden zwar gesellschaftliche und technische Verbindungen anerkannt. Jedoch wird dabei von immer noch in klar zu unterscheidende menschliche und nichtmenschliche Aren von Handlungsmacht ausgegangen. Zudem werden sie vor allem im Kern einer staatlicher Politik und Regulierung unterworfen bzw. als in diese eingebettet konzipiert (vgl. Grunwald, 2019, S.34).

Um die Funktionsweisen des Gendershades Projektes sichtbar zu machen und diese in Bezug zu

meiner Forschungsfrage zu klären, brauche ich einen soziotechnischen²² theoretischen Rahmen. Dieser umfasst einerseits die politischen Implikationen, die in Technikgestaltung und Anwendung von Technologie einfließen lässt und andererseits eine gesellschafts-/ politikwissenschaftliche Perspektive, die Technik analysiert und diese in ihrer spezifischen Arbeitsweise fassen kann. Ich möchte dem damit die Verbindung von Gender und dem Gendershades Projekt, aufgrund der Forschungsfrage aus einer gendersensiblen technowissenschaftlichen Perspektive beleuchten, die ich schon im Theorieteil angeschnitten habe. Dieser Rahmen ist von Nöten, da eine Prämisse meiner Arbeit die dynamische Verkettung technologischer Designs und politischen Praxen ist.

Ich verwende hierfür den theoretischen Ansatz „Gender als Technologie“, um die Verbindung von Artefakt, technischer Beschaffenheit und Politik zu untersuchen und diese Verbindung unter einem queer-theoretischen Aspekt, nämlich Gender, zu analysieren. Dafür werde ich diesen Teil des Kapitels in drei Teile aufgliedern.

Ich werde im ersten Teil ausführlich erläutern, warum ich die Perspektive „Gender als Technologie“ wähle, bzw. aus welchen theoretischen Überlegungen sich diese speist. Es handelt sich hierbei um eine theoretische Herleitung. Diese ergibt sich sowohl aus meiner Forschungsfrage und der darauf basierende Theoriematrix heraus als auch aus den antirassistischen Implikationen des Gendershades Projektes. Zweitens werde ich ein geisteswissenschaftliches Konzept von Gender als Technologie als Transdisziplinären einführen.

Drittens stelle ich die spezifischen politikwissenschaftlichen Implikationen dieser theoretischen Überlegung dar, die sich bis hierhin aus einem interdisziplinären Kanon von Medienwissenschaften, Philosophie und Soziologie ergeben.

a.) Herleitung

Warum also ist Gender als Technologie eine geeignete theoretische Basis?

Ich möchte hier zwei Begründungen anführen, die erste ist mein Forschungsinteresse und die verwendeten theoretischen Grundannahmen.

i.) Gender als Norm/ Gender als Automat?

In meiner Forschungsfrage frage ich weniger nach einer essentialisierten Form von Gender, als nach Funktionsweisen. Das „wie“ drückt eine Praxis aus und weniger einen abzählbaren Status quo (vgl. Landström 2007 S.19). So frage ich nicht etwa, wie viele Frauen an der Entstehung des Gendershades Projektes beteiligt sind, sondern welche Funktion Gender in dem Projekt erfüllt. Diese analysie-

²² Ich beziehe hier soziotechnisch auf den von mir im Theorie eingeführten Begriff der Techno Science. Ich nutze erstgenannten Begriff hier synonym als Übersetzung.

re ich mittels eines Genderbegriffs, der sich aus Butlers, Haraways und Halberstamms Ansätzen heraus entwickelt (vgl. Butler 1991, Halberstam 1991, Haraway 1995, 1997). Wie im Theorieteil erläutert, beziehe ich mich dabei stark auf die Idee von Gender als Norm von Butler, die mittels dieses Konzeptes verschiedenen Techniken der Subjektivierung analysiert. Gender als Norm wird nach Butler so zu einem Mechanismus, der Personen sichtbar werden lässt. Dabei ist Gender weniger eine stabile empirische Essenz einer Person, sondern eine spezifische Form regulatorischer Macht (vgl. Butler 1991, S.49). Allerdings verweist Butler nicht auf Computer Techniken, Informationsnetzwerke oder ähnliches, was sich für mein Untersuchungsfeld anbieten würde.

Hierzu habe ich auf Halberstams Essay „Automating Gender: Postmodern Feminism in the Age of the Intelligent Machine“ (ebd. 1991) verwiesen, in welchem der Theoretiker erklärt, das Gender einen funktionalen Charakter hat, der erlernt und ausführbar ist und dabei, wie eine künstliche Intelligenz funktioniert

Halberstam sieht Gender Anfang der 1990er Jahre jedoch nicht als Technik, sondern als etwas dem Technik innewohnt (vgl. ebd. 454). Ich möchte ausgehend von meiner Forschungsfrage und dem hier zu untersuchenden Forschungsfeld Gender als Norm mit Halberstams Idee des automating Gender zusammenbringen um „Gender als Technologie“ aus einer Gendertheoretischen Perspektive heraus verstehen zu können. Gender, so meine von Butler übernommene These, bringt spezifische Regulierungsformen hervor. Diese sind im Gendershadesprojekt auch auf spezielle Weise an der technischen Ebene eingebunden und übernehmen bestimmte Funktionen, wie etwa das Kategorisieren der für das Datenset verwendeten Bilder (vgl. Buolamwini 2017, S.57)

Das Funktionieren von Gender soll also auf die technischen Wirkweisen bezogen werden und nicht auf wie bei beispielsweise Butler sich auf juristische Praxen beziehen. Dafür brauche ich eine Theorie, die die spezifische Wirkweise von Gender berücksichtigt, aber wie Halberstam gezeigt hat, auch deren spezifische Funktionen innerhalb künstlich intelligenter Netzwerke, welche das Gendershades Projekt nutzt (vgl. ebd.). Gender als Technologie ermöglicht diese Verbindung und ist deshalb ausgehend von meiner Forschungsfrage meine gewählte theoretische Basis.

Dies ist jedoch nicht der alleinige Grund für die Auswahl der Theorie. Ein zweiter Grund betrifft die Arbeit von Buolamwini selbst, nämlich die antirassistischen Bemühungen, Tools zu kreieren, die für eine bessere Erkennung von Schwarzen Menschen durch automatische Genderclassifier sorgen könnten (vgl. ebd. S.13).

ii.) Race und Technologie

Race ist eine grundlegende Kategorie für das Gendershades Projekt. Hierbei beziehen sich Buolamwini und Gebru auf ein rassistisches Bias innerhalb konventioneller Gesichtserkennungssoftware,

die Gesichter von Menschen of Color und Schwarzen Menschen nicht erkennt bzw. sie nicht als menschliche Gesichter identifiziert (vgl. ebd. 30). Erst daraufhin entwickelten die Forscher:innen das Datenset, das der Kern des Gendershades Projektes ist und sich aus Schwarzen Menschen und Menschen of Color und *weißen* Menschen zusammensetzt. Race wird hierbei zu einer technischen Variable, die durch das Design des Datensets in Erscheinung tritt. Sie wird in automatische Gesichtserkennungssoftware durch Bilder von BiPOCs eingebunden (vgl. Buolamwini, Gebru 2018). Dabei wird die Beziehung von Race und Technik in den Sozialwissenschaften ebenfalls analysiert. Ein wichtiger Referenzpunkt hierbei sind die Texte der Medienwissenschaftler:innen Wendy Chun und Beth Coleman zu „Race as Technology“. Diese haben die spezifischen Wirkweisen und kolonialen Hintergründe von Technikgestaltung zusammengedacht und darauf aufbauend ein Konzept entwickelt. (vgl. Chun 2009, Coleman 2009). Diese theoretische Idee grenzt sich zu Konzepten, wie „Gender als Norm“ oder „automating Gender“, ab. Race als Technologie befasst sich explizit mit der technischen Wirkweise von sozialen Kategorien wie Race und dessen Implikationen für Theoriebildung (Chun 2009 S.8). Weiterhin entstehen auch Verbindungen zwischen Race und Gender als Technologien, wie sie beispielsweise der in den Transstudies forschende Julien Gill-Peterson (vgl. ebd. 2014) deutlich macht. Ich möchte daher auch Race als Technologie als einen wichtigen Aspekt meiner theoretischen Basis einarbeiten, als Schnittstelle zwischen meinem Fokus auf Gender und dem Schwerpunkt Race aus Buolamwinis Studien einerseits und andererseits aufgrund der konstitutiven Bedingung von Gender, Race und Technologie, wie sie auch Coleman in ihrem Ansatz Race as Technology explizit benennt (vgl. Coleman 2009, S. 193-4).

Doch was bedeutet Gender als Technologie überhaupt? Was meint in diesem Zusammenhang Technologie? Dies möchte ich in einem zweiten Teil herausarbeiten und eine Arbeitsdefinition von Gender als Technologie vorstellen, die für meine Arbeit von mir konzipiert wurde und keinen Anspruch auf Vollständigkeit oder Übertragbarkeit erhebt.

b.) Definition

i.) ...Technologie

In Beth Colemans Essay „Race as Technology“ schreibt die Musikwissenschaftlerin mit Bezug auf den Medientheoretiker McLuhan Technik einen konstitutiven Charakter für das menschliche Subjekt zu. Technik ist eine Erweiterung, eine Prothese des Menschen (vgl. Coleman 2009, S. 183). Die feministische Wissenschaftlerin Alison Kafer, die Colemans Ansatz der Technik als Prothese des Menschen übernimmt und ihn mit Gender und Disability zusammen denkt, expliziert diesen Zusammenhang als affektive Struktur, die nicht nur unsere Umwelt gestaltet, in Form von Infrastruktur

oder Architektur, sondern auch unser sozialen Beziehungen und unser Selbstverständnis als Mensch (vgl. Kafer 2017, S. 388). Diese Art der beidseitigen Verwicklung von Technik und Mensch ist für Kafer jedoch nicht nur eine Frage unterschiedlicher Gestaltungsweisen von Technik, sondern eine Art und Weise des Denkens und Handels.

„Technologies are not only tools used to solve problems but also ways of thinking, orientations, and modes of expression. [...] Gender for example, as a set of practices, norms and expressions, can be seen as a technology of the body.“ (Ebd. s. 387).

Dabei, so erklärt Kafer in diesem Beispiel, muss Technologie nicht immer ein Artefakt wie ein Telefon oder ein WLAN-Router sein, sondern ist viel eher eine Art des Einwirkens und Rückwirkens. In dem vorangestellten Zitat ist es das Einwirken von Gender auf menschlichen Körper Technologie. Technik ist dabei weniger ein konkretes Artefakt als viel eher die Verbindung von Artefakt, Wissen und Subjekt und Körper.

Diese Sichtweise ist prinzipiell nicht sonderlich spezifisch für den Rahmen meiner Forschungsfrage bzw. von technischen Artefakten wie Genderclassifier.

So weisen Butlers Ausführungen in Bezug auf Gender und Macht genau auf jene Verbindung von Wissen, Materialität und Subjektivität hin (vgl. Butler 1991). Oder aber auch intersektionale Überlegungen als sozialwissenschaftliche Perspektiven (ohne spezifischen Fokus auf Technologie), die, ausgehend von der Juristin Kimberle Crenshaw, sowohl Gender und Race zusammenführen. Nach Crenshaw wirken diese Kategorien auf Menschen und sind mit diesen fundamental verbunden, einerseits als politisches Handlungsfeld als politische Intersektionalität andererseits als eine strukturelle Intersektionalität, die auf die Verbindung von repressiven Systemen wie Rassismus und Sexismus beruht (vgl. Crenshaw. 1991, S.149-50, Chebout 2011).

Wo genau ist also die methodische Unterscheidung zwischen Crenshaws Ansatz Gender und Race analytisch zu verbinden und als Gesellschaft prägende Kategorien zu verstehen und Kafers Idee von Gender as Technology oder aber Beth Colemans prothetischem Verständnis von Race als mit dem Menschen verbundenen aber nicht der Mensch seiende Interpretation von Technik?

Anders formuliert, was macht Gender als Technologie zu einer speziellen Form der Analyse?

In meiner Analyse ist das Unterscheidungsmerkmal die Verwendung von Werkzeugen („Tools“) (Gill-Peterson 2014, S.415). So verweisen die bis jetzt genannten Autor:innen auf die soziale Konstruiertheit von Werkzeugen hin, aber auch hinsichtlich deren Gebrauch in Verbindung mit Subjekten und sozialen Kategorien wie Gender und Race. So beinhaltet Technologie immer auch das Werkzeug/Tool, mit dem wiederum Technologie produziert und verändert wird (vgl. Kafer 2017, S,

390). Anders formuliert: Erst das Werkzeug, als Grundlage der theoretischen Auseinandersetzung, lässt die Technologie dahinter sichtbar und somit untersuchbar werden. Technologie ist somit, wie bereits beschrieben, immer an spezifische sozio-materielle Bedingungen geknüpft. So erklärt die feministische Aktivistin und Literatur Professorin Helen Hester in ihrer Konzeption von Technik und Feminismus, dass diese Verbindung

„too often revolved around the metaphysically inflated phantom of technology as such rather than engaging the specific ways in which technologies are put to use for certain ends with distinct social assemblages.“ (Hester 2018, S. 70).

Gender as Technology und auch „Race as Technology“ als theoretischer Rahmen erlauben mir als Forscher:in, einen doppelten Analysestandpunkt einzunehmen. Die Konzepte ermöglichen mir einen Fokus auf spezifische gesellschaftliche Technologien, spezifische historische Konstruktionen von Race beispielsweise und andererseits soziotechnische Artefakte wie gebiaste Genderclassifys als Ausprägung dieser Technologie zu untersuchen.

Ich möchte diese These genauer ausführen, da sie aus der besonderen Verbingung von Gender; Race and/as Technologie hervorgeht und so noch einmal einen Schritt zurückgehen und erklären, was das spezifische Erkenntnisinteresse solcher Ansätze ist.

ii.) Gender als...

Bis hierhin war Technologie als theoretischer Begriff von Bedeutung und soll nun durch eine Betrachtung von „Gender als Technologie“ ergänzt werden. Hierbei möchte ich mich jedoch zuerst Wendy Chuns und Beth Colemans Ausführungen bezüglich der theoretischen Perspektive von Race as Technology bedienen, da diese beiden Ansätze von Chun und Coleman über dieses Verhältnis auch von Theoretiker:innen gebraucht werden, die sich auf „Gender als Technologie“ stützen (vgl. Hester 2018, Kafer 2017, Gill Peterson 2014)

Chun beschreibt die Verschiebung von Race und Technologie zu Race as Technology als einen Paradigmenwechsel der Forschung: “Crucially, race as technology shifts the focus from the *what* of race to the *how* of race, from *knowing* to doing race and technology. “ (Chun 2009, S.8, Hio)

Hier verweist Chun auf den Wechsel einer statischen Beschreibung, die auf das „was“ von Race abzielt, hin zu Praxen der Verkörperung bzw. der Funktionsweise von Race in Hinblick auf Technologie hin. Diese Äquivalenzen von technischer und kategorialer Wirkweise von Race beschreibt Coleman beispielhaft in Bezug auf digitale Datenverarbeitung und das Human Genom Projekt, welches sich mit der Sequenzierung der menschlichen DNA beschäftigt.

„In a sense race has been rendered a technology by science itself, since it’s now figured in sequences of code. There is no cranium to measure, but rather tendrils of information that cross-con-

continent. What used to be a matter of flesh and blood is now highly abstracted data. Race has been made information.“ (Coleman 2009 S.193)

Race wird in diesem Beispiel zur Verbindungstelle einer komplexen gesellschaftlichen und biotechnischen Konfiguration in Form eines Gensequenzierungsprojektes, welches mittels dieser Technologie (Race) Sinn in Form von Forschungsergebnissen schafft.

Hierbei wird Race als Technologie eine Funktionsweise zuerkannt, die im Produzieren von Daten liegt. Gender als Technologie kann als eine ähnliche Wirkweise zu erkannt werden. So schreibt Gill Peterson in Bezug auf Sex/(Gender), dass die Unterteilung in zwei verschiedene Gender erst die Möglichkeit der Hormontherapie für Transmenschen ermöglicht hat. In dem Gender sie als binäre Technik, feste Bezugspunkte und Standards geschaffen, in dem Grenzen gezogen bzw. dadurch Grenziehungen für Ärzt:innen, Staaten und andere Akteur:innen, zwischen Mann und Frau ermöglicht werden, wird ein Wechsel oder aber ein dazwischen erst möglich. Gender als Technik ermöglicht in diesem Fall eine Verbindung von Körper und dem Tool, das Hormonpräparat, welches Veränderungen erzeugt und eine Transition ermöglicht (vgl. Gill Peterson 2014, S.404). Gleichzeitig ermöglicht Gender als Technologie jedoch auch eine spezifische Trennung eines Tools und des Körpers, nicht etwa um die Natürlichkeit von Gender zu unterstreichen, beziehungsweise die Künstlichkeit von Trans, sondern um eine Verbindung beider sichtbar zu machen und ihre konkreten Existenzbedingungen.

„the body does not collapse life and technics into an undifferentiated organism. [...] the trans subject is likewise an artisan, and the material with which the subject is engaged is flesh of the body, with the hormone molecule serving as the interface relating the two without opposing them or collapsing their distinction.“ (Gill-Peterson 2014, 408).

Für meine Analyse ist dabei die Verbindung von Gender als Technologie mit „Tools“ als konkret Ausformung der Funktionsweise und Praxis von Technik entscheidend. Dabei wird Technologie selbst als eine Form des Handelns nur sichtbar in den jeweiligen spezifischen Kontexten, wie etwa einer Hormonbehandlung oder dem Humangenom Projekt. Dabei handelt es sich um einen theoretischen Ansatz, der Verbindungen von Subjekt, Technologie und Tool aufzeigt. Gleichzeitig sind alle diese Komponenten auf spezifische Weise miteinander verbunden, wie Gill Peterson in seinem Beispiel der Hormontherapie verdeutlicht (vgl. ebd.)

Jedoch muss an dieser Stelle auf eine Besonderheit von Race als Technologie eingegangen werden, die sich fundamental von Gender als Technologie abhebt, nämlich die Aufarbeitung kolonialer Vergangenheit und das andauernde Erbe der Kolonialzeit bis hinein in das 21. Jahrhundert (vgl. Chun 2009, S.8). Race, so schreibt Coleman, ist dabei nicht nur eine Technologie bzw. eine Theorie, die

sichtbar macht, wie Race funktioniert und weg von essentialistischen, oder wie Coleman schreibt, ontologischen Fragestellung führt, sondern auch ein kritischer Punkt des Erkennens der rassifizierenden und vor allem rassistischen Implementation in das westliche, *weiße* Technik Verständnis, „A structural position of mastery had been encoded in the machine itself. Under such circumstances, non-white peoples were often treated as objects for sale, as subjects, and this set in motion a binary logic of master/slave, man/machine“ (Coleman 2009, S.190).

Technik ist durch Rassismus in diess Grundverständnis eingeschrieben über Jahrhunderte hinweg, durch die Idee der Beherrschung von Technologie und die Hierarchisierung von Menschen (vgl. ebd. S.199). Gender als Technologie ist somit immer an einen rassismuskritisches Verständnis gebunden, da ich Coleman und Chun als Referenzrahmen nutze. Dabei sind Race und Gender keinesfalls zu getrennt betrachtende Variablen. Wie Gill Peterson schreibt, beruht Technik gerade auf diesen Kategorien “Technicity then [...] is how bodies assemble themselves simultaneously as racialized and gendered while preserving what makes each distinct. “ (Gill-Peterson 2014, S.409). In der kommenden Spezifizierung von Gender als Technologie in Bezug auf das Gendershades Projekts soll dies noch einmal veranschaulicht werden.

Ich möchte jedoch ausgehend von diesem Punkt auf eine politikwissenschaftliche Spezifizierung des Ansatzes Gender als Technologie kommen. Denn wie anhand von Colemans Idee der inhärenten oppressiven Struktur von Technologie ist Gender als Technologie immer mit Aspekten von Macht, Unterdrückung und Fragen der Subversion verbunden (vgl. Coleman 2009, S.184). Dieser Ansatz, so werde im Folgendem erläutern, ist konstitutiv für die hier erarbeitete Idee von Gender als Technologie und macht sie für eine politikwissenschaftliche Analyse besonders wertvoll.

c.) politikwissenschaftliche Implikationen/ Spezifizierung

Bis hierhin sind die Versatzstücke der verwendeten Theoretiker:innen aus einem breiten Feld verschiedener Fächer wie der Soziologie oder den Medienwissenschaften bzw. transdisziplinären Forschungsfeldern wie den Transstudies angeführt worden. Ich möchte im folgenden zwei Punkte vorstellen, die diesen Ansatz für eine politikwissenschaftliche Untersuchung nützlich machen: erstens Technologie und Macht in Form einer repressiven und einer emanzipatorischen Sichtweise auf Gender und Race als Technologie und zweitens: die Frage einer politischen Ebene dieser theoretischen Ansätze.

i.) Macht

Die vorliegenden Schriften stellen keinen konzeptionellen Entwurf eines Machtbegriffes zur Verfügung; dennoch stellen alle vorstellten Ansätze heraus, dass Technologie Machtstrukturen innewohnen, so schreibt Kafer beispielsweise, dass weiße Schönheitsstandards Technologien wie Hautauf-

heller oder aber Augenlid, „Korrekturen“ für asiatische Menschen erst möglich machen, da sie weißsein als Bedürfnis voraussetzen, welches die Cremes und Operationen erfüllen (vgl. Kafer 2017, S.391). Helen Hester, die in ihrem Technik bezogen feministischen Ansatz über die Frage nachdenkt, wie Technologie und Feminismus in Verbindung stehen, sieht in der Frage der Hormonbehandlung von Transpersonen und die Patentierung der Wirkstoffe von Pharmakonzernen eine wichtige politische Machtfrage und vor allem neue Möglichkeiten technisch vermittelter politischer Protestformen. Die politischen Fragen der Verteilung von Hormonbehandlungen (Wer ist antragsberechtigt?/ Was zahlen Krankenkassen?/ Warum zahlen sie? etc.), so Hester, lassen sich mit dem bewussten Hacken der Hormonpräparate bzw. das eigene Herstellen von Hormonen durch Transpersonen (quasi DIY Hormone) und einer Infrastruktur für die Verteilung dieser Produkte jenseits von staatlichen und medizinischen Institutionen effektiver begegnen als mit klassischer Parlamentsarbeit (zur allgemeinen Freigabe von Medikamenten) (Hester 2018, S.85). Technik und der Umgang mit Tools werden sowohl

Momente politischer Exklusions Prozesse, wie bei Kafer, können aber auch mittels technikvermittelter Strategien geöffnet werden, wie bei Hester. Ähnlich beschreibt diese ambivalente Wirkung auch Coleman in Anbetracht von Race. Dabei sei diese Technologie historisch immer mit Unterdrückung verbunden. Aber die Prämisse der kontextualisierten Funktionsweise lässt Race als kontingentes Moment auch als veränderbar erscheinen (vgl. Coleman 2009 S.178). Race as Technology, so Coleman, ermöglicht Race als Subjektivierungsstrategie zu begreifen, die durch verschiedenen Werkzeuge bestimmt, wie Wissenschaft eine neue Verortung der Subjekte ermöglicht. Race as Technology erlaubt durch den ambivalenten Charakter dieser gesellschaftlichen Strukturkategorie, die dieser Theoretische Ansatz zugrundelegt, einerseits den herabwürdigenden und den kritischen Charakter eben jener zu thematisieren. Folglich impliziert diese Technologie immer auch die Möglichkeit, zu handeln, da Race gesellschaftliche Strukturen stabilisiert und auch neuschafft. Dabei ist diese Ambivalenz immer an ein Subjekt als Träger:in von Handlungsmacht geknüpft.

„Race as technology formulates race as a proactive aspect of human subjectivity that is contingent to use or application like a tool.“ (ebd. S.199).

Dieses „like a tool“ ermöglicht die Verbindung von Veränderungsmöglichkeiten und politischer Gestaltung. Hierbei geht es Coleman, um eine Voraussetzung des Menschseins, die sich durch Technologie heraus ergibt und gleichzeitig ist diese Möglichkeit des Menschseins eine Technik an sich auch als eine funktioniert.

ii.) Politische Ebene

Ein weiterer politikwissenschaftlich wichtiger Punkt ist die Frage der Ebene, auf der Gender als Technologie operiert. Hierbei verweisen die bisher genannten Autor:innen einerseits immer wieder auf die Wichtigkeit der konkreten Situationen und Konfigurationen hin (vgl. Chun 2009, S.14). Ein wichtiger Ansatzpunkt ist Wendy Chuns Idee von Race als Heuristik: „Race, like media, is also a heuristic, a way to understand, to reveal the world around us.“ (ebd.).

Dies passiert in einer ständigen Pendelbewegung zwischen dem Beispiel wie der erwähnten Hormontherapie bei Gill Person und Hesters und dessen Bedeutung für gesellschaftliche Machtstrukturen, die über den jeweiligen Fall hinausgehen und in diesem nicht mehr aufgehen. Dabei formuliert Hester eine neue spezifische politische Ebene im Zeitalter intelligenter Computerprogramme und eines weltumspannenden Internets, die mesopolitische Ebene. Diese Ebene sei durch das Protokoll gekennzeichnet, einer Technik der Wissenvermittlung. Diese zielt auf wiederholbare Vorgänge ab. Es ist eine Methode, die zwischen dem individuellen Körper und einer Makrostruktur wie dem Nationalstaat operiert. Das Protokoll steht dabei exemplarisch für die neuen Herrschaftsformen, die sich innerhalb von Netzwerken als Paradigma des Globalen Nordens seit dem späten 20. Jahrhundert, verbreiten. Das Protokoll ist auf horizontale Machtverteilung aufgebaut (vgl. Hester 2018, S. 108). Die genauen politischen Praxen und eine Definition von Protokoll werden im dritten Teil des Kapitels im Zusammenhang mit Gender als Aparat diskutiert.

Hier möchte ich auf die der Theorie zugrundeliegende Eigenheit einer politischen Zwischenebene aufmerksam machen. Diese verbindet individuelle Nutzbarkeit von Technologie mit systemischen Fragen. „It's not restricted to individuals but also includes systems: concerns how beings are subjected in systems of power, ideology and other networks.“ (Coleman 2009, S.178).

Hier komme ich zu der anfänglichen Definition von Gender als Technologie zurück. Gender als Technologie setzt eine politische Ebene voraus, die Hester Mesoebene nennt und die eine Funktionsweise impliziert, die auf ein Tun abstellt, um Technologie als Symptom von Gesellschaft zu verstehen. Gender wird so als Technologie von mir verstanden, die sich aus einem konkreten Benutzen bzw. Verwenden ergibt, jedoch auch eine Struktur dahinter einschließt, so wie die Verwendung von „male“ und „female“ als einzige Kategorien für einen Genderclassifier auf eine strukturelle Bedingtheit von hegemonialer Zweigeschlechtlichkeit hinweist, aber nicht auf eine den Menschen zugrunde liegende Wahrheit der Geschlechter.

Ich habe bis jetzt aufgezeigt, dass sich Gender als Technologie aus verschiedenen Quellen speist und auf eine Funktionsweise von Gender verweist, die mittels Werkzeugen und insbesondere digitalen Werkzeugen erzeugt wird. Dabei spielen Exklusionsmomente genauso eine Rolle wie emanzipa-

torische Erkenntnisse alternativer Handlungsweisen. Beide verweisen auf eine politische Ebene, die entlang konkreter Werkzeuge analysiert werden muss.

Im jetzt folgenden Teil möchte ich das Gendershades Projekt als ein solches konkretes Fallbeispiel einführen und dann dieses mit Hilfe der vorgestellten Theorie analysieren. Um so die Besonderheit des Genderclassifier in Hinblick auf Gender als Technologie herauszuarbeiten. Diese Besonderheit wird von mir als spezifische Technologien herausgearbeitet, als kategorialer Standard und als agentielles Protokoll.

2.) Technische Grundlagen der Genderclassifier

In diesem Abschnitt der Arbeit werde ich die technischen Voraussetzungen des Gendershades Projektes kurz erläutern.

Das Gendershades Projekt ist eine Auditstudie, die einen sogenannten Benchmark einführt. Benchmarks sind in der Forschung innerhalb von Software und besonders von KI etablierte Evaluierungsmethoden. Dabei werden Leistungen von Software wie die Lerngeschwindigkeit von neuronalen Netzwerken anhand fester Parameter, zumeist ein Kontrollalgorithmus gemessen (Nie et. Al. 2020, S.1-2). Innerhalb des Gendershades Projektes soll das „Pilot Parliament Benchmark (PPB)“ (Buolamwini, Gebru; 2018, S.5) die Akkuranz von Genderclassifiern testen.

Ich möchte zuerst die Grundlagen von Gesichtserkennungsapplikationen erläutern, zu denen Genderclassifier gehören, dann Genderclassifier kurz in diesem Feld verorten und noch einmal auf die Methode des Benchmarks zurückkommen, die Buolamwini anwendet, um die Genderclassifier zu evaluieren. Dabei beziehe ich mich ausschlaggebend auf die negriffliche Zusammenfassung von Ethem Alpaydin. Ich habe diesen theoretischen Ansatzpunkt gewählt, da er sich speziell auf maschinelles Lernen bezieht, die grundlegende Technologie hinter Gesichtserkennungssoftware. Aufgrund des Einführungscharakters des verwendeten Begriffsrahmens möchte ich ferner auf den exemplarischen Charakter der Definition der technischen Grundlagen des Gendershades Projektes aufmerksam machen. Meine Analyse der Funktionsweisen des Genderclassifier und des Benchmarks erstrecken sich auf allgemeine Grundlagen des maschinellen Lernalers, seiner statistischen Grundannahmen und der dahinterliegenden Ausgestaltung in Form von Klassifikationsproblemen, datenbasierten Anwendungen, die Buolamwini durch ein Benchmark analysiert.

a.) Automatische Mustererkennung

Die automatische Gesichtserkennung ist ein Teilgebiet der datengetriebenen Ansätze künstlicher Intelligenz. Aus Bilddateien von Gesichtern sollen Muster herausgefiltert werden, die das Abgleichen und „Erkennen“ mit bzw. in anderen Bildern ermöglichen, beispielsweise das Abgleichen eines

Passfotos mit einer Videoaufzeichnung vom Flughafen (vgl. Nassehi 2019, S.241-242).

i.) Data Mining

Für diese Mustererkennung braucht es große Mengen an Daten, dabei ist das „Data Mining“ die methodische Grundlage dieser Form automatischer Mustererkennung (vgl. Alpaydin 2016, S. 14).

Das Data Mining ist eine Funktionsweise automatischer Gesichtserkennung. Dabei werden große Mengen an Bilddaten durch Algorithmen ausgewertet und nach Mustern abgesucht. Diese Suche geschieht ohne eine theoretische Vornahme. Es gibt keine vorher festgelegte Theorie, die beschreiben kann, was ein menschliches Gesicht erkennbar werden lässt (vgl. ebd.). Der Techniksoziologe Armin Nassehi geht dabei noch weiter und erklärt maschinelles Lernen als Form einer heuristischen Art des Lernens, die weder theoriegeleitet funktioniert, noch Theorien generiert. Algorithmen sollen Muster erkennen, wie, dass menschliche Gesichter zwei Augen eine Nase und einen Mund beinhalten. Diese Muster sind zu testende Hypothesen, die sich mit neuen Daten bestätigen lassen oder aber verworfen werden. Sie lassen sich nicht aus rationalen kausalen Grundannahmen ableiten (vgl. Nassehi, 2019, S. 240)

Die extrahierten Muster sollen dabei dazu führen, dass Algorithmen Bilder zuordnen können. Dabei beruht der Ansatz der Mustererkennung auf statistischen Annahmen. Muster sind verallgemeinerbare Beziehungen verschiedener Faktoren, wie das zwei Augenbeispiel, die immer wieder durch neue Daten aktualisiert werden. Der Algorithmus passt die Muster an die vorliegenden Daten an (vgl. Alpaydin 2016, S.16-17). Gesichtserkennung basiert dabei auf der Idee, durch die Speicherung all dieser Bilder und der dahinterliegenden Informationen nicht jedes existierende Bild einer Person mit der Person in Verbindung zu bringen, sondern mit Wahrscheinlichkeiten zu arbeiten, nach der Person X auf dem Foto Person X am Flughafen sein könnte (vgl. ebd. 42).

Buolamwini hat dabei in ihrer Studie aufgezeigt, dass die Faktoren, die Mustererkennung begünstigen oder aber erschweren, in gesellschaftliche Machtverhältnisse eingebettet sind. Mustererkennung ist nicht neutral, sondern erkennt beispielsweise Schwarze Gesichter signifikant schlechter (vgl. Buolamwini 2017, S.13).

ii.) Statistik als theoretische Grundlage

Gesichtsserkennung beruht auf der Annahme, dass Menschen kein Computerprogramm schreiben können, welches Gesichtserkennung durch Menschen nachahmt. Wie ich bereits geschrieben habe unterliegt Mustererkennung keiner theoriegeleiteten Programmierung. Was meint das?

Um Mustererkennung durch Computer automatisch betreiben zu lassen, braucht es Algorithmen. In-

nerhalb der Informatik sind Algorithmen Regeln, die Input in Output verwandeln. Ein bekanntes Alltagsbeispiel, das die Funktionen eines solchen Algorithmus beschreibt, ist das Kochrezept, das einzelne Zutaten (den Input) zu einem Gericht (Output) verwandelt (vgl. Bucher 2018, S.20)

Bei der Mustererkennung jedoch gibt es keine solchen von Menschen geschriebenen Algorithmus. Stattdessen wird mit sogenannten lernenden Algorithmen gearbeitet, die ein solchen erstellen. Jedoch sind diese Algorithmen aufgrund der Vielzahl an Daten und der Unbestimmtheit des Outputs, in dem Fall Gesichter zu erkennen, Näherungsalgorithmen, sie geben unter den gegebenen Daten Wahrscheinlichkeiten an (vgl. Alpydin 2016, S.30-32). Dies ist das maschinelle Lernen. Dabei bezieht sich dieser Zweig der künstlichen Intelligenz explizit auf Statistik.

„Where going from particular observations to general descriptions is called *inference* and learning is called *estimation*. Classification is called *discriminant analysis* in statistics. Statisticians mostly work on small samples and [...] mostly worked on simple models that could be analysed mathematically. In engineering, classification is called pattern recognition.“ (Alpaydin 2016, S. 27).

Genderclassiferyer stellen beruhen dabei auf einer überwachten Form des Lernens. Menschen können ihren Output als richtig oder falsch bewerten. In dem Fall beruht Buolamwinis Benchmark auf genau dieser Idee der menschlichen Überprüfbarkeit der Ergebnisse der Genderclassiferyer (vgl. Buolamwini, Gebur 2018, S. 8).

Dabei beruhen Genderclassiferyer auf Klassifikationsproblemen, die neben Regressionsproblemen auch ein Teilgebiet des überwachten Lernen ist. Es geht darum, einen Algorithmus zu finden, der richtig klassifiziert. Dabei gibt es abgegrenzte Klassen, in meinem Fall „female“ und „male“. Diese Klassen sind durch nicht bekannte Beziehungen, die alle Objekte dieser Klasse teilen, miteinander verbunden, den Diskriminanten. Diese gilt es durch Lernalgorithmen zu entdecken, welche später in Genderclassiferyern eingespeist werden (Alpydin 2016, S.48).

b.) Gesichtserkennungssaplikationen

Genderclassiferyer sind eine Komponente von Gesichtserkennungsoftware (Buolamwini, Gebur 2018, 2) Alpaydin unterscheidet dabei beispielsweise zwei Arten von Gesichtserkennungsssoftware. Die eine spezifische Anwendung ist die Identifikation, hierbei soll Gesichtserkennungsssoftware Identifikationsmerkmale eines Gesichts erkennen, um den vorliegenden Input (beispielsweise ein Foto einer Verkehrskontrolle) mit einem anderen Foto (Passfoto) zu vergleichen und als Ergebnis (Output) eine Übereinstimmung oder aber eine Differenz anzeigen (vgl. Alpydin 2016, S. 65). Gesichtserkennung fokussiert sich dabei auf Identifikationsmerkmale wie die Augenfarbe. Diese Technik kann nach Alpydin zu Technologien der biometrischen Überwachung angewendet werden (vgl.

ebd. S.66). Biometrische Techniken verweisen dabei auf eine Kombination aus Verhaltensdaten, wie Socialmediaüberwachung und Extrahierung sozialer Daten oder aber Gender als Faktor dieser Überwachung anhand von Gesichtsbildern (vgl. ebd.).

Ein zweiter Anwendungsbereich bezieht sich demgegenüber auf veränderbare Faktoren, wie den Gesichtsausdruck, und kann beispielsweise Anwendung in der online Lehre finden, indem automatisch durch Gesichtserkennungssoftware analysiert werden kann, ob Studierende an gewissen Stellen des online Programms frustriert oder besonders motiviert sind (vgl. ebd.).

Buolamwini unterteilt demgegenüber, Gesichtserkennungssoftware anhand verschiedener Anwendungsbereiche. So gibt es Gesichtserkennung- bzw. Abgleichsoftware, diese erfüllt die von Alpaydin beschreibende Funktion des Identifizierens. Zweiten gibt es die Funktion des Anzeigens von Gesichtern, der Facedetection (vgl. Buolamwini, Gebru 2018, S.2). Diese Anwendungen sollen erkennen, ob auf den zu untersuchenden Bildern überhaupt Gesichter sind. Buolamwini benutzt diese oft vor der automatischen Identifikation genutzten algorithmische Funktion nicht. Facedetection Algorithmen sind zu stark gebiast gegen Schwarze Gesichter und Gesichter of Color. Sie „erkennen“ prinzipiell nur weiße Gesichter (vgl. ebd.). Genderclassifier sind demgegenüber Funktionen, die sich an Gesichtserkennungsapplikationen anschließen, sie ordnen automatisch Genderkategorien zu Bildern (vgl. Keyes 2018, S.88:4)

c.) Benchmarks

Wie bereits geschrieben entwickelt Buolamwini in ihrer Studie keinen eigenen Genderclassifier, sondern ein sogenanntes Benchmark. Benchmarks sind Möglichkeiten der Evaluierung von künstlich intelligenten Anwendungen, wie Genderclassifier (vgl. Ignatov et. al 2019, Ren et. al. 2020, S. 2). Sie beinhalten ein Datenset, welches von den zu beurteilenden Genderclassifiern analysiert bzw. eingeordnet wird. Dabei bekommen alle zu untersuchenden Genderclassifier dasselbe Datenset (vgl. Buolamwini 2017 S.56). Die Bilder in diesem Datenset wurden zuvor entlang einer Hautfarbenskala und einer paritätischen Genderverteilung ausgewählt und mit Genderkategorien versehen, „Femal“ für weibliche Gesichter auf Bildern, male für männliche Gesichter auf Bildern (vgl. ebd. S.55). Ich werde im anschließenden Teil genauer auf diese Zuordnung eingehen.

Diese Bilder werden dann mittels der Genderclassifier analysiert. Die Ergebnisse der Genderclassifier wird mit der Zuordnung des Bildes verglichen (vgl. Buolamwini, Gebur 2017, S. 8).

Das Benchmark ist eine Möglichkeit einer überwachten und nachvollziehbaren Evaluation. Jedoch kann das Benchmark keine Erklärung dafür bieten, warum Genderclassifier die Bilder als female oder male gelabeelt hat .

Anders formuliert, das Genderbenchmark ermöglicht eine überwachte „Lernumgebung“, in der Menschen die Ergebnisse des Genderclassifier nachvollziehen können (vgl. Nie et. al. 2020). Das Benchmark ermöglicht es aber nicht, das dahinterliegende Muster zu erkennen, bzw. die Diskriminante nach der unterteilt wird (vgl. Alpaydin 2016, S.46-47).

Nach dieser kurzen Einführung möchte ich zwei verschiedene Interpretation von Gender anbieten, die dieses Konzept mit den hier vorgestellten Technologien verbinden: Gender als kategorialer Standart und Gender als agentielles Protokoll.

3.) Gender als Kategorialer Standart

Nachdem ich im letzten Abschnitt die technischen Grundlagen des Gendershades Projektes erläutert habe, werde ich in diesem Abschnitt Gender als Technologie spezifizieren, indem ich das „Wie“ von Gender im Zusammenhang mit den technischen Grundlagen des Gendershades Projektes genauer untersuche. Ich verwende in diesem Teil die Idee des kategorialen Standards, im anschließenden Teil verwende ich die Idee des agentiellen Protokolls. Das bedeutet, ich gehe von der theoriegeleiteten These aus, dass Gender als Standard funktioniert, der sich auf die binären Kategorien Mann/Frau stützt. Somit kann ich die binäre Genderkonzeption von Buolamwini als Funktion begreifen und nicht als einen ontologischen Fakt, ähnlich der Eingangs erläuterten Idee von Race as Technologie von Wendy Chun (vgl. Chun 2009, S.8)

Ich berufe mich dabei einerseits auf die von Buolamwini selbst verwendet Bezeichnung von Gender als Kategorien (vgl.) und zweitens. darauf aufbauend, auf die von mir verwendete Theoriematrix und hier auf Butlers Idee von Gender als Norm (vgl.).

Ich werde dabei die Idee des Standards der Soziolog:innen Goffery Bowker und Susan Light Star verwenden. Diese Theorie bietet sich an, da Bowker und Star erstens eine bis heute verwendete Definition von Kategorie und Standard anbieten und diese zweitens explizit mit staatlichen Akteur:innen und regulatorischen Prozessen wie internationalen Übereinkommen verbinden (vgl. Bowker, Star 2000).

Ich werde zuerst kurz einführen in die Bedeutung der Kategorie für die sozialwissenschaftliche Analyse von Computertechnik, dann definitivisch herausarbeiten, was kategorialer Standart meint und was dieses Konzept für Gender als Technologie aus einer politikwissenschaftlichen Perspektive bedeutet.

Dann folgt die Bezugnahme des Konzeptes auf das Gendershades Projekt, anhand derer ich die These bestätigen werde, das Gender als kategorialer Standard in dem Untersuchungsfeld eingelassen ist In einem abschließenden Teil möchte ich diese These spezifisch politikwissenschaftlich untersuchen.

a.) Definition des Konzeptes Kategorie

Kategorie als Technologie ist in der sozialwissenschaftlichen Forschung zu Informationstechnologie, Computerprogrammen und Künstlicher Intelligenz ein etabliertes Untersuchungsfeld. So beschreibt die Anthropologin Lucy Suchman bereits 1993 in ihrem Aufsatz „Do categories have politics“ die Verbundenheit von Computersoftware und Kategorisierung. Dabei verweist sie auf die konstruierte Notwendigkeit der Programme, formalisiertes Wissen verarbeiten zu müssen. Formalisierung meint hierbei die Verallgemeinerung von Wissen im Sinne kategorisierbarer Aussagen wie wahr oder falsch, ja oder nein, Aktion/keine Aktion (vgl. Suchman 1993 S. 178-9). In Rückgriff auf Foucault beschreibt sie dabei auch eine politische Technologie. Für Suchman etablieren Kategorien soziale Ordnungen, die Situationen, Menschen und Interaktionen in Beziehung setzten und diese ordnen.

„That is to say, systems of categorization are ordering devices, used to organize the persons, settings, events or activities by whom they are employed or to which they refer.“ (vgl. ebd. 182). Für die Techniken, auf die sich Buolamwini und Gebru beziehen, aktualisieren die Soziolog:innen Rena Bives und Oliver Haimson diese frühe Arbeit und bezeichnen Kategorisierung als fundamentale Technologie der sozialen Medien wie Facebook. Gender ist dabei als Kategorie von wesentlicher Bedeutung, denn da, so die Autor:innen, Gender unsere Gesellschaft strukturiert, ist diese Kategorisierung auch in Software zu finden. Software hat jedoch im 21. Jahrhundert eine besondere Bedeutung innerhalb westlicher kapitalistischer Staaten. „Since our social life are now peppered with digitally mediated encounters, software components literally join as we engage in everyday social practices“ (Bives, Haimson 2016, S.2). Die Frage, wie Geschlecht in diese Applikationen eingebunden ist, ist somit eine wichtige sozialwissenschaftliche Fragestellung. Ausgehend von dieser ist Kategorisierung und mit ihr Kategorien Grundlagen für Geschlecht

Gender als Kategorisierung wird von den Wissenschaftler:innen, wie auch von Suchman, dabei weniger als unveränderlicher Inhalt von Software begriffen, sondern als soziotechnische Praxis verstanden, die Subjekte wie etwa Facebookuser:innen hervorbringt. Diese Nutzer:innen müssen sich beispielsweise, unter festgelegten Genadernkategorien verorten, um einen Account auf Facebook zu erstellen (vgl. ebd. S. 4). Hierbei unterscheiden Bives und Haimson zwischen einer staatlichen Kategorisierung von Menschen, die auf die Bevölkerungskontrolle und Politik zurückgreift und statisch erscheint und algorithmischen Kategorisierungen, die sich aus den Verhaltensweisen, Namensgebung und Onlineverhalten der Nutzer:innen ergeben und von einer Software, bzw. einem Gender-classifyer/Dector ermittelt werden (vgl. ebd. S.3). Dennoch berauben auch diese letzteren Kategorisierungsversuche User:innen der Möglichkeit der Selbstbezeichnung, da sie nicht die Wahl haben, sich einer freien Genderkategorie zuzuordnen (vgl. ebd.).

Kategorisierung ist Zusammenfassen eine Komponente, die in Softwaregestaltung, ausgehen von den angeführten sozialwissenschaftlichen Untersuchungen, einfließt, (vgl. Suchman 1993, Bives, Haimson 2016, Bives 2015). Wie aber lässt sich diese Praxis definieren? Was sind Begriffsmerkmale dieses Konzeptes? Um das zu klären möchte ich auf die Arbeiten von Bowker und Star hinweisen.

i.) Herleitung (Praxis der Klassifikation)

In ihrem kanonischen Werk „Sorting things out. Classification and its Consequences,“ formulieren die Soziologieprofessor:innen Geoffrey Bowker und Susan Light Star in den späten 1990er Jahren erstmalig den Zusammenhang zwischen Staat, moderner Informationstechnologie und Kategorisierung bzw. Klassifikation, wobei begrifflich Kategorisierung und Klassifikation in diesem Werk gleichzusetzen sind, deshalb verwende ich die Begriffe in dieser Arbeit synonym (vgl. Star Bowker 2000, S. 10). Star und Bowker sehen in Kategorisierungen ein Ordnungsparadigma moderner Gesellschaften und begreifen diese als gesellschaftstheoretische relevante Forschungsfelder (vgl. ebd. S. 31). Dies ist auch ein Grund für meine Wahl, die Definitionen von Bowker und Star zu übernehmen; denn Klassifikationen und Standards sind dabei Technologien, welche politische Implikationen haben. „There are two processes associated with these politics: arriving at the categories and standards, and along the way, deciding what will be visible or invisible within the system“ (Bowker, Star 2000, S. 44). Die Verbindung der Wahl der Kategorien und der Frage des Zuordnens seitens des Staates oder von Algorithmen wird anhand der Frage der von Staatlicher Geschlechtsanerkennung und Grenzkontrollen, die auf Genderclassifyern beruhen im dritten Teil dieser Arbeit genauer analysiert. An dieser Stelle sei jedoch schon darauf verwiesen, dass sich die Trennung von algorithmischer und staatlicher Genderkategorisierung, wie sie Bives und Haimson vertreten, innerhalb der Funktionen von Gender als Kategorie als unzureichend herausstellt.

Doch wie funktionieren Kategorien genau? Was sind spezifische Begriffsmerkmale?

ii.) Begriffsbestimmung

Für meine Verwendung von Gender als Technologie ist zuerst anzumerken, dass Bowker und Star Klassifikation und Standards als Technologie verstehen, die entscheiden, „what will be visible or invisible within the system.“ (ebd. S.44). Diese Definition von Technologie kommt wie in vorangegangenen Definition von Gender als Technologie, der Idee Butler von Gender als Sichtbarmachung nahe (vgl.).

Nach Bowker und Star beruhen Klassifikationen auf dabei drei Merkmalen. „1. *There are consistent, unique classificatory, principles in operation [...]* 2. *The Categories are mutually exclusive [...]* 3. *The System is complete.*“ (ebd. S.10-11). Klassifikationen sind hierbei Praxen, die Kategori-

en verwenden. Gender wäre somit ein Klassifizierungssystem, das auf den Kategorien Mann/ Frau; weiblich/ männlich beruht.

Auffällig ist, dass sich diese Bestimmungen von Klassifikation mit Os Keyes Inhaltsanalyse von 400 Papern zu Genderclassifer deckt, die ich bereits im ersten Teil der Analyse ausgeführt habe, und in dem they beschreibt wie Gender als Klassifikationssystem exklusiv und stabil in Form von binären Genadern kategorien operationalisiert wird (vgl. Keyes 2018, S. 88:7-8).

Bowker und Star verfolgen jedoch einen pragmatisch, materiellen Ansatz der Klassifikation und beschreiben mit den drei Prinzipien eher ein Ideal als eine Praxis. Denn so begreifen die Autor:innen Klassifikationen als Verbindungen aus verschiedenen Komponenten, wie materiellen Teilen wie Disketten²³ und immateriellen Teilen wie Arbeitsverbindungen und Kategorien des laufenden Programms, die praktiziert werden. „Here we use the term practical classifying to mean how people categorize the objects they encounter in everyday situations, including formal classification.“ (Bowker, Star, 2000, S.59).

Klassifikationssysteme beruhen dabei immer auf einer Umwelt und sind stets verankert in dieser. Sie entstehen aus bestimmten Bedingungen heraus (vgl. ebd.).

Die Kategorien innerhalb dieser beziehen sich dabei gleichzeitig aufeinander. Das System entsteht also aus einer spezifischen gesellschaftlichen Konstellation und bringt dabei auch spezifische Kategorien hervor. So ist AIDS und dessen Bedeutungswechsel in den USA nach Bowker und Star eines dieser Beispiele. So wurde die Krankheit Anfang der 80er Jahre noch aufgrund homophober Vorurteile als eine nur schwule cis Männer betreffende Krankheit verstanden und so bezeichnet (Gay Related Immune Disorder (GRID)). Das änderte sich nach ein paar Jahren und GRID wurde zu einer neuen Krankheit AIDS, welche sich nicht mehr nur auf schwule Männer sondern auch auf Kinder, cis Frauen und allgemein die weiße amerikanische Vorstadtfamilie bezog, nachdem Fälle bekannt wurden, in denen Kinder in den USA verunreinigte Blutkonserven erhalten hatten und an AIDS erkrankten (vgl. Zivi 1998, S.39-44). Dabei wirken auch hier wieder Machtstrukturen, wie Klasse und Race; denn schwarze trans Personen waren Mitte der 80er bis in die 90er Jahre nicht medial präsent als Opfer der Krankheit und somit als Schutzbedürftige in staatlichen Kampagnen beispielsweise zur Sensibilisierung (vgl. ebd.) – ein weiterer Punkt, den Bowker und Star als Merkmal ihrer pragmatischen Theorie von Klassifikation ist die Idee der Ressourcen als gestaltenden Faktor bei der Etablierung und Entstehung von Klassifikationssystemen anführen (vgl. Bowker, Star 2000, S. 39).

23 In Anbetracht des Erscheinungsjahres der Veröffentlichung erscheint Diskette hier als integraler Bestandteil der Computertechnik. In Bezug auf das Gendershades Projekt könnte man an dieser Stelle eher von Serverfarmen sprechen, in denen die großen Mengen Daten heute gelagert werden (vgl. D'Glnazio, Klein 2020, S. 42).

iii.) Ressourcen, Grenzen und das Internet

Ihrem pragmatischen Ansatz folgend, sich auf die Arbeitsweisen von Klassifikationssystemen zu fokussieren, erklären Bowker und Star diese Verortung der Systeme auch als Abwägung zwischen dem Wunsch nach möglichst genauen und vollständigen Klassifikationssystemen und Fragen der (politischen) Kosten und Speicherung eben jener Information (vgl. ebd 68). Die Forscher:innen diskutieren dies am Beispiel der „International classification of diseases ICD“ einem System zur Klassifikation von Krankheiten. Einerseits müssen diese mit ihren verschiedenen Symptomen klar präzise und aussagekräftig sein und Information zur Erkennung dieser Krankheiten bereithalten, andererseits müssen sie aber auch so gestalten sein, dass beispielsweise Ärzt:innen sie in einer gewissen Zeit sowohl lesen als auch verstehen können, um Patient:innen zu helfen.

In Bezug auf die sich in den 90er Jahren entwickelnde Computertechnologie geben Bowker und Star jedoch auch an, dass eine ubiquitäre, also sich über einzelne Problemfelder hinweg setzende Klassifikation bzw. eine Klassifikation durchsetzen könnte, die eben nicht mehr auf Beschränkungen von Speicherkapazitäten und Rechenleistung beruht und somit eine neue Dimension der Vermessung erreichen könnte (vgl. Bowker, Star 2000, S.158). Diese Aussage möchte ich zu einem späteren Punkt noch einmal aufgreifen und die Frage stellen, inwieweit die kritische Auseinandersetzung und Verortung des Gendershades Projektes als materielle-technische Assamblage im 21. Jahrhundert in einem Grenzbereich zwischen universalen Kategorien und lokalen Verortungen wie dem politischen Aktivismus von Buolamwini als neue Form der politischen Ebene verstanden werden kann.

iv.) Gender als Standard

Bis jetzt habe ich die Kategorie als Technik definiert und genauer als Praxis der Klassifikation. Hierbei stellt sich in Bezug auf die von Bowker und Star vorgestellte Definition von Klassifikation die Frage: ist Gender in diesem Fall ein Klassifikationssystem oder aber etwas anders?

Wie ich bereits erläutert habe, gehe ich von einem performativen Verständnis von Gender aus, welches sich auf Butler bezieht. Dabei schreibt Butler Gender auch eine standardisierende Wirkung zu und bezieht sich dabei auf Foucault. Der Standard ist dabei nach Butler die Funktionsweise von Gender. Der Genderstandard macht Subjekte innerhalb des sozialen Feldes sichtbar. Für Bowker und Star hingegen ist diese Sichtbarmachung sowohl der Klassifikation als auch dem Standard inhärent (vgl. ebd., S. 44). Ein Standard unterscheidet sich, bezogen auf die Soziolog:innen, dabei von einer Klassifikation in Fragen der Reichweite und Verstetigung. Wie Butler anmerkt ist Gender ein Standard, der das gesamte gesellschaftliche Feld durchdringt, nach Bowker und Star sind Standards präziser formuliert normalisierte Klassifikationen. „[S]tandards are crucial components of the larger

argument [...] A standard spans more than one community of practice [...] It has a temporal reach as well in that persists over time.“ (ebd. S.13)

So definiert als Standard beinhaltet Gender ein Klassifikationssystem, die binäre Einteilung in Mann und Frau. Gender kann aber mehr als einer sozialen Gruppe zu geschrieben werden, wie beispielsweise niederösterreichischen Senior:innen oder Londoner Punks; beide Gruppen beinhalten Männer und Frauen. Bis vor wenigen Jahren wurde beispielsweise in Deutschland auch nur Mann und Frau als Geschlechtseintrag in staatliche Register zugelassen. Dies änderte sich erst mit einem Grundsatzurteil des Bundesverfassungsgericht (vgl. BVerfG, Beschluss des Ersten Senats 10.10.2017).

Standards funktionieren aufgrund ihrer Reichweite ferner im Gegensatz zu einfachen Klassifikation auch als Normalisierungsprozesse. Normativität ergibt sich dabei folglich nicht per se aus dem unterteilen/klassifizieren, sondern aus dem Verunsichtbaren und Naturalisieren dieser Prozesse (vgl. Butler 2011, S. 74) Standards werden so in Anlehnung an Butler und Foucault normalisiert. Sie gelten als „natürliche“ Einteilung und nicht als gesellschaftliche Prozesse. Os Keyes beschreibt diese Naturalisierung in Form der unhinterfragten Operationalisierung von Gender als binär. Die Forschenden müssen Geschlecht nicht mehr definieren und als Klassifikationmodell herleiten, vielmehr erscheint die Unterteilung in Mann/Frau ein sozialer Fakt (vgl. Keyes 2018, S. 88:4).

Dennoch beruhen Standards auf Klassifikationen. Sie sind übergeordnete Klassifikationen. So beinhalten sie nach Bowker und Star einen juridischen Unterbau, wie ihn auch Butler beschreibt. Auch hängen sie in ihrer Durchsetzung von verschiedenen Faktoren, wie die Vormachtstellung auf einem speziellen Markt ab, beispielsweise die Monopolstellung von Windows als Betriebsprogramm bis in die 90er Jahre (vgl. Butler 1991, S. 202-205; Bowker, Star 2000, S.3-4).

Zwei für meine weitere Analyse wichtige Punkte spezifizieren Bowker und Star hinsichtlich der Verwendung für die digitale Sphäre.

“Standarts are deployed in making things work together over distance and heterogeneous metrics. For example, computer protocols for internet communication involve a cascade of standards [...] that need to work together well for the average user to gain seamless access to the web of information.“ (Bowker Star, 2000, S.14).

In der Analyse des Gendershades Projektes möchte ich daran anschließen aufzeigen, dass die hier von den Autor:innen aufgezeigte notwendige Zusammenarbeit, durch eine Kompatibilität des Genderstandards gekennzeichnet ist. Damit meine ich, dass der Genderstandart durch seine Allgemeingültigkeit in dem Projekt funktioniert. Diese bezieht sich auf die von Buolamiwni und Gebur genutzte binäre Geschlechtereinteilung, als auch die zweigeschlechtlich operierenden Genderclassifier

(vgl. Buolamwini, Gebru 2018). Dabei zeichnen sich Standards durch eine relative Unveränderbarkeit aus. Standards sind stabil, so Bowker und Star. Durch ihre Anwendung innerhalb gesellschaftlicher Makrostrukturen, wie dem Staat und ihrer Funktion als Übersetzungsmedien, verursacht es hohe Kosten sie zu ändern. Viele Akteur:innen beziehen sich auf diese Standards und nicht etwa nur eine lokale Gruppe. Folglich bezieht sich das Zustandekommen von Standards und ihre Aufrechterhaltung auf gesellschaftliche und nicht allein intersubjektive Prozesse (vgl. ebd. S. 13)

Bezogen auf Gender als Standard wird die Frage der Diversifizierung von Genderkategorien in digitalen und smarten Technologien ein Akt kollektiver politischer Bemühungen, der in ein Netz aus Firmeninteressen, staatlichen Interessen und normierten Softwareapplikationen eingelassen ist. So beschreibt Rena Bives Facebooks Entscheidung über die wählbaren Genderkategorien für User:innen als Abwägung zwischen Werbekund:innen, bestehende Datenbestände und Fragen bezüglich der Akkuratheit der Genderkategorien. Diese Entscheidungen sind konfliktbeladen und Facebook verändert sie ständig (vgl. Bives 2015, S. 889)

v.) Exklusionsmechanismen

Eine weitere Wichtige Funktion von Standards nach Bowker und Star ist die Grenzziehung, die Standards erzeugen (vgl. Bowker, Star 2000, S. 14). Standards produzieren ob ihrer gesellschaftlichen Erzeugung Exklusion und wie Chun in Verbindung zu Race als Technologie deutlich macht, erzeugen sie auch Hierarchien (Chun 2009, S. 40).

Standards sind demnach, weil sie auf Strukturen beruhen, keine Selbstbezeichnungssysteme, in denen sich Menschen autonom einordnen. Vielmehr sind Standards Mechanismen, die soziale Ordnungen herstellen und das mittels verschiedener Akteur:innen, jedoch nicht mittels der bezeichneten Individuen. Für Bowker und Star sind deshalb staatliche Akteur:innen elementarer Bestandteil von Standards. Diese Akteur:innen stabilisieren Standards durch zumeist juridisch abgesicherte Grenzen, wie erlaubte Geschlechter am Beispiel des Bundesverfassungsgerichtsurteils. Butler weist in diesem Kontext staatlich juridischer und medizinischer Typologisierung des schon vorgestellten CDI darauf hin, dass die Diagnose einer Transgeschlechtlichkeit nicht der Selbstbezug einer Person auf die persönliche Transidentität darstellt, sondern eine Technik, die diese Person einordnet und zu Patient:innen macht. Diese somit klar von cis geschlechtlichen Menschen abgrenzt.

„Die Diagnose [Transgender zu sein] versucht, Kriterien aufzustellen, mit denen eine Person bestimmt werden kann, die sich mit dem anderen Geschlecht identifiziert. In dem sie die Kriterien dafür artikuliert, formuliert sie aber eine sehr rigide Version der Geschlechternorm.“ (Butler 2011, S. 157).

Gender wird, nach Butler, zu einem Standard für psychiatrische Intervention und bestätigt sich

durch die Anwendung durch Ärzt:innen immer wieder selbst, in dem abweichende Gender die scheinbare Notwendigkeit der Korrektur dieser Identitäten verlangt bekräftigt(vgl. ebd.)

Das macht politische Bemühungen, um wieder auf das Gendershades Projekt zurück zu kommen, für einen (a-, trans-, queeren) genderinklusiven Genderclassifier zu einem ambivalenten Unterfangen, den auch hier muss entschieden werden, wem welches Gender zugeordnet wird und von wem. Gleichzeitig ist der binäre Genderstandard und damit potentielle genderqueere Versuche diesen Standard abzuändern nicht nur in Praxen politischer Bewegungen eingelassen, wie der Kampf um beispielsweise die Anerkennung des dritten Geschlechts in Deutschland, sondern beruht gleichzeitig auf technischen Strukturen wie Software, wie Bives und Haimson verdeutlichten.

Gender ist zusammenfassend in diesem Abschnitt der Arbeit ein Standard, welcher sich auf binäre Kategorien stützt. Dieser ermöglicht, bezogen auf Computersoftware, aufgrund des Forschungsgegenstands, eine Systemkompatibilität von verschiedenen Softwareanwendungen und ist eine Technologie, die bezeichnet und so die User:innen bzw. die zu analysierenden Personen sichtbar werden lässt oder aber aus Netzwerken und Datensets verschwinden lässt, wie das Beispiel der Facebookkontos von Bives und Haimson verdeutlichte. Der Standard ist dabei stets eine verortete Praxis, die sich aus dem Wirken von Software, Akteur:innen und gesellschaftlichen Kontexten heraus entwickelt hat und selbst rekursiv stabilisierend auf diese einwirkt.

Im folgenden Abschnitt, der Verortung des theoretischen Konzepts im Gendershades Projekt, möchte ich auf diese Praxis des Standards eingehen und erklären in wieweit sich Gender als Standard mit binären Kategorien im Boulamwini Projekt wiederfindet. In einem dritten Teil werde ich dann einerseits analysieren, welche politikwissenschaftlichen Implikationen der Genderstandard im Gendershades Projekt aufweist und welche neuen Möglichkeiten der Technikgestaltung sich aus solch einer Betrachtungsweise entwickeln lassen.

b.) Empirie

Im Folgendem werde ich die Funktionsweise der technischen Gestaltung des Gendershades Projektes analysieren. Ich gehe dabei auf zwei Aspekte ein: erstens die Verwendung von Gender als binäre Kategorie durch Buolamwini und zweitens Gender als Standard, der eine Vermittlung des Datensets der Wissenschaftlerin und der evaluierten Genderclassifier ermöglicht.

i.) Gender als Kategorie des PPB

Das Benchmark, das Buolamwini kreiert, sowie die zu analysierenden Genderclassifier, verwen-

den eine binäre Kategorisierung von Gender. Buolamwini benennt diese Verwendung von Gender explizit; „An evaluation of gender classification performance currently requires reducing the construct of gender into defined classes“ (Buolamwini, Gebru 2018, S.3). Die Wissenschaftlerin kreiert ihr Benchmark so entlang den Kategorien Mann/Frau. Dabei sind diese Kategorien exklusiv, das bedeutet ein Bild einer Person wird einer Kategorie zugeordnet. Die Kategorisierung ist des weiteren vollständig, alle Bilder der Abgeordneten des jeweiligen Parlamentes wurden in dieses Schema eingestuft. Ferner ist sie eine geschlossene Kategorisierung, da die Personen während der Anwendung bzw. der Kategorisieren durch die Genderclassifier und nach dieser Prozedur nicht geändert werden (vgl. ebd. S. 8-10). Die genauere Analyse der untersuchten Genderclassifier kann jedoch weder ich noch Buolamwini evaluieren, da diese nicht zugänglich sind. Es ist mir nicht bekannt ob, die beispielsweise von Face++ verwendeten Datensets für die Programmierung der Genderclassifier Gender queere Personen in ihren Daten enthalten haben und diese dann herausgefiltert worden sind, als statistische Fehler eventuell, oder ob diese überhaupt nicht in den Samples auftauchen (vgl. Buolamwini, Gebru 2018, S.6; Buolamwini 2017, S.55). Ich vermute ausgehend von Studien zur Kategorisierung von Gender in Software, insbesondere zur automatischen Gendererkennung, dass die Binarität in den Designprozess mit eingeflossen ist und genderqueere Personen nicht erkannt wurden (vgl. Bives, Haimson 2016, Bives 2015, Keyes 2018, Johnson 2015). Aber aufgrund der nicht vorhandenen Daten bleiben das Spekulationen und sollen somit weiterhin nicht Gegenstand der Analyse sein.

Zum jetzigen Standpunkt bleibt festzuhalten das Gender die Funktionsweisen einer Kategorie erfüllt, nach der vorgelegten Definition von Bowker und Star. Ich möchte jetzt noch einen Schritt weiter gehen und argumentieren, dass Gender hier nicht nur als Klassifikation, sondern als Standard dient.

Gender als Standard

Wie bereits erläutert verweisen Kategorien auf verschiedene Wirkungsweisen. Ich habe hier die klassische Unterscheidung von Bowker und Star gewählt, weil sie einerseits aufzeigen, dass Kategorien in verschiedenen Zusammenhängen unterschiedlich wirken (als Klassifikationen oder Standards). Andererseits beziehen sich die Soziolog:inne verstärkt auf Foucaultsche Überlegungen zu Normierung als biopolitische Praxis, die ich selbst in der Spezifizierung von Gender als Norm anwende.

Meine These lautet daher, dass sich Genderclassifier und das Benchmark auf Gender als Standard und nicht als Klassifikation beziehen. Standards sind, wie schon beschrieben eine Art übergreifende Klassifikation, als solche beinhalten sie spezifische Funktionsweisen. Sie sortieren Informationen

(Bowker, Star 2000, S.10). Sie normieren diese Informationen (vgl. ebd.) und sie verallgemeinern sie (sie machen sie für verschiedene Kontexte lesbar) (vgl. ebd. S.14).

ii.) Sortierung

Die Sortierung als Funktion beziehe ich auf zwei Ebenen des Gendershades Projektes, einmal auf das Benchmark, das Buolamwini erstellt hat und zweitens auf die Funktionsweise datengetriebener automatischer Genderclassifier bzw. deren Arbeit mit Datensets wie dem Benchmark

Gender als Standard sortiert in dem Gendershades Projekt konkret Bilder von Abgeordneten. So ergibt sich aus der Stabilität der Kategorien, die sich innerhalb des Benchmarks zu keinem Zeitpunkt nach der Festlegung der Kategorien „male“ „female“ ändert, die Möglichkeit einer stabilen Ordnung, die Gender als Rahmen nutzt. Bilder können so in einer festen binären Struktur für Evaluation der Kategorie Race genutzt werden.

Die Sortierung zeigt sich, wenn man sich die Visualisierung des Gendershades Projektes in Abb. 3 anschaut.

Hierbei verteilen sich die Bilder der Abgeordneten und geben somit auch Informationen preis. Folglich wird aufgrund der Einsortierung in die binären Klassen „female“ und „male“ die Evaluation der Genderclassifier erst möglich (vgl. Buolamwini, Gebru 2018, S. 9). Denn ungeordnete (in diesem Fall gelabelte) Bilder könnten keinen Aufschluss darüber geben, ob und inwiefern die Genderclassifier akkurat Gender automatisch ordnen können, bzw. würden sie eine Evaluation verunmöglichen, die das korrekte Zuordnen von Gender erlauben soll. Anders ausgedrückt, nur durch die binär geordneten Daten im PPB kann eine Evaluation und damit eine kritische Intervention wie Buolamwini und Gebru sie einfordern, stattfinden.

Ferner und dies ist in Bezug auf die großen Mengen an Daten ein essentielles Kriterium: In dem Kontext künstlicher Intelligenzen, wie ich sie eingeführt habe, werden Daten maschinell ausgewertet und aus ihnen Muster herausgelesen (vgl. Alpaydin 2016, S.23-24). Diese automatische Auswertung braucht bestimmte inhaltliche Ordnung in Form festgelegter Referenzwerte, wie bei Genderclassifier Gender als Untersuchungsgegenstand und binäre Genderkategorien als Zielorte der maschinellen Mustererkennung. Gender ist in seiner strikten Kategorisierung und der Exklusivität solch ein Moment.

iii.) Normierung

Im folgendem Teil möchte ich die von Bowker und Star erwähnte Normierung als Funktion von

Standards in Bezug zum Gendershades Projekt analysieren. Dafür werde ich jedoch kurz Normierung genauer erläutern. Ich fokussiere mich hierbei nicht auf Butler oder Foucault, sondern auf den Disability Theoretiker Lennard J. Davis. Dieser bezieht das Konzept der Normierung konkret auf die im 19. Jahrhundert aufkommende Technik der Statistik, die das Normprinzip als gesellschaftliche Machtfunktion etabliert. (vgl. Davis 2017.)

Dabei ist die Norm im Gegensatz zum Ideal dadurch gekennzeichnet, dass ein Teil der Bevölkerung in die Norm fällt (vgl. ebd. S.3). Dabei entwickelt sich die Normidee, die Davis untersucht aus der Idee der Statistik, Bevölkerung zu normieren und sie somit in Standardkohorten aufzuteilen. “An important consequence of the idea of the norm is that it divides the population into standard and nonstandard subpopulations. “ (ebd. S.3-4). Davis verweist hier auf zwei wichtige Punkte, erstens die Verbindung soziodemografischer Marker wie etwa Körpergröße oder Alter mit der Norm und zweitens auf den Zusammenhang von Statistik als Methode diese hierarchisch zu gliedern. Hinsichtlich des Gendershades Projektes kann ich das Labeln der Bilder aufgrund einer solchen Definition als Normieren bezeichnen, da das binäre Genderkategorie Modell hier die vorherrschende Norm aufgrund von angenommener statistischer Relevanz nach Buolamwini ist und zweitens, weil diese Vorstellung von Norm an statistische Methoden geknüpft ist, die gerade für die hier untersuchten Verfahren der Gendererkennung eine tragende Rolle spielen. Denn diese Ansätze von Künstlicher Intelligenz, welche Buolamwini in Form von Genderclassifyern analysiert, beruhen auf Wahrscheinlichkeiten als Ergebnis (zu wie viel Prozent liegen Genderclassifier richtig bei der Erkennung von Gender) und zweitens beruhen sie auch grundlegend auf Statistiken als Arbeitsweise, in dem sie aus einer großen Anzahl Bilder statistische Muster ableiten (vgl. Buolamwini 2017, S.55, Appledyn 2016, S. 46-47)

Ausgehend von der Sortierung ergibt sich somit die Möglichkeit der Normierung in einem soziotechnischen Sinn: Statistische Prozesse innerhalb der automatischen Bildverarbeitung verbinden sich mit Genderideen über den vermeintlichen Standard binärer Cisgeschlechtlichkeit.

Gender als normierender Standard erfüllt folglich im Gendershades Projekt eine Doppelfunktion. Gender wird hierbei bezogen auf die normierende Funktion: erstens zu einem Gradmesser (Standard als Bewertungskriterium), der Performance von Genderclassifyern und zweitens zur Norm, die vor dem Einsetzen der Genderclassifier Bilder sortierbar und einordenbar für diese macht. Das bedeutet, einerseits ermöglicht der binäre Genderstandard die klare Zuordnung der Bilder innerhalb des PPB zu festen Begriffen (male/female) und somit einen Bewertungsmaßstab für die daran anschließende Bewertung der Ergebnisse der Classifier (vgl. Buolamwini, Gebru 2018, S. 6-8). Dieser normierende Genderstandard ordnet die Bilder des PPB als entweder falsch oder richtig „erkannt“ ein. Gender wird zum Bewertungsmaßstab der Performance der Genderclassifier und etabliert sich fer-

ner als entpersonalisierte Bewertungsstruktur innerhalb des Gendershades Projekts. Nicht mehr die einzelnen Personen Gebru oder Boulamwimi entscheiden darüber, ob der Genderclassifier richtig entschieden hat, sondern eine vorsortierte Menge an Fotos, die binär codiert sind. Normalisierende Prozesse wie Davis sie versteht, beziehen sich deshalb auf hegemoniale Praxen gesellschaftlicher Ordnungen. Gender als Standard verweist nicht auf eine individuelle Wahl Buolamwinis, sondern auf die hegemoniale zweigeschlechtliche Struktur, die westlich kapitalistischen Gesellschaften inneohnt (vgl. Ludwig 2011, S.191-7).. Standards sind als normalisierende Prozesse Technik, die in diesen Gesellschaften eingewoben sind als „part of the a notion progress, of industrialization“ (Davis 2017, S. 12,)

Die Normierung entzieht sich somit der rein menschlichen Einflussnahme von Buolamwini und Gebru, vielmehr werden sowohl das Erkennen des Genders (durch Genderclassiyfer) als auch die Bewertung (durch den Abgleich mit dem Benchmark) zu posthumanen Phänomenen. Dabei wirken Menschen, Gender als Technologie und konkret Tools wie das Auditdatenset und die Genderclassifier in verschiedenen Positionen. Wobei Gender als gewählter Evaluierungspunkt hierbei die Norm der Richtigkeit der Genderclassifier darstellt.

Damit wird „Gender queer“ jedoch zu einem nicht maschinenkompatiblen Fehler, welcher sich schon vor der Evaluierung der Genderclassifier herausfiltern lässt und auf das Paradox hinweist, einerseits durch statistische Methoden und Technologien, Standards zur Gendererkennung zu setzten und diese maschinell auch zu performen im Sinne des automatischen Gender Zuordnens, andererseits werden Gender, die nicht cis Männlichkeit oder Weiblichkeit „sind“, zu unsichtbaren Subkategorien, die zwar die Abgrenzung zu female/male erlauben, aber nicht die Einbeziehung in solche Standards (vgl. Davis 2017, S.4). Welche politischen Potentiale dieses Paradox für eine genderinklusive Politik des Gendershades Projektes bietet, möchte ich in der anschließenden Diskussion im letzten Teils dieses Abschnittes erörtern. Ich möchte jetzt noch einmal die Funktion von Gender als Standard diskutieren, welche die Zusammenarbeit von Genderclassifier und Benchmark ermöglicht.

iv.) Standardisierung als Wissenschaftlichkeit

Das ein Standard standardisierte Wirkung entfaltet, wirkt tautologisch. Ich möchte diese Wirkung jedoch für die wissenschaftliche Arbeit von Buolamwini spezifisch operationalisieren und den Genderstandard als Voraussetzung für die Reliabilität des Benchmark analysieren und zweitens den Genderstandard als Merkmal der Validität dieser Auditstudie von Buolamwini definieren. Dabei beziehe ich mich auf die Wirkung des Standrads als statistisches Werkzeug der Verallgemeinerung und des Transfers von Wissen zwischen verschiedenen sozialen Gruppen, aufgrund meiner gewähl-

ten Definition des Standards (vgl. Bowker, Star 2000, S. 13, Alpaydin 2016, S.27).

Der Standard, so Bowker und Star, entfaltet dabei einen spezifischen Nutzen für die Statistik, er ermöglicht die „stabile Lagerung“ von Informationen (vgl. Bowker, Star 1999, S.58). Ich möchte in den folgenden beiden Abschnitten diese Funktion genauer analysieren.

α.) Der reliable Genderstandard

Wie Bowker und Star formulieren, kommt dem Standard die Aufgabe zu, Informationen für verschieden Wirkungszusammenhänge lesbar und verwendbar zu machen (vgl. ebd. S.14). Diese Funktion möchte ich hier als Reliabilität definieren, da es sich um Konzepte der Nachvollziehbarkeit und Übertragbarkeit durch Wiederholbarkeit wissenschaftlicher Arbeiten handelt. Dies ist Güterkriterien quantitativer Forschung

Als Definition für Reliabilität dient folgendes Zitat: „Reliabilität (Zuverlässigkeit) einer Messung wird definiert als das Ausmaß, in dem wiederholte Messungen eines Einstellungsobjekts zu gleichen Werten führen.“ (ebd. S.427).

Die Definition habe ich gewählt, da sie aus einem Einführungswerk der empirischen Sozialwissenschaft stammt und somit als Grundwissen des wissenschaftlichen statistischen Arbeitens dienen kann, andere Zitate wären jedoch auch denkbar. Das Zitat ist hier aufgrund seines Beispielcharakters für diese Definition ausgewählt worden.

Ich möchte hier darstellen, dass Gender als Standard aufgrund einer natürlichen Begebenheit von Gender als binär nicht reliabel ist, sondern dass Gender als Standard die Voraussetzungen schafft, das das Gendershades Projekt, genauer das Benchmark, als wissenschaftliche Arbeit diese Reliabilitätskriterien erfüllt und so die von Bowker und Star formulierte Übersetzung von Wissen schafft. Dies geschieht durch die Wiederholbarkeit der Operationalisierung der Kategorie Gender (vgl. ebd., S. 14).

Geschlecht ermöglicht so erstens die Verbindung zwischen Benchmark und den ausgewählten Genderclassifern, indem Gender eine Bewertung der drei Genderclassifern entlang der standardisierten Kategorien „female“ und „male“ ermöglicht. So schafft der Genderstandard die Voraussetzung für die Zuverlässigkeit des Benchmarks als Evaluierungswerkzeug, nämlich die Vergleichbarkeit verschiedener Softwareapplikationen anhand eines festen Vergleichsrahmens (vgl. Buolamwini, Gebru 2018, S.4). Gender ist dabei ferner auch ein wissenschaftliches Gütekriterium, das Verständlichkeit für ein breites Fachpublikum ermöglicht, welches die Methodik des Benchmarks des Gendershades Projekts anhand der von Buolamwini aufgeführten Operationalisierungen von Gender und dessen Verwendung für das Daten Set nachvollziehen und wiederholen könnte (vgl. Buolamwini 2017,

S.55). Dabei wird auch deutlich, dass der Genderclassiferyer und die Benchmark durch den Genderstandard verbunden werden. So würde ein Genderclassiferyer mit spezifischen Genderkategorien jenseits der Mann/Frau Aufteilung nicht mit dem Benchmark von Buolamwini untersucht- bzw. evaluierbar sein, andersherum wären die untersuchten Genderclassiferyer nicht durch ein genderqueeres Benchmark analysierbar gewesen (vgl. Buolamwini, Gebru 2018, S.8).

Zweitens wäre fraglich, ob sich die im Sample befindlichen Bilder auf andere Gender Kategorien als die von Buolamwini verwendeten Genderkategorien beziehen könnten. Wenn alle Bilder aufgrund der Bezeichnung auf den Parlamentseiten als Mann oder Frau gelabelt werden, erscheint es als nicht stringent, diesen Menschen andere Kategorien zuzuordnen. Ich möchte im Folgendem darauf aufbauend die zweite Funktion des wissenschaftlichen Genderstandards erörtern.

β.) Gender als valider Standard

In diesem Abschnitt möchte ich auf ein zweites Gütekriterium statistischer Forschung eingehen, die Validität. Ich möchte dadurch die Idee der Funktionalität des Genderstandards schärfen. Ich beziehe mich dabei auf das theoretische Konzept von Landström, die Grundlage meiner Theoriematrix, in dem die Wissenschaftler:in beschreibt, dass Gender als Funktion und als eine Frage des „Wie“ verstanden werden muss. Diese Idee habe ich durch die Interpretation Gender als Technologie, wie ich sie in der Einleitung dieses Kapitels formuliert habe, genauer bestimmt. Ich möchte nun eine bestimmte Funktion von Gender herausarbeiten. die des validen Genderstandards

Doch was meint Validität und wie trifft dieses Konzept auf Gender innerhalb des Gendershades Projekts zu?

Nach den Regeln für wissenschaftliches Arbeiten an der Universität Wien meint Validität als Gütekriterium statistischer Forschung: „**Validität** liegt vor, wenn die gewählten Indikatoren, Fragen und Antwortmöglichkeiten wirklich und präzise das messen, was gemessen werden soll.“ (<https://www.univie.ac.at/ksa/elearning/cp/quantitative/quantitative-34.html> zuletzt abgerufen 04.12.2020)

Validität bezeichnet hier eine Beziehung von Untersuchungsgegenstand und Untersuchungsdesign. Gender ermöglicht bezogen auf das Gendershades Projekt, das Einsetzen von den Indikatoren Mann/Frau als Messindikatoren der Performance der Genderclassiferyer. Durch das Benchmark wird so genau die Funktion der Genderclassiferyer männlich und weiblich zu erkennen gemessen (vgl. Buolamwini und Gebru 2018, S.8)

Eine Änderung dieses Standards bei gleicher Funktionalität, also dem Erkennen von Gender, würde anders formuliert zu einer Änderung von Genderclassiferyer, Bezeichnungen für Parlamentarier, Ver-

gleichskategorien und vieles mehr führen, um valide messen zu können, was nicht bedeutet, dass dieses nicht möglich oder nicht wünschenswert wäre. Es würde jedoch bedeuten, dass sich dadurch der Inhalt des Genderstandards verändert.

Dies meint im Umkehrschluss, dass Kategorien, wie beispielsweise agender oder genderqueer hinzuzufügen nicht bloß eine formale Änderung der Codierung des Benchmarks enthalten würde bzw. das Verändern von binär standardisierten Genderkategorien nicht einfach bloße Addition von Kategorien beinhaltet. Das Inkludieren von genderqueeren Kategorien ist eine soziotechnische Frage, die das gesamten Forschungsdesigns und dessen Umwelt beinhaltet. Anders formuliert, Genderclassifyer und das Benchmark brauchen einen binären Standard, um als Auditstudie zu funktionieren (vgl. Bowker, Star 2000, S.14). Wie ich eingangs formulierte, ist der Genderstandard dabei aber nicht etwa der Ausdruck einer gegebenen binären Genderaufteilung, sondern durch seine Funktionalität als Grundlage wissenschaftlicher Gütekriterien zu verstehen.

Ich möchte hier überleiten zur politikwissenschaftlichen Dimension dieser Funktionen von Gender als binärer Standard. Denn wie ich bereits ausformuliert habe, verweist die spezifische Funktion von Gender als Technologie innerhalb konkreter Fallbeispiele auf politische Faktoren, wie etwa Hierarchisierung, Exklusion und Fragen der Subjektivierung, auch beinhaltet sie politische Faktoren, wie ich sie in der Definition von Gender als Technologie angeführt habe (vgl. Kafer 2017, Chun 2009, S. 8-9).

Ich möchte dafür in folgenden Teil noch einmal auf die eingeführte Frage der staatlichen und algorithmischen Genderkategorisierung eingehen und aufzeigen inwiefern diese Unterscheidungen, die Bivess und Haimson aufmachen., im Falle des Gendershades Projektes nicht standhalten. Vielmehr, so argumentiere ich, verbindet Gender aufgrund seiner Funktion staatliche und algorithmische Genderkategorisierung.

c.) Polittheoretische Analyse

Standard, wie ich ihn eingeführt habe, ist seiner Definition nach Star und Bowker als politische Technologie zu verstehen. Dabei ist ein Anhaltspunkt für diese Auslegung des Konzeptes die starke Verwobenheit von Standard als Technologie mit Staatlichen Strukturen als Absicherung (vgl. Bowker, Star 2000 S.44). Ich möchte im Folgendem deshalb in der politikwissenschaftlichen Analyse das Verhältnis von Staat und Genderstandard ausgehend von einer transfeministischen Perspektive hinsichtlich Gender spezifizieren und im Anschluss meinen Fokus auf Genderclassifyer als standardisierte Technologien staatlicher Genderpolitik lenken.

Dabei sind Standards als Untersuchungsfeld der Politikwissenschaft nicht ungewöhnlich und fo-

kussieren sich auf staatliches Handeln rund um Standards (vgl. Johnson 2018, Jasanof 2010, Scott 1999). Foucault, der schon in meinem Theorieteil als Referenz diente, beschreibt mit seinem Ansatz der Biopolitik, also der Bevölkerungsregulierung moderner Staaten anhand von Sex(ualität), einen Standard zur Subjektivierung von Bürger:innen im modernen Staat, jedoch ohne auf Informationstechnologie einzugehen. Die technischen Mittel der Erstellung und Aufrechterhaltung einer Norm sind hierbei Schulen, Gefängnisse, Patientenakten und Censusdaten. Butler wird hierbei für meine Fragestellung noch konkreter und bezieht dies auf Gender, verbleibt jedoch wie Foucault in einer staatlichen Betrachtung der Genderregulierung entlang analoger Werkzeuge. Sie konzentriert sich auf Gesetze, wie Heiratsgesetze, Familienkonstellationen im psychoanalytischen Untersuchungsrahmen und auch ärztlichen Diagnosen (vgl. Butler 2011, S.97-ff.)

Ein akademisches Feld, welches die Standardisierung von Geschlecht und Staat vielfältiger bzw. mehrstimmig ausleuchtet, sind die Trans(gender)studies, ein weiteres interdisziplinäres Forschungsfeld, welches sich mit Transgender als Untersuchungsgegenstand beschäftigt, jedoch spezifisch ausgehend von Transgender als Abweichung des Genderstandards. Fragen der Standardisierung, als Form staatlicher Gewalt, werden hier prioritär verhandelt (vgl. Fütty, 2019, S.18-19).

Dabei hat sich durch die AIDS Krise der 1990er Jahre die Bedeutung dieser kritischen akademischen Disziplin geschärft (Stryker, Aizura 2013, S.1) Ein besonderer Schwerpunkt ist das Verhältnis staatlicher Anerkennung und die Normierung bzw. die Subversion entgegen solcher Standards. So beschreibt Susan Stryker, eine Begründer:in der Disziplin in ihrer Einführung zu Transgender Studies, (Trans-)Gender aus dieser theoretischen Perspektive explizit als Ort der Aushandlung von Staatsbürger:innenschaft und des Staates an sich.

„*transgender* is intimately bound up with questions of nation, territory and citizenship, with *categories of belonging and exclusion* (eigene Hervorhebung), of excess and incorporation [...] All current transgender phenomena are thus local and national phenomena that encounter transnationalizing forces.“ (vgl. ebd. S.8).

So stehen binäre Standards ausgehend von den Transstudies exemplarisch für eine Grenzpraktik und resultierend aus der Grenze zwischen Cis und Trans auch als Grenze des Staates, wie die in Yale lehrende Literaturwissenschaftlerin Gayle Salamon anhand des Adverbs *trans* ausführt.

„First, understanding transgender as a “border crossing” figures gender itself as a kind of bounded territory—gender not just as a property of bodies but as property itself—and I’d like to explore both the possibilities and the difficulties of this metonymic landscape“ (Salamon 2010, S.172).

In diesem Abschnitt macht Salamon bewusst auf die Verkettung der Trope des Grenzübertritts des Trans Individuums und den Staat als Ort der Grenze aufmerksam. Später beschreibt sie diesen Grenzprozess exemplarisch am Beispiel des Eintritts in ein anderes Land am Flughafen.

„The peril comes if she is “read” as transsexual, as having no proper gender at all. We can understand this moment of anxious hesitation before the gendered bifurcation of the airport security line as exemplifying the process of gender itself. That process is neither purely bodily nor purely volitional, but created—and enforced—through other peoples’ expectations and interpretations of one’s bodily appearance and behaviour.“ (ebd., S. 181).

Dieses Beispiel recurriert auch wieder auf Strykers Verständnis von Transgender als politische Identität der Globalisierung und ist in Hinblick auf die an Flughäfen heute verwendete Technologie der Körperscanner und Genderrecognitiontools auch für meine Forschung im speziellen ein wichtiges Beispiel.

Die queer Theoretikerin Nicole Shephard spezifiziert in ihrem Aufsatz „Queer enough? Categories and norms at the border“ die Verbindung von Gesichtserkennungsoftware und (Trans)Gender, wie Salamon es versteht, als Grenzregime. Dieses unterscheidet zwischen Staatsbürger (hier bewusst gegendert)/ Fremden/ oder aber einen asylsuchenden Menschen der „queer enough“ für den besonderen Schutz durch das Gesetz ist. Dabei kommt es zu einer Art digitaler analoger Assemblage, ein gemischt aus automatischen biometrischen Technologien wie der Gesichtserkennungsoftware oder probeweise der Genderclassifier und Foucaultianischer Befragungsbögen, die die „wahre“ Sexualität als Akt des Geständnisses vor dem Staat hier vertreten durch die Grenzbeamten, preiszugeben verlangen/vorgeben. Dabei ist die standardisierte und rassifizierende Idee eines zu erkennenden queeren Subjektes Grundlage

“Gender and sexuality are mobilized as border technologies in the service of sorting non-normative bodies into deserving refugees and fraudsters based on (Western) stereotypical notions of queerness like ‘coming out’ or attending gay pride parades.” (vgl. Shepard <https://xyz.informationactivism.org/en/queer-enough/> zuletzt abgerufen am 19.11.2020).

Die Theoretikerin Nadine Dannenberg beschreibt diese Verkettung von Gender und staatlichen Grenzen als paradigmatische Verbindung für queere Theorie.

„Dasm was in den Surveillance Studies mit Fokus auf den Verwaltungs- und Kontrollcharakter der Bevölkerung herangezogen wird, dient der Queer Theory zur Ausarbeitung einer Theorie und Gesichte des Sex selbst als Verwaltungs- und Kontrollinstanz, die multiple Überwachungsmodi ebenso antreibt wie legitimiert.“ (Danneberg 2019, S. 30)

Die Transtudies bieten hier in Bezug auf das Thema meiner Forschungsarbeit und der Fokus des Gendershades Projektes noch einmal einen spezifischeren Fokus, der sich nicht so sehr auf Sexualität, sondern auf Gender als eigenständige Kategorie bezieht (vgl. Halberstam 2018). Um auf die

empirische Ausgangslage meiner Arbeit zurückzukommen bleibt die Frage, wie staatliche Grenzbeziehungen als dem Genderstandard inhärentes Moment mit dem Gendershades Projekt zusammen gedacht werden können und was das spezifische an dieser Verbindung ist.

Hierfür möchte ich auf die besondere Verwendung von Gesichtserkennungstechnologie und den Staat des 21. Jahrhundert eingehen.

Wie bereits deutlich wurde sind Genderstandards insbesondere in Bezug auf staatliche Praxen, wie das Begrenzen der Geschlechtsoptionen und die Regelung zu medizinischer Versorgung ein gut untersuchtes Terrain (vgl. Baumgartinger 2017). In Rahmen des 21. Jahrhunderts übernimmt jedoch immer mehr Software diese Aufgaben des Einordnens, in Form von biometrischen Techniken, die biometrische Daten wie Gender/ Alter/ race anhand von Gesichtserkennungssoftware und Körperscannern erfüllen (vgl. Pugliese 2014). Dabei werden diese Erkennungstechniken sowohl auf Überwachung von öffentlichen Plätzen, Terrorist:innenabwehr und -erkennung, aber auch in Anbetracht der aktuellen Anlässe zur Vorhersage von Pandemieverläufen verwendet (vgl. Weber, 2018; Galloway, Thacker 2014, S.289; Puar 2007). Diesen Technologien sind in ihrer Vielfältigkeit (bezogen auf Anwendungsgebiete und Untersuchungsgegenstände) vorrangig die zugrundeliegenden Informationsverarbeitungsmechanismen gemeinsam, auf denen explizit auch Genderclassifiers beruhen, nämlich auf datengetriebenen Ansätzen, Künstlicher Intelligenz., sogenannter Big Data Technologie. Die Soziolog:innen danach boyd und Kate Crawford weisen in einer Kritik an diesen Technologien auf eine paradigmatische Verschiebung der Verarbeitung von Daten als Erkennungsmerkmal von KI hin. „Big Data is less about data that is big than it is about a capacity to search aggregate and cross reference large data sets.“ (vgl. boyd, Crawford, 2012, S.663)

Im Gegensatz zu der Sichtweise von Foucault, der moderne Staaten seit dem 19. Jahrhundert, durch das Sammeln großer Mengen soziodemografischer Daten gekennzeichnet sah, sind Applikationen wie Genderclassifiers weniger durch ihr Vermögen gekennzeichnet, große Mengen an Daten zu generieren, als sie zu verknüpfen. boyd und Crawford weisen in diesem Zusammenhang darauf hin, dass mit Daten getriebene Künstlicher Intelligenz sogar weniger Daten verwendet werden als beispielsweise in einem Zensusdatensatz vorhanden ist. Foucault stellt dabei den modernen Staat als Daten-Generierungs-Apparat dar, der Daten erstmalig in einer großen Anzahl erhebt. Mit diesen arbeitet Künstliche Intelligenz als ein neues Tool staatlichen Handelns hingegen auf Basis der Möglichkeit, diese Daten in Verbindung zu setzen und sie skalierbar zu machen. Hier lässt sich Donna Haraway anführen, die in ihrem Ansatz von Macht auf genau dieses neue Paradigma des 21. Jahrhunderts herausarbeitet (vgl. Haraway 1995a, S.49). Dabei sind vorrangig auch die Ausbreitung der Hardware und Software von einem spezifischen universitären und militärischen Kontext in den 60-er und 70-er Jahren hin zu einer allgemeinen Anwendung auch in staatlichen Behörden heutzutage

materielle Voraussetzungen(vgl. boyd, Crawford, 2012, S. 664).

Das macht Genderclassifier, um abschließend auf die spezifische Politik hinter Gender als kategorialen Standard zu blicken, zu einer staatlich algorithmischen Technologie und nicht zu einer staatlichen UND algorithmischen Technik (vgl. Bives, Haimson 2016, Johnson 2018). Gender setzt dabei als quantitative Standardisierung eine Vielzahl komplexer (Bild)Daten miteinander In Verbindung, in dem Gender diese nach Mann/ Frau einordenbar macht. Dieser Standard ist folglich einerseits systemübergreifend, wie im Falle des Gendershades Projektes. So nutzten Face++; IBM und Microsoft alle denselben Genderstandard. Gleichzeitig reiht sich diese Gender Technik in eine staatliche standardisierte vergeschlechtlichte Politik ein, wie das Beispiel der Genderbezeichnungen der Parlamentarier:innen auf den von Buolamwini verwendeten Bildern zeigt. Diese Genderpolitik baut auf einer störungsfreien, hegemoniale Zweigeschlechtlichkeit auf (vgl. Ludwig, 2016) Sei es in Form von vergeschlechtlichten Grenzkontrollen, reguliertem Zugang zu Gendereintragungen in Dokumenten oder aber Zugang zu medizinischer Versorgung (Salomon 2010, S.172, Hicks 2019). Genderstandards sind somit elementarer Bestandteil der Vermittlung von Technologie und Staatlichem Handeln und sind konstitutiv für diese Technologien,

Wenn Bives und Haimson von algorithmischer einerseits Standardisiert und staatlicher Standardisierung andererseits sprechen, möchte ich dem widersprechen. Gender als Kategorie ist zu einem technologischen Standard geworden, der im Gendershades Projekt für das Funktionieren, also das Einordnen und Verstehen grundlegend ist, sowohl was technische Kompatibilität als auch die Wissenschaftlichkeit der beiden analysierten Papers anbelangt. Ferner ist aufgrund der engen Verzahnung des Nutzens biometrischer Technologien durch staatliche Behörden ein Auseinanderhalten der algorithmischen Strukturen und staatlichen Wirkweisen nicht zielführend, da sich in Bezug auf Genderclassifier erstens die Verwendung von staatlichen Akteur:innen mit Software überlagern, ganz einfach deswegen, weil der Staat solche Technologien nutzt. Zweitens sind Genderclassifier eben nicht auf ein qualitatives Verständnis von Gender ausgerichtet, sondern auf die quantitative Zuordnung von vorgefertigten Kategorien, die sich weder verändern noch Lücken zulassen (vgl. Bives, Haimson 2016). So bleibt die Frage, wie mit dieser Funktionsweise von Gender umzugehen ist, denn sie wird in Form der Genderclassifier angewendet und die wissenschaftlichen Prognosen gehen sogar von einer Erweiterung des Einsatzes aus.

Die Frage des Verbotes, die Os Keyes bezüglich der Transexklusion solcher Anwendungen in den Raum stellt, erscheint mir in diesem Zusammenhang wenig uieführend. Denn das Gendershades Projekt bzw. Buolamwini und Gebru arbeiten aktiv daran, Schwarze Menschen und Menschen of Color dieCis sind vor den Gefahren einer Nichterkennung durch Genderclassifier zu bewahren.

Eine komplette Ablehnung der kategorialen Technologie von Gender kommt einer Ablehnung dieser Systeme gleich, die vielleicht normativ wünschenswert wären (was an andere Stelle ausführlich diskutiert werden müsste) aber in der jetzigen Situation gefährlich für Schwarze- vis- Menschen und cis Menschen of Color wäre, da die Verbotsposition keine Möglichkeit der politischen Einmischung innerhalb der Genderclassifier bietet und auch keine weiteren Verbesserungen, im Sinne einer Inklusion von Gender, in das PPB Benchmark.

Ich möchte an dieser Stelle eine konkrete Maßnahme vorstellen, die sich mit der Subversion von Genderstandards auseinandersetzt. Doch was kann Subversion innerhalb kategorialer Standards meinen? In Lucy Suchmann Ausführungen zu Kategorien und Computersoftware formuliert die Wissenschaftler:in eine allgemeine Form der Emanzipation, nämlich die Verweigerung, welche die Allgegenwärtigkeit der Kategorien für Gesellschaften einbezieht.

„Non-compliance with the use of a particular category scheme, particularly one imposed from outside, or the adoption of an alternative are in this sense acts of resistance.“ (Suchman 1993, S.182).

Hier wird also gerade der Moment der externen Kategorisierung bzw. der Fremdbezeichnung angesprochen, der schon für Os Keyes ein entscheidender Faktor war, Genderclassifier abzulehnen. Jedoch verweist Suchmann hier auf eine Unausweichlichkeit. Mensch wird in ein System eingeordnet, ohne es zu wollen und muss innerhalb dieses Systems operieren. Wie jedoch kann eine solche „Non-compliance“ - Haltung aussehen?

Die Algorithmic Justice League, die von Buolamwini gegründet wurde, hat hierzu dieses Jahr einen Workshop mit dem Titel „Drag vs. AI“ veranstaltet(vgl. <https://www.ajl.org/drag-vs-ai#signup> zuletzt abgerufen 20.11.2020). Hierbei wird Drag, also die Aneignung von Gendermarkern in spielerischer Form, als Moment der „Verwirrung“ von Gesichtserkennungsoftware und somit auch von Genderclassifyern genutzt. Indem Personen Genderperformances aneignen, die sie selbst in einem System das auf Gender beruht wählen , stimmen sie aktiv gegen eine eindeutige Identifikation. Dies geschieht mittels genderqueeren Methoden (vgl. Mertlitsch 2016, S.162).

Auch Butler benennt Drag als explizite Methode der Subversion des rigiden Genderstandards. „In dem die Travestie die Geschlechtsidentität imitiert, offenbart sie implizit die Imitationstruktur der Geschlechtsidentität als solcher – wie auch die Kontingenz.“ (Butler 1991, S. 202). Dabei, so Butler, verweist die Drag Performance immer auf die Künstlichkeit von Gender und nicht etwa auf ein originellen/ natürlichen Zustand von Gender (vgl. ebd. 203) In Bezug auf Drag vs. AI wird hierbei Gender auch zur künstlich digitalen Kategorie, die als Empirie eine subversive Drag Queen heranzieht. Drag wird zu einer Art utopischer Korrektur des Genderstandards im Genderclassifier, in dem die Funktion beibehalten wird. Subjekte werden kategorisiert, jedoch ändern sich die Muster,

nach denen diese Daten ausgewertet werden (vgl. Mertlitsch 2016, S.167).

Das Gesicht des Drag Kings wird zu Referenz einer Männlichkeit, die weder außerhalb der Performance existiert, noch ein authentisches Zukunftsversprechen bezeugen soll, sondern den Status Quo des cis geschlechtlichen Standards als künstliche Praxis offenbart. Dabei stellt die Workshop Leitung Materialien für Communitys zur Verfügung, um diese Praxen anzuwenden.

Drag vs. AI wird somit in der theoretischen Praxis zu einer subversiven BigDataTechnologie (vgl. <https://www.ajl.org/drag-vs-ai#signup> zuletzt abgerufen am 30.11.2020). Nicht mehr auf einen Ort oder einen Workshop begrenzt, werden diese Praxen in verschiedenen Gemeinschaften getragen. Dabei orientierten sie sich gleichzeitig an bestimmten Vorlagen.

Ich möchte in der nächsten Sektion auf die Art der Informationsübermittlung eingehen und Gender als agentielles Protokoll fassen, was wiederum andere Aspekte technologischer Wirkweisen offenbart und auch andere politische Praxen involviert, bzw. eine neue Perspektive auf beispielsweise den Workshop eröffnet.

4.) Gender als agentielles Protokoll

Im Folgenden möchte ich eine zweite Interpretation der Funktionsweise von Gender innerhalb der technischen Grundlagen anbieten, die sich von dem Konzept „Gender als kategorialer Standard“ unterscheidet.

Dabei berufe ich mich auf zwei unterschiedliche Konzepte, die ich zusammenfügen werde: Einerseits beziehe ich mich auf den agentiellen Realismus von Karen Barad und hierbei voranging auf ihr Verständnis des Apparats als Praxis der Zusammenführung von Naturwissenschaft und politischer Ökonomie. Andererseits verwende ich das Konzepts des Protokollfeminismus von Michelle Murphy und auf die, von Helen Hesters davon abgeleitete Idee des Mesopolitischen als politikwissenschaftliche Interpretationsgrundlage zur Beantwortung der zweiten Forschungsfrage (vgl. Murphy 2012, Hester 2018, Barad 2007). Gender als agentielles Protokoll ist dabei für mich aus zwei Gründen eine wichtige Analyse Perspektive.

Der Protokoll-Begriff ermöglicht mir eine theoriegeleitete These für die Analyse der technischen wirkweisen des Gendershades Projektes zu verwenden, nämlich das Gender im Gendershades Projekt als Protokoll funktioniert. Diese These legt den Fokus in Hinblick auf technische Praxen, die in diesem Teil der Arbeit von mir untersucht werden auf das „Wie“ meiner Forschungsfrage,,. Zweitens kann ich Gender in Bezug zu Halberstamms Ansatz des computergleichen Funktionierens von Geschlecht untersuchen. Die Konzeption des Protokolls verweist hierbei auf die Untersuchung empirische Beispiele, wie sie das Gendershades Projekt darstellt. Die Protokoll-Idee nimmt dabei Bezug auf konkrete Handlungsweisen und verwendete Techniken.

Bevor ich mich jedoch dieser theoretischen Übertragung widme, möchte ich wie zuvor eine Definition von Protokollfeminismus in Zusammenhang mit dem Konzept des Apparates von Barads anbieten und diese theoretischen Überbelegungen mit der Vorgehensweise Buolamwini und Gebru im Gendershades Projekt verbinden, um dann politische Implikationen dieser Theorie aufzuzeigen. Weiter möchte ich deutlich machen, wie „Gender als Kategorialer Standard“ eine Erweiterung hinsichtlich nicht binärer Gender darstellt.

a.) Definition

Ich werde zuerst den Protokoll Feminismus definieren und in einem zweiten Schritt die Idee des Apparates näher definieren. Dafür möchte ich diesen jedoch zuerst theoretisch verorten, bevor ich mich der Definition nähere.

i.) Theoretische Herleitung

Protokoll Feminismus als Theorie bestimmt nach dem Medienwissenschaftler Alexander Galloway Das Verhältnis von Frauen, Maschinen und Technologie, wobei letztere das Protokoll sei. Dabei vereint das Protokoll hier vor Allem auf seine Funktion im Bereich des Internets oder des Computerprotokolls (vgl. Galloway 2004, S.188). Galloway beschreibt dabei den Protokoll Feminismus als Cyberfeminismus²⁴ (vgl. ebd. 186).

Dieser wiederum ist ein theoretisches Konzept, entstanden in den frühen 90er Jahren als Versuch, die neuen digitalen Technologien zur Überwindung des Patriarchats einzusetzen (vgl. Wilding, Critical Art Ensemble 1998, S.47). Die Medienwissenschaftlerin Susanna Paason beschreibt in ihrem Versuch einer Theoriegenese des Cyberfeminismus diesen als Schnittstelle feministischer Theoriebildung, Kunst und digitalem Aktivismus (vgl. Paason 2011, S.366). Dabei lasse er sich in drei Strömungen unterteilen:

In the first instance, cyberfeminism stands for feminist analyses of human-machine relations, embodiment, gender and agency in a culture saturated with technology [...]

A second possible definition of cyberfeminism points to critical analyses of cyberculture in relation to feminist thought. Here, cyberfeminism becomes a critical feminist position for interrogating and intervening in specific technological forms and practices [...]

In a third possible definition, “cyberfeminism” stands for analyses of the gendered user cultures of information and communication technologies and digital media, their emancipatory uses, as well as the social hierarchies and divisions involved in their production and ubiquitous presence“ (ebd. S.340).

24 Cyberfeminismus ist dabei keine Definition einer einheitlichen Theorie sondern vielmehr Ausdruck einer Theoriestörung mit verschiedensten disziplinären Hintergründen, aktivistischen Zielen und Methoden (vgl. Sollfrank 2018, S.12-13)

Der dritte Strang, so die Autor:in sei dabei gerade in einem US-amerikanischen Diskurs über die kritische Auseinandersetzung neuer Medien und sozialer Netzwerke dominant (vgl. ebd.). So sei der Cyberfeminismus in diesem Kontext vorrangig ein Zweig universitärer feministischer Theoriebildung, wo hingegen im westeuropäischen Raum nach wie vor eher ein künstlerisches Ideal des Cyberfeminismus vorherrsche. Dieses sei vor Allem durch die britische Aktivist:innen Gruppe Subrosa und das deutsche Künstler:innen Kollektiv OldBoyNetwork geprägt worden (vgl. ebd. S.341). Dabei ist diesem frühen Strang des Cyberfeminismus eine Prämisse gemein, die der Zweigeschlechtlichkeit und vor Allem der besonderen Stellung der Frauen innerhalb dieser neuen Techniken. Eine der Begründer:innen des Cyberfeminismus, die Medientheoretikerin Sandy Plant beispielsweise geht von besonderen Fähigkeiten der Frauen hinsichtlich der Benutzung digitaler Technologien wie dem Internet, aus (vgl. Plant 2000, 35-36). Frauen, so Plants Argument, seien aufgrund der anhaltenden Deprivilegierung immer schon gezwungen, schnell neue Fertigkeiten zu lernen und sich anzupassen, ferner haben sie sich immer schon in Techniken des Vernetzens und Verbindens geübt, wie dem Weben als Handwerk oder dem Dasein als Telefonistin oder menschlicher Computer seit den 50er Jahren. Diese Fähigkeiten würden die Frauen im Angesicht der neuen digitalen Transformation sowohl im Globalen Norden als auch im Globalen Süden zu einer Möglichkeit der Beendigung des Patriarchats führen (vgl. ebd., S. 51- 59). Dies bezieht wiederum Alexander Galloway dabei maßgeblich auf die Funktion des Protokolls., „The universality of protocol can give feminism something that it never had at its disposal, the obliteration of the masculine from beginning to end.“ (Galloway 2004, S. 188). Dieser naiv optimistische Gedanke der 90er Jahre und des frühen 21. Jahrhunderts im cyberfeministischen Diskurs, wie in Galloway hier formuliert, wurde nach Ausbleiben dieser Versprechungen und der starken Fokussierung auf die Frau als wenig differenziertes politisches Subjekt stark kritisiert aus der feministischen aktivistischen und akademischen Gemeinschaft. Der Cyberfeminismus geriet dabei als theoretische und politische Strömung innerhalb der feministischen Debatten um neue Informationstechniken ins Abseits (vgl. Sollfrank 2018, Avanessian, Hester 2015, S. 10-11). So kommentiert Judy Wajcman in ihrem theoretischen Konzept des Technofeminismus im Jahre 2006 bezüglich des Cyberfeminismus lakonisch, „Cyberfeminist have coffee in cyber-cafes, surf the internet, and imagine a gender-free future in cyberspace.“ (Wajcman 2004, S.3). Das Protokoll als Gegenstand der feministischen Auseinandersetzung mit Technologie/ Gender und insbesondere neuer Technologie wie KI ist jedoch weiterhin Bestandteil einer wissenschaftlich/aktivistischen Auseinandersetzung und theoretisch verankert in den Debatten um diese Themenkomplexe. So arbeitet die Literaturwissenschaftlerin Helen Hester mit dem Begriff Protokoll Feminismus als politische Theorie einer kritischen emanzipatorischen Technikgestaltung (vgl. Hester 2018,

S.84-85). Hester übergeordnetes Konzept ist der Xenofeminismus, dieser beruht anders als der frühe Cyberfeministische Ansatz auf einem genderabolitionistischen Ansatz. Dieser ist durch die Ablehnung der Vorstellung von einer „Natürlichkeit“ der Geschlechter gekennzeichnet, also der Prämisse, es gebe eine vorgesellschaftliche „wahre“ Geschlechtsidentität (vgl. ebd. S. 22-23). Hester beruft sich bei ihrer Vorstellung des Protokollfeminismus auf Michelle Murphys Arbeiten, die diesen Begriff zuerst geprägt hat. Murphy verweist dabei stärker noch als Hester auf die Verschränkung von Race und Gender in Bezug auf Technik und deren Gestaltung.

Ich werde mich im folgenden auf diesen Strang des Protokollfeminismus beziehen, da ich selbst eine queere und intersektionale theoretische Perspektive nutzte, um das Gendershades Projekt zu analysieren.

Ich möchte jedoch kurz erwähnen, dass auch der Cyberfeminismus sich weiterentwickelt hat als theoretische Strömung und genderqueere Position aufgreift wie etwa Glitch Feminisms von Russel Legacy.. Die*der Künstler:in beschäftigt sich mit Fragen der Auflösung von Gender in Bezug auf digitale Technologien (vgl. Russell, 2020).

ii.) Protokollfeminismus (Begriffsbestimmung)

Der Ansatz des Protokollfeminismus, denn ich hier also verwende, beruht nach Murphy auf die in den 70er Jahren in den USA verhaftete Women Health Bewegung. Dies ist eine politisch/aktivistische Strömung, die mittels medizinischen Wissens über Abtreibung, Periode und Fragen zu Themen rund um „Frauengesundheit“ feministische Praxen vermitteln wollte. Dies sollte nach den Aktivist:innen neue Zugänge zur Gesundheitsversorgung in den USA schaffen. Diese sollten es marginalisierten Gruppen von cis Frauen, besonders Schwarzen Frauen (Transfrauen wurden in Murphys Untersuchung nicht genannt) ermöglichen, einerseits medizinische Versorgung jenseits von Stigmatisierung zu erhalten, gleichzeitig diese Gruppen jedoch auch ermutigen, das gesammelte Wissen weiterzugeben (vgl. Murphy 2012, S.27-28)

Murphy verbindet dabei in ihrer Idee des Protokollfeminismus Fragen intersektional verbundener Unterdrückungen und gleichzeitig taktische Aktionen wie den Aufbau von Datenbanken, Vernetzung von einzelnen Wissensständen durch Manuals und Workshops zu einer computerisierten Funktion, der „Interwingularity as an attempt to capture the intermingled aspect and “cross- connections” of knowledge“ (ebd. 39).

Computertechnologie und die Frage der politischen Handlung stehen dabei in einer Spannung zwischen Unterdrückung und Emanzipation, jedoch leitet Murphy in ihrem Ansatz, anders als beispielsweise die Soziologin Judy Wajcman in ihrem kanonischen Werk *Techno Feminsim*, konkrete

Handlungsmöglichkeiten ab, die hier auf das Gendershades Projekt angewendet werden können (vgl. Wajcman 2004).

Der Protokollfeminismus beruht wenig überraschend auf der Idee des Protokolls. Für Murphy ist das Protokoll eine Zusammenführung verschiedener Praktiken:

„[A] procedural script that strategically assembles technologies, exchange, epistemologies, subjects, and so on. Put simply, a protocol establishes “how to” do something, how to compose the technologies, subjects, exchanges, affects, processes, “ (Murphy 2012, S.25)

Das Protokoll kann hier nach dieser allgemeinen Beschreibung auf verschiedenen Arten der Zusammenführung hinweisen, ist jedoch nicht paradigmatisch für beispielsweise computerbasierte Informationstechnologie, wie Genderclassifier, sondern könnte auch auf Architektur, landwirtschaftliche Bearbeitung von Feldern und ähnliches bezogen werden. Murphy bezieht diese Zusammenführung beispielsweise, vorrangig aufgrund ihrer eigenen Forschung, auf die feministischen „Self Help“ Bewegung der 70er Jahre in den USA (vgl. ebd.) Deshalb möchte ich hier Helen Hester folgen, die den Ansatz von Murphy für ihr Konzept den Xenofeminismus aufgreift und das Protokoll als die paradigmatische politische Struktur des 21. Jahrhunderts versteht.

„I would argue that the protocol as a social or political technology has greater resonance as a social or political resonance with current discussion than does the protocol in its merely technical sense.“(Hester 2018, S.108).

Hester beruft sich bei dieser These auf die Analyse der Medienwissenschaftler:innen Alexander Galloway und Eugene Thacker. Diese beziehen das Protokoll auf Netzwerke. Netzwerke, so die Wissenschaftler:innen, seien eine paradigmatische soziale Struktur des 21. Jahrhunderts. Als Beispiele dafür führen sie Phänomene wie „Infektionsnetzwerke“, „politische Schwarmbildung“ und „Onlinerollenspiele“ an (vgl. Galloway, Thacker 2014, S. 289). Dabei entpolitisierten die Debattenbeiträge um dieses Phänomen, das Netzwerk in dem sie dieses als technische Bedingtheit setzten, unabhängig von sozialen Prozessen und nicht veränderbar durch politische Interventionen. Eine wichtige theoretische Notwendigkeit sei daher Untersuchung „der mikrotechnischen Ebene nichthumaner Maschinenpraktiken“ (vgl. ebd. S. 290). Diese Untersuchung benötige dafür aber weniger eine politische Theorie althergebrachter Konzepte, wie Demokratie und Staat, sondern eine Idee von Kontrolle, die sich auf diese neue Ebene anwenden lasse. Dies sei das Protokoll, so Galloway und Thacker. Dieses stelle die Regeln für das Netzwerk bereit, beispielsweise ist das http Protokoll eine Form der Kontrolle im Internet. Dabei hat das Protokoll „eine Doppelfunktion. Es ist sowohl ein Apparat, der Netzwerke ermöglicht, wie eine Logik, die Aufgaben und Abläufe in diesem Apparat regelt.“ (vgl. ebd. S.291-292). Thacker und Galloway deuten hier schon eine Verbindung des Protokolls mit dem Konzept des Apparates an, führen diese Verbindung jedoch nicht weiter aus. Ich

möchte in einem zweiten Teil dieser definitorischen Näherung an das Konzept des Protokolls mit Hilfe von Barad nachgehen.

Dennoch möchte ich mich im Folgendem stärker auf Hester und Murphy als Referenz beziehen. Anders nämlich als Galloway und Thacker beziehen beide Wissenschaftler:innen nicht nur maschinelle Aspekte des Protokoll in ihre Überlegungen zur Funktion des Protokolls ein. Des Weiteren ist Gender als Protokoll bei Hester und Murphy von entscheidender Bedeutung.

Dabei zeichnet sich diese Definition des Protokolls, so Hester, durch drei Komponenten aus. Protokolle sind, nach Hester, nicht nur „Gebrauchsanleitungen“. Sie sind (a) relational, d.h. sie beschreiben nicht nur Verbindungen verschiedener Werkzeuge, Subjekte Artefakte und Technologien. Sie entstehen viel mehr durch andere Netzwerke. Erst das Zusammenführen der Netzwerke lässt ein Protokoll entstehen (vgl. Hester 2018, S.109).

Ferner sind Protokolle (b) adaptiv. Sie müssen Verbindungen zwischen verschiedensten Akteur:innen schaffen, so muss bezogen auf das Gendershades Projekt die Operationalisierung von Gender zwischen verschiedenen Softwares, Staaten (aus denen die Bilder der Parlamentarier:innen stammen wie beispielsweise Island und Uganda (vgl. Buolamwini 2017, S.51) und Einzelpersonen eine Vorgehensweise schaffen, welche alle diese Akteur:innen in Beziehung setzt (vgl. Hester 2018, S.109).

Und drittens muss das Protokoll (c) „organizational“ (ebd.) sein. Das bedeutet, sie sind Ausdruck von Macht bzw. wie ich erläutern werde von Kontrolle innerhalb von Netzwerken. Protokolle sind „emergent property of organization and control in networks that are radically horizontal and distributed“ (ebd.). Hierbei bezieht sich Hester auf die von Galloway und Thacker genutzte Definition der Kontrollgesellschaft des Philosophen Gilles Deleuze. Diese, so Galloway und Thacker, sei durch Computernetzwerke gekennzeichnet, anstatt durch staatliche und wirtschaftliche Akteur:innen wie Behörden. Sie zeichne sich durch horizontale Machtgefälle in Form von beispielsweise organisationalen Hierarchien aus. Kontrolle, so Deleuze, denn die drei Autor:innen hier zugrundelegen, um Protokoll zu definieren „ist kurzfristig und auf schnellen Umsatz gerichtet, aber auch kontinuierlich und unbegrenzt“ (Deleuze 2016 [1993], S.260).

Protokolle sind hierbei die Ordnungsmechanismen bzw. Formen der Kontrolle, die sich weder ausschließlich auf Institutionen noch auf menschliche Subjekte als Ausgangspunkt von Kontrolle beziehen (vgl. Galloway, Thacker 2014, S.292). In diesem Sinne subjektiviert nicht die institutionelle Macht, wie sie beispielsweise Foucault und Butler beschreiben, die von empirischen Einheiten wie der Schule oder Fabrik ausgeht, ihre unterworfenen Individuen (vgl. Butler 2011 S.72-73). Nach Galloway und Thacker organisiert zum jetzigen Zeitpunkt das Protokoll das Netzwerk und moduliert die Verbindungen und Handlungen der Akteur:innen. Kontrolle ist dabei die Vorausset-

zung der Modulation (vgl. Galloway, Thacker 2014, S. 293).

Modulation als Organisationsform des Protokoll beschreibt dabei ausgehend von dem von Hester, Galloway und Thacker zugrunde gelegten Konzept eine Form der Individualisierung, die nie abgeschlossen ist: „die permanente *Weiterbildung* [löst] tendenziell die *Schule* ab, und die kontinuierliche Kontrolle das Examen.“ (Deleuze 1993 [1993], S.257). Dabei stellt das Protokoll eine inhärente Eigenschaft dieser neuen Kontrollgesellschaft dar, in dem es selbst nicht mehr der einzelnen Institution zugeschrieben werden kann, da diese nicht mehr ausschlaggebend sind für die Konstruktion von Gesellschaft. In diesem Sinne ist das Protokoll organisierend und nicht subjektivierend (vgl. Galloway 2004, S. 121).

Diese drei Merkmale sind nach Hester, die sich auf Murphy bezieht, Erweiterungen des Protokollbegriffs auf Informationstechnologien und stehen nicht in Konkurrenz zu den Überlegungen des Protokollfeminismus.

So fragt Hester insbesondere in Bezug zu emanzipatorischer Geschlechter Politik (Gender politics) (ebd. S.77) was genau eigentlich kontrolliert wird in solchen Netzwerksituationen und verweist dabei wieder auf Murphys Protokollfeminismus als eine mögliche Definition.

Dabei bedient sich der Protokollfeminismus nach Murphy bewusst Technologien als politischer Akte und kommt somit einer kritischen Technologie gleich, welche ich im ersten Abschnitt dieses Kapitels erläutert habe.

Protokollfeminismus nimmt dabei die Politik von Technologien sowohl zum Ausgangspunkt der wissenschaftlichen Analyse, als auch Form des praktischen Handelns.

„[P]rotocol feminism—a form of feminism concerned with the recrafting and distribution of technosocial practices by which the care and study of sexed living-being, assembling protocol Feminism could be conducted. In other words, the politicization of life in the formulation of “women’s health” was simultaneously bound to the politicization of the details of the techniques by which it was known and materially altered“ (Murphy 2012, S.28-29)

Das Aneignen bereits vorhandener Werkzeuge ist dabei eine wichtige Unterscheidung zwischen einem feministischen und einem, wie Murphy es beschreibt, rein technischen Protokoll, wie Murphy in dem eben angeführten Zitat verdeutlicht. Dabei ist für Murphy, das Verteilen des feministischen Protokolls und somit des Wissens, wie etwas gemacht wird, und das gleichzeitig bewusste Verändern dieses je nach Kontext der Grundgedanke dieser Art von feministischen Technologien (vgl. ebd. S.28). Dabei verweist sie auf die Standardisierung als Voraussetzung des feministischen Protokolls. SdDenn dieses muss sich an Umstände anpassen, aber immer noch Wissen vermitteln. So sind die Beispiele, die Murphy hierbei einbringt, wie das Untersuchen der Vulva durch einen kleinen Spiegel (das Spekulum). einerseits durch individuelle Erkenntnisinteressen, Körperbeschaffenheit

und Umgebung geprägt und müssen auf diese Faktoren angepasst werden, andererseits müssen diese Anpassungen in dem Fall auch der Untersuchung der Vulva dienen. Eine Untersuchung der Vulva ohne Spiegel beispielsweise würde den Gedanken der Untersuchung selbst zuwider laufen (vgl. ebd. 29). Nach Hester kann diese Ambivalenz als eine Spannung zwischen der Adaptivität und der Funktion der Organisation spezifiziert werden (vgl. Hester 2018, S.110). Das Protokoll muss Verbindlichkeiten für die Akteur:innen schaffen, mit denen gearbeitet werden kann und gleichzeitig muss es sich je nach Kontext ändern. Diese Spannung richtet sich an den jeweiligen relationalen Voraussetzungen aus. Beispielsweise müsste ein Kollektiv, welches nur aus trans Frauen besteht und sich Methoden der Womens Health Bewegung anschließen wollen würde noch einmal andere Wissensbestände aneignen. beispielsweise was die Pflege von Neovulven anbelangt. Genau dies wird auch getan im 21. Jahrhundert (vgl. ebd. S..84-5).

Der Protokollfeminismus in dieser konstitutiven Spannung zwischen Adaption und Organisation, nimmt hierbei jedoch nicht nur ein präexistentes Subjekt und Werkzeuge als Faktoren auf, , sondern auch Carearbeit²⁵, Gefühle und Machtrelationen, die sich entlang von Gender, Class und Race verstärken (vgl. Murphy 2012, S.32). Dabei sind Protokolle die Möglichkeit, solche Verbindungen zu schaffen (vgl. ebd. 30). Und gleichzeitig sind diese Verbindungen die Möglichkeiten, in denen Subjekte entstehen. Die gesunde Frau entsteht durch das Erkennen der eigenen Vulva innerhalb eines „Self help“ Workshops zur Befreiung von patriarchalen Ideen über weibliche Körper (vgl. ebd.). Aber, so schreibt Murphy innerhalb dieser spezifischen Untersuchung in den 70iger Jahren, entstehen auch Hierarchien. So entsteht ein weißer Frauenkörper der durch die Exklusion Schwarzer Frauen und ihrer Bedürfnisse konstruiert wird. Die Frage der Selbsterkundung und DIY Abtreibung, die durch weiße Frauengruppen angeboten wurde, vernachlässigten das koloniale Erbe von Zwangsterilisation Schwarzer Frauen und Frauen of Color, sowie Experimente zur Empfängnisverhütung, die ohne Einwilligung und Information von Frauen of Color in den 60ern Jahren in Puerto Rico durchgeführt wurden (vgl. ebd.).

Murphy verweist in diesem Zusammenhang auf die Frage des Subjektes bzw. der Subjektwerdung innerhalb der theoretischen Perspektive des Protokollfeminismus hin. Das Subjekt selbst wird hier zu einem Widerspruch zwischen den einzelnen Machtstrukturen patriarchaler, rassistischer und kapitalistischer Art. Die Idee Identität ist keine Form des einen revolutionären Subjektes, sondern, „but rather draws out the contradictions formed at the axis of race, class, sexuality, and gender, as

25 Arbeit die auf das Handeln für und mit anderen ausgelegt ist:

„Diese Handlungen basieren explizit oder implizit auf ethisch-normativen Grundlagen, die die jeweilige Konstruktion von Care leiten. Im Gegensatz zu Caritas ist der Anspruch von Care weniger selbstlose Sorge als ein Ausdruck zwischenmenschlicher Interdependenz, der eine Achtsamkeit für sich selbst und den jeweils anderen voraussetzt und daher einzubeziehen ist.“(Brückner 2011, S.78)

well as logics of capitalism and state violence, that require disidentifications from any singular identity and thus the recognition of contradictory difference.“ (ebd. S.39)

Als weiße cis Frau in den 70ern Jahre erlebe ich beispielsweise Unterdrückung aufgrund meines Geschlechtes und kann mittels des Self-Help-Angebots feministischer Gruppen einen Weg finden, patriarchalen Strukturen zu begegnen, aber ich bin immer auch Teil einer rassistischen Struktur, die zu einer Negation von Bedürfnissen andere cis Frauen führt und von ihrherrührt. Ferner sind diese Zusammenhänge niemals ein individuelles Handeln bzw. Sein, sondern ein Verweis auf Beziehungen zu Andern. Protokolle verweisen so, um noch einmal Hester zu bemühen, auch auf Relationen, die ein einzelnes Subjekt als theoretischen Paradoxon erscheinen lassen, ohne aber die Ideen von Gender/ Race/ Class als strukturellen Kategorien aufzugeben (vgl. Hester 2018, S.29-32)

Protokollfeminismus ist zusammenfassend eine Art der Theorie, die sich auf Adaption, Rekontextualisierung und Politisierung soziotechnischer Zusammenhänge beruft. Dabei sind diese Zusammenhänge erst durch ein feministisches Protokoll möglich, wie das Beispiel der „self help“ Workshops gezeigt hat. Das feministische Protokoll zeichnet sich dabei durch eine intendierte Weiterentwicklung durch die Nutzenden aus und ist nicht nur eine Art Gebrauchsanleitung. Protokollfeminismus zeigt Verteilung von Macht durch Übertragbarkeit von Wissen durch das konkrete Verbinden von Praktiken, Werkzeugen und gesellschaftlichen Kontexten auf.

iii.) Der Apparat (Begriffsbestimmung)

In einem letzten definitorischen Schritt möchte ich diese Theorie um die Idee des Apparates von Barad erweiterten und auf das Konzept des dahinterliegenden agentiiellen Realismus eingehen. Jedoch begrenze ich dieses auf die politischen Implikationen des Apparates und seiner Funktionen. Denn das Konzept des Apparats stellt eine wichtige theoretische Verbindung zwischen der Funktion von Technik, wie ich sie im ersten Teil beschrieben habe und der besonderen disziplinärer Verbindung von Informationstechnologie, die ich untersuche und den Sozialwissenschaften dar, auf die ich mich als theoretischen Referenzrahmen beziehe.

Für Barad ist der Apparat dabei ein Mechanismus, welcher Verbindungen erzeugt und weniger ein Medium, welches Konzepte eines Bereiches (wie den Naturwissenschaften) auf ein anderen Bereich (wie die Sozialwissenschaften) überträgt. Diese Verbindungen sind für Barad eine politischen Theorie des Apparates. Diese sei eine Verbindung der Naturwissenschaften mit politischer Ökonomie, welche beispielsweise technologischen Wandel während des 20. Jahrhunderts erklärt. Dieser hätte weniger mit einer Vermittlung politischer und technischer Systeme zu tun als mit „the material

entanglement of political and scientific practices (Barad 2007, S: 231). Gender ist nach Barad einer dieser Apparate (vgl. ebd. S.242). Dabei wirken Apparate nicht wie ein Detektor für Unterschiede, wie etwa eine politikwissenschaftliche Vergleichsmatrix innerhalb der Subdisziplin Vergleich politischer Systeme, sondern als analytische Werkzeug, welches diese Unterschiede mitproduziert und aktualisiert (vgl. ebd.). Gender als ein solcher Apparat würde dann nicht real existierende Männer und Frauen „entdecken“, sondern durch spezifische Praxen, wie etwa dem Recht auf Ehe, zusammen mit anderen Apparaten wie Sexualität und Nationalität gestalten. Gender kann mit der Apparat-Idee wiederum als Rückgriff auf Gender als Technologie auf die Funktion von Gender hinweisen (vgl. ebd.).

Dabei sind Apparate keinesfalls deterministische Diskurspraxen, sondern ermöglichen eine Form der Kausalität die Barad „causal intra-actions“ nennt, wobei diese als Kausalstrukturen auftreten, welche einige Aspekte der Intraaktion als Grund und manche Punkte als Effekt markieren (vgl. ebd. S.149). Am Beispiel eines binären Genderapparates innerhalb des Gendershades Projektes können die Kategorien Mann/Frau als Grund unterschiedlicher Gesichter verstanden werden. Wobei der Effekt des vergeschlechtlichten Bildes von der Verwendung der Genderclassifier als Gesichtsanalyse-Software von verschiedenen Akteur:innen wie Buolamwini und Gebru sowie der von ihnen zitierten Wissenschaftlerinnen im Aufsatz und ebenso den bereitstellenden Firmen dieser Technologie geformt wird (vgl. Buolamwini, Gebru 2018)

Dabei sind Strukturen wie Genderapparate ebenfalls sowohl innerhalb eines Phänomens wie dem Gendershades Projekts spezifisch realisiert und immer mit Bezug auf andere Phänomene wie patriarchale Genderregulierung westlicher Nationalstaaten (vgl. Barad 2007, S.242)

Der Apparat als Interaktion von politischer Ökonomie und Naturwissenschaft schafft dabei eine Verbindung, die wie Barad es nennt, Phänomene sichtbar werden lässt, weil sie durch Ausschluss andere Phänomene begrenzend wirken. Diese Funktion sieht Barad auch in Kategorisierungsprozessen. Hierbei habe ich die Funktion von Exklusion als Mittel der Erstellung von Kategorien schon beispielhaft aufgeführt. Nach Barad jedoch ist ein Apparat nicht auf kategoriale Prozesse beschränkt. Apparate können durch Genzziehungen auch wissenschaftliche Erkenntnisse produzieren wie Interpretationen innerhalb der Quantenphysik (vgl. ebd. S.146).

Was Barad Apparat-Idee dabei in den Vordergrund rückt, ist die Idee, die Murphy als „Interwingularity“ (Murphy 2012, S.39) bezeichnet, also die Möglichkeit, Werkzeuge (Apparate) zu schaffen (vgl. ebd.). Jedoch ist Barad an dieser Stelle konsequenter, in dem ihre Idee der Intraaktion nicht nur Tools und Subjekte einbezieht. Die Intraaktion schafft eine Offenheit, die sich in ständiger Performativität der letztgenannten Komponenten ausdrückt. Subjekt und Werkzeug wer-

den durch den Apparat in Verbindung gesetzt. Diese Verbindung ist jedoch immer ein Werden und nie abgeschlossen. Dies bietet eine zeitliche Komponente, die sich so nicht bei Murphy findet und Handlungen auch für nicht-menschliche Akteurinnen öffnet, da diese teil der performativen Intraaktion sind, die Subjekt und Objekt hervorbringen (vgl. Barad 2007, S.231). Diese Idee stellt in Bezug auf Künstliche Intelligenz eine wichtige theoretische Bereicherung dar. Handelt es sich bei Genderclassifyer doch um Software die autonom, ohne menschliche Kategorisierung der Bilder in gewissen Grenzen handeln soll. Barad beschreibt dabei Handlungsmöglichkeit als „changing possibilities of change entailed in reconfiguring materialdiscursive apparatuses of bodily production, including the boundary articulations and exclusions.“ (vgl. ebd. S.235).

Apparate fassen dabei nicht-menschliche und menschliche Akteur:innen mit ein bzw. erzeugen sie, ohne sich auf ein vorangegangenes Subjekt festzulegen und werden nicht zu einem Medium, welches „Sozial“ und „Technisch“ als feste Gegenpole behandelt, sondern soziotechnische Muster erzeugt, wie beispielsweise die von Buolamwini und Gebru verwendeten Kategorien female/male. Gender als spezifischer Apparat rekurriert dabei nicht auf vorgefasste Subjekte, sondern dient der Beschreibung von Fotos, der Zurordenbarkeit von Fotos durch Maschinen und der Nachvollziehbarkeit wissenschaftlichen Arbeitens durch ein Fachpublikum, welches das Gendershadesprojekt rezipieren soll, wie ich im vorangegangenen Teil erläutert habe. Ich möchte jedoch in der nachfolgenden Analyse der These von Gender als Protokoll im GendershadesProjekt nachgehen.

Dabei möchte ich das Konzept des agentielle Protokolls einführen, welches ich Barads theoretische Überlegungen zum Apparat entnehmen. Dabei sind Apparate wie Protokolle Orte der möglichen Verbindung. Apparate sind jedoch nicht nur auf das Zusammensetzen von Werkzeug und Subjekt begrenzt, sondern setzen diese beiden Akteur:innen als performativ voraus (in einem Prozessbegriff, der das Phänomen darstellt). Apparate ermöglichen somit eine theoretische Perspektive, die nicht-menschlichen Komponenten Handlungsmacht zuschreibt. Deshalb meint der Präfix agentiell in meinem theoretischen Rahmen die Erweiterung des Protokollbegriffes um nicht-menschliche Handlungsmacht und die Performativität von Subjekten. Agentiell ist hier eine Erweiterung des Protokollbegriffes.

Ich möchte in einem nächsten Schritt die Idee des feministischen Protokolls als spezifischen Apparat des Gendershades Projektes einführen und diskutieren. Darauf aufbauend möchte ich eine polit-theoretische Interpretation dieser Sichtweise erarbeiten.

Ich werde dafür das erarbeitete Konzept des agentiellen Protokolls auf das Gendershades Projekt beziehen.

b.) Empirie

Im Folgendem möchte ich den Prozess des Labelns analysieren, um die These des Genderrotokolls zu erörtern. Stand bei dem vorangegangenen Abschnitt zu Gender als kategorialer Standard die Unterteilung und in female/ male im Zentrum, möchte ich hier auf den dahinterliegenden Prozess eingehen.

i.) Das Label als Genderprotokoll

Buolamwini und Gebru erstellen im Gendershades Projekt wie beschrieben ein Datenset, welches sich aus einem zweistufigen Labelprozess ergibt, einmal das Klassifizieren von Bildern entlang verschiedener Hauttypen (mithilfe der Fitzpatrick Skala) und einmal das Labeln derselben Bilder entlang der Kategorie Mann/Frau.

Das Einordnen der Hautfarbe auf den jeweiligen Bildern, erfolgt dabei durch Buolamwini und Gebru mit einer einheitlichen Vorgehensweise.

„We chose the Fitzpatrick six-point labeling system to determine skin type labels given its scientific origins. Dermatologists use this scale as the gold standard for skin classification and determining risk for skin cancer (TB, 1988). The six-point Fitzpatrick classification system which labels skin as Type I to Type VI is skewed towards lighter skin and has three categories that can be applied to people perceived as White.“ (Buolamwini, Gebru, 2018, S.6)

Auch Gender wird hierbei nach einer bestimmten Prozedur bestimmt.

„Gender labels were determined based on the name of the parliamentarian, gendered title, prefixes such as Mr or Ms, and the appearance of the photo.“ (ebd.).

Diese Prozedur wird für eine bestimmte Anzahl von Bildern (n= 1270) (vgl ebd. S.5) durch die beiden Wissenschaftler:innen durchgeführt und dies in ein festes Datenset zusammengeführt, dem PPB (vgl. ebd.). Danach beziehen Buolamwini und Gebru die zu evaluierenden Genderclassifier ein. Diese sollen das Datenset klassifizieren, bzw. jedem Foto des PPB eine der zwei Geschlechterkategorien zuweisen. Die Performance der drei Anwendungen wird dann anhand der richtig identifizierten Genderkategorie der Bilder von Buolamwini und Gebru bewertet. Die „Erkennungsraten der Genderclassifier werden dabei wiederum von den Wissenschaftler:innen nach Hautfarbe unterteilt (vgl. ebd. S. 8).

ii.) Protokoll als „how to do benchmarks“

Entsprechen der Idee Murphys, ein Protokoll als Set von Praktiken zu beschreiben, verstehe ich somit den Prozess des Labelns von Buolamwini und Gebru und ihrer Evaluation der automatischen

Genderzuweisungen durch die Genderclassifier. Dazu möchte ich weiterführend erklären, welche Praxen hinter dem Genderstandard im Gendershades Projekt stehen, bzw. wie genau der Genderstandard in dieses Projekt eingelassen ist. Gender als Protokoll im Genderclassifier beinhaltet, wie oben beschrieben die Einordnung der Bilder nach den oben genannten Kategorien Hautfarbe und Gender (vgl. ebd.). Diese Labels sind Werkzeuge des Protokolls. Sie sind eine Art Gebrauchsanweisung der Zuordnung und der Wiederholbarkeit der Evaluation von Genderclassifier und der Herstellung eines geeigneten Evaluationswerkzeugs wie dem PPB.. Dies habe ich bereits theoretisch anhand des Genderstandards als Grundlage der Validität und der Reliabilität des Datensets aufgezeigt. Gender als Protokoll ermöglicht demgegenüber aber auch die Analyse des Prozesses. Anders formuliert greift Gender als Standard auf die Label als Untersuchungsgegenstand zurück, Gender als Protokoll hingegen bezieht sich auf das Labeln, also den Prozess der Standardisierung. Gender gibt so beispielsweise Buolamwini und Gebru einen Rahmen vor, der sich auf Punkte wie Mr, Ms beziehen kann (vgl. ebd. S.6). Gender beinhaltet also auch die von Murphy erwähnte Möglichkeit der politischen Aktion. Durch die Vorgabe einer bestimmten Praxis wie das binäre Gendern können Buolamwini und Gebru beispielsweise rassistisch begründete Fehlerraten der Genderclassifier aufzeigen (vgl. Murphy 2012, S.30, Buolamwini, Gebru 2018, S.8).

Gender als Set von Praktiken aufzufassen, ermöglicht hier als Ergebnis eine Evaluation Strategie für Genderclassifier. Diese wird entlang der Achse Race und Gender durch Buolamwini herausgearbeitet (vgl. Buolamwini 2017). Gender als agentielles Protokoll verbindet damit auch Halberstamms Ansatz des Automating Gender als Analogie zu Software mit künstlicher Intelligenz. Denn so wird Gender als Protokoll nicht Software ähnlich, in dem es genauso ausführbar ist wie diese, sondern ermöglicht das Wirken von Software, in dem Fall Genderclassifier. Gender wird zu einer posthumanistische Technologie, die nicht nur subjektiviert oder wie Halberstam schreibt, den „female Cyborg“ ermöglicht, sondern auch in dem Gender als Protokoll Voraussetzung für intelligente Software jenseits des konkreten Subjektes ist (vgl. Halberstam 1991, S.454)

Gender erfüllt hierbei eine Funktion des Protokoll, in dem eine Reihe von Praxen durch Gender angeleitet werden können, im Beispiel des Gendershades Projektes sind dies Kategorisierungsprozesse.

iii.) protokollarische Relation

Das Protokoll verweist jedoch nicht nur auf ein Set von Praktiken, sondern auch auf eine Kontextualisierung dieser. Hester macht in ihrer Konzeption des Protokolls als informationstechnologische Notwendigkeit auf diese Verortung aufmerksam. Protokolle sind immer auch kontextualisierte Ar-

tefakte, die sich aus anderen über das Protokoll hinausreichenden Bedingungen speisen. Die Zusammenstellung und Ermöglichung von Praxen sind auf andere Verhältnisse angewiesen, die ihnen vorrangig sind (vgl. Hester 2018, S.110).

Gender als Protokoll wäre innerhalb des Gendershades Projekt erst durch vorangegangene Faktoren bestimmt, wie beispielsweise die binäre Bildsprache der verwendeten Fotos für das Datenset von Bulamwini (vgl. Bulamwini 2017, S.55). Die Wissenschaftler:in verweist dabei explizit auf solche Faktoren, wie den binär operierenden zu zu untersuchenden Genderclassifier.

„All evaluated companies provided a „gender classification „ feature that uses the binary sex labels of female and male“ (Bulamwini, Gebru 2018, S.6).

Hierbei wird aus der Perspektive von „Gender als agentielles Protokoll“ Gender nicht nur zu einem Medium, sondern verbindet verschiedene Funktionsorte von Gender (Staat/ Firma/ Algorithmus) zu einem Ausgangsgrund, welcher dem Kategorisieren der verwendeten Bilder vorrangig ist. Gender wird dabei zu einem Fakt, der sich aus sich selbst ergibt; aus der Zusammenführung der Genderclassifier und des Benchmarks von Bulamwini entsteht eine binäre Genderkonzeption. Das Genderprotokoll (nicht Gender an sich) entsteht dabei aus vorrangigen Beziehungen, die als solche definiert werden (vgl. Barad 2007, S.231).

Dies verweist ferner auf Barads Konzeption der intraaktiven Kausalität: indem Gender bestimmte Exklusionen nach sich zieht, entsteht eine Kausalität. Das binäre Gendermodel muss von Bulamwini verwendet werden, weil es schon verwendet wird. Diese Kausalität trennt Gender im Gendershades Projekt von Gender in den Genderclassifiern, insofern als dass es eine zeitliche Unterscheidung gibt (Bulamwini verwendet das binäre Konstrukt von Gender erst nach dem Genderclassifier entstanden sind) und eine normative Trennung (Bulamwini verweist hier auf die Trennung zwischen der „technischen“ Notwendigkeit einer gender Binarität und einer „sozialen“ Welt, die eine Vielzahl von Gender beinhaltet (vgl. ebd. S.237). Diese Trennung wird wiederum im Moment der von Bulamwini und Gebru durchgeführten Evaluation zusammengebracht, in dem Genderclassifier das Benchmark analysieren, ohne dass diese beiden Genderoptionen im Moment der Ergebnisse zusammenfallen würden.

Denn die Genderbinarität der Genderclassifier und die des Datensets blieben in dieser Sichtweise des Phänomens bestehen, sie sind ja der Grund für die technische Ausgestaltung des Benchmarks und somit protokolarische Notwendigkeit. Beide Genderprotokolle führen in der Evaluation aber zu einer Verkettung der beiden in Form einer Performancerate, die angibt inwiefern der Genderclassifier Bildern der Genderkonzeption von Bulamwini entsprach (Bulamwini, Gebru 2018, S. 9-10)

iv.) Das Benchmark als agentieller Apparat

Zu dieser Betrachtungsweise möchte ich ferner auf die Verbindung von Nicht-Humanen und Humanen Akteur:innen kommen bzw. die nicht-humanen Akteur:innen als aktive Faktoren zu Bildung des Gendershades Projektes herausarbeiten. Ich möchte mittels dieser analytischen Perspektive die vielfältigen sozio-technischen Verbindungen innerhalb des Gendershades Projektes sichtbar machen. Dabei möchte ich zeigen, dass das Gendershades Projekt Gender in Form eines agentiellen Protokolls als Technologie verwendet, welches sich aber konstitutiv auf andere Protokolle und Phänomene beruft.

Das Protokoll als theoretischer Rahmen, verweist dabei auf eine Reihe an Verbindungen von unterschiedlichsten Komponenten innerhalb einer so geschaffenen Assemblage „as composed, not by the list of tools and components, but by the interminglings that make the tools possible“ (Murphy 2012, S.29) oder aber von Barad als Phänomen.

Im Falle des Gendershades Projektes sind diese verschiedene Faktoren und auch Strukturen, welche nicht Gender sind. Ich habe drei dieser Komponenten ausgewählt, da sie Grundlegend für das Verständnis des Gendershades Projektes sind: (a) Rassismus als Grund für das Gendershadesprojekt nach Buolamwini, (b) Künstliche Intelligenz als zugrundeliegende Technologie, welche Buolamwini untersucht und (c) Datensets, die neben Buolamwini entwickelten Benchmarks als Referenz für die Wissenschaftler:in hinsichtlich einer Unterrepräsentation Schwarzer Gesichter und Gesichter of Color dienen.

Rassismus

Eine wichtige und grundlegende Komponente ist der Genderclassifyern zugrundeliegende rassistische Bias, Schwarze Gesichter und Gesichter of Color mittels Gesichtserkennungsoftware „schlechter“ zu erkennen als weiße (vgl. Buolamwini 2017, S.13). Dies hat sich nach dem Projekt 2018 immer noch nicht grundlegend geändert und ist ein Problem, dass bis heute forstbesteht (<https://www.bitchmedia.org/article/coded-bias-racist-artificial-intelligence-interview> zuletzt abgerufen am 8.12.2020). Wie Chaterine D'Ignazio und Lauren F. Klein schon anhand des Gendershades Projektes analysierten, ist dieser Bias Ausdruck einer grundlegend rassistischen Struktur innerhalb westlicher Gesellschaften, die eingelassen sind in digitale Applikationen wie dem Genderclassifier und dort aktualisiert und synchronisiert werden (D'Ignazio, Klein 2020, S.33). Hierbei zeigte Buolamwini, dass Schwarze Frauen und Frauen of Color im Gegensatz zu weißen Frauen und Männern und auch zu Männern of Color und Schwarzen Männern öfter falsch gelabelt werden. Genderclassify erkennen Schwarze Frauen und Frauen of Color „schlechter“. Race und Gender als Set von Praxis der Bezeichnung und technischer Evaluation von Genderclassyfer stellen dabei Ergebnisse her,

die intersektionale Unterdrückung von jenen marginalisierten Gruppen sichtbar machen (vgl. Buolamwini 2017, S.13-15).

Künstliche Intelligenz

Eine weitere wichtige Komponente ist Künstliche Intelligenz, welche in den Genderclassifyern verwendet ist und Einfluss auf die Gestaltung der Bilder für das Datenset hatte (vgl. Alpaydin 2016, S.17). So schreiben Buolamwini und Gebru, dass sie Bilder von Parlamentarier:innen auch ausgewählt haben, weil diese gut ausgeleuchtet waren, neutrale Gesichtsausdrücke enthielten und ohne Zwang entstanden sind (vgl. Buolamwini, Gebru 2018, S.6.). Dies ist für die Bilderkennung als intelligente Technologie ein großer funktionaler Vorteil, da Gesichtserkennungssoftware mit solch genormten und ausgeleuchteten Bildern am verlässlichsten arbeitet. Neben Gender als technisches Merkmal der Bilder für das Benchmark ist auch computergenerierte Bilderkennung eine wichtige Leitlinie für das Auswählen und Verwenden der Bilder und damit für das Benchmark und das Gendershades Projekt (vgl. ebd.).

Datensets

Auch andere Datensets sind wichtig für das Gendershades Projekt und das Gender Protokoll. So stellen beispielsweise andere zugrundegelegte gegenderte Datensets eine wichtige Referenz für das „funktionieren“ des PPB eine wichtige Rolle dar. Die Datensets wurden aufgrund ihrer eigenen Wirkung als Evaluationsdatensets ausgewählt, sie sind auch Auditdatensets, mittels derer die Performance von Genderclassifyern gemessen wird. Diese wurden ob ihrer intersektionalen Verteilung von Race und Gender analysiert und dienten dadurch auch einer Standort Bestimmung des PPB (vgl. Buolamwini 2017, S.51)

Diese Beispiele sollen verdeutlichen, dass Gender als Protokoll als theoretische Basis auf anderen Faktoren beruht, die das Kategorisieren und Verwenden von Gender beschränken, wie ich es in der Analyse von Gender als kategorialer Standard getan habe. Gender als Protokoll und hier als agentielles Protokoll bezieht diese Verbindungen als Notwendigkeit der Beschreibung des Forschungsgegenstands (Forschungsphänomen nach Barad) mit ein. Gleichzeitig würde sich die Analyse und somit die Ergebnisse des Gendershades Projektes ändern, wenn mensch ein anders agentielles Protokoll zur Untersuchung heranziehen würde und Fragen würde, wie beispielsweise Race in das Gendershades Projekt eingelassen ist. Gender als Protokoll, wie ich es verstehe, ist auch mit der Kategorie Race und auch mit Unterdrückungsstrukturen konfrontiert. Denn Genderclassifier weisen einen Bias hinsichtlich Schwarzer „Frauen“ auf. Diese sind eine Kategorie, die sich durch die höchste An-

zahl an Missklassifikation durch Genderclassifier auszeichnet (vgl. ebd.S.64). Durch das Gendershades Projekt werden diese rassistischen, sexistischen Technologien sichtbar und Gebru und Buolamwini eröffnen eine Möglichkeit einer Lösung zu zukünftigen besserer Zuordnung Schwarzer Frauen und Frauen of Color. Jedoch formuliert Buolamwini keine queeres Genderprotokoll, da genderqueere Personen bzw. Bilder exkludiert werden (vgl.ebd.). Dabei ist es keine Race vs. Gender Sichtweise, die ich hier vertrete, vielmehr können unterschiedliche Politiken denkbar werden, die beispielsweise die besondere Verwundbarkeit Schwarzer trans Personen berücksichtigt.

Ich habe hier nun versucht Gender als agentielles Protokoll auf das Gendershades Projekt zu beziehen. Dabei habe ich mich auf Gender als Protokoll im Sinne einer Funktionsanleitung bezogen und die speziellen agentiellen Voraussetzungen des Gendershades Projektes hinsichtlich der von Buolamwini und Gebru verwendete Technik angeführt. Die These, dass Gender als Protokoll funktioniert, ist damit bestätigt. Ich möchte nun die polittheoretischen Überlegungen dazu erläutern, die sich aus den verwendeten Ansätzen von Murphy, Hester und Barad ergeben und diese Implikationen beispielhaft an einer spekulativen Erweiterung der Benchmark für das Gendershades Projekt erläutern.

c.) Polittheoretische Analyse

Die bisherigen Ausführungen haben polittheoretische Ideen wie die protokollarischer Macht und Hierarchisierungsprozesse gestreift, aber noch nicht konzeptioniert. Was also bedeutet es, Gender als agentielles Protokoll in meiner politikwissenschaftlichen Arbeit zu verwenden? Was sind die politikwissenschaftliche Implikationen, die sich aus solch einer theoretischen Perspektive, die sich aus transdisziplinären Ansätzen speist, ergeben?

Die polittheoretische Bedeutung des Ansatzes eines agentiellen Protokoll ergibt sich ausgehend von meiner theoretischen Basis aus drei Faktoren, (a) aus einer dezidiert technischen Idee des Politischen, welche Murphy als Technopolitics (Murphy 2012, S.39) einführt, (b) aus einer Fokussierung auf Agency und der Überlegung der Anwendung oder Operationalisierung, wie Hester dies formuliert und (c) in einer technisch vermittelten politischen Ebene, die Hester wiederum als Mesebene bezeichnet und die ich mit Barads Idee der Intraaktion verbinde.

Ich möchte in diesem Abschnitt der Arbeit politikwissenschaftliche Impulse ableiten und sie in Form eines alternativen Benchmarks auf die technische Funktion des PP-Benchmarks von Buolamwini beziehen. Anders formuliert: Die politikwissenschaftliche Interpretation der Funktion von Gender als agentielles Protokoll werde ich in Form eines queeren Benchmarks erarbeiten.

i.) Technopolitics

Murphys Ansatz des Protokollfeminismus bietet einen konkreten Einsatz des Politischen innerhalb des Protokolls. Sie nennt diesen Punkt „technopolitics—a sense of the immanence of politics in practices and technologies” (ebd.). Technologien und deren Anwendung sind im Protokoll verankert und stellen die politische Handlung dar. Politik selbst wird hier zu formgebenden Funktion der Anwendung, das halten des Spekulum in Murphys Beispiel ist ein politischer Akt, der sich aus dem Zusammenschluss von Spekulum, patriarchaler medizinischer Praxis und dem Zusammenschluss der Workshop Teilnehmenden ergibt. Der (männliche. weiße) Arzt wird aus der Praxis des Entdeckens verbannt. An die Stelle rückt ein Kollektiv, welches sich das Spekulum, als verortete Praxis untereinander beibringt und Räume jenseits der Klinik eröffnet, um zusammen zu arbeiten. Das Politische ist hier nicht externalisiert, sondern auf die Zusammenführung von Körper und kollektiven brüchigen Identitäten durch technische Praxen bezogen.

„The small self help group heralded this self-possessed subject through technical acts on the body, binding the figure of the self-determined person, not to nation or capitalism but to a “community of women” or “sisterhood” as its condition of possibility.” (ebd. S.38).

Technopolitik verweist auf das Zusammenkommen und Handeln. Dabei ist die Form, die das Politische hier annimmt, materiell, sprich verkörpert. Der untersuchte Körper wird das Politische.

Mit Barads Idee des Apparates möchte ich hinzufügen, dass in solchen protokollarischen Prozessen nicht ausschließlich ein Subjekt als Ergebnis stehen muss. Die Technopolitik eines agentiellen Protokolls verweist somit auf eine Verbindung verschiedenster Akteur:innen, jedoch nicht mit dem Subjekt als Mittelpunkt bzw. Output des Protokolls, wie etwa bei Murphy (vgl. ebd.). Vielmehr verweist das technopolitische in Anlehnung an den Theoretiker José Esteban Muñoz, welcher eine Theorie Queerer Utopie entwirft, auf ein „doing somethingelse“ (Muñoz 2009, S.154) Muñoz sieht die Stärke queerer Politiken in einem Unterbechern vorherrschender (heterosexueller) Protokolle (vgl. ebd. S.155). Das Techno-Politische ist somit ein Handeln, das sich durch den Gebrauch von Technologien herrschenden Protokollen entzieht.

ii.) Operationalisierung und Agency

Helen Hester und Murphy erklären ferner, dass die Besonderheit der politischen Theorie des Protokolls in seiner Anwendung liegt (vgl. Hester 2018, S. 114, Murphy 2012, S.28). Im Gegensatz zum Standard, den ich im vorangegangenen Teil vorgestellt habe, beschreibt das agentielle Protokoll als politische Technik Prinzipien der Selbstermächtigung und des Unterbrechens bestehender Praxen,

anstatt wie etwa Star und Bowker, Standards und Kategorien als fundamental für die Organisation einer Gesellschaft sehen (vgl. Muñoz 2009, S.155, Bowker, Star 2000, S.60-61).

Protokollfeminismus demgegenüber stützt sich auf marginalisierte Gruppen, die Wissen austauschbar und anwendbar machen. Die konkrete Umsetzung und Anpassung des Wissens ist entscheidend für diese politische Strategie, wobei ein Hauptaugenmerk auf die einfache Umsetzung des Protokolls und das Hinführen zu Technologien liegt. Dabei werden Technologien marginalisierten Personen und besonders Gruppen zugänglich gemacht, so dass diese Technologie anwenden und gebrauchen.

„Further, the ease of vaginal self-examination suggested the possibility of learning and hence appropriating other techniques, such as abortion. It taught, according to Downer, “how easy it was to learn these things—that they were learnable. They were not rocket science.” (Murphy 2012, S.48).

So könnte der als erweiterte politische Strategie von mir eingeführte Workshop Drag vs. Ai vor dem Hintergrund des Protokollfeminismus auch durch die Zirkulation des Wissens als feministischer Akt gesehen werden. In diesem sollen die Teilnehmenden mittels Drag Makeup lernen, Gesichtserkennungssoftware mittels fluider Genderidentitäten zu verwirren, so die Initiator:innen (vgl. <https://www.ajl.org/drag-vs-ai> zuletzt abgerufen am 03.02.2021). Denn die Workshopinstruktionen kann mensch sich frei zugänglich zuschicken lassen und selbst anbieten. Ferner kann Schminken und Verkleiden als technische Praxis innerhalb des Feldes von automatischen Genderclassifier eine Funktion der Operationalisierung darstellen. Sind Lidstrich Ziehen und Make up Auftragen doch Fähigkeiten, die sich durchaus einfacher erlernen lassen als das Programmieren einer KI, einfach aufgrund von Makeup als verbreiteten und zugänglichere Technologie in Bezug auf Gender (vgl. Wolf 1996).

iii.) Mesopolitik

Ich möchte auf einen weiteren wichtigen Punkt der politikwissenschaftlichen Bedeutung von Gender als Protokoll eingehen und dieses Konzept von der vorangegangenen Analyse von Gender als kategorialer Standard abgrenzen. Während der Genderstandard sich dabei, wie ich gezeigt habe, auf staatliche Strukturen und deren Wirken auf das Individuum bezieht, greift das Protokoll eine neue Ebene der Politik auf. Diese Ebene bezieht sich auf politische Handlungen zwischen staatlichen Institutionen und Individuen

Bei diesen Überlegungen beziehe ich mich auf Helen Hester. Diese fügt dem Protokoll einen neuen politischen Ort hinzu. Dieser ergebe sich aus den Eigenschaften des Protokolls, als adaptiv und rela-

tional, also der Anpassung an jeweilige lokale Begebenheiten und den Beziehungen zwischen den Akteur:innen und Strukturen, die diese in die Notwendigkeit des Handelns zwingen (vgl. Hester 2018, S.110). Wie Murhy sieht Hester dabei einen politischen Zusammenhang zwischen der Skalierbarkeit des Problems, also des verallgemeinerbaren Problems, das spezifische gesellschaftliche Gruppen betrifft (in Murphys Beispiel, die Unterversorgung von Frauen im Bereich der „Frauen Gesundheit“) und der jeweiligen Notwendigkeit der Anpassung (vgl. Murphy, S.40-41). Das Recht auf Abtreibung mag, so Murphy, dabei für weiße Frauen ein Ziel sein, da ungewollte Schwangerschaften das Problem sind. Für Schwarze Frauen jedoch macht dies nur ein Teil des Problems aus, sind sie doch auch stärker mit Fragen der staatlichen Geburtskontrolle und Fragen der Finanzierbarkeit eines Kindes konfrontiert, da sie proportional stärker von Armut und staatlicher Kontrolle betroffen sind (vgl. ebd.). Hester leitet daraus die Idee der Mesoebene des Protokolls bzw. die Mesoebene als die Ebene der Politik des Protokolls ab. Diese versteht Hester als Verbindung zwischen dem Individuum und Makrostrukturen wie dem Staat, jedoch nicht als ein abzählbares oder begrenztes Dazwischen, sondern als Vernetzung dieser beiden Pole durch die Übertragbarkeit und somit Skalierbarkeit des Protokolls:

„[T]he mesopolitical operates between atomized, hyper-local interventions at the level of, for example individual embodiment (micropolitics), on the one hand, and big-picture, speculative projects permitted on the whole sale over trowel of power“ (Hester 2018, S.114).

Dabei ist Protokoll durch seine Anwendung, Modifizierung und Verhandlung über den politischen Sinn immer in einem performativen Prozess (vgl. ebd. S.112-3). Diese performative Ebene der Politik des Protokolls kann auch auf Barads Konzept des Apparats als Vermittlung von politisch und technischen Phänomenen angewendet werden. Somit kann das Protokoll noch stärker auf die post-humane Handlungsfähigkeit der Genderclassifayer innerhalb des Gendershades Projektes gelenkt werden. So entwickelt Barad in ihrer Theorie des agentiellen Realismus eine Idee soziotechnischer Kategoriebildung (Barad verhandelt diese Idee am Beispiel des Arbeiters in einer fordistischen Fabrik), die sich nicht auf ein Individuum wie „den Arbeiter“ bezieht, sondern auf eine diskursive materielle Praxis. Diese besteht aus Subjekten, wie in ihrem Beispiel Arbeiter:innen, technischen Artefakten wie der Fabrik und anderen Apparaten wie Gender und Class. In Barads Beispiel konturieren diese Akteur:innen in intraaktiven Beziehungen die Kategorie des „Arbeiters“ (vgl. Barad 2007, S.242). Barad wird hierbei konkreter als Hester in ihre kurzen Erklärungen zu Mesopolitiken und erklärt den Prozess des Entstehens von Makrostrukturen wie Kapitalismus durch konkrete materielle Praxen, wie dem Arbeiten in einer Fabrik und „dem Arbeiter“ als Verbindung konkreter Begebenheiten und abstrakten Kategorien, welche über die materiellen Begebenheiten hinausgehen. Dabei sind diese Strukturen ihrerseits wieder angebunden sind an materielle Bedienungen. Anders ausge-

drückt, der Arbeiter entsteht durch einen Ort, die Fabrik und die Ideen kapitalistischer Produktion, welche ihrerseits wieder in Maschinen, Schichtplänen und ähnlichen in der Fabrik als materiellen Ort eingelassen sind (vgl. ebd. 244). Der Arbeiter wird ferner durch Kategorien von Race, Gender und Class konstruiert, die ihn als ein weißes männliches Subjekt innerhalb dieses Kontextes erscheinen lassen. Der Protokollfeminismus als Erweiterung von Barads Idee der Intraaktion bietet eine Idee, wie diese Verbingung als adaptive und relationale Praxis aussehen und verändert werden könnte. Ein Streik innerhalb der Fabrik könnte als ein Computerprotokoll zwischen kapitalistischen Strukturen, Kollektiver Willensbildung, individuellen Bedürfnissen und der materiellen Umgebung der Fabrik dienen.

Barads Idee der Intraaktion macht so die ständige Verbindung von Makrostrukturen und materieller Verortung anhand des Beispiels Fabrikarbeiter deutlich und ermöglicht eine Analyse, die es erlaubt, das Verbinden von Mikro- und Makroebene durch Komponenten, die Handlungsmacht in diese konkreten Situationen erhalten, zu fassen. Diese lassen sich auch nicht völlig auf menschliche Handlungen reduzieren, wie das Beispiel des Fabrikgebäudes als Ort der materiellen Ordnung verdeutlichen sollte (vgl. ebd. S.239). Hesters und Murphys Idee des Protokolls macht diese Verbindung zu konkreten Strategie emanzipatorischer Politik, die Barad antizipiert.

Dabei erörtert Murphy die Idee der Kleingruppe als konkrete lokale Akteur:in jener Strategien und für Hesters Idee der Mesopolitik wird diese Überlegung von zentraler Bedeutung, ermöglicht die Gruppe doch eine verhaftete lokale Möglichkeit des Handelns. (vgl. Hester 2018, S.107; Murphy 2012, S.48-49)

Ausschlaggebend nach Murphy ist bei der Umsetzung eines feministischen Protokolls die Idee der Anwendung innerhalb und durch eine Gruppe und nicht die individuelle Handlung. So weist Murphy explizit auf die politische Taktik von Kleingruppen als Ort politischer Umwandlung hin. Diese Taktik, so Murphy, wendet sich einerseits gegen eine neoliberaler Idee des Individuums als Handlungsträger:innen, welche im historischen Kontext der USA der 70er Jahren als Kontraerzählung zu kommunistischen Ideen der Gruppe als politische handelnde und im Zuge des Kalten Kriegs politisch genutzt wurde, das Kollektiv als antiamerikanische politische Formation zu etablieren. Gleichzeitig bedienten sich die „self-help“ Gruppen jedoch Ideen der Wirtschaftspsychologie, die wirtschaftliche Innovation am ehesten durch Kleingruppen verwirklicht sah (vgl. Murphy 2012, S.61). Die Kleingruppe bildete in diesem konkreten Beispiel eine politische Einheit der Möglichkeit aber auch der Bestärkung hegemonialer Denkweisen bezüglich wirtschaftlicher Effizienz.

Weiter formuliert für meine Arbeit bleibt festzuhalten, dass der Zusammenschluss kleiner Gruppen nicht gleichbedeutend mit politischer Emanzipation ist, sondern durch historische Bedingungen ge-

formt ist. In Anbetracht von Corona kann heute beispielsweise gefragt werden, ob die physische Präsenz einer Gruppe nicht an sich ein unsolidarischer Akt der Gefährdung Anderer in Beziehung zu dieser Gruppe ist.

Zusammenfassend stärkt der Protokollfeminismus eine Theorie der Handlung, welche lokal und temporal verortet sind. Die Idee der Gruppe als Ort der Handlung und Möglichkeit politisch zu agieren, ist dabei ein ambivalenter Ort, welcher sowohl inklusive als auch exklusive Momente beinhaltet (vgl.ebd.).

Was können diese Ideen für die politikwissenschaftliche Interpretation der Genderfunktion innerhalb des Gendershades Projektes jedoch genau bedeuten? Um dieser Frage nachzugehen, möchte ich in einem zweiten Schritt eine spekulative Protokollpraxis einführen, die sich auf gegenderte Benchmarks bezieht, wie sie Buolamwini verwendet und analysiert und gleichzeitig auf die hier besprochenen politischen Möglichkeiten eingehen.

Gender als Protokoll verbindet dabei, wie ich bereits gezeigt habe, für das bestehende Benchmark verschiedene Komponenten und ermöglicht gleichzeitig eine Anpassung des Gendershades Projektes an technische, materielle. sowie politischen Strukturen. in die das Projekt interveniert. Was würde es aber bedeuten, Gender als Protokoll zu betrachten, das Genderbinarität queert? Wie könnte Buolamwinis Vorschlag, weitere Benchmarks zu kreieren, auch genderqueere Menschen mit einzu-beziehen gelingen und wie könnte eine Praxis aussehen, die solch ein Protokoll antizipiert?

Dies werde ich in einem abschließenden Schritt darstellen.

d.) Queeres Benchmark als feministisches Protokoll (Polittheoretische Intervention)

Ich möchte hier ein beispielhaftes Protokoll für die Erstellung queere Benchmarks angeben. Dabei beziehe ich mich auf den von Buolamwini formulierten Anspruch ,weitere Benchmarks zu gestalten, die Evaluationen von Genderclassiifyer ermöglichen und dabei intersektional verortet sind. Das meint hier die Verbindung verschiedener Kategorien, wie Race und Gender, zur Beurteilung von Genderclassiifyern in die Gestaltung eines Datensets aufzunehmen.

Mittlerweile gibt es sowohl akademische Forschung als auch feministische aktivistische Projekte, die sich mit Fragen der Inklusion genderqueerer Menschen in digitale Strukturen beschäftigen und dies auch vor dem Hintergrund der besonderen Vulnerabilität von Schwarzen trans Personen. So hat beispielsweise die Künstler:in Danielle Brathwaite Shirley ein Schwarzes trans online Archiv ins Leben gerufen, um einen digitalen Ort der Erinnerung und Sichtbarmachung für eine Schwarze Trans Community zu schaffen (vgl. <https://blacktransarchive.com/> zuletzt abgerufen am 7.12.2020).

D'Ignazio und Klein haben des Weiteren auf Projekte verwiesen, die vorrangig Praxen der Datengenerierung und Verarbeitung durch marginalisierte Gemeinschaften fördern (D'Ignazio, Klein 2020).

Ferner hat die Genderforscherin Hemangini Gupta in einem kürzlich veröffentlichten Beitrag ein Protokoll entworfen, um in Bangalore durch die Entwicklung von technischen Artefakten wie Robotern, die globalisierte Techindustrie zu stören. So schreibt sie zu dieser politischen und Theoretischen Arbeit:

„This manual offers a startup kit to assemble a team, invent a robot, pitch to funders, and potentially disrupt an industry!“ (Gupta 2020, S.2) .

Die Idee des queeren Benchmarks ist eingebettet in diese Praktiken. Es ist eine darin verortete Skizze, die nicht in einem luftleeren Raum existiert, sondern an jetzige Projekte anknüpfen kann.

Ich möchte diese Idee in vier Schritten erläutern. Dabei setzte ich die Arbeitsweise von Buolamwin als zu adaptierende Praxis voraus. Ich werde deshalb zur Erstellung eines queeren Protokolls für ein Benchmark zuerst die Kategorisierung von Bildern erläutern. Darauf aufbauend werde ich zweitens die Bildproduktion erläutern bzw. die Auswahlprozesse für die verwendeten Daten. Dabei werde ich einerseits erklären, durch wen diese Schritte durchzuführen sind und zweitens herausarbeiten, von wem die Bilder des Benchmarks stammen könnten. Drittens werde ich aufgrund des gewählten Protokollansatzes und Buolamwinis Arbeit auf die Wiederholbarkeit dieser Prozesse eingehen bzw. erklären, inwieweit sie nachahmbar sein könnten und zum Schluss erklären, welche politischen Prozesse dieses Protokoll konturieren.

i.) How to queer AI

Buolamwini und Gebru beschreiben, wie ich bereits dargestellt habe, den Labelprozess als Bestimmen einer Datengrundlage und das Weiterverarbeiten dieser, etwa Bildauswahl und Kategorisierung. Diesen Prozess möchte ich hier beispielhaft öffnen und eine auf queere Gemeinschaftsarbeit beruhende Idee eines feministischen Protokolls für ein Benchmark erörtern.

Als Grundannahme bleibt dabei die Kategorisierung von Bildern aufgrund bestehender Genderclassifiers, wie sie Buolamwini annimmt (vgl. Buolamwini, Gebru 2018, S.6). Das Erstellen eines queeren Benchmarks ist immer an einen binären Genderclassifier gebunden.

Labeln als queering

Das Kategorisieren ist ein wichtiger Punkt der Idee des queeren Benchmarks, da Genderkategorien eine wichtige politische Rolle in feministischen Theorien spielen. So schreibt Jack Halberstamm bei-

spielsweise, dass das Benennen der vielfältiger Genderkategorien essentielles politisches Handeln sei:

„That said, one of the biggest innovations of the past two decades in relation to gendered expression indeed has been the production, circulation, and usage of just such a vernacular language for non normative gender and sexual expression” (Halberstam 2018, S.10).

Dabei bezieht sich Halberstam vor Allem auf den selbstgewählten Namensgebungsprozess, der entgegen den Klassifikationen, die ich im vorangegangenen Teil vorgestellt habe, nicht auf eine externe Instanz verweist, die Betroffene zu einer Unterordnung zwingt.

Ich möchte hier zwei Ideen, wie diese Praxis für ein queeres Benchmark genutzt werden könnte, einführen. (a) Eine Strategie könnte ein Aneignungsprozess der vorhandenen Kategorien durch genderqueere Menschen sein, (b) eine andere Taktik könnte eine Störung der Genderclassifier sein.

Eine aneignende Strategie der Kategorisierung der Daten könnte das Aufbrechen tradierte Genderrollen Erwartungen sein. So könnten beispielsweise die lesbisch gelebten Genderkonzeptionen von Frau von betroffenen Personen genutzt werden, um die Kategorie „female“ die Genderclassifier, wie Buolamiwni sie untersucht hat, aufzubrechen.

Hierzu könnten beispielsweise male/ female Kategorien als Platzhalter dienen. Unterkategorien wie die lesbischen Bezeichnungen Butch/ femme²⁶ könnten von betroffenen Personen genutzt werden um Weiblichkeit aufzubrechen. Butch Lesben könnten sich in einem Selbstbezeichnung Akt als male codieren, ohne dass sie Männer sind. Dies könnte nur mit dem absoluten Einverständnis der jeweiligen Personen passieren, andernfalls wäre dies Misgendering und ein Akt transfeindlicher oder queerfeindlicher Gewalt (vgl. Halberstam 2018, 118-9).

Queere Benchmarks als Störung von Genderclassifiern

Zweitens könnten auch als Art „library of missing Data Sets“ solche Bilder kategorisiert werden, die dezidiert queer/ trans/ nicht binäre Personen darstellen und diese Kategorien als Evaluationskriterien einführen. Dies würde zu einer völligen Außerkraftsetzung der gewohnten Performancesung der Genderclassifier führen mit 0 Prozent Trefferquote.

In einer der aktuellsten Studien zum Thema Benchmarks und Künstliche Intelligenz kommen Nie et. Al. zu dem Schluss, dass es neue Datensets braucht, die Fehler der zu evaluierenden Programme nachbereitet, um weiterhin wirkungsvolle Evaluationen durch Benchmarks zu generieren, also die Technologien zu verbessern. So könnte eine erweiterte Methode der Evaluation folgende sein:

“For instance, we collected annotator-provided explanations for each example that the model got

²⁶ Die Beziehungen Butch und Femme beziehen sich auf eine lesbische Bezeichnungspolitik. Lesben, die sich als männlich (nicht als Mann) verstehen sind Butch und Lesben die sich auf eine feminine Identität beziehen Femme. Ausschlaggebend für diese Einteilung ist die Selbstbezeichnung dieser Personen (vgl. Halberstam 2018, S.143)

wrong. We provided inference labels for the development set, opening up possibilities for interesting more fine-grained studies of NLI model performance.” (Nie et. Al 2020, S.9).

Nie et. Al verweisen dabei auf die Möglichkeit, schon innerhalb der Trainingsdatensätze die Bilder, die falsch eingeordnet wurden, von lernenden Algorithmen zu korrigieren.

Queere Benchmarks könnten solche Evaluationen dabei einerseits bereits im Vorfeld einarbeiten, in Form neuer Kategorien für „falsch klassifizierte Bilder“ genderqueerer Menschen. Dies geschähe jenseits der von Genderclassifyern angebotenen Genderbinarität und mit einer, wie Halberstam beschreibt, jahrzehntelanger Erfahrung und Praxis der Genderkategorisierung und ermöglichte eine fundierte Fehleranalyse für die an Benchmark getesteten Genderclassier anzu bieten (vgl. Halberstam 2018, S.8)

Beide Vorgehensweisen würden in der von Buolamwini und Gebru etablierten Arbeitsschrittform durchgeführt werden können und würden eine Erweiterung durch einen queeren Standard für Genderclassier ermöglichen. Beide Vorgehensweisen schließen sich dabei nicht aus, sondern könnten ergänzend verwendet werden.

Datensets und gemeinschaftliche Praxen

Ein weiterer wichtiger Punkt ist die Frage, wer kategorisieren soll? D’Ignazio und Klein verweisen in ihrem Konzept des Data Feminism auf genau diese Frage, wenn sie schreiben, dass das Arbeiten mit Daten immer auch eine Form von Macht beinhaltet. Dennoch verweist die Idee des queeren Benchmarks auf einen Widerspruch (vgl. D’Ignazio, Klein 2020, S.23). Wie bereits angeklungen ist, behandelt das Erarbeiten alternativer Genderkategorien eine Selbstbezeichnungspraxis. Gleichzeitig verweisen jedoch Benchmarks auf eine große Anzahl an Daten. Wie könnte ein queeres Benchmark diese Gegensätze feministisch vereinen?

Ich möchte hier auf die Idee des Protokollfeminismus zurückkommen. Murphy hat in ihrer theoretischen Konzeption auf die notwendige Idee der Praxen und des ermöglichen von Versammlungen gesprochen. Diese Idee könnte auch in die Produktion einer queeren Benchmark aufgenommen werden. Ein Beispiel können dafür schon jetzt beispielsweise trans Hackathons sein, die sich explizit mit Gender und dessen Wirkung auf Technik beschäftigen (vgl. Christensen, Conradi 2017, S.86). Ein Datenset bzw. Benchmark könnte so in Zusammenarbeit mit Akteur:innen der jeweiligen queeren Gemeinschaften erarbeitet werden. Das jeweilige queere Benchmark könnte sich so anhand bestehender lokaler Strukturen, wie Gemeindezentren, queeren Bars und ähnlichen Orten anpassen.

Bildproduktion

Innerhalb dieser gemeinschaftlichen Prozesse wäre dann des Weiteren zu fragen, wie die Bilder, die

für die Benchmarkproduktion genutzt werden würden, entstehen sollten, bzw. woher sie stammen sollten. Bilder könnten einerseits selbst aufgenommen werden und so sowohl die Bezeichnung des Fotos an die jeweiligen Personen, die die Kategorisierung betrifft binden und zweitens könnten auch Formen der gewünschten Repräsentation gewählt werden, beispielsweise wäre denkbar das trans Personen sich eventuell ungern in Transition für Bilder zur Genderclassiyyer Evaluation zur Verfügung stellen würden, da das eventuelle Misgendern traumatische Folgen haben könnte oder aber als politischer Akt der Sichtbarmachung der Transition als vergeschlechtlichtem Raum und wichtigen Teil eines persönlichen Gender Protokoll untermauern (Preciado 2016,

Keyes 2018). Gleichzeitig könnten auch intergenerationelle Prozesse in Betracht kommen, die sich beispielsweise auf queer Archive vergangener Jahrzehnte beziehen und so noch einmal andere Ausdrucksformen von geschlechtlicher Identität aufzeigen könnten. Dabei rekurrieren queere Archive zumeist auf eine Verbindung von Privat und Öffentlich, die sie als Anknüpfungspunkte für Arbeiten innerhalb queere Gemeinschaften prädestinieren. So beschreibt die Medientheoretikerin Kate Eichorn in ihrer Anslyse zu dieser Form der Archive folgendermaßen: „Queer collections not only frequently begin in the private sphere but also have often taken shape under the unique constraints of queer life. Many started out in the closet, so to speak, and originally functioned as sites of retreat for their collectors.” (Eichorn 2018, S.124).

Queere Archivarbeit und Fotografien könnten so eine politische Praxis des gemeinsamen Erinnerns werden, welche für Eichorn ein wesentliches Merkmal dieser Archive ist. Ferner könnte diese Zusammenführung zu einer historischen Kontextualisierung queerer Identitäten innerhalb der Genderclassiyyer führen. Denn beide Bildproduktionsstratgeien würden ein über Jahrzehnte hinweg spannenden Benchmark queerer Bilder generieren, welches sich nicht nur auf queere Menschen des 21. Jahrhunderts bezöge.

Diese Generierung von Bildern würde schlussendlich, über die von Buolamwini gestellten Fragen von Datenschutz als Frage der Einwilligung hinausgehen (vgl. Buolamwini 2017, S.53). Die Bildproduktion in einem queeren Benchmark entspräche einer Art der Konsensfindung hinsichtlich politischer Fragen von Repräsentation queerer Gesichte und queerer Subjektivität innerhalb verwundbarer Zustände. Diese Idee des Erarbeitens eines Datensets würde Care Arbeit hinsichtlich des Kuratieren eines solchen Benchmarks ebenso beinhalten, wie die Erarbeitung transinklusive Methoden der Dokumentation der eigenen Genderidentität.

Ressourcen

Neben der Bildreproduktion sind auch Fragen nach dem Zugang der Ressourcen für ein feministisches Protokoll zur Erstellung eines Benchmarks nach Murphy von entscheidender Bedeutung.. Murphy

phy erläutert in ihrem Konzept des Protokollfeminismus, dass ein feministisches Protokoll immer wieder kritisch befragt werden muss hinsichtlich seiner Reproduktion von rassistischen oder sexistischen Strukturen. Deswegen ist die Frage vor Allem nach materiellen Ressourcen wie Geld und die von ideellen Ressourcen, wie Studienabschlüsse für ein queeres Benchmark von entscheidende Bedeutung, soll sich dieses nicht in die Reproduktion bereits erläuterter Ungleichheiten beteiligen. So sind marginalisierte Gruppen besonders von Armut betroffen und verfügen nicht über ausreichend Mittel, um Studienabschlüsse zu erreichen und sind prozentual am wenigsten vertreten in hohen Positionen innerhalb der Wirtschaft und Forschung zu Künstlicher Intelligenz (vgl. D'Ignazio Klein 2020, S.28).

Ein wichtiger Punkt wäre daher einerseits, die Arbeit an Benchmarks und vor allem die Evaluation zu vergüten. Es wäre eine Idee beispielsweise im nordeuropäischen Kontext diese Vergütung durch staatliche Förderung abzusichern.

Operationalisieren

Bis jetzt bin ich vor allem auf die einzelnen zu beachtenden Praxen eingegangen, die eine genderqueeres Protokoll hinsichtlich eines Benchmarks bräuchte. Ich möchte jetzt noch einmal auf die zugrundeliegenden protokollarischen Funktionen eingehen, die Murphys Idee des Protokollfeminismus anleiten: Die Idee der Operationalisierbarkeit der Praxis und der Nachahmbarkeit (vgl. Hester 2018, S.103-104, Murphy 2012, S.48)

Die Operationalisierbarkeit des queeren Benchmarks hätte zwei konkrete Ausformungen in dieser spekulativen Idee. Zuerst würde die bereits vorhanden Praxis des eigenen Bezeichnens von Gender und das intensive Auseinandersetzen mit Genderkategorien als einfache bzw. gewohnte und geübte Praxis in queeren Kreisen als niederschwelliger Einstieg für das Arbeiten mit KI bzw. Genderclassifizieren dienen. Zeitens würde diese Praxis technisches Arbeiten an KI ermöglichen, was nicht die queere Community Technologie lernen lässt, bzw. Queere Menschen als Lernende einordneten, sondern ihrer gelebte Erfahrung als Technologie in KI Produktion einführt .

Ein zweiter wichtiger Punkt ist die Einfachheit der zu erledigenden Aufgaben. Nach Murphy verweist das feministische Protokoll auf eine Anleitung zu einer Praxis, die möglichst inklusiv und für viele Menschen offenstehen soll. Dies lässt sich vorrangig dadurch erreichen, dass diese Praxen so einfach gehalten sind, dass viele Menschen sie nachahmen können. Murphy hat dies wie bereits geschrieben deutlich gemacht anhand der Selbstuntersuchung der eigenen Vulva mit einem Spekulum (vgl. Ebd.). Hierbei würde ich jedoch Murphys Ausspruch des „einfachen Untersuchens“ spezifizieren als Tätigkeit, die von queeren Menschen oft ausgeführten Tätigkeiten, nämlich das Selbstbezeichnung und die Arbeit mit Genderkategorien (vgl. Halberstam 2018, S. 10).

Gendern als Technik 2.0

Wie Murphy schreibt, hilft die Einfachheit, Menschen an Technik zu gewöhnen bzw. diese mit ihr vertraut zu machen. Ich würde im Bezug der queeren Benchmarks jedoch weitergehen. Nicht die Technik an sich wird Personen nahegebracht, sondern ihre eigenen Praxen und die politischen Kämpfe um Anmerkung und Sichtbarkeit sind Techniken, die wirksam werden in Bezug auf Genderclassifyer. Hier kann Gender nach Murphy selbst als technopolitische Intervention verstanden werden, die sich in den Akt der Erschaffung des Benchmark einschreibt, als Kontrollvariable und als Praxis innerhalb der Data Science (vgl. Murphy 2012, S.48). Gender als Protokoll ermöglicht ein politisches Eingreifen in Genderclassifyer durch die Praxis des queeren Benchmarks.

Die hier aufgeworfenen Ideen des queeren Benchmark sind dabei durch die protokollarischen Funktionen der Adaption und Relation, die Hester als Hauptmerkmal der Politiken rund um Protokolle aufgeworfen hat, geprägt (vgl. Hester 2018, S. 110). Einerseits bleibt das Benchmark der Idee der Repräsentation und der Wirkungsweise von Genderclassifyern verhaftet. Andererseits habe ich in dieser Projektskizze die Idee hinsichtlich der Störung und Verwirrung jener Technologien durch eine queere, konsensorientierte und selbstbestimmte kollektive Praxis erweitert.

Die Idee des queeren Benchmark liegt somit zwischen der strikten und unveränderlichen Beibehaltung der binären Genderklassifikation von Buolamwini und Gebru, die sich mit ihrem Benchmark stärker auf den Evaluationprozess und damit auf Faktengewinnung für eine parlamentarische Debatte rund um die Verwendung von Genderclassifyern bemühen einerseits und der von Os Keyes vertretenden Idee der strikten Ablehnung von Genderclassifyern andererseits (vgl. Keyes 2018, S.88:13; D'Ignazio, Klein 2020, S.32). Jedoch ist die Konzeption dieses queeren Benchmarks anpassbar und steht einem Verbot oder der parlamentarischen Arbeit von Buolamwini nicht entgegen. Die jeweiligen Gruppenkonstellationen innerhalb der Benchmarkgestaltung können über gewünschte Verbindungen, Koalitionen oder aber gänzlich andere Verbindungen nachdenken, das queere Benchmark als feministisches Protokoll soll hierbei einen Möglichkeitsraum schaffen, eigene politische Bedürfnisse zu artikulieren (vgl. Murphy 2012, S.30).

Ich möchte abschließend noch einmal auf die Verbindung dieser spekulativen Idee mit dem bereits existierenden Gendershades Projekt zurückkommen und die Beziehung von Gender mit der Idee des queeren Benchmarks kurz verdeutlichen anhand der Frage, ob Gender ein feministisches Protokoll sein kann bzw. was dies bedeuten würde als theoretische Idee für ein queeres Benchmark.

Gender als feministisches Protokoll?

Ich habe in der Analyse des Gendershades Projektes Gender als Protokoll vorgestellt und diese Perspektive angewandt auf das Projekt. Nun möchte ich kurz erläutern, warum Gender auch als Protokoll im queeren Benchmark wirkt. Gender funktioniert als spezifisches Protokoll innerhalb des queeren Benchmarks, da es sich auf Selbstbezeichnungspraxen und intersektionale Ideen von Verknüpfungen von Gender mit Race und Class bezieht. Es besteht aus verschiedenen Komponenten, wie Geld, Fotos und queeren Archiven. Gender ist dabei keine dem einzelnen Akteur zuzuschreibende Komponente, sondern sowohl ein Zuschreibungsprozess innerhalb kollektiver Bemühungen. Gender entfaltet dabei seine Wirkweise durch Technologien wie Fotografieren oder aber einen Diskurs zur Ermittlung verschiedener Genderkategorien und kann gleichzeitig als gemeinsames verbindendes Element unter verschiedenen Kollektiven gelten, die sich mit Gender als Parameter des Benchmarks auseinandersetzen, als auch mit dem Ausgangsbenchmark von Buolamwini und Gebru.

Der Protokoll Ansatz stellt somit eine Form transformativer politischer Arbeit dar, welche nutzbar gemacht werden könnte für weitere Auseinandersetzungen mit Genderclassiferyer.

Ich möchte abschließen noch einmal die Ergebnisse des gesamten Kapitels vorstellen.

5.) Zwischenfazit

In diesem Teil der Arbeit habe ich untersucht, inwiefern Gender und die technischen Wirkungsweisen des Gendershades Projektes zusammenhängen.

Ich habe ausgehend von Wendy Chun und Beth Colemans Theorie „Race as Technologie“ eine theoretische Grundlage geschaffen, Gender als Technologie zu fassen und so ermöglicht, die Forschungsfrage, „wie ist Gender in das Gendershades Projekt eingelassen“ auf eine spezifische Weise zu untersuchen und zwar in Hinblick auf die Verbindung von Gender und Technologien (vgl. Chun 2009, Coleman 2009).

Ich habe so zwei Konzepte entworfen, die mir eine genaue Funktionsbestimmung von Gender als Technologie ermöglichen.

Zuerst habe ich anhand der Konzeption des Standards von Bowker und Star herausgearbeitet, inwiefern die gewählten Genderkategorien innerhalb des Gendershades Projektes statistische Analysen durch Genderclassiferyer ermöglichen (vgl. Bowker, Star 2000, S.13). Gender ist als Standard in das Gendershades Projekt eingelassen und ermöglicht eine Kompatibilität des Genderbenchmarks mit den zu untersuchenden Genderclassiferyer. Ich habe ferner in einem zweiten Schritt aufgezeigt, inwiefern diese Funktion politikwissenschaftlich zu interpretieren sein könnte und habe mich auf eine transfeministische Perspektive bezogen, die Standard als staatliche Genderegulierung begreift. Mittels queerer Surveillance Studies habe ich dann aufgezeigt, dass Staaten im 21. Jahrhundert vermehrt auf automatische Gesichtserkennung und Genderclassiferyer setzen, sodass eine Verbindung zwi-

schen Staat, Genderclasiyfer und Gender hergestellt werden konnte (vgl. Danneberg 2019; StrykerAizura 2013). Darauf bezugnehmend habe ich den Workshop Drag vs. Ai vorgestellt, als Möglichkeit innerhalb des kategorialen Genderstandards subversiv zu operieren und diesen wieder auf Butler Konzept des Drag als Subversion der Gendernorm bezogen (vgl. Butler 1991, S.202).

Zweitens habe ich eine weitere Funktion von Gender herausgearbeitet. Meine theoretischen Ideen bezogen sich auf das Konzept des Protokollfemismus von Michelle Murphy und Hellen Hester, sowie eine posthumanistische Erweiterungsmittels der Idee des Apparates von Karen Barad (vgl. Barad 2007, Hester 2018, Murphy 2012).

Das Protokoll beschreibt dabei eine Form der politischen Machtausübung (vgl. Galloway, Thacker 2014, S.290). Ich habe herausgearbeitet, dass Gender innerhalb dieser theoretischen Betrachtungsweise das Gendershadesprojekt ermöglicht, indem es Regeln für den Umgang mit und das Arbeiten an Genderclassiysern ermöglicht. In einer politikwissenschaftlichen Interpretation, habe ich die spezifische politische Ebene des Genderprotokolls nach Hester Idee der Mesopolitik analysiert (vgl. Hester 2018, S.107). Ich habe als Implikation dieser Interpretation eine Skizze zur Erstellung weiterer Benchmarks als queer inklusiv entworfen und dabei ein Hauptaugenmerk auf die Praxis queerer Selbstbezeichnungspraxen gelegt (vgl. Halberstam 2018, S.10).

Gender ist in diesem Kapitel als Technik eingebettet in das Gendershadesprojekt. Es erfüllt eine kategorial standardisierende Funktion und dient als Protokoll.

Durch diese Sichtweise lassen sich Handlungsanweisungen gewinnen, die eine Inklusion von trans Personen und queeren Menschen gewährleisten können.

Ich möchte in einem nächsten Schritt zu einem Fazit der gesamten Arbeit kommen und die bisher gefundenen Ergebnisse diskutieren.

C.) Arbeitsteilung und Gender im Gender Shades Projekt

Ich möchte mich im folgendem Kapitel mit der Arbeitsteilung innerhalb des Gender Shades Projektes beschäftigen und meine Forschungsfrage anhand dieses Fokus versuchen zu beantworten. Arbeitsteilung ist dabei eine wichtige Analyseebene für die Arbeit mit und nur um Künstliche Intelligenz, aus zwei Gründen. Erstens ist Künstliche Intelligenz, technisch betrachtet, ein arbeitsteiliger Prozess. KI besteht aus verschiedenen Programmen, benötigt vorbereitete Daten und muss dabei immer wieder evaluiert werden, um eine schadhafte Anwendung zu vermeiden und oder Computerleistung einzusparen oder auf Datenschutzverordnungen reagieren zu können (vgl. Russell, Norvig 2012, S.808ff, Russell 2019, S. 13ff). Dies ist auch einer der Gründe warum ich Buolamwinis Arbeit als Arbeit an Künstlicher Intelligenz verstehe, da sie diese evaluiert.

Zweitens beruht Künstliche Intelligenz dabei auf Arbeitsteilung als sozialwissenschaftlichem Konzept, das sich entlang von Ungleichheitsbedingungen und globalen Positionierungen entwickelt und fortbesteht (vgl. Butollo 2020, Amrute 2020). Diese Betrachtungsweise ist auch in meiner Theoriematrix zu finden, so analysieren Cheney-Lippold und Haraway diese Technologie anhand materialistischer Perspektiven. Haraway geht dabei von einer feministisch sozialistischen Theorien aus, die ihr eine Analyse der Herstellung von KI anhand geschlechtlicher Arbeitsteilung erlauben, genauso wie Halberstam der die Ausbreitung des Computers als Massenproduktionsware für den Grundstein für KI interpretiert. (vgl. Haraway 1995a, S. 54, Halberstam 1991, S.441, Cheney-Lippold 2017).

Diese beiden Bedeutungen von Arbeitsteilung sind miteinander verbunden und dienen hier zur analytischen Unterscheidung. Heutige genutzte Künstliche Intelligenz beruht auf dem Wunsch, „reale Probleme“ zu lösen, beispielsweise automatisierte Programme zur Preisermittlung von gebrauchten Autos, Wettervorhersage-Apps und vieles mehr (vgl. Alpaydin 216, S. 39-40). Dafür braucht es Entscheidungen, die Lösungen für diese Probleme anbieten. Diese sind komplex und beinhalten nicht nur Trainingsdaten, mit denen Software lernt, sondern Entscheidungen, die eine unendliche Zahl an externen Faktoren mit berücksichtigt. Dabei muss KI immer wieder auf menschliche Arbeit zurückgeführt werden, um Entscheidungen zu verifizieren bzw. Entscheidungslücken, für die Programme keine Lösungen finden, zu füllen (vgl. Suri, Gray 2019, S. xii).

Schon in den 90iger Jahren wurde dabei von Wissenschaftler:innen auf die besondere Produktionsstätte von Künstlicher Intelligenz, die Universität, hingewiesen und hierbei vor allem auf (a) eine Verunsichtbarung vieler Tätigkeiten rund um KI und (b) die Exklusion von Frauen. So wies die Anthropologin Diana Forsythe in einer der ersten ethnografischen Studien zu KI-Forschung darauf hin, dass als Arbeit mit KI nur das Programmieren, aber beispielsweise nicht der Austausch mit Kolleg:innen bei einer Teamsitzung, von den beteiligten Akteur:innen benannt wurde (vgl. Forsythe 2001, S. 33-34). Dabei machten Forscher:innen wie Lucy Suchman und Alison Adams immer wieder auf den spezifischen Ausschluss von Frauen in diesen universitären Kontexten aufmerksam (Adam 1998, S.45-46; Suchman 2007, S.200-2005).

Deshalb möchte ich in diesem Kapitel, ausgehend von der Prämisse, dass Arbeit an und mit KI vergeschlechtlicht und, wie ich gleich erläutern werde, rassifiziert ist, untersuchen, wie Gender sich an Arbeit im Gender Shades Projekt durch Buolamwini und Gebru organisiert.

Ich werde dafür zuerst eine Definition von Arbeitsteilung in Bezug auf Künstliche Intelligenz einführen und Schwerpunkte in Bezug auf Daten getriebende Künstliche Intelligenz und vergeschlechtlichte und rassifizierte Arbeitsteilung der Forschung vorstellen und diese in einem

anschliessenden Schritt auf die Arbeit von Buolammini und Gebru anwenden.

In einem letzten Schritt möchte ich die Ergebnisse diskutieren und sie kritisch einordnen. Ich möchte jedoch, bevor ich dies tue, eine Einschränkung vorwegnehmen. Meine Studie wird sich ausschließlich mit der Frage der Verteilung von Arbeit entlang der Kategorien Gender und Race beschäftigen. Ich werde mich nicht mit Fragen der vergeschlechtlichten Subjektivierung durch verschiedene Techniken beschäftigen. Ich analysiere in dieser Arbeit die wissenschaftlichen Arbeiten von Buolamwini und Gebru. Anhand dieser lassen sich nicht ausreichende Erkenntnisse bezüglich einer Subjektivierung durch Arbeit ablesen.

1.) Definition

Arbeitsteilung ist ein weitreichendes Konzept der Sozialwissenschaften, insbesondere der Politikwissenschaft (vgl. Schäfers 2019). Ich möchte hier eine Definition von Arbeitsteilung als Konzept anbieten, die der begrifflichen Abgrenzung zu der vorgestellten geteilten Arbeit als technische Voraussetzung bei Künstlicher Intelligenz dient.

Arbeitsteilung verstehe ich, ausgehend von Shoshana Zuboff, die dieses Konzept auf das Arbeiten mit neuen digitalen Technologien überführt, als Wissensteilung. Für Zuboff ist Arbeitsteilung ein Konzept, welches, ausgehend von Dirckheim und Marx, den Wechsel einer feudalen zur fordistischen²⁷ Gesellschaft markiert und nicht mehr die Idee der Warenproduktion, sondern die gesellschaftliche Ordnung an sich gestaltet (vgl. Zuboff 2018, S.50-51).

Für Zuboff befinden wir uns wiederum in einer neuen Umbruchphase von einer postfordistischen Gesellschaftsrodung in einen Überwachungskapitalismus, welcher nach Zuboff nicht mehr durch Arbeitsteilung, sondern durch Wissensteilung markiert ist.

„Die beispiellose Konzentration von Wissen sorgt für eine nicht weniger beispiellose Konzentration von Macht – für Asymmetrien, die als unautorisierte Privatisierung der Wissensteilung in der Gesellschaft zu verstehen sind. Das bedeutet, dass mächtige private Interessen das maßgebliche gesellschaftliche Ordnungsprinzip unsere Zeit kontrollieren“ (vgl. ebd. 225).

Zuboff begreift hierbei Arbeit nunmehr als ein Wissen über neue Technologien, voranging des Codierens. Nach Zuboff unterteilt sich die Gesellschaft innerhalb des globalen Nordens voranging durch das Wissen über das Programmieren und das Nichtwissen der Anderen (hier ein bewusst breit gewählter Ausspruch, der eine Vielzahl an Personen erfassen soll) (vgl. ebd.).

27 Fordismus bezeichnet dabei eine kapitalistische Wirtschaftsform mit standardisierten Beschäftigungsverhältnissen (Beispiel der Autofabrikarbeiter), angebotsorientierter Massenproduktion und staatlicher Anteilseignung an Konzernen sowie sozialer Absicherung der Arbeiter:innen durch langjährige Beschäftigungsverhältnisse und wohlfahrtsstaatliche Sicherungsnetze (vgl. Michalitsch 2004, S.147-153)

Diese werden einerseits durch Daten getriebene KI ausgebeutet, indem ihre Daten kostenlos von Firmen erlangt und diese dann wiederum in Gewinn umgewandelt werden, und andererseits durch eine Teilung der Arbeit in Gutverdiener (wenige Programmierer:innen) und Menschen die diesen zuarbeiten (vgl. ebd. S. 60ff).

Anders als Zuboff möchte ich jedoch weiterhin den Begriff Arbeitsteilung verwenden, da Menschen weiterhin Arbeiten für das Funktionieren neuer smarterer Technologien leisten und das oftmals unter prekären Bedingungen, wie Scheinselbstständigkeiten, fehlendem Versicherungsschutz und unsicheren Arbeitsaufträgen (vgl. Gray, Suri 2019). Dabei wird auf vergeschlechtlichte Arbeitsteilung zurückgegriffen, also die Idee, dass Arbeit nicht nur aufgrund von Funktionen geteilt wird, sondern auch aufgrund von Geschlechternormen. Dies untersuchen Wissenschaftler:innen in Bezug auf Computer getriebene Automatisierungsprozesse²⁸ seit den 80ern (vgl. Haug et al. 1987, Spertus 1991, Misa et al 2010). Zudem weisen ältere Studien, die sich explizit mit vergeschlechtlichten Arbeitsweisen innerhalb von Künstlicher Intelligenz befassen, auf die Exklusion von Frauen innerhalb des akademischen Feldes hin. So zeigen die Anthropologin Diana Forsythe und die Medientheoretikerin Gertraud Koch für Deutschland und die USA auf, wie Frauen durch sexistische Arbeitsumfelder aus den jeweiligen Arbeitskontexten gedrängt wurden (Forsythe 2002, Koch 2005). So erklärt Forsythe die Exklusion von Frauen in akademischen Laboren, die zu KI forschen, anhand der Entwertung ihrer Abschlüsse gegenüber und durch männliche Kollegen, sowie die Konstruktion von Frauen „als die Anderen“ durch männlich besetzte Forschungsgruppen und das Verunsichtbaren bestimmter weiblich konnotierter Arbeiten wie Sekretariatstätigkeiten (vgl. Forsythe 2002, S. 169-173).

In diesem Zusammenhang führt Sarah Myers West in ihrer Studie zur Fairness und KI, einer Studie des MIT aus den 80iger Jahren, an, dass

„Many women are treated as if they were invisible in technical situations ... [O]verlooked in technical discussions and excluded from group efforts“ (MIT 1983, S. 9 nach Sarah Myers West 2020, S. 11).

Besonders interessant an diesem Umstand ist die direkte Verbindung zum Gender Shades Projekt, da dieses knapp 40 Jahre später an jener Universität durchgeführt wurde (vgl. <http://gendershades.org/> zuletzt aufgerufen am 14.12.2020).

Dabei stellt dieser Fakt keine kausale Verbindung da, dennoch bleibt festzuhalten das Künstliche Intelligenz und Geschlechterdiskriminierung innerhalb akademischer Strukturen ein Jahrzehnte

28 Ich beziehe hier die frühen Automatisierungsstudien von Frigga Haug und Gerhard Brosius in Bezug auf KI mit ein, da diese schon Tendenzen der vergeschlechtlichten Arbeitsteilung aufwiesen, die heute in Bezug auf moderne Ansätze Künstlicher Intelligenz zu beobachten sind: Weiblichkeit als Marker unterbezahlter administrativer Arbeiten und Männlichkeit als Marker gut bezahlter Arbeit an den Maschinen/Programmen der Maschinen (vgl. Hauget.al. 1987, S. 50-53)

altes Problem darstellt, das sich in den Zahlen, die D'Ignazio und Klein bezüglich der geringen Frauenquote in Tech-Unternehmen wie Facebook und Google wiederzufinden scheinen, als korrelative Beziehung ausdrückt (vgl. Klein, D'Ignazio 2020, S.27).

Ich möchte in einem nächsten Schritt weg von dieser historischen Basis, hin zu neuen Entwicklungen im Bereich Arbeitsteilung und Künstliche Intelligenz.

a.) Arbeitsteilung und Themengebiete

Dabei beziehe ich mich voranging, aufgrund der von mir gewählten Untersuchungsgegenstände, auf die Arbeitsteilung um Daten getriebene KI-Ansätze, mit denen sich das Gender Shades Projekt in der Form von Genderclassifyern auseinandersetzt.

Innerhalb neuerer Forschung der letzten Jahre gibt es drei große Teilgebiete: (a) Die Frage der Herstellung der verwendeten Hardware und deren Wiederverwendung, (b) die Unterteilung von prekariisierter und elitärer Arbeit entlang der Unterteilung Daten bearbeiten und Programme bzw. Software schreiben und (c) die Verunsichtbarung bzw. der Wegbruch der Arbeit durch neue smarte Technologien (vgl. Frey 2019, Schmidt 2018, Waycman 2019).

i.) Hardware und Arbeitsteilung

Haraway wies bereits in den frühen 1990igern auf die Ressourcenintensivität von Computer-Technologie und deren paradoxe Erscheinung als „durchsichtig“ hin (vgl. Haraway 1995a, S.39). Dabei erklärte sie, dass für die verwendete Hardware, auf der die neuen Programme und vor allem Künstliche Intelligenz zu laufen habe, aus dem globalen Süden kämen, hergestellt durch Women of Color unter ausbeuterischen Arbeitsverhältnissen (vgl. ebd. S.54-5).

In ihrer Studie zum Lebenszyklus moderner smarterer Endgeräte arbeitete die in Cambridge lehrende Medienwissenschaftlerin Jennifer Gabrys heraus, dass neue Technologien heute immer mehr auf eine Einwegidee von Hardware bauen. Einmal verwendet, schon sind moderne Handys, Computer und Tablets nicht mehr tragbar für eine weitere, über Jahre anhaltende Verwendung, da sie nicht mehr mit den neuesten technischen Entwicklungen im Bereich der Software kompatibel wären (vgl. Gabrys, 2011, S.2). Dabei sind gerade die Weiterentwicklungen der Hardware dafür verantwortlich, dass Technologien, wie Künstliche Intelligenz und somit Genderclassifier verwendet werden können (vgl. Russel, S. 68-71, Sejnowski 2018).

Dabei verweist Gabrys auf die globale Verwertungskette und darauf, dass vorrangig der globale Süden diese Produkte recycelt und aus ihnen neue Ressourcen für eine nächste Generation smarterer Endgeräte herstellt, und das unter prekären und gefährlichen Bedingungen, die das Leben der

Arbeiter:innen gefährden und sie zu Niedriglöhnen ausbeuten (vgl. Gabrys 2011, S.29).

Haraway untersucht hierbei die, vor allem durch Frauen, ausgeführte Fabrik­tätigkeit des Zusammensetzens der immer kleiner werdenden technischen Hardware, voranging der verwendeten Mikrochips und den dahinter liegenden rassifizierten und vergeschlechtlichten Ideen. Sie verdeutlicht dabei die These, dass diese Arbeit, aufgrund ihrer Miniaturisierung, an Frauen aus dem globalen Süden ausgelagert wird (vgl. Haraway 1995a, S.55-57).

Ich möchte hierbei vorwegnehmen, dass zwar Boulamwinis Hardware auf dieser Arbeitsteilung beruhen könnte, da die von Gabrys angeführte Studie aber auch beispielsweise die Zahlen der Umweltbelastung hinsichtlich der Computer-Hardware von Klein und D'Ignazio auf ein systemischen Nutzen beziehungsweise das Ausnutzen der natürlichen und menschlichen Ressourcen verweisen, ich diese jedoch nicht weiter untersuchen kann.

Ich habe keine Interviews geführt und kann nicht wissenschaftlich darlegen, inwiefern welche Geräte von Buolamwini benutzt wurden. Deshalb werde ich den hier angeführten Strang der Idee der vergeschlechtlichten Arbeitsteilung entlang der Hardware, die vielleicht im Gender Shades Projekt verwendet wurde, nicht weiterverfolgen.

ii.) *Prekäre und elitäre Arbeit mit und an KI*

Wie ich bereits erläutert habe, zeichnet sich KI durch eine technische Arbeitstrennung aus, die sich, entlang der hier vorgestellten Idee von KI als Daten getriebener Ansatz, entlang der Linie von Daten, aufbereitet für Künstliche Intelligenz und das Schreiben von lernenden Algorithmen, abzeichnet (vgl. Alpaydin 2016, S. 14-17).

Der Soziologe Florian Schmidt erläutert in seiner Studie zur Arbeit an Bilddaten für das autonome Fahren²⁹. Dabei, so Schmidt, arbeiten sogenannte Crowdworker:innen auf Online-Plattformen, wie (Amazons Plattform zur Vergabe digitaler Kleinstaufgaben, wie dem Beschriften von Bildern oder Untertiteln von Filmsequenzen). Sie werden als selbständige Arbeiter:innen für zumeist Mikroarbeitsaufträge bezahlt, die sich in Minuten erledigen und pro erledigtem Auftrag in Cent-Beträgen auszahlen lassen (vgl. Schmidt 2018, S. 7).

Gerade bei der Bilderkennung und dem Vorsortieren der Bilder, also das Beschriften Beschriften der Bilder, analog zur Kategorisierung der Bilder für das Datenset bei Buolamwini und Gebru, sind gerade im universitären Umfeld Aufgaben, die an solche Plattformen weitergeleitet werden. So beschreiben die Anthropolog:innen Mary Gray und Siddharth Suri in ihrer Studie zur Crowd-Arbeit,

29 Autonomes Fahren meint: The algorithmic architecture of self-driving-car programming begins with 'if-then-else' rules, which define actions under certain conditions. The car senses and classifies the world around it before probabilistically making choices based upon what it has learnt. (Stilgoe 2018, S. 29)

dass vor allem universitäre Studien zur Bilderkennung auf genau jene Ressourcen zurückgreifen, um weniger Menschen in festen, an der Universität angegliederten Arbeitsverhältnissen zu beschäftigen. Hierbei können Universitäten bzw. Wissenschaftler:innen einerseits den enormen Arbeitsaufwand, große Studien mit mehreren Tausend zu klassifizierenden Bildern zu erstellen, in kurzer Zeit erledigen und gleichzeitig diese Arbeit sehr kostengünstig durchführen lassen (Gary, Suri 2019, S. xvii). Dabei zeichnen die beiden Autor:innen hinsichtlich der Geschlechteraufteilung dieser Arbeit ein indifferentes Bild. Es sei beispielsweise keine rein auf weibliche Arbeit basierende Sparte oder eine mit einer Weiblichkeit assoziierte Arbeit. Vielmehr, so die Autor:innen, lassen sich geschlechtliche Unterschiede in der Begründung der Wahl der Arbeit feststellen und diese unterscheiden sich noch einmal stark hinsichtlich globaler Verortung. Gray und Suri stellen jedoch heraus, dass die Beschäftigten solcher Plattformen hochgebildet sind und rund 80 % der Befragten ihres untersuchten Panels mindestens einen Bachelor-Abschluss hatte (vgl. ebd. S. 29). Die Forscher:innen fanden zudem heraus, dass Frauen, die solche Aufgaben annehmen, vor allem in den USA dabei unter anderem den Vorteil des von zuhause Arbeitens sehen, bei gleichzeitiger Pflegeverantwortung. Frauen aus Indien, die sie befragten, sahen ihre Arbeit als selbständige Arbeit und Emanzipationsbestrebung an, hinsichtlich einer Tätigkeit, fernab der Sorgearbeit innerhalb familiärer Verbünde (vgl. ebd.). Im Gegenzug gibt es jedoch eine klare Geschlechteraufteilung hinsichtlich gutbezahlter Arbeitsplätze in der Entwicklung von intelligenter Software (D'Ignazio, Klein 2020, S.29).

Firmen, die mit Waren und Dienstleistungen in diesem Bereich arbeiten, sind für massive Gewinne verantwortlich. So erwirtschaftete Microsoft, die auch eine der untersuchten Genderclassifier stellen, während der Finanzkrise 2008 einen Kassenbestand von 18 Milliarden Euro (vgl. Schiller 2010, S.141). Die vier wertvollsten Unternehmen sind 2020 reine Tech-Unternehmen und mit Google, Amazon und Microsoft Unternehmen, die sich auf Künstliche Intelligenz spezialisieren (vgl. <https://de.statista.com/statistik/daten/studie/162524/umfrage/markenwert-der-wertvollstenunternehmen-weltweit/> zu letzt geprüft am 05.03.2020). Dabei stellen diese Unternehmen relativ wenige feste Mitarbeiter:innen ein, welche gut bezahlt werden. Diese sind zumeist weiße Männer (vgl. D'Ignazio, Klein 2020)

Der Journalistik-Professor Douglas Rushkoff beschreibt den Zusammenhang zwischen den prekären Arbeitsverhältnissen und den extremen Gewinnen und Gehältern der Mitarbeiter:innen dieser Konzerne als neue Form eines Wirtschaftssystems, dass auf maschinellen Handlungen und dem menschlichen Vertrauen in seinen Wert als wirtschaftliche Ware baut. Dabei, so Rushkoff, folgt diese Entwicklung keiner böswilligen Absicht einzelner Akteur:innen, sondern bettet sich in diese Struktur ein, die Rushkoff digitalen Kapitalismus nennt (vgl. Rushkoff 2016, S. xiii).

Ähnliche Analysen dieses Phänomens sind derzeit in wissenschaftlichen Diskursen zu Arbeit und Künstlicher Intelligenz zu finden - so spricht die posthumanistische Theoretikerin Rosi Braidotti von advanced capitalism - indem sie die Ungleichzeitigkeit des freien digitalisierten Marktes und die Freiheit der Manager:innen den strengen Grenzregimen für arme Menschen, Menschen of Color und Frauen entgegensetzen (vgl. Braidotti 2019, 29-30).

Diese neuen Kapitalismen sind jedoch weiterhin eingebunden in eine vergeschlechtlichte Trennung, in welcher weiße US amerikanische Cismänner, statistisch gesehen, Geld erwirtschaften und eine große und stetig wachsende Zahl an Menschen prekär beschäftigt ist (vgl. West 2020, Cherry 2016, S. 71).

In einem dritten Punkt möchte ich nun erläutern, weshalb diese Trennung speziell BIPOCs trifft.

iii.) Revolution und das Verschwinden der (weißen, westlichen) Arbeit.

Einen letzten wichtigen Strang der Literatur zu Arbeit und Künstlicher Intelligenz möchte ich an den vorangegangenen Absatz anschließen, und zwar nicht nur Prekarisierung von Arbeit, sondern auch die Verunsichtbarmachung von Arbeit und deren Entkontextualisierung.

Ausgehend von der These der Revolution der Arbeit, gibt es zum jetzigen Zeitpunkt unzählige Literatur zum Thema Zukunft der Arbeit durch und mit KI. Im Vordergrund steht dabei die zentrale Idee, dass, durch die Einführung smarter Technologie, menschliche Arbeit obsolet wird (dritte Revolution der Arbeit). Einerseits wird dies als positive Entwicklung aufgefasst, indem Menschen von leidiger und unbefriedigender Arbeit befreit werden. Die andere Interpretation befasst sich stark mit der Vorstellung der Massenarbeitslosigkeit und Armut jener dann lohnarbeitslosen Menschen (vgl. Waycman 2019; Wilson, Daugherty 2018, Davenport 2018). Dabei ist die Prämisse immer, dass die Arbeit für den Menschen verschwindet (vgl. ebd.).

Feministische und postkoloniale Wissenschaftler:innen haben dabei herausgearbeitet, dass diese Idee nur bedingt auf die jetzige Arbeitssituation zutrifft (vgl. Cherry 2016, S.71, Kutzner, Roski 2019, 364-5; Mavhunga 2017, S.4-6).

In ihrer Studie zur Arbeit in dieser dritten Revolution der Arbeit schreiben die Soziolog:innen Neda Atanasoski und Kalinda Vora über die Auswirkungen des Narrativs der dritten Revolution der Arbeit auf BIPOC und Frauen. So analysieren sie den Zusammenhang vom Ende der menschlichen Arbeit als Privileg weißer Mittelschichtmänner, die vorrangig als Mensch gezählt werden und deren Arbeit entlohnt wurde und als solche betrauert bzw. verändert wird (Atanasoski, Vora 2015, S.9). Demgegenüber würden die Ideen einer arbeitsfreien Gesellschaft immer auf dem Erbe kolonialer Ausbeutungsverhältnissen beruhen, da diese nicht berührt würden von den neuen smarten

Technologien, und durch ihre oftmals nicht institutionalisierte Struktur (nicht angemeldete Tätigkeiten, keine gewerkschaftliche Vertretung und Ähnliches) nicht als Arbeit gewertet werden, beziehungsweise die Menschen, die diese Arbeit verrichten, nicht als Menschen und Arbeiter:innen in dem Diskurs um die neue Art der Arbeit vorkommen (vgl. ebd. 11). Vielmehr würden die bisher prekären Ausformungen der Care-Arbeit, wie Haushälter:innen, Kellner:innen, Kinder“Mädchen“, nach wie vor bestehen bleiben als Arbeit, die nur mittels smarterer Technologie gemanagt würde und auch entpersonalisiert. So werden Arbeiter:innen, die auf Mturk ihre Dienste anbieten, nicht mit Namen sondern persönlichen Codes betitelt. Für die Wissenschaftler:innen ergibt sich daraus ein nicht aufzulösendes Paradoxon:

„intertemporal imaginaries of freedom from work points to how contemporary platforms also engender a different time-space that virtualizes the racialized and gendered materiality of service work through a dynamics of connectivity and the ideals of the sharing economy. These are liberal fantasies that seek to include racialized and gendered bodies into freedom (women are now professionals, 3D printing can help the global South) while invisibilizing the racialized and gendered architectures of their revolutionary claims to remaking the human.“ (vgl. ebd. S.30)

Wichtig sind dabei die Verbindungen von Race und Gender als Strukturen sozialer Ungleichheit. Die Autor:innen stellen vor allem den Zusammenhang vergeschlechtlichter und rassifizierter Carearbeit im globalen Norden heraus. Frauen of Color würden dabei unverändert die Arbeit übernehmen, die ihnen seit der Kolonisation durch die westlichen Gesellschaften aufgebürdet wurde. Die Revolution der Arbeit ist somit keine Revolution, da sie gesellschaftliche Machtverhältnisse beibehält und diese durch die Verunsichtbarmachung der Plattform-Infrastrukturen, wie etwa Mturk verschleiert. Gleichzeitig revolutioniert sie bestehende Verhältnisse für eine exklusive Gruppe Menschen, ohne diese Exklusion zu thematisieren (vgl. ebd. S. 28)

Zusammenfassend möchte ich festhalten, die Arbeitsteilung rund um Künstliche Intelligenz ist vergeschlechtlicht organisiert. Es gibt seit den 80er Jahren Forschungen, die auf den Zusammenhang gerade gut bezahlter Programmierjobs und der Exklusion von Frauen hinweisen (vgl. Chun 2013, S.29-35, Light 1999). Mittlerweile sind dabei postkoloniale Auseinandersetzungen mit Race und Gender als Kategorien der Arbeitsteilung rund um Künstliche Intelligenz in die wissenschaftliche Öffentlichkeit gelangt (vgl. Arora 2016, Couldry, Mejias 2019, Amrute 2020), Dabei spielen vor allem Umbrüche in der Gestaltung der Arbeitsverteilung (Plattformen) und der hohen Gewinnerzielung durch Tech-Unternehmen eine große Rolle. Die vorgestellten Studien berufen sich hierbei auf eine neue Form des Kapitalismus der konstitutiv mit den neuen Technologien verbunden ist, sei es durch die Ausbeutung bestehender sozialer Daten von Social Media, dem Nutzen prekärer Arbeit um Künstliche Intelligenz zu verbessern oder aber durch die

unsichtbare prekäre, rassifizierte vergeschlechtlichte Arbeitsverteilung (vgl. Zuboff 2018, Waycman 2019, Rushkoff 2016). Ich möchte im anschließend Teil diese Ergebnisse auf das Gender Shades Projekt übertragen. Ich werde dabei die materiellen Grundvoraussetzungen aufgrund der von mir gewählten Untersuchungsmaterialien, wie bereits erläutert, auslassen.

2.) Empirie

Ausgehend von den vorangegangenen Analysenergebnissen, werde ich das Gender Shades Projekt anhand dreier Schwerpunkte untersuchen. Zuerst werde ich darstellen (a) wer das Gender Shades Projekt erarbeitet hat. Dabei habe ich im vorangegangenen Teil dargestellt, dass vor allem Frauen in dem Bereich der Datenbereinigung und -bearbeitung tätig seien und nicht als Programmierer:innen auftreten (vgl. D'Ignazio, Klein 2020).

In einem zweiten Punkt möchte ich genauer herausarbeiten, (b) was genau bearbeitet wurde innerhalb des Gender Shades Projekts und wie sich dies zu der These der Exklusion von Frauen innerhalb der wissenschaftlichen Arbeit zu KI verhält. In einem dritten Schritt werde ich (c) auf den genauen Umgang der Projektbeteiligten mit den verarbeiteten Daten eingehen und diesen analysieren. Hierbei werde ich mich vor allem auf den innerhalb meiner Definition genannten Bereich der Verunsichtbarung der Arbeit marginalisierter Gruppen beziehen.

a.) Die besondere Stellung Schwarzer KI Wissenschaftlerinnen

Das Gender Shades Projekt wurde von Joy Buolamwini und Timnit Gebru in Leben gerufen, beides Schwarze Frauen. Buolamwini ist MIT-Absolventin, eine der anerkanntesten Universitäten für die Forschung zu Künstlicher Intelligenz, und Timnit Gebru frühere leitende Mitarbeiter:in in Googles Firmensparte Ethik und KI (vgl. Buolamwini 2017, Buolamwini, Gebru 2018 S.1). Erst vor wenigen Wochen wurde jedoch bekannt, dass Timnit Gebru aufgrund kritischer Auseinandersetzung mit Googles Produkten zur Spracherkennung fristlos entlassen wurde (vgl. <https://www.nytimes.com/2020/12/03/technology/google-researcher-timnit-gebru.html> zuletzt abgerufen am 15.12.2020). Dies hatte ein breites Medienecho zur Folge, durch das noch einmal, für ein weltweites Publikum, auf Gebrus herausragende Stellung als schwarze Frau und eine der führenden Expert:innen im Bereich Künstliche Intelligenz und Ethik hingewiesen wurde (vgl. <https://www.bbc.com/news/technology-55281862>, <https://googlewalkout.medium.com/standing-with-dr-timnit-gebru-isupporttimnitbelieveblackwomen-6dad300d382> zuletzt abgerufen am 15.12.2020).

Eine besondere Aufmerksamkeit verdankt das Gender Shades Projekt hierbei den beiden Forscher:innen als Expert:innen in einem vorwiegend männlichen, weißen Arbeitsfeld und der daraus resultierenden Erstellung des ersten intersektionalen Benchmarks für Genderclassifier (vgl. Nkonde 2019, S.30, Buolamiwini 2017, S.17-18).

Entgegen der vorgestellten Befunde der Exklusion von Minderheiten kann das Gender Shades Projekt zuerst einmal als Gegenbeweis erscheinen. Es wird jedoch in Zusammenhang mit der wissenschaftlichen und medialen Rezeption zu einem Moment der Sichtbarmachung weißer, männlicher Privilegien, die verdeutlicht wie stark Schwarze Wissenschaftler:innen bis ins Jahr 2020 hinein Ausnahmerecheinungen der akademischen Welt sind (vgl. D'Ignazio, Klein 2020, S. 27; Benjamin 2019, S. 77-79).

b.) Elitäre Wissensarbeit

Dies lässt mich überleiten zu der Frage, was Gebru und Buolamwini, als Wissenschaftler:innen erarbeitet haben. Sie entwickelten, wie schon beschrieben, eine Benchmark, also eine Werkzeug zur Evaluation von Genderclassifiern. Sie haben dafür Bilder zusammengestellt und diese technisch aufgearbeitet, indem sie sie kategorisiert haben. Ferner haben sie diese Benchmark angesetzt, die drei untersuchten Genderclassifier zu testen, und haben anschließend diese Ergebnisse statistisch ausgewertet (vgl. Buolamwini, Gebru 2018, S 12).

Dabei haben die beiden Wissenschaftler:innen sowohl einen politischen als auch wissenschaftlichen Diskurs über die Produktion von Benchmarks angestoßen.

In dem Fall des Gender Shades Projektes haben sie so maßgeblich zu der Arbeit an Künstliche Intelligenz beigetragen und zwar im externen Entwicklungsprozess von Genderclassifiern (vgl. <http://gendershades.org/overview.html> zuletzt abgerufen am 15.12. 2020). Dabei haben sie ferner einen politischen Diskurs angeregt, welcher sich unter anderem in Hearings des US Kongress ausdrückte, die sich mit Gesichtserkennung und deren Einsatz durch staatliche Behörden auseinandersetzte (vgl. D'Ignazio, Klein 2020, S. 32; Federal Trade Commission 14 11 2018, S. 166ff). Im Hinblick auf die wissenschaftliche Einflussnahme des Gender Shades Projektes können beispielsweise die darauf folgenden Studien zu Gesichtserkennung entlang der Achsen Race und Gender als Folgestudiengelten (vgl. Raji et. al. 2020, Gebru 2020, Hooker et al. 2020). Wie ich bereits i vorangegangenen Teil erläutert habe, handeln Boulamwini und Gebru hierbei als Wissenschaftler:innen und Aktivist:innen. Die Forschenden entwickeln eignen Konzepte und Handlungsanweisungen für das Handeln mit KI. Sie nicht nur Betroffene oder aber prekariisierte Arbeiter:innen, welchen die Sichtbarkeit fehlt, sondern leisten renommierte Wissensarbeit (vgl

D'Ignazio, Klein 2020, S. 32).

Dabei stellen sie die von mir vorgestellten Ergebnisse von Atanasoski, und Vora nicht in Frage, sondern eröffnen eine Blick auf eine widerständige Praxis, indem Schwarze Frauen nicht nur prekäre Care-Arbeit in anonymisierten Plattformstrukturen verrichten, sondern Technik gestalten und diese Ergebnisse prominent vertreten (vgl. Atanasoski, Vora 2015). In einem letzten Abschnitt möchte ich auf diese Besonderheit in Bezug auf den konkreten Umgang mit den von Gebru und Buolamwini erhobenen Daten eingehen und erläutern, dass das Gender Shades Projekt auch eine neue Blickweise auf die Trennung zwischen exklusiver, elitärer Wissensarbeit und der vorrangig prekarierten Arbeit in der Datenverarbeitung ermöglicht.

c.) Der Umgang mit Daten

Wie ich bereits in der Begründung zur Fallauswahl dargelegt habe, sind der Umgang und die Gewinnung der Daten von Buolamwini und Gebru ausschlaggebend für mich gewesen, diese Studie zu verwenden (vgl. Buolamwini, Gebru 2018, S. 6). In diesem Teil möchte ich auf eine zweite Besonderheit eingehen, nämlich die Arbeitsteilung entlang dieser Datenerhebung.

Wie dargestellt, ist die Kategorisieren von Daten, wie sie Buolamwini und Gebru betreiben, eine Tätigkeit, die sowohl im akademischen Bereich als auch in wirtschaftlichen Bereich ausgelagert wird an Plattformen, die das Kategorisieren von Bildern in einzelne Arbeitsaufträge untergliedern (vgl. Gray, Suri 2019 S. 6-7), oftmals mit einer globalen Komponente, bei der dieses Kategorisieren, aufgrund geringerer Lohnzahlung von Seiten der Auftraggeber:innen, an Länder des globalen Südens geht (vgl. Schmidt 2018, S.37).

Diese Vorgehensweise ist auch prägend für das Erstellen von Benchmarks, die speziell für das automatische Erkennen von Gesichtern ausgelegt sind.

Buolamwini und Gebru verzichten auf diese Auslagerung der Arbeit, beide kategorisieren ihre Bilder selbst, anhand von Geschlecht und Race, und legen dabei ihre Kriterien offen (vgl. Buolamwini, Gebru 2018, S. 6-7). Dabei stellt vor allem die Kategorisierung von Hautfarbe für die Wissenschaftler:innen einen Notwendigkeit dar, diese selbst vorzunehmen. Dabei sei diese Arbeit, nach den Autor:innen, ihrerseits eine sensible Aufgabe, da sie mit Hautfarben als Charakteristikum arbeiten und diese phänotypische Auswertung politische Sensibilität erfordere, da sie sich auch auf koloniale Praktiken der Eugenik beruft (vgl. ebd.; James 2008, 34-35).

Des Weiteren kategorisieren Buolamwini und Gebru nicht nur ihre Bilde selbst, sondern sie markieren ihre Bilder auch als Bilder von Gesichtern. Diese Arbeit kann technisch betrachtet eine automatisierte Aufgabe sein. Sie wird von automatischen „Face Detectoren“ übernommen, die

erkennen sollen, ob auf den jeweiligen Bildern ein Gesicht gezeigt wird oder nicht. Buolamwini und Gebru haben sich jedoch gegen diese automatische Arbeit gewendet und werten die Bilder selbst als Gesicht, da frühere Studien zu Face Detectoren zeigten, dass diese nur weiße Gesichter als Gesichter anzeigen (vgl. Benjamin 2019, S. 112-113). Gebru und Buolamwini entschieden sich deshalb bewusst für diesen Schritt der Kategorisierung, um Schwarze Gesichter in ihre Benchmark aufzunehmen, da eine automatische Erkennung von Gesichtsbildern höchstwahrscheinlich ein rein weißes Datenset für die Benchmark ergeben hätte (vgl. Buolamwini, Gebru 2018, S.3). Hier beschreiben, anders als das Narrativ der Ersetzung der menschlichendurch maschinelle Arbeit, Gebru und Buolamwini eine Gegenpraxis, die sich die Bereiche des maschinellen Erkennens wieder aneignet als menschliche und auch als antirassistische Praxis. Hier fügt das Gender Shades Projekt der These von Atanasoski und Vora einen weiteren Punkt hinzu. Gebru und Buolamwini verweisen dabei in ihrer Praxis auf die Notwendigkeit der kritischen Hinterfragung der Notwendigkeit bereits bestehender maschineller Arbeit (vgl. Atanasoski, Vora 2015, Buolamwini, Gebru 2018, S. 2-3). Hierdurch machen Buolamwini und Gebru mit ihrer wissenschaftlichen Arbeit sichtbar, wie diese Exklusion von marginalisierten Gruppen, als Betroffene der automatischen Gesichtserkennung, konkret funktioniert. Andererseits erlaubt Buolamwinis und Gebrus Arbeit aber auch eine Möglichkeit und ein Beispiel der kritischen Intervention, indem die Wissenschaftler:innen Alternativen zu maschinellern Handeln und der Verunsichtbarung von Arbeit aufzeigen, die auf wissenschaftlicher und teilweise gut bezahlter Arbeit (Gebru war zu diesem Zeitpunkt in einer führenden Position bei Google angestellt (vgl. Buolamwini, Gebru 2018, S. 1)) und eben nicht auf unsichtbarer, prekärer Care-Arbeit beruht.

3.) Polittheoretische Analyse

Wie diese Untersuchung gezeigt hat, stellt das Gender Shades Projekt eine Ausnahme hinsichtlich aktueller wissenschaftlicher Debatten um das Themengebiet Künstliche Intelligenz und Arbeitsteilung dar. Ich möchte jedoch an dieser Stelle zwei Einschränkungen meiner bisherigen Ergebnisse vornehmen: einerseits eine globale Verortung bezüglich der Rolle von Frauen innerhalb der universitären Ausbildung hinsichtlich Computer-Technologie und Informatik und zweitens die Frage der Kategorie Gender in diesem Teil der Arbeit.

a.) *Der globale Süden als Vorbild*

Ich habe geschrieben, dass Frauen als Wissenschaftler:innen im Bereich der Künstlichen Intelligenz

Ausnahmeerscheinungen sind (vgl. D'Ignazio, Klein 2020, S. 60, West 2020, S. 12). Dies betrifft jedoch nur einen spezifischen Teil der Welt. Wissenschaftliche Studien in Bezug auf den Gender Gap innerhalb der Studiengänge zur Informationstechnologie zeigen, dass im globalen Süden in verschiedenen Ländern ein relativ ausgeglichenes Verhältnis von Frauen und Männern hinsichtlich der Abschlüsse besteht (vgl. Mellström 2009, Yardi et. al. 2017, Rezai-Rashti 2014).

In Ländern des globalen Südens, wie Indien oder Malaysia, sind, hinsichtlich der Studierendenzahlen und Studienabschlüsse innerhalb des großen Forschungsbereichs, Frauen keinesfalls Ausnahmeerscheinungen. So sind in Indien 45% der Abschlüsse in „Computer Science“ Frauen zuzurechnen, in den USA sind es hingegen 21% (vgl. Yardi et. al. 2017, S. 2).

Daher möchte ich an dieser Stelle mein Argument präzisieren. Frauen sind im globalen Norden Ausnahmen, was die universitäre Bildung in Bezug auf Computertechnologie betrifft. Ferner deuten die Besetzungen gerade in den großen Technologiekonzernen, wie Google und Facebook, auf eine exklusiv westliche, männliche Einstellungspraxis hin (vgl. D'Ignazio, Klein 2020, S. 27).

Hinzu kommt, dass mittlerweile ein Großteil der Forschung zu dem Thema Künstliche Intelligenz von diesen Firmen und deren Wissenschaftler:innen kommt. Nach Shoshana Zuboff publizieren Forscher:innen der großen Tech-Unternehmen Google und Facebook vier- bis fünfmal mehr Publikationen als solche an Universitäten (vgl. Zuboff 2018, S. 222). Wissenschaftliche Erkenntnisse werden so in einem sehr homogenen Umfeld erzeugt und produziert (vgl. Daub 2020, S.18-19). Was jedoch nicht bedeutet, dass es nicht auch Wissenschaftler:innen im globalen Süden gäbe, die sich mit Fragen von Künstlicher Intelligenz und deren Entwicklung beschäftigen. Vielmehr geschieht dies weniger prominent, in Form von Journalbeiträgen oder der Einbeziehung von Wissen außerhalb eines kleinen universitären und privatwirtschaftlichen Kreises.

b.) Frau sein und Arbeitsteilung

Eine weitere wichtige Einschränkung ist, das ich vergeschlechtlichte Arbeitsteilung hier entlang der Kategorie Gender und der Frage nach der Stellung bzw. hier der Exklusion von Frauen verhandelt habe. Dies ergibt sich vor allem aus der hier von mir gewählten Perspektive, Gender als Verteilungspunkt von Arbeit zu fassen, also aus der Frage, wie oder ob sich Arbeitsteilung entlang von Gender differenzieren lässt, beispielsweise im Hinblick auf die Verteilung von Frauen innerhalb der Arbeit an Computern (vgl. Chun 2013, S. 31).

Diese Sichtweise, die ich gewählt habe, nimmt dabei nicht in den Blick, wie im Umkehrschluss Arbeitsverhältnisse Rollenerwartungen konstruieren, beispielsweise, wie sich durch die Arbeit mit Computern die Idee von weiblicher Arbeit oder Weiblich sein wandelt. Noch kann analysiert

werden, von mir, inwiefern beispielsweise diese Form der Arbeit rund um KI auch Geschlechtlichkeit hervorbringt, was etwa in den frühen Studien der Soziologin Cynthia Cockburn diskutiert wird, die auf den konstitutiven Moment von Arbeit als Moment der Gender-Differenz hinweisen. Sie zeigt auf, dass beispielsweise Männlichkeit gerade durch den Moment der scheinbaren Beherrschung der Maschine durch die jeweiligen Arbeiter ausgedrückt wird (vgl. Cockburn, 1988).

4.) Zwischenfazit

Ich möchte nach diesen Einschränkungen nun zu einem Zwischenfazit kommen. Gender weist in dem Zusammenhang von Arbeitsteilung mit Bezug auf das Gender Shades Projekt auf eine Art westliches „Best Practice“-Beispiel, welches vielen bisherigen wissenschaftlichen Kritikpunkten an der Arbeit mit KI widerspricht (vgl. Arora 2016). So weisen Boulamwini und Gebru auf Momente der automatisierten Exklusion von BIPoCs hin. Gleichzeitig stellen sie selber eine Ausnahmehinsichtlich einer homogenen wissenschaftlichen Community zu Künstlicher Intelligenz dar und leben in ihrer Arbeit ein Modell der Arbeit an Daten vor, welches nicht auf Prekariat und Outsourcing beruht, wie es die Forschung gerade in Hinblick auf Kategorisierungsaufgaben feststellt (vgl. Gray, Suri 2019, Schmidt 2018). Gender ist dabei nicht gleichzusetzen mit diesen Ergebnissen, Gender ist, anders formuliert, keine kausale Beziehung zwischen der Arbeit der Forscher:innen und ihren emanzipatorischen Bestrebungen, dafür habe ich keine Anhaltspunkte innerhalb der vorliegenden Materialien gefunden (was nicht bedeutet, das es diesen Zusammenhang nicht gibt). Folglich stellt Gender hier einen Marker für die progressive Arbeitsteilung dar, die Buolamwini und Gebru durch das Gender Shades Projekt leben. Sie entwickeln nicht nur, wie sie selber schreiben, eine der ersten intersektionalen Benchmark-Studien, sondern geben gleichzeitig ein Beispiel für eine gerechtere Verteilung von Arbeit innerhalb der Zusammenhänge von Arbeitsteilung und Daten getriebener Künstlicher Intelligenz, bleiben dabei jedoch in einem US-amerikanischen Kontext verhaftet.

IV.) Fazit

Ich werde im folgenden Teil ein detailliertes Fazit ziehen und meine Forschungsfrage beantworten. In einem zweiten Schritt werde ich meine Ergebnisse und den Forschungsprozess durch eine Methodenkritik noch genauer einordnen und in einem letzten Schluss werde ich einen Ausblick auf

weitere Fragestellung so wie eine potentielle Verwendung meiner Forschungsergebnis geben.

A.) Vorstellung der Forschungsergebnisse

Ich möchte nun die Forschungsfrage beantworten, inwiefern Gender in Gesichtserkennung eingelassen ist und wie diese Ergebnisse politikwissenschaftlich zu interpretieren sind.

1.) Antwort zur Forschungsfrage 1

F1: Wie ist Gender in die automatische Gesichtserkennung eingelassen, am Beispiel des Gender Shades Projektes?

Ich habe in meiner Arbeit nachgewiesen, dass Gender multiple Funktionen insbesondere im Hinblick auf die Zielsetzung und Funktionsweisen des Gender Shades Projektes erfüllt.

Einerseits ist die binäre Gender-Konstruktion die Buolamwini und Gebru nutzen performativ. Das bedeutet, dass Geschlecht mittels maschineller und menschlicher Aktionen immer wieder als Kategorie hergestellt wird, beispielsweise durch das Beschriften der Bilder im PPB, das Buolamwini und Gebru erstellten (vgl. Buolamwini, 2017). Dabei habe ich herausgearbeitet, dass diese Performativität nicht nur in einem rein menschlichen Kontext ausgeübt wird, sondern beispielsweise auch durch das Klassifizieren von Bildern durch die untersuchten Genderclassifier; auch maschinelle Agency greift. Der von mir verwendete Theorierahmen, hinsichtlich der von mir genutzten Butlerschen Konzeption von Gender als Norm, hat sich dabei als unzureichend herausgestellt (vgl. Butler 2011, S. 71-73). Am Beispiel der von Buolamwini und Gebru verwendeten FAT-Ansätze zur Charakterisierung ihrer Zielvorstellung, faire Algorithmen zu erstellen bzw. genaue Handlungsbedarfe durch Evaluation zu ermitteln, habe ich aufgezeigt, dass diese Konzepte politischer Aktionen und Möglichkeiten immer noch stark in einem rein menschlichen und juristischen Rahmen operieren (vgl. D'Ignazio, Klein 2020, S. 60-61). Gender, so habe ich argumentiert, ist dabei eine Komponente, die den Handlungsraum hinsichtlich nichtmenschlicher Akteur:innen und algorithmischer Handlungen durchaus aufbrechen kann (vgl. Bath 2017). Ich habe deshalb auf die speziellen Bedingungen der Gesichtserkennungssoftware als Untersuchungsgegenstand aufmerksam gemacht; dies mittels Baths Konzept der hybriden Objekte und darauf aufbauend Karen Barads Konzept der posthumanistischen Performativität, welches ich noch einmal von Butlers Begriff der Performativität abgrenzt habe (vgl. Bath 2013, S.99; Barad 2007, S. 135, Butler 1991, S. 205). Geschlecht ist nach meiner Analyse also posthumanistisch performativ in das Gender Shades Projekt eingelassen.

Aufgrund dieser Interpretation konnte ich ferner als zweiten Schritt Buolamwinis und Gebrus Idee

der fairen Gestaltung von KI weiterführen und das Handeln der Wissenschaftler:innen als politisch definieren, angelehnt an Donna Haraways Idee der Informatik der Herrschaft und meiner zugrundelegten Definition der Techno-Science als politische Designpraxen (vgl. Haraway 1995a, S. 48-50). Die politikwissenschaftliche Dimension des Gender Shades Projekts hinsichtlich seiner Zielsetzung, ergibt sich dabei für mich, ausgehend von meiner theoretischen Setzung, aus dem verantwortungsvollen und nachvollziehbaren Prozess der Gestaltung des PPB durch Buolamwini und Gebru. Der Prozess und nicht das Ergebnis ist dabei das Augenmerk meiner Machtanalyse gewesen. Dadurch biete ich eine Perspektive, die Technikgestaltung als politisch definiert, und somit im Folgeschluss gestaltbar macht, da, so meine Argumentation, das Design der PPB immer auch anders hätte geschehen können, beispielsweise durch das Verwenden nicht öffentlich zugänglicher Bilder (vgl. Haraway 1995b, S. 77-79). Dabei, so habe ich argumentiert, verschiebt sich der Ausgangspunkt politikwissenschaftlicher Untersuchungen und Policy-Vorschläge weg von einer datenschutzrechtlichen Schwerpunktsetzung hin zu einer prozessualen Idee, Codierpraxen politischen Regeln zu unterwerfen, wie kollektive Designstrategien, Grundrechte als Gestaltungsgrundlage für Datensets und Benchmarks und die Idee eines Aushandlungsprozesses für die Verwendung potentieller Genderclassifier, wie sie auch Buolamwini in einem kritischen Folgeartikel über das Gender Shades Projekt beschreibt (vgl. Raji, Buolamwini 2019).

Ich habe dabei aufgezeigt, dass sexistische und rassistische Strukturen in die Praxen zur Erstellung von Künstlicher Intelligenz (ich habe hierbei nicht nur Gesichtserkennungssoftware eingeschlossen) eine Grundbedingung für aktuellen Anwendungen dieser Art sind, ferner, dass Gender als Ungleichheitskategorie eine Bedeutung hat, die über das Gender Shades Projekt hinausragt in Form des Gender Data Gap, der Unterrepräsentation von Cis-Frauen und insbesondere BIPoCs, und schließlich auf das Privilege Hazard hingewiesen, die privilegierte und unterbewusste Perspektive sozial dominanter Gruppen, hier insbesondere weißer Cis-Männer. Diese politikwissenschaftlich relevanten Konzepte sind, so mein Fazit der Analyse, konkreter bezüglich intersektional angelegter Untersuchungen hinsichtlich schadhafter KI-Anwendungen, als beispielsweise die Fokussierung auf Fairness und Unfairness, da diese Konzepte marginalisierte Gruppen explizit als Ausgangspunkt ihrer Analyse setzen (vgl. D'Glnazio, Klein 2020, S. 29, Criado-Perez 2020, S.15 - 20).

Ich habe ferner im zweiten Teil der Arbeit argumentiert, dass Gender als Technologie bestimmte Funktionen innerhalb des Gender Shades Projektes erfüllt. Gender, so meine Antwort auf die Forschungsfrage, ist nicht nur durch Technologie posthumanistisch in das Gender Shades Projekt eingeschrieben sondern ist selbst eine Technologie innerhalb des Projektes. Dafür habe ich die Konzeption von „Race as Technologie“, der Medienwissenschaftler:innen Beth Coleman und Wendy Chun zugrundelegt und mittels meiner Theoriematrix zusätzlich auf Gender bezogen (vgl.

Chun 2009; Coleman 2009).

Hierbei habe ich zwei spezifische Funktionsweisen von Gender als Technik herausgearbeitet, die des Standards und die des Protokolls. Gender ist also als Protokoll und als Standard in das Gender Shades Projekt eingeschrieben.

Hinsichtlich des theoretischen Konzepts Standard habe ich dabei auf die Arbeiten von Star und Bowker zurückgegriffen und Gender als Standard erfasst (vgl. Bowker, Star 1999). Ich habe diese Funktion von Gender noch einmal in zwei Aspekte aufgeteilt. Erstens wirkt der Gender-Standard, wie ich das von Buolamwini und Gebru verwendet Geschlechterkonzept benannt habe, auf einer basal technischen Ebene. Genderclassifier funktionieren demnach nur als binäre Zuordnungsmaschinen, weitere Gender-Kategorien kann es in diesem System nicht geben, da alle Genderclassifier auf dem gleichen Gender-Modell basieren. Folglich ist eine Evaluation dieser Genderclassifier mit Buolamwinis und Gebrus Prämisse, Genderclassifier auf ihre Akkuranz hinsichtlich Schwarzer Menschen zu prüfen, nicht anders möglich. Zweitens ermöglicht der binäre Gender-Standard eine Wissenschaftlichkeit, die sowohl Validität als auch Reliabilität schafft. Erst durch die Übernahme der binären Gender-Kategorie können die zu evaluierenden Genderclassifier dann auch von Buolamwini und Gebru untersucht und verglichen werden (vgl. Buolamwini, Gebru 2018, S. 6). Der von mir herausgearbeitete Gender-Standard schafft somit eine Realität, in der das binäre Konzept von Geschlecht ein selbstwirksamer Fakt wird, der gemessen werden kann, durch Software und Forscher:innen. Grundlegend habe ich mich dabei bewusst für den Standard als Ausgangspunkt meiner Analyse der technischen Funktionen im Gender Shades Projekt entschieden. Die bisherigen Studien zu KI und Klassifikation in Bezug auf Gender und Race gingen dabei zumeist von einer einfachen Klassifikation aus bzw. davon, dass Gender dieses Klassifikationssystem darstellt (vgl. West 2020, S.3; Bives, Haimson 2016, S.443-444). Ich habe Gender hingegen als Standard aufgefasst und auf die rechtlichen und staatlichen Zwangsmechanismen hingewiesen, die mit der Durchsetzung eines Standards im Zusammenhang mit einem binären Gender-System einhergehen, und ferner auf die Besonderheit der Normierung die dem Konzept des Standards, wie ich ihn definiert habe inhärent ist. Dafür habe ich einerseits den Normbegriff von Butler und Foucault angeführt und in einem zweiten Spezifikationsschritt auf den Zusammenhang statistischer Standardisierung und Normalisierung durch eben jene hingewiesen, ausgehend von kritischen Disability-Studien zum Zusammenhang von Standards und Unterdrückungsmechanismen (vgl. Davis 2017, Butler 2011, S.91)

Des Weiteren habe ich Gender als Technologie spezifiziert und, aufbauend auf der Idee von Eugene Thacker und Alexander Galloway, Gender als Protokoll begriffen bzw. die These untersucht, dass Gender als Protokoll innerhalb des Gender Shades Projektes zu verstehen sei. Diese These habe ich

aus meiner Theoriematrix abgeleitet, genauer, anhand Jacks Halberstams Idee des Automating Gender hergeleitet (vgl. Halberstam, 1991). Gender als Protokoll stellt dabei eine Analyseperspektive dar, die es mir erlaubte Boulamwinis und Gebrus Arbeit als Zusammenführen von menschlichen und maschinellen Akteur:innen sowie gesellschaftlichen Strukturen zu beschreiben und, anders als der Standard, verstärkt auf die Prozesse hinter eben diese zu blicken (vgl. Hester 2018, S. 110). Dabei habe ich die Idee der posthumanistischen Performativität von Barad wieder aufgegriffen und mittels ihrer Idee des Apparates die spezifische Verbindung von Mensch und Maschine erläutert, dabei habe ich die zugrunde liegende Theorie des Protokolls Feminismus von Murphy eingeführt. Murphy bezog dieses Konzept noch auf die Frauenbewegung der 70er Jahre in den USA und weniger auf Genderclassifery (vgl. Murphy 2012, S. 21-22). Dafür habe ich Helen Hesters Aktualisierung genutzt, die diesen Ansatz auf Open Source Bewegungen und Transfeminismus bezog (vgl. Hester 2018, S. 110-118). Barads Konzeption des Apparats hat dabei die Verbindung von Murphys Idee und KI als Anwendungsfeld als Art Zwischenschritt in meiner Arbeit ermöglicht (vgl. Barad 2007, S.232). Gender stellt demnach spezielle Praxen von Boulamwini und Gebru hinsichtlich des Beschriftens der Bilder, der Verwendung der Genderclassifery und der Evaluierung dar. Das Protokoll ermöglicht hierbei die Idee der emanzipatorischen Handlung und des Weitertragens in andere Kontexte (vgl. Murphy 2012, S. 27-28). Das Gender-Protokoll des Gender Shades Projektes könnte beispielsweise weitergetragen werden, indem andere Wissenschaftler:innen, diese Designpraxis, die Boulamwini und Gebru ausformulieren, als Anleitung nutzen, entweder inklusivere Datensets zu erstellen oder aber auch neue Benchmarks zu gestalten, möglicherweise mit anderen, zusätzlichen Kategorien, wie Be:hinderung. Das Protokoll ermöglicht dabei, anders als das Konzept des Standards, eine lokal verortbare Praxis kleinerer Gruppen und zielt weniger auf Makrostrukturen wie den Staat als Akteur:in ab (vgl. Hester 2018, 114).

Zusammenfassend habe ich mittels dieser bisherigen Analyseabschnitte, zu Zielsetzung und technischer Funktion des Gender Shades Protokolls drei Funktionen ausmachen können: erstens Gender als posthumanistische performative Kategorie, dann Gender als Technik, die ich spezifiziert habe als kategorialen Standard, und als agentielles Protokoll.

Ich möchte in einem nächsten Schritt meine zweite Forschungsfrage beantworten. Diese Trennung meiner Ergebnisse entlang der beiden Forschungsfragen ist dabei bis jetzt nicht immer stringent gewesen. So habe ich bereits anklingen lassen, inwiefern politikwissenschaftliche Aspekte meine Studie konturieren. Die ließ sich aufgrund meiner Fachdisziplin nicht verhindern. Ich möchte jedoch im nachfolgendem Schritt eine ausformulierte Antwort geben.

2.) Antwort zur Forschungsfrage 2

F2: Wie lassen sich die Ergebnisse politikwissenschaftlich interpretieren?

Ich werde mich hierbei stärker auf die Ergebnisse des zweitens Teils der Arbeit stützen, da ich, wie bereits erläutert, gerade in Bezug auf die Zielstellung des Gender Shades Projektes deutlich gemacht habe, dass die Einbettung von Gender aufgrund althergebrachter Konzepte der Disziplin politikwissenschaftlich analysiert werden muss, wie meine Auseinandersetzung mit Transparenz, Fairness und Verantwortlichkeit gezeigt hat.

Ich werde deshalb in diesem Teil der Antwort zur Forschungsfrage auf zwei Punkte eingehen: erstens, ausgehend von einer transfeministischen Analyseebene, die Verbindung von Staat, Gender und Genderclassifier und zweitens die Möglichkeit einer emanzipatorisch politischen Praxis am Beispiel eines von mir entwickelten queeren Protokolls zur Herstellung eines Benchmarks für Genderclassifier.

Im Hinblick auf die posthumanistische Performativität von Gender habe ich bereits argumentiert, dass diese einen weiteren polittheoretischen Horizont eröffnet. Im Hinblick auf politikwissenschaftliche Interpretationen habe ich ferner staatliches Handeln in Bezug auf die Aufrechterhaltung und digitale Intensivierung einer Gender-Binarität diskutiert (vgl. Hoffmann 2020, S. 2). Ich habe hierbei aus einer transfeministischen Perspektive argumentiert, dass staatliche Regulierungen in Hinblick auf Gender immer auf die Funktion eines Standard verweisen. Gender, so habe ich mit Rückgriff auf Stryker und Aizura ausgeführt, wird dabei zu einer mehrdimensionalen staatlichen Grenzziehungspraxis (vgl. Stryker, Aizura 2008). Diese Praxis drückt sich in vergeschlechtlichem Zugang zur Staatsbürger:innenschaft aus, die beispielsweise in Deutschland bis vor kurzem noch exklusiv zweigeschlechtlich reguliert war (vgl. Fütty 2019). Ich habe des weiteren argumentiert, dass sich heutige, westlich kapitalistische Staaten immer stärker auf eine Verwendung und Intensivierung von Gesichtserkennungs- und Gendererkennungssoftware beziehen. Hoffman hat hierfür den theoretischen Begriff Data Violence geprägt, um diese Zunahme zu beschreiben, und den ich als theoretische Brücke genutzt habe um die transfeministischen Ansätze mit meinem Untersuchungsgegenstand zu verbinden (vgl. Hoffmann 2020, S. 3-4). Gender als Standard ist dabei, nach meiner Analyse, eine Verbindung zwischen neuen Technologien und staatlicher Gender-Regulierung. Dabei ist die Verbindung des kategorialen Standards, welche Gender in meiner Untersuchung darstellt, eine neue Konfiguration von Geschlecht und digitalen Technologien, die sich beispielsweise von Foucaults Deutung des Daten erfassenden Staates abhebt (vgl. boyd, Crawford, 2012, S. 663-665). KI beruht dabei, so mein Argument, anders als staatliches Handeln, viel mehr auf dem Wiedererkennen von Mustern durch Mensch/Maschine/Netzwerke und weniger auf rein menschliche Kontrolle. Der kategoriale Gender-Standard ermöglicht dabei das

Generieren dieser wiedererkennbaren Muster.

Ich habe deshalb in einem zweiten Schritt eine Praxis der Subversion dieses Standards aus einer genderqueeren Perspektive erläutert, am Beispiel des Workshops „Drag vs. AI“, welcher von Buolamwinis gegründeter Agorithmic Justice League ins Leben gerufen wurde (vgl. <https://www.ajl.org/drag-vs-ai>, zuletzt abgerufen am 8.2.2020, Butler 1991, S. 202-204). In Anlehnung an Butlers Idee von Drag als Politik für queere Gender-Identitäten habe ich argumentiert, dass der Workshop, welcher dazu anleitet, mittels Drag Make-up Gesichtserkennungssoftware zu „verwirren“, also der Kategorisierungsversuche zu unterwandern, als eine Form dieser performativen Politik gesehen werden kann. Dabei arbeiten die Workshopteilnehmer:innen mit dem explizit gemachten Gender-Standard, welcher nur Mann und Frau zulässt, und erschaffen Identitäten, die nicht in diese Anwendung von KI hineinfallen (vgl. <https://www.ajl.org/drag-vs-ai>, zuletzt abgerufen am 8.2.2020,). Sie, die Workshopteilnehmenden, erschaffen ein Außerhalb des Standards, um Überwachung und Sortierungspraxen zu entgehen und nutzen dabei die Norm der binären Zweigeschlechtlichkeit als Abgrenzungsmöglichkeit (vgl. ebd.). Zweitens habe ich in Bezug auf Gender als agentielles Protokoll argumentiert, dass die Protokoll-Idee eine Möglichkeit bietet, konkrete Handlungen in betroffenen Gemeinschaften zu initiieren, von und mit Datenwissenschaftler:innen. Dabei habe ich ferner aufgezeigt, dass Gender eine Praxis ist, die KI anleitet und nicht nur eine messbare Variable, die diskriminierenden Output von KI hinsichtlich Geschlecht hervorbringt.

Dabei habe ich die Idee des Gegenprotokolls von Thacker und Galloway als konkreten Vorschlag für eine politische Theorie in bio-digitalen Zeiten als Grundlage genommen und das Gender-Protokoll als queeres Protokoll in Verbindung mit ähnlichen, theoretisch spekulativen Ideen, wie etwa dem kritischen, postkolonialen Protokoll von Hemangini Gupta, gebracht (vgl. Gupta 2020, Thacker, Galloway 2014).

Ich habe hier, vorrangig für genderqueere Gemeinschaften und deren Handlungsmöglichkeiten jenseits der Einordnung in nicht zutreffende Gender-Kategorien, ein Konzept der Erweiterung des verwendeten Protokolls von Boulaqmini und Gebru beschrieben. Ausgehend von der von Hester entworfenen Ebene des Mesopolitischen, die zwischen den Lokalen und Globalen Strukturen liegt, habe ich argumentiert, dass Gender als queer-feministisches Protokoll immer kontextualisiert ist, jedoch als Operationalisierungsversuch zwischen Kollektiven weitergegeben wird (vgl. Hester 2018, S. 82-86). Dabei funktionieren Protokolle im praktischen Sinne, indem sie angewendet und je nach Bedarf auch angepasst werden. Andererseits werden Protokolle verbreitet, sie sind operationalisierbare Handlungsanleitungen, die sich zwischen Netzwerken und Gruppen bewegen können, bzw. ausgetauscht werden durch ihrer Verwender:innen (vgl. ebd.).

Ich habe diese konzeptuelle Ambivalenz des Gender-Protokolls genutzt, um eine fiktionale genderqueere Benchmark zu beschreiben, welche durch eine gemeinschaftliche Datenproduktion, in Form von Bildarchiven und Neuproduktionen, gekennzeichnet ist. Des Weiteren habe ich eine gemeinschaftsorientierte Praxis der Benennung für Gender-Kategorien entworfen und, darauf aufbauend, ein Konzept der Zuordnung von Kategorien und Bildern. Diese Praktiken sollten innerhalb konkreter Gemeinschaften erarbeitet werden. Ferner habe ich angeregt diese Arbeit durch staatliche Förderung zu bezahlen. Ich beziehe mich dabei auf die frühen Überlegungen der Freien Software- Bewegung, die Software in Gemeinschaftsprozessen entwickelte und dafür einen Lohn forderte, um leben zu können von den Produkten und unabhängig zu sein von großen Konzernen. Eine politische Strömung, die sich in den 1980er in den USA formierte. Heutige Free Software Anwendungen sind beispielsweise die Programmpaket GNU, welches einen Bestandteil der Opensource Software Linux ist (vgl. Webb, 2020, S.160-170). Dabei habe ich ausgearbeitet, dass politische Implikationen, in wie fern die Selbstermächtigung marginalisierter Gruppen innerhalb des Designprozesses rund um KI, ein wichtiger Punkt ist. Genderqueers und Queers of Color, welche noch nicht einmal in statistischen Erhebungen zu der Geschlechterverteilung innerhalb von Designteams erfasst werden, könnten so eigenständig KI gestalten. Dabei folge ich auch Buolamwinis und Gebrus weiterführenden Gedanken. Die Wissenschaftler:innen sprechen sich in Folgepublikation zu dem Gender Shades Projekt, für ein prozessuales Verständnis von Fairness aus, welches den Designprozess mit einschließt und nicht nur auf ein fertiges Produkt fokussiert ist (vgl. Raji et. al. 2020, S.4-5). Meine Protokollidee kommt dieser Forderung nach.

Als letzten grundlegenden Punkt habe ich innerhalb der drei vorgestellten Funktionen herausgearbeitet, dass Haraways Konzeption der Informatik der Herrschaft, Macht als Möglichkeit der Verbindung der Akteur:innen Kapital und Techniken am Beispiel von Gender definieren und vor allem spezifizieren lässt, indem ich die speziellen Mechanismen der Verbindung in Form des Standards und des Protokolls beschrieben habe, und auch als grundlegenden Eigenschaft von Gender als posthumanistisch performativ. Gender, um Butler noch einmal zu bemühen, weist demnach nicht nur einen spezifischen Regulierungsapparat auf, der sich durch Rechtsnormen und staatliche Gewalt nachweisen lässt, sondern auch durch intelligente Software, dabei ist weniger Foucaults Verständnis von Norm prägend, welches Butler nutzt, sondern Haraways Idee der Verbindung von Mensch und Maschine. Ich habe argumentiert, dass diese Verbindung im Fallbeispiel des Gender Shades Projektes Gender sein kann.

In einem letzten Abschnitt möchte ich auf meine Ergebnisse hinsichtlich Gender und Arbeitsteilung kommen, welche, aufgrund des gewählten Fallbeispiels, weniger aussagekräftig sind. Ich habe hinsichtlich der von Buolamwini und Gebru geleisteten Arbeit argumentiert das Gender hier die

Funktion eines Differenzmarkes einnimmt. Ich habe dafür die wissenschaftliche Arbeit zu Gender und AI hin Hinblick auf die Frage von Frauen in technologischen Berufen betrachtet, um die Stellungen von Buolamwini und Gebru genauer beschreiben zu können (vgl. Chun 2013, S.30-33; Light 1999). Ich habe aufgezeigt, dass Boulamwini und Gebrus Stellung als Schwarze weibliche Informationswissenschaftler:innen in zweierlei Weise marginalisiert ist, was sich einerseits durch die bereits beschriebene homogene Entwickler:innenschaft zeigt. Andererseits habe ich aber auch , darauf aufmerksam gemacht, dass gerade Arbeiten rund um KI, die schlecht bezahlt sind und prekäre Arbeitsbedingungen bereithalten, oftmals von marginalisierten Gruppen ausgeführt werden (vgl. Gray, Suri 2019, Schmidt 2019). Ferner haben Boulamwini und Gebru, in ihrer Funktion als Autor:innen der Studie, bestimmte Arbeiten wieder sichtbar werden lassen, die mittlerweile in anderen Studien mit KI externalisiert werden, wie das Beschriften der Bilder für die Benchmark (vgl. Gray, Suri 2019, S. 6-7).

Gender und Race stellen in Bezug auf Arbeitsteilung im Gender Shades Projekt eine Aussicht dar, die die bis jetzt bestehenden Ungleichverteilungen von universitären und wirtschaftlichen Positionen umkehrt und Ausblick auf eine akademische Arbeitsweise ermöglicht, die weniger auf unterbezahlte, verunsichtbarende menschliche Arbeitskraft setzt (vgl. Atanasoski, Vora 2015). Gleichzeitig ist das Gender Shades Projekt jedoch eine Ausnahme und fällt statistisch nicht ins Gewicht, im Hinblick bezogen auf weiße Männlichkeit als Paradigma der Computerwissenschaftler, insbesondere im Bereich Künstliche Intelligenz (vgl. Ensmenger 2015).

Dazu kommt, dass Timnit Gebru vermutlich aufgrund kritischer Äußerungen im Hinblick auf diskriminierende Spracherkennungsanwendung von Google entlassen wurde (vgl. <https://www.nytimes.com/2020/12/03/technology/google-researcher-timnit-gebru.html>, zuletzt geprüft am 05.03.201).

Zusammenfassend ist Gender als posthumanistische Performativität und Technik in das Gender Shades Projekt eingeschrieben. Im Hinblick auf Arbeitsteilung kann Gender als Perspektive dienen, bestehende Forschung hinsichtlich der Arbeit an Künstlicher Intelligenz einzuordnen.

Politikwissenschaftliche Interpretationen dieser Funktionen haben in meiner Arbeit ferner ergeben, dass Gender eine bestimmte Form der Macht, nämlich der Informatik der Herrschaft, beinhaltet und sich diese durch verschiedene Betrachtungsweisen der genauen Funktion von Gender untersuchen ließen. Dabei habe ich durch die Perspektive von Gender als Standard und als Protokoll einerseits aufgezeigt, inwiefern staatliche Regulation von Bedeutung für Gender als Standard ist, und andererseits politische Interpretationen angeboten, die neue Handlungsmöglichkeiten in Bezug auf Genderclassifyer ermöglichen, wie beispielweise das beschriebene Protokoll für eine genderqueere Benchmark. Dabei gibt es jedoch Einschränkungen für die Gültigkeit meiner Analyse, die über die

Begrenztheit aufgrund meines gewählten Forschungsgegenstandes hinaus ragen. Ich werde diese im folgendem Teil ausführen.

B.) Methodenkritik

Ich möchte in diesem Schritt meine Methoden und forschungsethisch relevanten Punkte für diese vorliegende Arbeit besprechen, dabei konzentrieren ich mich voranging auf drei Punkte: erstens die Ansiedlung der Arbeit in nordamerikanische und europäische Kontexte, zweitens die Vorrangstellung bestimmter zugrunde gelegter Theorien, sowohl in meiner Literatúrauswahl als auch in den von mir zitierten Quellen und drittens die Frage nach Interessenskonflikten hinsichtlich meiner verwendeten Literatur.

1.) Geograpische Einschränkungen

Zuerst möchte ich auf meinen geographischen Fokus eingehen. Für meine Studie habe ich dabei das Gender Shades Projekt gewählt, welches zwar ein Teil seiner verwendeten Bilddateien aus Südafrika und Ruanda bezieht, aber in einen eurozentristischen und besonders US-amerikanischen Diskurs rund um Gesichtserkennungssoftware eingeschrieben ist (vgl. ebd. Buolamwini 2018, S.4; Couldry 2019, S.38-42). Des Weiteren habe ich, darauf aufbauend, vor allem wissenschaftliche Arbeiten aus dem globalen Norden verwendet, um meine Argumentation fallspezifisch zu stützen. Dies trägt unvermeidlich zu einem Diskurs bei, der vor allem den globalen Norden als Ort der Erstellung und Verwendung von KI setzt (vgl. ebd. Amrute, Murillo 2020, S.3). Dabei bin ich zwar kurz auf eine internationale Arbeitsteilung eingegangen, die entlang dieser Differenzlinie (globaler Norden/ globaler Süden) zu analysieren ist. Ferner habe ich auf die kolonialen Erbschaften der Idee von Gesichtserkennung oder aber auch der Frage der Grenzziehung des Staates durch den Gender-Standard hingewiesen. Dies war notwendig, um das Gender Shades Projekt in breite politikwissenschaftliche Debatten einzuordnen und weist auf gesellschaftlich verortete Unterdrückungsstrukturen hin. Dennoch bleibt der globale Süden in meiner Arbeit unterrepräsentiert als Ort kritischer technischer Entwicklung bzw. geht meine Arbeit nicht näher auf Kämpfe hinsichtlich von Technikgestaltung im globalen Süden ein, die es zweifelsohne genauso gibt wie wissenschaftliche Literatur darüber (vgl. Russel, Stolfi 2020; Shaw 2011; Kwet 2019, S.19-21). Ich habe den Fokus meiner Arbeit so gewählt, dass diese Fragen nicht im Zentrum der Analyse standen; dennoch stellt meine Ausklammerung eine weitere Leerstelle innerhalb eines nordamerikanischen und eurozentristischen Bildes von Künstlicher Intelligenz dar. Dies ist

problematisch, da es innerhalb der kritischen wissenschaftlichen Auseinandersetzung um Künstliche Intelligenz, so immer wieder das Narrativ eines technisch entwickelten globalen Nordens und eines passiven abgehängten Südens wiedergibt. Was einerseits die komplexen Abhängigkeitsverhältnisse innerhalb der Gestaltung von KI verschleiern, wie ich sie anhand der Arbeitsteilung angerissen habe, oder aber Chinas ambivalente Rolle als Vorreiterin in Bezug auf KI und der Frage des Umgangs mit intelligenter Software in autoritären Staaten (vgl. Larson 2018; Zeng 2020). Andererseits erlauben meine Fallauswahl und mein Fokus wenig Weitblick hinsichtlich der vielen Bestrebungen von Wissenschaftler:innen und Aktivist:innen das Potential subalterne Technologie voll auszuschöpfen, da ich als Gegenstand eine Evaluationsstudie von Genderclassifyern gewählt habe.

2.) Theoretische Homogenität

Zweitens möchte ich die von mir gewählte Literatur kritisch betrachten. Ich beziehe mich dabei vor allem auf die US-amerikanische Literatur die ich benutzt habe. Hierbei fiel mir eine große Übereinstimmung hinsichtlich der Verwendung des Intersektionalitätskonzeptes und der Verwendung von Macht auf. Fast ausschließlich haben die Autor:innen, die sich mit schadhafter KI auseinandergesetzt haben, in ihren Werken Crenshaw als Referenz für ihre intersektionalen Ansätze, also das Verbinden der Ungleichheitskategorien von Race und Gender verwendend, und sind im Anschluss auf die Matrix of Domination von Patricia Hill als Definition für die Wirkweise von Macht in solch intersektionalen Zusammenhänge eingegangen (vgl. D'gnazio, Klein 2020, S.60-61, Costanza-Chock 2018). Diese Wahl erscheint sinnvoll in Anbetracht der Urheber:innenschaft des theoretischen Konzeptes von Kimberlé Crenshaw. Dennoch möchte ich hier anmerken, dass die Verbindung der Kategorien Race und Gender vielstimmiger vertreten ist im wissenschaftlichen Diskurs. Die Politikwissenschaftlerin Leslie McCall hat darauf hingewiesen, dass die Verwendung von Kategorien innerhalb intersektionaler Ansätze auf verschneiden Theoretische Traditionen aufbaut und diese auch von verschiedenen Wissenschaftler:innen in unterschiedlichen Funktionen genutzt wird. So gibt es intersektionale Ansätze, die die Dekonstruktion von Kategorien als Grundprämisse folgen oder aber Strömungen, die Gesellschaft entlang solcher Kategorien, wie Race und Gender analysieren (vgl. McCall 2005, S.178). Dies habe ich in meiner Arbeit zwar an manchen Stellen einfließen lassen, wie beispielsweise der Einbezug des Combahee River Collective als theoretische Vordenker:innen des Protokoll-Feminismus nach Murphy (vgl. Murphy 2012, S.42-44). Dennoch möchte ich in diesem Zusammenschluss diese von mir auch genutzte theoretische Homogenität benennen. Ich möchte deshalb diesen von mir gewählten theoretischen Zugang, der sich sowohl auf Buolamwini und Gebrus Konzept als auch auf Sekundärliteratur, wie der Analyse

von D'Ignazio und Klein, stützt, als MIT-Schule bezeichnen. Folgende Gründe sind für mich dabei ausschlaggebend: erstens ist eine Vielzahl der Autor:innen, die sich kritisch mit KI, und vor allem Facial Recognition, im Rahmen intersektionaler Analysen auseinandersetzt, am MIT beschäftigt sind oder waren, und/oder das MIT bzw. Projekte des MIT als empirischen Ausgangspunkt von Studien nutzten (vgl. Buolamwini 2017, D'Ignazio, Klein 2020, Constanza-Chock 2018).

Die Benennung dieser Theorieschule gilt dabei nur für meine Analyse, da ich sie im Rahmen der von mir gewählten und zitierten Literatur festgestellt habe. Ich treffe hierbei keine theoretisch abgesicherte Allgemeinaussage über das Feld der intersektionalen Analyse von KI und insbesondere von Gesichtserkennungsanwendungen.

3.) Interessenkonflikt

Als Dritten Punkt möchte ich transparent machen, dass viele zitierte Quellen von Wissenschaftler:innen verfasst wurden, die in der freien Wirtschaft, genauer gesagt für Konzerne arbeiten, die intelligente Software vertreiben, nicht zuletzt Timnit Gebru, die zum Zeitpunkt der Studie Gender Shades bei Google beschäftigt war (vgl. Buolamwini, Gebru 2018, S. 1). Zwar bekundeten die Wissenschaftler:innen, die für ein solches Unternehmen gearbeitet haben, dass keine Interessenskonflikte innerhalb der Forschung auftraten, jedoch verweist der Sozialwissenschaftler Thilo Hagendorf vor diesem Hintergrund auf das Weglassen von kritischen Forschungsansätzen, die sich mit unternehmenskritischen Forderungen hinsichtlich der Gestaltung von KI befassen, gerade auch im Forschungsbereich der Fairen Algorithmen (vgl. Hagendorf 2020, S.100-105). Es bleibt aber auch fraglich, inwiefern beispielsweise die Studie von Gray und Suri zu Ghostwork wirklich frei von Interessenkonflikten ist, wenn beide Beteiligten für einen der von ihnen untersuchten Konzerne arbeiten.

Dies ist nicht nur ein Problem für meine jetzige Forschungsarbeit, sondern wird auch von anderen Wissenschaftler:innen kritisiert hinsichtlich weitreichender Fragen zu Forschungsergebnissen von Konzernmitarbeiter:innen großer digitaler Unternehmen wie Facebook oder Google (vgl. Zuboff, 2018, S.222)

4.) Ein generelles Verbot

Und viertens bleibt die Frage, ob es in Anbetracht der Auswirkungen von Genderclassifyern überhaupt vertretbar ist, diese zu erforschen und sie weiterzuentwickeln. Buolamwini und Gebru weisen in ihrer Studie auf dieses Paradoxon hin, indem sie beschreiben, dass Transpersonen möglicherweise so noch höheren Risiken von Gewaltanwendungen durch staatliche Behörden oder

kommerzielle Nutzer:innen der Genderclassifier ausgesetzt seien, aber diese, da sie verstärkt eingesetzt werden, das Leben zumeist Schwarzer Menschen gefährde, und deshalb diese Technologie weiterhin evaluiert und nicht nur kritisiert oder abgelehnt werden darf (vgl. Buolamwini, Gebru 2017, S. 1-3, Buolamwini 2017, S.89-99). Dieser Widerspruch lässt sich für mich nicht auflösen und stellt eher einen Kompromiss auf Zeit da. Weil jetzt Genderclassifier möglichst inklusive gestaltet werden, kann so vielleicht größerer Schaden von besonders stark betroffenen Gruppen wie BIPOCs oder aber Transpersonen abgewendet werden. Dennoch sind Genderclassifier als Teil von Gesichtserkennungssoftware höchst risikobehaftete Technologien, die der Kontrolle von Menschen dient, die bisher schon faktisch übermäßiger Kontrolle durch Staaten ausgesetzt waren (vgl. Browne 2015, S.129; Keyes 2018, S.88:13). Dabei sind diese Technologien auf höchst prekären Arbeitsteilungsprozessen aufgebaut und demokratische Kontrollen sind mit jetzigen Gesetzen wenig zielführend, da bis jetzt keine gesetzlichen Reglementierungen zur Erstellung von Datensets oder aber verbindlichen und folgensweren Evaluationen verabschiedet worden sind (vgl. Jobin et al. 2019).

Deshalb erachte ich meine Arbeit zwar als wichtigen Diskursbeitrag, welcher auch weiterführende Gedanken zur Einbeziehung genderqueerer Menschen bereithält, jedoch sehe ich diesen Beitrag kritisch und plädiere in Anbetracht der vorangegangenen Erläuterungen für ein Verbot solcher Technologien. Solange ein solches aber nicht wirksam ist, ist es unerlässlich, marginalisierte Gruppen vor Schaden durch Nicht-Klassifikation oder maschinelle Fehlurteile zu schützen.

Dies führt mich zum letzten Teil des Fazits, in dem ich meinen Beitrag vor genau jener beschriebenen Ambivalenz einordnen werde und eruiere, wie meine Ergebnisse zu weiterführenden Fragen beitragen bzw. welche dies sein könnten.

C.) Ausblick

Ich möchte meinen Ausblick anhand dreier Punkte ausführen. Die ersten beiden Ausblicke hängen dabei maßgeblich zusammen.

Ich möchte hierbei erörtern, inwiefern wissenschaftliche Arbeiten einerseits weiterführend gestaltet werden können entlang der Prämisse, dass die automatische Geschlechtererkennung weiterhin existiert, und andererseits einen Ausblick geben unter der Prämisse, dass diese Technologie prinzipiell abgelehnt werden muss und nicht weiter evaluiert werden sollte.

Drittens werde ich noch eine Frage unabhängig zu diesen beiden Prämissen erarbeiten

1.) Das Fortbestehen der Genderclassifier

In diesem Abschnitt erörtere ich, inwiefern meine Arbeit weitere Anknüpfungspunkte für zukünftige Forschung im Bereich der automatischen Gender-Erkennung, unter der Prämisse vertiefender intersektionaler und inklusiver wissenschaftlicher Arbeit, stellen könnte. Zweitens werde ich erörtern, inwiefern meine entworfene Idee des Gegenprotokolls zu einer ethisch vertretbaren Datengewinnung führen könnte.

a.) Intersektionale Inklusion

Das Gender Shades Projekt stellt eine einzigartige wissenschaftliche Arbeit im Bereich der Gender-Erkennung dar, indem es die erste Studie ist, die sich kritisch mit dem Zusammenhang zwischen Race und Gender im Hinblick auf die Akkuranz der Performance von Genderclassifiern auseinandersetzt (vgl. Buolamiwini 2017, S.).

Es ist daher naheliegend, als Ausblick zu schreiben, dass mehr Studien benötigt werden, um diese Technik, die in sensiblen Bereichen des alltäglichen Lebens eingesetzt wird, zu erweitern (vgl. Hamidi et. Al. 2018). Dazu könnten Studien zählen, die sich auf weitere Anwendungen beziehen, beispielsweise Benchmark-Studien, die über Gesichtserkennung, wie Buolamiwini und Gebru sie durch geführt haben, hinausgehen und beispielsweise Sprachanalyse anhand der Akkuranz entlang von Gender und Race analysieren, was bereits in wenigen Pionierstudien getan wird (vgl. Tatman 2017; Koenecke et. Al. 2020). Andererseits könnten auch weitere Kategorien bzw. eine gender-sensiblere Forschung zur Generierung verlässlicherer Anwendungen innerhalb marginalisierter Gruppen eine Möglichkeit der Erweiterung der kritischen Auseinandersetzung mit Genderclassifier sein.

Ein wichtiger Punkt, den ich bereits benannt habe, ist dabei die jetzige Exklusion von genderqueeren Menschen und Transpersonen als Betroffene von Genderclassifiern und die nicht aussagekräftigen Erhebungen zu den Bedürfnissen genderqueerer Menschen hinsichtlich Genderclassifier und Gesichtserkennung (vgl. Keyes 2018, 88:14).

Ein verstärkter Fokus auf die Inklusion genderqueerer Menschen als Personen, die von Genderclassifiern betroffen sind bzw. die Anerkennung ihrer Existenz innerhalb der Forschung zu Gender-Erkennungssoftware, scheint im Hinblick auf diese Zahlen in Kombination mit den Befunden von Os Keyes wenig fraglich. Jedoch ist die Einbeziehung dieser marginalisierten Gruppe ein ambivalentes Unterfangen, wie ich bereits darlegte. So gaben in der qualitativen Studie von Hamiti et. al. von 12 genderqueeren Menschen in Bezug auf Genderclassifier alle Personen an, a) keinen Mehrwert dieser Technologie zu erkennen, b) keine positiven Effekte der Falschbenennung

ihres eigenen Genders zu erhalten (was in Anbetracht ihrer Nichtexistenz bzw. durch das Einbeziehen nicht gender-konformer Personen eine weitverbreitete Folge ist und c) Genderclassifier für eine Gefahr ihrer eigenen Sicherheit zu halten (Hamidi et al. 2018, S. 6-7). Ferner weist die Soziologin Kate Crawford darüber hinaus darauf hin, dass die Inklusion vulnerabler Gruppen in automatische Gesichtserkennung prinzipiell wenig progressiv ist, bzw. keine Garantie für einen Schutz vor Misgendering, Nichterkennung und somit Ausschluss und Gewalt ist (vgl. Hoffman 2019, S.907-908)

Doch geben Hamidi et al. Beispiele für eine Inklusion von genderqueeren Identitäten innerhalb der Gender-Erkennungssoftware, die zu einer Inklusion führen könnten, wenn diese Technologien weiterhin bestehen sollten.

“P5 [Interviewte Person], P7, and T1 recommended allowing users to explicitly consent and confirm their own identities, supporting their choice and autonomy in the interaction. “I would definitely recommend having an option ... for the person to be able to confirm their identity or have an option ... for people to address that before it affects them.” –P7” (Hamidi et. al. 2018, S.8)

In Kombination mit meinen Vorschlägen zur Ausarbeitung genderqueerer Politiken jenseits einer Debatte um ethische Leitlinien für Technikdesign, könnten diese konkreten Vorschläge weiter ausgearbeitet werden, um responsivere Systeme zu erstellen und beispielsweise ein Eingreifen seitens der Betroffenen zu ermöglichen.

Ferner könnten weitere Arbeit betrieben werden, hinsichtlich der bereits bestehenden Technologie und deren Schrittweiser Transformation zu inklusiveren und faireren Anwendungen. Wie ich bereits gezeigt habe, ist Gender nicht nur eine spezifische Kategorisierungsweise des Gender Shades Projektes sondern eine standardisierte Funktion von Genderclassifiern, deren Wirkungsweise von einem binär ausgeführten und operationalisierten Geschlechtsbegriff abhängt. Es wäre folglich eine weitere wichtige Forschungsfrage, wie dieser Gender-Standard aufgebrochen werden könnte, bzw. inwiefern gender-binäre Systeme und dahinter liegende Datenbanken synchron umgeformt werden könnten, um genderqueere Menschen mit einzubeziehen.

Dies müsste in Anlehnung an das Gender Shades Projekt immer mit der kritischen Hinterfragung eines hegemonialen Weißseins passieren, so dass solche Anwendungen zwar nicht transinklusiv jedoch rassistisch arbeiten. Damit wären die Ziele des Gender Shades Projektes, welche ich als absolut notwendig erachte, konterkariert.

b.) Alternative Datengewinnung

Ein zweiter Punkt unter der Prämisse des Fortbestehens der automatischen Geschlechtererkennung wäre die Ausarbeitung eines neuen Konzeptes der Datengewinnung und somit der Funktionweise für solche Anwendungen.

Wie in der Begründung meiner Fallauswahl schon beschrieben, habe ich mich, aufgrund forschungsethischer Bedenken, gegen das Verwenden von Gender-Erkennungssoftware ausgesprochen und versucht, mittels der Erläuterung eines Protokolls für eine queere Benchmark, diese Bedenken zu minimieren. Eine Frage, die ich aufgrund meiner Fragestellung jedoch nicht bearbeitet habe, ist, wie die Nutzung der potentiellen queeren Datensets oder die des bestehenden, von Boulamwini und Gebru erstellten Datensets, geregelt werden könnte, um den Betroffenen dieser Technologie Gestaltungsmöglichkeiten zu bieten.

Die Weiterverarbeitung von Datensets zur Testung und zum Training von Gesichtserkennungs-Technologien ist eine gängige Praxis so sind Benchmarks beispielsweise oftmals frei zugänglich für potentielle Nutzer:innen, die ihre Anwendungen evaluieren und optimieren möchten. Neuere Studien zeigen dabei, dass Benchmarks und die zugrunde gelegten Datensets immer wieder für ethisch fragwürdige Zwecke gebraucht worden sind (vgl. Weinberg 2019). So zeigten die Künstler:innen Adam Harvey und Jules LaPlace, dass ein von der Universität Stanford entwickeltes Datenset für das Trainieren von Gesichtserkennungssoftware nicht nur eine breite wissenschaftliche Rezeption in über hundert anderen wissenschaftlichen Aufsätzen nach sich zog, sondern dass es auch als Grundlage für Überwachungstechnologien des Chinesischen Militärs zur Unterdrückung der dortigen muslimischen Minderheit der Uiguren führte. Dabei hatten die Forscher:innen, die dieses Datenset erstellten, Kameras auf dem Campus der Universität Stanford aufgestellt und die vorbeigehenden Menschen gefilmt und deren Bilder für das Trainingsdatenset genutzt, welches sie dann öffentlich zugänglich hochluden. Ferner wurde es von der amerikanischen Grenzschutzbehörde genutzt als technische Hilfe zur Migrationsüberwachung. Des Weiteren wurde es von IBM und Microsoft für kommerzielle Zwecke eingesetzt, ohne dass dafür eine explizite Zustimmung der gezeigten Personen in diesem Datenset gegeben war (vgl. Harvey, LaPlace 2019, Weinberg 2020, S. 20).

Ausgehend von diesem Beispiel muss weitere Forschung spezifischer auf die Frage eines sicheren Zugangs zu sensiblen Datensets eingehen. So könnte weitere Forschung ausgehen von den Ideen bezüglich der ethischen Datengewinnung, wie ich sie anhand von Buolamwinis und Gebrus Arbeiten aufgezeigt und durch das spekulative queerfeministische Protokoll weiter geführt habe, Lizenzierungsmodelle erarbeiten und Fragen der Offenlegung des Zwecks der Verwendung der Datensets ermöglichen. Als Ansatzpunkt könnten dabei beispielsweise der Code of Research Ethics der San in Südafrika gelten. Die San sind eine Gemeinschaft Indigener, die bis zum jetzigen Zeitpunkt beforscht worden sind, aufgrund ihrer DNA. Diese Forschung viel immer wieder durch koloniale Machtverhältnisse auf, in dem beispielsweise dem Oberhaupt der San in verschiedensten Forschungsunterfangen, die Einsicht in die erhobenden Daten verwehrt wurde (vgl. Schröder et. Al.

2019, 73-74). Dieser Code beruht dabei weder ausschließlich auf individueller Zustimmung als Ausgangsvoraussetzung für ethisches Forschen, sondern erfordert das Einbeziehen der Gemeinschaft in den Prozess der Einverständniserklärung, noch ist er lediglich auf den Aspekt des informierten Konsens beschränkt. Vielmehr fließen hier Überlegungen zu Care und Respekt vor der Gemeinschaft als grundlegende Faktoren für die Zusammenarbeit mit externen Forscher:innen ein (vgl. Schörder et. al. 2019, S.85).

So heißt es in den Regelungen zu Care „Research should be aligned to local needs and improve the lives of San. This means that the research process must be carried out with care for all involved, especially the San community. The caring part of research must extend to the families of those involved, as well as to the social and physical environment“ (San Code of Research Ethics 2017, S.3)

Bezogen auf die Arbeit an einem oder mit einem queerefeministischen Protokoll zur Erstellung des Benchmarks könnten marginalisierte Communities in ihrer Funktion als Gruppe und als Sorgengemeinschaft bedacht werden und sich aktiv am Forschungsprozess beteiligen in bereits erarbeiteten Strukturen, wie diesem Ethischen Code of Research.

Betroffene Menschen und Gemeinschaften könnten so feststellen an wen und wie sie ihre Bildrechte abtreten und zu welchem Verwendungszweck und der andererseits aber auch nachvollziehbar bleibt.

Die Frage, die sich jedoch stellt, ist es überhaupt möglich Datensets und Anwendungen der automatischen Gender-Erkennung zu generieren, die Datenschutz und den Grundrechten marginalisierter Gruppen Rechnung tragen können? Oder ist die Funktionsweise solcher Systeme nicht inhärent konträr zu solchen Forderungen? Ich möchte in einem nächsten Abschnitt weiterführenden Gedanken dazu ausführen, ausgehend von der Prämisse, dass Gender-Erkennungssoftware verboten werden sollte.

2.) Das Verschwinden der Genderclassifier

Ausgehend von der zugrunde gelegten Literatur weist ein Großteil kritischer Wissenschaftler:innen auf die Gefahr von automatischer Geschlechtererkennung und deren zugrundeliegenden technischen Faktoren, wie der Datengewinnung und der generellen Probleme dieser Systeme. auf ein Verbot dieser Anwendungen hin (vgl. ebd. Brown 2015, S.129, Hoffman 2019, S.908; Benjamin 2019, S.53-55; Costanza-Chock 2018, Hamidi et. Al. 2018, S.9).

Os Keyes weist hier beispielgebend daraufhin, dass inklusivere Genderclassifier ein Oxymoron seien und fragt nach ihrem vermeintlichen Nutzen (vgl. Keyes 2018, S.88:12)

Dabei kann diese Prämisse des Verbotes von automatischer Gender-Erkennungssoftware auch zu Fragen alternativer Technikgestaltung führen, indem das Produkt abgelehnt wird, aber alternative Technologien erprobt und erdacht werden. Ich beziehe mich hier auf Haraways Ausführungen bezüglich der Möglichkeit von Cyborgs als revolutionäre Techniken. Diese, so Haraway, seien zwar immer geprägt durch Ungleichheitsverhältnisse wie Rassismus, Militarismus und Sexismus, da sie in solchen Gesellschaften entworfen wurden, dennoch können sie auch zu gesellschaftlichen Veränderungen führen, wie sie wie folgt beschreibt:

„Das große Problem mit Cyborgs besteht allerdings darin, daß sie Abkömmlinge des Militarismus und patriarchalen Kapitalismus sind, vom Staatskapitalismus ganz zu schweigen. Aber illegitime Abkömmlinge sind ihrer Herkunft gegenüber häufig nicht allzu loyal.“ (Haraway 1995a, S.36).

In diesem abschließenden Teil möchte ich ausgehen von dieser These, im Hinblick auf AGR, weiterführende Überlegungen darlegen, die diese technische Ausgestaltung verwerfen und KI-Systeme, im Hinblick auf auf genderqueer, feministisch fassen könnten. Solches, in Kombination mit meinen Ausführungen zu genderqueeren Funktionsweisen und Anwendungsmöglichkeiten von Datensets, die in betroffenen Communitys entworfen werden könnten, ließen weitere Arbeiten zur Produktion von KI, beispielsweise nicht nur Gender-Erkennung, zu. Bereits jetzt gibt es kleine Projekte, die sich beispielsweise mit der Möglichkeit auseinandersetzen, Techniken wie Gesichts- bzw. Körpererkennung zu modulieren, und die Geschlecht nicht nur binär fassen Dabei experimentieren Forscher:innen im Rahmen eines Workshops an der Technischen Universität Chemnitz mit Gender-Normen hinsichtlich der Bewegung im Raum.

„As an illustrative subversion of a simple and highly rational device, one group created the concept of “transwalking”. Being a heterogeneous consortium of people, they attached the device to their hips and used the motion sensor to measure and document the patterns of their different walks. They then overlaid the data to create diverse merges and attempted, in turn, to perform these new patterns.“ (vgl. Christensen, Conradi 2017, S.84).

Dies wären Anknüpfungspunkte zu einem feministischen Technikdesign. Des Weiteren könnten Anschlussforschungen im polittheoretischen Feld verstärkt auf die gesellschaftlichen Gründe eingehen, die hinter der scheinbaren Notwendigkeit von Genderclassifyern steht, wie beispielsweise: Auf welches gesellschaftliche Problem antwortet die Entwicklung von Genderclassifyern? Gibt es normative Unterschiede für die Begründung eines Verbots von Genderclassifier und dem Verbot anderen künstlicher intelligenter Software, wie der Bilderkennung in selbstfahrenden Autos? Würde das Verbot von Genderclassifyern, andere Anwendungen, wie beispielsweise Stimmerkennung, tangieren bzw. welche Wechselwirkungen auf andere

Technologien hätte das Verbot von Genderclassifyern und welche Auswirkungen hätte ein solches Verbot auf verschiedene marginalisierte Gruppen? Hierbei könnten diese Forschungen wichtige grundlegende Prämissen für transformatorische Techniken liefern, die sich nicht auf Geschlechterbinarität bezieht.

So beschrieben kann also auch die Prämisse eines Verbotes von Genderclassifyern, ausgehend von meiner Analyse, zu weitreichenden neuen Fragestellungen hinsichtlich dieser Technologie führen.

In einem dritten und letzten Abschnitt möchte ich auf eine Problematik eingehen, die sich nicht auf die Prämissen hinsichtlich des Nutzens von Gender-Erkennung richtet

D.) Abschluss

In meiner Forschungsarbeit habe ich die Funktion von Gender entlang der Frage analysiert, wie Gender in das Gender Shades Projekt eingelassen ist, dabei habe mich bewusst gegen Fragen der vergeschlechtlichten Subjektivierung, beispielsweise der betroffenen Personen, die im Datenset mit ihren Bildern vertreten sind, gewandt. Eine weiterführende Untersuchung könnte diese Lücke schließen und eventuelle Funktionsambivalenzen zwischen der Ordnungsfunktion von Gender in meiner Arbeit und empirischen Geschlechterrollen herausarbeiten.

Andererseits, so möchte ich erläutern, wären in Anbetracht meiner verwendeten Theorien, vor allem der posthumanistischen Gender-Performance von Bath, auch Studien denkbar, die erörtern, inwiefern Gender immer auf menschliche Akteur:innen Bezug nehmen muss (vgl. Bath 2013). So könnte, mit dem Fokus auf Gender-Erkennung, Fragen erörtert werden, die die Beziehung von Gender als Subjektivierungsstrategie und Subjekt, wie beispielsweise bei Butler, als eine Möglichkeit, jedoch nicht ausschließliche Perspektive verhandeln. Dabei könnten folgenden Fragen weiterbearbeitet werden: Braucht Gender ein menschliches Subjekt als Zielrichtung? Wie können beispielsweise Gender-Regulationen jenseits konkreter menschlicher Referenz analysiert werden, wenn Butlers Gender-Performance als theoretische Referenz weiterhin verwendet wird?

Genderclassifier könnten dafür eine fruchtbaren empirischen Referenten bieten und Forschung rund um Geschlecht weiter vorantreiben, indem feste Prämissen, wie Gender als Subjektivierungsweise, hinterfragt werden könnten und zwar mittels technischer Begebenheiten, die vor 20 Jahren beispielsweise nicht existierten, als Butler ihrer bis heute einflussreiche Studie Gender Trouble veröffentlichte (vgl. Butler 1991).

Zusammenfassend erachte ich weitere Forschung rund um Genderclassifier als notwendig, sie sind Teil einer immer stärker einsetzenden Technologie, die explizit Bereiche der politikwissenschaftlichen Analyse betreffen, wie Wahlsystem, Wahlwerbung, Grenzregime,

Unsicherheit und ferner Fragen von Datenschutz, Schutz von marginalisierten Personengruppen und Teilhabe von verschiedenen Akteur:innen hinsichtlich der Gestaltung und Einsatzbestimmung dieser Technologien (vgl. Hoffmann 2019).

Wissenschaftliche Untersuchungen, die sich mit der Verbindung zwischen Gestaltung dieser Technologien und ihren gesellschaftlichen Folgen befassen, sind bis jetzt nicht sehr zahlreich und ethische Überlegung zur zukunftsorientierten und nachhaltigen Ausgestaltung von KI wenig effektiv (vgl. West et. Al. 2019, S.10). Es ist daher von unbedingter Wichtigkeit, weitere Anstrengungen hinsichtlich der Problematiken der Exklusion bestimmter marginalisierter Gruppen (genderqueere Menschen, BIPOCs) und der anhaltenden Diskriminierung dieser entgegenzuwirken.

Um es mit dem Manifest des transfeministischen Kollektivs Laboria Cuboniks zu sagen:

„Statt der endlosen Vermehrung technischer Spielereien liegt die eigentliche Aufgabe in der Entwicklung von Technologien, die den ungleichen Zugang zu reproduktiven und pharmakologischen Werkzeugen, ökologische Katastrophen, ökonomische Instabilität sowie gefährliche Formen der unbezahlten/unterbezahlten Arbeit bekämpfen können.“ (Laboria Cuboniks 2015, S. 17).

Denn diese beschriebenen Ungleichheiten sind keine gesellschaftliche Notwendigkeit. Sie gehören abgeschafft!

V.) Literaturverzeichnis

(2018) Michigan professor embroiled in Israel boycott row. *BBC News*, September 21.

<https://www.bbc.com/news/world-us-canada-45592553>. Accessed 3/12/2021.

Adam, A. (1998) *Artificial knowing: Gender and the thinking machine*. Routledge, London.

Alaimo, S. (ed.) (2017) *Gender: Matter Macmillan Interdisciplinary Handbooks*. Macmillan Reference USA, Farmington Hills, MI.

Albu, O. B. & Flyverbom, M. (2019) Organizational Transparency: Conceptualizations, Conditions, and Consequences. *Business & Society* **58** (2), 268–297.

Alexander, M. (2016) *The new Jim Crow: Masseninhaftierung und Rassismus in den USA*. Verlag Antje Kunstmann, München.

Algorithmic Justice League DRAG VS AI. <https://www.ajl.org/drag-vs-ai#signup>. Accessed 2/19/2020.

Alpaydin, E. (2016) *Machine Learning: The New AI*. The MIT Press, Cambridge.

Ames, M. G. (2018) Deconstructing the algorithmic sublime. *Big data & society* **5** (1), 205395171877919.

Amoore, L. & Piotukh, V. (eds.) (2016) *Algorithmic life: Calculative devices in the age of big data*. Routledge Taylor & Francis Group, London, New York.

- Amrute, S. (2020) Follow the Money. Follow the Bodies. Follow the Design. *Catalyst: Feminism, Theory, Technoscience* **6** (2).
- Amrute, S. & Murillo, L. F. R. (2020) Introduction: Computing in/from the South. *Catalyst: Feminism, Theory, Technoscience* **6** (2).
- Ananny, M. & Crawford, K. (2018) Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society* **20** (3), 973–989.
- Andrejevic, M. (2020) Data Civics: A Response to the "Ethical Turn". *Television & New Media* **21** (6), 562–567.
- Arora, P. (2016) Bottom of the data pyramid: Big data and the global South. *International journal of communication* **10**, 1681–1699.
- Atanasoski, N. & Vora, K. (2015) Surrogate Humanity: Posthuman Networks and the (Racialized) Obsolescence of Labor. *Catalyst: Feminism, Theory, Technoscience* **1** (1), 1–40.
- Avanessian, A. & Hester, H. (2015) *Dea ex machina*, Originalausgabe. Merve Verlag, Berlin.
- Avanessian, A. & Hester, H. (2015) Einleitung. In: Avanessian, A; Hester, H. *Dea ex machina*. Edited by A. Avanessian, H. Hester, Originalausgabe. Merve Verlag, Berlin, pp. 7–14.
- Barabas, C. (2020) Beyond Bias: Re-Imagining the Terms of 'Ethical AI' in Criminal Law. *SSRN Electronic Journal*, 83–111.
- Barabas, C., Doyle, C., Rubinovitz, J. B. & Dinakar, K. (2020) Studying up: Reorienting the Study of Algorithmic Fairness around Issues of Power. In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery, New York, NY, USA, pp. 167–176.
- Barad, K. (2007) *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. Duke Univ. Press, Durham.
- Barad, K. (2013) Diffractionen: Differenzen, Kontingenzen und Verschränkungen von Gewicht. In: Bath, C., Meißner, H., Trinkaus, S. & Völker, S. (eds.) *Geschlechter Interferenzen: Wissensformen - Subjektivierungsweisen - Materialisierungen*. Lit, Berlin, pp. 27–68.
- Bath, C. (2013) Semantic Web und Linked Open Data: Von der Analyse technischer Entwicklungen zum "Diffractive Design". In: Bath, C., Meißner, H., Trinkaus, S. & Völker, S. (eds.) *Geschlechter Interferenzen: Wissensformen - Subjektivierungsweisen - Materialisierungen*. Lit, Berlin, pp. 69–117.
- Bath, C., Meißner, H., Trinkaus, S. & Völker, S. (eds.) (2013) *Geschlechter Interferenzen: Wissensformen - Subjektivierungsweisen - Materialisierungen*. Lit, Berlin.
- Baumgartinger, P. P. (2017) *Trans studies: Historische, begriffliche und aktivistische Aspekte*, 1. Auflage. zaglossus, Wien.
- Beauchamp, T. (2016) 6. When Things Don't Add Up: Transgender Bodies and the Mobile Borders of Biometrics. In: Martínez-San Miguel, Y. & Tobias, S. (eds.) *Trans studies: The challenge to hetero/homo normativities*. Rutgers University Press, New Brunswick, New Jersey, London, pp. 103–112.
- Beauchamp, T. (2018) *Going Stealth: Transgender Politics and U. S. Surveillance Practices*. Duke University Press, Durham.

- Benjamin, R. (2019) *Race after technology: Abolitionist tools for the new Jim code*.
- Bivens, R. (2017) The gender binary will not be deprogrammed: Ten years of coding gender on Facebook. *New Media & Society* **19** (6), 880–898.
- Bivens, R. & Haimson, O. L. (2016) Baking Gender Into Social Media Design: How Platforms Shape Categories for Users and Advertisers. *Social Media + Society* **2** (4), 205630511667248.
- Bovens, M. (2010) Two Concepts of Accountability: Accountability as a Virtue and as a Mechanism. *West European Politics* **33** (5), 946–967.
- Bovens, M. (2014) *The Oxford handbook of public accountability*, 1. ed. Oxford Univ. Press, Oxford [u.a.].
- Bowker, G. C. & Star, S. L. (2000) *Sorting things out: Classification and its consequences*, First paperback edition. The MIT Press, Cambridge, Massachusetts, London, England.
- boyd, d. & Crawford, K. (2012) CRITICAL QUESTIONS FOR BIG DATA. *Information, Communication & Society* **15** (5), 662–679.
- Braidotti, R. (2019) *Posthuman knowledge*. Polity, Cambridge.
- Braidotti, R. & Laugstien, T. (2014) *Posthumanismus: Leben jenseits des Menschen*. Campus-Verlag, Frankfurt am Main.
- Browne, S. (2015) *Dark matters: On the surveillance of blackness*. Duke University Press, Durham u. a.
- Brückner, M. (2018) Care – Sorgen als sozialpolitische Aufgabe und als soziale Praxis. In: *Handbuch Soziale Arbeit, 6. überarbeitete Auflage*. Ernst Reinhardt Verlag, München.
- Bucher, T. (2018) *If...then: Algorithmic power and politics*. Oxford University Press, New York.
- Bundesverfassungsgericht (2017) Beschluss des Ersten Senats vom 10. Oktober 2017. https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/DE/2017/10/rs20171010_1bvr201916.html.
- Buolamwini, J. (2017) *Gender Shades: Intersectional Phenotypic and Demographic Evaluation of Face Datasets and Gender Classifiers*. MIT, Massachusetts.
- Butler, J. (1991) *Das Unbehagen der Geschlechter*, Deutsche Erstausgabe, 19. Auflage. Suhrkamp, Frankfurt am Main.
- Butler, J. (2015) *Die Macht der Geschlechternormen und die Grenzen des Menschlichen*, 3. Auflage. Suhrkamp, Frankfurt am Main.
- Butollo, F. (2020) Digitalization and the geographies of production: Towards reshoring or global fragmentation? *Competition & Change*, 1–20.
- Butollo, F. & Nuss, S. (eds.) (2019) *Marx und die Roboter: Vernetzte Produktion, künstliche Intelligenz und lebendige Arbeit*, 1. Auflage. Dietz, Berlin.
- Campolo, A. & Crawford, K. (2020) Enchanted Determinism: Power without Responsibility in Artificial Intelligence. *Engaging Science Technology and Society* **6**, 1–19.
- Carpenter, L. & Marshall, R. B. (2018) Walking While Trans: Profiling of Transgender Women by Law Enforcement, and the Problem of Proof. *William & Mary Journal of Women and the Law* **24**, 5–38.
- Castelvecchi, D. (2020) Beating Biometric Bias: Facial recognition is improving - but the bigger issue is how it used. *Nature* (587), 347–349.

- Castoriadis, C. (1990) *Gesellschaft als imaginäre Institution: Entwurf einer politischen Philosophie*. Suhrkamp, Frankfurt am Main.
- Chebout, L. N. (2011) Wo ist Intersectionality in bundesdeutschen Intersektionalitätsdiskursen?: - Exzerpte aus dem Reisetagebuch einer Traveling Theory. In: Smykalla, S. & Vinz, D. (eds.) *Intersektionalität zwischen Gender und Diversity: Theorien, Methoden und Politiken der Chancengleichheit*, 1. Auflage. Westfälisches Dampfboot, Münster, pp. 43–57.
- Cheney-Lippold, J. (2017) *We are data: Algorithms and the making of our digital selves*. New York University Press, New York.
- Cherry, M. A. (2016) Virtual Work and Invisible Labor. In: Crain, M. G., Poster, W. R. & Cherry, M. A. (eds.) *Invisible labor: Hidden work in the contemporary world*. University of California Press, Oakland, California, 71–86.
- Choudhury, B. D. (2020) Undetectable: "Coded Bias" Challenges the Racism of Facial-Recognition Software. *Bitch Media*, November 24.
<https://www.bitchmedia.org/article/coded-bias-racist-artificial-intelligence-interview>. Accessed 2/23/2021.
- Christensen, M. & Conradi, F. (2017) Open-source cyborgs and DIY data: Chances and challenges for a democratisation of gender. *GENDER – Zeitschrift für Geschlecht, Kultur und Gesellschaft* **9** (3), 81–90.
- Chun, W. H. K. (2006) *Control and freedom: Power and paranoia in the age of fiber optics*. The MIT Press, Cambridge, Massachusetts, London.
- Chun, W. H. K. (2009) Introduction: Race and/as Technology; or, How to Do Things to Race. *Camera Obscura: Feminism, Culture, and Media Studies* **24** (1), 7–35.
- Chun, W. H. K. (2013) *Programmed visions: Software and memory*, 1. MIT Press paperback ed. MIT Press, Cambridge, Mass.
- Clayton, J. (2020) Timnit Gebru: Google and big tech are 'institutionally racist'. *BBC News*, December 14. <https://www.bbc.com/news/technology-55281862>. Accessed 3/5/2021.
- Cockayne, D. G. & Richardson, L. (2017) A queer theory of software studies: software theories, queer studies. *Gender, Place & Culture* **24** (11), 1587–1594.
- Cockayne, D. G. & Richardson, L. (2017) Queering code/space: the co-production of socio-sexual codes and digital technologies. *Gender, Place & Culture* **24** (11), 1642–1658.
- Cockburn, C. (1988) *Die Herrschaftsmaschine: Geschlechterverhältnisse und technisches Know-how*, 1. Aufl., 1.- 3. Tsd. Argument-Verl., Berlin.
- Coleman, B. (2009) Race as Technology. *Camera Obscura: Feminism, Culture, and Media Studies* **24** (1), 177–207.
- Conitzer, V., Hadfield, G. & Vallor, S. (eds.) (01272019) *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, New York, NY, USA.
- Cooper, M. (2008) *Life as surplus: Biotechnology and capitalism in the neoliberal era*. University of Washington Press, Seattle.
- Costanza-Chock, S., Philip, N. & Ahearn, C. (2018) Design Justice, A.I., and Escape from the Matrix of Domination. *Journal of Design and Science*.

- Couldry, N. & Mejias, U. A. (2019) *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford University Press; De Gruyter, Stanford California, Berlin.
- Crain, M. G., Poster, W. R. & Cherry, M. A. (eds.) (2016) *Invisible labor: Hidden work in the contemporary world*. University of California Press, Oakland, California.
- Crenshaw, K. (1991) Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color. *Stanford Law Review* **43** (6), 12–41.
- Criado-Perez, C. (2020) *Unsichtbare Frauen: Wie eine von Daten beherrschte Welt die Hälfte der Bevölkerung ignoriert*, Deutsche Erstausgabe, 1. Auflage. btb, München.
- Danielle Brathwaite Shirley Black Trans Archive. <https://blacktransarchive.com/>. Accessed 2/23/2021.
- Dannenberg, N. (2019) Queer Surveillance Studies. Überlegungen zu den Schnittstellen von Queer Theory und Surveillance Studies. *GENDER – Zeitschrift für Geschlecht, Kultur und Gesellschaft* **11** (3-2019), 26–40.
- Daub, A. & Gebauer, S. (2020) *Was das Valley denken nennt: Über die Ideologie der Techbranche*, Deutsche Erstausgabe, erste Auflage. Suhrkamp, Berlin.
- Daugherty, P. R. & Wilson, H. J. (2018) *Human + Machine: Reimagining Work in the Age of AI*. Harvard Business Review Press, La Vergne.
- Davenport, T. H. (2018) *The AI advantage: How to put the artificial intelligence revolution to work*. The MIT Press, Cambridge, Massachusetts.
- Davis, L. J. (2017) Introduction: Disability, Normality, and Power. In: Davis, L. J. (ed.) *The disability studies reader*, Fifth edition. Routledge Taylor & Francis Group, New York, London, pp. 1–16.
- Davis, L. J. (ed.) (2017) *The disability studies reader*, Fifth edition. Routledge Taylor & Francis Group, New York, London.
- Deleuze, G. (2016) Postskriptum über die Kontrollgesellschaften. In: Klimke, D. & Legnaro, A. (eds.) *Kriminologische Grundlagentexte*. Springer Fachmedien Wiesbaden, Wiesbaden, pp. 345–352.
- D'Ignazio, C. & Klein, L. F. (2020) *Data feminism*.
- Dubber, M. D., Pasquale, F., Das, S., Powers, T. M. & Ganascia, J.-G. (eds.) (2020) *The Oxford Handbook of Ethics of AI*. Oxford University Press.
- Ebermann, E. (2010) Gütekriterien quantitativer Untersuchungen. <https://www.univie.ac.at/ksa/elearning/cp/quantitative/quantitative-34.html>.
- Edelman, L. (2007) *No future: Queer theory and the death drive*. Duke University Press, Durham.
- Eglash, R. (2002) Race, Sex, and Nerds: FROM BLACK GEEKS TO ASIAN AMERICAN HIPSTERS. *Social Text* **20**, 49–64.
- Eichhorn, K. (2018) Queer Archives: From Collections to Conceptual Framework. In: Romesburg, D. (ed.) *The Routledge History of Queer America*. Routledge, New York, NY Routledge, 2018. |, pp. 123–134.
- Engelmann, P. (ed.) (1998) *Cyborgs & Designerbabies: Flesh Machine: die biotechnologische Revolution*, Dt. Erstausg. Passagen-Verl., Wien.

- Engemann, C. & Sudmann, A. (eds.) (2018) *Machine Learning - Medien, Infrastrukturen und Technologien der Künstlichen Intelligenz*. transcript-Verlag.
- Ensmenger, N. (2010) *The computer boys take over: Computers, programmers, and the politics of technical expertise*. MIT Press, Cambridge, Mass.
- Ensmenger, N. (2015) "Beards, Sandals, and Other Signs of Rugged Individualism": Masculine Culture within the Computing Professions. *Osiris* **30**, 38–65.
- Eubanks, V. (2018) *Automating inequality: How high-tech tools profile, police, and punish the poor*, First edition. St. Martin's Press, New York.
- Evans, W. & Rangarajan, S. (2017) Hidden figures: How Silicon Valley keeps diversity data secret. <https://revealnews.org/article/hidden-figures-how-silicon-valley-keeps-diversity-data-secret/>.
- Faulkner, W. (2001) The technology question in feminism: A view from feminist technology studies. *Women's studies international forum* **24**, 79–95.
- FEDERAL TRADE COMMISSION (2018) *COMPETITION AND CONSUMER PROTECTION IN THE 21ST CENTURY*, Howard University School of Law.
- Firestone, S. (2015) *The dialectic of sex: The case for feminist revolution*, Paperback edition. Verso, London, New York.
- Forsythe, D. E. (ed.) (2001) *Studying those who study us: An anthropologist in the world of artificial intelligence*, Orig. printing. Stanford Univ. Press, Stanford, Calif.
- Forsythe, D. E. & Hess, D. J. (eds.) (2001) *Studying those who study us: An anthropologist in the world of artificial intelligence*, Orig. printing. Stanford Univ. Press, Stanford, Calif.
- Foucault, M. (1990) *The history of sexuality*, Vintage Books ed. Vintage Books, New York.
- Frey, C. B. (2019) *The technology trap: Capital, labor, and power in the age of automation*. Princeton University Press, Princeton, Oxford.
- Fuller, M. (ed.) (2008) *Software studies: A lexicon*. MIT Press, Cambridge, Mass.
- Funk, W. (2018) *Gender studies*. Wilhelm Fink, Paderborn.
- Fussey, P. & Murray Daragh (2019) *Independent Report on the London Metropolitan Police Service's Trial of Live Facial Recognition Technology*, Essex.
- Fütty, T. J. J. (2019) *Gender und Biopolitik: Normative und intersektionale Gewalt gegen Trans*Menschen*, 1. Auflage. transcript, Bielefeld.
- Gabrys, J. (2011) *Digital Rubbish: A Natural History of Electronics*. University of Michigan Press, s.l.
- Galloway, A. R. (2004) *Protocol: How control exists after decentralization*. The MIT Press, Cambridge, Massachusetts, London, England.
- Galloway, A. R. (2007) *Gaming: Essays on algorithmic culture*, [2.] Nachdruck. University of Minnesota Press, Minneapolis.
- Galloway, A. R. & Thacker, E. (2007) *The exploit: A theory of networks*. Univ. of Minnesota Press, Minneapolis, Minn.
- Galloway, Alexander, Thacker, Eugene (2014) Protokoll, Kontrolle und Netzwerke. In: Reichert, R. (ed.) *Big Data*. transcript-Verlag, pp. 289–312.
- Gates, K. (2011) *Our biometric future: Facial recognition technology and the culture of surveillance*. New York University Press, New York.

- Gebru, T. (2020) Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning. In: Gupta, R. (ed.) *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Association for Computing Machinery, New York, NY, United States, p. 3609.
- Gebru, T., Morgenstern, J. & Vecchione, B. et al. (2018) *Datasheets for Datasets*.
- Gill-Peterson, J. (2014) The Technical Capacities of the Body: Assembling Race, Technology, and Transgender. *TSQ: Transgender Studies Quarterly* **1** (3), 402–418.
- Goffey, A. (2008) Algorithm. In: Fuller, M. (ed.) *Software studies: A lexicon*. MIT Press, Cambridge, Mass, pp. 15–20.
- Google Walkout For Real Change (2020) Standing with Dr. Timnit Gebru — #ISupportTimnit #BelieveBlackWomen. *Medium*, December 04.
<https://googlewalkout.medium.com/standing-with-dr-timnit-gebru-isupporttimnit-believeblackwomen-6dad300d382>. Accessed 3/5/2021.
- Graf, D. & Kaser, K. (eds.) (2004) *Vision Europa: Vom Nationalstaat zum Europäischen Gemeinwesen*, [1. Aufl.]. Czernin, Wien.
- Gray, M. L. & Suri, S. (2019) *Ghost work: How to stop Silicon Valley from building a new global underclass*. Houghton Mifflin Harcourt, Boston.
- Grunwald, A. (2019) Technikfolgenabschätzung bei der Mensch-Maschine-Interaktion. In: Liggieri, K. & Müller, O. (eds.) *Mensch-Maschine-Interaktion: Handbuch zu Geschichte – Kultur – Ethik*. J.B. Metzler, Stuttgart, pp. 338–344.
- Gupta, H. (2020) Postcolonial Assembly Protocols for Unnamed Automation Projects [Special Section]. *Catalyst: Feminism, Theory, Technoscience* **6** (2).
- Gupta, R. (ed.) (2020) *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Association for Computing Machinery, New York, NY, United States.
- Hagendorff, T. (2020) The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines* **30** (1), 99–120.
- Halberstam, J. (1991) Automating Gender: Postmodern Feminism in the Age of the Intelligent Machine. *Feminist Studies* **17** (3), 439.
- Halberstam, J. (2018) *Trans*: A quick and quirky account of gender variability*. University of California Press, Oakland, California.
- Hall, S. (2017) *The Fateful Triangle: Race, Ethnicity, Nation*. Harvard University Press, Cambridge, Mass.
- Hall, S. (2018) *Das verhängnisvolle Dreieck: Rasse, Ethnie, Nation*, 1. Auflage. Suhrkamp, Berlin.
- Halpern, O., Mitchell, R. & Geoghegan, B. D. (2017) The Smartness Mandate: Notes toward a Critique. *Grey Room* **68**, 106–129.
- Hamidi, F., Scheuerman, M. K. & Branham, S. M. (2018) Gender Recognition or Gender Reductionism? In: Mandryk, R., Hancock, M., Perry, M. & Cox, A. (eds.) *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, pp. 1–13.
- Haraway, D. J. (ed.) (1995) *Die Neuerfindung der Natur: Primaten, Cyborgs und Frauen*. Campus-Verl., Frankfurt/Main.

- Haraway, D. J. (1995) Manifest für Cyborgs: Feminismus im Streit mit den Technowissenschaften. In: Haraway, D. J. (ed.) *Die Neuerfindung der Natur: Primaten, Cyborgs und Frauen*. Campus-Verl., Frankfurt/Main, pp. 33–72.
- Haraway, D. J. (1995) Situiertes Wissen: Die Wissenschaftsfrage im Feminismus und das Privileg einer partialen Perspektive. In: Haraway, D. J. (ed.) *Die Neuerfindung der Natur: Primaten, Cyborgs und Frauen*. Campus-Verl., Frankfurt/Main, pp. 73–97.
- Haraway, D. J. (1997) *Modest_WitnessSecond_Millennium. FemaleMan©_Meets_OncoMouseTM: Feminism and technoscience*. Routledge, New York, London.
- Harvey, A. & LaPlace, J. (2021) Exposing.ai. <https://exposing.ai/megaface/>.
- Haug, F. (ed.) (1987) *Widersprüche der Automationsarbeit: E. Handbuch*. Argument, Berlin.
- Hayles, N. K. (1999) *How we became posthuman: Virtual bodies in cybernetics, literature, and informatics*. Univ. of Chicago Press, Chicago, Ill.
- Hester, H. (2018) *Xenofeminism*. Polity, Cambridge, UK, Medford, MA.
- Hicks, M. (2019) Hacking the Cis-tem. *IEEE Annals of the History of Computing* **41** (1), 20–33.
- Hildebrandt, M., Castillo, C., Celis, E., Ruggieri, S., Taylor, L. & Zanfir-Fortuna, G. (eds.) (01272020) *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. ACM, New York, NY, USA.
- Ho, D. & Teo, G. (2020) Digital Medicine – The New Frontier for AI in Healthcare. *Advanced therapeutics* **3**, 1–5.
- Hoffmann, A. L. (2019) Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication & Society* **22** (7), 900–915.
- Hoffmann, A. L. (2020) Terms of inclusion: Data, discourse, violence. *New Media & Society*, 146144482095872.
- Hooker, S., Moorosi, N., Clark, G., Bengio, S. & Denton, E. (2020) *Characterising Bias in Compressed Models*.
- Hovy, D., Spruit, S., Mitchell, M., Bender, E. M., Strube, M. & Wallach, H. (eds.) (2017) *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*. Association for Computational Linguistics, Stroudsburg, PA, USA.
- Hutchinson, B. & Mitchell, M. (01292019) 50 Years of Test (Un)fairness. In: *Proceedings of the Conference on Fairness, Accountability, and Transparency*. ACM, New York, NY, USA, pp. 49–58.
- Ignatov, A., Timofte, R. & Kulik, A. et al. (2019) *AI Benchmark: All About Deep Learning on Smartphones in 2019*.
- James, A. (2008) Making Sense of Race and Racial Classification. In: Zuberi, T; Bonilla-Silva, E. *White Logic, White Methods: Racism and Methodology*. Edited by E. Bonilla-Silva. Rowman & Littlefield Publishers, Lanham, pp. 31–45.
- Jasanoff, S. (ed.) (2010) *States of knowledge: The co-production of science and social order, transferred to digital print*. Routledge, London.
- Jaton, F. & Bowker, G. C. (2021) *The constitution of algorithms: Ground-truthing, programming, formulating*. The MIT Press, Cambridge, Massachusetts.

- Jennifer S. Light (1999) When Computers Were Women. *Technology and culture* **40**, 455–483.
- Jobin, A., Ienca, M. & Vayena, E. (2019) The global landscape of AI ethics guidelines. *Nature Machine Intelligence* **1** (9), 389–399.
- Johnson, J. A. (2015) Information Systems and the Translation of Transgender. *TSQ: Transgender Studies Quarterly* **2** (1), 160–165.
- Johnson, J. A. (2018) *Toward Information Justice: Technology, Politics, and Policy for Data in Higher Education Administration*. Springer, Cham.
- Joy Buolamwini & Timnit Gebru (2018) Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In: Sorelle A. Friedler & Christo Wilson (eds.) *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*. PMLR, New York, NY, USA, pp. 77–91.
- Kafer, A. (2017) Bodily Technologies. In: Alaimo, S. (ed.) *Gender: Matter Macmillan Interdisciplinary Handbooks*. Macmillan Reference USA, Farmington Hills, MI, pp. 387–400.
- Keyes, O. (2018) The Misgendering Machines. *Proceedings of the ACM on Human-Computer Interaction* **2** (CSCW), 1–22.
- Keyes, O., Stevens, N. & Wernimont, J. (2019) The Government Uses Images of Abused Children and Dead People to Test Facial Recognition Tech. *Slate*, March 18. <https://slate.com/technology/2019/03/facial-recognition-nist-verification-testing-datasets-children-immigrants-consent.html>. Accessed 3/3/2021.
- Kitchin, R. (2014) *The data revolution: Big data, open data, data infrastructures & their consequences*. Sage, Los Angeles, London, New Dehli.
- Kitchin, R. (2017) Thinking critically about and researching algorithms. *Information, Communication & Society* **20** (1), 14–29.
- Kitchin, R. & Perng, S.-Y. (eds.) (2016) *Code and the city*. Routledge Taylor & Francis Group, London, New York NY.
- Klävers, S. (2018) *Decolonizing Auschwitz?: Komparativ-postkoloniale Ansätze in der Holocaustforschung*. De Gruyter Oldenbourg, Berlin.
- Klimke, D. & Legnaro, A. (eds.) (2016) *Kriminologische Grundlagentexte*. Springer Fachmedien Wiesbaden, Wiesbaden.
- Knapp, G.-A. (2005) »Intersectionality« ein neues Paradigma feministischer Theorie? ; Zur transatlantischen Reise von »Race, Class, Gender«. *Feministische Studien Zeitschrift für interdisziplinäre Frauen- und Geschlechterforschung* **23** (1), 68–8.
- Koch, G. (2005) *Zur Kulturalität der Technikgenese: Praxen, Policies und Wissenskulturen der künstlichen Intelligenz*. Röhrig, St. Ingbert.
- Koenecke, A., Nam, A. & Lake, E. et al. (2020) Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences of the United States of America* **117** (14), 7684–7689.
- Kroll, J. A. (2020) Accountability in Computer Systems. In: Dubber, M. D., Pasquale, F., Das, S., Powers, T. M. & Ganascia, J.-G. (eds.) *The Oxford Handbook of Ethics of AI*. Oxford University Press, pp. 179–196.

- Kurzweil, R. (2005) *The singularity is near: When humans transcend biology*. Viking, New York.
- Kushner, S. (2013) The freelance translation machine: Algorithmic culture and the invisible industry. *New Media & Society* **15** (8), 1241–1258.
- Kwet, M. (2019) Digital colonialism: US empire and the new imperialism in the Global South. *Race & Class* **60** (4), 3–26.
- Laboria Cuboniks (2015) Xenofeminismus. In: Avanessian, A; Hester, H. *Dea ex machina*. Edited by A. Avanessian, H. Hester, Originalausgabe. Merve Verlag, Berlin, pp. 45–57.
- Landström, C. (2007) Queering feminist technology studies. *Feminist Theory* **8** (1), 7–26.
- Larson, C. (2018) China's AI imperative. *Science (New York, N.Y.)* **359** (6376), 628–630.
- Liggieri, K. & Müller, O. (eds.) (2019) *Mensch-Maschine-Interaktion: Handbuch zu Geschichte – Kultur – Ethik*. J.B. Metzler, Stuttgart.
- Loh, J. (2019) *Roboterethik: Eine Einführung*, Erste Auflage, Originalausgabe. Suhrkamp, Berlin.
- Loh, J. (2019) *Trans- und Posthumanismus zur Einführung*. Junius, Hamburg.
- Ludwig, G. (2011) *Geschlecht regieren: Zum Verhältnis von Staat, Subjekt und heteronormativer Hegemonie*. Campus-Verl., Frankfurt am Main.
- Ludwig, G. (2016) Das „liberale Trennungsdispositiv“ als staatstragendes Konstrukt. Eine queer-feministische hegemonietheoretische Perspektive auf Öffentlichkeit und Privatheit. *Politische Vierteljahresschrift* **57** (2), 193–216.
- Ludwig, G. (2016) Das „liberale Trennungsdispositiv“ als staatstragendes Konstrukt. Eine queer-feministische hegemonietheoretische Perspektive auf Öffentlichkeit und Privatheit. *Politische Vierteljahresschrift* **57** (2), 193–216.
- Maciel, M. L. & Albagli, S. (eds.) (2010) *Information, power, and politics: Technological and institutional mediations*. Lexington Books, Lanham, MD.
- Mandryk, R., Hancock, M., Perry, M. & Cox, A. (eds.) (04212018) *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA.
- Margolis, J. & Fisher, A. (2003) *Unlocking the clubhouse: women in computing*, 1. pbk. ed. MIT Press, Cambridge, Mass.
- Marino, M. C. (2020) *Critical code studies*. The MIT Press, Cambridge Massachusetts, London England.
- Mark, G., Fussell, S. & Lampe, C. et al. (eds.) (2017) *Explore, innovate, inspire: CHI 2017 May 6–11, Denver, CO, USA*. Association for Computing Machinery Inc. (ACM), New York, NY.
- Markham, A., Powles, J., Walsh, T. & Washington, A. L. (eds.) (02072020) *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. ACM, New York, NY, USA.
- Martínez-San Miguel, Y. & Tobias, S. (eds.) (2016) *Trans studies: The challenge to hetero/homo normativities*. Rutgers University Press, New Brunswick, New Jersey, London.
- Marx, G. T. (2016) *Windows into the soul: Surveillance and society in an age of high technology*. The University of Chicago Press, Chicago, IL, London.
- Mavhunga, C. C. (2017) Introduction. In: Mavhunga, C. C. (ed.) *What do science, technology, and innovation mean from Africa?* The MIT Press, Cambridge, Massachusetts, pp. 1–29.

- Mavhunga, C. C. (ed.) (2017) *What do science, technology, and innovation mean from Africa?* The MIT Press, Cambridge, Massachusetts.
- McCall, L. (2005) The Complexity of Intersectionality. *Signs: Journal of Women in Culture and Society* **30** (3), 1771–1800.
- Meier, L. (2018) *MRX-Maschine*, Erste Auflage. Matthes & Seitz, Berlin.
- Meißner, H. (2013) Feministische Gesellschaftskritik als onot-epistemo-logisches Projekt. In: Bath, C., Meißner, H., Trinkaus, S. & Völker, S. (eds.) *Geschlechter Interferenzen: Wissensformen - Subjektivierungsweisen - Materialisierungen*. Lit, Berlin, pp. 163–209.
- Mellström, U. (2009) The Intersection of Gender, Race and Cultural Boundaries, or Why is Computer Science in Malaysia Dominated by Women? *Social studies of science* **39** (6), 885–907.
- Mertlitsch, K. (2016) *Sisters - Cyborgs - Drags: Das Denken in Begriffspersonen der Gender Studies*, 1st ed. transcript, Bielefeld.
- Metz, C. & Wakabayashi, D. (2020) Google Researcher Timnit Gebru Says She Was Fired For Paper on AI Bias. *The New York Times*, December 03.
<https://www.nytimes.com/2020/12/03/technology/google-researcher-timnit-gebru.html>. Accessed 3/5/2021.
- Michalitsch, G. (2004) Was ist Neoliberalismus? Genese und Anatomie einer Ideologie. In: Graf, D. & Kaser, K. (eds.) *Vision Europa: Vom Nationalstaat zum Europäischen Gemeinwesen*, [1. Aufl.]. Czernin, Wien, pp. 144–163.
- Mike Ananny & K. Crawford (2018) Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society* **20**, 973–989.
- Misa, T. J. (ed.) (2010) *Gender codes: Why women are leaving computing*. Wiley IEEE Press; IEEE Xplore, Hoboken, New Jersey, Piscataway, New Jersey.
- Misselhorn, C. (2018) *Grundfragen der Maschinenethik*, 2., durchgesehene Auflage. Reclam, Ditzingen.
- Mittelstadt, B. (2019) Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence* **1** (11), 501–507.
- Muñoz, J. E. (2009) *Cruising utopia: The then and there of queer futurity*. New York University Press, New York.
- Murphy, M. (2012) *Seizing the means of reproduction: Entanglements of feminism, health, and technoscience*. Duke University Press, Durham and London.
- Nassehi, A. (2019) *Muster: Theorie der digitalen Gesellschaft*. C.H. Beck, München.
- Nicholson, L. J. (ed.) (1997) *The second wave: A reader in feminist theory*. Routledge, New York.
- Nie, Y., Williams, A., Dinan, E., Bansal, M., Weston, J. & Kiela, D. (2019) *Adversarial NLI: A New Benchmark for Natural Language Understanding*.
- Nissenbaum, H. (1996) Accountability in a computerized society. *Science and engineering ethics* **2**, 42.
- Nkonde Mutale (2019) Automated Anti-Blackness: Facial Recognition in Brooklyn, New York. *Kennedy School review* **20**, 30–36.

- Noble, S. U. (2018) *Algorithms of oppression: How search engines reinforce racism*. New York University Press, New York.
- Ochigame, R. (2020) The Invention of "Ethical AI," How Big Tech Manipulates Academia to Avoid RegulTION. *The Intercept*, December 20.
<https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/>. Accessed 2/24/2020.
- Okune, A. (2020) Open Ethnographic Archiving as Feminist, Decolonizing Practice. *Catalyst: Feminism, Theory, Technoscience* **6** (2).
- O'Neil, C. (2017) *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown, New York.
- Paasonen, S. (2011) Revisiting cyberfeminism. *Communications* **36** (3).
- Plant, S. (2000) *Nullen + Einsen: Digitale Frauen und die Kultur der neuen Technologien*, Vollst. Taschenbuchausg. Goldmann, München.
- Powers, T. M. & Ganascia, J.-G. (2020) The Ethics of the Ethics of AI. In: Dubber, M. D., Pasquale, F., Das, S., Powers, T. M. & Ganascia, J.-G. (eds.) *The Oxford Handbook of Ethics of AI*. Oxford University Press, pp. 25–51.
- Prainsack, B. (2019) Logged out: Ownership, exclusion and public value in the digital data and information commons. *Big data & society* **6**, 1–15.
- Preciado, B. & Preciado, P. B. (2016) *Testo Junkie*, Erweiterte Übersetzung von Beatriz Preciado, Testo Junkie, Paris 2001. b_books, Berlin.
- Puar, J. K. (2007) *Terrorist assemblages: Homonationalism in queer times*. Duke University Press, Durham.
- Pugliese, J. (2007) Biometrics, Infrastructural Whiteness, and the Racialized Zero Degree of Nonrepresentation. *boundary 2* **34** (2), 105–133.
- Pype, K. (2017) Smartness from Below: Variations on Technology and Creativity in Contemporary Kinshas. In: Mavhunga, C. C. (ed.) *What do science, technology, and innovation mean from Africa?* The MIT Press, Cambridge, Massachusetts, pp. 99–115.
- Raghavan, M., Barocas, S., Kleinberg, J. & Levy, K. (01272020) Mitigating bias in algorithmic hiring. In: Hildebrandt, M., Castillo, C., Celis, E., Ruggieri, S., Taylor, L. & Zanfir-Fortuna, G. (eds.) *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. ACM, New York, NY, USA, pp. 469–481.
- Raji, I. D. & Buolamwini, J. (01272019) Actionable Auditing. In: Conitzer, V., Hadfield, G. & Vallor, S. (eds.) *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, New York, NY, USA, pp. 429–435.
- Raji, I. D., Gebru, T., Mitchell, M., Buolamwini, J., Lee, J. & Denton, E. (02072020) Saving Face. In: Markham, A., Powles, J., Walsh, T. & Washington, A. L. (eds.) *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. ACM, New York, NY, USA, pp. 145–151.
- Reichert, R. (ed.) (2014) *Big Data*. transcript-Verlag.
- Reichle, I. (2009) *Art in the age of technoscience: Genetic engineering, robotics, and artificial life in contemporary art*, [Rev. and expanded English ed.]. Springer, Wien, New York.
- Ren, Z., Liu, Y. & Shi, T. et al. (2020) *AIPerf: Automated machine learning as an AI-HPC benchmark*.

- Rezai-Rashti, G. M. (2015) The politics of gender segregation and women's access to higher education in the Islamic Republic of Iran: the interplay of repression and resistance. *Gender and Education* **27** (5), 469–486.
- Richardson, R., Schultz, J. & Crawford, K. (2019) Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice. *New York University Law Review* (94), 193–232.
- Romesburg, D. (ed.) (2018) *The Routledge History of Queer America*. Routledge, New York, NY Routledge, 2018. |.
- Roth, N. L. & Hogan, K. (eds.) (1998) *Gendered epidemic: Representations of women in the age of AIDS*. Routledge, New York.
- Roussel, N. & Stolfi, A. (2020) Taking Back the Future. *Catalyst: Feminism, Theory, Technoscience* **6** (2).
- Rushkoff, D. (2016) *Throwing rocks at the Google bus: How growth became the enemy of prosperity*. Portfolio/Penguin, New York, NY.
- Russell, L. (2020) *Glitch feminism: A manifesto*. Verso, London.
- Russell, S. J. (2019) *Human compatible: Artificial intelligence and the problem of control*. Viking, New York NY.
- Russell, S. J., Norvig, P., Davis, E. & Edwards, D. (2016) *Artificial intelligence: A modern approach*, Third edition, Global edition. Pearson, Boston, Columbus, Indianapolis, New York, San Francisco, Upper Saddle River, Amsterdam, Cape Town, Dubai, London, Madrid, Milan, Munich, Paris, Montreal, Toronto, Delhi, Mexico City, Sao Paulo, Sydney, Hong Kong, Seoul, Singapore, Taipei, Tokyo.
- Sachverständigenkommission für den Dritten Gleichstellungsbericht der Bundesregierung (2021) *Digitalisierung geschlechtergerecht gestalten: Gutachten für den Dritten Gleichstellungsbericht der Bundesregierung*.
- Salamon, G. (2010) *Assuming a body: Transgender and rhetorics of materiality*. Columbia University Press, New York.
- Samudzi, Z. (2019) Bots Are Terrible at Recognizing Black Faces. Let's Keep it That Way. *The Daily Beast*, February 09. <https://www.thedailybeast.com/bots-are-terrible-at-recognizing-black-faces-lets-keep-it-that-way>. Accessed 2/25/2021.
- Schäfers, B. (2019) *Einführung in die Soziologie*, 3rd ed. 2019. Springer Fachmedien Wiesbaden; Imprint: Springer VS, Wiesbaden.
- Schiller, D. (2010) Digital Capitalism In Crisis. In: Maciel, M. L. & Albagli, S. (eds.) *Information, power, and politics: Technological and institutional mediations*. Lexington Books, Lanham, MD, pp. 130–145.
- Schlesinger, A., Edwards, W. K. & Grinter, R. E. (2017) Intersectional HCI. In: Mark, G., Fussell, S. & Lampe, C. et al. (eds.) *Explore, innovate, inspire: CHI 2017 May 6–11, Denver, CO, USA*. Association for Computing Machinery Inc. (ACM), New York, NY, pp. 5412–5427.
- Schmidt, F. A. (2019) *Crowdproduktion von Trainingsdaten: Zur Rolle von Online-Arbeit beim Trainieren autonomer Fahrzeuge*. Hans-Böckler-Stiftung, Düsseldorf.
- Schroeder, D., Chatfield, K., Singh, M., Chennells, R. & Herissone-Kelly, P. (2019) *Equitable Research Partnerships: A Global Code of Conduct to Counter Ethics Dumping*. Springer International Publishing AG, Cambridge.

- Schuklenk, U. (2020) On the ethics of AI ethics. *Bioethics* **34** (2), 146–147.
- Scott, J. C. (ca. 2008] c1998) *Seeing like a state: How certain schemes to improve the human condition have failed*, [Nachdr.]. Yale Univ. Press, New Haven, Conn.
- Sejnowski, T. J. (2018) *The deep learning revolution*. The MIT Press, Cambridge, Massachusetts, London, England.
- Shaw, A. (2011) Insurgent Expertise: The Politics of Free/Libre and Open Source Software in Brazil. *Journal of Information Technology & Politics* **8** (3), 253–272.
- Shephard, N. (2020) Queer enough? Categories and norms at the border.
<https://xyz.informationactivism.org/en/queer-enough/>.
- Smykalla, S. & Vinz, D. (eds.) (2011) *Intersektionalität zwischen Gender und Diversity: Theorien, Methoden und Politiken der Chancengleichheit*, 1. Auflage. Westfälisches Dampfboot, Münster.
- Snow, J. (2018) "We're in a diversity crisis": cofounder of Black in AI on what's poisoning algorithms in our lives. *MIT Technology Review*, February 14.
<https://www.technologyreview.com/2018/02/14/145462/were-in-a-diversity-crisis-black-in-ais-founder-on-whats-poisoning-the-algorithms-in-our/>. Accessed 2/23/2021.
- Sollfrank, C. (ed.) (2018) *Die schönen Kriegerinnen: Technofeministische Praxis im 21. Jahrhundert*, 1. Auflage. transversal texts, Wien.
- Sorelle A. Friedler & Christo Wilson (eds.) (2018) *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*. PMLR, New York, NY, USA.
- South African San Institute (2017) *San Code of Research Ethics*.
- Spertus, E. (1991) *Why Are There So Few Female Computer Scientists?* Massachusetts Institute of Technology, USA.
- statista (2020) Größte Unternehmen der Welt nach ihrem Marktwert im Jahr 2020.
<https://de.statista.com/statistik/daten/studie/12108/umfrage/top-unternehmen-der-welt-nach-marktwert/>.
- Stilgoe, J. (2018) Machine learning, social learning and the governance of self-driving cars. *Social studies of science* **48** (1), 25–56.
- Striphas, T. (2015) Algorithmic culture. *European Journal of Cultural Studies* **18** (4-5), 395–412.
- Stryker, S. & Aizura, A. Z. (2013) Introduction: Transgender Studies 2.0. In: Stryker, S. & Aizura, A. Z. (eds.) *The transgender studies reader*. Routledge, New York, pp. 1–12.
- Stryker, S. & Aizura, A. Z. (eds.) (2013) *The transgender studies reader*. Routledge, New York.
- Suchman, L. (1993) Do categories have politics? *Computer Supported Cooperative Work (CSCW)* **2** (3), 177–190.
- Suchman, L. (2009) *Human-machine reconfigurations: Plans and situated actions*, 2nd edition. Cambridge University Press, Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi, Daubei, Tokyo, Mexico City.
- Suchman, L. (2019) Feministische Science & Technology Studies (STS) und die Wissenschaften vom Künstlichen. *GENDER – Zeitschrift für Geschlecht, Kultur und Gesellschaft* **11** (3), 56–83.
- Suchman, L. A. (1987) *Plans and situated actions: The problem of human-machine communication*. Cambridge Univ. Pr, Cambridge.

- Sudmann, A. (2018) Einleitung: Zur Einführung Medien, Infrastruktur und Technologien des Maschinellen Lernens. In: Engemann, C. & Sudmann, A. (eds.) *Machine Learning - Medien, Infrastrukturen und Technologien der Künstlichen Intelligenz*. transcript-Verlag, pp. 9–36.
- Tatman, R. (2017) Gender and Dialect Bias in YouTube's Automatic Captions. In: Hovy, D., Spruit, S., Mitchell, M., Bender, E. M., Strube, M. & Wallach, H. (eds.) *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*. Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 53–59.
- Thacker, E. (2004) *Biomedica*. University of Minnesota Press, Minneapolis.
- The Combahee River Collective (1997) A Black feminist statement. In: Nicholson, L. J. (ed.) *The second wave: A reader in feminist theory*. Routledge, New York, pp. 63–70.
- van Dijck, J. (2014) Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society* **12** (2), 197–208.
- Vertesi, J. & Ribes, D. (2019) *DigitalSTS: A field guide for science & technology studies*. Princeton University Press, Princeton.
- Wajcman, J. (2004) *TechnoFeminism*. Wiley, Hoboken.
- Wajcman, J. (2019) Automatisierung: Ist es diesmal wirklich anders. In: Butollo, F. & Nuss, S. (eds.) *Marx und die Roboter: Vernetzte Produktion, künstliche Intelligenz und lebendige Arbeit*, 1. Auflage. Dietz, Berlin, pp. 22–37.
- Waldby, C. & Mitchell, R. (2006) *Tissue economies: Blood, organs, and cell lines in late capitalism*. Duke University Press, Durham N.C.
- Webb, M. (2020) *Coding democracy: How hackers are disrupting power, surveillance, and authoritarianism*. The MIT Press, Cambridge Massachusetts, London England.
- Weber, J. (2018) Big-Data-Kriege: Über Tötungslisten, Drohnen und die Politik der Datenbanken. In: Engemann, C. & Sudmann, A. (eds.) *Machine Learning - Medien, Infrastrukturen und Technologien der Künstlichen Intelligenz*. transcript-Verlag, pp. 219–246.
- Weinberg, L. (2020) Feminist Research Ethics and Student Privacy in the Age of AI. *Catalyst: Feminism, Theory, Technoscience* **6**.
- Weiser, M. (1991) The Computer for the 21st Century. *Scientific American*, 94–104.
- West, S. M. (2020) Redistribution and Rekognition. *Catalyst: Feminism, Theory, Technoscience* **6** (2).
- West, S. M., Whittaker, M. & Crawford, K. (2019) *Discriminating systems: Gender, race, and power in artificial intelligence*.
- Wilding, F. & Ensemble, C. A. (1998) Notes on the Political Condition of Cyberfeminism. *Art Journal* **57** (2), 46.
- Wolf, N. (1996) *Der Mythos Schönheit*, 18. - 20. Tsd. Rowohlt, Reinbek bei Hamburg.
- Yardi, P., Marathe, M. & Toyama, K. (2017) Differences in STEM Gender Disparity Between India and the United States. http://www.hcixb.org/papers_2017/hcixb17-final-37.pdf.
- Zeng, J. (2020) Artificial intelligence and China's authoritarian governance. *International affairs (London)* **96**, 1459.

- Zivi, K. (1998) Constituting the "Clean and Proper" Body: Convergences between Abjection and AIDS. In: Roth, N. L. & Hogan, K. (eds.) *Gendered epidemic: Representations of women in the age of AIDS*. Routledge, New York, pp. 33–62.
- Zuberi, T. & Bonilla-Silva, E. (2008) *White Logic, White Methods: Racism and Methodology*. Rowman & Littlefield Publishers, Lanham.
- Zuboff, S. (2018) *Das Zeitalter des Überwachungskapitalismus*. Campus Verlag, Frankfurt, New York.
- Zuboff, S. (2018) *Das Zeitalter des Überwachungskapitalismus*. Campus Verlag, Frankfurt, New York.

VI Anhang

A.) Zusammenfassung

Diese Arbeit untersucht die Frage wie Gender in Genderclassifier eingelassen ist und welche politikwissenschaftliche Interpretation diese Ergebnisse zulassen. Dabei wird die Auditstudie Gendershades Projekt als Fallbeispiel analysiert. In einem Dreischritt werden dabei die a.) Ethischen Zielvorstellung des Gendershades Projekt sogenannte FAT Algorithmen b.) die Funktionsweisen automatischer Gesichtserkennungsanwendungen c.) die Arbeitsteilung innerhalb des Gendershades Projekt analysiert.

Dabei wird Gender als eine Technologie verstanden die es künstlicher Intelligenz ermöglicht zu funktionieren. Ausgehen von einem Genderbegriff, der sich an Butlers Idee von Gender als Norm und Barads Konzeption des agentiellen Realismus orientiert, wird binäre Zweigeschlechtlichkeit, als Funktion und nicht als empirischer Fakt der KI Anwendung sichtbar. Dadurch werden Handlungsmöglichkeiten für ein demokratisches Technikdesign möglich, die sich vorrangig auf intersektionale feministische Intervention rund um KI beziehen, wie Community basiertes Benchmark Design.

B.) Abstract

This thesis examines the question: how gender is integrated into gender classifiers and which political science interpretation these results allow. The Gendershades Project is analyzed as a case study. In a three-step process, (a) ethical objectives of the Gendershades project, so-called FAT algorithms, (b) the functionality of automatic facial recognition applications and (c) the division of labor within the Gendershades project are analyzed. Gender is understood as a technology that

enables artificial intelligence to function. Focusing on a gender concept that is based on Butler's gender as the norm and Barad's idea of agential realism, gender binary becomes visible as a function and not as an empirical fact of the AI application. This enables options for action for a democratic technology design that primarily relate to intersectional feminist intervention around AI, such as community-based benchmark design.