universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Modifikationen sozialer Ungleichheit durch KI“

verfasst von / submitted by

Charlotte Fleischmann, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien, 2022 / Vienna 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 589
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Internationale Entwicklung
Betreut von / Supervisor:	Dr. Johannes Knierzinger

Zusammenfassung

Schlagwörter: Künstliche Intelligenz; Algorithmen; soziale Ungleichheit; Ungleichheitsforschung; Dimensionen der Ungleichheit; Ethnizität; Gender; Klasse; Staatszugehörigkeit

Für diese Arbeit bringe ich Erkenntnisse der Gesellschafts-Technik-Forschung zu KI mit denen der Ungleichheitsforschung zusammen. Die interdisziplinäre Betrachtung eröffnet neue Perspektiven und wird der Vielzahl an Anwendungsbereichen sowie der Verwobenheit von KI mit der Gesellschaft gerecht. Systeme künstlicher Intelligenz funktionieren datenbasiert, das heißt sie greifen auf von Menschen generierte Daten zurück. Zudem besteht die der Technologie inhärente Verwobenheit durch die Anwendung der Technologie in gesellschaftsrelevanten Bereichen und deren Wechselwirkungen mit dem Menschen.

Die Beantwortung der Forschungsfrage <*Wie modifiziert KI die Ungleichheitsdimensionen Ethnizität, Gender, Klasse und Staatszugehörigkeit?*> basiert auf einer Literaturarbeit und untersucht Thesen, Konzepte und Anwendungsbeispiele von KI-Systemen der letzten vier Jahre. Entlang des entwickelten Kategoriensystems ordne ich die gesammelte wissenschaftliche Sekundärliteratur den jeweiligen Kategorien qualitativ-interpretativ zu.

Abstract

Keywords: Artificial intelligence; algorithms; social inequality; inequality research; dimensions of inequality; ethnicity; gender; class; nationality

For this work, I bring together findings from Science and Technology studies AI with those from inequality research. The interdisciplinary approach opens up new perspectives and does justice to the multitude of application areas and the interconnectedness of AI with society. Artificial intelligence systems function in a data-based manner, i.e., they draw on data generated by humans. In addition, there is the inherent interconnectedness of the technology through the application of the technology in areas relevant to society and its interactions with humans.

The answer to the research question <*How does AI modify the inequality dimensions of ethnicity, gender, class and nationality?*> is based on a literature review and examines theses, concepts and application examples of AI systems from the last four years. Along the developed category system, I qualitatively-interpretatively assign the collected academic secondary literature to the respective categories.

Inhaltsverzeichnis

1. Einleitung

Die Verselbständigung von Algorithmen oder: Wie Mystifizierung uns davon abhält, den Fokus auf akute gesellschaftlich relevante Themen zu behalten.

Künstliche Intelligenz (=KI) ist – egal ob als automatisierte Gesichtserkennung, bei Vorhersagen oder anderen Formen des Einsatzes – ein Instrument, welches in seinem Zweck von Menschen bestimmt wird. Die Tendenz, der Funktionsweise von Algorithmen ein autonomes Eigenleben zuzuschreiben, zeigt sich in den filmischen Aufarbeitungen von KI. Sie machen den Eindruck, dass KI nicht nur komplett eigenmächtig handelt, sondern sich zusätzlich auch von einem auf den anderen Moment verselbstständigen kann. Filme wie „Lucy“, „Matrix“ oder „I, Robot“ rücken die Verselbständigung von KI in den Vordergrund. Sie laden dadurch Begriffe wie Roboter und KI mit Emotionen auf. Während wir also Filme anschauen, in denen Roboter die Welt übernehmen und danach bei hitzigen Diskussionen darüber rätseln, ob und wann eine verselbstständigte KI die Menschheit auslöscht, verpassen wir es, über aktuelle, gesellschaftsrelevante Themen in Bezug auf KI zu sprechen.

Relevanz der Forschungsfrage

KI ist ein Thema, das in den letzten Jahren immer mehr Aufmerksamkeit bekommen hat (vgl. European Commission 2020, S.1). Im akademischen Bereich forschen Wissenschaftler*innen nicht nur an der Entwicklung der Systeme, sondern setzen sich aus gesellschaftswissenschaftlicher Perspektive mit den Wirkmechanismen von KI auseinander (vgl. Ligo et al. 2021, o.S).

KI-Systeme können nicht nur große Mengen von Daten analysieren, sondern auch unstrukturierte Daten sowie unterschiedliche Typen von Daten verarbeiten (vgl. Stanford University 2022, o.S.). In Estland verteilt ein KI-System Familienleistungen. So können Eltern die Ihnen zustehenden Leistungen für ihr Kind beziehen, ohne einen Antrag stellen zu müssen. Dafür sammelt der Staat von Geburt an Informationen über jedes Neugeborene und seine Eltern (vgl. Algorithm Watch 2020b, S. 6). In Italien wird momentan ein KI-System entwickelt, welches Richter bei der Rechtsprechung unterstützen soll (ebd.). Hierfür werden Trends aus früheren Gerichtsentscheidungen zu einem bestimmten Thema analysiert und Empfehlungen abgeleitet (ebd.). Solche KI-Systeme gehören beispielsweise in den USA bereits zum Alltag (vgl. Gless et al. 2019, S.147ff). In Dänemark setzten Universitäten ein KI-System ein, welches – bis es

durch Studierenden Proteste gestoppt wurde – Tastatur- und Mausklicks während der Prüfungen auf den Computern der Studierenden überwachte (vgl. Algorithm Watch 2020b, S. 6).

KI-Systeme sind keine alleinstehenden technischen Artefakte. Wie in den vorigen Beispielen sichtbar ist, werden sie mit direktem Bezug zur Gesellschaft eingesetzt und besitzen Handlungs- und Wirkungsmacht (netzforma* e.V. 2020, S. 8). Die Einbettung von KI-Systemen in die Gesellschaft bedingt auch ihre Verwobenheit mit gesellschaftlichen Macht- und Ungleichheitsverhältnissen (vgl. Weber und Prietl 2021, S. 71). Ungleichheitsforschung untersucht soziale Ungleichheit nicht als objektive Tatsache der sozio-ökonomischen Lage, sondern analysiert das „aufeinander bezogene Handeln von Menschen“ (Behrmann et al. 2018, S.1ff). Typische Fragen der Ungleichheitsforschung sind: Wer profitiert von Handlungen und ihren Wirkmechanismen und wer nicht? Wer wird ausgeschlossen und wer wird mitgedacht?

Diese Forschungsarbeit verfolgt das Ziel KI und ungleichheitsrelevante Merkmale ihrer technisch-gesellschaftlichen Wirkmechanismen zu untersuchen. Folgende Fragen weisen die Richtung dieser Arbeit: Wer ist an der Entwicklung und dem Einsatz von KI beteiligt und wer nicht? Welche Perspektiven und wessen Interessen spielen bei der Entwicklung und dem Einsatz von KI-Systemen eine Rolle? Welche menschlich bestimmten Normen und Werte werden in KI-Systeme integriert und welche nicht?

Trotz der Aktualität und Relevanz, sind diese Fragen in den etablierten Wissenschaften noch nicht angekommen (vgl. Algorithm Watch 2020b, S. 6). Obwohl KI-Systeme Auswirkungen auf fast alle menschlichen Aktivitäten, wie beispielsweise die Verteilung von staatlichen und unternehmerischen Leistungen sowie auf den Zugang zu Rechten haben, sind diese Fragen mehrheitlich unbeantwortet (ebd.).

Im deutschsprachigen Raum entsteht erst seit einigen Jahren ein interdisziplinärer Gesellschafts-Technik-Forschungsbereich in dem die Wechselwirkungen zwischen Technologie und Gesellschaft analysiert werden (vgl. Lengersdorf und Weber 2019, S. 7). Die interdisziplinäre Erforschung, wie auch die Betrachtung von KI als ein sozio-technisches System, ist elementar. Denn, neben der technischen sind viele weitere Perspektiven, Bereiche und Sphären von Relevanz: KI umfasst technische, rechtliche, ethische und gesellschaftliche Themen (vgl. AI Watch 2020, S. 5). Während technische und rechtliche Perspektiven die Debatte um KI dominieren, gewinnt die

ethische Perspektive zwar an Aufmerksamkeit, die gesellschaftliche Perspektive bleibt jedoch marginal (vgl. Algorithm Watch 2020b, S. 84).

Die Betrachtung der Schnittstelle zwischen Ungleichheitsforschung und Gesellschafts-Technik-Forschung in Bezug auf KI ermöglicht die Analyse ungleichheitsrelevanter Merkmale technisch-gesellschaftlicher Wirkmechanismen von KI.

Davon ausgehend untersuche ich im Rahmen dieser Arbeit folgende Frage: <*Wie modifiziert KI die Ungleichheitsdimensionen Ethnizität, Gender, Klasse und Staatszugehörigkeit?*>

Vorgehensweise und inhaltliche Gliederung der Forschungsarbeit

Für die Beantwortung der Forschungsfrage untersuche ich Argumente und Konzepte von Wissenschaftler*innen und Expert*innen, sowie praktische Anwendungsbeispiele von KI-Systemen auf Basis einer systematischen Literaturrecherche. Ergänzend zu wissenschaftlicher Sekundärliteratur, habe ich bei Bedarf Kontextmaterial hinzugezogen. Beispielsweise wenn die Sekundärliteratur keine ausreichenden Informationen über ein spezifisches KI-System gegeben hat. Bei dem Kontextmaterial handelt es sich um Studien oder Berichte von NGOs sowie um Zeitungsartikel.

Die Untersuchung umfasst den gesamten Anwendungsbereich von KI. Somit stammen Literatur und Anwendungsbeispiele aus unterschiedlichen Bereichen und zeigen vielfältige Perspektiven auf. Eine Forschungsfrage zu einem einzelnen Anwendungsbereich von KI oder aus der Perspektive einer einzelnen wissenschaftlichen Disziplin würde den übergeordneten Diskurs zu KI beschneiden. Zudem wären komplexe Argumente, die mehrere Forschungsbereiche umfassen, weniger gut einzuordnen.

Ausgehend von der systematischen Literaturrecherche habe ich manche Kategorien direkt aus der Literatur übernommen. Andere Kategorien haben sich auch aus der Zusammenschau der Literatur ergeben.

Ich untersuche entlang der Ungleichheitsdimensionen Ethnizität, Gender, Klasse und Staatszugehörigkeit die aus der Literatur abgeleiteten Kategorien. Ich habe das Kategoriensystem qualitativ-interpretativ entwickelt. Die Kategorien beschreiben, wie die jeweiligen Ungleichheitsdimensionen durch KI modifiziert werden. In der Zusammenfassung der Ergebnisse beantworte ich die Forschungsfrage und gehe auf

die Unterschiede und Zusammenhänge der jeweiligen Modifikationen der vier Dimensionen der Ungleichheit ein. Im Fazit halte ich weitere Erkenntnisse der Arbeit fest.

Im Kapitel 2, Methodik schildere ich nicht nur das wissenschaftliche Vorgehen beim Schreiben der Arbeit. Ich beschreibe auch meine eigene Positionalität, den definitorischen Zugang zu sozialer Ungleichheit und künstlicher Intelligenz, sowie die Limitationen der Arbeit.

2. Methodik

Motivation, Themenfindung & Positionalität

Ich belegte in meinem Bachelorstudium Geographie die Spezialisierung Geoinformation. Hier hatte ich meine ersten Berührungspunkte mit Data Science. Meine Bachelorarbeit habe ich über die Bedeutung von raumbezogenen Daten für politische Kampagnen geschrieben. In meinem Masterstudium Internationale Entwicklung belegte ich ein Individualmodul über die digitale Transformation. Im Rahmen dieses Individualmoduls habe ich eine Seminararbeit über Chancen und Herausforderungen der digitalen epidemiologischen Überwachung geschrieben. Darüber hinaus habe mich bereits mehrere Jahre mit dem Thema KI beschäftigt. Im April 2021 veröffentlichte die EU den ersten Vorschlag für die Regulierung von KI mittels EU-Rechtsakten. Im Rahmen des ersten Entwurfs haben viele Panel-Diskussionen und andere öffentlich zugängliche Veranstaltungen stattgefunden, an denen ich teilgenommen habe. Zu dieser Zeit begann ich meine Auseinandersetzung mit KI aus einer entwicklungspolitischen Perspektive. Um die Modifikationen sozialer Ungleichheit durch KI zu untersuchen, ist es wichtig, die Wirkmechanismen zu verstehen. Dafür fehlte mir zwar das Informatik- oder KI-Studium, jedoch habe ich mich mit Hilfe von Literatur und Universitäts-Vorlesungen in das Thema eingearbeitet. Weiters zu meiner Positionalität: Ich bin in Deutschland geboren, weiß und weiblich.

Wissenschaftliches Vorgehen

Ausgehend von der Forschungsfrage <*Wie modifiziert KI die Ungleichheitsdimensionen Ethnizität, Gender, Klasse und Staatszugehörigkeit?*> und der Literaturrecherche entschied ich mich für eine Literaturarbeit als zweckgemäße Methode. Angelehnt an Mayring und Fenzl entwickelte ich ein Kategoriensystem, in welchem die Modifikationen von KI auf soziale Ungleichheit zusammengefasst sind

(Mayring und Fenzl 2014, S. 552). Die Auswahl der untersuchten Dimensionen der Ungleichheit (Ethnizität, Gender, Klasse und Staatszugehörigkeit) erfolgte in Anlehnung an Hradil, der die Begrifflichkeit der „Dimensionen der sozialen Ungleichheit“ (Hradil 2016, S.247ff) prägte. Mit der Unterscheidung zwischen Klasse und Staatszugehörigkeit ist die Unterscheidung zwischen binnenstaatlicher, also materieller Ungleichheit in Staaten und internationaler Ungleichheit, also Ungleichheit zwischen Staaten gemeint (vgl. Dannecker, Gächter et al. 2010, S.53). Diese Unterscheidung entspringt der Ungleichheitsforschung (ebd.). Die Unterscheidung ist für die Forschungsarbeit zweckgemäß, weil sie der Unterschiedlichkeit der Ausprägungen materieller Ungleichheit in Bezug auf den Analyserahmen binnenstaatlicher Ungleichheit und internationaler Ungleichheit gerecht wird (vgl. Milanovic 2009, S.2ff). Entlang der Dimensionen der Ungleichheit führe ich Konzepte, Anwendungsbeispiele und Studien aus der Literaturrecherche auf, welche die Modifikationen der jeweiligen Ungleichheitsdimension aufzeigen. Diese sind einzelnen Kategorien zugeordnet. Der Titel der Kategorie beschreibt den Kern der jeweils zusammengeführten Argumente und Konzepte. Um die Forschungsarbeit einzugrenzen habe ich das Kategoriensystem und die entwickelten Kategorien definiert und die Analyseeinheit der Literaturrecherche festgelegt (Mayring und Fenzl 2014, S. 547). Die Festlegung der Analyseeinheit umfasst eine Beschreibung der Kodiereinheit, der Auswertungseinheit und der Kontexteinheit. Diese Definitionen und Festlegungen führe ich in den nächsten Absätzen aus.

Als Kodiereinheit der Forschungsarbeit stellt eine semantische Einheit die zweckgemäße Wahl dar: Bereits die erste Sichtung der wissenschaftlichen Sekundärliteratur aus der systematischen Recherche hat gezeigt, dass die untersuchten Texte teilweise seitenlange Beschreibungen eines relevanten Arguments oder Konzepts enthielten. Um durch die Auswahl der Kodiereinheit nicht die Auswahl der für die Beantwortung der Forschungsfrage relevanten Argumente zu limitieren, habe ich die „semantische Einheit“ als Kodiereinheit gewählt. Somit habe ich eine umfassende Erforschung von Argumenten und Konzepten unterschiedlicher Ausprägungen vornehmen können. Akremi zur Folge macht die Auswahl der semantischen Einheit die Analyse weniger sensibel, weil die Entscheidung, welche Informationen Teil der semantischen Einheit sind und welche nicht, dem oder der Forscher*in überlassen sind (vgl. Akremi 2014, S. 265ff).

Die Auswertungseinheit, also der Korpus der Arbeit umfasst die vorgenommene systematische Literaturrecherche. Bei den gesammelten Argumenten und Konzepten handelt es sich somit um wissenschaftliche Sekundärliteratur. Die einzelnen Argumente können aus demselben Text stammen und trotzdem unterschiedlichen Kategorien zugeordnet werden. Somit sind Mehrfachzuordnungen von sich überschneidenden Argumenten und Konzepten eines Texts, also beispielsweise Textstellen, die unterschiedliche Argumente darstellen, zu unterschiedlichen Kategorien zulässig. Die Vorgehensweise für die Literaturrecherche erläutere ich im folgenden Absatz.

In einer ersten Sichtung habe ich Basisliteratur zum Thema KI gelesen und zusammengefasst. Nach dem Festlegen der Forschungsfrage <*Wie modifiziert KI die Ungleichheitsdimensionen Ethnizität, Gender, Klasse und Staatszugehörigkeit?*> und der Auswahl der Literaturarbeit als zweckgemäße Methode, habe ich eine systematische Literaturrecherche vorgenommen. Mit den Suchbegriffen „Ungleichheit; Repräsentativität; Internet of Things; Vernetzung; Big Data; Industrie 4.0; algorithmische Diskriminierung; Rassismus; KI; internationale Ungleichheit; intelligente Technologien; digitale Kluft; Technologie“ habe ich Literatur aus dem Zeitraum zwischen 2018 und 2022 herausgefiltert und nach relevanten Argumenten, Studien und Konzepten durchsucht. Bei den Datenbanken für die Literaturrecherche handelte es sich um die Datenbank der Universität Wien, die Datenbank der österreichischen Nationalbibliothek, dem „web of science“ und „google scholar“.

Wenn ich in der Literatur Argumente und Konzepte fand, welche für die Beantwortung der Forschungsfrage nützlich waren, ordnete ich diese der jeweiligen Kategorie zu. Ich habe entweder konkrete Anwendungsbeispiele von KI-Systemen gesucht, die exemplarisch eine Modifikation der jeweiligen Ungleichheitsdimension durch KI anzeigen. Oder ich ging Argumenten und Konzepten nach, die den Wirkmechanismus zwischen KI und sozialer Ungleichheit thematisierten. Hier waren nicht nur technische Funktionsweisen von KI, sondern in Anlehnung an den definitorischen Zugang von KI als sozio-technisches System, auch die Wechselwirkungen, die Einbettung und die Entwicklung der technischen Systeme durch und in Wechselwirkung mit der Gesellschaft relevant (vgl. Algorithm Watch 2020b, S. 84; Lücking 2020, S. 73). Ich habe alle Argumente und Konzepte aus der wissenschaftlichen Sekundärliteratur, welche die Modifikationen von KI auf die Ungleichheitsdimensionen Ethnizität, Gender,

Klasse und Staatszugehörigkeit thematisieren, in einem ersten Kategorien-Entwurf gesammelt. Weiters habe ich Quellenangaben aus den semantischen Einheiten der jeweiligen recherchierten Literatur nachverfolgt. Einerseits um die Quellen zu überprüfen, andererseits um die Argumente und Konzepte für die Kategorien zu ergänzen.

Bei Bedarf, habe ich die Recherche um Kontextmaterial erweitert. Das Kontextmaterial umfasste neben wissenschaftliche Sekundärliteratur auch Studien und Berichte von NGOs, Veröffentlichungen von überstaatlichen Institutionen sowie Zeitungsartikel. Die recherchierte Kontextliteratur reicht über das Jahr 2018 hinaus, weiter zurück. Im Fall von Donna Haraway beispielsweise bis in das Jahr 1988. Für Kontextmaterial, welches die Anwendung von KI-Systemen thematisiert, habe ich jedoch nicht auf Literatur, die vor 2017 veröffentlicht wurde, zurückgegriffen. Aufgrund des dynamischen Fortschritts von KI würde Literatur von vor 2017 nicht mehr dem aktuellen technologischen Entwicklungsgrad entsprechen (vgl. Algorithm Watch 2020b, S. 6). Für die Kontexteinheit habe ich mich auf vollständige Dokumente – was zusammengehörige Absätze, die in der Regel aus einer Kommunikationsquelle oder Entstehungssituation stammen, einschließt (vgl. Mayring 2014, S.32), – festgelegt. Bei Bedarf nach mehr Informationen wurde diese Festlegung auf ein Dokument um zusätzliches Kontextmaterial erweitert. Das ist nur bei den Veranschaulichungen der wissenschaftlichen Sekundärliteratur mittels Anwendungsbeispiele von KI-Systemen der Fall. Ausführliche Beschreibungen der Funktionsweise dieser exemplarisch aufgeführten KI-Systeme habe ich meistens auf Webseiten der produktanbietenden Unternehmen, in wissenschaftlichen Studien zu den jeweiligen KI-Systemen oder in NGO-Berichten gefunden. Die zwei NGOs, deren Studien ich als Kontextmaterial benutze, beschreibe ich in einem eigenen Absatz, in Kapitel 2, Methodik.

Zusammenfassend handelt es sich bei der Analyseeinheit der Arbeit um wissenschaftliche Sekundärliteratur aus einer systematischen Literaturrecherche. Ausgehend von diesem Textkorpus leitete ich das Kategoriensystem ab. Zusätzlich band ich bei Bedarf Kontextmaterial ein. Das Kontextmaterial ergänzt die sekundärwissenschaftliche Literatur. Das Kontextmaterial umfasst NGO-Berichte, NGO-Studien, Zeitungsartikel und Webseiten von Unternehmen, die KI-Systeme anbieten.

Entlang des im folgenden Absatz festgehaltenen Kategoriensystems beantworte ich die Forschungsfrage in Kapitel 4, Zusammenfassung der Ergebnisse. Je nach Abstraktionsniveau ist ein Teil des Kategoriensystems auch durch das Inhaltsverzeichnis wiedergegeben.

Modifikationen der Ungleichheitsdimension Gender durch KI
3.1.1 Datenbasiertheit von KI und Verzerrungen von Algorithmen durch Daten
a. Verzerrungen durch Unsichtbarkeit mancher Gruppen in Daten – digitale Kluft
b. Einschreiben bestehender genderspezifischer Diskriminierung in Algorithmen durch Daten
3.1.2 Algorithmische Kodierung von Realität basiert auf Kategorisierung und Normierung
3.1.3 Ergebnisse von KI-Systemen sind kein situiertes Wissen
3.1.4 Fehlende Gerichtsbarkeit und Regulierung - KI als Blackbox
3.1.5 Kann KI sozialer Ungleichheit entgegenwirken?

Modifikationen der Ungleichheitsdimension Ethnizität durch KI
3.2.1 Datenbasiertheit von KI und Verzerrungen von Algorithmen durch Daten
a. Verzerrungen durch Unsichtbarkeit mancher Gruppen in Daten – digitale Kluft
b. Einschreiben bestehender rassisitischer Diskriminierung in Algorithmen durch Daten
3.2.2 Algorithmische Kodierung von Realität basiert auf Kategorisierung und Normierung – Die Norm ist der weiße Mann
3.2.3 Überwachung durch KI-Systeme betrifft marginalisierte Gruppen stärker
3.2.4 Fehlende Gerichtsbarkeit und Regulierung - KI als Blackbox
3.2.5 Kann KI sozialer Ungleichheit entgegenwirken?

Modifikationen der Ungleichheitsdimension Klasse durch KI
3.3.1 Datenbasiertheit von KI und Verzerrungen von Algorithmen durch Daten
3.3.2 Überwachung durch KI-Systeme betrifft marginalisierte Gruppen stärker
3.3.3 Ideologische Einbettung von KI in kapitalistische Prinzipien
3.3.4 Arbeitsverlust durch KI?

Modifikationen der Ungleichheitsdimension Staatsbürgerschaft durch KI
3.4.1 Datenbasiertheit von KI und Verzerrungen von Algorithmen durch Daten
3.4.2 Internationaler Wettbewerb um KI
3.4.3 KI als Projekt reicher Länder

Positionalität der NGOs und einer weiteren Organisation

Im folgenden Absatz will ich die zwei NGOs, deren Studien und Berichte als Kontextmaterial die Arbeit ergänzen, kurz vorstellen:

MarkUp ist eine gemeinnützige Organisation, die investigativen Journalismus auf Basis datengestützter Untersuchungen betreibt (vgl.The MarkUp 2022). MarkUp

veröffentlicht zusätzlich zu ihren Studien die jeweilige Methodik, die verwendeten Daten und ihre Herkunft sowie die bei der Analyse verwendeten statistischen Verfahren. Sie arbeiten international und recherchieren als auch erforschen, wie mächtige Institutionen Technologien nutzen. Eines ihrer Projekte war beispielsweise in Zusammenarbeit mit der Süddeutschen Zeitung: Sie untersuchten den Einfluss von Facebook auf die Bundestagswahl im September 2021 (vgl. The MarkUp 2022). Ich greife für die Veranschaulichung eines Beispiels auf eine Studie von MarkUp zurück. In dem Beispiel beschreibe ich Zusammenhänge von Vorwürfen rassistischer Diskriminierung und der Anwendung des KI-Systems „Navigate“ an US-amerikanischen Universitäten.

Die gemeinnützige Organisation Algorithm Watch besteht seit 2015 und bezeichnet sich als eine evidenzbasierte Advocacy-Organisation (vgl. Algorithm Watch 2022). Sie setzt sich für einen Einsatz von KI-Systemen ein, der Mensch und Gesellschaft nützt, und nicht schadet. Sie analysieren die Auswirkungen algorithmischer Entscheidungs-findungsprozesse und zeigen ethische Konflikte auf. Zudem unterstützen sie die Entwicklung von Ideen und Strategien, „die eine Nachvollziehbarkeit dieser Prozesse ermöglichen - mit einem Mix aus Technologie, Regulierung und geeigneten Aufsichtsinstitutionen“ (Algorithm Watch 2022). In einer großen Studie, untersuchten Algorithm Watch und ihre Projektpartner die Anwendung von KI-Systemen in 16 Ländern der EU. Alle KI-Systeme, die von den Regierungen der jeweiligen Länder eingesetzt werden, wurden in dem Bericht auf ihre Auswirkungen untersucht. In der abschließenden Publikation, dem „Automating Society Report“ (Algorithm Watch 2020b) hält Algorithm Watch die Ergebnisse ihrer mehrjährigen Forschung fest. Auf die Ergebnisse des im Jahre 2020 veröffentlichten Berichts greife ich zur Veranschaulichung von Argumenten und Konzepten aus der wissenschaftlichen Sekundärliteratur ergänzend zurück.

Weiters stelle ich das Gremium „AI Watch“ der EU vor. Das Gremium wird von der Europäischen Kommission zur Überwachung von KI eingesetzt (vgl. AI Watch 2020, S. 3). Die Zuständigkeitsbereiche der Expert*innen von AI Watch sind erstens, das Beobachten des Einsatzes von KI weltweit mit einem Fokus auf die europäischen Mitgliedsstaaten. Zweitens, das Abschätzen des Einflusses der Systeme und drittens, die (Weiter-)Entwicklung der Technologie (vgl. AI Watch 2020, S. 3). Um diese Aufgaben zu meistern, haben sie KI im Rahmen eines 2020 veröffentlichten 90 Seiten

langen Berichts definiert. Die Definition schafft ein Verständnis und strukturiert die Möglichkeiten des Einsatzes der Technologie. Der Bericht basiert auf einer Studie, in der mithilfe eines KI-Systems eine große Menge Literatur analysiert wurde. Die Literaturrecherche umfasst somit Definitionen von KI aus den Anwendungsfeldern Politik, Forschung und Wirtschaft (vgl. AI Watch 2020, S. 3). Ich nutze die Ergebnisse der Studie in Kapitel 2, Methodik, weil sie einen Überblick über die Themen- und Einsatzbereiche von KI sowie einen definitorischen Zugang für diese Arbeit geben.

Limitationen

Im folgenden Absatz schildere ich zwei Limitationen der Arbeit. Die Logik des Intersektionalitäts-Ansatzes steht dem der voneinander getrennt betrachteten Kategorien entgegen: Eine intersektionale Betrachtung von Ungleichheit bedeutet, dass Ungleichheitsdimensionen weder nebeneinander bestehen noch additiv subsumiert werden können (vgl. Aulenbacher und Riegraf 2012, S.1). Sie beeinflussen sich wechselseitig und wirken somit in jeder Konstellation individuell. (ebd.) Jalloh zufolge bedeutet das, dass eine bestimmte Kombination von Ungleichheitsdimensionen durch den sozialen Kontext einer Person auf bestimmte Erfahrungen von Diskriminierung verstärkend wirken kann (vgl. Jalloh 2022, S.48). Dadurch steht Intersektionalität im Gegensatz zu den Kategorien, in denen die Modifikationen von KI auf Gender, Ethnizität, Klasse und Staatszugehörigkeit getrennt voneinander betrachtet werden. Eine weitere Schwachstelle der Methodik ist das Auslassen von Ungleichheitsdimensionen und der damit verbundenen Auslassung von Formen der Marginalisierung und Diskriminierung. Ich habe Ableismus, Antisemitismus, Ageismus oder LGBTQIA+-Feindlichkeit in dieser Arbeit nicht untersucht.

In den folgenden Absätzen erläutere ich die theoretischen Hintergründe, sowie die gewählten definitorischen Zugänge, der in der Forschungsfrage vorkommenden Begriffe KI und soziale Ungleichheit. Weil sich keiner der Begriffe auf ein Theoriekonstrukt bezieht, sondern es sich um Konzepte und Forschungszugänge handelt, beschreibe ich sie als „definitorische Zugänge“.

Definitorischer Zugang zu dem Begriff der sozialen Ungleichheit

Christian Suter definiert soziale Ungleichheit als „ungleichen Zugang von Menschen und sozialen Gruppen zu gesellschaftlich hoch bewerteten sozialen Gütern“ (Dannecker, Gächter et al. 2010, S.50). In der Ungleichheitsforschung gibt es verschiedene Zugänge zu sozialer Ungleichheit. Für diese Arbeit habe ich in Anlehnung an Hradil die „Dimensionen der sozialen Ungleichheit“ als definitorischen Zugang gewählt (vgl. Hradil 2016, S.247ff). Jedes Individuum erfährt Begünstigungen oder Benachteiligungen durch Zugehörigkeiten zu sozialen Kategorien (vgl. Dannecker, Gächter et al. 2010, S.52). Es ist wichtig anzuerkennen, dass in einem Individuum unterschiedliche Ungleichheitsdimensionen zusammenkommen und miteinander wechselwirken. Die Individualität von Ungleichheits-Verschränkungen in jedem Individuum ist ein wichtiges Merkmal des Begriffs der Dimensionen sozialer Ungleichheit: Eine hierarchische Schichtung von sozialer Ungleichheit bei der Beschreibung von Wechselwirkungen und Verschränkungen der verschiedenen Formen von sozialer Ungleichheit greift zur kurz. Die Bezeichnung „Dimensionen“ begreift soziale Ungleichheiten als intersektional. Die Dimensionen der Ungleichheit bestehen weder nebeneinander, noch lassen sie sich additiv subsumieren. Stattdessen beeinflussen sie sich wechselseitig und wirken somit in jeder Konstellation individuell (vgl. Aulenbacher und Riegraf 2012, S.1). Kimberlé Crenshaw, die den Begriff der Intersektionalität prägte, verstand soziale Ungleichheiten als Achsen der Ungleichheit, [sex (später gender), race und class] welche durch ihre Überkreuzungen Menschen gesellschaftlich ungleich positionieren (vgl. Crenshaw 2018, S.57ff). In Anlehnung an diese Definition, grenze ich für die Beantwortung der Forschungsfrage, meine Betrachtung auf die vier Ungleichheitsdimensionen Ethnizität, Gender, Klasse und Staatszugehörigkeit ein.

Die Dimensionen der Ungleichheit, die im Rahmen der Arbeit untersucht werden, teilen sich in horizontale und vertikale soziale Ungleichheiten auf. Zu den horizontalen Ungleichheiten zählt Ethnizität und Gender. Diese beiden Dimensionen „bezeichnen zugeschriebene Merkmale sozialer Ungleichheit, [womit] individuell unbeeinflussbare Merkmale“ (Dannecker, Gächter et al. 2010, S.52) gemeint sind. Sie werden auch als „neue Ungleichheiten“ bezeichnet, weil sie in der Ungleichheitsforschung erst später den Stellenwert erfahren, der ihnen aktuell zukommt (vgl. Kreckel 1983, S. 7). Sie basieren auf der gesellschaftlichen Konstruktion von Ethnizität und Gender und sind

verschränkt mit rassistischer und genderspezifischer Diskriminierung (vgl. Kohli et al. 2000, S. 318).

Vertikale soziale Ungleichheiten sind individuell veränderbare Merkmale (vgl. Dannecker, Gächter et al. 2010, S.52). In diese Kategorie fällt die Ungleichheitsdimension Klasse, welche im Rahmen dieser Arbeit einerseits als materielle Ungleichheit zwischen Individuen innerhalb von Staaten und andererseits als materielle Ungleichheit zwischen Staaten untersucht wird. Die Ungleichheit innerhalb von Ländern, wird als intranationale oder binnenstaatliche Ungleichheit bezeichnet (vgl. Greve 2010, S. 64ff). Die Unterscheidung ist für die Forschungsarbeit zweckgemäß, weil sie der Unterschiedlichkeit von Ausprägungen materieller Ungleichheit in Bezug auf den Analyserahmen binnenstaatlicher materieller Ungleichheit und internationaler Ungleichheit gerecht wird. Die Relevanz dieser Unterscheidung liegt darin begründet, dass durch die Betrachtung von internationaler Ungleichheit die soziale Ungleichheit innerhalb eines Landes unsichtbar bleibt. Milanovic zeigte mittels Studien zur Ungleichheitsforschung auf wie unterschiedlich die Ergebnisse sozialer Ungleichheit je nach Analyserahmen sind (vgl. Milanovic 2009, S.2ff). Er bewies dadurch die Relevanz der getrennten Untersuchung von inter- und intranationaler Ungleichheit.

Im Rahmen dieser Arbeit können die Modifikationen globaler Ungleichheit durch KI nicht untersucht werden, da sie in der recherchierten Literatur nicht ausreichend abgebildet waren. Die soziale Ungleichheit zwischen Ländern unterscheidet sich von der globalen Ungleichheit wie folgt: Die internationale Ungleichheit nutzt den Nationalstaat als Bezugsrahmen für die sozialen Ungleichheiten. Bei globaler Ungleichheit bilden einzelne Haushalte den Referenzrahmen (vgl. Dannecker, Gächter et al. 2010, S.53).

Definitorischer Zugang zu Künstlicher Intelligenz

KI ist aufgeladen mit Ideen von Science-Fiction-Romanen und Filmen. Oft verschwimmen Mythos und Realität auch im wissenschaftlichen Diskurs (vgl. Unesco Courier 2018, S. 7). Eine präzise Benennung der Technologie ist deshalb wichtig. Algorithm Watch spricht beispielsweise von automatisierter Entscheidungsfindung um sich von dem verschwommenen Diskurs von KI abzugrenzen (vgl. Algorithm Watch 2020a, S. 4). Ich habe mich trotzdem für den Begriff KI entschieden, weil der wissenschaftliche Diskurs unter dem Schlagwort KI relevant ist und wissenschaftliche

Arbeiten, die sich gegen die Nutzung des Begriffes entscheiden, zwar Teil des Diskurses sind, aber von Suchfiltern übersehen werden. Um bei der Benennung zu spezifizieren, ob es um eine konkrete Anwendung von KI oder um den allumfassenden generischen Begriff geht, beziehe ich mich auf KI-Systeme, wenn ich von einer Software künstlicher Intelligenz spreche. Wenn KI als Begriff alleinstehend ist, meine ich den gesamten Bereich KI inklusive die im nächsten Absatz beschriebenen Teil- und Querschnittsthemen. Für diese Arbeit habe ich mich zudem für den definitorischen Ansatz entschieden, welcher KI nicht als ein alleinstehendes technisches System begreift, sondern als ein sozio-technisches System, welches sowohl die gesellschaftliche Reflektion und Aushandlung während des Entwicklungs- und Implementierungsprozesses, wie auch die Funktionsweise und die Wirkmechanismen eines KI-Systems umfasst (vgl. Algorithm Watch 2020b, S. 84).

In den folgenden Absätzen stelle ich Thesen und Konzepte vor, welche für die Beantwortung der Forschungsfrage relevant sind: Für einen Überblick in das Feld von KI stelle ich eine Taxonomie im Absatz „*KI als generischer Begriff“* vor. Dann zeige ich die Funktionen im Absatz „*Die Funktion von KI“* und die Anwendungen im Absatz „*Die Anwendungsbereiche von KI-Systemen“* auf. Daraufhin grenze ich den gewählten definitorischen Zugang von allgemeiner künstlicher Intelligenz im Absatz zu „*Künstliche allgemeine Intelligenz und KI*“ ab. Abschließend erläutere ich den gewählten definitorischen Zugang zu künstlicher Intelligenz im Absatz „*KI als sozio-technisches System“*.

KI als generischer Begriff

KI umfasst mehrere Anwendungsfelder und Themen. Diese hängen auf unterschiedliche Weise miteinander zusammen. Die folgende Strukturierung der Bereiche von KI ist angelehnt an die Taxonomie für KI von AI Watch und gibt einen Überblick über die Kern- und Querschnittsbereiche von KI. Die Kernbereiche, welche die Ziele von verschiedenen KI-Systemen beschreiben, sind „Begründen“, „Planen“, „Lernen“, „Kommunizieren“ und „Wahrnehmen“. Die Querschnittsthemen oder Anwendungsfelder umfassen keine Ziele, die in dem Einsatz eines spezifischen KI-Systems fokussiert wird, sondern Themen an denen KI teilhat oder Anwendungen in denen KI mehrfach zum Einsatz kommt. Als Querschnittsbereiche zählen „Integration & Interaktion“, „Dienstleistungen“ sowie „Ethik & Philosophie“ (frei übersetzt nach AI Watch 2020, S. 5). Diese Strukturierung gibt einen ersten Eindruck, warum ich KI als

einen generischen Begriff beschreibe: Eine Vielzahl unterschiedlicher Themen und Anwendungsfelder können mit dem Begriff angesprochen werden. Bei der Strukturierung ist es jedoch problematisch, dass sie den Eindruck vermittelt die einzelnen Bereiche wären auch in der Realität voneinander abgegrenzt (vgl. AI Watch 2020, S. 12). Hier beweist das Beispiel des autonomen Fahrens das Gegenteil: Verschiedene Kernthemen und Anwendungsfelder von KI kommen in einem selbstfahrenden Auto zusammen (vgl. Nolting 2021, S. 113ff): Das „Wahrnehmen“, wie auch das „Lernen“ und „Planen“. Zugleich wird hinsichtlich des autonomen Fahrens das Trolley-Problem als ethisches Dilemma diskutiert und ist somit dem Querschnittsbereich von KI „Ethik und Philosophie“ zuzuordnen (vgl. Sommaggio und Marchiori 2020, S. 91). Dieses Beispiel zeigt, dass die Bereiche von KI miteinander verwoben sind und dadurch nicht wie in der hier vorgestellten Gliederung klar voneinander abzugrenzen sind. Im Folgenden werden die oben aufgezählten Themen und Bereiche vorgestellt. Die Kernbereiche sind **fett gedruckt** und unterstrichen geben einen Eindruck über die Ziele und Fähigkeiten von KI. Die Querschnittsbereiche sind nur unterstrichen und zeigen miteinander verschränkte Bereiche von KI auf.

Begründen

Unter den Bereich des Begründens oder dem „Reasoning“ fallen Systeme künstlicher Intelligenz, die aus Daten, Informationen und Wissen ableiten. Das kann unterschiedlich passieren: Meistens handelt es sich um eine Darstellung von Wissen oder eine automatisierte Form des Schlussfolgerns/Begründens.

Planen

Wie der Begriff schon verrät, handelt es sich um die automatisierte Planung und/oder Ausführung von Strategien und Tätigkeiten. Das „Planning“ von einem KI-System kommt meistens bei komplexen Kontroll- und Klassifizierungsproblemen zum Einsatz, beispielsweise beim autonomen Fahren oder in der Robotik (vgl. AI Watch 2020, S. 12).

Lernen

Das Lernen oder “Learning” bezieht sich hauptsächlich auf den Bereich des maschinellen Lernens, dem „Machine Learning“. AI Watch bezeichnet das Ziel des automatisierten Lernens nicht als Teilbereich von KI, sondern erkennt ihn als eine der generellen Grundlagen von allen KI-Systemen an. Der Vorgang des

Lernens beinhaltet, Entscheidungen zu treffen, Vorhersagen zu machen, adaptiv auf einen sich verändernden Kontext reagieren zu können und/oder sich durch Erfahrung zu optimieren (vgl. AI Watch 2020, S. 9). In einem Strategiebericht der Europäischen Kommission von 2018 ist „Machine Learning“ die am weitesten verbreitete Technik (vgl. Europäische Kommission 2018a, S. 18).

Kommunikation

Im Bereich Kommunikation finden sich Techniken des „Natural Language Processing“ wieder. Hierbei geht es darum, dass ein System künstlicher Intelligenz menschliche Sprache und Schrift versteht, verarbeitet und/oder wiedergibt (vgl. AI Watch 2020, S. 12–13).

Wahrnehmung

Der Bereich des „Wahrnehmens“ bezieht sich auf Systeme Künstlicher Intelligenz, die ihre Umgebung erfassen, beispielsweise bei der Gesichts-, Körper- aber auch Sprach- oder Stimmerkennung. Vor allem die Bereiche des Sehens und des Hörens schreiten in der Entwicklung immer weiter fort. Die Systeme erkennen Muster in ihrer Umgebung und klassifizieren und/oder identifizieren sie (vgl. AI Watch 2020, S. 13).

Integration and Interaktion

Der übergreifende Bereich Integration und Interaktion fasst Kombinationen zuvor beschriebener Kernthemen zusammen, beispielsweise in der Robotik oder beim autonomen Fahren. Systeme künstlicher Intelligenz, die in diesen Bereich fallen, automatisieren komplexe Abläufe und Prozesse. Es werden dadurch neue Möglichkeitsräume geöffnet. Tätigkeiten, die zuvor dem Menschen vorbehalten waren, können von den Systemen übernommen werden (ebd.).

Services

Der Bereich „Services“ oder KI-Dienste beschreibt Systeme künstlicher Intelligenz, die es zum Ziel haben, die Verwaltung komplexer Infrastrukturen zu vereinfachen. Somit fallen Infrastrukturen, Software oder Plattformen unter die KI-Dienste, welche oft in Verbindung mit „Cloud-Computing“ Anwendung finden.

Cloud-Computing-Dienste bieten Server, Speicher und Netzwerkkapazitäten für Daten an (ebd.).

Ethik und Philosophie

Dieser Bereich gewinnt mit dem wachsenden Einsatz von Systemen Künstlicher Intelligenz immer mehr an Bedeutung und beschreibt Reflexionsprozesse dazu. AI Watch hat in der Studie zwei Themenbereiche herausgefiltert: Erstens ethische Grundsätze und Vorschriften für KI und zweitens Auswirkungen auf Mensch und Gesellschaft, wobei „die Schaffung von Vertrauen in KI im Mittelpunkt mehrerer Rahmenwerke und Initiativen von politischen Gremien und Institutionen [steht]“ (AI Watch 2020, S. 14). In diesem Bereich von KI kommen KI-Systeme selbst zwar nicht zur Anwendung, trotzdem wird der Bereich als ein Teil von KI begriffen. Das Inkludieren dieses Bereichs in den Bedeutungszusammenhang von KI steht für mich für die konzeptuelle Erweiterung der Technikbegriffs um die sozio-technische Perspektive: Technik wird hier nicht als alleinstehend, sondern immer in Zusammenhang mit der Gesellschaft betrachtet (vgl. Lücking 2020, S. 73) und kann damit als Bereich „Ethik und Philosophie“ auch ohne die Anwendung der Technik selbst fungieren.

Funktion von KI

Es unterscheiden sich nicht nur die Anwendungsbereiche, sondern auch KI als Technologie funktioniert unterschiedlich. KI-Systeme arbeiten als Gesichtserkennungssoftware, in Übersetzungsprogrammen oder in der Berechnung von Vorhersagen unterschiedlich (vgl. Kirste und Schürholz 2019, S.24ff). Im Rahmen dieser Arbeit ist eine Erklärung der Unterschiedlichkeit der technischen Funktionsweisen nicht relevant, jedoch soll in den nächsten Punkten kurz gezeigt werden, wozu KI-Systeme auf Basis ihrer unterschiedlichen technischen Funktionsweise fähig sind. Hierfür greife ich ein weiteres Mal auf die Studie von AI Watch zurück, in der einerseits mithilfe einer KI eine große Menge Literatur analysiert wurde, andererseits konnten durch qualitative Methoden Schlüsselliteratur von Expert*innen und Definitionen aus den Anwendungsfeldern Politik, Forschung und Wirtschaft miteinbezogen werden (vgl. AI Watch 2020, S. 3). Aus den zwischen 1955 bis 2019 aufgestellten Definitionen konnten vier vermehrt vorkommende Merkmale von KI abstrahiert werden:

i) *"Wahrnehmung der Umwelt und der Komplexität der realen Welt,*
ii) *Informationsverarbeitung: Sammeln und Interpretieren von Informationen,*
iii) *Entscheidungsfindung, einschließlich des Denkens, Lernens und Handelns;*
iv) *und das Erreichen von vordefinierten Zielen"* (frei übersetzt nach AI Watch 2020, S. 4)

Anwendungsbereiche von KI-Systemen

Es gibt nicht „den einen" Anwendungsbereich von KI. KI-Systeme kommen in der Kunst, dem Marketing, der Wohlfahrt, der Bildung, dem Gesundheitssystem, im Rechtsbereich und vielen anderen Anwendungsfeldern zum Einsatz (vgl. Algorithm Watch 2020b, S. 6). Jeder Bereich bringt eigene Standpunkte, Definitionen und dadurch unterschiedliche Perspektiven mit sich. Ausgehend von Stellungnahmen der Europäischen Kommission, nationalen Strategien und politischen Dokumenten (europäische und außereuropäische) sowie anderen internationalen Institutionen wie der OECD, der UNESCO, dem Weltwirtschaftsforum, der ISO, etc. leitet AI Watch folgende Perspektive ab (vgl. AI Watch 2020, S. 7): Die Politik und die eben genannten Institutionen betrachten KI als ein Instrument für Wachstum und technologische Entwicklung betrachten (ebd.). Das Verständnis der Forschung für KI steht dem etwas neutraler entgegen: Hier ist das Verständnis für KI eines, dass in ihr ein Forschungsgebiet, sowie die Entwicklung der Technologie für allgemeine Zwecke anerkennt (ebd.). Aus Sicht des Marktes liegt der Studie von AI Watch zufolge der Fokus auf der industriellen Entwicklung der Technologie und der Bewertung des wirtschaftlichen Werts sowie künftiger Marktaussichten (ebd.). In diesen aufgeführten Anwendungsbereichen wird der militärische Bereich ausgespart. Weber und Prietl beschreiben, dass neben kapitalistischen Prinzipien auch militärische Interessen auschlaggebend die Entwicklung von KI ist: Sie beschreiben KI als „das Zentrum eines militärisch-industriellen Komplexes" (frei übersetzt nach Weber und Prietl 2021, S. 63). Die militärische Nutzung von KI ist im Gegensatz zu den anderen Anwendungsbereichen eine, die in der wissenschaftlichen Literatur wenig behandelt wird (vgl. Krause 2021, S. 5). Das macht diesen Einsatzbereich von KI weniger transparent und nachvollziehbar. Für die Beantwortung der Forschungsfragen habe ich im Rahmen der gegebenen Möglichkeiten versucht, die Modifikationen auf soziale Ungleichheit so weit wie möglich auch für den militärischen Bereich abzudecken.

Künstliche allgemeine Intelligenz und KI

1965 wurde der Begriff Künstliche Intelligenz benutzt, um folgende Möglichkeit damit zu beschreiben: KI baut Funktionsweisen realer Prozesse nach. Dadurch ist es möglich, Abläufe im digitalen Raum zu simulieren, die den kognitiven Fähigkeiten eines Menschen nahekommen (vgl. Unesco Courier 2018, S. 7). Die Umsetzung war technisch zwar noch nicht möglich, aber durch die digitale Transformation und den technologischen Fortschritt kommen wir ebendieser Simulation von Realität im digitalen Raum immer näher (vgl. Kirste und Schürholz 2019, S.10). Ein KI-System, welches die vollumfänglichen kognitiven Fähigkeiten eines Menschen simulieren kann, nennt man allgemeine künstliche Intelligenz. Aktuell ist es unmöglich und Expert*innen wie beispielsweise Fjelland zufolge wird es auch in Zukunft nicht möglich sein ein solches System zu entwickeln (vgl. Fjelland 2020, S. 8). Die Unterscheidung zwischen allgemeiner KI und KI ist wichtig, weil der Mythos, der sich um allgemeine KI rankt die Diskussion über KI verzerrt (vgl. Unesco Courier 2018, S. 7). Zudem rückt der Begriff der allgemeinen KI die Verantwortlichkeit für KI-Systeme weg von den Menschen, welche diese Systeme entwickeln und einsetzen: Während allgemeine KI die Fähigkeit hat, eigenständig, autonom und verselbständigt zu agieren, handelt es sich bei den aktuell eingesetzten KI-Systemen um technische Systeme, die von Menschen eingesetzt und verantwortet werden (vgl. Unesco Courier 2018, S. 41). Diese Unterscheidung zeigt, dass es wichtig ist, KI-Systeme, als in die Gesellschaft eingebettete, sozio-technische Systeme zu kontextualisieren, die in ihrer Anwendung und Funktionsweise vom Menschen bestimmt werden.

KI als sozio-technisches System

KI-Systeme sind verwoben mit gesellschaftlichen Strukturen und Wirkmechanismen. Sie können entkoppelt von ihnen nicht bestehen (vgl. Lücking 2020, S. 73). Das liegt unter anderem darin begründet, dass es Daten braucht, deren Inhalte von den Systemen verarbeitet werden können. Daten sind Informationen, die unter anderem Praktiken in Gesellschaften abbilden (vgl. Nassehi 2020, S.108ff). Sie werden größtenteils von und durch Individuen passiv oder aktiv generiert und produziert (ebd.). Ausgehend von dieser Betrachtungsweise lässt sich KI mehreren Expert*innen zufolge als ein sozio-technisches System begreifen (vgl. Algorithm Watch 2020b, S. 84; Lücking 2020, S. 73; Horwath 2022, S. 72-73). Pinto nennt KI-Systeme „aus mehreren Komponenten bestehende sozio-technologische Rahmenwerke“ (Pinto 2020, S. 14).

Algorithm Watch betont nicht nur die Rolle der Gesellschaft und des Individuums für KI, sondern auch, dass die Art und Weise, wie die Technologie implementiert wird, ausschlaggebend für ihre Wirkung auf die Gesellschaft ist (vgl. Algorithm Watch 2020a, S. 4). Statt von Künstlicher Intelligenz, spricht die die NGO von Systemen automatisierter Entscheidungsfindung und bezeichnet diese als soziotechnische Systeme. Das liegt darin begründet, dass ein System automatisierter Entscheidungsfindung „keine eigenständige Technologie [ist], sondern ein Rahmen, der das Entscheidungsmodell und den politischen, wirtschaftlichen und organisatorischen Kontext seiner Nutzung umfasst“ (vgl. Algorithm Watch 2020b, S. 84). Automatisierte Entscheidungsfindung lässt sich auch in der Taxonomie von AI Watch wiederfinden. Dort wurde sie jedoch nur als eines von vier Charakteristika von KI-Systemen herausgearbeitet (vgl. AI Watch 2020, S. 4). Auch Lücking betont, dass „im öffentlichen Diskurs oft von ‚Algorithmen‘ gesprochen wird, wenn eigentlich das Zusammenspiel verschiedener Teile eines sozio-technischen Systems, also einer Mischung aus technischen und menschlich-sozialen Teilen, gemeint ist“ (Lücking 2020, S. 73). Horwath schreibt, dass das Einbeziehen sozio-materieller Kontexte von KI-Systemen essenziell ist, um „die (Re-)Produktion sozialer Ungleichheiten und Ausbeutungsverhältnisse erkennen zu können“ (Horwath 2022, S. 72-73). Es geht den genannten Expert*innen darum, die technische Perspektive auf KI-Systeme zu erweitern. Sie finden in der Implementierung und dem Aufbau von KI-Systemen eine gesellschaftliche Reflektion und Aushandlung wieder. Sie identifizieren den gesellschaftlichen Anteil als essenzielle Komponente der ganzheitlichen Betrachtung von KI. Für die Untersuchung der Forschungsfrage wähle ich deshalb jenen definitorischen Ansatz für KI. Mittels diesem Zugang erfasse ich KI-Systeme im Rahmen dieser Arbeit als in gesellschaftliche Macht- und Ungleichheitsverhältnisse verwobene sozio-technische Artefakte.

Nomenklatur

Ich habe mich für BIPoC als Bezeichnung entschieden, weil das die politische Selbstbezeichnung von Black, Indigenous und People of Color ist, die eine Rassismus-Erfahrung teilen und diese Nomenklatur diese kollektive Positionierung einfließen lässt (vgl.netzforma* e.V. 2020, S. 9).

Ich entscheide mich für diese Arbeit *weiß* immer klein und kursiv zu schreiben, da es im Gegensatz zu Schwarz kein Begriff für eine politisch empowernde Selbstdarstellung

ist (netzforma* e.V. 2020, S. 17) und die Schreibweise auf die meist unmarkierten Merkmale sozialer Konstruktionen hinweist (vgl. Eggers et al. 2005, S. 24ff).

Englischsprachige Terminologie im deutschen Fließtext

Viele der Wörter, die benutzt werden, um die neuen Entwicklungen und Bereiche der Technologie KI und die damit einhergehenden Möglichkeiten zu beschreiben, haben ihren Ursprung in der Informatik (vgl. Zuckarelli 2021, S. 25). Größtenteils werden in der Quellenliteratur trotz deutschem Fließtext englische Begriffe für die Bezeichnung von Fachbegriffen aus dem Bereich KI benutzt. Auch ich habe mich für die Nutzung englischsprachiger Fachbegriffe entscheide, um Ungenauigkeit im Ausdruck zu vermeiden.

3. Modifikationen der Dimensionen von Ungleichheit durch KI

Dieses Kapitel ist der Hauptteil dieser Arbeit. Ich beschreibe hier das für die Beantwortung der Forschungsfrage entwickelte Kategoriensystem.

3.1 Modifikationen der Ungleichheitsdimension Gender durch KI

In diesem Kapitel führe ich Argumente und Thesen hinsichtlich der Implikationen von KI auf die Ungleichheitsdimension Gender auf. Die Argumente sind den für die Beantwortung der Forschungsfrage entwickelten Kategorien zugeordnet. Ich fasse die Argumente zusammen und beschreibe ihre Zusammenhänge. Manche der Wirkmechanismen veranschauliche ich mittels Anwendungsbeispielen von KI. Es ist zu beachten, dass Gender – wie bereits in Kapitel 2. Methodik beschrieben – intersektional wirkt, d.h. in Verwobenheit mit weiteren Kategorien sozialer Differenzierung wie Ethnizität und Klasse, Alter, soziale Herkunft oder sexuelle Orientierung ist (vgl. Horwath 2022, S.75).

3.1.1 Datenbasiertheit von KI und Verzerrungen von Algorithmen durch Daten

Aufgrund der Betrachtung von KI als sozio-technisches System ist die Kontextualisierung der Einbettung der Technologie in die Gesellschaft immanent. Ein Charakteristikum von KI-Systemen, welches die Wechselwirkung zwischen KI und Gesellschaft bedingt, ist die Datenbasiertheit. Daten sind für KI-Systeme in zwei Phasen elementar: Einmal in der Entwicklungsphase des KI-Systems in Form von Trainingsdatensätzen und ein zweites Mal, wenn Daten von dem fertigen KI-System algorithmisch verarbeitet werden (vgl. Europäische Kommission 2018b, S. 3-4). Ich will die erste Phase detailliert beschreiben: Mithilfe von Trainingsdatensätzen lernt ein zu entwickelndes KI-System die Logik nach der es analysiert, kategorisiert und bewertet (vgl. Ertel und Black 2016, S. 285ff). Der Trainingsdatensatz ist den Softwareentwickler*innen bekannt. Entwickler*innen wissen also, welche Analyseergebnisse von ihrem zu entwickelnden KI-System zu erwarten sind. Sie vergleichen den Output des zu entwickelnden KI-Systems solange mit dem gewünschten Output, bis eine zufriedenstellende Übereinstimmung vorliegt (vgl. Kirste und Schürholz 2019, S. 32; Stanford University 2022). Nach diesem Prinzip werden die Stellschrauben, auch Hyperperameter genannt, solange gedreht, bis die Ergebnisse des KI-Systems mit den erwarteten Ergebnissen übereinstimmen und die Systementwicklung damit abgeschlossen werden kann (vgl. Kirste und Schürholz 2019, S. 32; Stanford University 2022). Zusammenfassend gesagt, baut ein KI-System

auf Basis der Informationen, die der jeweilige Trainingsdatensatz enthält, die Regelsysteme nach, auf denen es später Daten analysieren und Entscheidungen treffen wird (vgl. Pinto 2020, S. 14).

Exkurs: Beispiele für Trainingsdatensätze

Für das Training von KI-Systemen gibt es Trainingsdatensätze, die von Anbietern teilweise kostenfrei zur Verfügung gestellt werden, wie beispielsweise von Regierungen. In Österreich existiert beispielsweise das von der Regierung bereitgestellte, öffentlich zugängliche Datenportal Statistik-Austria[1]. Auch Institutionen wie die NASA und die WHO sowie private Unternehmen wie Twitter oder Google[2] bieten Datensätze zum Kauf oder kostenfrei an. Im Jahr 2018 machte Twitter einen Umsatz von 90 Millionen Dollar durch den Verkauf von Daten (vgl. Frankfurter Allgemeine 2018). Es existieren Unternehmen, die sich darauf spezialisieren, Datensätze mit themenspezifischen Inhalten zu verkaufen wie beispielsweise das Unternehmen statista[3].

Trainingsdatensätze bilden jedoch oft eine verzerrte gesellschaftliche Realität ab (vgl. Weber und Prietl 2021, S. 64). Das bedeutet, dass der Algorithmus durch Trainingsdatensätze ein Modell der Realität kennenlernt, das nicht repräsentativ für die Gesellschaft ist (ebd.). Die folgenden zwei Kapitel der Kategorie „Datenbasiertheit von KI und Verzerrungen von Algorithmen durch Daten“ zeigen zwei Arten ungleichheitsmodifizierender Wirkmechanismen, die durch Verzerrungen in Daten auftreten können.

3.1.1.a Verzerrungen durch Unsichtbarkeit mancher Gruppen in Daten - Digitale Kluft

Wie eine Verzerrung in Datensätzen zustande kommt, zeigt das folgende Beispiel: Aiello et al. stellen fest, dass geolokalisierte Daten von Twitter hauptsächlich von *weißen* Männern über vierzig stammen (Aiello et al. 2020, S. 109). Das heißt, dass diese Gruppe von Twitter-Nutzern häufiger ihren aktuellen Standort angibt als andere soziale Gruppen. Wird ein KI-System auf Basis eines solchen Datensatzes trainiert, wird dieses Muster inhärenter Bestandteil seiner Funktionsweise. In diesem Fall führt der Einsatz eines KI-Systems zu potenziell ungleichen Ergebnissen für die Angehörigen der verschiedenen repräsentierten sozialen Gruppen. Wie eine ungleiche

[1] Webseite von Statistik-Austria: https://www.statistik.at/en
[2] Webseite zum Abrufen der Datensätze von google: Dataset Search (google.com)
[3] Webseite von statista: https://de.statista.com/

Abbildung von sozialen Gruppen in Daten zu einer Verzerrung in einem KI-System führen kann, wird in den nächsten Absätzen geschildert.

Zwischen Verzerrungen in Daten, die sich in den KI-Systemen einschreiben, und bestehenden sozialen Ungleichheiten gibt es Zusammenhänge und Wechselwirkungen: Der Begriff digitale Kluft definiert Ungleichheiten in der Repräsentativität und im Zugang zu digitalen Ressourcen (vgl. Aiello et al. 2020, S. 108). Diese kann unterschiedliche Ausprägungen und Folgen haben: Die Dimensionen sozialer Ungleichheit haben Einfluss auf die Art und Weise, wie Individuen durch Daten repräsentiert sind und ob sie Zugang zu Software, Hardware und damit zur digitalen Welt haben (vgl. Nassehi 2020, S.108ff). Individuen, soziale Gruppen oder Bewohner*innen von Regionen sind ungleich genau oder von und durch Daten verzerrt repräsentiert (vgl. Rubinov und Hagerty 2019, S. 2). Beispielsweise sind ältere Menschen in der digitalen Welt weniger präsent und nur knapp die Hälfte der Menschen weltweit besitzen ein Smartphone (vgl. Budd et al. 2020, S. 1189). Das verringert die Repräsentativität derjenigen Personen in Daten. Rubinov und Hagerty sowie andere Expert*innen zeigen auf, dass im Gegensatz zu privilegierten Mitgliedern der Gesellschaft ältere Menschen, weniger wohlhabende Menschen, Menschen mit Behinderungen, Frauen, BIPOC und Menschen aus sozioökonomisch benachteiligten Ländern in Daten unterrepräsentierter sind (vgl. Weber und Prietl 2021, S. 64; Rubinov und Hagerty 2019, S. 2).

Die digitale Kluft steht in enger Verbindung mit der digitalen Spaltung. Das Konzept der digitalen Spaltung beschreibt, dass nicht nur eine ungleiche Verfügbarkeit, sondern auch ungleiche Nutzungsroutinen von Informations- und Kommunikationstechnologien entlang verschiedener sozialer Gruppen bestehen (vgl. Zillien 2009, S. 2). Diese unterschiedlichen Nutzungsweisen von Technik durch unterschiedliche soziale Gruppen werden ebenfalls in dem oben genannten Beispiel sichtbar: Die gesellschaftlich privilegierte soziale Gruppe *weißer* Männer nutzt öfter die Geolokalisierungsfunktion für ihre Tweets als Frauen und BIPoC (Aiello et al. 2020, S. 109). Dass die Nutzung von Technologie geschlechtsspezifisch und abhängig vom sozioökonomischen Hintergrund ist, zeigen auch Karusala et al. am Beispiel von Indien (vgl. Karusala et al. 2019, S. 511). Eine 2019 veröffentlichte Studie zeigt, dass Frauen im Durchschnitt 10 % seltener als Männer ein Mobiltelefon besitzen. Sie nutzen das mobile Internet zu 23 % weniger als Männer (vgl. GSMA 2019, S. 4). Die Beispiele

zeigen, dass „Datensätze bzw. ‚fehlende Daten' […] immer gesellschaftliche Machtverhältnisse wieder[spiegeln]" (Horwath 2022, S.77). Dieses geschlechtsspezifische Gefälle bildet sich in den Daten ab. Infolgedessen verstärken Daten, welche innerhalb von Trainingsdatensätzen die Grundlage der Entwicklung eines KI-Systems sind, den Ausschluss der Interessen und Bedürfnisse sozialer Gruppen**,** die in Daten nicht abgebildet sind. Das reproduziert die Marginalisierung bestimmter sozialer Gruppen (vgl. Weber und Prietl 2021, S. 64). Weiters sind die Ergebnisse von KI-Systemen für minder repräsentierte Gruppen weniger akkurat, wie ich in Kapitel 3.1.1.a am Beispiel von Hypothekenbewilligungen durch KI-Systeme zeige. Die geringere Sichtbarkeit von Frauen in Daten hat je nach KI-System unterschiedliche Konsequenzen. Beispielsweise führt Horvath als Konsequenz die Entstehung von Trennlinien auf: Diese kann von einer „Zutrittsverweigerung zum Fitnessstudio bis zur Einstufung als terroristische Gefahr reichen" (Horwath 2022, S. 77).

3.1.1.b Einschreiben bestehender genderspezifischer Diskriminierung in Algorithmen durch Trainingsdatensätze und Daten

Nicht nur die Unsichtbarkeit bestimmter Individuen und sozialer Gruppen in Daten ist problematisch, auch andere Wirkmechanismen der Datenbasiertheit von KI-Systemen reproduzieren bestehende Machtrelationen der Gesellschaft. Das folgende Beispiel beschreibt einen Mechanismus, der zum Einschreiben relationaler Machtverhältnisse in KI-Systeme führt. Zudem erkläre ich, wie es bei Verzerrungen in Trainingsdatensätzen zu einem langfristigen Einschreiben, also einer Fundamentierung dieser Machtverhältnisse kommen kann.

KI-System des AMS

Im Jahr 2020 hat ein Algorithmus des AMS auf Basis unterschiedlicher Daten arbeitssuchender Österreicher*innen entschieden wie vermittlungsfähig die Betroffenen sind (vgl. Berner und Schüll 2020, S. 3ff). Anhand der Kategorisierung der arbeitssuchenden Person durch das KI-System werden dem oder der Betroffenen Ressourcen, wie beispielsweise der kostenfreie Zugang zu einem Lehrgang zugeteilt. Das KI-System entscheidet nicht alleinig über die Ressourcen für die Re-integrierung in den Arbeitsmarkt. Die Kategorisierung sollte die zugeteilte Betreuer*in der arbeitssuchenden Person bei ihrer Entscheidung unterstützen (vgl. Wagner et al. 2020, S.191ff). Bei der Kategorisierung nach Vermittlungsfähigkeit wurden

Datenpunkte wie weiblich oder drittstaatangehörig von dem KI-System kategorisch schlechter bewertet als männlich oder österreichisch (vgl. Guijarro-Santos 2020, S. 55 ff). Es ist - wie oben beschrieben - KI-Systemen inhärent, dass sich ihre Regeln, nach denen sie funktionieren, auf Basis von Trainingsdatensätzen formieren. In dieser Phase der Entwicklung des Algorithmus kann es zu falschen oder generalisierenden Annahmen und Stereotypisierungen kommen. Das zeigen die genderspezifisch diskriminierenden Ergebnisse des KI-Systems des AMS (vgl. Berner und Schüll 2020, S. 9). Formal neutral wirkende Kriterien können mit Entscheidungskriterien korrelieren, die sich in der Folge benachteiligend für soziale Gruppen auswirken: Anhand des Trainingsdatensatzes hat das KI-System des AMS gelernt, Betreuungspflicht als negativ gewichtetes Kriterium für die Vermittlungsfähigkeit einer arbeitssuchenden Person einzustufen (vgl. Guijarro-Santos 2020, S. 53–54). Trotz der formalen Neutralität des Merkmals „betreuungspflichtig“ bei einer arbeitssuchenden Person, wirkt es sich im Ergebnis auf Frauen aus, weil Betreuungsarbeit mehr von Frauen als Männern getragen wird (vgl. Villa 2020, S.433). So kommt es Guijarro-Santos zufolge im Fall des KI-Systems des AMS zu Beeinträchtigungen der Grundrechte auf informationelle Selbstbestimmung und Datenschutz sowie des Grundrechts auf Gleichbehandlung und Antidiskriminierung (vgl. Guijarro-Santos 2020, S. 50). Berner und Schüll warnen vor der „Verdinglichung des Menschen sowie der Gefährdung der Menschenwürde und des Gleichbehandlungsgrundsatzes“ (Berner und Schüll 2020, S.3) im Fall des KI-Systems des AMS. Wegen der fehlenden Rechtsgrundlage wurde der Einsatz des KI-Systems kurz nach seiner Einführung im Jahr 2020 gestoppt (ORF 2020).

Das Beispiel zeigt wie es bei der Nutzung von Daten, die ein Ungleichgewicht enthalten oder reale Diskriminierungen abbilden, zu einem Fortschreiben von Ungleichbehandlung kommt (Schmidt und Mellentin 2020, S. 22).

Die folgenden Beispiele veranschaulichen, wie sich bestehende genderspezifische Diskriminierung auf unterschiedliche Weise in Algorithmen von KI-Systemen einschreiben. Die Beispiele stammen aus Anwendungen von KI-Systemen für die Unterstützung von Einstellungsprozessen für Arbeitgeber*innen. Hier finden KI-Systeme beispielsweise beim Suchen potenzieller Bewerber*innen Einsatz: Mithilfe des „Micro-Targetings“, dem zielgerichteten Ansprechen bestimmter Bevölkerungsgruppen, werden Nutzer*innen verschiedenster Plattformen auf Basis

aufgezeichneter Online-Aktivitäten kategorisiert (vgl. Bogen und Rieke 2018, S. 17). Je nach demographischer Zielgruppe werden so potenzielle Bewerber*innen aus- oder eingegrenzt. Die KI-Systeme leiten beim „Micro-Targeting“ die Kategorisierung direkt vom Individuum ab, während beim sogenannten „Matching“ eine Kategorisierung auch durch übereinstimmende Muster von Online-Aktivitäten zweier Nutzer*innen abgeleitet werden (vgl. Bogen und Rieke 2018, 19ff). Deniz Erden zeigt, wie es bei beiden Varianten zu genderspezifischer Diskriminierung kommen kann:

> *„Erstens könnten Frauen aufgrund ihrer tatsächlichen Fähigkeiten auf Stellen in niedrigeren Positionen klicken, so dass ein Modell, das nur aus diesem Verhalten lernt, ihre Positionen, die ihren Fähigkeiten entsprechen, möglicherweise nicht zeigt. Andererseits könnte eine Frau, die sich für Führungspositionen interessiert auch aus der Matching-Gruppe ausgeschlossen werden, weil andere Frauen mit ihren Fähigkeiten kein Interesse an solchen Positionen zeigen“ (Erden 2020, S. 80).*

Erden beschreibt, wie ein potenzielles Muster der androzentristischen Gesellschaft, in diesem Fall eine angelernte Zurückhaltung oder Bescheidenheit von Frauen bei der Einstufung ihrer Qualifikationen und Fähigkeiten, in den Algorithmen festgeschrieben und verstärkt wird. Das KI-System schreibt einer einzelnen Person Gruppeneigenschaften zu und kategorisiert und bewertet auf Basis dieser Zuschreibungen die Person. Diese Bewertung bestimmt über den Zugang zu Ressourcen, im Fall dieses Beispiels den Zugang zu einer Arbeitsstelle (vgl. Erden 2020, S. 80). In Stellensuch-Webseiten wie LinkedIn oder ZipRecruiter kommt es nach ähnlichem Prinzip zu diskriminierenden Wirkmechanismen für Frauen, wie Bogen und Rieke in ihren Untersuchungen zeigen (Bogen und Rieke 2018). Im folgenden Absatz gebe ich die Ergebnisse ihrer Studie zu Einstellungsalgorithmen wieder: Nutzer*innen werden auf der Grundlage anderer, mit ihnen assoziierten Gruppen von Nutzer*innen, Informationen zugeschrieben (vgl. Bogen und Rieke 2018, S. 22). Diese zugeschriebenen Informationen basieren auch auf Verhaltensdaten von Nutzer*innen, die auf deren Interaktionen und Reaktionen auf der Website basieren (ebd.). Aus diesen Verhaltensdaten wurde abgeleitet, dass Männer sich tendenziell mehr auf Stellen bewerben, die über ihre Qualifikationen hinausgehen (ebd.). Der Algorithmus interpretiert diese Verhaltensunterschiede und passt seine Empfehlungen in einer Weise an, die Frauen benachteiligt (vgl. Bogen und Rieke 2018, S. 19–25). In diesem,

wie auch im nächsten Beispiel kommt es zur Diskriminierung von Frauen, da der Einzelnen eine Gruppeneigenschaft zugeschrieben wird (vgl. Martin 2022, S. 291).

Folgendes Beispiel zeigt, dass ein KI-Algorithmus trotz des Ausschlusses des Merkmals „Frau“ oder „weiblich“ als Analysekriterium zu diskriminierenden Ergebnissen für Frauen kommt. Im Jahr 2014 entwickelte Amazon ein KI-System, welches Lebensläufe von Bewerber*innen überprüfte**,** um die besten Kandidat*innen für eine bestimmte Stelle herauszufiltern (vgl. Kodiyan 2019, S. 1). Im Jahr 2015 stellte Amazon die Entwicklung ein, da das System Bewerber*innen für Softwareentwicklung und andere technische Stellen nicht genderneutral bewertete (vgl. Dastin 2022, S. 296). Lebensläufe, die das Wort "Frau" enthielten, wie z. B. "Kapitänin des Frauenschachclubs" und Absolventinnen von zwei reinen Frauen-Colleges wurden als negativ bewertet (vgl. Kodiyan 2019, S. 2). Das KI-System hat also Stellvertretermerkmale für Gender und Geschlecht, wie beispielsweise der Mitgliedschaft in einem Frauenschachclub, negativ bewertet. So konnte es, trotz des Ausschlusses gendersensibler Attribute, zu diskriminierenden Ergebnissen für Frauen kommen. Dastin zufolge, war der Grund für diese diskriminierenden Bewertungen der Trainingsdatensatz. Dieser bestand aus den Lebensläufen der eingestellten Personen der letzten zehn Jahre bei Amazon: Der Großteil der eingestellten Menschen waren Männer, was die generelle männliche Dominanz in der Technologiebranche widerspiegelt (vgl. Dastin 2022, S. 296). Durch dieses, in den Daten abgebildete Ungleichgewicht kam es zu einer algorithmischen Diskriminierung des KI-Systems gegenüber Frauen.

Zudem gibt es KI-Systeme auf dem Markt, die vorhersagen, welche Bezahlung eine Bewerber*in dazu bringt, das Jobangebot anzunehmen (vgl. Erden 2020, S. 81). Aufgrund von Trainingsdatensätzen, welche den real existierenden Gender Pay Gap abbilden, können KI-Systeme dementsprechend niedrigere Gehälter für Frauen vorschlagen (ebd.). Hier kommt es ebenfalls aufgrund der in den Daten abgebildeten geschlechtsspezifischen Ungleichheit zu einer Diskriminierung von Frauen.

Beide Beispiele zeigen, wie sich in Daten abgebildete soziale Ungleichheiten in KI-Systeme einschreiben können und wie der Einsatz von KI-Systemen zu genderspezifischen und anderen Formen der Diskriminierung führt.

Eine zusätzliche Fundamentierung der in den KI-Systemen eingeschriebenen diskriminierenden Wirkmechanismen passiert, wenn Algorithmen von KI-Systemen

nicht nachbearbeitet und auf den neuesten Stand gebracht werden. Also, wenn ihre Entwicklung mit dem Beginn ihrer Anwendung endet. Ich will mittels eines kurzen Beispiels den Mechanismus der Fundamentierung veranschaulichen: Stellen wir uns ein KI-System ähnlich dem des AMS vor. Ein Trainingsdatensatz bestehend aus Daten des AMS aus den Jahren 2015 bis 2020 wird für die Entwicklung des Systems eingesetzt. Das KI-System lernt aus den Daten dieses Zeitraums, welche Empfehlungen es in Zukunft aussprechen wird. Wird das KI-System nach einem Jahr mit einem aktuellen Trainingsdatensatz aktualisiert? Oder wird eine solche Aktualisierung erst nach fünf oder vielleicht 15 Jahren vorgenommen? Je nachdem wann und ob eine solche Aktualisierung stattfindet, können die Daten und somit auch die ihnen zugrundenliegenden gesellschaftlichen Verhältnisse, die sie abbilden langfristig ausschlaggebend für die Entscheidungslogik eines solchen KI-Systems sein. Diese potenzielle Fundamentierung von algorithmischer Diskriminierung bezeichnen Lücking und Prietl als eine ungleichheitskonservierende Tendenz von KI-Systemen (vgl. Lücking 2020, S. 70; Weber und Prietl 2021, S. 66). Bei den meisten KI-Systeme bleibt nach der Fertigstellung des Designs der Algorithmus auf diese Weise bestehen und die Phase des Lernens und der Softwareentwicklung ist abgeschlossen (vgl. DSK 2019, S.2ff). Bei allen Anwendungsbeispielen von KI, die ich im Rahmen dieser Arbeit aufführe, handelt es sich um KI-Systeme deren Algorithmen ab dem Zeitpunkt der Fertigstellung des Systems statisch bestehen bleiben. Demnach nehmen KI-Systeme keine aktuellen gesellschaftlichen Trends und Entwicklungen auf, sondern konservieren jene aus der Vergangenheit (AI Watch 2020, S. 4). Daten sind nach Prietl und Weber Abbilder gesellschaftlicher Strukturen und stellen damit soziokulturelle und hochpolitische Konstrukte dar (vgl. Weber und Prietl 2021, S. 65). Sie beschreiben KI-Systeme zudem als soziomaterielle Apparate selbsterfüllender Prophezeiungen (vgl. Weber und Prietl 2021, S. 65). Weitere Autor*innen erkennen an, dass KI-Systeme konservative Tendenzen entfalten (AI Watch 2020; vgl. Prietl 2019, S. 316; Lücking 2020, S. 70).

3.1.2 Algorithmische Kodierung von Realität basiert auf Kategorisierung und Normierung – Die Norm ist der *weiße* Mann

Das zweite technische Charakteristikum von KI-Systemen, das sich auf genderbasierte Ungleichheit auswirkt, ist eines, das allen algorithmischen Systemen inhärent ist: Ihr auf Spannungsunterschieden basierendes Funktionieren bedingt eine Übersetzung der nicht-digitalen Welt in Nullen und Einsen (vgl. Nassehi 2019, S. 152

ff). Dabei werden Strukturen bevorzugt, die sich leicht in numerische Daten und eindeutige Kategorien umwandeln lassen, da diese sich leichter algorithmisch verarbeiten lassen (vgl. Weber und Prietl 2021, S. 68). Zudem sind für das Kodieren und Dekodieren Sortierungs-, Kategorisierungs- und Bewertungsprozesse nötig (vgl. Pinto 2020, S. 14). Die Autorinnen Weber und Prietl warnen davor, dass Gesellschaften, die durch Daten beschrieben werden, auf kodierbare Handlungen reduziert werden und "wechselseitige Beziehungen, Komplexität, Variabilität und Disparität" (frei übersetzt nach Weber und Prietl 2021, S. 65) unsichtbar bleibt. Beim Fortführen dieser Argumentation kommen Prietl und Weber zu dem Schluss, dass algorithmische Verarbeitung von Daten eine „Normierung menschlichen Verhaltens" (frei übersetzt nach Weber und Prietl 2021, S. 62) bedingt. Sie fragen, welches Verhalten als Norm identifiziert und durchgesetzt wird und welches menschliche Verhalten von dieser Konzeptualisierung ausgeschlossen wird und wie diese beiden Elemente mit „genderspezifischen, rassistischen und ableistischen Strukturen und Symbolsystemen verwoben sind" (frei übersetzt nach Weber und Prietl 2021, S. 62). In einer androzentristisch geprägten Gesellschaft, deren Normen sich – wie oben beschrieben – in Daten und Algorithmen einschreiben, liegt es nahe, die Frage nach der Identifikation einer Norm mit einer Fortschreibung der bestehenden Machtverhältnisse zu beantworten. Das würde in Bezug auf genderspezifische soziale Ungleichheit bedeuten, dass der bestehende Androzentrismus, also „die geschlechtsbezogenen Strukturierungen kultureller Ordnungen und Legitimationsmuster" (Knapp 2018, S. 2) als Normen von KI-Systemen identifiziert werden und Frauen somit als eine Abweichung von der männlichen Norm verstanden werden. Donna Haraway hat bereits vor der Jahrtausendwende für technowissenschaftliche Artefakte im Allgemeinen argumentiert, dass diese nicht neutral sind, sondern vielmehr bestimmte Weltanschauungen, Wahrnehmungsstile und die Reproduktion bestehender sozialer Ungleichheiten begünstigen (Haraway 1988, S. 587). Auch Horvath beschreibt, dass die mit der „Klassifikation verbundenen Risiken der Hierarchisierung und Unterdrückung […] für bereits marginalisierte Gruppen wesentlich höher, als für sozial etablierte und privilegierte Gruppen" (Horwath 2022, S.78) sind.

3.1.3 Ergebnisse von KI-Systemen sind kein situiertes Wissen

Wegen der Verortung des Diskurses von situiertem Wissen in der feministischen Theorie und im Diskurs um Feminismus (vgl. Haraway 1988, S. 587; Ziai et al. 2018, S. 11–12) ordne ich diese Kategorie der Modifikation der Ungleichheitsdimension Gender durch KI zu. Die Kategorie betrifft jedoch nicht nur die Ungleichheitsdimension Gender, sondern auch die Dimensionen Klasse und Ethnizität.

Ein weiteres wichtiges Argument, das von Prietl, aber auch von anderen Autor*innen beschrieben wird, ist, dass die Situiertheit von Wissen im Prozess der algorithmischen Verarbeitung durch das KI-System verloren geht. Ergebnisse oder Entscheidungen von KI-Systemen geben eine vermeintlich unpolitische Analyse sogenannter Fakten vor (Prietl 2019, S. 303 ff; Weber und Prietl 2021, S. 65 ff). Auch Lücking beschreibt, dass bei KI, wie auch bei anderen Formen von Technik folgender Glaubenssatz eine wichtige Rolle spielt: „Technik könne tiefe, letzte und unabänderliche Wahrheiten bezüglich sozialer Phänomene aufdecken“ (Lücking 2020, S. 66). Wie bereits gezeigt, reflektieren KI-Systeme und Daten relationale Machtverhältnisse in Gesellschaften. Die Situiertheit von Wissen ist für feministische Analysen ein wichtiger Anhaltspunkt. Er entspringt der feministischen Kritik der 80er Jahre (vgl. Haraway 1988, S. 587). Feminist*innen dieser Zeit kritisierten „tiefere Wahrheiten und objektives Wissen […] als patriarchale Herrschaftstechnik“ (Lücking 2020, S. 72). Es wird damit kritisiert, dass Wissensproduktion lange Zeit als vermeintlich objektiv dargestellt wurde, obwohl es von Männern produziert wurde und ignorierte, dass die Perspektive von Frauen ausgeschlossen war (vgl. Haraway 1988, S. 587; Ziai et al. 2018, S. 11–12). Haraway beschreibt, dass jedes Wissen unter materiellen, kulturellen und politischen Bedingungen produziert ist und auch so verstanden werden muss (vgl. Haraway 1988, S. 587). Zudem beschreibt sie rationales Wissen als ein „machtbewusstes Gespräch“ (frei übersetzt nach Haraway 1988, S. 590). Sie, wie auch andere Feminist*innen

dieser Zeit, forderten eine kritischen Reflexion der Geschichte der Wissenschaften: Es wurde vermeintlich objektives Wissen produziert, welches in der Realität jedoch aus einer mehrheitlich männlichen Perspektive entstanden war. Ziai et al. wenden sich in ihrem Buch nicht nur gegen den „expliziten und meist unreflektierten Eurozentrismus in den Sozialwissenschaften“ (Ziai et al. 2018, S. 11–12), sondern auch gegen die Verallgemeinerung männlicher Erfahrungen, da diese eine Marginalisierung jeder anderen Erfahrungen reproduziert. Aus dieser kritischen Perspektive lassen sich auch

KI-Systeme betrachten. Denn sie produzieren ebenfalls Wissen, bewerten und kategorisieren, wobei eine kritische Betrachtung der Situiertheit dieses Wissens Prietl zufolge nicht passiert (Prietl 2019, S. 303 ff). Dabei kommt es zu einer Marginalisierung weiblicher und weiblich gelesener Erfahrungen. Ähnlich wie Prietl kommt auch Lücking zu dem Schluss, dass die Objektivierung oder Neutralisierung von Technologien wie KI dazu führt, dass bestehende Machtverhältnisse zementiert werden (vgl. Lücking 2020, S. 73).

3.1.4 Fehlende Gerichtsbarkeit und Regulierung - KI als Blackbox

Ausgehend von der obigen Herleitung lassen sich genderspezifische Diskriminierungen, wie andere Formen der Diskriminierung in den Wirkmechanismen von KI-Systemen in ihrer Wechselwirkung mit der Gesellschaft erkennen. Antidiskriminierungsgesetze, die Datenschutzgrundverordnung der EU oder die Menschenrechte der UN-Charta bilden einen gesetzlichen Rahmen, der Menschen vor Diskriminierungen dieser Art schützen soll. Die Gesetzgebungen unterscheiden sich auf nationaler Ebene: In den meisten Ländern gibt es noch keine Gesetze, die spezifisch vor algorithmischer Diskriminierung schützen. In manchen Ländern gibt es jedoch Versuche, einen gesetzlichen Rahmen aufzubauen oder auszubauen. Zum Beispiel, ist in Artikel 22 der DSGVO festgelegt, dass Entscheidungen nicht ausschließlich automatisiert getroffen werden dürfen (vgl. Bauer et al. 2021, S. 16). Zudem müssen von KI-Systemen getroffene Entscheidungen immer von Menschen verantwortet werden (ebd.). Damit ist aber noch kein Schutz vor diskriminierenden Einschätzungen oder Beschneidungen des Datenschutzes durch KI gewährleistet.

Ekker schildert 2020 in einer Onlinekonferenz des Projekts „Getting the future right – Artificial intelligence and fundamental rights[4]“ der Fundamental Rights Agency und Projektpartner*innen von seiner Arbeit als unabhängiger Anwalt für Technik und Datenschutz mit Sitz in Amsterdam. Er berät und verhandelt über Datenschutz und Sicherheit bei KI-Systemen und vertritt Betroffene bei Anklagen diskriminierender Ergebnisse algorithmischer Systeme. Einer seiner Fälle wird in Kapitel 3.3 diskutiert. Eine Problematik, die er schildert und die auch von der Fundamental Rights Agency und vielen anderen Institutionen an bestehenden KI-Systemen kritisiert wird, ist die Intransparenz der Algorithmen von KI-Systemen (European Union Agency for Fundamental Rights 2020b, 2018; European Commission 2020; UNICRI 2020; United

[4] https://fra.europa.eu/en/publication/2020/artificial-intelligence-and-fundamental-rights

Nations 2020). Auch die Studie von Algorithm Watch, für die eine Vielzahl von KI-Systemen in europäischen Ländern untersucht wurden, kommt zu dem Ergebnis das es nicht genug Transparenz bei KI-Systemen – „weder im öffentlichen noch im privaten Sektor“ (Algorithm Watch 2020b, S. 7), gibt. Von der Fundamental Rights Agency wird einerseits das Fehlen passender Leitlinien für Betroffene und Entwickler*innen zu Datenschutz und Antidiskriminierung bei KI-Systemen problematisiert, andererseits steht der fehlende Zugang zur Justiz in der Kritik: Wenn Menschen KI-basierte Entscheidungen anklagen wollen, ist schwer bis gar nicht möglich. Im Zuge dessen werden KI-Systeme als Blackboxes bezeichnet. Insbesondere solche des maschinellen Lernens (vgl. Kapitel 3.1) wie Lena Simon in folgendem Zitat erklärt:

> *„KI besteht […] aus Algorithmen. Ihre Besonderheit besteht darin, dass sie ihre Wenn-Dann-Beziehungen […] selbst herleiten. […] Das hat […] den großen Nachteil, dass hier falsche Schlüsse entstehen können, die wir leider nur sehr schwer entdecken können, da wir nicht mehr nachvollziehen können, wie die Entscheidung eigentlich zu Stande kam“ (vgl. Simon 2020, S. 32).*

Dieser Funktionsmechanismus von KI-Systemen bedeutet für den Menschen, dass die Zwischenschritte, die vom Eingangsbild zur schlussendlichen Klassifikation und damit zum Ergebnis führen, nicht mehr ersichtlich sind. Hier stellt sich die Frage, ob die Zugänglichkeit der Algorithmen durch eine entsprechende Anpassung der Software möglich ist, oder ob ein KI-System eine intransparente Blackbox ist.

Der Begriff der Black Box hat seinen Ursprung in der Kybernetik. Nach dem 2. Weltkrieg wurde der Begriff hier erstmalig verwendet, um technologische Kriegsartefakte zu bezeichnen, deren Funktionsweise unzugänglich gemacht wurde (vgl. Schäffner et al. 2019, S. 135). „Blackboxing“ war damals eine Praxis von Wissenschaftlern und Entwicklern in der Radar- und Bombenzielgeräte verplombt wurden. Die Versiegelung von Schlüsseltechnologie bewahrte den technologischen Fortschritt vor der jeweilig gegnerischen Seite (vgl. Ashby 1961, o.S.). Der Begriff Blackbox wird heutzutage nicht mehr nur benutzt, um Technik zu beschreiben deren Funktionieren wenig bis gar nicht begreifbar oder zugänglich ist, sondern auch um Phänomene wie beispielsweise „neuronale Netze, äußerliches Verhalten und psychische Vorgänge“ (Schäffner et al. 2019, S. 153) zu beschreiben. In der Informatik nimmt der Begriff Blackbox jedoch immer noch Bezug auf die ursprüngliche Praxis der

Versiegelung und stellt damit ein Grundprinzip der Computerentwicklung und der Technikentwicklung da: „Blackboxing‘ ist angewendetes und positives Wissen, das etwa zum Standard von neueren Programmiersprachen geworden ist, die auf Kosten der Effizienz die Kapselung von Unterroutinen vorschreiben“ (Schäffner et al. 2019, S. 141–142). Mit Kapselung der Unterroutinen meinen Schäffner et al., dass mehrere programmierende Personen in geringerem Maße wissen müssen, was die jeweils andere Person in den Programmier-Code geschrieben hat. Das ermöglicht eine leichtere Arbeitsteilung bei der Entwicklung von Algorithmen.

Die Frage, ob KI-Algorithmen Intransparenz und fehlende Nachvollziehbarkeit als unveränderliche Merkmale ihrer technischen Funktionsweise in sich tragen, oder ob die beiden Merkmale viel mehr Ergebnis der Art und Weise sind, wie KI-Systeme und Software entwickelt werden, lässt sich jedoch nicht allein mittels der Etymologie des Begriffs beantworten. Die Forderung nach mehr Transparenz und Nachvollziehbarkeit bei Algorithmen in KI-Systemen zeigt dennoch, dass der Gesellschaft die Einsicht in algorithmische Prozessen und Systeme verwehrt ist. Nichtregierungsorganisationen wie beispielsweise Algorithm Watch fordern Regierungen und private Unternehmen auf, ihre Algorithmen offenzulegen (vgl. Algorithm Watch 2020b, S. 6).

Solange die Zugänglichkeit zu Algorithmen von KI-Systemen nicht gegeben ist, ist die Gerichtsbarkeit einer algorithmischen Entscheidung nicht möglich: Für Personen, die eine genderspezifische Diskriminierung erfahren, ist es durch die Intransparenz der KI-Systeme unmöglich sich Gewissheit darüber zu verschaffen, ob ihnen eine Diskriminierung widerfahren ist oder nicht. Vor Gericht stellt die Intransparenz der KI-Systeme dieselbe Problematik dar: Für ein Gerichtsurteil ist es schlecht, wenn es nicht möglich ist, die Funktionsweise und Auswirkungen auf Grundlage genauer Informationen über den Algorithmus des KI-Systems zu bewerten (European Union Agency for Fundamental Rights 2020c, S.16). Die Undurchsichtigkeit verhindert das Sammeln von Beweisen, die notwendig sind, um ein fundiertes Urteil über die Logik der Funktionsweise des KI-Systems zu fällen (vgl. Algorithm Watch 2020b, S. 7). Ein Argument mit dem Unternehmen die Unzugänglichkeit der Algorithmen ihrer KI-Systeme begründen, ist das Betriebsgeheimnis: Timo Daum berichtet, dass immer mehr Unternehmen ihre Algorithmen über Patente schützen und sie somit geheim halten dürfen (vgl. Daum 2019, o.S.). Für die betroffenen Menschen ist schwer zu

erkennen, ob sie ausgegrenzt oder diskriminiert werden, wenn die die Logik hinter den algorithmischen Regelsystemen nicht preisgegeben wird (vgl. Pinto 2020, S. 14).

Eine weitere rechtliche Problematik liegt Fröhlich und Spiecker zufolge in der Zuschreibung von Gruppenmerkmalen durch KI-Systeme (vgl. Fröhlich und Spiecker 2018, o.S.). Dieses Charakteristikum von KI-Systemen habe ich bereits in Kapitel 3.1.1.b erklärt. Fröhlich und Spiecker beschreiben diesen Mechanismus von KI-Systemen als externe Persönlichkeitskonstruktion, welche diskriminierende Ergebnisse durch die Zuschreibung von Gruppenmerkmalen auf einzelne Individuen hervorbringen kann (ebd.). Die Charakterisierung anhand gruppenbildender Merkmale durch Algorithmen greift nach Fröhlich und Spiecker in das Recht auf informationelle Selbstbestimmung (nach Artikel 1, Absatz 1, Grundgesetzbuch) bzw. das Recht auf Datenschutz (Artikel 7 bzw. Artikel 8 Europäische Grundrechtecharta) ein (ebd.).

3.1.5 Kann KI sozialer Ungleichheit entgegenwirken?

Ausgehend von bisherigen Argumenten lässt sich die Wirkung von KI auf die Ungleichheitsdimension Gender als eine ausmachen, die bestehende genderspezifische Ungleichheiten reproduziert und durch ihre Einschreibung in die algorithmischen Systeme fundamentiert. Es gibt jedoch auch Argumente, die dem Entgegenstehen: Mackey et al. erkennen in Algorithmen selbst die Lösung zur Bewältigung der diskriminierenden Ergebnisse von Algorithmen (Mackey et al. 2020, S. 1). Sie schätzen KI-Systeme, die diskriminierende Ergebnisse hervorbringen, als nicht gut genug entwickelt ein. Wissenschaftler*innen erforschen bereits Ansätze, mithilfe derer KI-Systeme potenzielle Verzerrungen in Daten erkennen und darauf reagieren können. Ein Beispiel für einen Lösungsansatz für KI-Systeme, deren algorithmische Entscheidungssysteme Frauen benachteiligen, wäre ein Gegenvektor (vgl. Sutton et al. 2018, S.328). Sutton et al. beschreiben, dass ein Gegenvektor dort ansetzt, wo der Algorithmus Ungleichheiten reproduziert. Der Gegenvektor gleicht automatisiert mittels neuer Gewichtungen die Ungleichheiten aus. Für den AMS-Algorithmus hätte das bedeutet, dass Frauen im gebärfähigen Alter als grundsätzlich vermittlungsfähiger eingestuft würden, um die diskriminierenden Verzerrungen in den Datensätzen und damit auch in den Ergebnissen ihrer Analyse auszugleichen.

Neben Gegenvektoren gibt es noch weitere technische Lösungsansätze, mit denen gegen das Einschreiben sozialer Ungleichheit in den Algorithmen von KI-Systemen vorgegangen werden kann (vgl. Sutton et al. 2018, S. 338). Aiallo et al. kritisieren, dass

diese Lösungsansätze nicht ausreichend erforscht sind (vgl. Aiello et al. 2020, S. 109-110). Nassehi meint, dass der gesellschaftliche Ursprung des Problems dieser technischen Lösung entgegensteht. Er argumentiert, dass reproduzierte Ungleichheit nur diejenige Ungleichheit widerspiegelt, die bereits vorhanden ist (vgl. Nassehi 2020, S. 2). Demnach ist ein Ausgleich der Verzerrungen von Datensätzen zwar ein Ansatz, der ein KI-System von einer spezifischen, abgrenzbaren und quantifizierbaren Ungleichheitsreproduktion kurieren kann, jedoch ist damit nicht der Kern, und zwar der gesellschaftliche Ursprung des Problems gelöst. Pinto argumentiert, dass ein KI-System, wenn es nur hinreichend entwickelt wird, das Potenzial birgt „veraltete Annahmen über Geschlechterrollen in Frage zu stellen“ (Pinto 2020, S. 14). Er verweist hier beispielhaft auf ein soziales Sicherungssystem, welches „Frauen tatsächlich mehr Macht und Eigenverantwortung gibt und zu einer gerechteren Gesellschaft für alle Menschen führt“ (Pinto 2020, S. 14). Für die Realisation eines solchen KI-Systems braucht es jedoch nicht nur das technologische Potenzial, sondern Nassehi zufolge auch eine Gesellschaft, eine Politik oder eine Wirtschaft die Interesse an der Entwicklung und dem Einsatz eines solchen Systems hat (vgl. Nassehi 2020, S. 2).

3.2 Modifikationen der Ungleichheitsdimension Ethnizität durch KI

Dieses Kapitel behandelt die Modifikationen der Ungleichheitsdimension Ethnizität durch KI. Nach gleichem Prinzip wie in Kapitel 3.2 führe ich Argumente und Thesen hinsichtlich der Implikationen von KI auf. Die Argumente sind den einzelnen Kategorien zugeordnet und ich beschreibe innerhalb der Kategorien die Zusammenhänge. Manche der Wirkmechanismen veranschauliche ich mittels Anwendungsbeispielen von KI-Systemen. Es ist zu beachten, dass Ethnizität – wie bereits in Kapitel 2. Methodik beschrieben – intersektional wirkt, d.h. verwoben ist mit weiteren Kategorien sozialer Differenzierung wie Gender, Klasse, Staatszugehörigkeit, Alter, soziale Herkunft oder sexuelle Orientierung (vgl. Horwath 2022, S.75).

3.2.1 Datenbasiertheit von KI und Verzerrungen von Algorithmen durch Daten

3.2.1.a. Verzerrungen durch ungleiche Repräsentativität mancher Gruppen in Daten

Daten sind – wie bereits in Kapitel 3.1.1.a beschrieben – ungleich repräsentativ für Menschen und soziale Gruppen. Die Dimensionen sozialer Ungleichheit bilden sich durch ungleiche Repräsentativität in Daten ab. Ethnizität, als eine gesellschaftlich konstruierte Ungleichheitsdimension, spiegelt sich nicht nur in der fehlenden Sichtbarkeit bestimmter ethnischer Gruppen in Daten wider, sondern auch in anderen

Formen der Verzerrung von Daten. Diese führt durch ein ähnliches Prinzip – wie bereits in Kapitel 3.1.1. in Bezug auf Gender beschrieben – zu diskriminierenden Ergebnissen von KI-Systemen. Im folgenden Kapitel zeige ich mittels verschiedener Beispiele, wie KI-Systeme auf diese Art und Weise die Ungleichheitsdimension Ethnizität modifizieren.

Exkurs KI-Systeme für Gesichtserkennung

KI-Systeme zur Gesichtserkennung haben sich im letzten Jahrzehnt schnell entwickelt und weit verbreitet (vgl. European Union Agency for Fundamental Rights 2020a, S. 2). Je nach Region werden die KI-Systeme sehr unterschiedlich eingesetzt. Mit Blick auf Europa finden sich Systeme beispielsweise in Schulen, Stadien, Flughäfen und Kasinos (vgl. Algorithm Watch 2020b, S. 7). Anwendungsbereiche von KI-Systemen für Gesichtserkennung liegen beispielsweise in der Unterstützung von Polizeiarbeit bei der Identifikation von vermissten Personen oder Kriminellen, oder – im Hinblick auf die COVID-19-Pandemie – in der Durchsetzung der sozialen Distanzierung, sowohl in Apps als auch durch intelligente Videoüberwachung des öffentlichen Raums (ebd.).

Joy Buolamwini, eine Forscherin am MIT wählte für ihre Studie über diskriminierende Gesichtserkennungssoftware drei KI-Systeme aus, deren Algorithmen öffentlich zugänglich waren (vgl. Buolamwini und Gebru 2018, S. 5–6). Für ihre Tests erstellten sie und ihre Kollegin Gebru einen Trainingsdatensatz aus 1270 Gesichtern, sodass die Trainingsdaten eine große Diversität wiedergeben (ebd.). Dafür verwendeten sie Profilbilder von Amtsträger*innen aus Ländern mit überwiegend dunkelhäutiger Bevölkerung sowie überwiegend hellhäutigen Einwohner*innen mit einem hohen Frauenanteil im Amt (ebd.). Microsofts Fehlerquote für dunkelhäutige Frauen lag bei 21 Prozent, während die von IBM und Megvii bei fast 35 Prozent lagen (vgl. Buolamwini und Gebru 2018, S. 10). Alle hatten Fehlerquoten unter 1 Prozent für hellhäutige Männer (ebd.). In ihren Ergebnissen beschreiben die beiden Forscherinnen, dass der Algorithmus der KI-Systeme ungleiche Ergebnisse hervorbringen und die überproportionale Repräsentativität *weißer* männlicher Gesichter in Trainingsdatensätzen Ursprung dieser Diskriminierung ist (vgl. Buolamwini und Gebru 2018, S. 11). Das begründen sie damit, dass Gesichtserkennungssysteme vor allem mit Bildern von *weißen*, männlichen Gesichtern trainiert werden, und deshalb die Erkennung für weibliche BIPoC weniger gut funktioniert (ebd.).

Steve Lohr schreibt, dass in einem weit verbreiteten Datensatz für die Gesichtserkennung zu mehr als 75 Prozent männliche und zu mehr als 80 Prozent *weiße* Personen vertreten sind (vgl. Lohr 2022, S. 143). Wie bereits in Kapitel 3.1.1.a beschrieben, stellen Aiello et al. fest, dass geolokalisierte Daten von Twitter hauptsächlich von *weißen* Männern über vierzig stammen (Aiello et al. 2020, S. 109). KI-Systeme für Bildklassifizierung und Gesichtserkennung werden von Datensätzen der Webseite ImageNet[5] trainiert, einer Sammlung von mehr als 14 Millionen gekennzeichneten Bildern (vgl. Zou und Schiebinger 2018, S. 324). Mehr als 45 % der ImageNet-Daten, die Shankar et al. zufolge die Forschung auf dem Gebiet der Computervision vorantreiben, stammen aus den USA, in denen nur 4% der Weltbevölkerung leben (vgl. Shankar et al. 2017, S. 2–3). Im Gegensatz dazu steuern China und Indien zusammen nur 3 % der ImageNet-Daten bei, obwohl diese Länder 36% der Weltbevölkerung ausmachen (vgl. Shankar et al. 2017, S. 2–3). Dieser Mangel an Diversität in den Trainingsdatensätzen erklärt, warum Algorithmen für Computervision das Foto einer traditionell *weiß* gekleideten US-amerikanischen Braut als "Braut", "Kleid", "Frau" und "Hochzeit" bezeichnen, das Foto einer nordindischen Braut aber als "Performance-Kunst" und "Kostüm " (vgl. Zou und Schiebinger 2018, S. 324) Wird ein KI-System auf Basis dieses Datensatzes trainiert, werden diese Muster zum inhärenten Bestandteil seiner Funktionsweise. Im Ergebnis kommt es, sobald die KI entwickelt und im Einsatz ist, zu benachteiligenden Ergebnissen für die jeweiligen Angehörigen der in den Daten unterrepräsentierten Gruppen.

Im Jahr 2019 kam es als Folge eines auf verzerrten Datensätzen trainierten KI-Systems zu der Verhaftung von Robert Julian-Borchak Williams (vgl. Schelenz 2021, S. 84). Ein KI-System für Gesichtserkennung hatte ihn fehlerhaft als eine für einen Ladendiebstahl gesuchte Person identifiziert (vgl. Schelenz 2021, S. 84). Das Standbild aus dem Überwachungsvideo des Geschäfts, in dem 2018 ein Ladendiebstahl begangen worden war, wurde von dem KI-System fälschlicherweise als die Person Robert Julian-Borchak Williams identifiziert (vgl. Hill 2022, S. 139). Es ist der erste bekannte Fall, bei dem ein Afro-Amerikaner aufgrund fehlerhafter Gesichtserkennung verhaftet wurde. Hill vermutet jedoch, dass es bereits zu mehr solcher Fälle gekommen ist (vgl. Hill 2022, S. 139). KI-Systeme für Gesichtserkennung werden beispielsweise auch zur Suche vermisster Personen (vgl. Wennker 2020, S.

[5] https://www.image-net.org/index.php

154) oder im Versicherungswesen zur Analyse des Gesundheitszustandes eingesetzt. Das Unternehmen Lapetus aus den USA bietet mit ihrem KI-System „Janus[6]“ ein KI-System zur Gesichtserkennung an, das mit zwei Fotos und neun Fragen bewertet, ob und für wie viel Geld eine Person eine Versicherung bekommt. Die zwei Bilder müssen zu zwei unterschiedlichen Zeitpunkten aufgenommen worden sein, denn sie werden von dem KI-System auf den Alterungsprozess und den Gesundheitszustand hin analysiert (vgl. Wennker 2020, S. 114). Wenn KI-Systeme, wie von Buolamwini und Gebru festgestellt, für BIPoC weniger gut funktionieren (vgl. Buolamwini und Gebru 2018, S. 12), bekommen sie eine weniger akkurate Einschätzung ihres Gesundheitszustandes oder im Falle des KI-Systems, welches vermisste Personen aufsucht, können diese weniger schnell identifiziert und gefunden werden.

Ein weiteres Beispiel einer Folge verzerrter Trainingsdatensätze aufgrund von ungleicher Repräsentativität für KI-Systeme stammt aus dem Bereich des Gesundheitswesens. Es werden KI-Systeme für die Erkennung von Hautkrebs eingesetzt (vgl. Haggenmüller et al. 2021, S.202-203). Eines solcher KI-Systeme hat bei einer Überprüfung gezeigt, dass die Erkennung von Hautkrebs bei hellen Hautfarben besser als bei dunkleren Hautfarben funktionierte (vgl. Weber und Prietl 2021, S.64). Das KI-System wurde Zou und Schiebinger zur Folge, mit einem Trainingsdatensatz trainiert, welcher zu 60% aus Bildern von google images bestand (vgl. Zou und Schiebinger 2018, S. 324). Dieser Datensatz enthält zu weniger als 5% Bilder von Menschen mit dunkler Hautfarbe (ebd.).

Die ungleiche Repräsentativität in Trainingsdatensätzen lässt sich in ihrem Wirkmechanismus nicht auf eine Dimension der Ungleichheit beschränken, sondern muss intersektional begriffen werden. Weber und Prietl schreiben, dass Trainingsdatensätze große Teile der Weltbevölkerung - ältere Menschen, weniger wohlhabende Menschen, Menschen mit Behinderungen, Frauen, BIPoC und Menschen aus sozioökonomisch benachteiligten Ländern weniger repräsentieren als Menschen aus den USA, (West-)Europa, sowie privilegierte Mitglieder dieser Gesellschaften (vgl. Weber und Prietl 2021, S.64). Weber und Prietl argumentieren, dass die Datensätze, welche die Grundlage für KI-Entwicklung sind, somit den Ausschluss der Interessen und der Bedürfnisse marginalisierter sozialer Gruppen verstärken und historisch gewachsene Beziehungen von Dominanz, Marginalisierung

[6] https://www.lapetussolutions.com/industry/lifeinsurance/

und Unterordnung reproduzieren (ebd.). Viele weitere Autor*innen warnen vor der algorithmischen Diskriminierung von Gesichtserkennungssystemen (Bacchini und Lorusso 2019, S.3ff; Weber und Prietl 2021, S. 64; European Union Agency for Fundamental Rights 2020a, S. 4; Unesco Courier 2018, S. 31).

3.2.1.b Einschreiben von bestehender rassistischer Diskriminierung in Algorithmen durch Trainingsdatensätze und Daten

Wie ich bereits in Kapitel 3.1.1.b zur Modifikation der Ungleichheitsdimension Gender veranschaulicht habe, schreiben sich bestehende Formen der Diskriminierung in KI-Systeme ein. Dasselbe gilt auch für die Ungleichheitsdimension Ethnizität: Wenn die soziale Realität von BIPoC diskriminierend ist, können Algorithmen, die diese Realität abbilden, ebenfalls diskriminierend wirken (vgl. Simon 2020, S. 40). Dazu kommt, dass Rassismen langfristig wirken, wenn sie einmal in die Algorithmen der KI-Systeme eingeschrieben sind und ihre Algorithmen nicht aktualisiert werden (vgl. Fröhlich und Spiecker 2018, o.S.; Lücking 2020, S. 70). Im Folgenden beschreibe ich anhand dreier Beispiele, wie sich bestehende Formen der Diskriminierung in die Algorithmen der KI-Systeme einschreiben und welche potenzielle Modifikation der Ungleichheits-dimension Ethnizität, das mit sich bringt. Die Beispiele stammen aus drei verschiedenen Anwendungsbereichen von KI und bilden das breite Spektrum des Einsatzes solcher Systeme ab.

KI-System „OptumIQ“

KI-Systeme unterstützen das Gesundheitswesen und dessen Verwaltungsaufgaben (vgl. Goodman et al. 2020S. 230ff). So sagt das System „OptumIQ[7]“ des Unternehmens Optum in den USA vorher, welche medizinische Versorgung eine Person benötigt (vgl. Alugubelli 2016, S. 4). Der Trainingsdatensatz mit dem „OptumIQ“ trainiert wurde bestand aus Krankenhaus Historien der vergangenen Jahre (vgl. Panesar 2019, S. 38–39). Es wurden die Kosten, die Patient*innen für das Gesundheitssystem in der Vergangenheit verursacht haben benutzt, um den Algorithmus zu trainieren, Vorhersagen für Patient*innen der Zukunft zu treffen (vgl. Johnson 2022a, S. 10). Auf der Basis von scheinbar neutralen statistischen Indikatoren trainierte der Trainingsdatensatz dem KI-System eine von rassistischer Diskriminierung verzerrte Realität an: BIPoC-Patient*innen erleben Hindernisse beim

[7] https://www.optum.com/business/solutions/health-plans/data-analytics.html

Zugang zu Gesundheitsversorgung und zudem eine weniger effektive Gesundheitsversorgung (vgl. Johnson 2022a, S. 10). Beispielsweise haben Studien gezeigt, dass BIPoC im Vergleich zu *weißen* Patienten seltener eine Schmerzbehandlung, eine potenziell lebensrettende Lungenkrebsoperation oder cholesterinsenkende Medikamente erhalten, wenn es solchen Behandlungen bedarf (vgl. Johnson 2022a, S. 11). Diese Benachteiligungen sind in den Krankenhausdaten abgebildet und führten im Fall von „OptumIQ" dazu, dass die gesundheitlichen Bedürfnisse von BIPoC-Patient*innen von dem KI-System falsch bewertet werden (vgl. Johnson 2022a, S. 11). Weiters zeigt die Nachforschung von Johnson, dass „OptumIQ" BIPoC-Patient*innen dieselbe Pflege-Stufe wie *weißen* Patent*innen zuteilt, obwohl sie im Vergleich zu den gleich kategorisierten *weißen* Patient*innen 48.772 zusätzliche chronische Krankheiten hatten und Ihnen demzufolge eine höhere Pflegestufe zustehen würde (ebd.).

KI-System „Navigate"

Dieses Beispiel zeigt, wie soziale Ungleichheiten basierend auf ethnischer Zugehörigkeit mit dem Einsatz von KI-Systemen wechselwirken, und es, ähnlich wie im vorherigen Beispiel, zu einer Übertragung und einem Festschreiben bestehender Formen der Diskriminierung in den Algorithmen von KI-Systemen kommen kann. Die im folgenden Absatz festgehaltenen Ergebnisse stammen aus einer Studie von MarkUp, einer gemeinnützigen Organisation, die investigativen Journalismus auf Basis datengestützter Untersuchungen betreibt (vgl.The MarkUp 2022). Ich habe die Organisation in Kapitel 2, Methodik genauer beschrieben.

In den USA verwenden mehr als 500 Universitäten das KI-System „Navigate[8]" des Bildungsforschungsunternehmens EAB, um unter Anderem vorherzusagen, wie wahrscheinlich es ist, dass Studierende ihr Studium abbrechen (vgl. Feathers 2022, S. 268). Ähnlich wie im vorigen Beispiel nutzt EAB Daten aus der Vergangenheit, um ihr KI-System zu trainieren. EAB entwickelt für jede Universität, die ein solches KI-System erwerben will, einen an ihren Kontext angepassten Algorithmus. Dafür nutzt EAB die Daten der Studienhistorie der jeweiligen Universität als Trainingsdatensatz (vgl. Feathers 2022, S.272). Das KI-System nutzt die ethnische Zugehörigkeit von Studierenden als Indikator für die Bewertung des Studienabbruchrisikos (vgl. Feathers

[8] https://eab.com/products/navigate/

2022, S.268). EAB beschreibt, dass der Indikator Ethnizität für die Vorhersage und Identifizierung von Studierenden, die Bedarf an gezielter Unterstützung haben sehr wirksam ist (ebd.). Die Ergebnisse des Systems, also die Risikobewertungen für einzelne Studierende liegen ihren unterrichtenden Professor*innen vor (vgl. Feathers 2022, S.269). Wie die jeweiligen unterrichtenden Personen mit den Ergebnissen umgehen, ist Ihnen überlassen, die Empfehlung des Unternehmens ist jedoch die gezielte Unterstützung derjenigen Studierenden, die Hilfe bedürfen (vgl. Feathers 2022, S.270). In der Studie untersucht MarkUp die Risikobewertungen für das Herbstsemester 2020 von sieben Universitäten, darunter große öffentliche Universitäten wie die University of Massachusetts Amherst, die University of Wisconsin-Milwaukee, die University of Houston und die Texas A&M University (vgl. Feathers 2022, S. 268). Mindestens vier der sieben Universitäten beziehen die ethnische Zugehörigkeit ihrer Studierenden als Prädiktor mit ein. Bei zwei Universitäten benennt das KI-System die ethnische Zugehörigkeit ihrer Studierenden als "hochwirksamen Prädikator" (vgl. Feathers 2022, S.269). Zwei Universitäten machen keine Angaben zu den Variablen, die in ihre Modelle einfließen (vgl. Feathers 2022, S.268).

Die Ergebnisse der Studie zeigen, dass Studierende je nach ethnischer Zugehörigkeit sehr unterschiedlich bewertet werden: Das KI-System mancher Universitäten stuft die Studienabbruchwahrscheinlichkeit bei BIPoC bis zu viermal so hoch ein wie für *weiße* Studierende (vgl. Feathers 2022, S.269). „An der University of Massachusetts Amherst beispielsweise ist die Wahrscheinlichkeit, dass schwarze Frauen als hoch riskant eingestuft werden, 2,8-mal so hoch wie bei *weißen* Frauen, und bei schwarzen Männern ist die Wahrscheinlichkeit, dass sie als hoch riskant eingestuft werden, 3,9-mal so hoch wie bei *weißen* Männern“ (frei übersetzt nach Feathers 2022, S.269). Wenn eine studierende Person für ein bestimmtes Fach mit einem hohen Risiko des Studienabbruchs bewertet ist, schlägt das KI-System andere Fächer vor, für die die Studienabbruchwahrscheinlichkeit niedriger wäre (ebd.). Feathers hält fest, dass eine solche Praxis dazu führen könnte, dass BIPoC in "leichtere" Klassen und Studiengänge gedrängt werden (vgl. Feathers 2022, S.271). Dass die Ergebnisse des KI-Systems für BIPoC höhere Studienabbruchrisiken vorhersagen, liegt an den Trainingsdatensätzen, in denen je nach Universität bis zu 10 Jahre Studienhistorie abgebildet ist (vgl. Feathers 2022, S. 272). Wenn das KI-System mit Statistiken trainiert wird, welche die rassistischen und diskriminierenden Praktiken vergangener Jahre

widerspiegeln, werden diese in die Algorithmen des KI-Systems übertragen und festgeschrieben (ebd.).

Das Unternehmen gibt an, dass ihre Kunden-Universtäten selbst entscheiden, ob sie die Ethnizität ihrer Studierenden als Indikator einbeziehen wollen, zudem offerieren sie ihren Kunden Schulungen und Beratungen unter anderem zur Vermeidung von Vorurteilen (ebd.). Nicht nur für Professor*innen sind solche rassistisch gefärbten Vorhersagen problematisch, auch universitären Berater*innen für junge Collegeabsolvent*innen liegen solche scheinbar neutralen Risikovorhersagen vor, die zu einer von Rassismen gefärbten Empfehlung für BIPoC führen können (vgl. Feathers 2022, S.269-270).

KI-System „COMPAS“

Auch in der Justiz finden KI-Systeme unter anderem für die Vorhersage von Straffälligkeit Anwendung. Das KI-System „COMPAS[9]“ des Unternehmens Equivant ist in den USA weit verbreitet und unterstützt Beamt*innen bei Entscheidungen über Kaution, Verurteilung und vorzeitige Entlassung von inhaftierten Personen (vgl. Angwin und Larson 2022, S. 265). In einer Studie testen Angwin und Larson die Ergebnisse des KI-Systems und kommen auf Basis einer Analyse von 10.000 dokumentierten Straffälligkeitseinschätzungen zu dem Ergebnis, dass sich überproportional mehr BIPoC unter denjenigen Einschätzungen befinden, die das KI-System als risikoreich für einen Rückfall einstuft, die jedoch nicht wegen erneuter Straftaten verhaftet wurden (vgl. Angwin und Larson 2022, S. 265).

> *„Die Daten zeigten, dass die Wahrscheinlichkeit, dass angeklagte BIPoC fälschlicherweise als risikoreicher eingestuft wurden, doppelt so hoch war wie bei weißen Angeklagten. Umgekehrt wurden weiße Angeklagte, die mit geringem Risiko eingestuft wurden, viel häufiger wegen neuer Straftaten angeklagt als BIPoC mit vergleichsweise niedrigen COMPAS-Risikowerten“ (frei übersetzt nach Angwin und Larson 2022, S. 266).*

Das Unternehmen bewirbt das KI-System nicht nur als neutral gegenüber ethnischer Herkunft, sondern verweist zudem auf die Trefferquote des KI-Systems. Diese liegt bei etwa 60 Prozent und ist bei BIPoC und *weißen* Angeklagten gleich hoch (vgl. Angwin

[9] https://www.equivant.com/category/compas/

und Larson 2022, S. 266). Das KI-System müsste jedoch auch eine gleich hohe Fehlerquote aufzeigen, würde es wirklich neutral gegenüber der ethnischen Herkunft funktionieren (ebd.). Dadurch, dass das Unternehmen nur die Trefferquote veröffentlicht und diese diskriminierungsfrei scheint, ignoriert oder verschleiert es die falschen Ergebnisse des KI-Systems für BIPoC und somit auch die diskriminierenden Auswirkungen seines KI-Systems für BIPoC.

Es stellt sich die Frage, ob ein KI-System ethischen Standards und dem Anspruch der Antidiskriminierung genügen kann ohne hohe Treffer- und Fehlerquoten zu generieren. Sam Corbett-Davies et al. beschreiben diese Frage als inhärentes Spannungsverhältnis zwischen der Minimierung der erwarteten Gewaltkriminalität und der Erfüllung gängiger Gerechtigkeitsvorstellungen (vgl. Corbett-Davies et al. 2017, S. 804). In ihrer Forschung, die sie unter dem Titel „Algorithmic decision making and the cost of fairness“ veröffentlichen, analysieren sie die algorithmische Verarbeitung von Strafakten. Sie kommen zu dem Ergebnis, dass es bei der Optimierung der öffentlichen Sicherheit zu Unterschieden in der Kategorisierung in Bezug auf Indikatoren der ethnischen Herkunft von Straftäter*innen kommt (vgl. Corbett-Davies et al. 2017, 797ff). Für die Erfüllung gängiger Gerechtigkeitsvorstellungen und die Minimierung des Risikos der Treffer- und Fehlerquote bedeutet das, dass mehr hochgefährdete Angeklagte freigelassen werden müssten. Das würde sich im Umkehrschluss negativ auf die öffentliche Sicherheit auswirken (vgl. Corbett-Davies et al. 2017, S. 804).

Dieses Beispiel, wie auch die Kritik von Corbett-Davies et al. zeigt, dass die Übertragung bestehender Diskriminierung in die Algorithmen von KI-Systemen komplex ist: Diese Komplexität erschwert die Gestaltung eines KI-Systems frei von Rassismen. Wie in den vorigen zwei Beispielen ist auch im Fall von „COMPAS“ der Trainingsdatensatz, auf Basis dessen das KI-System sich sein Regelsystem baut, ausschlaggebend für die Ergebnisse.

Beispielsweise kann die Anzahl an Verhaftungen einer Person, Corbett-Davies et al. zufolge ein rassistisch verzerrter Indikator sein. Eine „[e]ine verstärkte Polizeipräsenz in Minderheitenvierteln kann dazu führen, dass schwarze Angeklagte häufiger verhaftet werden als *weiße*, die das gleiche Verbrechen begehen“ (frei übersetzt nach Corbett-Davies et al. 2017, S. 804). Im Fall des KI-Systems, welches einschätzt wie hoch das Risiko ist, das eine Person wieder straffällig wird, kann folgendes passieren: Das KI-System nutzt die Anzahl an Verhaftungen einer Person als Indikator und stuft

demzufolge Personen, die bereits öfter verhaftet wurden als risikoreicher für wiederholte Straffälligkeit ein. Eine Person aus einem Minderheitenviertel mit erhöhter Polizeipräsenz wurde öfter verhaftet als eine Person, die aus einem anderem Viertel kommt. Beide Personen waren in selbem Maße straffällig, aber nur bei der einen Person kam bei wiederholtem Mal zu einer Verhaftung.

Der bewusste Ausschluss von Merkmalen ethnischer Herkunft, Hautfarbe oder Stellvertretermerkmale, die mit den eben genannten korrelieren, werden als „colorblind design“ bezeichnet (vgl. Daniels 2015, S. 1377). Ich habe in meiner Recherche keine Studien zu der Wirksamkeit von „colorblind design“ von KI-Systemen gefunden. Schelenz kritisiert „colorblind design“: Ihr zufolge verschleiert es historisch gewachsene Ungleichheiten (vgl. Schelenz 2021, S. 83). Eine kritische Frage, die sich für mich bei „colorblind design“ stellt, ist, wie eine Umsetzung dessen im Fall von KI-Systemen funktioniert. Bei der nicht-intelligenten algorithmischen Verarbeitung von Daten ist es leicht, aus strukturierten Daten sensible Datenpunkte auszuschließen und somit ein „colorbind design“ zu realisieren. Diese Vorgehensweise lässt sich jedoch nicht auf KI-Systeme übertragen: Die unstrukturierten Daten, die von KI-Systemen verarbeitet werden, wie auch die Art und Weise mit der KI-Systeme Korrelationen zwischen Datenpunkten analysieren, macht die Überwachung der Datenanalyse kompliziert. Das KI-System ist eine Blackbox, wie ich es bereits in Kapitel 3.1.4 und 3.2.4 beschrieben habe. Aus diesen zwei Punkten schlussfolgere ich, dass ein Eingriff, also das Herausstreichen sensibler Datenpunkte und die Identifikation von Stellvertretermerkmalen sowie andere Strategien der Umsetzung eines „colorblind designs“ schwer umsetzbar sind.

Das „colorblind design“ ist eine technische Lösung für das Problem der algorithmischen Diskriminierung. Noble argumentiert, dass KI-Systeme strukturell gedacht werden müssen und dass solche technischen Lösungsansätze keine Antwort auf diskriminierende Entscheidungen der KI-Systeme sein können (vgl. Noble 2019, S. 166–168). Alle Beispiele zeigen, wie nach ähnlichen Wirkmechanismen bestehende Formen der Diskriminierung und dadurch bestehende soziale Ungleichheiten in KI-Systeme übertragen und in ihren Algorithmen langfristig festgeschrieben werden. Der Rückgriff auf Daten, welche die Vergangenheit abbilden, trägt zu der Verzerrung der KI-Systeme bei. Lücking spricht in diesem Fall von einem tendenziellen Konservatismus bei KI-Systemen, sowie Projektionen der Vergangenheit in die

Zukunft (Lücking 2020, S. 70). Es ist wichtig zu erkennen, dass es die Datenbasiertheit der KI-Systeme ist, die in den Beispielen eine Übertragung von Ungleichheiten bewirkt hat**.** Datensammlungen, wie beispielsweise die Studienhistorie unterschiedlicher Universitäten, justizielle Strafakten oder Krankenakten aus Krankenhäusern bilden Entscheidungen von Menschen ab, die Rassismen und andere Formen von Diskriminierung verinnerlicht haben. Dementsprechend bilden sie strukturelle Exklusion und Marginalisierung ab. Eine KI, die aus diesen Daten ihr algorithmisches Regelsystem ableitet, automatisiert diese Formen der Diskriminierung (vgl. Simon 2020, S. 40). Neben Simon, Lücking und Nassehi warnen viele weitere Autor*innen vor dem Ein- und Festschreiben problematischer Macht- und Hierarchiestrukturen sowie Mustern der Marginalisierung und der Ungleichheit in die Technik (Lücking 2020; Weber und Prietl 2021; Simon 2020; Nassehi 2019; Algorithm Watch 2020b).

3.2.2 Algorithmische Kodierung von Realität basiert auf Kategorisierung und Normierung – Die Norm ist der *weiße* Mann

Ähnlich der Argumentation in 3.1.2 wird die Ungleichheitsdimension Ethnizität von KI-Systemen durch die algorithmische Funktionsweise insofern modifiziert, als dass sie von einer durch Daten erlernten Norm ausgehend kategorisieren und bewerten. Androzentrismus und andere Normen wie heterosexuelle Dominanz und Heteronormativität prägen die Gesellschaft (vgl. Hartmann et al. 2008, S.9ff). Schelenz erklärt, dass dies dazu führt, dass *weiß*-sein, Männlichkeit und Heterosexualität als „normal“ wahrgenommen wird und damit für Angehörige dieser Gruppe unsichtbar ist (vgl. Schelenz 2021, S. 86).

Die Norm des *weißen* Mannes spiegelt sich nicht nur in KI-Systemen wider, sondern auch in anderer Technik, wie folgendes Beispiel eines elektronischen Seifenspenders in öffentlichen Toiletten zeigt: Im Jahr 2019 gingen mehrere Videos viral, in denen BIPoC zeigen, wie der Mechanismus des Seifenspenders nicht ausgelöst wird, wenn sie ihre Hand darunter halten (vgl. ECIAIR 2019, S. 297). Erst die Hand einer *weißen* Person oder eines Taschentuchs aktiviert die Seifenspender Funktion (ebd.). Die technische Funktionsweise und deren ungleichheitsmodifizierenden Auswirkungen sind für KI komplexer. Mit folgendem Beispiel veranschauliche ich den ungleichheitsmodifizierenden Wirkmechanismus der Funktionsweise der Kategorisierung und Bewertung ausgehend von der Norm des *weißen* Mannes von KI-Systemen. Die KI-Anwendung stammt aus dem Akquise Bereich von Unternehmen:

KI-Systeme können in der ersten Phase eines Einstellungsprozesses mittels Gesichts- und Spracherkennung aus vielen Bewerber*innen die herausfiltern, die dem algorithmischen Entscheidungssystem die gefragten Kompetenzen und Charaktereigenschaften liefern (vgl. Erden 2020, S. 81).

KI-System „HireVue“

Mehr als 100 große Unternehmen, wie beispielsweise Unilever und Hilton nutzen dafür das KI-System der Recruiting-Firma HireVue[10]. Das KI-System analysiert Gesichtsbewegungen, Wortwahl und Sprechstimme von Bewerber*innen, und bewertet damit deren Beschäftigungsfähigkeit. Trotz kritischer Stimmen wurden bereits mehr als eine Million Arbeitssuchende analysiert (vgl. Harwell 2022, S. 206). Was als gute – beziehungsweise schlechte Leistung bewertet wird, leitet das KI-System von Interviews erfolgreicher Arbeitnehmer*innen ab (vgl. HireVue 2022). Deniz Erden erklärt, wie eine solche Software beispielsweise nicht-westliche Frauen diskriminieren kann: Während Lächeln in Deutschland und China als intelligent gelte, bedeutet es im Iran das Gegenteil (vgl. Erden 2020, S. 80–81). Ähnlich verhält es sich mit Blickkontakt: In westlichen Kulturen ist Blickkontakt zwischen Frauen und Männern anders attribuiert als in nicht-westlichen Ländern (ebd.).

Kulturelle Unterschiede werden von KI-Systemen nicht als solche kontextualisiert und wenn die Norm ausgehend von verzerrten Trainingsdatensätzen die des *weißen* Mannes ist, kann das KI-System die von der Norm abweichenden Merkmale als negativ bewerten (vgl. Erden 2020, S. 80–81). Harwell führt weitere potenzielle „Normabweichungen“ auf die von einem als KI-System dementsprechend negativ bewertet würden, wie beispielsweise Nicht-Muttersprachler*innen oder nervöse Gesprächspartner*innen (vgl. Harwell 2022, S. 206). Zusammenfassend leite ich ab, dass alle Menschen deren Gesichtsbewegungen, Wortwahl und Sprechstimme nicht der Norm entsprechen, negativ bewertet würden. Die Webseite des Unternehmens HireVue bewirbt ihr KI-System gegenteilig: Sie argumentiert, dass mit ihrem System Voreingenommenheit reduziert wird und Bewerber*innen mehr Konsistenz und Fairness im Vorstellungsgespräch entgegenbracht wird (vgl. HireVue 2022). Das Unternehmen bietet jedoch keinen Einblick in das algorithmische Regelsystem, nach dessen Bewertungen Bewerber*innen als fähig eingestuft werden. Somit existiert

[10] https://www.hirevue.com/platform/online-video-interviewing-software

keine Nachvollziehbarkeit für die Argumente von „HireVue“ und es ist unklar, ob das KI-System wirklich mehr Fairness und weniger Voreingenommenheit im Einstellungsprozess bringt.

Ähnlich wie „HireVue“ argumentieren auch Altenburger und Schmidpeter, dass KI-Systeme förderlich für die Gleichbehandlung von Bewerber*innen sein können, weil diese Systeme im Gegensatz zu einem Menschen Vorurteile und Bias aufgeben können (vgl. Altenburger und Schmidpeter 2021, S. 333). Ich erkenne jedoch eine Widersprüchlichkeit in dieser Position: KI-Systeme beruhen wie in Kapitel 3.1.1. bereits gezeigt, auf Sammlungen von Entscheidungen von Menschen in der Vergangenheit, deren inhärente Bestandteile Rassismen sind (vgl. Simon 2020, S. 40; Beattie und Johnson 2012, S.7). Die Anwendungsbeispiele von KI haben gezeigt, dass KI-Systeme in ihrer Datenanalyse kategorisieren und bestimmte Merkmale als Norm identifizieren. Obwohl KI-Systeme wie „HireVue“, „COMPAS“ oder „Navigate“ von ihren Firmen damit beworben werden, dass sie für Fairness stehen, dass sie gegen Ungleichheiten im Bildungsbereich vorgehen oder, dass sie (im Fall von „COMPAS“) frei von rassistischer Ungleichbehandlung sind, stehen die KI-Systeme trotzdem in der Kritik algorithmisch zu diskriminieren. Hier stellt sich mir die Frage, ob eine diskriminierungsfreie Gestaltung von KI-Systemen, solange diese auf Kategorisierung und Normierung beruhen, überhaupt möglich ist.

Ein weiterer Bereich für den KI-Systeme eingesetzt werden, ist für die Bewertung der Kreditwürdigkeit (vgl. European Union Agency for Fundamental Rights 2020b, S. 68). Beim Beantragen eines Kredites prüft die Bank oder eine andere kreditgewährende Institution die Wahrscheinlichkeit, mit der die beantragende Person den Kredit zurückzahlen kann (vgl. European Union Agency for Fundamental Rights 2020b, S. 71–72). Ist das Risiko, dass die Person das nicht kann, zu hoch, wird der Kredit nicht bewilligt. Die Überprüfung der Bonität der Kredit-beantragenden Person ist mit Einsatz eines KI-Systems schneller und weniger kostenaufwendig (vgl. Aggarwal 2021, S.43). Immer mehr Banken kaufen daher sogenannte „computergenerierte Ratingsysteme“ von externen Unternehmen wie OnDeck[11] ein oder lassen sich intern ein System programmieren (vgl. Aggarwal 2021, S. 42). Zur Beurteilung der Kreditwürdigkeit werden Daten verwendet, die in Zusammenhang mit personenbezogenen Daten

[11] https://www.ondeck.com/

stehen (vgl. Müller 2021, S.275ff; European Union Agency for Fundamental Rights 2020b, S. 71–72).

KI-System der Schufa

In Deutschland ist die Aktiengesellschaft Schufa[12] für die Bewertung der Kreditwürdigkeit zuständig (vgl. Schufa 2022). Daten zu Nationalität, Beruf, Einkommen oder Familienstand werden nicht als Indikatoren in das algorithmische Regelsystem der Schufa einbezogen (vgl. Eschholz und Djabbarpour 2018, S. 64–65). Jedoch kann der Wohnort einer Person, wenn das von Kunden der Schufa, also Banken oder Unternehmen, gewollt ist, als Indikator miteinbezogen werden (vgl. Verbraucherzentrale 2014). Der Wohnort ist, wie Poster in seinem Beitrag zu rassifizierter Überwachung beschreibt, genauso wie der Name oder Bilder einer Person, ein Stellvertretermerkmal für die Zugehörigkeit einer Person zu einer sozialen Gruppe, unter anderem für die der Ethnizität (vgl. Poster 2019, S. 134). Im Jahr 2014 und 2015 wurde der Einsatz des Bonitätsprüfungsindikators „Wohnort“ medial kritisiert (vgl. Welt 2015). Trotz Kritik habe ich keine Äußerungen zu einer Änderung dieser Vorgehensweise seitens Schufa finden können. Mehr Informationen zu dem algorithmischem Regelsystem von Schufa oder anderen KI-Systemen, die für die Bewertung der Kreditwürdigkeit eingesetzt werden, waren im Rahmen meiner Recherche nicht zugänglich. Im Fall von Schufa ist es zwar möglich einmal im Jahr Einsicht in die Daten zu bekommen, welche für die Bewertung zur Verfügung stehen (vgl. Schufa 2022), jedoch hält das Unternehmen die Logik, nach denen die Entscheidungssysteme funktionieren, unter Verschluss (vgl. Eschholz und Djabbarpour 2018, S. 64–65).

Das Risiko, dass es zu einem Einschreiben von Rassismen in die Algorithmen der KI-Systeme zu Kreditwürdigkeitseinschätzung kommt und dadurch zu einer niedrigeren Kreditwürdigkeitseinschätzung aufgrund von ethnischer Herkunft oder Stellvertretermerkmalen desselben, ist gegeben. Die Intransparenz der Logik, nach der ein solches System die Bewertungen vornimmt, lässt jedoch keine weiteren Rückschlüsse ziehen. In Kapitel 3.1.1 wird das Beispiel von KI-Systemen für die Kreditwürdigkeitsprüfung wieder aufgenommen, jedoch nicht unter dem Augenmerk der Funktionsweise von KI-Systemen, welche im Mittelpunkt der Kategorie in diesem

[12] https://www.schufa.de/

Kapitel steht, sondern mit dem Fokus auf die ungleiche Repräsentativität von sozialen Gruppen und Individuen in Daten.

KI-System von Airbnb

Will eine Person eine Unterkunft bei Airbnb[13] zur Vermietung anbieten, schlägt Airbnb auf Basis vorliegender Daten zu der Unterkunft, zur Lage und zur vermietenden Person einen Mietpreis für die Unterkunft vor (vgl. Zhang et al. 2021, S.3). Die vermietende Person kann nicht mehr Miete für die Unterkunft verlangen als den Preis, den das KI-System von Airbnb für die jeweilige Unterkunft vorgeschlagen hat. Eine Studie kommt zu dem Ergebnis, dass die Häuser und Wohnungen von BIPoC anders bewertet werden, wenn sie auf dem virtuellen Markt zum Tausch angeboten werden: Beispielsweise können *weiße* Gastgeber auf Airbnb für gleichwertige Wohnungen mit ähnlichen Fotos in New York City 12 % mehr Miete für ihre Unterkünfte verlangen als BIPoC (vgl. Edelman und Luca 2014, S. 1ff). Die Intransparenz der Logik, nach der ein solches System die Bewertungen vornimmt (vgl. Zhang et al. 2021, S.14), lässt – ähnlich im vorigen Beispiel – zwar keine weiteren Rückschlüsse zu, jedoch stehen die Ergebnisse der Studie für sich: Sie lassen erkennen, dass *weiß*-sein eine Normierung erfährt. Diese Normierung bringt Formen rassistischer Diskriminierung mit sich und hat somit eine ungleichheitsverstärkende Wirkung. Die Fundamental Rights Agency erklärt in einem Vortrag zu künstlicher Intelligenz und Grundrechten, dass wenn Personen aufgrund einer mittels statistisch-algorithmischer Analyse definierten Norm kategorisiert und bewertet werden, sowohl Persönlichkeitsrecht, Datenschutz als auch Gleichheit – als zentrale Rechtsgüter – zur Disposition stehen (vgl. European Union Agency for Fundamental Rights 2020c, S.12-13).

3.2.3 Überwachung durch KI-Systeme betrifft marginalisierte Gruppen stärker

Wegen der Verortung des Diskurses von Überwachung in black feminist theory und im postkolonialen Diskurs sind folgende Argumente dem Kapitel 3.1.2 und damit der Ungleichheitsdimension Ethnizität zugeordnet. Das folgende Argument betrifft jedoch nicht nur die Ungleichheitsdimension Ethnizität, sondern im weiteren Sinne auch die Dimensionen Klasse, Staatszugehörigkeit und Gender.

Durch KI-Systeme ist die Vernetzung von intelligenten und nicht-intelligenten Technologien möglich (vgl. AI Watch 2020, S. 13). Zudem machen KI-Systeme die

[13] https://www.airbnb.com/

Analyse unstrukturierter Datensätze möglich, was die Auswertung verschiedenster Formen von persönlichen Daten über Menschen möglich macht (vgl. Punia et al. 2021, S. 60). Poster führt hier Sound (Akzente oder Dialekte, Arten des Sprechens), Bilder (Gesichts- und Körpermerkmale, Klamotten), Schrift (Lebensläufe, Informationen in sozialen Medien), physische Bewegung und Aufenthalt als gesammelte, in Daten gespeicherte und für die Ungleichheitsdimension Ethnizität sensible Informationen über Menschen auf (vgl. Poster 2019, S. 134). KI-Systeme kommen in unterschiedlichen Anwendungsbereichen zum Einsatz und bewerten die gesammelten Daten und im Umkehrschluss den Menschen (vgl. European Commission 2020, S. 2). Poster überträgt die Funktionsweisen von KI auf das Konzept der Mehrfachüberwachung (vgl. Poster 2011, S.868ff). Der Begriff der Mehrfachüberwachung erweitert die gängige Auffassung von Überwachung, in der Gruppen von Menschen mit Macht, andere Menschen mit weniger Macht überwachen (ebd.): Die Erweiterung des Konzepts der Überwachung wird dem Intersektionalitäts-Ansatz gerecht, in dem Machtdynamiken, wie im Konzept der Mehrfachüberwachung nicht binär, sondern als komplexes System sich überschneidender Machtrelationen aufgefasst werden (vgl. Aulenbacher und Riegraf 2012, S.1). Poster zufolge zeigt das Konzept der Mehrfachüberwachung, dass digitale Systeme und insbesondere KI-Systeme das Spektrum der Agenten, die überwachen, erweitern und damit neue Quellen für Überwachung bieten (vgl. Poster 2011, S.868ff). Als neue Akteure der Mehrfachüberwachung führt Poster auf: „Unternehmen, die Dienstleistungen in Auftrag geben; Anbieter*innen, die die Technologien bereitstellen, Arbeitnehmer*innen, die die Dienstleistungen erbringen sowie die Verbraucher und Nutzer*innen der Dienstleistungen“ (frei übersetzt nach Poster 2019, S. 135). Das Konzept bricht die gegenseitige Exklusivität der Rollen auf, denn es besagt, dass eine Person von einer Gruppe beobachtet werden kann, während sie gleichzeitig der Beobachter einer anderen Gruppe ist (ebd.). Schmidt und Mellentin beschreiben ähnlich zu Poster eine kategorische Überwachung durch den Einsatz von KI-Systemen, weil eine große Menge an Daten aus unterschiedlichen Quellen genutzt wird, die grundsätzlich alle Menschen miteinschließt, deren Aktivitäten in Daten abgebildet sind (Schmidt und Mellentin 2020, S. 16). Auch Shoshana Zuboff stellt in ihrem Buch „Überwachungskapitalismus“ eine ähnliche These auf. Sie begründet diese jedoch mit den Wirkmechanismen der Wertschöpfungsprozesse im heutigen kapitalistischen System (vgl. Zuboff 2019, S. 11). Eine weitere, an die vorigen Thesen und Konzepte

anschließende Argumentation findet sich in den Veröffentlichungen der Nichtregierungsorganisation Algorithm Watch wieder, die aufgrund der aktuellen Entwicklungen in Bezug auf Gesichtserkennung davor warnt, dass es zu einer undurchsichtigen Massenüberwachung kommt, wenn dem Trend der Installation von Gesichtserkennungssoftware und Kameras keine greifenden Datenschutzrechte entgegengesetzt werden (vgl. Algorithm Watch 2020b, S. 7).

Für die Beantwortung der Forschungsfrage ist vor allem relevant, dass Überwachung nicht für alle Menschen die gleichen regulierenden Eingriffe zur Folge hat, sondern konstituiert ist von Macht- und Hierarchiestrukturen (vgl. Schmidt und Mellentin 2020, S. 17; Poster 2019, S. 134; Roberts 2019, S. 341; Scannell 2019, S. 108; Schelenz 2021, S. 79ff). De-privilegierte, marginalisierte Gruppen erfahren somit andere Einschränkungen durch Überwachung als privilegiertere Mitglieder der Gesellschaft.

3.2.4 Fehlende Gerichtsbarkeit und Regulierung - KI als Blackbox

Das Grundrecht auf Antidiskriminierung wurde bereits in Kapitel 3.1.1 dargelegt und ist für diese Kategorie gleich relevant wie für die der Ungleichheitsdimension Gender. Im Folgenden soll unter anderem am Beispiel eines KI-Systems zum Zweck der Gesichtserkennung gezeigt werden, wie ein weiterer Artikel des europäischen Grundgesetzes durch KI-Systeme beschnitten werden kann: Das Recht auf informationelle Selbstbestimmung (vgl. Europäische Union 2020).

Kommt es zu einer diskriminierenden Bewertung oder Kategorisierung durch ein KI-System kann diese entweder unerkannt bleiben oder der betreffenden Person auffallen. Im zweiten Fall kommt ein weiteres, für die Beantwortung der Forschungsfrage relevantes Argument zum Tragen, und zwar die fehlende Gerichtsbarkeit und Regulierung von KI-Systemen, welche eng mit der Intransparenz und fehlenden Nachvollziehbarkeit der KI-Systeme zusammenhängt.

Eine der Folgen der Datenbasiertheit von KI-Systemen ist, dass viele Daten von Menschen gesammelt und analysiert werden müssen. Je nach Zweck und Einsatzbereich des Systems werden unterschiedliche Mengen und Typen von Daten gesammelt. Siems und Repka zufolge gibt es aus juristischer Perspektive drei relevante Kategorien für Daten (vgl. Siems und Repka 2021, S. 3). Für jede dieser drei Kategorien sind jeweils unterschiedliche Gesetze und Richtlinien relevant: Die erste Kategorie umfasst Inhalte wie zum Beispiel Texte oder Bilder. Bei der zweiten handelt es sich um Rohdaten, wie beispielsweise von Maschinen aufgezeichnete Daten. Die

dritte Kategorie ist die der personenbezogenen Daten (vgl. Siems und Repka 2021, S. 3). Daten sind je nach Kategorie in ihrer Verwendung durch Datenschutzgesetze, wie beispielsweise die Datenschutzgrundverordnung (=DSGVO) der Europäischen Union auf unterschiedliche Weise geschützt und dadurch beschränkt nutzbar (vgl. Europäische Union 2020). Inhalte finden sich vor allem auf Plattformen wie Twitter und Facebook wieder und unterstehen dementsprechend der Verfügung der jeweiligen dahinterstehenden Unternehmen (vgl. Wenger 2020, S. 173). Rohdaten sind Daten, die von Maschinen generiert werden (vgl. Siems und Repka 2021, S. 3), beispielsweise einem Fahrkartenautomaten, der aufzeichnet wann welche Tickets gekauft werden. Personenbezogene Daten sind Daten, die nicht anonym, sondern einer bestimmten Person zugeordnet sind und Rückschlüsse auf deren Persönlichkeit erlauben. Nach der, im Jahr 2018 in Kraft getretenen DSGVO wird innerhalb der Europäischen Union zwischen allgemeinen personenbezogenen Daten und besonderen personenbezogene Daten unterschieden (vgl. Europäische Union 2020). Die besonderen personenbezogenen Daten sind Informationen über die ethnische und kulturelle Herkunft, politische, religiöse und philosophische Überzeugungen, Gesundheit, Sexualität, Gewerkschaftszugehörigkeit, genetische Daten, biometrische Daten zur eindeutigen Identifizierung einer Person und Gesundheitsdaten (ebd.). Die Verarbeitung dieser besonderen personenbezogenen Daten ist mit wenigen Ausnahmen verboten (ebd.). Eine solche Ausnahme ist beispielsweise der Schutz eines lebenswichtigen Interesses der betroffenen Person. Es reicht jedoch auch die Zustimmung der betroffenen Person, um ihre besonderen personenbezogenen Daten verarbeiten zu dürfen (ebd.). Anonymisierte Daten fallen nicht unter die DSGVO (ebd.). Die Bestimmungen der europäischen DSGVO ersetzen das nationale Recht für Datenschutz der Staaten der EU (vgl. Schunicht 2018, S. 2). Sie schreibt fest, dass Menschen das Recht auf informationelle Selbstbestimmung haben, was bedeutet, dass die Preisgabe und die Verwendung von personenbezogenen Daten selbstbestimmt funktionieren muss (vgl. Schunicht 2018, S.309). Das Recht hat Schunicht zufolge eine „Sicherungsfunktion für die individuelle Freiheitsausübung“ (Schunicht 2018, S.309). Guijarro-Santos führt an, dass es beim Einsatz von KI-Systemen jedoch beschnitten werden kann (vgl. Guijarro-Santos 2020, S. 50). Folgendes Beispiel einer KI für Gesichtserkennung veranschaulicht das Argument von Guijarro-Santos. Es zeigt, wie das KI-System PimEyes eine Form der Überwachung ermöglicht, die das Recht auf informationelle Selbstbestimmung beschneidet. Zudem

handelt es sich bei PimEyes[14] um ein KI-System, welches das Argument, dass KI-Systeme das Spektrum der Überwachung erweitern, untermauert: Es ist ein öffentlich zugängliches und in beschränktem Ausmaß kostenfrei nutzbares KI-System zur Gesichtserkennung. Es gleicht ein auf der Website hochgeladenes Bild mit 900 Millionen Bildern aus dem Internet ab und stellt so ein digitales Persönlichkeitsprofil zusammen (vgl. PimEyes 2022). Während Gesichtserkennungssoftware lange Zeit nur für bestimmte Gruppen zugänglich war, ist sie mit PimEyes für jede*n frei verfügbar (vgl. PimEyes 2022). Ohne eine gesetzliche Regelung für die Nutzung der von PimEyes analysierten Daten kommt es zur Missachtung des Rechts auf Privatheit. Das betrifft alle Personen, die Fotos in sozialen Netzwerken hinterlegt haben oder deren Gesichter auf hochgeladenen Bildern anderer Personen sichtbar sind. Wie im vorigen Kapitel 3.2.3 bereits beschrieben, wirkt Überwachung für marginalisierte Gruppen gravierender (vgl. Schelenz 2021, S. 84).

Die Fundamental Rights Agency fordert für den Einsatz von Systemen künstlicher Intelligenz eine legale Verarbeitung von Daten, die nicht zu unfairer Behandlung führt; die Anfechtbarkeit oder Gerichtsbarkeit von Bewertungen von KI-Systemen; die Bewusstmachung von Betroffenen, dass ein KI-System eingesetzt wird, wenn es eingesetzt wird und die Nachvollziehbarkeit und Transparenz der Logik nach der die algorithmischen Regelsysteme funktionieren (vgl. European Union Agency for Fundamental Rights 2020c, S.12-13). Die Umsetzung solcher Forderungen können der Fundamentalt Rights Agency zufolge die Beschneidung der Persönlichkeitsrechte wie auch die Modifikation der sozialen Ungleichheit durch KI abfedern (ebd.).

3.2.5 Kann KI sozialer Ungleichheit entgegenwirken?

Die Aufdeckung von algorithmischer Diskriminierung durch KI-Systemen hat, beispielsweise durch den im Kapitel 3.2.1.a geschilderten Fall der fehlerhaften Verhaftung von Robert Julian-Borchak Williams wiederholt zu Diskussionen über Rassismen in der Gesellschaft geführt. Auch im wissenschaftlichen Bereich, insbesondere in den Technik-Gesellschafts-Wissenschaften findet ein Diskurs zu den Fällen algorithmischer Diskriminierung statt (vgl. Prietl 2019, S. 305), der wiederum zu einer Reflektion bestehender Ungleichheitsverhältnisse führt. Die Komplexität der Wechselwirkungen bestehender sozialer Ungleichheiten und deren Übertragung in die algorithmischen Regelsysteme der Technik zwingt zu einer genauen

[14] https://pimeyes.com/en

Auseinandersetzung mit den Wirkmechanismen von Machtrelationen und -hierachien, sowie institutionalisierten Formen der Diskriminierung wie Ntoutsi et al., Weber, Lücking und andere Wissenschaftler*innen zeigen (Prietl 2019, S. 303; Weber und Prietl 2021, 58ff; Ntoutsi et al. 2020, S. 3–4; Frey et al. 2021, 19ff; Eglash 2019, 227ff; Lücking 2020, 65ff). Im Falle des KI-Systems des AMS in Österreich, der in Kapitel 3.1.1 beschrieben wird, argumentieren Befürworter*innen von KI, das gerade durch das Sichtbarmachen statistisch belegbarer Unterschiede eine effektivere Verteilung staatlicher Ressourcen gelingen kann (vgl. Fröhlich und Spiecker 2018, o.S.). Es gibt zudem KI-Systeme deren Funktion die Bekämpfung von Rassismen ist: Beispielweise werden in Italien KI-Systeme für Gesichts- und Stimmerkennung in Fußballstadien eingesetzt, um rassistische Übergriffe zu bekämpfen (vgl. Algorithm Watch 2020b, S. 7). Durch dieses KI-System werden Personen, denen der Zutritt zu Fußballspielen untersagt ist, beim Zutritt in Sportstadien identifiziert. Zudem erkennt und überwacht das System Personen in den Stadien, die in ihrer Vergangenheit in rassistische Übergriffe verwickelt waren (ebd.).

3.3 Modifikationen der Ungleichheitsdimension Klasse durch KI

In diesem Kapitel thematisiere ich n die Implikationen von KI auf die Ungleichheitsdimension Klasse. Ich fasse die gesammelten Argumente und Thesen in den jeweiligen Kategorien zusammen und beschreibe ihre Zusammenhänge. Manche der Wirkmechanismen veranschauliche ich mittels Anwendungsbeispielen von KI-Systemen. Es ist zu beachten, dass Klasse – wie bereits in Kapitel 2. Methodik beschrieben – intersektional wirkt, d.h. verwoben ist mit anderen Kategorien sozialer Differenzierung wie Ethnizität und Gender, Alter, soziale Herkunft oder sexuelle Orientierung (vgl. Horwath 2022, S.75).

3.3.1 Datenbasiertheit von KI und Verzerrungen von Algorithmen durch Daten

Wie bereits in Kapitel 3.1.1.a beschrieben, bezeichnet der Begriff der digitalen Kluft die ungleiche Repräsentativität bestimmter sozialer Gruppen in Daten (vgl. Aiello et al. 2020, S. 108). Menschen mit niedrigem und mittlerem Einkommen sind in Trainingsdatensätzen im Gegensatz zu privilegierten Mitgliedern der Gesellschaft unterrepräsentiert (vgl. Rubinov und Hagerty 2019, S. 2). Diese ungleiche Repräsentativität ist unter anderem Grund für inakkurate Ergebnisse und Bewertungen von KI-Systemen. Im Falle von KI-Systemen, die für Bonitätsprüfung eingesetzt werden, hat das für betroffene Personen große Auswirkungen. Beispielsweise bei der

Vergabe von Krediten oder Mietwohnungen, bei der Einstufung für Versicherungen oder bei der Einstellung von Mitarbeitern können die Einschätzungen von KI-Systemen dieser Art benutzt werden, um Bewerber*innen einzuschätzen. Wie bereits in Kapitel 3.2.2 beschrieben, können diskriminierende Bewertungen dieser KI-Systeme strukturelle Ungleichheiten vertiefen, da sie (mit)entscheiden wer in der Lage ist, einen Kredit aufzunehmen, ein Haus zu mieten, versichert zu werden oder ein Stellenangebot zu erhalten (vgl. Blattner und Nelson 2021, S.29).

Blattner und Nelson analysieren in ihrer Studie Kreditberichte von 50 Millionen anonymisierten US-Verbraucher*innen und verknüpften jede*n dieser Verbraucher*innen mit den sozioökonomischen Angaben aus einem Marketing-Datensatz, ihren Eigentumsurkunden und Hypothekentransaktionen sowie Daten über die Hypothekenkreditgeber, die ihnen Darlehen gewährten (vgl. Blattner und Nelson 2021, S.7-8). Ausgehend von dieser Analyse ist es möglich nachzuvollziehen wer, auf Basis welcher Daten, wie von den KI-Systemen zur Kreditwürdigkeitseinschätzung bewertet wurde (ebd.). Die Ergebnisse der Studie zeigen, dass Minderheiten und einkommensschwache Gruppen weniger Daten in ihrer Kredithistorie haben und es dadurch für Angehörige dieser Gruppe zu inakkuraten Ergebnissen bei der Kreditwürdigkeitseinschätzung kommt (ebd.). Das bedeutet, dass KI-Systeme die Kreditwürdigkeit dieser Personen als kreditunwürdiger oder kreditwürdiger einschätzen, als das bei anderen Personen der Fall ist deren Daten repräsentativer für ihre Person sind (vgl. Blattner und Nelson 2021, S.1). Ob es im Durchschnitt zu mehr fehlerhaften Einstufungen der Kreditunwürdigkeit kommt, als zu fehlerhaften Einschätzungen der Kreditwürdigkeit wird in der Studie nicht gemessen. Auch mittels weiterer Recherchen konnte ich keine Informationen zu dieser Frage generieren. Das Thema der Datenungenauigkeit und Fehleranfälligkeiten bei KI-Systemen wird jedoch nicht nur von Blattner und Nelson thematisiert. Auch die Nichtregierungsorganisation Algorithm Watch kritisiert, dass KI-Systeme ihre verfolgten Ziele beim Einsatz oft nicht realisieren, unter anderem weil die Datenlage schlecht ist (Algorithm Watch 2020b, S. 6). Unter dem Begriff „noisy data“ wird eine mit diesem Phänomen zusammenhängende Problematik diskutiert: Chen und Beam zufolge bringt der Zugang zu unstrukturierten, webbasierten Daten Herausforderungen mit sich, weil die Informationen sich in ihrer Reliabilität von gesicherten Quellen unterscheiden (vgl. Chen und Beam 2019, S. 93). Sind Informationen weniger zuverlässig und bringen dadurch fehlerhafte Ergebnisse hervor, ist das durch die Intransparenz von KI-

Systemen nicht nachvollziehbar für Betroffene. Aus demselben Grund sind „noisy data“ auch für Entwickler*innen und KI-anbietende Unternehmen problematisch. Algorithm Watch berichtet in Ihrer Studie über Anwendungen von KI-Systemen in europäischen Ländern, dass KI-Systeme von Unternehmen weiter verwendet und oder angeboten werden, obwohl Nachforschungen zeigen, dass die Analysen oder Ergebnisse der KI-Systeme fehlerhaft sind (vgl. Algorithm Watch 2020b, S. 7).

3.3.2 Überwachung durch KI-Systeme betrifft marginalisierte Gruppen stärker

Wie die Argumente aus Kapitel 3.2.3 bereits gezeigt haben, betrifft Überwachung durch KI-Systeme marginalisierte Gruppen stärker. Während ich in Kapitel 3.2.3 die Modifikationen sozialer Ungleichheit auf die Ungleichheitsdimension Ethnizität thematisiert habe, geht es im folgenden Kapitel um die Auswirkungen auf die Ungleichheitsdimension Klasse.

KI-System „SyRI“

In den Niederlanden wird an der Entwicklung eines KI-Systems zur Aufdeckung von Betrug für Sozialgeld gearbeitet. Die Regierung will Daten unterschiedlicher Quellen nutzen, auf Basis derer das KI-System entscheidet, ob eine Person womöglich das soziale Auffangnetz des Staates missbraucht (vgl. Algorithm Watch 2020b, S.161). In Dänemark wurde ein KI-System „SyRI“ (= System Risico Indicatie) mit demselben Zweck eingesetzt (vgl. Algorithm Watch 2020c). Nach Protesten der Zivilgesellschaft ist die Funktionsweise von „SyRI“ vor Gericht gekommen und im Jahr 2020 wurde das gerichtliche Urteil gesprochen (ebd.): Die Rechtsvorschriften, die die Verwendung von SyRI regeln, verstoßen gegen höheres Recht (vgl. De Rechtspraak 2020). Das Gericht hat geprüft, ob das KI-System „SyRI“ mit der Gesetzgebung des Artikel 8 Absatz 2 EMRK vereinbar ist (ebd.). Diese Bestimmung verlangt ein vernünftiges Verhältnis zwischen dem sozialen Interesse, dem das KI-System dienen soll, und dem Eingriff in das Privatleben durch die Verarbeitung der personenbezogenen Daten durch das System (ebd.). Das Gericht hat die Ziele von „SyRI“, nämlich die Verhinderung und Bekämpfung von Betrug im Interesse des wirtschaftlichen Wohlergehens, mit dem Eingriff in die Privatsphäre, den der Einsatz von „SyRI“ mit sich bringt verglichen und ist zu dem Ergebnis gekommen, dass keine ausreichende Rechtfertigung für den Eingriff in das Privatleben vorliegt (ebd.). KI-Systeme mit diesem Zweck kommen meinen Recherchen zufolge in keinem anderen Europäischem Land zum Einsatz. Ich

habe keine Informationen darüber gefunden, ob in Ländern außerhalb der Europäischen Union KI-Systeme zu diesem Zweck eingesetzt werden.

Der Einsatz eines KI-Systems zum Aufdecken von Sozialbetrug hat Menschen aus niedrigen Einkommensklassen im Visier**.** Das vorige Beispiel, wie auch die Beispiele und Darlegungen in Kapitel 3.2.3 zu Überwachung zeigen, dass KI-Systeme neue Formen der Überwachung möglich machen. Wird KI, wie es bei dem KI-System „SyRi" der Fall ist, gegen den Betrug von Sozialgeld, der tendenziell einkommensschwache Schichten betrifft, eingesetzt, wirkt sich das auf die die materielle Ungleichheit in Staaten aus. Denn ein Algorithmus mit ähnlicher Funktionsweise könnte benutzt werden, um beispielsweise den Betrug von Steuergeldern aufzudecken. Hier wären nicht einkommensschwache Bevölkerungsgruppen, wie diejenigen die von Sozialbetrug profitieren im Visier der Algorithmen, sondern Menschen aus höheren Einkommensschichten. Solch ein KI-System, welches vom Staat im Einsatz oder in Entwicklung ist, habe ich im Rahmen meiner Recherche jedoch nicht finden können.

3.3.3 Ideologische Einbettung von KI in kapitalistische Prinzipien

Die momentane Entwicklung von KI-Systemen, wie auch andere Formen der Technikentwicklung funktionieren Lücking zufolge nach neoliberalen und kapitalistischen Prinzipien (vgl. Lücking 2020, S. 73). Auch Lengersdorf und Weber beschreiben Technikentwicklung nach 1989 als verwachsen mit der neoliberalen Wirtschaftspolitik (vgl. Lengersdorf und Weber 2019, S. 7). Horvath beschreibt, dass die Entwicklung und Anwendung von KI „großteils in global aktiven […] Unternehmen" (Horwath 2022, S. 72) stattfindet. Weber und Prietl benennen neben den kapitalistischen Prinzipien auch die Relevanz militärischer Interessen als auschlaggebend für den soziokulturellen Kontext der Entwicklung von KI. Sie beschreiben KI als „das Zentrum eines militärisch-industriellen Komplexes" (frei übersetzt nach Weber und Prietl 2021, S. 63). Warum KI kapitalistisch eingebettet ist, erklären sie damit, dass KI-Systeme mit großen Mengen an Daten funktionieren und die Entwicklung von KI-Systemen somit von diesen großen Unternehmen, welche die Daten besitzen, abhängt (ebd.). Die kapitalistischen Wertesysteme nach denen Unternehmen handeln, finden sich dementsprechend auch in KI wieder (ebd.). Die militärischen Interessen rühren den Autorinnen zufolge daher, dass die technologische Entwicklung von KI „stark vom Militär, Verteidigungs- und Geheimdienst-organisationen finanziert wird" (frei übersetzt nach Weber und Prietl 2021, S. 63).

Simon beschreibt, dass sich dieser Kontext und die kapitalistisch-militärischen Interessen in die Funktionsweise von KI-Systemen einschreiben (vgl. Simon 2020, S. 38–40). Sie erklärt, dass die Definitionsmacht von KI in den Händen der großen Firmenimperien liegt, da diese über die Technikgestaltung bestimmen (ebd.). Sie führt aus, dass wenn KI-Systemen immer mehr wichtige Entscheidungsverantwortungen übertragen werden und ihre Gestaltung dadurch langfristige Entwicklungen bestimmt, sie dadurch zu einem Machterhalt großen Firmenimperien sowie zur Konservierung bestehender Macht- und Hierarchiestrukturen beitragen (ebd.). Solche Argumente werden durch die nationalen und übernationalen Strategien für KI untermauert: Beispielsweise beschreibt die EU KI als einen Bereich von strategischer Bedeutung und identifiziert ihn als potenziellen Hauptmotor für die wirtschaftliche Entwicklung (vgl. AI Watch 2020, S. 6). Frey et al. beschreiben, dass die europäische Debatte um die „Industrie 4.0“ mit ihrem Fokus auf globale Wettbewerbsfähigkeit die Technikdebatte verengt: Gesellschaftliche Bedürfnisse, die mit der Zielsetzung der Wettbewerbsfähigkeit unvereinbar sind, hätten Frey et al. zufolge keinen Platz in der Debatte (vgl. Frey et al. 2021, S. 20). Nicht nur in den Argumenten von Expert*innen, auch in den praktischen Anwendungen von KI-Systemen lassen sich kapitalistisch motivierte Begründungen für ihren Einsatz erkennen: Motive der Kosteneinsparung und der Effizienz spielen hier die vordergründige Rolle, wie die Beispiele von KI-Systemen dieser Arbeit zeigen: Feathers führt beispielsweise den finanziellen Druck auf die öffentlichen Universitäten aufgrund jahrelanger staatlicher Sparmaßnahmen und die rückläufige Zahl der Studierenden, als entscheidenden Faktor für den Einsatz von „Navigate“ an den US-amerikanischen Universitäten auf (vgl. Feathers 2022, S.273). Ich erkenne darin eine ideologische Einbettung von KI in die kapitalistischen Prinzipien. Für die materielle Ungleichheit in Staaten bedeutet diese ideologische Einbettung das Weiterführen und Vertiefen kapitalistischer Praktiken. Für die Beantwortung der Forschungsfrage ist dieses Argument insofern relevant, als dass eine Vertiefung und das Weiterführen kapitalistischer Praktiken eine weiter andauernde Zunahme der sozioökonomischen Ungleichheit in Staaten, wie sie Haller et al. beschreiben, mit sich bringt (vgl. Haller et al. 2015, S.1).

3.3.4 Arbeitsverlust durch KI?

Die Anwendungsbeispiele der KI-Systeme, die ich im Rahmen dieser Arbeit beschreibe, veranschaulichen die Bandbreite der Anwendungsfelder von KI. Viele KI-Systeme unterstützen durch ihre Analysen und Bewertungen die Entscheidungen von Menschen. Sie können Daten systematisieren und aus den Informationen schneller und effizienter Wissen ableiten als Menschen. In Nahezu jedem Arbeitssektor gibt es Bereiche in denen Entscheidungen getroffen werden müssen. Es werden vermehrt Arbeitsplätze durch KI unterstützt oder ganz von ihnen besetzt. Ob und wie sich diese Entwicklung in der Zukunft fortsetzt ist unklar. Es gibt eine Vielzahl von Modellvorhersagen, die zeigen welche Arbeitsbereiche betroffen sein werden und welche nicht. Eine der bekanntesten und größten Studien zu dieser Thematik stammt aus dem Jahr 2013. In der Studie kommen Die Forscher zu dem Ergebnis, dass für die drei Tätigkeitsfelder Dienstleistung, Verkauf und Bürotätigkeit/administrative Unterstützung die größte Wahrscheinlichkeit besteht zukünftig von KI-Systemen übernommen zu werden (vgl. Frey und Osborne 2013, S. 40). Frey und Osborne leiten von den Ergebnisse ihrer vielzitierten Studie ab, dass die Entwicklung von KI in naher Zukunft „einen beträchtlichen Teil der Beschäftigung in einem breiten Spektrum von Berufen gefährden würden“ (frei übersetzt nach Frey und Osborne 2013, S.43). Eine andere große Studie kommt zu dem Ergebnis, dass der Einsatz von KI-Systemen sich in naher Zukunft auf Arbeitsplätze auswirken wird, da Arbeitsaufgaben, sowohl Routine- als auch Nicht-Routineaufgaben automatisiert werden (vgl. Poba-Nzaou et al. 2021, 60ff). Es handelt sich bei Studien dieser Art um Vorhersagen. Petropoulus kritisiert, dass für diese Prognosen müssen bestimmte „Annahmen getroffen werden, deren Gültigkeit nicht mit Sicherheit beurteilt werden können“ (Petropoulos 2018, S. 128). Petropoulus, wie auch andere Wissenschaftler*innen betonen, dass Vorhersagen über Arbeitsverluste durch KI auf Modellen basieren und von subjektiven Einschätzungen verfälscht sind (vgl. Petropoulos 2018, S. 128; Hunt et al. 2022, o.S.). Es fehlen Hunt zufolge Analysen dessen, was tatsächlich in denjenigen Unternehmen passiert, die KI-gestützte Technologien einführen (vgl. Hunt et al. 2022, o.S.). Somit ist es schwer Aussagen darüber zu tätigen, ob und wie sich der Einsatz von KI-Systemen in Zukunft auf die Arbeitsplätze von Menschen auswirkt (ebd.).

Dannecker et al. beschreiben, dass aktuelle Befunde aus der empirischen Forschung zeigen, dass technologische Innovations- und Modernisierungsschüben sektorale Ungleichheiten hervorrufen können (vgl. Dannecker et al. 2010, S.70). Das passiert,

indem der Modernisierungsschub zu einem Übergang von Beschäftigten von traditionellen zu modernen Wirtschaftsbereichen führt. Die sektorale Ungleichheit beschreibt den Zeitpunkt, in dem die Nachfrage an Arbeiter*innen in den modernen Wirtschaftsbereichen steigt und die Nachfrage in den traditionellen Wirtschaftsbereichen sinkt. Die veränderte Nachfrage führt zu sinkenden Löhnen und Arbeitsplatzverlusten in den traditionellen Wirtschaftsbereichen. Die sektorale Ungleichheit hat somit Auswirkungen auf die soziale Ungleichheit (vgl. Dannecker et al. 2010, S.70). Ob KI einen Modernisierungs- und Innovationsschub hervorruft, der zu einer sektoralen Ungleichheit und somit zu Arbeitsverlusten und sinkenden Löhnen führt, ist meinen Recherchen zufolge noch unklar. Modelle sagen zwar solche Entwicklungen vorher, stehen jedoch in der Kritik auf Annahmen zu basieren (vgl. Petropoulos 2018, S. 128). Trotzdem ist es sichtbar, dass KI das Potenzial eines solchen technologischen Innovations- und Modernisierungsschubs mit sich bringt: Der Bericht von Algorithm Watch, aber auch die Aussagen der EU und anderer Institutionen zeigen, dass nicht nur die Entwicklung von KI in den letzten Jahren einen großen Sprung gemacht hat, sondern auch die Anwendung von KI-Systemen gestiegen ist (Bundesregierung der Bundesrepublik Deutschland 2020; Europäische Kommission 2020; European Union Agency for Fundamental Rights 2020b; Algorithm Watch 2020b; UNICRI 2020; Interpol und UNICRI 2020).

Dannecker et al. beschreiben, dass eine sektorale Ungleichheit nicht gezwungenermaßen zu steigender materieller Ungleichheit führen muss: Sektorale Ungleichheit kann durch Politik und Zivilgesellschaft (auf nationaler und internationaler Ebene) verringert und verträglicher gemacht werden (vgl. Dannecker et al. 2010, S.70). Beispielsweise können Menschen aus dem traditionellen Sektor umgeschult werden, sodass sie den neuen Anforderungen des Arbeitsmarktes gerecht werden. Dannecker et al. zeigen zudem auf, dass die Handlungsspielräume im Vergleich zu früheren Globalisierungs- und Modernisierungsschüben in viel größerem Maße vorhanden sind (ebd.). Ob KI-Systeme in Zukunft die Arbeitskraft von Menschen ersetzen und es dadurch zu gesteigerter materieller Ungleichheit kommt, hängt also unter anderem von den Reaktionen der Politik und der Zivilgesellschaft ab. Hier kann es sich beispielsweise um rechtliche Einschränkungen der Entscheidungsspielräume von KI handeln, die vorgenommen werden, um die Anwendungsfelder von KI-Systemen zu regulieren.

3.4 Modifikationen der Ungleichheitsdimension Staatsbürgerschaft durch KI

In diesem Kapitel thematisiere ich die Auswirkungen von KI auf die Ungleichheitsdimension der Zugehörigkeit von Menschen durch ihre Staatsbürgerschaft. Ich fasse die Argumente, die für diese Ungleichheitsdimension relevant sind, zusammen und beschreibe ihre Zusammenhänge in den jeweiligen Kategorien. Auch für dieses Kapitel ist zu beachten, dass Zugehörigkeit – wie bereits in Kapitel 2. Methodik beschrieben – intersektional wirkt, d.h. verwoben ist mit anderen Kategorien sozialer Differenzierung wie Ethnizität und Gender, Alter, soziale Herkunft oder sexuelle Orientierung (vgl. Horwath 2022, S.75).

3.4.1 Datenbasiertheit von KI und Verzerrungen von Algorithmen durch Daten

Für dieses Kapitel konnte ich wenig wissenschaftliche Sekundärliteratur finden. Ich vermute, dass liegt darin begründet, dass die Ungleichheiten, welche sich in den algorithmischen Systemen von KI einschreiben, auf gesellschaftlichen Zuschreibungen basieren. Diese haben in Bezug auf die Ungleichheitsdimension Ethnizität, rassistische Diskriminierung zur Folge, welche sich in den Daten abbildet. In Bezug auf die Ungleichheitsdimension Gender, haben sie sexistische Diskriminierung zur Folge, welche sich ebenfalls in Daten abbildet. In Bezug auf die Ungleichheitsdimension Klasse können sie zwar Klassismus zur Folge haben, diese Form der Diskriminierung ist jedoch nicht analog zur Ungleichheitsdimension Klasse: Während es bei Klassismus um die Wahrnehmung und Bewertung von Klassenunterschieden geht, handelt es sich bei der Ungleichheitsdimension Klasse um die Erzeugung dieser Unterschiede an sich (vgl. Kemper und Kuleßa 2020, S. 11). Die vertikale Ungleichheitsdimension Klasse steht dadurch in einer anderen Wechselwirkung mit KI, wie die horizontalen Ungleichheitsdimensionen Ethnizität und Gender. Wie diese Wechselwirkung und die Modifikation der Ungleichheitsdimension Klasse funktioniert, beschreibe ich in den folgenden Absätzen.

Wie bereits in den Kapiteln 3.1.1.a und 3.3.1 beschrieben, existiert eine digitale Kluft. Diese hat Implikationen auf die materielle Ungleichheit zwischen Staaten. Menschen aus Ländern, deren pro-Kopf-Einkommen niedriger ist, sind durch Daten tendenziell weniger sichtbar als Menschen aus Ländern mit hohem pro-Kopf-Einkommen (vgl. Rubinov und Hagerty 2019, S. 2). Das Weißpapier des Weltwirtschaftsforums aus dem Jahr 2018 zeigt auf, dass ein durchschnittlicher US-Haushalt alle sechs Sekunden einen Datenpunkt erzeugt (vgl. World Economic Forum 2020, S. 8). In Mosambik, wo

etwa 90 % der Menschen keinen Internetzugang haben, erzeugt der durchschnittliche Haushalt null digitale Datenpunkte (ebd.). 30% der Weltbevölkerung haben keinen Zugang zu 3G/4G-Breitband-Internet, vor allem im ländlichen Asien und Afrika südlich der Sahara (ebd.). Ungleiche Repräsentativität in Daten ist einer der Gründe für inakkurate Ergebnisse und diskriminierende Bewertungen von KI-Systemen, wie Kapitel 3.3.1.a gezeigt hat. Zudem fehlen Ländern deren Bewohner*innen weniger akkurat durch Daten abgebildet sind, eine auf ihren Kontext zutreffende Grundlage für die Entwicklung und den Einsatz datenbasierter KI-Systeme.

Einschätzungen internationaler Institutionen wie der OECD, der UNESCO, dem Weltwirtschaftsforum, der ISO und der Europäischen Kommission bewerten KI als ein Instrument für Wachstum und technologische Entwicklung (vgl. AI Watch 2020, S. 7). Beispielsweise erwartet sich die Finanzbranche durch den Einsatz von KI-Systemen „Effizienz- und Effektivitätsgewinne, Rentabilitätssteigerungen, Prozessautomatisierungen, Kostenreduktionen, Qualitätssteigerungen und Angebotserweiterungen“ (Bauer et al. 2021, S. 16). Wenn KI-Systeme keine akkurate Datenbasis haben, funktionieren sie, wie in Kapitel 3.3.1.a gezeigt, weniger gut. Die digitale Kluft bedeutet also für Länder, deren Gesellschaften weniger gut in Daten abgebildet sind, einen Wettbewerbsnachteil. Es kann sein, dass ein solcher Nachteil zu steigender internationaler Ungleichheit führt. Weber und Prietl argumentieren, dass in der Entwicklung von KI-Systeme die Interessen, Perspektiven und Meinungen der Menschen ausgeschlossen sind, die nicht oder unpräzise durch Daten abgebildet sind und dass es dadurch zu einer Reproduktion historisch etablierter machtrelationaler Strukturen von Dominanz, Marginalisierung und Subordination kommt (vgl. Weber und Prietl 2021, S. 64). Im nächsten Kapitel erweitere ich dieses Argument um andere Thesen und Konzepte und führe damit die Implikationen von KI auf die internationale Ungleichheit weiter aus.

3.4.2 Internationaler Wettbewerb um KI

Um die Modifikationen internationaler Ungleichheit durch KI zu verstehen, ist es wichtig anzuerkennen, wie anders KI in verschiedenen Ländern eingesetzt wird. Nicht nur die Quantität und Qualität des Einsatzes von KI unterscheidet sich von Staat zu Staat, sondern auch die Auffassung über das Zusammenspiel von Technologie und Recht (Algorithm Watch 2020a, S. 3). Unterschiedliche Regierungsformen und -werte spiegeln sich in den KI-Systemen wider. Das zeigt sich beispielsweise darin, wie ein

Staat Datenschutz und Grundrechte durchsetzt. Mittels folgendem Beispiel will ich die unterschiedliche Weise der Technikanwendung und der Spiegelung der Staatsform in derjenigen KI Anwendung zeigen: Im Zusammenhang mit COVID-19 kann zum selben Zweck unterschiedliche Technologie[15] eingesetzt werden. Ausgehend von den Prioritäten, die in der Entwicklung der Software gesetzt werden, finden unterschiedliche Anwendungsformen in verschiedenen Ländern Einsatz. Um gegen die Virusübertragung vorzugehen, setzen Länder Smartphones und digitale Technologie ein, um Menschen zu warnen, wenn sie in Kontakt mit einer infizierten Person hatten. Über Bluetooth können Daten der getrackten Personen freiwillig und anonym entweder mit einem zentralen Server oder nur mittels den Smartphones potenziell infizierter Personen geteilt werden (Algorithm Watch 2020a, S. 5). Diese Form des Technologieeinsatzes ist anonym und es werden keine sensiblen Daten von der Software gesammelt. Eine andere Technologie mit demselben Zweck kann auch im Rahmen einer sehr viel stärker in die Persönlichkeitsrechte eingreifenden Lösung eingesetzt werden: Beispielsweise bei Systemen die mit GPS den Standort des Smartphones kontinuierlich an Behörden weitergeben (Algorithm Watch 2020a, S. 5). Manche Staaten priorisierten den Datenschutz der Zivilgesellschaft und setzen datensensible Varianten, wie die der anonymisierten Datenübertragung mittels Bluetooth ein. Andere Staaten tracken die raumbezogenen Daten ihrer Bürger*innen und Bürger mit weniger Sensibilität.

Dieser unterschiedliche Einsatz von KI-Systemen hat zwar keine direkte Auswirkung auf die materielle Ungleichheit zwischen Staaten, jedoch schränken strenge Datenschutzregelungen die Entwicklung von datenbasierten KI-Systemen ein. Je mehr Daten ein KI-System zur Verfügung hat, desto besser und akkurater sind die Ergebnisse des Systems (vgl. Blattner und Nelson 2021, S.1). Die Einschränkung von KI zum Schutz der Privatsphäre, für Antidiskriminierung und für die informationelle Selbstbestimmung steht damit der uneingeschränkten Entwicklung von KI entgegen. Im Weißpapier der Europäischen Kommission zu KI wird hinsichtlich dieses Spannungsfeldes folgendes festgehalten:

[15] Bei Software zur epidemiologischen Überwachung handelt es sich nicht ausschließlich um KI-Systeme. KI ergänzt aber oft andere Technologien und analysiert die überwachten Daten, um die Ausbreitung einer Viruserkrankung nachzuverfolgen.

„Angesichts der rasanten Entwicklung der KI muss der Rechtsrahmen Raum für weitere Entwicklungen lassen. Jegliche Änderungen sollten sich auf eindeutig identifizierte Probleme beschränken, für die es machbare Lösungen gibt.“ (frei übersetzt nach European Commission 2020, S. 9).

Ob Wachstum und Datenschutz gleichermaßen realisierbar sind, lässt sich noch nicht sagen. Eine Studie, die Datenschutzvorschriften und den Einsatz von KI im Bankenwesen analysiert, besagt, dass Datenschutzvorschriften den Einsatz von KI im Bankensektor möglicherweise erschweren können (vgl. Bauer et al. 2021, S. 16). In einer umfassenden Analyse von mehr als 35 000 Schlüsselakteuren der globalen KI-Landschaft werden die USA, Europa und China als führende Länder des internationalen Wettbewerbs um die Vorreiterstellung für KI identifiziert (vgl. Europäische Kommission 2018a, S. 35). Im Rahmen des Programm "Digitales Europa" werden mehr als 4 Mrd. EUR zur Förderung von Hochleistungsrechnern und Quantencomputern, einschließlich Edge Computing, KI, Daten und Cloud Infrastruktur investiert (vgl. European Commission 2020, S. 8). Nicht alle Länder können sich gleichermaßen Investitionen in KI leisten. Wenn KI in dem vorhergesagten Ausmaß zum wirtschaftlichen Wachstum beiträgt (vgl. AI Watch 2020, S. 7), werden bestehende materielle Ungleichheiten zwischen Ländern fortgeführt und vertieft (vgl. Schelenz 2021, S. 83).

Neben den wirtschaftlichen Vorteilen, die der Einsatz von KI, den Vorhersagen nach mit sich bringt, ist auch eine militärische Überlegenheit durch den Einsatz von KI-Systemen gegeben. KI findet zu militärischen Zwecken Einsatz in der Krisenfrüherkennung, im Cyber- und Informationsraum, im Militärischen Führungsprozess, beispielsweise bei Strategischer Entscheidungsfindung oder der Organisation von Streitkräften und beim Waffeneinsatz, beispielsweise für autonome Drohnen (vgl. Masuhr 2019S. 2ff). Die Investitionen, die es braucht, um KI-Systeme für solche Zwecke zu entwickeln, haben nicht alle Länder zur Verfügung. Aktuell sind es die USA und China, die in der Entwicklung und dem Einsatz von KI-Systemen für militärische Zwecke ein Rennen um die Vormachtstellung haben, aber auch Russland und andere Länder kündigen den militärischen Einsatz von KI an (vgl. Johnson 2021, S. 370). Eine Leistungssteigerung durch militärisch-technologische Innovationen, wie sie durch KI gegeben ist, kann zu einem Rüstungswettlauf führen, welcher Krause zufolge eine „potentielle Destabilisierung des sicherheitspolitischen Gleichgewichts“

(Krause 2021, S. 28) zur Folge haben kann. Die militärische Überlegenheit von Staaten, die sich die Entwicklung von KI-Systemen leisten können, produziert ein Machtgefälle zwischen Staaten mit militärischer KI und ohne militärische KI. Im Falle einer militärischen Auseinandersetzung sind Staaten ohne KI-Systeme unterlegen und bestehende materielle Ungleichheiten zwischen Ländern können sich durch dieses Machtgefälle vertiefen. Im speziellen Fall von autonomen Drohnen argumentiert Krause jedoch entgegen einer Vertiefung dieser Machtgefälle: Denn Angriffe mit autonomen Drohnen können mit „relativ einfachen und billigen Mitteln auch gegen vergleichsweise moderne und teure Luftverteidigungssysteme erfolgreich sein“ (Krause 2021, S. 9). Das eröffnet Ländern mit geringen finanziellen Mitteln neue militärische Möglichkeiten. Weitere Schlussfolgerungen hinsichtlich der Modifikation von internationaler Ungleichheit durch den Einsatz von KI-Systemen im militärischen Bereich sind im Rahmen dieser Arbeit nicht möglich. Wie bereits in Kapitel 2. Methodik beschrieben, ist die militärische Nutzung von KI im Gegensatz zu den anderen Anwendungsbereichen eine, die in der wissenschaftlichen Literatur wenig berücksichtigt wird. Das erschwert weitere Einblicke.

3.4.3 KI als Projekt reicher Länder

Schelenz hält fest, dass die Wechselwirkungen zwischen KI und Gesellschaft, wie auch die Einbettung von KI-Systemen global gesehen zu einer Verfestigung bestehender Machtrelationen und Hierarchiestrukturen zugunsten westlicher Akteure führen (vgl. Schelenz 2021, S. 83), was auch unter dem Begriff des digitalen Kolonialismus diskutiert wird (Tuzcu 2021; Coleman 2018; Schelenz 2021). Weber und Prietl zufolge besteht eine geographische Situiertheit der Entwicklung von KI in Ländern mit hohem durchschnittlichem Einkommensniveau, da sich vor allem in diesen Ländern einflussreiche militärische und unternehmerische Akteur*innen befinden. Folgendes Beispiel zeigt wie sich eine solche Situiertheit in einem KI-System realisiert.

KI-System „HealthMap“

„HealthMap“[16] ist ein KI-System, das in Echtzeit die Ausbreitung von infektiösen Krankheiten anzeigt. Es analysiert dafür unterschiedlichste Quellen, wie beispielsweise lokale Nachrichten oder Daten aus soziale Medien (vgl. HealthMap

[16] https://www.healthmap.org/formobile/

2022). Lokale Medienberichte aus Ländern mit niedrigem Einkommensniveau konnten bisher mehrheitlich nicht analysiert werden, weil die Nachrichtendienste nicht darauf programmiert waren, mehr als 9 Sprachen zu verarbeiten (vgl. Brownstein et al. 2008, o.S.). Die analysierten Sprachen stammten aus Ländern mit hohem durchschnittlichem Einkommensniveau (ebd.). Dadurch waren die Warnmeldungen von „HealthMap“ für Länder deren Sprachen nicht verarbeitet werden konnten weniger akkurat oder teilweise auf lokaler Ebene nicht vorhanden (ebd.).

Weber und Prietl beschreiben zudem die (post-)koloniale Struktur, die sich in der technischen Entwicklung von KI widerspiegele. Sie erklären, dass einerseits die Konzeptualisierung, das Design und die Entwicklung von KI durch hochqualifizierte, junge, sozio-ökonomisch privilegierte Männer passiert, während die sogenannte „content-moderation“ und das einfache Datenmanagement von niedrig-qualifizierten, gering entlohnten, meist online und anonym arbeitenden Menschen aus Ländern mit niedrigem Einkommensniveau vorgenommen wird (vgl. Weber und Prietl 2021, S. 63–64). Diese strukturelle Ungleichheit in KI spiegele sich den Autorinnen zufolge in der Zusammensetzung der für die Entwicklung von KI verantwortlichen Arbeitskräfte, sowie in den Perspektiven und Interessen wider, die bei der Entwicklung von KI-Technologien berücksichtigt werden (Weber und Prietl 2021, S. 63).

Ein weiteres für die materielle Ungleichheit zwischen Staaten relevantes Argument, welches Kwet in Beziehung zu postkolonialen Strukturen setzt, ist, dass KI einem offenen Wissens- und Informationsaustauschs entgegensteht: KI-Systeme können große Mengen von Daten auswerten, die Ergebnisse sind aber oft exklusiv und nicht für alle Menschen zugänglich, wie folgende Argumente zeigen (vgl. Kwet 2019, S.3ff). Mehreren Autorinnen zufolge liegen sowohl die für KI-Systeme essenziellen Datensätze, als auch die informationstechnischen Infrastrukturen für ihre Erhebung, Verarbeitung und Analyse entweder in den Händen von Internet- und Datenkonzernen, oder aber in denen staatlicher Nachrichtendienste und Verteidigungseinrichtungen (vgl. Prietl 2019, S.314; Zuboff 2019, S.20ff). Diese Art der Wissensproduktion vertiefe Tuzcu zufolge die geopolitischen Hierarchien zwischen sozioökonomisch bevorteilten – und sozioökonomisch benachteiligten Ländern, da vor allem große multinationale Konzerne die Datenmengen besitzen, die es benötigt um Wissen mit KI-Systemen zu produzieren (vgl. Tuzcu 2021, 514ff).

4. Zusammenfassung der Ergebnisse

<Wie modifiziert KI die Ungleichheitsdimensionen Ethnizität, Gender, Klasse und Staatszugehörigkeit?>

Für die Zusammenfassung strukturiere ich die Ergebnisse des Hauptteils nicht entlang der Ungleichheitsdimensionen wie im Hauptteil, sondern entlang der einzelnen Kategorien. Somit können die Unterschiede und Gemeinsamkeiten der Modifikationen der einzelnen Ungleichheitsdimensionen durch KI hervortreten. Dieses Kapitel gibt einen abschließenden Überblick der Antwort auf die Forschungsfrage und fasst die Argumente und Beispiele des Hauptteils zusammen.

Die erste Kategorie thematisiert die Modifikationen von Ethnizität, Gender, Klasse und Staatszugehörigkeit durch die Datenbasiertheit von KI und findet sich als einzige Kategorie in jedem der vier Kapitel wieder. Es geht in dieser Kategorie um eine Funktionsweise, die der Technologie inhärent ist: Die Datenbasiertheit ist eine technische Bedingtheit von KI-Systemen und bringt ungleichheitswirksame Modifikation auf zwei Arten und Weisen mit sich.

Einerseits die Modifikation der jeweiligen Ungleichheitsdimension aufgrund von fehlender Repräsentativität bestimmter sozialer Gruppen in Daten. Diese fehlende Repräsentativität lässt sich in diesem Fall auch als Unsichtbarkeit bestimmter sozialer Gruppen in Daten beschreiben. Der Begriff der digitalen Kluft benennt diesen Zusammenhang. Somit ist die digitale Kluft ein wichtiges Schlüsselwort in dieser Kategorie, was die Nennung im Titel der Kategorie begründet. Durch die (Un-)Sichtbarkeit mancher sozialer Gruppen in Daten werden auch deren Bedürfnisse unsichtbar werden. Weil (Un-)Sichtbarkeit in Daten in Relation zu den Achsen der Ungleichheit verläuft, hat sie Auswirkungen auf die untersuchten Ungleichheitsdimensionen. Diese Modifikation beschreibe ich in der ersten Unterkategorie „Verzerrung durch Unsichtbarkeit mancher Gruppen in Daten“.

Andererseits die Modifikation der jeweiligen Ungleichheitsdimension in Bezug auf das Einschreiben bestehender genderspezifischer oder rassistischer Diskriminierungen in Algorithmen durch Trainingsdatensätze und Daten. Dieser Sachzusammenhang bezieht sich jedoch im Gegensatz zu der vorigen Kategorie auf eine andere Funktionsweise von KI. Bei der ersten Kategorie wird durch das Fehlen von Daten eine Verzerrung der Algorithmen verursacht, die sich wiederum auf die soziale Ungleichheit auswirkt, während bei der zweiten Kategorie eine in Daten abgebildete konstruierte

soziale Ungleichheit in das Regelsystem der Algorithmen der KI übertragen wird. Diese Modifikation beschreibe ich in der zweiten Unterkategorie „Einschreiben bestehender Diskriminierung in Algorithmen durch Daten“.

In 3.1.1.a, führe ich die erste Unterkategorie, also die „Verzerrung durch Unsichtbarkeit mancher Gruppen in Daten“ in Bezug auf Gender auf. Hier sind vor Allem die zwei Begriffe digitale Kluft und digitale Spaltung relevant, weil diese beschreiben, wie es zu der Unsichtbarkeit bestimmter Gruppen in Daten kommt und wie sich die Unsichtbarkeit in KI-Systemen ausprägt. Ich beschreibe hier nicht nur die geschlechtsspezifischen Gefälle in der Repräsentativität von Daten, sondern führe auch die Ausprägungen der Unsichtbarkeit anderer sozialer Gruppen auf. Frauen besitzen beispielsweise im Durchschnitt 10% seltener ein Mobiltelefon und benutzen zu 23% weniger mobiles Internet (vgl. GSMA 2019, S. 4). Zusammengefasst zeigen die Beispiele und Erklärungen in diesem Unterkapitel auf, wie „‚fehlende Daten‘ […] gesellschaftliche Machtverhältnisse wieder[spiegeln]“ (Horwath 2022, S.77). Wenn Menschen in Daten nicht repräsentiert sind, sind sie für KI-Systeme nicht erkennbar. Das macht auch die Ergebnisse der KI-Systeme für diese Menschen weniger akkurat.

Für die in Kapitel 3.2.1.a beschriebenen Modifikationen der Ungleichheitsdimension Ethnizität ist vor allem das Anwendungsfeld der Gesichtserkennung mit KI relevant. BIPoC werden von KI-Systemen weniger gut erkannt als andere Gruppen (vgl. Buolamwini und Gebru 2018, S. 11). Grund hierfür ist die fehlende Repräsentativität von BIPoC in den Trainingsdatensätzen, die bei der Entwicklung von KI-Systemen eingesetzt werden (ebd.). Das hat unterschiedliche Auswirkungen, wie beispielsweise folgende: In den USA hat ein KI-System einen gesuchten Kriminellen falsch identifiziert und es kam zu einer Verhaftung dieser vermeintlich kriminellen afro-amerikanischen Person (vgl. Hill 2022, S. 139). Weitere ungleichheitsmodifizierende Auswirkungen hat die Anwendung von KI im Gesundheitsbereich, beispielsweise bei der Erkennung von Hautkrebs oder der Bewertung von Personen für Lebensversicherungen. Bei dem KI-System für die Erkennung von Hautkrebs wurde in einer Studie nachgewiesen, dass sie für BIPoC weniger gut funktioniert (vgl. Zou und Schiebinger 2018, S. 324). Bei dem KI-System für die Kategorisierung der Konditionalitäten für eine Lebensversicherung, schlussfolgere ich, dass es zu einer ähnlichen Diskriminierung kommt: Auch dieses KI-System basiert auf Gesichtserkennung (vgl. Wennker 2020, S. 114).

In Bezug auf die in Kapitel 3.3.1 beschriebenen Auswirkungen auf Klasse zeigen KI-Systeme für die Kreditwürdigkeitseinschätzung ungleichheitsmodifizierende Auswirkungen: Bei schlechter Datenlage kommt es zu fehlerhaften Ergebnissen und Menschen werden Kredite zu falschen Konditionen vermittelt (vgl. Blattner und Nelson 2021, S.29). Ob es eher zu fehlerhaften Einstufungen der Kreditunwürdigkeit oder der Kreditwürdigkeit kommt, konnte ich im Rahmen meiner Recherche nicht herausfinden. Meine Schlussfolgerung ist jedoch, dass in beiden Fällen die inakkuraten Ergebnisse bei schlechter Datenlage eine ungleichheitsverstärkende Wirkung haben. Für Menschen, die durch eine fehlerhafte Kreditwürdigkeitseinschätzung einen Kredit bekommen und ihn nicht zurückzahlen können, kann sich diese Kreditvergabe zur Schuldenfalle entwickeln. Andersherum hat es auch für Personen, die einen Kredit nicht bekommen, obwohl sie ihn zurückzahlen könnten, negative Konsequenzen. Denn sie können die geplante Investition oder Zahlung mit dem Kredit nicht tätigen.

In Kapitel 3.1.1 beschreibe ich die ungleichheitsmodifizierende Wirkung der Datenbasiertheit von KI auf internationale Ungleichheit. KI wird als ein Instrument für Wachstum und technologische Entwicklung aufgefasst (vgl. AI Watch 2020, S. 7). Länder können es sich in unterschiedlichem Maß leisten in diese Technologie zu investieren. Wenn Länder den Technologie-Vorsprung durch KI für sich nutzen können, verstärkt das die internationale Ungleichheit. Ländern deren Bewohner*innen weniger akkurat durch Daten abgebildet sind fehlt eine auf ihren Kontext zutreffende Datengrundlage für die Entwicklung und den Einsatz von KI-Systemen. Studien zeigen, dass beispielsweise in den USA viele Daten gesammelt werden können, was dem Land eine gute Grundlage in der Entwicklung von KI verschafft (vgl. World Economic Forum 2020, S. 8). Die Auswirkungen der Datenbasiertheit von KI-Systemen in Bezug auf die (Un-)Sichtbarkeit von Menschen in Daten sind für die drei Ungleichheitsdimensionen Gender, Ethnizität und Klasse vom Prinzip ähnlich, jedoch sind ihre Ausprägungen, wie die Beispiele zeigen, sehr unterschiedlich. Im Fall der internationalen Ungleichheit beschränkt sich der modifizierende Effekt der Datenbasiertheit von KI im Gegensatz zu den anderen untersuchten Ungleichheitsdimensionen auf den entwicklungstechnischen Vorsprung, den Länder mit vielen Daten haben.

Weitere Modifikationen, die mit der Datenbasiertheit von KI-Systemen einhergehen, fasse ich – wie bereits zu anfangs des Kapitels beschrieben – separat in einem zweiten

Unterkapitel zusammen. In diesem Unterkapitel thematisiere ich das Einschreiben bestehender Formen der Diskriminierung in Algorithmen durch Daten: Durch Daten lernen KI-Systeme über die gesellschaftliche Realität. In den Daten zeichnen sich Diskriminierung und Marginalisierung ab. Diese sind jedoch für die KI-Systeme nicht als solche gekennzeichnet, sondern werden als eine verzerrte Realität Teil der Logik nach derer sie funktionieren.

Der in diesem Argument beschriebene Wirkmechanismus modifiziert die horizontalen Ungleichheitsdimensionen Gender und Ethnizität anders als die Ungleichheitsdimension Klasse und die internationale Ungleichheit, weshalb ich sie nur für die Ungleichheitsdimensionen Gender und Ethnizität als gesonderte Kategorie aufführe. Genderspezifische und rassistische Diskriminierung bilden sich in Daten ab. Beide Formen der Diskriminierung basieren auf gesellschaftlichen Zuschreibungen und der Konstruktion von Stereotypen. Das zeige ich in den zwei Unterkapiteln 3.1.1.b „Einschreiben bestehender genderspezifischer Diskriminierung in Algorithmen durch Daten“ und 3.2.1.b „Einschreiben bestehender rassistischer Diskriminierung in Algorithmen durch Daten“ anhand von Beispielen auf:

In Kapitel 3.1.1.b schildere ich die Funktionsweise eines KI-Systems, welches in Österreich vom Arbeitsmarktservice für die Unterstützung der Reintegration am Arbeitsmarkt eingesetzt wurde. Der Algorithmus des KI-Systems des AMS lernte Betreuungspflicht als negativ gewichtetes Kriterium für die Vermittlungsfähigkeit einer arbeitssuchenden Person einzustufen (vgl. Guijarro-Santos 2020, S. 53–54). Es kam dadurch zu diskriminierenden Ergebnissen bei der Bewertung der Vermittlungsfähigkeit von Frauen (ebd.). Mithilfe weiterer Beispiele aus dem Bereich der KI für Arbeitsvermittlung, zeige ich auf, wie Frauen durch die Zuschreibung von Gruppeneigenschaften algorithmisch diskriminiert werden (vgl. Martin 2022, S. 291). Zudem veranschauliche ich, dass es auch bei Ausschluss der Merkmale Gender und Geschlecht zu diskriminierenden Bewertungen für Frauen kommen kann: KI-Systeme identifizieren Stellvertretermerkmale. Diese bringen durch korrelierende Datenpunkte mit gendersensiblen Merkmalen ähnlich diskriminierende Ergebnisse hervor, wie sie durch Korrelationen mit den ausgeschlossenen Merkmalen zustande kommen.

In Kapitel 3.2.1.b „Einschreiben bestehender rassistischer Diskriminierung in Algorithmen durch Daten“ beschreibe ich die Modifikation der Ungleichheitsdimension Ethnizität durch KI. Anhand von drei KI-Systemen aus dem Gesundheitswesen, dem

Bildungssektor und dem Justizbereich veranschauliche ich die Wirkmechanismen algorithmischer Diskriminierung. Jedes der drei KI-Systeme nimmt Einschätzungen vor und unterstützt mit den Einschätzungen die Entscheidungen von Menschen. Das KI-System „OptumIQ“ spricht Empfehlungen in Krankhäusern aus und analysiert auf Basis von Krankenhausakten den Gesundheitszustand von Patient*innen. Eine Studie über das KI-System zeigt, dass BIPoC-Patient*innen 50% niedriger für den Bedarf einer komplizierten medizinischen Behandlungen eingestuft werden, wie *weiße* Patient*innen mit der gleichen Erkrankung (vgl. Johnson 2022a, S. 10). Das KI-System „Navigate“ bewertet die Studienabbruchwahrscheinlichkeit an US-amerikanischen Universitäten. Eine Studie über das KI-System zeigt, dass Studierende je nach ethnischer Zugehörigkeit sehr unterschiedlich bewertet werden. Das KI-System mancher Universitäten stuft die Studienabbruchwahrscheinlichkeit bei BIPoC bis zu viermal so hoch ein wie für *weiße* Studierende (vgl. Feathers 2022, S.269). Das dritte Beispiel veranschaulicht algorithmische Diskriminierung durch das KI-System „COMPAS“. Dieses wird im Bereich der Strafjustiz eingesetzt. Es bewertet die Wahrscheinlichkeit eines Rückfalls von straffälligen Menschen. Eine Studie zu dem KI-System zeigt, dass sich überproportional mehr BIPoC unter denjenigen Einschätzungen befinden, die das KI-System als risikoreich für einen Rückfall einstuft, die jedoch nicht wegen erneuter Straftaten verhaftet wurden (vgl. Angwin und Larson 2022, S. 265). In jedem der drei KI-Systeme sind die Trainingsdatensätze der Grund für die diskriminierenden Ergebnisse der KI-Systeme. Datensammlungen, wie beispielsweise die Studienhistorie unterschiedlicher Universitäten, justizielle Strafakten oder Krankenakten aus Krankenhäusern bilden Entscheidungen von Menschen ab, die Rassismen und andere Formen von Diskriminierung reproduzieren. Dementsprechend bilden sie strukturelle Exklusion und Marginalisierung ab. Eine KI, die daraus ihr algorithmisches Regelsystem ableitet, wird diese Diskriminierung automatisieren (vgl. Simon 2020, S. 40). Viele Expert*innen für KI warnen vor dem Ein- und Festschreiben problematischer Macht- und Hierarchiestrukturen sowie Mustern der Marginalisierung und der Ungleichheit in die Technik (Lücking 2020; Weber und Prietl 2021; Simon 2020; Nassehi 2019; Algorithm Watch 2020b).

Auch bei der nachfolgenden Kategorie handelt es sich um eine der Technologie inhärenten Funktionsweise und auch sie zeigt insbesondere für die horizontalen Ungleichheitsdimensionen Gender und Ethnizität eine modifizierende Wirkung. Ich nenne diese Kategorie „Algorithmische Kodierung von Realität basiert auf

Kategorisierung und Normierung – Norm ist der *weiße* Mann“. Die Kategorie steht in Verbindung zu der vorig beschriebenen, jedoch ist der Ausgangspunkt dieser Kategorie ein anderer: Argumente dieser Kategorie gehen von einer algorithmischen Diskriminierung aufgrund ihrer Funktionsweise der Definition einer Norm aus. Während diese Kategorie also die Normierung und Kategorisierung als eine Funktionsweise von KI-Systemen fokussiert, ist es in der vorigen Kategorie anders: Hier steht die Übertragung bestehender Formen der Diskriminierung durch die Datenbasiertheit der Systeme im Zentrum der Argumentationsweise. Prietl und Weber argumentieren, dass algorithmische Verarbeitung von Daten eine „Normierung menschlichen Verhaltens“ (frei übersetzt nach Weber und Prietl 2021, S. 62) bedingt. Sie fragen, welches Verhalten als Norm identifiziert und durchgesetzt wird und welches menschliche Verhalten von dieser Konzeptualisierung ausgeschlossen wird und wie diese beiden Elemente mit „genderspezifischen, rassistischen und ableistischen Strukturen und Symbolsystemen verwoben sind“ (frei übersetzt nach Weber und Prietl 2021, S. 62). In einer androzentristisch geprägten Gesellschaft, deren Normen sich – wie oben beschrieben – in Daten und Algorithmen einschreiben, liegt es nahe, die Frage nach der Identifikation einer Norm mit einer Fortschreibung der bestehenden Machtverhältnisse zu beantworten. Die Norm des *weißen* Mannes spiegelt sich nicht nur in KI-Systemen wider, sondern auch in anderer Technik, wie beispielsweise bei einer Version eines elektronischen Seifenspenders in öffentlichen Toiletten: Dieser funktionierte, wie die viral gegangenen Videos aus dem Jahr 2019 zeigen, nur für Menschen mit heller Hautfarbe (vgl. ECIAIR 2019, S. 297). Donna Haraway hat bereits vor der Jahrtausendwende für technowissenschaftliche Artefakte im Allgemeinen argumentiert, dass diese nicht neutral sind, sondern vielmehr bestimmte Weltanschauungen, Wahrnehmungsstile und die Reproduktion bestehender sozialer Ungleichheiten begünstigen (Haraway 1988, S. 587). Auch Horvath beschreibt, dass die mit der „Klassifikation verbundenen Risiken der Hierarchisierung und Unterdrückung […] für bereits marginalisierte Gruppen wesentlich höher, als für sozial etablierte und privilegierte Gruppen“ (Horwath 2022, S.78) sind.

Die nachfolgende Kategorie „Ergebnisse von KI sind kein situiertes Wissen“ umfasst Argumente und Konzepte, welche die (nicht-)Situiertheit von Wissen im Prozess der algorithmischen Verarbeitung thematisieren. Die Kategorie ist für die Ungleichheitsdimension Gender als auch für die anderen Ungleichheitsdimensionen relevant und bezieht sich auf den Begriff des situierten Wissens aus der feministischen

Theorie. Trotz der Relevanz für alle Ungleichheitsdimensionen, ist diese Kategorie nur in dem Kapitel „Modifikationen der Ungleichheitsdimension ‚Gender' durch KI" aufgeführt, da der Ursprung der Argumente und Konzepte, wie auch die Verortung des Diskurses über „situiertes Wissen" dem feministischen Diskurs zugeordnet ist (vgl. Haraway 1988, S. 587; Ziai et al. 2018, S. 11–12). Bei „nicht-situiertem Wissen" ist die Perspektive aus der gesprochen und bewertet wird nicht definiert. Es kann dabei zu einer Verschleierung der Wertehaltung kommen. Wenn KI-Systeme zur Anwendung kommen, bringen sie scheinbar (wert-)neutrale und objektive Ergebnisse hervor. Diese Wertneutralität und Objektivität treffen nicht zu, denn KI-Systeme werden – wie bereits mehrfach gezeigt – von Menschen entwickelt und eingesetzt. Die Logik, nach der ihre Regelsysteme funktionieren sind zwar meist undurchsichtig, jedoch sind die Anforderungen der Entwicklung eines Systems vorgegeben. Zudem funktionieren sie datenbasiert, was bedeutet, dass die in den Daten festgehaltene gesellschaftliche Realität abgebildet wird und zum inhärenten Bestandteil der Logik der KI-Systeme wird.

Die Kategorie der „fehlenden Gerichtsbarkeit und Regulierung – KI als Blackbox" beschreibe ich für die zwei Ungleichheitsdimensionen Gender und Ethnizität. Meine Recherche hat in Bezug auf die Modifikation der Ungleichheitsdimension Klasse und internationaler Ungleichheit zu keinem Ergebnis geführt. Ich nehme dennoch an, dass fehlende Gerichtsbarkeit und Regulierung von KI-Systemen Auswirkungen auf diese beiden Ungleichheitsdimensionen haben. Eine prekäre wirtschaftliche Situation erschwert den Zugang zu Gerichtsbarkeit: Für eine Person ohne finanzielle Mittel ist es schwerer ihre Rechte durchzusetzen als für eine Person mit finanziellen Mitteln. Wenn zu dieser Hürde noch eine weitere dazukommt, und zwar die fehlende Gerichtsbarkeit von KI-Systemen, hat das einen ungleichheitsverstärkenden Effekt. Auch für Länder, in denen die Gerichtsbarkeit weniger stark durchgesetzt wird, wirkt sich die fehlende Gerichtsbarkeit zusätzlich benachteiligend aus. Jedoch stehen diese Modifikationen eher indirekt in Verbindung mit der Gerichtsbarkeit von KI-Systemen. Deshalb habe ich weder für die internationale Ungleichheit noch für die Ungleichheitsdimension Klasse, die Kategorie „fehlende Gerichtsbarkeit und Regulierung - KI-System als Black-Box" aufgeführt. Die Frage, ob KI-Algorithmen Intransparenz und fehlende Nachvollziehbarkeit als unveränderliche Merkmale ihrer technischen Funktionsweise in sich tragen, oder ob die beiden Merkmale viel mehr

Ergebnis der Art und Weise sind, wie KI-Systeme und Software entwickelt werden, konnte ich im Rahmen meiner Recherche nicht beantworten.

In Kapitel 3.1.4 fasse ich Argumente der Kategorie in Bezug auf die Modifikation der Ungleichheitsdimension Gender zusammen. Weil es in den wenigsten Ländern Gesetze gibt, die spezifisch vor algorithmischer Diskriminierung schützen, lege ich den Fokus auf den Rechtsrahmen der EU. In Artikel 22 der DSGVO festgelegt, dass Entscheidungen nicht ausschließlich automatisiert getroffen werden dürfen (vgl. Bauer et al. 2021, S. 16). Zudem müssen von KI-Systemen getroffene Entscheidungen immer von Menschen verantwortet werden (ebd.). In manchen Ländern gibt es jedoch Versuche, einen gesetzlichen Rahmen aufzubauen oder auszubauen und auch die EU arbeitet aktuell an einem neuen Gesetzentwurf für KI. Zusammenfassend lässt sich aus der Recherche ableiten, dass eine algorithmischen Diskriminierung selten, bis gar nicht rechtlich anfechtbar ist. Eng in Verbindung mit der fehlenden Regulierung und Gerichtsbarkeit steht die Intransparenz von KI-Systemen: Ein Gericht kann keine Entscheidung über eine algorithmische Diskriminierung treffen, wenn die Logik, mit der das KI-System zu der Bewertung kommt, nicht nachvollziehbar ist. Die Intransparenz von KI-Systemen wird von vielen großen Institutionen kritisiert (European Union Agency for Fundamental Rights 2020b, 2018; European Commission 2020; UNICRI 2020; United Nations 2020). Warum also wird die Transparenz und Nachvollziehbarkeit von KI-Systemen nicht gewährleistet, um Menschen vor algorithmischer Diskriminierung besser schützen zu können? Um dieser Frage nachzugehen, versuche ich einerseits darzulegen, inwiefern es sich bei KI um eine Blackbox handelt und andererseits warum die Logik nach der KI-Systeme ihre Entscheidungen treffen nicht nachvollziehbar gestaltet werden kann. Die Frage, ob KI-Algorithmen Intransparenz und fehlende Nachvollziehbarkeit als unveränderliche Merkmale ihrer technischen Funktionsweise in sich tragen, oder ob die beiden Merkmale viel mehr Ergebnis der Art und Weise sind, wie KI-Systeme und Software entwickelt werden, konnte ich im Rahmen meiner Recherche nicht beantworten. Ich habe der Quellenliteratur hierzu keine konkreten Informationen entnehmen können. Unternehmen schützen ihre KI-Systeme beispielsweise mit Patenten, sodass sie als Betriebsgeheimnis gelten und nicht offengelegt werden dürfen (vgl. Daum 2019, o.S.).

In Kapitel 3.2.4 gehe ich auf die Modifikation der Ungleichheitsdimension Ethnizität durch die fehlende Gerichtsbarkeit und Regulierung von KI-Systemen ein. Anfangs

beschreibe ich, welche Datentypen wie geschützt werden. In Staaten der EU ist die Verarbeitung besonderer personenbezogener Daten ohne die Genehmigung der betroffenen Person durch das DSGVO verboten (vgl. Europäische Union 2020). Die besonderen personenbezogenen Daten sind Informationen über die ethnische und kulturelle Herkunft, politische, religiöse und philosophische Überzeugungen, Gesundheit, Sexualität, Gewerkschaftszugehörigkeit, genetische Daten, biometrische Daten zur eindeutigen Identifizierung einer Person und Gesundheitsdaten (ebd.). Wenn die Person jedoch Daten in sozialen Netzwerken hochlädt oder anderen Unternehmen eine solche Genehmigung für die Verarbeitung ihrer Daten gibt, gehören die Daten dem jeweiligen Unternehmen. Die DSGVO schreibt zudem fest, dass Menschen das Recht auf informationelle Selbstbestimmung haben, was bedeutet, dass die Preisgabe und die Verwendung von personenbezogenen Daten selbstbestimmt funktionieren muss (vgl. Schunicht 2018, S.309). Guijarro-Santos wie auch andere Autor*innen führen an, dass beim Einsatz von KI-Systemen diese Grundrechte beschnitten werden können (vgl. Guijarro-Santos 2020, S. 50). Die Anwendungsbeispiele in dieser Arbeit untermauern das. Abschließend für diese Kategorie veranschauliche den unzureichenden Schutz durch diesen gesetzlichen Rahmen am Beispiel eines KI-Systems für Gesichtserkennung. Dieses KI-System ist öffentlich zugänglich, kostenlos, identifiziert eine Person mithilfe eines Profilbilds und stellt ein Persönlichkeitsprofil dieser Person zusammen. Dafür greift es auf eine Vielzahl von Quellen zurück, wie beispielsweise Webseiten von sozialen Netzwerken. Das Beispiel macht sichtbar, dass sensible Daten nicht geschützt werden.

Die Fundamental Rights Agency fordert für den Einsatz von Systemen künstlicher Intelligenz nicht nur eine legale Verarbeitung von Daten, die zu keiner unfairen Behandlung führt sondern auch die Anfechtbarkeit und Gerichtsbarkeit von Bewertungen von KI-Systemen sowie die Bewusstmachung von Betroffenen, dass ein KI-System eingesetzt wird, wenn es eingesetzt wird (vgl. European Union Agency for Fundamental Rights 2020c, S.12-13). Außerdem soll eine Nachvollziehbarkeit und Transparenz der Logik, nach der die algorithmischen Regelsysteme funktionieren gegeben sein (ebd.). Die Umsetzung solcher Forderungen können der Fundamental Rights Agency zufolge die Beschneidung der Persönlichkeitsrechte wie auch die Modifikation der sozialen Ungleichheit durch KI abfedern (ebd.). Die gesamte Kategorie der fehlenden Gerichtsbarkeit von algorithmischer Diskriminierung steht in

enger Verbindung zu der Kategorie „Überwachung durch KI-Systeme betrifft marginalisierte Gruppen stärker“.

In der Kategorie „Kann KI sozialer Ungleichheit entgegenwirken?“ versammele ich Argumente, die zeigen, wie KI-Systeme sozialer Ungleichheit entgegenwirken. Eine wichtige Erkenntnis ist, dass ich in der Recherche eine Vielzahl von Argumenten gefunden habe, die eine ungleichheitsverstärkende Modifikation beschreiben. Ich habe im Rahmen der Recherche weniger Argumente der Gegenposition gefunden. Für die Ungleichheitsdimension Klasse und Staatsbürgerschaft habe ich keine Argumente gefunden. Für diese beiden Kapitel habe ich die Kategorie deshalb auch nicht aufgeführt. Argumente dieser Kategorie fokussieren zumeist technische Lösungsansätze: Mackey et al. erkennen beispielsweise in Algorithmen selbst die Lösung zur Bewältigung der diskriminierenden Ergebnisse (Mackey et al. 2020, S. 1). Ein Beispiel für einen solchen technischen Lösungsansatz stellt der Gegenvektor dar. Dieser kann, wenn Frauen von einem KI-System benachteiligend bewertet werden, ein Gegengewicht setzen: Frauen werden durch einen Gegenvektor also automatisiert bevorteilt (vgl. Sutton et al. 2018, S.328). Aiallo et al. kritisieren, dass diese Lösungsansätze nicht ausreichend erforscht sind (vgl. Aiello et al. 2020, S. 109-110). Ein weiteres Argument gegen die ungleichheitsverstärkende Wirkung von KI ist, dass ein KI-System das Potenzial birgt „veraltete Annahmen über Geschlechterrollen in Frage zu stellen“ (Pinto 2020, S. 14). Pinto betont aber, dass es für die Erfüllung seines Arguments ein KI-System bedarf, dass hinreichend entwickelt ist (ebd.).

Im Kapitel 3.2.5 beschreibe ich Argumente gegen die ungleichheitsverstärkende Wirkung von KI-Systemen auf die Ungleichheitsdimension Ethnizität. Die Argumente in diesem Kapitel schließen an die aus 3.1.5 an. Ähnlich wie Pinto argumentiert auch Fröhlich und Spiecker: Der Fall des KI-Systems des AMS in Österreich (siehe Kapitel 3.1.1) zeigt den Wissenschaftlerinnen zufolge, dass gerade durch das Sichtbarmachen statistisch belegbarer Unterschiede eine effektivere Verteilung staatlicher Ressourcen gelingen kann (vgl. Fröhlich und Spiecker 2018, o.S.). Weiters führe ich ein Beispiel aus Italien an. Hier wird in Fußballstadien ein KI-System eingesetzt, sodass Personen, die rassistische Übergriffe begangen haben, in Echtzeit erkannt und überwacht werden (vgl. Algorithm Watch 2020b, S. 7).

Für die Kategorie „Überwachung von KI-Systemen trifft marginalisierte Gruppen mehr“ versammele ich Argumente und Konzepte, welche die Anwendung von KI-Systemen

zum Zweck der Überwachung in Zusammenhang mit sozialer Ungleichheit thematisiert. Ich verwende den Begriff der Überwachung in Anlehnung an die „black feminist theory“ und dem postkolonialen Diskurs (vgl. Poster 2011, S.868ff). Diese Kategorie betrifft nicht nur die Ungleichheitsdimension Ethnizität, sondern auch die Dimensionen Klasse und Gender, da Marginalisierung ein Phänomen aller Ungleichheitsdimensionen darstellt. Trotzdem habe ich die Kategorie nur in zwei der Kapiteln aufgeführt. Und zwar in den Kapiteln zu den Ungleichheitsdimensionen Ethnizität und Klasse. Die Recherche hat nicht zu genug Ergebnissen in Bezug auf die Modifikation der Ungleichheitsdimension Gender geführt.

In Kapitel 3.2.3 „Überwachung durch KI-Systeme betrifft marginalisierte Gruppen stärker“ zur Modifikation der Ungleichheitsdimension Ethnizität beleuchte ich das Thema Überwachung und Marginalisierung. Ich stelle die Funktionsweise von KI-Systemen dem Konzept der Mehrfachüberwachung entgegen. Das Konzept der Mehrfachüberwachung beschreibt, dass digitale Systeme das Spektrum der Agenten, die überwachen, erweitern und damit neue Quellen für Überwachung bieten (vgl. Poster 2011, S.868ff). Schmidt und Mellentin beschreiben eine kategorische Überwachung durch den Einsatz von KI-Systemen, weil eine große Menge an Daten aus unterschiedlichen Quellen genutzt wird, die grundsätzlich alle Menschen miteinschließt, deren Aktivitäten in Daten abgebildet sind (Schmidt und Mellentin 2020, S. 16). Auch Shoshana Zuboff stellt in ihrem Buch „Überwachungskapitalismus“ eine ähnliche These auf. Sie begründet diese jedoch mit den Wirkmechanismen der Wertschöpfungsprozesse im heutigen kapitalistischen System (vgl. Zuboff 2019, S. 11). Weiters erkläre ich, wie KI-Systeme die Möglichkeiten der Überwachung zusätzlich erweitern: Beispielsweise mithilfe der vielen Datentypen, die KI analysieren kann. KI-Systeme sind imstande alle Formen digitaler Information zu verarbeiten. Diese Konzepte und Argumente zeigen, dass die Möglichkeiten der Überwachung durch KI gewachsen sind. Für die Beantwortung der Forschungsfrage <*Wie modifiziert KI die Ungleichheitsdimensionen Ethnizität, Gender, Klasse und Staatszugehörigkeit?*> ist dieses Argument relevant, weil Überwachung marginalisierte Gruppen stärker betrifft, und eine Erweiterung des Spektrums an Möglichkeiten der Überwachung somit einen ungleichheitsverstärkenden Effekt besitzt: Überwachung ist von Macht- und Hierarchiestrukturen konstituiert (vgl. Schmidt und Mellentin 2020, S. 17; Poster 2019, S. 134; Roberts 2019, S. 341; Scannell 2019, S. 108; Schelenz 2021, S. 79ff).

Für Kapitel 3.3.2 „Überwachung durch KI-Systeme betrifft marginalisierte Gruppen stärker“ zur Modifikation der Ungleichheitsdimension Klasse durch KI, veranschauliche ich das Argument durch ein Beispiel. Das KI-System „SyRi“ wurde in den Niederlanden zur Aufdeckung von Sozialbetrug eingesetzt. Der Einsatz dieses KI-Systems hatte Menschen aus niedrigen Einkommensklassen im Visier, denn das KI-System überwachte Menschen die Sozialgeld vom Staat beziehen. Ein Algorithmus mit ähnlicher Funktionsweise könnte benutzt werden, um beispielsweise den Betrug von Steuergeldern aufzudecken. Hier wären nicht einkommensschwache Bevölkerungsgruppen, wie diejenigen, die von Sozialbetrug profitieren, im Visier der Algorithmen, sondern Menschen aus höheren Einkommensschichten. Solch ein KI-System, welches vom Staat im Einsatz oder in Entwicklung ist, habe ich im Rahmen meiner Recherche jedoch nicht finden können. Dieses Argument bezieht sich auf die gesellschaftliche Einbettung von KI. Es veranschaulicht, dass wie und für welchen Zweck KI-Systeme eingesetzt werden eine modifizierende Wirkung auf soziale Ungleichheit haben kann. Diese Form des Einsatzes hat eine ungleichheitsverstärkende Wirkung für die Ungleichheitsdimension Klasse.

In Bezug auf die Ungleichheitsdimension Klasse haben sich mehrfach Kategorien aus der Analyse ableiten lassen, die für die anderen Ungleichheitsdimensionen weniger oder gar nicht relevant sind. Eine Modifikation beschreibe ich mittels der Kategorie „Ideologische Einbettung von KI in kapitalistische Prinzipien“. Der Einsatz von KI zum Zweck der Effizienz und der Kosteneinsparung ist eine klassisch kapitalistisch motivierte Vorgehensweise (vgl. Dörre et al. 2013, S. 24). Der Zweck der Kosteneinsparung und der Effizienz ist in den meisten Fällen der Grund für des Einsatz von KI-Systemen. Ich erkenne hier eine ideologische Einbettung von KI in kapitalistische Prinzipien**.** Für die materielle Ungleichheit in Staaten bedeutet diese ideologische Einbettung das Weiterführen und Vertiefen kapitalistischer Praktiken. Für die Beantwortung der Forschungsfrage ist dieses Argument insofern relevant, als dass eine Vertiefung und das Weiterführen kapitalistischer Praktiken eine weiter andauernde Zunahme der sozioökonomischen Ungleichheit in Staaten mit sich bringt (vgl. Haller et al. 2015, S.1).

Unter der Kategorie „Arbeitsverlust durch KI?“ fasse ich Argumente und Konzepte zusammen deren inhaltliche Auseinandersetzung die Auswirkungen von KI-Systemen auf den Arbeitsmarkt und damit auch auf einen möglichen Arbeitsverlust durch

sektorale Ungleichheiten thematisieren (vgl. Dannecker et al. 2010, S.70). Das Fragezeichen unterstreicht, dass die Meinungen von Expert*innen in Bezug auf Arbeitsverluste durch KI unterschiedliche Standpunkte einnehmen. Es wird jedoch betont, dass es an Analysen fehlt um eine konkrete Vorhersage tätigen zu können (vgl. Hunt et al. 2022, o.S.). Wissenschaftler*innen kritisieren zudem, dass Vorhersagen über Arbeitsverluste durch KI auf Modellen basieren und von subjektiven Einschätzungen verfälscht sein können (vgl. Petropoulos 2018, S. 128; Hunt et al. 2022, o.S.). Das Fragezeichen zeigt also, dass es auf die Frage nach der Modifikation materieller Ungleichheit aufgrund von Arbeitsplatzverlusten durch KI noch keine klare Antwort gibt.

Die Kategorien „Ideologische Einbettung von KI in kapitalistische Prinzipien“ und „Internationaler Wettbewerb um KI“ ähneln sich. In der Kategorie „Ideologische Einbettung von KI in kapitalistischen Prinzipien“ habe ich Argumente und Konzepte gesammelt, die in der Einbettung von KI-Systemen in das kapitalistische System Modifikationen für die Ungleichheitsdimension Klasse erkennen. Während ich bei dieser Kategorie tendenziell mehr Argumente gefunden habe, die das Streben nach Effizienz als Merkmal des Kapitalismus fokussieren (vgl. Dörre et al. 2013, S. 24), habe ich für die Kategorie „Ideologische Einbettung von KI in kapitalistische Prinzipien“ mehr Argumente gefunden, die den Wettbewerb als eines der Merkmale des kapitalistischen Systems fokussieren (vgl. Dörre et al. 2013, S. 14). Deshalb habe ich mich für eine unterschiedliche Benennung der Kategorien entschieden.

Die Modifikationen der internationalen Ungleichheit sind in drei Kategorien zusammengefasst. Die erste Kategorie habe ich bereits beschrieben (siehe Anfang, Kapitel 4). Sie thematisiert die Modifikationen durch die Datenbasiertheit von KI. Die beiden anderen Kategorien versammeln Argumente in Bezug auf den internationalen Wettbewerb um KI und auf KI als Projekt reicher Länder. In Kapitel 3.4.2 „Internationaler Wettbewerb um KI“ geht es um die Modifikation materieller Ungleichheit zwischen Staaten durch ungleiche finanzielle Möglichkeiten in der Investition in KI. Wenn KI in dem vorhergesagten Ausmaß zum wirtschaftlichen Wachstum beiträgt (vgl. AI Watch 2020, S. 7), kann es zu Vertiefung oder Fortführung bestehender materieller Ungleichheiten zwischen Ländern kommen (vgl. Schelenz 2021, S. 83).

In Kapitel 3.4.3 „KI als Projekt reicher Länder“ geht es darum, dass von den arbeitsteiligen Prozessen der Entwicklung von KI mehrheitlich Länder mit hohem Einkommensniveau profitiert. Diese Entwicklung wird unter dem Begriff des digitaler Kolonialismus diskutiert (Tuzcu 2021; Coleman 2018; Schelenz 2021). Weiters beschreibe ich, wie KI eine neue Form der Wissensproduktion prägt: In dieser Wissensproduktion besitzen vor allem große multinationale Konzerne die Datenmengen, die es benötigt um Wissen mit KI-Systemen zu produzieren (vgl. Tuzcu 2021, 514ff). Beide Argumente haben eine Vertiefung geopolitischer Hierarchien zwischen Ländern mit hohem und niedrigem Einkommensniveau zur Folge und modifizieren so die internationale Ungleichheit.

5. Fazit: Ist Technik immer nur so gut, wie die Gesellschaft, die sie hervorbringt?

Eine Erkenntnis, die ich im Rahmen dieser Arbeit gewonnen habe, hält mich davon ab Aussagen über die zukünftige Entwicklung von KI und Gesellschaft zu machen. Die Entwicklungspfade von KI in der Gesellschaft sind aufgrund der komplexen systemischen Zusammenhänge schwer zu verstehen. Trotzdem wagen viele Personen Vorhersagen über ebenjene Entwicklungspfade. Manche dieser Vorhersagen sind von subjektiven Einschätzungen verfälscht (vgl. Petropoulos 2018, S. 128; Hunt et al. 2022, o.S.) und machen teilweise einen reißerischen Eindruck. In der Frage nach der zukünftigen Entwicklung von KI gibt es viele gegenläufige Positionen. Ich will das mittels folgendem Beispiels veranschaulichen:

Der Einsatz von KI im Bereich der Pflege kann zwei gänzlich gegensätzliche Effekte haben: Der Einsatz von KI-gestützten Pflegerobotern könnte zu einem Anstieg der Isolation und Arbeitsplatzverlusten führen, „wenn beispielsweise Krankenschwestern durch Pflegeroboter ersetzt werden (Kaplan und Haenlein 2020, S. 43). Es kann jedoch auch bedeuten, dass einsamen Menschen durch den Einsatz von humanoiden Pflegerobotern geholfen werden kann und das Problem des Arbeitskräftemangels in Krankenhäusern, Sozialstationen oder Altenheimen beseitigt wird (ebd.). Je nachdem ob und wie der Einsatz von Robotern im Bereich der Pflege funktioniert und wie sich dieser Einsatz auf den Arbeitskräftemangel auswirkt, kann dies gegensätzliche Effekte auf die soziale Ungleichheit haben.

Die systemischen Zusammenhänge von KI und Gesellschaft sind komplex. Es ist schwer generalisierende Aussagen zu den Auswirkungen von KI auf die soziale Ungleichheit zu treffen. Im Rahmen dieser Arbeit habe ich aktuelle Entwicklungen beschrieben und hier eine klare Trennung zu Vorhersagen vorgenommen.

Eine weitere Erkenntnis, die mich in der Auseinandersetzung mit KI und sozialer Ungleichheit begleitet hat, will ich im Folgenden festhalten.

Technikentwicklungen bergen nicht nur neue Lösungen und Probleme, sondern auch das Potenzial transformatorischer Prozesse. Dies wird offensichtlich, wenn man sich die Funktionsweise von Algorithmen vor Augen ruft: Sie besteht aus der Automatisierung von Abläufen. Ist ein Algorithmus einmal installiert und funktioniert, verzichtet er auf Konsenszwänge. So zumindest definiert Nassehi das

Charakteristikum der Technik: Solange sie funktioniert, wird ihre Funktionsweise nicht mit den Nutzenden ausgehandelt (vgl. Nassehi 2019, S. 196). Gepaart mit dem heutigen Vertrauen unserer Gesellschaft in Zahlen, das in den Publikationen von Felt & Davies mehrmals aufgegriffen wird (vgl. Davies und Felt 2020), entsteht eine gefährliche Mischung: Ein Algorithmus birgt, wenn er einmal installiert ist und funktioniert, die Gefahr, dass er langfristig auf Menschen wirkt, ohne dass es einen wiederkehrenden Konsens über dessen Funktionsweise oder Wirkung gibt.

Laut einer Umfrage des Pew Research Center aus dem Jahr 2018 glauben 40 % der Menschen, dass KI-Systeme objektiv und frei von den Verzerrungen sein können, welche die menschliche Entscheidungsfindung beeinträchtigen (vgl. Johnson 2022b, S. 27) . Menschen nehmen mehrheitlich an, dass KI neutraler sei als das menschliche Analyse- und Urteilsvermögen (vgl. Wachter-Boettcher 2017, 14ff). Diese Haltung ist im Bereich des kommerziellen Technikdesigns weit verbreitet und verbunden mit vorherrschenden Werten wie Meritokratie und Gleichheit (vgl. Schelenz 2021, S. 83).

In der Technik-Gesellschafts-Wissenschaft beschreiben die Begriffe Technik- und Soziodeterminismus zwei gegensätzliche Standpunkte: Der Technikdeterminismus geht davon aus, dass sich der Raum der Möglichkeiten, der sich durch eine neue Technik öffnet, eigendynamisch und unabhängig von gesellschaftlichen Prozessen entfaltet (vgl. Grunwald und Hubig 2018, S.323). Soziodeterministische Perspektiven betonen im Gegensatz dazu die Dominanz der gesellschaftlichen Prozesse und damit die Bestimmtheit von Technik durch bestehende gesellschaftliche Dynamiken. Aus der technikdeterministischen Perspektive betrachtet, scheint es, als würde KI ihre Entwicklungspfade in der Gesellschaft eigendynamisch prägen. Sie eröffnet einen Raum an technischen Möglichkeiten und diese bestimmen den Pfad, den die Gesellschaft mit der Technik geht. Im technikdeterministischen Ansatz besitzt die Zivilgesellschaft keine Partizipationsmöglichkeit in der Auswahl der Entwicklungspfade mit der Technik. Bei soziodeterministischen Ansätzen prägt die Gesellschaft die Nutzung und die Art und Weise des Einsatzes von Technologien. Die Technologie ist ein Ausdruck des Zeitgeists und passiert nicht eigendynamisch (Weber und Prietl 2021, S. 68–69).

Neben den Begriffen Sozio- und Technikdeterminismus gibt es noch einen weiteren wichtigen Begriff, der eine kritische Position gegenüber der gesellschaftlichen Umgangsweise mit Technologie einnimmt. Morozov prägt das Konzept des

„technological solutionism“. Er beschreibt damit eine Beobachtung der Gesellschaft: Es würde sich „[…] jedes Problem, auch das gesellschaftliche, letztlich auf eine Reihe kleiner und damit überschaubarer Probleme reduzieren [lassen], für die dann technische Lösungen gefunden werden können“ (frei übersetzt nach Weber und Prietl 2021, S. 67). Das Konzept hat seinen Ursprung in der Ideologie des Silicon Valley, in der der optimistische Glaube an die Technologie Hand in Hand mit liberalen Prinzipien geht (ebd.). Die Autorinnen Weber und Prietl kritisieren diese Ideologie und den „technological solutionism“ für ihr Verkennen bestehender Ungleichheiten im Zugang zu Technik und Bildung sowie der Reproduktion dieser Ungleichheiten durch und in digitalen Technologien (vgl. Weber und Prietl 2021, S. 67–68). Algorithm Watch beschreibt den „technological solutionism“ als „eine fehlerhafte Ideologie, die jedes soziale Problem als ‚Fehler‘ ansieht, der durch Technologie ‚behoben‘ werden muss“ (Algorithm Watch 2020a, S. 3).

Der Begriff der digitalen Mündigkeit greift diese Phänomene der aktuellen sozio-technischen Entwicklungen auf. Er ruft die Gesellschaft auf, sich aktiv an Technik zu beteiligen. Es sollte nach Simon den Anspruch geben, Technik zu verstehen und damit Eigenverantwortung für Handeln und Agieren mit Technik zu übernehmen (vgl. Simon 2020, S. 37). Ich erkenne in dieser Forderung eine Gegenposition zu der Theorie der digitalen Gesellschaft von Armin Nassehi. In seiner Theorie beschreibt Nassehi die Funktion von Technik als das Funktionieren selbst (vgl. Nassehi 2019, S. 196). Technik folgt daher keinem Konsenszwang. Ein kurzes Beispiel: Du kommst zu spät zu einer Theatervorführung. Mit einem Menschen kannst du aushandeln, ob du trotzdem noch eine Eintrittskarte zu der Vorführung bekommst. Mit einem Ticketautomat geht das nicht. Dieses Beispiel zeigt, dass Technik, wenn sie einmal installiert ist, auf Konsenszwang verzichtet. Diese Umgangsweise mit Technik trifft Nassehi zufolge auf jede Technik zu. Für KI-Systeme und ihre ungleichheitsverstärkenden Wirkmechanismen ist dieser Zusammenhang von Gesellschaft und Technik ein Risiko. Ebenso die Beobachtungen von Morozov über den „technological solutionism“:

Erst wenn Technik nicht mehr nur funktioniert, wie Nassehi es in seiner Definition festhält; wenn technikdeterministische Ansichten nicht mehr den Blick auf technik-gesellschaftliche Entwicklungen prägen, wenn „technological solutionism“ nicht die alleinige Strategie zur Lösung von Problemen ist – dann würde KI aktiver diskriminierungsfrei gestaltet werden können. Denn: „Technik ist am Ende schließlich

immer nur so gut oder so schlecht, wie die Gesellschaft, die sie hervorbringt“ (netzforma* e.V. 2020, S. 8–9).

Was sind weitere Stellschrauben, die es zu drehen gilt, um der ungleichheitsverstärkenden Tendenz von KI etwas entgegenzuhalten? Die Ergebnisse dieser Arbeit zeigen, dass viele der ungleichheitsverstärkenden Mechanismen von KI eng gekoppelt an das kapitalistische System und an bestehende, gesellschaftlich konstruierte Formen der Diskriminierung und Marginalisierung sind. Vielleicht macht erst eine Veränderung dieser gesellschaftlichen Strukturen eine Veränderung in KI möglich. Eine Neuerung, durch die KI-Systemen abseits von kapitalistischen Prinzipien wie Wettbewerb und Effizienz gedacht werden können. Denn die Art und Weise wie KI-Systeme eingesetzt werden ist entscheidend für die Auswirkungen solcher Systeme.

Bestehende, gesellschaftlich konstruierte Formen der Diskriminierung stellen eine weitere Hürde da: Solange Rassismen und Sexismen in unserer Gesellschaft existieren, sind sie in Daten und dadurch auch in datenbasierten KI-Systemen abgebildet. Technische Lösungen zur Überwindung dieser Problematik bringen noch nicht den gewünschten Erfolg, wie die Beispiele in dieser Arbeit zeigen.

Eine Chance könnte die Neuerung der Beziehung zwischen Gesellschaft und Technik bergen. Beispielsweise die Stärkung der digitalen Mündigkeit in der Zivilgesellschaft oder auch eine Sensibilisierung für die Auswirkungen von KI-Systemen könnten Ergebnis von oder Impuls zu einer solchen Neuerung sein.

6. Literaturverzeichnis

Aggarwal, Nikita (2021): The norms of algorithmic credit scoring. In: *The Cambridge Law Journal* 80 (1), S. 42–73.

AI Watch (2020): Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence. Unter Mitarbeit von Europäische Kommission.

Akremi, Leila (2014): Stichprobenziehung in der qualitativen Sozialforschung. In: Nina Baur und Jörg Blasius (Hg.): Handbuch Methoden der empirischen Sozialforschung: Springer, S. 265–283.

Algorithm Watch (2020a): Automated Decision-Making Systems in the COVID-19 Pandemic: A European Perspective.

Algorithm Watch (2020b): Automating Society Report. Unter Mitarbeit von Fabio Chiusi, Sarah Fischer, Nicolas Kayser-Bril und Matthias Spielkamp. Berlin. Online verfügbar unter https://automatingsociety.algorithmwatch.org.

Algorithm Watch (2020c): How Dutch activists got an invasive fraud detection algorithm banned. Unter Mitarbeit von Koen Vervloesem. Online verfügbar unter https://algorithmwatch.org/en/syri-netherlands-algorithm/, zuletzt aktualisiert am 15.08.2022.

Algorithm Watch (2022). Online verfügbar unter https://algorithmwatch.org/de/was-wir-tun/, zuletzt geprüft am 12.09.2022.

Altenburger, Reinhard; Schmidpeter, René (2021): CSR und Künstliche Intelligenz. Berlin, Heidelberg: Springer Berlin Heidelberg.

Alugubelli, Raghunandan (2016): Exploratory Study of Artificial Intelligence in Healthcare. In: *International Journal of Innovations in Engineering Research and Technology* 3 (1), S. 1–10.

Angwin, Julia; Larson, Jeff (2022): Bias in Criminal Risk Scores Is Mathematically Inevitable, Researchers Say *. In: Kirsten Martin (Hg.): Ethics of Data and Analytics. Boca Raton: Auerbach Publications, S. 265–267.

Ashby, W. Ross (1961): An introduction to cybernetics: Chapman & Hall Ltd.

Aulenbacher, Brigitte; Riegraf, Birgit (2012): Intersektionalität und soziale Ungleichheit.

Bacchini, Fabio; Lorusso, Ludovica (2019): Race, again: how face recognition technology reinforces racial discrimination. In: *Journal of information, communication and ethics in society*.

Bauer, Kevin; Hinz, Oliver; Weber, Patrick (2021): KI in der Finanzbranche: Im Spannungsfeld zwischen technologischer Innovation und regulatorischer Anforderung (Nr. 80).

Beattie, Geoffrey; Johnson, Patrick (2012): Possible unconscious bias in recruitment and promotion and the need to promote equality. In: *Perspectives: Policy and Practice in Higher Education* 16 (1), S. 7–13.

Behrmann, Laura; Eckert, Falk; Gefken, Andreas (2018): Prozesse sozialer Ungleichheit aus mikrosoziologischer Perspektive – eine Metaanalyse qualitativer Studien. In: Laura Behrmann, Falk Eckert, Andreas Gefken und Peter A. Berger (Hg.): ‚Doing Inequality': Prozesse sozialer Ungleichheit im Blick qualitativer Sozialforschung. Wiesbaden: Springer Fachmedien Wiesbaden, S. 1–34.

Berner, Heiko; Schüll, Elmar (2020): Bildung nach Maß. Die Auswirkungen des AMS-Algorithmus auf Chancengerechtigkeit, Bildungszugang und Weiterbildungsförderung. In: *Magazin erwachsenenbildung. at* (40).

Blattner, Laura; Nelson, Scott (2021): How costly is noise? Data and disparities in consumer credit.

Bogen, Miranda; Rieke, Aaron (2018): Help Wanted - An Exploration of Hiring Algorithms, Equity and Bias. In: *Upturn*.

Brownstein, John S.; Freifeld, Clark C.; Reis, Ben Y.; Mandl, Kenneth D. (2008): Surveillance Sans Frontieres: Internet-based emerging infectious disease intelligence and the HealthMap project. In: *PLoS medicine* 5 (7), e151.

Bundesregierung der Bundesrepublik Deutschland (2020): Stellungnahme der Bundesregierung der Bundesrepublik Deutschland zum Weißbuch zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen.

Buolamwini, Joy; Gebru, Timnit (Hg.) (2018): Gender shades: Intersectional accuracy disparities in commercial gender classification. Conference on fairness, accountability and transparency: PMLR.

Chen, Jonathan; Beam, Andrew (2019): Potential trade-offs and unintended consequences of AI. In: *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril*, S. 89.

Coleman, Danielle (2018): Digital colonialism: The 21st century scramble for Africa through the extraction and control of user data and the limitations of data protection laws. In: *Mich. J. Race & L.* 24, S. 417.

Corbett-Davies, Sam; Pierson, Emma; Feller, Avi; Goel, Sharad; Huq, Aziz (2017): Algorithmic Decision Making and the Cost of Fairness. In: Stan Matwin, Shipeng Yu und Faisal Farooq (Hg.): Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '17: The 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Halifax NS Canada, 13 08 2017 17 08 2017. New York, NY, USA: ACM, S. 797–806.

Crenshaw, Kimberlé (2018): Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory, and antiracist politics [1989]. In: Katharine T. Bartlett und Rosanne Kennedy (Hg.): Feminist legal theory: Routledge, S. 57–80.

Daniels, Jessie (2015): “My Brain Database Doesn’t See Skin Color” Color-Blind Racism in the Technology Industry and in Theorizing the Web. In: *American Behavioral Scientist* 59 (11), S. 1377–1393.

Dastin, Jeffrey (2022): Amazon Scraps Secret AI Recruiting Tool that Showed Bias against Women *. In: Kirsten Martin (Hg.): Ethics of Data and Analytics. Boca Raton: Auerbach Publications, S. 296–299.

Daum, Timo (2019): Die künstliche Intelligenz des Kapitals: Edition Nautilus.

Davies, Sarah R.; Felt, Ulrike (2020): Exploring science communication: A science and technology studies approach. In: *Exploring Science Communication*, S. 1–264.

De Rechtspraak (2020): SyRI legislation in violation of the European Convention on Human Rights. Hg. v. Rechtbank Den Haag. Online verfügbar unter https://www-rechtspraak-nl.translate.goog/Organisatie-en-contact/Organisatie/Rechtbanken/Rechtbank-Den-Haag/Nieuws/Paginas/SyRI-wetgeving-in-strijd-met-het-Europees-Verdrag-voor-de-Rechten-voor-de-Mens.aspx?_x_tr_sl=nl&_x_tr_tl=en&_x_tr_hl=en.

Dörre, Klaus; Lessenich, Stephan; Rosa, Hartmut (2013): Soziologie-Kapitalismus-Kritik: eine Debatte: Suhrkamp Verlag.

DSK (Hg.) (2019): Positionspapier der DSK zu empfohlenen technischen und organisatorischen Maßnahmen bei der Entwicklung und dem Betrieb von KI-Systemen. Datenschutzkonferenz.

ECIAIR (Hg.) (2019): Racist Soapdishes and Rebellious (?) Children: Towards Human/AI Cooperation. ECIAIR 2019 European Conference on the Impact of Artificial Intelligence and Robotics: Academic Conferences and publishing limited.

Edelman, Benjamin G.; Luca, Michael (2014): Digital discrimination: The case of Airbnb.com. In: *Harvard Business School* (14-045), S. 1–21.

Eggers, Maureen Maisha; Kilomba, Grada; Piesche, Peggy; Arndt, Susan (2005): Mythen, Masken und Subjekte. In: *Kritische Weißseinsforschung in Deutschland. Münster: Unrast*.

Eglash, Ron (2019): Anti--Racist Technoscience A Generative Tradition. In: Ruha Benjamin (Hg.): Captivating Technology: Duke University Press, 227-251.

Erden, Deniz (2020): KI und Beschäftigung. Das Ende menschlicher Voreingenommenheit oder der Anfang von Diskriminierung 2.0. In: netzforma* e.V. (Hg.): Wenn KI dann feministisch. Impulse aus Wissenschaft und Aktivismus, S. 77–90.

Ertel, Wolfgang; Black, Nathanael T. (2016): Grundkurs Künstliche Intelligenz: Springer.

Eschholz, Stefanie; Djabbarpour, Jonathan (2018): Big Data and Scoring in the Financial Sector. In: Thomas Hoeren und Barbara Kolany-Raiser (Hg.): Big Data in Context. Cham: Springer, S. 63–71.

Europäische Kommission (2018a): Artificial intelligence : a European perspective: Publications Office.

Europäische Kommission (2018b): Eine Definition der KI: Wichtigste Fähigkeiten und Wissenschaftsgebiete. Unter Mitarbeit von hochrangige Expertengruppe für künstliche Intelligenz.

Europäische Kommission (2020): AI watch. 2019 activity report. Luxembourg: Publications Office of the European Union (JRC technical report).

Europäische Union (2020): Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung). In: Fabian Schuster und Malte Grützmacher (Hg.): IT-Recht: Verlag Dr. Otto Schmidt, S. 47–194.

European Commission (2020): White Paper on Artificial Intelligence. A European approach to excellence and trust.

European Union Agency for Fundamental Rights (2018): #BigData: Discrimination in data-supported decision making.

European Union Agency for Fundamental Rights (2020a): Facial recognition technology: fundamental rights considerations in the context of law enforcement. In: *Eur. Data Prot. L. Rev.* 6.

European Union Agency for Fundamental Rights (2020b): Getting the future right – Artificial intelligence and fundamental rights.

European Union Agency for Fundamental Rights (2020c): Getting the Future Right – Artificial Intelligence and Fundamental Rights. report presentation. Unter Mitarbeit von Joanna Goodey. Berlin, 14.12.2020.

Feathers, Todd (2022): Major Universities Are Using Race as a “High Impact Predictor” of Student Success *. In: Kirsten Martin (Hg.): Ethics of Data and Analytics. Boca Raton: Auerbach Publications, S. 268–273.

Fjelland, Ragnar (2020): Why general artificial intelligence will not be realized. In: *Humanities and Social Sciences Communications* 7 (1), S. 1–9.

Frankfurter Allgemeine (2018): Twitter hat Daten an Cambridge Analytica verkauft. Hg. v. Frankfurter Allgemeine. Online verfügbar unter https://www.faz.net/aktuell/wirtschaft/digitec/twitter-hat-daten-an-cambridge-analytica-verkauft-15567421.html, zuletzt aktualisiert am 30.04.2018, zuletzt geprüft am 26.02.2022.

Frey, Carl Benedikt; Osborne, Michael (2013): The future of employment.

Frey, Philipp; Schaupp, Simon; Wenten, Klara-Aylin (2021): Towards Emancipatory Technology Studies. In: *Nanoethics* 15 (1), S. 19–27. DOI: 10.1007/s11569-021-00388-6.

Fröhlich, Wiebke; Spiecker, Indra (2018): Können Algorithmen diskriminieren? Online verfügbar unter https://verfassungsblog.de/koennen-algorithmen-diskriminieren/.

Gless, Sabine; Wohlers, Wolfgang; Böse, Martin; Schumann, K. H.; Toepel, Friedrich (2019): Subsumtionsautomat 2.0 Künstliche Intelligenz Statt Menschlicher Richter? In: *Festschrift für Urs Kindhäuser zum* 70, S. 147–165.

Goodman, Kenneth; Zandi, Diana; Reis, Andreas; Vayena, Effy (2020): Balancing risks and benefits of artificial intelligence in the health sector. In: *Bulletin of the World Health Organization* 98 (4), S. 230.

Greve, Jens (2010): Globale Ungleichheit: Weltgesellschaftliche Perspektiven. In: *Berlin J Soziol* 20, S. 87. DOI: 10.1007/s11609-010-0117-9.

Grunwald, Armin; Hubig, Christoph (Hg.) (2018): Technikhermeneutik: ein kritischer Austausch zwischen Armin Grunwald und Christoph Hubig. Arbeit und Spiel: Nomos Verlagsgesellschaft mbH & Co. KG.

GSMA (2019): The Mobile Gender Gap Report 2019. Connected Women. In: *GSMA*.

Guijarro-Santos, Victoria (2020): Effiziente Ungleichheit. In: netzforma* e.V. (Hg.): Wenn KI dann feministisch. Impulse aus Wissenschaft und Aktivismus, S. 47–64.

Haggenmüller, Sarah; Maron, Roman C.; Hekler, Achim; Utikal, Jochen S.; Barata, Catarina; Barnhill, Raymond L. et al. (2021): Skin cancer classification via convolutional neural networks: systematic review of studies involving human experts. In: *European Journal of Cancer* 156, S. 202–216.

Haller, Max; Eder, Anja; Kmet, Bernadette Müller (2015): Drei Wege zur Zähmung des Kapitalismus. In: *Österreichische Zeitschrift für Soziologie* 40 (1), S. 1–31. DOI: 10.1007/s11614-015-0153-y.

Haraway, Donna (1988): Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective (14), Artikel 3, S. 575–599. Online verfügbar unter http://www.jstor.org/stable/3178066.

Hartmann, Jutta; Klesse, Christian; Fritzsche, Bettina; Wagenknecht, Peter; Hackmann, Kristina (2008): Heteronormativität: Empirische Studien zu Geschlecht, Sexualität und Macht: Springer.

Harwell, Drew (2022): A Face-Scanning Algorithm Increasingly Decides Whether You Deserve the Job *. In: Kirsten Martin (Hg.): Ethics of Data and Analytics. Boca Raton: Auerbach Publications, S. 206–211.

HealthMap (2022): The Desease Daily. Online verfügbar unter https://www.diseasedaily.org/about/, zuletzt geprüft am 15.08.2022.

Hill, Kashmir (2022): Wrongfully Accused by an Algorithm *. In: Kirsten Martin (Hg.): Ethics of Data and Analytics. Boca Raton: Auerbach Publications, S. 138–142.

HireVue (2022): Video interviewing software. Screen People, Not CVs. Hg. v. HireVue. Online verfügbar unter https://www.hirevue.com/platform/online-video-interviewing-software, zuletzt geprüft am 19.07.2022.

Horwath, Ilona (2022): Algorithmen, KI und soziale Diskriminierung. In: Kordula Schnegg, Julia Tschuggnall, Caroline Voithofer und Manfred Auer (Hg.): Inter- und multidisziplinäre Perspektiven der Geschlechterforschung. Innsbrucker Gender Lectures IV. Innsbruck, S. 71–102.

Hradil, Stefan (2016): Soziale Ungleichheit, soziale Schichtung und Mobilität. In: Einführung in Hauptbegriffe der Soziologie: Springer, S. 247–275.

Hunt, Wil; Sarkar, Sudipa; Warhurst, Chris (2022): Measuring the impact of AI on jobs at the organization level: Lessons from a survey of UK business leaders. In: *Research Policy* 51 (2), S. 104425.

Interpol; UNICRI (2020): Towards Responsible AI Innovation. Second Interpol-UNICRI report on artificial intelligence for law enforcement. Unter Mitarbeit von Odhran McCarthy und Maria Eira.

Jalloh, Linda (2022): # BLACKFEMINISM. In: *FKW//Zeitschrift für Geschlechterforschung und visuelle Kultur* (70).

Johnson, Carolyn Y. (2022a): Racial Bias in a Medical Algorithm Favors White Patients over Sicker Black Patients *. In: Kirsten Martin (Hg.): Ethics of Data and Analytics. Boca Raton: Auerbach Publications, S. 10–12.

Johnson, Gabbrielle M. (2022b): Excerpt from Are Algorithms Value-Free? Feminist Theoretical Virtues in Machine Learning *. In: Kirsten Martin (Hg.): Ethics of Data and Analytics. Boca Raton: Auerbach Publications, S. 27–35.

Johnson, James (2021): The end of military-techno Pax Americana? Washington's strategic responses to Chinese AI-enabled military technology. In: *The Pacific Review* (3), S. 351–378.

Kaplan, Andreas; Haenlein, Michael (2020): Rulers of the world, unite! The challenges and opportunities of artificial intelligence. In: *Business Horizons* 63 (1), S. 37–50.

Karusala, Naveena; Kumar, Neha; Bhalla, Apoorva (2019): Privacy, patriarchy, and participation on social media. In: *Genderer Considerations*.

Kemper, Andreas; Kuleßa, Peter (2020): Gegen die Missachtung von Armut. Klassismus endlich ernst nehmen. In: *Theorie und Praxis der sozialen Arbeit 1/2020*.

Kirste, Moritz; Schürholz, Markus (2019): Einleitung: Entwicklungswege zur KI. In: Volker Wittpahl (Hg.): Künstliche Intelligenz. Technologie Anwendung Gesellschaft. Berlin: Springer, S. 21–35.

Knapp, Gudrun-Axeli (2018): Auf ein Neues!? Feministische Kritik im Wandel der Gesellschaft. In: *Geschlecht, Differenz und Identität. Darmstadt*.

Kodiyan, Akhil Alfons (2019): An overview of ethical issues in using AI systems in hiring with a case study of Amazon's AI based hiring tool. In: *Researchgate Preprint*, S. 1–19.

Kohli, Martin; Künemund, Harald; Motel, Andreas; Szydlik, Marc (2000): Soziale Ungleichheit. In: Harald Künemund (Hg.): Die zweite Lebenshälfte: Springer, S. 318–336.

Krause, Ulf von (2021): Potenziale der KI im Militär. In: Ulf von Krause (Hg.): Künstliche Intelligenz im Militär. Chancen und Risiken für die Sicherheitspolitik: Springer, S. 9–23.

Kreckel, Reinhard (1983): Soziale Ungleichheiten: Schwartz Göttingen.

Kwet, Michael (2019): Digital colonialism: US empire and the new imperialism in the Global South. In: *Race & Class* 60 (4), S. 3–26.

Lengersdorf, Diana; Weber, Jutta (2019): Gender, Technik und Politik 4.0. Über digitalen Kapitalismus, disruptive Technologien und neue Regime der Unsicherheit. In: *Special issue of the journal "Gender"* (11), S. 7–10.

Ligo, Alexandre K.; Rand, Krista; Bassett, Jason; Galaitsi, Stephanie Elisabeth; Trump, Benjamin D.; Jayabalasingham, Bamini et al. (2021): Comparing the emergence of technical and social sciences research in artificial intelligence. In: *Frontiers in Computer Science* 3, S. 653235. Online verfügbar unter https://www.frontiersin.org/articles/10.3389/fcomp.2021.653235/full#h1, zuletzt geprüft am 04.08.2022.

Lohr, Steve (2022): Facial Recognition Is Accurate, If You're a White Guy *. In: Kirsten Martin (Hg.): Ethics of Data and Analytics. Boca Raton: Auerbach Publications, S. 143–147.

Lücking, Phillip (2020): Automatisierte Ungleichheit. Wie algorithmische Entscheidungssysteme gesellschaftliche Machtverhältnisse (re)produzieren. In: netzforma* e.V. (Hg.): Wenn KI dann feministisch. Impulse aus Wissenschaft und Aktivismus, S. 65–76.

Martin, Kirsten (2022): Discrimination and Data Analytics. In: Kirsten Martin (Hg.): Ethics of Data and Analytics. Boca Raton: Auerbach Publications, S. 291–295.

Masuhr, Niklas (2019): KI als militärische Befähigungstechnologie. In: *CSS Analysen zur Sicherheitspolitik* 251.

Mayring, Philipp (2014): Qualitative content analysis: theoretical foundation, basic procedures and software solution. Klagenfurt. Online verfügbar unter https://nbn-resolving.org/urn:nbn:de:0168-ssoar-395173, zuletzt geprüft am 28.07.2022.

Mayring, Philipp; Fenzl, Thomas (2014): Qualitative Inhaltsanalyse. In: Nina Baur und Jörg Blasius (Hg.): Handbuch Methoden der empirischen Sozialforschung: Springer, S. 543–558.

Milanovic, Branko (2009): Global inequality and the global inequality extraction ratio: the story of the past two centuries. In: *World Bank Policy Research Working Paper* (5044).

Müller, Hanns-Ferdinand (2021): Hype oder Realität – KI als Instrument zur Optimierung des vertrieblichen Portfolios eines juristischen Finanzdienstleisters. In: Meike Terstiege (Hg.): KI in Marketing & Sales – Erfolgsmodelle aus Forschung und Praxis: Konzepte und Instrumente zum erfolgreichen Einsatz künstlicher Intelligenz. Wiesbaden: Springer Fachmedien Wiesbaden, S. 275–291.

Nassehi, Armin (2019): Muster. Theorie der digitalen Gesellschaft. München: C.H. Beck.

netzforma* e.V. (Hg.) (2020): Wenn KI dann feministisch. Impulse aus Wissenschaft und Aktivismus. Online verfügbar unter https://netzforma.org/publikation-wenn-ki-dann-feministisch-impulse-aus-wissenschaft-und-aktivismus.

Noble, Safiya Umoja (2019): Noble, Safiya Umoja. Algorithms of Oppression: How Search Engines Reinforce Racism. New York University Press, 2018. In: *WHY POPULAR CULTURE MATTERS*, S. 166.

Nolting, Michael (2021): Autonomes Fahren und Künstliche Intelligenz. In: Michael Nolting (Hg.): Künstliche Intelligenz in der Automobilindustrie: Springer, S. 113–129.

Ntoutsi, Eirini; Fafalios, Pavlos; Gadiraju, Ujwal; Iosifidis, Vasileios; Nejdl, Wolfgang; Vidal, Maria-Esther et al. (2020): Bias in data-driven artificial intelligence systems—An introductory survey. In: *WIREs Data Mining Knowl Discov* 10 (3). DOI: 10.1002/widm.1356.

ORF (2020): Zukunft für AMS-Algorithmus ungewiss. Online verfügbar unter https://orf.at/stories/3178348/, zuletzt geprüft am 20.02.2022.

Panesar, Arjun (2019): Machine learning and AI for healthcare: Springer.

Petropoulos, Georgios (2018): The impact of artificial intelligence on employment. In: *Praise for Work in the Digital Age*.

PimEyes (2022): Face Search Engine - Reverse Image Search. Online verfügbar unter https://pimeyes.com/en, zuletzt geprüft am 14.08.2022.

Pinto, Renata Ávila (2020): Globale Trends - Analysen. Tech Power to the people! Demokratisierung von Zukunftstechnologien im Dienst der Gesellschaft.

Poba-Nzaou, Placide; Galani, Malatsi; Uwizeyemungu, Sylvestre; Ceric, Arnela (2021): The impacts of artificial intelligence (AI) on jobs: an industry perspective. In: *Strategic HR Review* 20 (2), S. 60–65.

Poster, Winifred R. (2011): Emotion detectors, answering machines, and e-unions: Multi-surveillances in the global interactive service industry. In: *American Behavioral Scientist* 55 (7), S. 868–901.

Poster, Winifred R. (2019): Racialized Surveillance in the Digital Ser-vice Economy. In: Ruha Benjamin (Hg.): Captivating Technology: Duke University Press.

Prietl, Bianca (2019): Algorithmische Entscheidungssysteme revisited: Wie Maschinen gesellschaftliche Herrschaftsverhältnisse reproduzieren können. In: *Feministische Studien* 37 (2), S. 303–319. DOI: 10.1515/fs-2019-0029.

Punia, Sanjeev Kumar; Kumar, Manoj; Stephan, Thompson; Deverajan, Ganesh Gopal; Patan, Rizwan (2021): Performance analysis of machine learning algorithms for big data classification: Ml and ai-based algorithms for big data analysis. In: *International Journal of E-Health and Medical Communications (IJEHMC)* 12 (4), S. 60–75.

Roberts, Dorothy (2019): Reimagining Race, Resistance, and Technoscience. In: Ruha Benjamin (Hg.): Captivating Technology: Duke University Press, S. 328–348.

Rubinov, Igor; Hagerty, Alexa (2019): Global AI Ethics: A Review of the Social Impacts and Ethical Implications of Artificial Intelligence. Online verfügbar unter https://arxiv.org/ftp/arxiv/papers/1907/1907.07892.pdf.

Scannell, R. Joshua (2019): This Is Not Minority Report. Predicitive policing and population racism. In: Ruha Benjamin (Hg.): Captivating Technology: Duke University Press, S. 107–130.

Schäffner, Wolfgang; Hörl, Erich; Krajewski, Markus; Chadarevian, Soraya de; Bredekamp, Horst; Ernst, Wolfgang et al. (2019): Rekursionen: Von Faltungen des Wissens. Leiden, Niederlande: Brill | Fink. Online verfügbar unter https://brill.com/view/title/40848.

Schelenz, Laura (2021): Schwarzfeministische Perspektiven auf Künstliche Intelligenz: Erkenntnisse und neue Fragen zu KI-gestützter Gesichtserkennung und Überwachung. In: *FemPol* 30 (2-2021), S. 73–93. DOI: 10.3224/feminapolitica.v30i2.07.

Schmidt, Francesca; Mellentin, Johanna Luise (2020): Überwachung und Künstliche Intelligenz. Wer überwacht hier eigentlich wen? In: netzforma* e.V. (Hg.): Wenn KI dann feministisch. Impulse aus Wissenschaft und Aktivismus, S. 15–30.

Schufa (2022): Scoring und Daten bei der Schufa. Hg. v. Schufa Holding AG. Online verfügbar unter https://www.schufa.de/scoring-daten/, zuletzt geprüft am 03.08.2022.

Schunicht, Barbara Elisabeth (2018): Informationelle Selbstbestimmung in sozialen Netzwerken: Mehrseitige Rechtsbeziehungen und arbeitsteilige Verantwortungsstrukturen als Herausforderung für das europäisierte Datenschutzrecht: Bucerius Law School.

Shankar, Shreya; Halpern, Yoni; Breck, Eric; Atwood, James; Wilson, Jimbo; Sculley, D. (2017): No classification without representation: Assessing geodiversity issues in open data sets for the developing world. In: *arXiv preprint arXiv:1711.08536*.

Siems, Jasper; Repka, Thomas (2021): Unrechtmäßige Nutzung von KI-Trainingsdaten als Gefahr für neue Geschäftsmodelle? Compliance-Anforderungen bei der Entwicklung von KI-Systemen. Herbstakademie. Deutsche Stiftung für Recht und Informatik, 2021. Online verfügbar unter https://dsri.de/storage/siems_repka.pdf, zuletzt geprüft am 26.02.2022.

Simon, Leena (2020): Kontrollverlust und (digitale) Entmündigung. Das Gewaltpotential Künstlicher Intelligenz. In: netzforma* e.V. (Hg.): Wenn KI dann feministisch. Impulse aus Wissenschaft und Aktivismus, S. 31–46.

Sommaggio, Paolo; Marchiori, Samuela (2020): Moral dilemmas in the AI era: A new approach. In: *Journal of Ethics and Legal Technologies* 2 (1).

Stanford University (2022): Artificial Intelligence Graduate Program. Artificial Intelligence: Principles and Artificial Intelligence: Principles and Techniques. Weitere Beteiligte: Andrew Ng, Christopher Manning, Chelsea Finn, Percy Liang und Jeanette Bohg. California: Stanford School of Engineering. Online verfügbar unter https://online.stanford.edu/programs/artificial-intelligence-graduate-program, zuletzt geprüft am 14.07.2022.

Sutton, Adam; Lansdall-Welfare, Thomas; Nello, Cristianini (2018): Biased embeddings from wild data: Measuring, understanding and removing. In: Wouter Duivesteijn, Arno Siebes und Antti Ukkonen (Hg.): Advances in Intelligent Data Analysis XVII (11191).

The MarkUp (2022): Big Tech Is Watching You. We're Watching Big Tech. Online verfügbar unter https://themarkup.org/, zuletzt geprüft am 14.08.2022.

Tuzcu, Pinar (2021): Decoding the cybaltern: cybercolonialism and postcolonial intellectuals in the digital age. In: *Postcolonial Studies* 24 (4), S. 514–527.

Unesco Courier (2018): Artificial intelligence: the promises and the threats (Vol.:3).

UNICRI (2020): Towards Responsible AI Innovation. Second Interpol - UNICRI.

United Nations (2020): First draft of the recommendation on the ethics of artificial intelligence. Ad Hoc Expert Group (AHEG) for the preparation of a draft text of a recommendation on the ethics of artificial intelligence.

Verbraucherzentrale (2014): Schufa und Co: Kredit-Scoring-Verfahren undurchsichtig. Hg. v. Verbraucherzentrale Bundesverband e.V. Berlin. Online verfügbar unter https://www.vzbv.de/pressemitteilungen/schufa-und-co-kredit-scoring-verfahren-undurchsichtig.

Villa, Paula-Irene (2020): Corona-Krise meets Care-Krise–Ist das systemrelevant? In: *Leviathan* 48 (3), S. 433–450.

Wagner, Ben; Lopez, Paola; Cech, Florian; Grill, Gabriel; Sekwenz, Marie-Therese (2020): Der AMS-Algorithmus. In: *zeitschrift für kritik, recht, gesellschaft* 2020 (2), S. 191–202.

Weber, Jutta; Prietl, Bianca (2021): AI in the Age of Technoscience. On the rise of data-driven AI and its episteme-ontological foundations. In: *Routledge Social Science Handbook of Artificial Intelligence*, S. 58–73.

Welt (2015): Wenn Ihr Wohnort Ihre Kreditwürdigkeit bestimmt. Online verfügbar unter https://www.welt.de/wirtschaft/article141192184/Wenn-Ihr-Wohnort-Ihre-Kreditwuerdigkeit-bestimmt.html.

Wenger, Thomas (2020): Digitale Geschäftsmodelle und asymmetrische Information in der Finanzierung von Unternehmen. In: Hans-Jörg Weiß (Hg.): Innovationen für eine digitale Wirtschaft: Springer, S. 173–197.

Wennker, Phil (2020): Künstliche Intelligenz in der Praxis. Wiesbaden: Springer Fachmedien Wiesbaden.

World Economic Forum (2020): A Framework for Responsible Limits on Facial Recognition Use Case: Flow Management. White Paper.

Zhang, Shunyuan; Mehta, Nitin; Singh, Param Vir; Srinivasan, Kannan (2021): Can an AI algorithm mitigate racial economic inequality? An analysis in the context of Airbnb. In: *An Analysis in the Context of Airbnb (January 21, 2021). Rotman School of Management Working Paper* (3770371).

Ziai, Aram; Ataç; Kraler, Albert; Schaffar, Wolfgang (Hg.) (2018): Politik und Peripherie. eine politikwissenschaftliche Einführung: mandelbaum.

Zillien, Nicole (2009): Digitale Ungleichheit. Neue Technologien und alte Ungleichheiten in der Informations- und Wissensgesellschaft. 2. Auflage. Wiesbaden: VS Verlag für Sozialwissenschaften.

Zou, James; Schiebinger, Londa (2018): AI Can Be Sexist and Racist – It's Time to Make it Fair. In: *Nature* (559), S. 324–326.

Zuboff, Shoshana (2019): Surveillance Capitalism and the Challenge of Collective Action. In: *New Labor Forum* 28 (1), S. 10–29. DOI: 10.1177/1095796018819461.

Zuckarelli, Joachim L. (2021): Was ist eine Programmiersprache? In: Joachim L. Zuckarelli (Hg.): Programmieren lernen mit Python und JavaScript: Springer, S. 25–32.

universität wien

Eidesstattliche Erklärung im Rahmen von schriftlichen Arbeiten

Angaben zur Studierenden

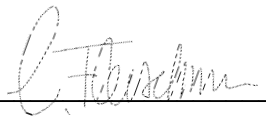
Matrikelnummer: 01617065
Zuname: Fleischmann
Vorname(n): Charlotte
Studienkennzahl: UA 066 589

Erklärung

Ich erkläre eidesstattlich, dass ich die Arbeit selbständig angefertigt, keine anderen als die angegebenen Hilfsmittel benutzt und alle aus ungedruckten Quellen, gedruckter Literatur oder aus dem Internet im Wortlaut oder im wesentlichen Inhalt übernommenen Formulierungen und Konzepte gemäß den Richtlinien wissenschaftlicher Arbeiten zitiert, durch Fußnoten gekennzeichnet bzw. mit genauer Quellenangabe kenntlich gemacht habe.

05.08.2022 _____

Datum



Unterschrift der / des Studierenden