

Zeitschriftenartikel*Begutachtet***Begutachtet:**

Dr. Steffen Rudolph 
HAW Hamburg
Deutschland

Erhalten: 27. Mai 2023**Akzeptiert:** 16. Juni 2023**Publiziert:** 29. Juni 2023**Copyright:**

© Annika Christina Trzensimiech.
Dieses Werk steht unter der Lizenz
Creative Commons Namens-
nennung 4.0 International (CC BY 4.0).

**Empfohlene Zitierung:**

TRZENSIMIECH, Annika Christina,
2023: Umfang und Reduktions-
möglichkeiten der Geschlechts-
diskriminierung in KI-gestützten
Auswahlprozessen. In:
API Magazin 4(2) [Online]
Verfügbar unter: [DOI 10.15460/
apimagazin.2023.4.2.153](https://doi.org/10.15460/apimagazin.2023.4.2.153)

Umfang und Reduktionsmöglichkeiten der Geschlechtsdiskriminierung in KI-gestützten Auswahlprozessen

Annika Christina Trzensimiech^{1*} ¹ Hochschule für Angewandte Wissenschaften Hamburg, Deutschland

Studentin im 2. Semester Digitale Transformation der Informations- und Medienwirtschaft

* Korrespondenz: redaktion-api@haw-hamburg.de

Zusammenfassung

Die vorliegende Arbeit untersucht den Zusammenhang zwischen KI-gestützten Auswahlprozessen und der Geschlechterdiskriminierung. Die Personalbeschaffung deckt den Personalbedarf eines Unternehmens durch Sichtung von Bewerbungen und Identifizierung qualifizierter Bewerber*innen. Unternehmen nutzen KI-Systeme, um Gender Bias zu beseitigen, aber es besteht die Beobachtung, dass Frauen trotzdem aufgrund ihres Geschlechts benachteiligt werden können. Es wird daher der Frage nachgegangen, inwieweit Frauen aufgrund eines Bias von KI-gestützten Auswahlverfahren in Bewerbungsprozessen diskriminiert werden. Das Ziel dieser Hausarbeit ist es, den Forschungsstand und die Anwendungsbereiche von KI-gestützten Methoden in Bewerbungsprozessen von Unternehmen zu untersuchen und anschließend auf potenzielle Gender Bias zu analysieren.

Schlagwörter: Künstliche Intelligenz, Bias, Auswahlprozesse, Bewerbungsprozesse

Extent and potential reduction strategies of gender discrimination in AI-assisted selection processes

Abstract

The present study examines the relationship between AI-assisted selection processes and gender discrimination. The process of personnel recruitment involves screening applications and identifying qualified candidates to meet a company's staffing needs. Companies use AI systems to eliminate gender bias, but there is an observation that women may still experience discrimination based on their gender. The research question addressed is: To what extent are women discriminated against in job application processes due to bias in AI-assisted selection procedures? The aim of this study is to investigate the current state of research and the application areas of AI-assisted methods in job application processes of companies, followed by an analysis of potential gender bias.

Keywords: Artificial Intelligence, Bias, Selection Processes, Application Processes, Recruiting

1 Einleitung

1.1 Problemstellung

Mit der Personalbeschaffung, auch unter dem englischen Begriff Recruiting bekannt, wird der Personalbedarf eines Unternehmens gedeckt. Dazu werden Bewerbungen gesichtet und in Zusammenarbeit mit der entsprechenden Fachabteilung qualifizierte Bewerber*innen identifiziert (vgl. [Buller 2021](#)). Bei der Auswahl der Bewerber*innen muss stets das Allgemeine Gleichbehandlungsgesetz (AGG)¹ beachtet werden:

Ziel des Gesetzes ist es, Benachteiligungen aus Gründen der Rasse oder wegen der ethnischen Herkunft, des Geschlechts, der Religion oder Weltanschauung, einer Behinderung, des Alters oder der sexuellen Identität zu verhindern oder zu beseitigen.

Laut der Antidiskriminierungsstelle des Bundes sind Frauen am häufigsten von Diskriminierung am Arbeitsplatz betroffen (vgl. [Antidiskriminierungsstelle des Bundes 2017](#)). Unterstützende Technologien kommen dabei bis dato selten zum Einsatz. Um diese Benachteiligungen zu überwinden und Gender Bias gegenüber Frauen zu beseitigen, nutzen jedoch manche Unternehmen KI-Systeme im Auswahlverfahren ihres Bewerbungsprozesses. Sie erhoffen sich, dass die digitalen Systeme nicht nur erheblich Zeit und Arbeit für die Recruiter*innen einsparen, sondern auch möglichst vorurteils- und verzerrungsfrei ihre Entscheidungen fällen. Jedoch ist zu beobachten, dass selbst KI-Systeme nicht frei von (Gender) Bias sind und Frauen aufgrund ihres Geschlechts anders oder benachteiligt behandeln. Daher stellt die Arbeit folgende Forschungsfrage: Inwieweit werden Frauen aufgrund eines Bias von KI-gestützten Auswahlverfahren in Bewerbungsprozessen diskriminiert?

1.2 Aufbau und Zielsetzung

Aufgrund der Komplexität der Thematik wird im ersten Schritt zunächst ein einheitliches Verständnis von künstlicher Intelligenz in Unternehmen geschaffen. Dazu wird eine Begriffserklärung und die Funktionsweise dargestellt. Des Weiteren wird untersucht, inwieweit künstliche Intelligenz im Recruiting eingesetzt werden kann. Im zweiten Teil der Arbeit liegt der Fokus auf den Herausforderungen, die sich angesichts der diversen Bias stellen und besonders den Unconscious Bias und den Gender Bias. Um ein Verständnis von Bias zu erhalten, wird der Ursprung und Hintergrund betrachtet. Dieser Schritt ist notwendig, um ein Grundverständnis von Voreingenommenheiten und Verzerrungen zu ermöglichen und die Wichtigkeit der Thematik zu unterstreichen. Basierend auf dieser Grundlage werden Bias betrachtet, die KI-Systeme über die Gender Bias hinaus besitzen können. Anschließend werden Diskriminierungen beim Einsatz von KIs betrachtet und diese im Anschluss diskutiert.

Ziel dieser Arbeit ist es, den aktuellen Forschungsstand und die Einsatzfelder von KI-gestützten Methoden von Unternehmen in Bewerbungsprozessen herauszuarbeiten und diese anschließend auf mögliche Gender Bias zu prüfen.

¹ Allgemeines Gleichbehandlungsgesetz vom 14.08.2006 (BGBl. I S. 1897), zuletzt geändert 19.12.2022.

2 Künstliche Intelligenz in Unternehmen

Vor allem in den letzten Jahren wurde der Begriff künstliche Intelligenz im privaten sowie wirtschaftlichen und wissenschaftlichen Kontext immer prominenter. Das Themengebiet und die technische Funktionsweise erweisen sich als sehr komplex und vielfältig, weshalb im folgenden Kapitel sowohl der Terminus als auch dessen Funktionsweisen definiert werden. Dabei liegt der Fokus auf dem Einsatz in Unternehmen, besonders im Recruiting.

2.1 Begriffserklärung und Funktionsweisen

Unter Künstliche Intelligenz (engl. Artificial Intelligence) können digitale Systeme verstanden werden, die „menschenähnlich“ mit intelligenten Fähigkeiten ausgestattet sind und sich eigenständig mit Problemen auseinandersetzen können (vgl. [Teetz 2018](#), S. 225). Bereits in den 1950er Jahren hat der britische Mathematiker Alan Turing wichtige Grundlagen für die KI-Entwicklung formuliert. Turing setzte sich in seinem Aufsatz mit dem Titel „Computing Machinery and Intelligence“ mit der Frage auseinander, inwieweit Maschinen denken können (vgl. [Turing 1950](#), S. 445 ff.). Seit dem Jahr 1956 forschen Wissenschaftler*innen im akademischen Kontext, wie denkende, interagierende Systeme erstellt und sinnvoll genutzt werden können (vgl. [Verhoeven 2019](#), S. 115).

Klassische lineare Algorithmen bearbeiten Probleme, indem sie bestimmte Handlungsanweisungen immer und immer wieder auf die gleiche Art und Weise durchführen. Im Vergleich zu einer KI kann ein linearer Algorithmus keine Schlussfolgerungen ziehen (vgl. [Teetz 2018](#), S. 225ff.).

Künstliche neuronale Netze (KNN) sind KI-Systeme, die sich an den Organisationsprinzipien und Abläufen biologischer neuronaler Netze orientieren und modellieren diese mit Hilfe von mathematischen Formeln. Hierbei hat ein Neuron die Aufgabe, Informationen oder Reize zu empfangen, zu verarbeiten und weiterzuleiten. Wie das biologische Vorbild haben KNN die Aufgabe, Informationen zu empfangen, zu verarbeiten und weiterzuleiten. KIs hingegen sind in der Lage zu lernen und logische Schlussfolgerungen zu ziehen (vgl. [ebd.](#), S. 227ff.). Es ergeben sich vier verschiedene Eigenschaften, die eine KI haben kann, neben dem Verstehen und dem Schlussfolgern kann diese auch Lernen und Interagieren. Eine Kombination aus diesen vier Eigenschaften unterstreicht die Besonderheit der KI (vgl. [Semet & Hilberer 2018](#), S. 183).

Basis jeder KI ist die menschliche Programmierung der Software sowie eine Datenbasis, die ebenfalls durch menschliche Einflüsse, wie etwa der Datenerhebung, geprägt ist. Die Emotionen, Werte, Ängste, aber auch Wissenslücken und Vorurteile eines Menschen werden bewusst oder unbewusst im Programmier-Prozess in die Lern- und Verarbeitungsmuster integriert. Als Ergebnis können Anwendungen ent-

stehen, die durch bestimmte sexistische oder rassistische Muster Personengruppen benachteiligen oder diskriminieren. Die Qualität der Programmierung und des Designs ist damit von den Trainingsdatensätzen abhängig. Spiegeln diese jedoch nicht die Wirklichkeit, sondern ein unrealistisches und verzerrtes Bild wider, so kann dies ein Grund für falsche Schlüsse eines KI-Systems sein (vgl. [Langer & Weyerer 2020](#), S. 224 ff.). Crawford drückt dieses Dilemma wie folgt aus: „*Like all technologies before it, artificial intelligence will reflect the values of its creators.*“ ([Crawford 2016](#), S. 11)

KIs funktionieren demnach auf Basis von Datensätzen und sich daraus ergebender Strukturen. Sie orientieren sich an menschlichen Entscheidungen und versuchen, diese nachzuvollziehen und auf zukünftige Entscheidungen anzuwenden (vgl. [Arel et al. 2010](#), S. 13 ff.). Die Funktionsweise kann damit in drei Phasen unterteilt werden. In der ersten Phase wird ein Datensatz aufgenommen, erkannt und übernommen. Aus diesem Grund kann die erste Phase unter dem Namen Erkennen (Input) zusammengefasst werden. Die zweite Phase ist die Phase des Verstehens (Logik), hier wird der Datensatz analysiert und auf Muster geprüft. Die letzte Phase ist das Produzieren (Aktion), hierbei werden abhängig von der Aufgabenstellung zuvor erkannte Muster entsprechend angewendet (vgl. [Lloyd 2018](#), S. 229). Abbildung 1 zeigt die drei Phasen des KI-Lernens und Verarbeitens. Zu betonen ist, dass in allen Phasen verzerrte Ergebnisse von KI-Anwendungen liegen könnten.

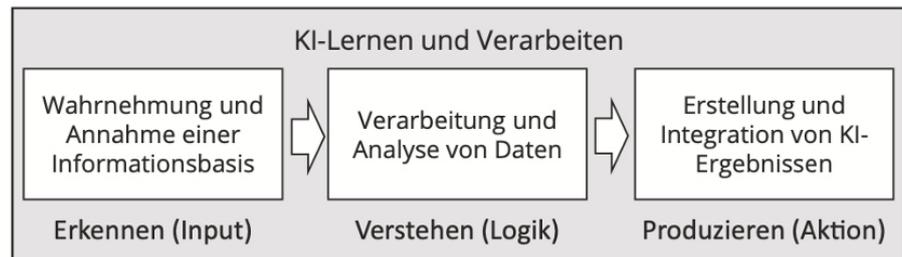


Abb. 1: Phasen des KI-Lernens ([Langer & Weyerer 2020](#), S. 230)

2.2 Das Black-Box-Problem

Eine Herausforderung beim Einsatz von KI-Systemen ist, dass selbstlernende Algorithmen meist wie eine Black Box arbeiten. In der Regel geben sie keine Auskunft darüber, wie sie Entscheidungen treffen und damit auf ein bestimmtes Ergebnis kommen (vgl. [Buxmann & Schmidt 2019](#), S. 15 f.). Daher ist es nicht ausgeschlossen, dass Entscheidungen in Abhängigkeit von Geschlecht, Hautfarbe oder Religion getroffen werden. Es werden simultan mehr Variablen verarbeitet, als klassische statische Verfahren dies können. Das stellt eine Herausforderung dar, bestimmte Korrelationen zu gestützten Merkmalen zu entdecken (vgl. [Orwat 2019](#), S. 82).

Die ermittelten Zusammenhänge sind für einen* eine Recruiter*in nicht unmittelbar ersichtlich (vgl. [Wolfangel 2018](#)). Besonders bei komplexen maschinellen Lernverfahren, die auf neuronalen Netzen basieren, ist die Wahrscheinlichkeit, dass selbst

der*die Programmierer*in jede Entscheidung der KI nachvollziehen kann, sehr gering (vgl. [Knobloch & Hustendt 2019](#), S. 13).

2.3 Recruiting mit Hilfe von künstlicher Intelligenz

Für den Einsatz künstlicher Intelligenz bieten sich im gesamten Personalbeschaffungsprozess vielfältige Möglichkeiten. Das Einsetzen von KI wird auch als „Robot-Recruiting“ bezeichnet, da die Beurteilung und Auswahl von Bewerber*innen teilweise anhand von Algorithmen stattfindet (vgl. [Schütz 2021](#)). Der Begriff „Data Driven Recruiting“ bezieht sich auf Recruiting-Prozesse, die nicht „vom reinen Bauchgefühl“, sondern von Daten und Fakten bestimmt werden (vgl. [Data Driven Recruiting 2021](#)).

Zur Vereinfachung wird in den folgenden Kapiteln der Begriff künstliche Intelligenz als Oberbegriff verwendet. Dabei wird zwischen vier Dimensionen unterschieden:

- Von Unternehmen zu Bewerbenden
KI-basierte Lösungen, welche die Adressierung von potenziellen Mitarbeitenden übernimmt.
- Prozesse der Bewerbenden
Durch Anwendungen kann beispielsweise die Erstellung von Bewerbungsunterlagen optimiert werden.
- Prozesse des Unternehmens
Sämtliche Lösungen, die das Recruiting aus Arbeitgebersicht optimieren. Dazu gehören z.B. Matching-Systeme oder automatisierte Vorauswahlen.
- Von Bewerbenden zu Unternehmen
Maßnahmen, die es den Bewerber*innen erleichtern, eine passende Stelle bzw. das richtige Unternehmen ausfindig zu machen.

Voraussetzung für den Einsatz einer KI bei einem Auswahlverfahren im Bewerbungsprozess ist, dass ein Trainingsdatensatz mit genügend Daten zur Verfügung gestellt wird. Hierbei ist die Qualität der Daten ausschlaggebend für das Ergebnis. Ziel des Einsatzes eines KI-Systems ist es, aus einer Auswahl von vielen Bewerber*innen, die ideale Person für die ausgeschriebene Stelle zu finden, wobei die KI eine Assistenzrolle einnehmen kann oder gar den kompletten Teilbereich des Bewerbenden-Screenings (vgl. [Teetz 2018](#), S. 237). Im Folgenden wird die Dimension der Prozesse in Unternehmen und der dortige Einsatz von KI-Systemen betrachtet und durch vier Möglichkeiten beschrieben.

Analyse der Bewerbungsunterlagen

Durch KI-gestützte Systeme können eingehende Bewerbungsunterlagen automatisch in qualitativ hochwertige Daten strukturiert werden. Dazu wählt die künstliche Intelligenz die relevanten Daten aus den Bewerbungsunterlagen aus. Hierzu können akademische Bildung, fachliche Erfahrungen sowie Soft- und Hard-Skills zählen.

Diese werden dann geordnet, strukturiert und vor allem anonymisiert dargestellt. In einem weiteren Schritt können eingehende Bewerbungsunterlagen automatisch durch öffentliche Daten, wie zum Beispiel Social-Media-Profile, analysiert und verglichen werden. Fachlich ungeeigneten Bewerber*innen kann direkt eine Absage erteilt werden, die KI übernimmt in diesem Fall eine aktive Rolle im Bewerbungsprozess. Alternativ kann eine künstliche Intelligenz auch eine assistierende Rolle einnehmen und einen Matching-Score ermitteln, um Recruiter*innen eine passende Vorauswahl bereitzustellen. In diesem Szenario dient die eingeleseene Stellenausschreibung oder aussagekräftige, manuell bestimmte Keywords als Entscheidungsgrundlage der KI. Zusätzlich können Daten aus erfolgreichen Bewerbungen oder von aktuellen Mitarbeiter*innen der Abteilung als Vergleichsdaten dienen (vgl. [Verhoeven 2019](#), S. 18 ff.). Studien zeigen, dass das automatische Scoring von Lebensläufen und Anschreiben durch eine KI verlässlich ist und valide Ergebnisse hervorbringt, die Menschen ähnlich hervorgebracht hätten (vgl. [Campion et al. 2016](#), S. 960).

Durchführung von Assessments

Assessments werden bei einer hohen Anzahl an potenziellen Kandidat*innen mit dem Ziel eingesetzt, die Bewerber*innen fachlich sowie kognitiv auf die Probe zu stellen und für den weiteren Bewerbungsprozess zu selektieren. Mithilfe der technologischen Unterstützung werden Reaktionen, Urteilsvermögen und Entscheidungsverhalten der Kandidat*innen geprüft. Die Antwortzeiten und Anzahl der Klicks geben Rückschlüsse auf das gewünschte Antwortverhalten. Zudem kann die KI ermitteln, inwieweit benötigtes Wissen für die entsprechende Stelle vorhanden ist. Die Ergebnisse der Assessments können dann entweder durch die KI oder einen*eine Recruiter*in genutzt werden, um das weitere Vorgehen mit einem Bewerber*innen einzuleiten (vgl. [Lochner & Preuß 2018](#), S. 193 ff.).

Erstellung von Persönlichkeitsprofilen

Bewerbungen mittels eingereicherter Videobotschaft oder klassische Video-Calls mit Recruiter*innen können durch eine künstliche Intelligenz, genauer eine Sprachanalyse-Software, ausgewertet werden. Dabei werden Mikroausdrücke, Sprache, Stimme sowie Schlüsselworte analysiert. Diese Daten werden anschließend mit Werten von leistungsstarken Mitarbeiter*innen verglichen. Außerdem können psychologische Merkmale mit den Anforderungen einer offenen Stelle und der Unternehmenskultur abgeglichen werden (vgl. [Adelmann & Wiedmer 2017](#)). Das System funktioniert sowohl mit gesprochenen Worten wie auch mit schriftlichen Texten. In der vereinfachten Variante können daher auch Bewerbungsunterlagen analysiert werden, wodurch ein erster Einblick in die Fähigkeiten allein auf Basis des Lebenslaufs erstellt wird. Auch Ergebnisse der im vorherigen Abschnitt beschriebenen Assessments können mit einbezogen werden und das Gesamtbild eines Kandidaten erweitern. Durch die umfassende Analyse aller vorhandenen Daten erstellt die KI ein Persönlichkeitsprofil (vgl. [Verhoeven 2019](#), S. 23 ff.).

Automatisierte Interviews und Einsatz von Robotern

Das russische Start-Up „Strafory“ entwickelte auf Russisch und Englisch eine Software mit dem Namen „Robot Vera“, welche bereits in mehr als 300 Unternehmen beim Recruiting eingesetzt wurde. „Vera“ prüft demnach eingehende Bewerbungen und ruft passende Kandidat*innen selbstständig an. Bei den ersten Informationsgesprächen wird nicht nur die unbesetzte Position vorgestellt, sondern es werden zugleich Fragen zum Unternehmen beantwortet. Mithilfe eines Avatars können zudem erste Videointerviews geführt werden. „Vera“ führt laut einem Mitgründer rund 50.000 Interviews am Tag und identifiziert dabei die besten 10% der Kandidat*innen bevor ein menschlicher Recruiter*innen die Bewerbungen sichtet (vgl. [Rottwilm 2018](#)).

Einen Schritt weiter geht die schwedische Personalvermittlung „The Next Generation“ (TNG). In Kooperation mit dem Tech-Unternehmen „Furhat Robotics“ wurde „Tengai“, ein androider Roboter, entwickelt. Die Besonderheit von „Tengai“ besteht darin, dass es sich nicht nur um eine Software, sondern um einen physischen Roboter und damit eine On-Premise-Lösung handelt. Durch diesen werden mit Hilfe von Natural Language Processing (NLP) schriftliche Bewerbungen anonymisiert, damit lediglich erforderliche Kompetenzen im Vordergrund stehen. Die anschließenden automatischen Roboterinterviews werden von „Tengai“ persönlich und vor Ort durchgeführt. Dafür wird er den Kandidat*innen gegenüber auf einem Tisch platziert. Durch eine integrierte Diversity and Inclusion Software schafft „Tengai“ ein innovatives Interviewerlebnis, welches fair und objektiv sein soll. „Tengai“ wird als sozialer KI-Roboter bezeichnet, da er nicht nur sprechen, zuhören und Aussagen bestätigen kann, sondern auch Emotionen zeigt. Zudem misst er Soft-Skills und persönliche Eigenschaften. Das strukturierte, 20-minütige Interview wird außerdem verschriftlicht. So wird eine Vorauswahl an geeigneten Bewerber*innen getroffen, an welche die Recruiter*innen anknüpfen können (vgl. [Tengai Robot 2021](#)).

3 Diskriminierung durch Bias

Der Ursprung von Diskriminierungen gegenüber Frauen im Berufsalltag sind häufig Verzerrungen, Voreingenommenheiten und Stereotypisierungen. Im folgenden Kapitel wird auf die verschiedenen Arten von Bias eingegangen und wie diese zu einer Diskriminierung von Frauen beitragen. Im Fokus steht hierbei der Gender Bias. Der Vollständigkeit halber werden auch weitere Bias, die eng mit dem Gender Bias in Verbindung stehen oder häufig bei KI-gestützten Auswahlverfahren in Bewerbungsprozessen auftreten, genannt.

3.1 Ursprung: Unconscious Bias

Der englischsprachige Begriff „bias“ beschreibt ein Vorurteil oder einen entsprechenden Denkfehler. Unter „Unconscious Bias“ kann ein unbewusstes Vorurteil verstanden werden. Die unbewusste Voreingenommenheit wird als kognitive Wahrneh-

mungsverzerrung definiert (vgl. [Domsch et al. 2019](#), S. 6). Vergleichbar mit Stereotypen stellen Unconscious Bias mentale Abkürzungen dar, welche als schnelle Assoziationen auf frühere Erfahrungen zurückzuführen sind. Die Verknüpfungen beeinflussen jede individuelle Entscheidung auf einer emotionalen Ebene (vgl. [Roxburgh & Hansen 2016](#)). Unbewusste Vorurteile werden bereits in jungen Jahren gebildet. So bilden Kinder ab einem Alter von drei Jahren bestimmte Voreingenommenheiten, welche durch Erfahrungen und die Erziehung verstärkt und geprägt werden. Viele Faktoren spielen hierbei eine Rolle, zum Beispiel werden Assoziationen und Emotionen mit individuellen und persönlichen Erfahrungen entweder positiv oder negativ verknüpft. So werden etwa unbewusst Menschen aus dem eigenen Kulturkreis präferiert, da mit diesen bereits Erfahrungen verknüpft wurden (vgl. [Wondrak 2020](#)).

Kahneman ([2011](#)) weist jedoch darauf hin, dass Unconscious Bias keine psychologischen Zwänge sind, sondern biologische Prozesse, die anhand seines Zwei-System-Modells erklärt werden können. Das Modell basiert auf zwei Systemen, die eng miteinander verknüpft sind. Das erste System wird als das unbewusste System beschrieben, welches intuitiv und instinktiv agiert. Die Fähigkeiten, die mit diesem ersten System verbunden sind, sind entweder angeboren oder durch häufiges Wiederholen zu automatischen Routinen geworden. Das erste System greift demnach auf gespeicherte Erfahrungen zurück. Tägliche Wahrnehmungs- und Denkprozesse laufen mit mehr als 90% als Teil des unbewussten Systems ab.

Das zweite System ist das bewusste System, in das das langsame und logische Denken wie auch bewusste Kontrolle und Entscheidungen eingeordnet werden. Als Basis dienen die Eindrücke und Emotionen aus dem ersten System. Hierbei wird die Aufmerksamkeit willentlich auf eine mental anstrengende Aktivität gelenkt, wie zum Beispiel das aufmerksame Zuhören während eines Gesprächs. Da das bewusste Denken eher als anstrengend empfunden wird, nutzen Menschen bevorzugt das unbewusste System, in das auch der Unconscious Bias eingeordnet ist. Unbewusste Vorurteile sind demnach tief in der Biologie des Menschen verankert, auch wenn sie fehlerhaft sind oder entgegen bewusster Überzeugungen und Werten ablaufen. In Bezug auf die Berufswelt treten, über das Recruiting hinaus, Verzerrungen in Teams auf aber auch in implizierten Normen oder Machtvorstellungen im Zusammenhang mit Rollenbildern (vgl. [Wondrak 2020](#)). Folglich sind Expert*innen davon überzeugt, dass rund 80% der Einstellungsentscheidungen von Unconscious Bias beeinflusst werden (vgl. [Nalty 2016](#)).

3.2 Gender-Bias

Der Gender Bias beschreibt die systematischen Verzerrungseffekte, die aufgrund von geschlechtsbezogenen Stereotypisierungen und Vorurteilen getroffen werden. Gender Bias treten nicht nur im Alltag auf, sondern sind auch in Wissenschaft und Forschung wiederzufinden (vgl. [Nalty 2016](#)). Das Bureau of Employers' Activities beschreibt einen unbewussten Gender Bias (unconscious bias) wie folgt: „[...] unab-

sichtige und automatische mentale Assoziation basierend auf dem Gender, die von Traditionen, Normen, Werten, Kultur und/ oder Erfahrungen herrührt“ ([International Labour Organization 2017](#), S. 3).

Resultat des Gender Bias sind Benachteiligungen oder Bevorzugung bestimmter Personen aufgrund ihres Geschlechts. Eichler et al. ([2002](#)) unterscheidet drei Arten des Gender Bias. Die erste Art des Gender Bias wird als Androzentrismus definiert. Hierbei wird die Sicht aus einer männlichen Perspektive als Norm gehandhabt. Dies kann einen Ausschluss oder Unterrepräsentanz von Frauen und weiblichen Perspektiven verursachen. Eine Sonderform ist der paradoxe Gynozentrismus, bei dem Männer aus Bereichen oder Branchen ausgeschlossen werden, weil diese als traditionell weiblich angesehen werden (vgl. [Draude et al. 2019](#), S. 224 ff.).

Die zweite Form von Gender Bias kann als Geschlechtsinsensibilität oder Geschlechtsneutralität definiert werden. Das Geschlecht, das als bedeutsame Variable genutzt werden müsste, wird in dem dafür relevanten Kontext nicht wahrgenommen oder ignoriert. Das bedeutet, dass in diesen Fällen eine Unterscheidung der Geschlechter sinnvoll wäre, aufgrund von zum Beispiel körperlichen Attributen. Die dritte Form des Gender Bias ist der doppelte Bewertungsmaßstab. Dieser kann entweder offenkundig oder versteckt auftreten und liegt vor, wenn eine Situation substantiell identisch ist, aber die Geschlechter unterschiedlich behandelt oder untersucht werden. Es werden demnach fälschlicherweise Annahmen basierend auf gleichen Verhaltensweisen unterschiedlichen interpretiert (vgl. [Eichler et al. 2002](#)). Als Beispiel kann hierfür die Geburt eines Kindes herangezogen werden. Konzentriert sich eine Frau nach der Geburt ihres Kindes auf die Erziehung, so kann dies so ausgelegt werden, dass sie die Familie über die Karriere stellt, was häufig einer negativen Zuschreibung gleichkommt. Befindet sich ein Mann in der gleichen Situation, so wird dieser als liebevoller Familienvater beschrieben und ihm werden positive Eigenschaften zugeschrieben. Der Gender Bias steht darüber hinaus eng mit der Bestätigungsvoreingenommenheit und der Attributionsvoreingenommenheit in Beziehung:

Bestätigungsvoreingenommenheit (Confirmation Bias)

Die Bestätigungsvoreingenommenheit beschreibt, wie Menschen gezielt auf der Suche nach Meinungen oder Denkmustern sind, die der eigenen entsprechen. Basis dieser Voreingenommenheit sind Stereotype. Es werden Beweise herangezogen, die die eigene Meinung stützen. Oftmals werden durch die selektive Begutachtung vermeintlich unwichtige Aspekte wie beispielsweise nützliche Skills und Informationen außer Acht gelassen und übersehen.

***Beispiel:** Echo Chambers oder Filter-Bubbles (Nachrichtenströme, die die bereits bestehende Meinung durch ähnliche Artikel oder Neuigkeiten bestätigen) verstärken im virtuellen Raum eine Bestätigungsvoreingenommenheit, indem durch einen bestimmten Algorithmus die eigene Meinung bekräftigt wird (vgl. [Anti-Bias 2022](#)).*

Attributionsvoreingenommenheit (Attribution Bias)

Bei der Attributionsvoreingenommenheit werden die Handlungen von anderen Personen und der eigenen Person verzerrt wahrgenommen. Es werden die Gründe und Intentionen von Handlungen oft fehlerhaft wahrgenommen und interpretiert. Erfolg wird auf die eigene Leistung und Fähigkeiten zurückgeführt, während Misserfolg externen Faktoren zugeschrieben wird. In einer Entscheidungssituation kann das Konzentrieren auf Fehler von der eigentlichen Leistung und Talenten ablenken.

Beispiel: Ein*e Bewerber*in verspricht sich aufgrund von Nervosität zu Beginn des Bewerbungsgesprächs. Der*Die Entscheidungsträger:in misst diesem Verhalten unverhältnismäßig viel Gewicht bei und führt es auf die Fähigkeiten und Kompetenzen des Bewerbenden zurück (vgl. [ebd. 2022](#)).

3.3 Bias in KIs

Ergebnisse oder Auswertungen von Computern werden als besonders objektiv angesehen und sollen laut Porter von Alltagsrassismus und -sexismus bestenfalls nicht betroffen sein: „*The ideal calculator is a computer, widely revered in part because it is incapable of subjectivity. [...] the desires and biases of individuals are screened out.*“ (Porter 1996, S. 74)

Die vermeintlich neutrale Instanz der Technik kann jedoch Diskriminierungseffekte durch geringe Nachvollziehbarkeit entstehen lassen und bestätigen. KI lässt Diskriminierungseffekte entstehen und KI bestätigt bestehende Diskriminierungseffekte bei Menschen. Sexistisch oder rassistisch eingestellte Recruiter*innen können ihre Vorurteile, wie zum Beispiel, dass Frauen oder Minderheiten eine mangelnde Kompetenz besitzen, von einem KI-Recruiting-Tool bestätigt sehen, wenn dieses seltener Frauen oder Migranten vorschlägt (vgl. [Dastin 2018](#)). Abbildung 2 macht deutlich, wie Bias in KI-Systemen entstehen. Wenn die KI in ihren Inputdaten mit menschlichen Bias trainiert wird, so verarbeitet und übernimmt sie diese. Aus den voreingenommenen Verarbeitungen entstehen damit Bias durch die KI, die sich in den Ergebnissen widerspiegeln. Die neuen Ergebnisse, wie zum Beispiel eine erfolgreiche Bewerbungssituation, werden dann wiederum als Inputdaten genutzt, um weitere Bewerber*innen zu akquirieren. Es entsteht somit ein Kreislauf, der durch Bias gestützt wird (vgl. [Langer & Weyerer 2020](#), S. 224).

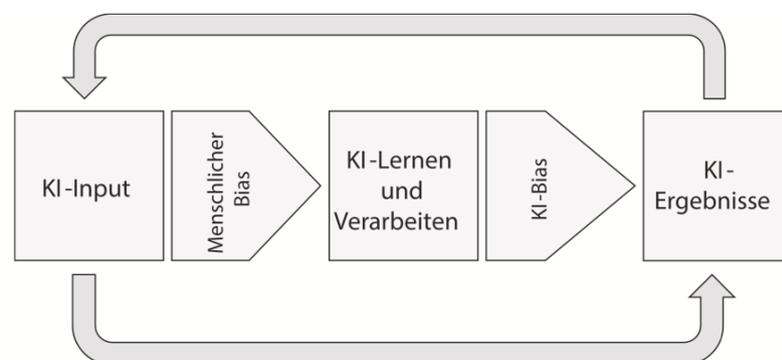


Abb. 2: KI-basierte Reproduktion von Vorurteilen und diskriminierendem Verhalten ([Langer & Weyerer 2020](#), S. 224)

Beim Einsatz von KI-Systemen können verschiedene Bias für individuelle Wahrnehmungsmuster von Bewerber*innen im Recruiting verantwortlich sein:

Datensatzverzerrung (Data Bias)

Die Datensatzverzerrung tritt auf, wenn in einem Datensatz die Gesellschaft nicht ausreichend repräsentiert wird. Hierbei kann es zu verzerrten Verallgemeinerungen kommen (z.B. Selektionsverzerrung durch das Geschlecht).

Beispiel: Wenn Internetdaten (zum Beispiel: Online-Zeitungen) herangezogen werden, sind diese oft nicht geschlechtsneutral, da Männer häufiger in Inhaltsbeiträgen repräsentiert werden als Frauen. Aus der Verzerrung könnte eine KI den Schluss ziehen, dass Frauen im Vergleich einen geringeren Beitrag leisten können (vgl. [Yong 2018](#), S. 203f.).

Assoziationsverzerrung (Association Bias)

Bei der Assoziationsverzerrung werden Verzerrungen in Trainingsdaten für KI-Systeme erzeugt, die nicht auf kausalen Fakten basieren. Zusammenhänge und ursächliche Wirkungsbeziehungen zwischen Merkmalen, Ereignissen und Zuständen werden geschlossen, ohne dass diese direkt gegeben sind.

Beispiel: Hohe Durchschnittsgehälter von männlichen Mitarbeitenden lassen nicht direkt auf eine höhere Leistungsstärke im Vergleich zu Frauen schließen (vgl. [Dastin 2018](#)).

Automatisierungsverzerrung (Automation Bias)

Ein Automatisierungsverzerrung tritt bei teilautonomen Systemen auf, die falsche oder unerwünschte Ergebnisse produzieren und nicht durch einen Menschen zusätzlich kontrolliert werden. Eine letzte Prüfung durch eine menschliche Entscheidungsinstanz ergänzt die Arbeitsweise durch eine soziale, moralische und kulturelle Prüfung und Korrektur. Wird die Kontrolle durch einen Menschen vernachlässigt, so können Verzerrungen entstehen, die bestimmte Minderheiten diskriminieren.

Beispiel: Die Foto-App von Google kennzeichnete Bilder automatisch mit verschiedenen Labels, um sie zu kategorisieren. Dabei trat 2015 das Problem auf, dass afrikanisch-stämmige Menschen mit dem Wort Gorilla klassifiziert wurden. Da eine technische Lösung seitens Google nicht gefunden werden konnte, musste das Wort Gorilla als Label entfernt werden (vgl. [Kasperkevic 2015](#)).

Interaktionsverzerrung (Interaction Bias)

Bei einer Interaktionsverzerrung handelt es sich um eine Verzerrung, die durch die nicht vorhandene moralische und soziale Kompetenz einer KI entsteht. So lernt ein KI-System auf Basis von Kommunikationsdaten menschengemachte Muster und leitet diese ab.

Beispiel: Der selbstlernende Twitter-Chatbot „Tay“ von Microsoft orientierte sich bei seinen Veröffentlichungen an den Tweets, die die meisten Reaktionen und Interaktionen auf der Plattform auslösten. Innerhalb weniger Stunden wurde „Tay“ offline genommen, da dieser antisemitische, rassistische und sexistische Inhalte teilte (vgl. [Lloyd 2018](#), S. 228).

4 Diskriminierungen beim Einsatz von KIs

Die Diskriminierung durch KI kann verschiedene Formen annehmen, wie zum Beispiel die Verunglimpfung von Minderheiten oder benachteiligten Personengruppen in Sprache oder im visuellen Kontext. Hierbei wird von einer Ehrverletzung, Herabsetzung oder Hetze (hate speech) gesprochen (vgl. [Davidson et al. 2017](#), S. 512f.). Darüber hinaus kann bei der Diskriminierung durch den Einsatz einer KI von direkter oder indirekter Diskriminierung gesprochen werden. Bei der direkten Diskriminierung ist eine direkte Form von Diskriminierung zu beobachten, die eine unmittelbare Form der Benachteiligung beschreibt, wie zum Beispiel im Beruf, vor Gericht oder bei Polizeieinsätzen. Hier steht ein bewertungsunabhängiges persönliches Merkmal, wie das Geschlecht oder die Ethnie im Mittelpunkt (vgl. [Dalenberg 2018](#), S. 616 ff. und [Ellis & Watson 2012](#), S. 143). Diese Art der Diskriminierung bei der KI-Anwendung wurde zum Beispiel beim Bewerbungsmanagement-Algorithmus von Amazon im Jahr 2018 beobachtet. Im beschriebenen Fall wurden Rankings von eingehenden Bewerbungen erstellt, die im Anschluss anhand bestimmter Merkmale und Schlüsselwörter untersucht wurden. Die KI wurde mit den Daten und Informationen trainiert, die in den letzten zehn Jahren zu erfolgreichen Neueinstellungen, die überwiegend Männer waren, geführt hatten und verglich diese mit allen eingehenden Bewerbungen und stellte so Zusammenhänge her. Das System zog aus den Daten anschließend den Schluss, dass Männer sich häufiger und erfolgreicher bei Amazon beworben haben und schloss somit Frauen aus dem Bewerbungsverfahren aus (vgl. [Wilke 2018](#)).

Bei der indirekten Diskriminierung ist eine ungerechte Behandlung aller zu beobachten, die jedoch nur negative Auswirkungen auf eine bestimmte Gruppe hat (vgl. [Ellis & Watson 2012](#), S. 148). Ein Beispiel für die indirekte Diskriminierung durch den Einsatz von KI kann das Priorisieren von bestimmten Merkmalen im Bewerbungsprozess sein. Orientiert sich eine KI bei der Bewerbendenauswahl im Lebenslauf an bestimmten Schulen oder Universitäten, entsteht eine Benachteiligung, zum Beispiel für Quereinsteiger.

Eine weitere Form der Diskriminierung ist die intersektionale Diskriminierung. Hierbei wird aufgrund von individuellen Merkmalen unterschieden, die nicht getrennt auftreten, sondern in Zusammenhang miteinander stehen, wie zum Beispiel das Geschlecht in Verbindung mit der ethnischen Herkunft oder der sexuellen Orientierung. Diese Art der Diskriminierung tritt zum Beispiel bei Frauen mit Migrationshintergrund auf. Diese werden benachteiligt, obwohl in separaten Messungen weder Benachteiligungen gegenüber Frauen noch Personengruppen mit Migrationshintergrund zu erkennen waren. In KI-Anwendungen ist die intersektionale Diskriminierung zu beobachten, wenn KI-Systeme darauf programmiert sind, besonders Minderheiten zu differenzieren (vgl. [Raji & Buolamwini 2019](#), S. 5f.).

Lineare Algorithmen lassen sich anhand des Codings häufig nachverfolgen. Beim Deep Learning oder neuronalen Netzen ist die Nachverfolgung von computerbasierten Entscheidungsprozessen nicht immer nachvollziehbar. Daher werden Entscheidungsprozesse solcher Algorithmen oftmals als Blackbox beschrieben (vgl. [Castelvecchi 2016](#), S. 21 ff.).

5 Diskussion

Im Rahmen dieser Hausarbeit wurden verschiedene Arten der Diskriminierung beim Einsatz von KI-Systemen ausgearbeitet. Um die Forschungsfrage beantworten zu können, wird im Folgenden untersucht, inwieweit Frauen aufgrund eines Gender Bias in Bewerbungsprozessen von KI-Systemen diskriminiert werden. Laut Kahneman ([2011](#)) handelt es sich bei Bias um biologische Abläufe, die tief in jedem Menschen verankert sind. Unconscious Bias sind demnach schwer zu erkennen oder nachzuweisen. Dies stellt generell eine große Herausforderung im Umgang mit Bewerber*innen im Bewerbungsprozess dar und kann auf den Einsatz eines KI-Systems übertragen werden. Wie Crawford feststellt, reflektieren KIs die Ansichten ihres Programmierenden oder Recruiter*innen: „Like all technologies before it, artificial intelligence will reflect the values of its creators.“ ([Crawford 2016](#), S. 11). Auch die Emotionen, Ängste, Wissenslücken oder Vorurteile können bewusst oder unbewusst in den Lern- und Verbreitungsmustern eingebaut werden und zu Gender Bias im Prozess führen (vgl. [Langer & Weyerer 2020](#), S. 124 ff.). Es können laut Langer und Weyerer ([2020](#)) in jeder Phase des KI-Lernens und -verarbeitens (Gender) Bias eingearbeitet werden, die zur Benachteiligung von Frauen führen können. So können bereits Inputdaten mit unvollständigen oder verzerrten Wirklichkeiten darstellen oder verzerrende Algorithmen bei der Verarbeitung und Analyse der Daten vorkommen.

Als Beispiel werden die vorgestellten Möglichkeiten eines Einsatzes von KI-Systemen im Recruiting-Prozess auf ihr Risiko eines Gender Bias untersucht:

Analyse der Bewerbungsunterlagen

Der Vorteil an der Analyse von Bewerbungsunterlagen liegt darin, dass Bewerber*innen anonymisiert dargestellt werden. Das bedeutet, dass auch das Geschlecht in den Hintergrund tritt und anhand der Qualifikationen und Erfahrungen gemessen wird, ob ein*e Bewerber*in zur ausgeschriebenen Stelle passt. Mit diesem Schritt ist die KI dem Menschen voraus, da (unbewusst) voreingenommene Recruiter*innen unbewusst zwischen der Bewerbung einer Frau und der eines Mannes unterscheiden könnten, sodass sie trotz gleicher Qualifikationen unterschiedlich bewertet werden und somit ein Gender Bias vorliegt. Die Analyse der Bewerbungsunterlagen durch ein KI-System birgt jedoch auch einige Risiken für einen Gender Bias. Die KI entscheidet auf Grundlage der Inputdaten und der Programmierung, wie sie mit den Daten umgehen soll. Anhand von Keyword-Scans werden Matching-Scores aufge-

stellt und fachlich ungeeigneten Bewerber*innen direkt eine Absage erteilt. Hier stellt sich jedoch die Frage, wie unvoreingenommen Keywords ausgesucht und geschlechtliche Unterschiede berücksichtigt werden (Geschlechtsinsensibilität). Darüber hinaus besteht das Risiko, dass Recruiter*innen unter der ersten Form von Gender Bias nach Eichler leiden, dem Androzentrismus, und Stellenanzeigen oder Keywords aus der männlichen Perspektive aufstellen. So können traditionell männliche Attribute als Voraussetzung für die ausgeschriebene Stelle formuliert werden, während bei traditionell weiblichen Merkmalen kein Matching entsteht, obwohl diese aus einer weiblichen Sicht genauso zu dem Job passen würden. Hinzukommt, dass Frauen in diesem Kontext sich selbst mit einem anderen Wording beschreiben würden. Stimmt das Wording jedoch mit den zuvor festgelegten Keywords nicht überein, wird ihnen schlimmstenfalls direkt eine Absage erteilt. Ein weiterer Aspekt wären die Lebensläufe von Frauen, die sich nach der Geburt und Erziehung eines Kindes bewerben. Werden anonym Lebensläufe verglichen und die geschlechtlichen Unterschiede, dass nur eine Frau in der Lage ist ein Kind zu gebären, vernachlässigt, so profitieren Männer mit der gleichen Qualifikation schlichtweg durch mehr Berufserfahrung, da sie keine Auszeit trotz Kinderplanung nehmen mussten. Darüber hinaus kann ein Gender Bias entstehen, wenn Bewerbungsunterlagen von bisherigen Mitarbeitenden als Inputdaten genutzt werden, diese aber nicht vollständig oder verzerrt sind, wie zum Beispiel im oben geschilderten Fall von Amazon. Außerdem werden durch diese Methode Bewerber*innen mit unkonventionellen Qualifikationen und Erfahrungen niedriger gerankt und scheiden somit aus. Das Risiko, dass aus der Belegschaft in einem Unternehmen eine homogene Arbeitnehmerschaft entsteht, wird dadurch erhöht.

Durchführung von Assessments

Bei der Durchführung von Assessments durch KI-Systeme können geschlechtsunabhängig Qualifikationen und Wissen unter Beweis gestellt werden, ohne dass Frauen bestimmte Kenntnisse, die als traditionell männlich angesehen werden, abgelehnt werden. Eine Herausforderung bei dieser Methode stellen erneut die Inputdaten und Bewertungsschemata der KI dar. Sind diese von den Recruiter*innen und Programmierer*innen unter Einfluss eines Gender Bias getroffen worden, so werden die Ergebnisse und Entscheidungen der KI ebenfalls unter einer Verzerrung durch einen Gender Bias leiden.

Erstellung von Persönlichkeitsprofilen

Durch das Erstellen eines Persönlichkeitsprofils von Bewerber*innen können positive Merkmale, die in einer üblichen Bewerbungsgesprächssituation nicht genügend Aufmerksamkeit bekämen, berücksichtigt werden. Nachteil des Einsatzes einer KI ist, dass häufig Gesichts- und Spracherkennungssysteme weibliche Stimmen und Mimik schwerer erkennen können, da diesbezügliche Trainingsdaten nicht ausreichend zur Verfügung stehen. Aus diesem Grund besteht das Risiko, dass aufgrund von mangelnden Inputdaten Frauen im Bewerbungsprozess und bei der Erstellung von Per-

sönlichkeitsprofilen benachteiligt werden. Ein besonderes Risiko besteht für Frauen, die einen Migrationshintergrund haben und bei der Gesichtserkennung anders wahrgenommen und aus diesem Grund unterschiedlich bewertet werden. Es besteht das Risiko eines Gender Bias in Form einer intersektionalen Diskriminierung.

Automatisierte Interviews

Der Vorteil von automatisierten Bewerbungsprozessen ist nicht nur, dass Recruiter*innen, wie bei den vorherigen Methoden, viel Zeit und Arbeitsaufwand bei der Einsicht von Bewerbungsunterlagen sparen, sondern, dass durch die ersten Informationstelefonate Bewerber*innen, die kein Interesse an der Stelle oder dem Unternehmen haben, direkt ausscheiden können. Darüber hinaus kann mit der vorgestellten Software „Vera“ eine Vielzahl an Gesprächen parallel geführt werden, was durch Recruiter*innen persönlich nicht geleistet werden könnte. Der*Die Recruiter*in setzt als menschliche Instanz damit weiter hinten im Bewerbungsprozess wieder ein. Nachteil und Gefahr ist, wie auch bei der Analyse von Bewerbungsunterlagen durch KI-Systeme, dass sich in sämtlichen Programmierungsphasen Gender Bias einschleichen können. Zudem besteht das Risiko, dass aufgrund mangelnder Trainingsdaten die KI die Frauenstimmen im Informationsgespräch aufgrund eines Androzentrismus anders einstuft, verzerrt wahrnimmt oder bewertet.

Einsatz von Robotern

Die schwedischen Entwickler des androiden Roboters „Tengai“ versprechen selbst, dass ihr Roboter besonders fair und objektiv sei. Vergleichbar mit der Analyse von Bewerbungsunterlagen durch KI-Systeme anonymisiert „Tengai“ die Bewerbungen, um das Risiko von Diskriminierungen zu verringern. Darüber hinaus verfügt „Tengai“ über eine integrierte Diversity und Inclusion Software, die besonders auf die Vermeidung von Gender Bias achten soll. Im Vergleich zu „Vera“ verschriftlicht „Tengai“ die Interviewgespräche mit den Bewerber*innen und bietet so eine Möglichkeit der Nachverfolgung der Entscheidungen, die zusätzlich von einem*einer Recruiter*in geprüft werden können. Ein Risiko für einen Gender Bias stellt die Genauigkeit und Verlässlichkeit der Diversity und Inclusion Software dar. Ist diese durch verschiedene Bias verzerrt, so kann sie ihre Bias an „Tengai“ weitergeben und diese können in den Bewerbungsprozess gelangen.

In allen Verfahren besteht darüber hinaus grundsätzlich ein Risiko auf einen Gender Bias aufgrund des Black-Box-Problems. Je weniger ein System nachzuvollziehen ist, desto wahrscheinlicher kann ein Bias unbemerkt auftreten.

6 Lösungswege

Um einen (Gender-) Bias zu überwinden, müssen verschiedene Maßnahmen in Betracht gezogen werden. Verschiedene Unternehmen und Organisationen befassen sich mit einer Lösung für dieses Problem. Die Unternehmensberatung „Accenture“ bietet einen KI-Test-Service mit einer Teach- und Test-Methodik an, um Unternehmen KI-Systeme ohne Vorurteile und diskriminierende Inhalte bereitzustellen (vgl. [Accenture 2018](#)). Auch Google verfolgt einen Forschungsansatz namens „Testing with concept activation vectors“ (TCAV). Dieser soll Techniker*innen zeigen, wie eine KI Entscheidungen trifft (vgl. [Google 2019](#)). Mit dem „AI fairness 360 Toolkit“ von IBM wurde ein Open-Source-Projekt geschaffen, um weltweit Forschungs- und Entwicklungsgemeinschaften zu fördern. Dabei handelt es sich um einen Satz von über 70 Fairness-Metriken, die zur Überprüfung und Vermeidung von Verzerrungen in KI-Systemen und Algorithmen genutzt werden können (vgl. [IBM 2018](#)). Krüger und Lischka ([2018](#)) behandeln das Problem auf konzeptioneller Seite: Sie prüfen und messen die KI-Systeme hin auf ihre gesellschaftliche Angemessenheit und ihre Wirkungseffekte, um Diversität in algorithmischen Systemen sicher zu stellen. Die größte Herausforderung bei der Beseitigung von Bias liegt in der Nachvollziehbarkeit, da nicht sichergestellt werden kann, ob eine Entscheidung ethisch legitim getroffen wurde (Black-Box-Herausforderung).

Eine weitere Maßnahme, um Gender Bias vorzubeugen, ist die Einführung einer Diversity-Management-Abteilung oder einer einzelnen Person, die für das Diversity-Management verantwortlich ist. Das Konzept des Diversity Managements hat seinen Ursprung in den USA der 1990er Jahre, wo es von Organisationsberater*innen als Antwort auf den demografischen Wandel entwickelt wurde. Die Einführung und Umsetzung des Diversity-Managements (DiM) lässt sich anhand fünf übergeordneter Leitfragen charakterisieren. An erster Stelle steht der Nutzen, ein Bewusstsein für die Chancen von Diversität in Hinblick auf Stakeholder zu schaffen (vgl. [Charta der Vielfalt 2022](#)). Das Ziel ist es, die Produktivitätseffekte durch die gezielte Nutzung von persönlichen Kompetenzen zu steigern und den Abbau von Vorurteilen, Stereotypen und Verzerrungen zu fördern. Zentrale Werte wie Toleranz, Fairness und Chancengleichheit sollen mit Hilfe von DiM transportiert werden. Um die besten Ergebnisse aus der Vielfalt der Diversity zu ziehen und für die Unternehmensziele nutzen zu können, müssen diese koordiniert und strukturiert werden. Das DiM bietet in seiner Umsetzung eine Vielzahl an Instrumenten aus den Bereichen Personalbeschaffung, -entwicklung und -beurteilung, etwa in Form von Diversity-Trainings, Sprachkursen oder Mentoringprogrammen (vgl. [Vedder 2005](#), S. 35 ff.).

7 Fazit

Vorurteilsfreies Recruiting kann nur dann funktionieren, wenn sich Recruiter*innen ihren Unconscious Bias bewusstwerden und aktiv versuchen, diese zu minimieren. Daher sind die aufgezeigten bisherigen Maßnahmen gegen den Unconscious Bias nicht zu vernachlässigen und sollten kontinuierlich durchgeführt werden. Erst dann kann eine künstliche Intelligenz eine sinnvolle und vorurteilsfreie Ergänzung für einen bewussten Bewerbendenumgang darstellen. Zudem ist vor allem der Einsatz der KI in einer ersten Vorauswahl sehr sinnvoll, wobei hier ebenfalls die Unterlagen der Bewerbenden gegengeprüft werden sollten, bevor Absagen versendet werden. Wesentliche Erkenntnisse aus der Arbeit sind, dass der Ursprung jeder KI ein Mensch ist, der zwar das Ziel hat, ein KI-System verzerrungs- und vorurteilsfrei zu kreieren, jedoch werden dennoch die menschlichen Bias einbezogen.

Das Wahrnehmen, Erkennen und Beseitigen von Bias und besonders Unconscious Bias kann nicht erst in einem KI-System anfangen, sondern muss bereits beim Programmieren und Erheben der Trainingsdaten stattfinden. Je besser die Trainingsdaten einen diversen und vorurteilsfreien Bewerber*innenmarkt darstellen, desto höher ist die Chance, Gender Bias zu überwinden.

Eine weitere Erkenntnis aus der Arbeit ist, dass die vorgestellten Auswahlverfahren im Bewerbungsprozess ein gewisses Risiko eines Gender Bias aufweisen, je nachdem, wie intensiv verzerrte und voreingenommene Trainingsdaten berücksichtigt werden. Darüber hinaus stellt das Black-Box-Problem eine Herausforderung dar, die aktuell und zukünftig ein Risiko für Gender Bias darstellt. Zusammenfassend kann die Forschungsfrage anhand der vorliegenden Literatur eindeutig dahingehend beantwortet werden, dass auch KI-gestützte Auswahlverfahren Frauen aufgrund eines Gender Bias in Bewerbungsprozessen diskriminieren.

Literatur

ACCENTURE, 2018. Accenture launches new artificial intelligence testing services. [online]. [Zugriff am: 15. 11. 2022]. Verfügbar unter: <https://newsroom.accenture.com/news/accenture-launches-new-artificial-intelligence-testing-services.htm>

ADELMANN, Lena & WIEDMER, Jennifer, 2017. Der Einsatz von Künstlicher Intelligenz in der Rekrutierung [online]. Universität Basel. [Zugriff am: 15. 11. 2022]. Verfügbar unter: https://wwz.unibas.ch/fileadmin/user_upload/wwz/00_Professuren/Beckmann_Personal_und_Organisation/Lehre/Digital_Transformation/Der_Einsatz_von_Kuenstlicher_Intelligenz_in_der_Rekrutierung_Adelmann_und_Wiedmer.pdf

ANTI-BIAS, 2022. Biases von A-Z [online]. [Zugriff am: 20.11.2022]. Verfügbar unter: <https://www.anti-bias.eu/wissen/biases-von-a-z/>

ANTIDISKRIMINIERUNGSSTELLE DES BUNDES, 2017. Diskriminierung in Deutschland. Dritter Gemeinsamer Bericht der Antidiskriminierungsstelle des Bundes und der in ihrem Zuständigkeitsbereich betroffene Beauftragten der Bundesregierung und des Deutschen Bundestages [online]. Berlin [Zugriff am: 27.11.2022]. Verfügbar unter: https://www.antidiskriminierungsstelle.de/SharedDocs/downloads/DE/publikationen/BT_Bericht/gemeinsamer_bericht_dritter_2017.pdf;jsessionid=7DA2CEA95D-FA7B772DB307036320A6DD.intranet242?__blob=publicationFile&v=6

AREL, Itamar, ROSE, Derek C. und KARNOWSKI, Thomas P., 2010. Deep machine learning – A new frontier in artificial intelligence research. IEEE computational intelligence magazine 5 (4): 13-18 [Zugriff am 27.06.2023]. Verfügbar unter: DOI: [10.1109/MCI.2010.938364](https://doi.org/10.1109/MCI.2010.938364)

BULLER, Michelle, 2021. Recruiting: Definition, Prozess und Instrumente [online]. Personio. [Zugriff am: 06.11.2022]. Verfügbar unter: <https://www.personio.de/hrlexikon/recruiting/>

BUXMANN, Peter, und SCHMIDT, Holger, 2019. Künstliche Intelligenz: Mit Algorithmen zum wirtschaftlichen Erfolg. Springer VS: Wiesbaden

CAMPION, Michael C., CAMPION, Michael A., CAMPION, Emily & REIDER, Matthew, 2016. Initial investigation into computer scoring of candidate essays for personnel selection. Journal of Applied Psychology 101(7), 958-975. DOI: [10.1037/apl0000108](https://doi.org/10.1037/apl0000108)

CASTELVECCHI, Davide. 2016. Can we open the black box of AI? Nature News 538 (7623): 20-23 [Zugriff am 27.06.2023]. Verfügbar unter: DOI: [10.1038/538020a](https://doi.org/10.1038/538020a)

CHARTA DER VIELFALT, 2022. Charta der Vielfalt [online]. [Zugriff am: 06.11.2022]. Verfügbar unter: <https://www.charta-der-vielfalt.de/ueber-uns/ueber-die-initiative/die-urkunde-im-wortlaut/>

CRAWFORD, Kate, 2016. Artificial Intelligence's White Guy Problem [online]. The New York Times, New York Edition, [Zugriff am: 20.11.2022]. Verfügbar unter: <https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html>

DALENBERG, D.J. 2018. Preventing discrimination in the automated targeting of job advertisements. Computer Law & Security Review 34(3), S. 615-627 [Zugriff am: 20.11.2022]. Verfügbar unter: DOI: [10.1016/j.clsr.2017.11.009](https://doi.org/10.1016/j.clsr.2017.11.009)

DASTIN, Jeffrey, 2018. Amazon scraps secret AI recruiting tool that showed bias against women [online]. Reuters. 10.10.2018 [Zugriff am: 20.11.2022]. Verfügbar unter: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>

DATA DRIVEN RECRUITING. 2021. HR monkeys [online]. [Zugriff am: 23.11.2022]. Verfügbar unter: <https://hr-monkeys.de/data-driven-recruiting/>

DAVIDSON, Thomas, WARMSLEY, Dana, MACY, Michael & WEBER, Ingmar 2017. Automated hate speech detection and the problem of offensive language. Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017), 512–516

DOMSCH, Michael, LADWIG, Desiree, WEBER, Florian, Hrsg., 2017. Cross Mentoring [online]. Ein erfolgreiches Instrument organisationsübergreifender Personalentwicklung. Berlin u. a.: Springer Gabler. [Zugriff am: 16.11.2022]. PDF-E-Book. Verfügbar unter: DOI: [10.1007/978-3-662-53184-6](https://doi.org/10.1007/978-3-662-53184-6)

DRAUDE, Claude, MAAß, Susanne, WAJDA, Kamila, 2014. GERD: ein Vergehensmodell zur Integration von Gender/Diversity in der Informatik. Vielfalt der Informatik: Ein Beitrag zu Selbstverständnis und Außenwirkung

EICHLER, Margrit et al., 2002. Zu mehr Gleichberechtigung zwischen den Geschlechtern: Erkennen und Vermeiden von Gender Bias in der Gesundheitsforschung: deutsche Bearbeitung eines vom kanadischen Gesundheitsministerium herausgegebenen Handbuchs. Berlin Zentrum Public Health.

ELLIS, Evillin & WATSON, Philippa. 2012. EU Anti-Discrimination Law, 2. Aufl. Oxford, UK: Oxford University Press.

GOOGLE. 2019. At I/O '19: Building a more helpful Google for everyone [online]. [Zugriff am: 20.11.2022]. Verfügbar unter: <https://www.blog.google/technology/developers/io19-helpful-google-everyone/>

IBM. 2018. AI Fairness 360 [online]. [Zugriff am: 20.11.2022]. Verfügbar unter: <https://developer.ibm.com/open/projects/ai-fairness-360/>

INTERNATIONAL LABOUR ORGANIZATION 2017. ACT/EMP Research Note [online]. Breaking barriers: Unconscious gender bias at workplace. [Zugriff am: 20.11.2022]. Verfügbar unter: https://www.ilo.org/actemp/publications/WCMS_601276/lang-en/index.htm

KAHNEMAN, Daniel, 2011. Schnelles Denken, Langsames Denken. Siedler Verlag: München

KASPERKEVIC, Jana, 2015. Google says sorry for racist auto-tag in photo app [online]. The Guardian. [Zugriff am: 20.11.2022]. Verfügbar unter: <https://www.theguardian.com/technology/2015/jul/01/google-sorry-racist-auto-tag-photo-app>

KNOBLOCH, Tobias, HUSTENDT, Carla, 2019. Der maschinelle Weg zum passenden Personal [online]. Bertelsmannstiftung. [Zugriff am 05.11.2022]. Verfügbar unter: DOI: [10.11586/2019027](https://doi.org/10.11586/2019027)

KRÜGER, Julia & LISCHKA, Konrad, 2018. Damit Maschinen den Menschen dienen. Lösungsansätze, um algorithmische Entscheidungen in den Dienst der Gesellschaft zu stellen. Impuls Algorithmenethik #6, Gütersloh

LANGER, Paul & WEYERER, Jan, 2020. Diskriminierungen und Verzerrungen durch Künstliche Intelligenz. Entstehung und Wirkung im gesellschaftlichen Kontext. In: OSWALD, Michael, und BORUCKI, Isabelle, Hrsg. Demokratietheorie im Zeitalter der Frühdigitalisierung. Wiesbaden: Springer VS. Verfügbar unter: DOI: [10.1007/978-3-658-30997-8_11](https://doi.org/10.1007/978-3-658-30997-8_11)

LLOYD, Keith 2018. Bias Amplification in Artificial Intelligence Systems. AAAI FSS-18: Artificial Intelligence in Government and Public Sector, Arlington, Virginia, USA

LOCHNER, Katharina & PREUß, Alan, 2018. Digitales Recruiting. Gruppe. Interaktion. Organisation. Zeitschrift für Angewandte Organisationspsychologie 49, S. 193-202. [Zugriff am 05.11.2022]. Verfügbar unter: [10.1007/s11612-018-0425-7](https://doi.org/10.1007/s11612-018-0425-7)

NALTY, Katharina 2016. Strategies for confronting unconscious Bias. The Colorado Lawyer [online]. 45(5) [Zugriff am: 05.11.2022]. Verfügbar unter: <https://kathleen-naltyconsulting.com/wp-content/uploads/2016/05/Strategies-for-Confronting-Unconscious-Bias-The-Colorado-Lawyer-May-2016.pdf>

ORWAT, Carsten, 2019. Diskriminierungsrisiken durch Verwendung von Algorithmen [online]. [Zugriff am: 20.11.2022]. Verfügbar unter: https://www.antidiskriminierungsstelle.de/SharedDocs/downloads/DE/publikationen/Expertisen/studie_diskriminierungsrisiken_durch_verwendung_von_algorithmen.pdf?__blob=publicationFile&v=3

PORTER, Theodore, 1996. Trust in numbers: The pursuit of objectivity in science and public life: Princeton University Press

RAJI, Inioluwa Deborah & BUOLAMWINI, Joy 2019. Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial ai products. In: ASSOCIATION FOR THE ADVANCEMENT OF ARTIFICIAL INTELLIGENCE (AAAI) (Hrsg.), Proceedings of the 2019 AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (Bd. 1, S. 1–7).

ROTTWILM, Christoph, 2018. Schlauer Roboter sucht Personal für Ikea, Pepsi und Co [online]. manager magazin. 26.04.2018 [Zugriff am: 15.11.2022]. Verfügbar unter: <https://www.manager-magazin.de/unternehmen/artikel/roboter-vera-software-sucht-personal-fuer-ikea-pepsi-und-co-a-1204888.html>

ROXBURGH, Erin & HANSEN, Kate, 2016. Bias in Recruitment and Selection [online]. Victoria University of Wellington. [Zugriff am: 19.11.2022]. Verfügbar unter: https://www.wgtn.ac.nz/__data/assets/pdf_file/0008/1816604/bias-in-recruitment-and-selection.pdf

SCHÜTZ, Martin, 2021. Robot-Recruiting – Employer Branding Lexikon [online]. Employer Branding NOW. [Zugriff am: 18.11.2022]. Verfügbar unter: <https://www.employer-branding-now.de/employer-branding-wiki-lexikon>

SEMET, Sven, & HILBERER, Laura, 2018. Potenziale von künstlicher Intelligenz in HR. In P. Thorsten & J. Wolfgang (Hrsg.), Digital HR – Smarte und agile Systeme. Prozesse und Strukturen im Personalmanagement. Freiburg: Haufe-Lexware GmbH & Co. KG

TEETZ, Ingolf 2018. Künstliche Intelligenz im Recruiting. In: PETRY, Thorsten und JÄGER, Wolfgang, Hrsg. Digital HR: Smarte und agile Systeme, Prozesse und Strukturen im Personalmanagement. Freiburg u. a.: Haufe Group. S. 225-240.

TURING, Alan 1950. Computing machinery and intelligence. In: Mind, New Series. 59 (236), S. 433-460 [Zugriff am 27.06.2023]. Verfügbar unter: DOI: [10.1093/mind/LIX.236.433](https://doi.org/10.1093/mind/LIX.236.433)

TENGAI ROBOT. 2021. Tengai Unbiased [online]. [Zugriff am: 25.11.2022]. Verfügbar unter: <https://www.tengai-unbiased.com/tengai-robot/>

VEDDER, Günther, 2005. Denkanstöße zum Diversity Management. In: Arbeit 14(1), S. 34-43. Verfügbar unter: DOI: [10.1515/arbeit-2005-0104](https://doi.org/10.1515/arbeit-2005-0104)

VERHOEVEN, Tim, 2019. Digitalisierung im Recruiting: Wie sich Recruiting durch künstliche Intelligenz, Algorithmen und Bots verändert (1. Aufl. 2020 Aufl.) [E-Book]. Springer Gabler. Verfügbar unter: DOI: [10.1007/978-3-658-25885-6](https://doi.org/10.1007/978-3-658-25885-6)

WILKE, Felicitas, 2018. Bewerbungsroboter: Künstliche Intelligenz diskriminiert (noch) [online]. ZEIT.de. 18.10.2018 [Zugriff am: 22.11.2022]. Verfügbar unter: <https://www.zeit.de/arbeit/2018-10/bewerbungsroboter-kuenstliche-intelligenz-amazon-frauen-diskriminierung/komplettansicht>

WOLFANGEL, Eva, 2018. Programmierter Rassismus. Zeit online. 19.06.2018 [Zugriff am: 24.11.2022]. Verfügbar unter: <https://www.zeit.de/digital/internet/2018-05/algorithmen-rassismus-diskriminierung-daten-vorurteile-alltagsrassismus>

WONDRAK, Manfred, 2020. Denkfehler! Wie UnconsciousBias entstehen [online]. Anti-Bias. [Zugriff am: 19.11.2022]. Verfügbar unter: <https://www.anti-bias.eu/>

YONG, Hannah, 2018. Das Internet ist nicht genderneutral. Über Geschlechterethik in Netz- öffentlichkeit. In P. Otto & E. Gräf (Hrsg.), 3TH1CS. Die Ethik der digitalen Zeit Schriftenreihe/Bundeszentrale für Politische Bildung, Band 10181, Sonderausgabe für die Bundeszentrale für Politische Bildung. Bonn: Bundeszentrale für Politische Bildung