




POLICY BRIEF

Lara Fleck, Nicholas Rounding, Pelin Özgül

Künstliche Intelligenz in der Personalauswahl



Diese Publikation ist im Rahmen des Forschungsprojekts ai:conomics – eine Kooperation der Universität Maastricht, dem Institut für Arbeitsmarkt- und Berufsforschung (IAB) sowie der zukunft zwei GmbH – entstanden und wurde von der Abteilung Denkfabrik Digitale Arbeitsgesellschaft des BMAS finanziell gefördert. Die Durchführung der Untersuchungen sowie die daraus folgenden Schlussfolgerungen sind von den Autor*innen in eigener wissenschaftlicher Verantwortung vorgenommen worden.


 **Maastricht University** **IAB** INSTITUT FÜR ARBEITSMARKT- UND BERUFSFORSCHUNG
Die Forschungserrichtung der Bundesagentur für Arbeit**zukunft zwei**

Gefördert durch:



Bundesministerium
für Arbeit und Soziales

aufgrund eines Beschlusses
des Deutschen Bundestages



Künstliche Intelligenz in der Personalauswahl

Citation for published version (APA):

Fleck, L., Rounding, N., & Özgül, P. (2022). *Künstliche Intelligenz in der Personalauswahl*. ROA. ROA External Reports Nr. ai:conomics Kurzdossier Mai 2022

Document status and date:

Published: 31/05/2022

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Künstliche Intelligenz in der Personalauswahl

Fleck, L., Rounding, N., und Özgül, P.

Kernaussagen

- Der Einsatz Künstlicher Intelligenz (KI) in der Personalauswahl kann sich erheblich auf das Ausmaß von Diskriminierung im Arbeitskontext auswirken. Algorithmen, die menschliche Vorurteile aufrechterhalten, können Ungleichheiten auf dem Arbeitsmarkt begünstigen.
- Durch objektive Entscheidungs- und Vorhersagekraft, kann KI potenziell die Vielfalt innerhalb eines Bewerber:innen-Pools erhöhen und somit Diskriminierung reduzieren. In der Praxis gilt dies jedoch nicht in jedem Fall.
- KI kann die Diskriminierung von Arbeitskräften weiter verschärfen, indem menschliche Vorurteile und unfaire Einstellungsentscheidungen auf Kosten bestimmter Personengruppen reproduziert werden.
- Die Ergebnisse KI-basierter Personalauswahl hängen weitgehend vom Design des angewandten Algorithmus, der Trainingsdaten und Stichproben ab.
- Wenn eine KI, unabhängig von ihrer tatsächlichen Fairness, von betroffenen Bewerber:innen als weniger fair wahrgenommen wird, könnte dies ihre Arbeitgeberwahl bereits im Vorhinein beeinflussen, oder gar nach erfolgreicher Anstellung das Vertrauensverhältnis zum Arbeitgeber negativ beeinträchtigen.

1. Einleitung

Künstliche Intelligenz (KI) als Prognosetechnologie kann von Firmen eingesetzt werden, um ihre Effizienz innerhalb verschiedener Prozesse zu erhöhen¹, z.B. in der Personalauswahl (Bogen & Rieke, 2018; Gonzalez et al., 2019; Tambe et al., 2019). Da es eine Vielzahl unterschiedlicher Meinungen bezüglich der Nutzung von KI in Auswahlverfahren gibt, möchten wir uns auf eine spezifische Debatte fokussieren: ob sie das Potenzial hat, Diskriminierung zu verstärken oder zu reduzieren. Diskriminierung innerhalb der Personalauswahl ist ein schon seit Langem untersuchtes Phänomen, das häufig auf unbewusster, zwischenmenschlicher Voreingenommenheit von Recruiter:innen bei subjektiver Beurteilung von Bewerber:innen aufgrund unterschiedlicher persönlicher Merkmale basiert². Weiter gefasst trägt Diskriminierung zu Ungleichheiten auf dem Arbeitsmarkt bei (Banks & Ford, 2008), indem Personen aus überrepräsentierten Gruppen wiederholt Zugang zu attraktiven Beschäftigungsmöglichkeiten bekommen, während dieser gleichzeitig für andere Gruppen begrenzt

1. Die aktuelle Debatte bezüglich KI-Implementierung und ihrer Auswirkungen auf Arbeitsmärkte wird im [ersten ai:conomics Kurzdossier](#) beleuchtet.

2. Für eine Erläuterung von Diskriminierung in Personalauswahlverfahren, siehe Box 1.

wird. Diskriminierung führt auch zu Qualifikationsmissverhältnissen und ineffizienter Zuteilung von Ressourcen (McGowan & Andrews, 2015). Prinzipiell sollte KI-Technologie vorurteilsfrei sein und somit Bewerber:innen nicht beispielsweise aufgrund ihres Aussehens, ihrer am Namen abgelesenen Nationalität oder ihres Bildungsabschlusses aussortieren. Einer der stärksten potenziellen Vorteile von KI-Technologie im Vergleich zu Menschen besteht in ihrem Potenzial, unter bestimmten Voraussetzungen objektive und verlässliche Entscheidungen zu treffen. Algorithmen können Vorhersagen aufgrund von großen Datenmengen modellieren, zu deren vollständiger Erfassung Menschen nicht fähig sind (Agrawal et al., 2019; van den Broek et al., 2021). Aspekte, die fragwürdig bleiben, sind jedoch nicht nur, ob die Objektivität und Wirksamkeit von KI in der Personalauswahl tatsächlich evidenzbasiert ist, sondern auch, wie die betroffe-

nen Personen algorithmische Entscheidungsprozesse wahrnehmen (Gonzalez et al., 2022).

Dieses Kurzdossier skizziert evidenzbasierte Erkenntnisse zum Einsatz von KI in Personalauswahlverfahren und damit verbundene Auswirkungen auf Qualität und Vielfalt der eingestellten Bewerber:innen. In Abschnitt 2.1 wird das Potenzial der KI, Diskriminierung zu reduzieren, beleuchtet. Darüber hinaus wird in Abschnitt 2.2 ein Problem bei der Anwendung von KI bei Einstellungen ausgeführt: KI-Bias. Anschließend werden in Abschnitt 3 die Relevanz der Wahrnehmung algorithmischer Entscheidungsprozesse durch Arbeitskräfte betont und die potenziellen Folgen auf individueller und betrieblicher Ebene erörtert. Abschließend werden regulatorische Implikationen der Erkenntnisse der Forschenden aus dem ai:conomics-Team diskutiert, um für mehr Fairness in algorithmischen Personalauswahlverfahren zu sorgen.

Was ist Diskriminierung?

Diskriminierung ist die unfaire oder nachteilige Behandlung von Menschen und Gruppen aufgrund ihrer Rasse oder ethnischen Herkunft, ihres Geschlecht, ihrer Religion oder Weltanschauung, einer Behinderung, ihres Alters oder ihrer sexueller Identität (§1, §2 Abs. 1, Allgemeines Gleichbehandlungsgesetz, AGG, 2006; Art. 1 Abs. 1, Übereinkommen über die Diskriminierung in Beschäftigung und Beruf, C111, 1958). Die unfaire Behandlung im Kontext der Personalauswahl basiert darauf, dass zwei Personen aufgrund von Merkmalen, die nicht zu ihrer Produktivität beitragen, unterschiedlich behandelt werden (§3 Abs. 1, AGG, 2006; Becker, 2010). Um dies konkret zu demonstrieren, kann es z.B. sein, dass voreingenommene Recruiter:innen weiße Bewerber:innen schwarzen Bewerber:innen vorziehen, obgleich Letztere einen besseren Lebenslauf vorweisen oder besser im Vorstellungsgespräch abschneiden. Forscher:innen identifizierten diskriminierende Verhaltensweisen im Einstellungsprozess und fanden heraus, dass weiße Bewerber:innen eher zu Bewerbungsgesprächen eingeladen werden und Jobangebote erhalten als schwarze Bewerber:innen mit ähnlichen Merkmalen

(Bendick et al., 1994). Dieses Vorurteil kann sich bereits beim Screening der Lebensläufe äußern: einige Studien zeigen, dass Einladungen zu Vorstellungsgesprächen signifikant vermehrt werden könnten, wenn Namen, die auf eine bestimmte Nationalität, Rasse oder ethnische Herkunft hinweisen, verändert würden (Bertrand & Mullainathan, 2004; Kaas & Manger, 2012; McGinnity et al., 2009). Das Bestehen eines sog. „Bias“ innerhalb des Auswahlverfahrens kann sich signifikant auf das Einkommen von Personen, ihre Karriereaussichten und die Entwicklung ihres Humankapitals auswirken. Darüber hinaus entsteht potenziell ein suboptimales Marktergebnis infolge eines mangelnden Abgleiches zwischen Kompetenzen und Anforderungen, welcher wiederum wirtschaftliches Wachstum einschränken könnte (McGowan & Andrews, 2015). Aufgrund dieser Überlegungen ist es von entscheidender Bedeutung, herauszuarbeiten, wie und in welchem Umfang Diskriminierung in der Personalauswahl vorliegt, in welchen Phasen des Auswahlprozesses sie entsteht und wie sie abgeschwächt werden kann.

2. Kann KI Diskriminierung reduzieren?

Die Forschung hat wiederholt festgestellt, dass Personalauswahlverfahren häufig mit Diskriminierung aufgrund persönlicher Merkmale verbunden sind, die in Wirklichkeit nichts mit der Produktivität der Bewerber:innen zu tun haben; als solche potenziellen Merkmale sind beispielsweise Rasse oder Nationalität zu nennen, die vom Namen oder vom Aussehen von Bewerber:innen abgeleitet werden (Becker, 2010; Bendick et al., 1994; Bertrand & Mullinathan, 2004; Kaas & Manger, 2012). In letzter Zeit haben Forscher:innen die Integration von KI in den Auswahlprozess vorgeschlagen, um auf unbewussten Vorurteilen von Menschen basierende Diskriminierung zu überwinden (van den Broek et al., 2021). In Anbetracht der Tatsache, dass Menschen nur eine vergleichsweise begrenzte Fähigkeit zur Verarbeitung von Informationen aus unterschiedlichen Quellen haben, kann KI objektiveres Wissen bieten: intelligente Algorithmen in der Personalauswahl ziehen große Datenbestände aus verschiedenen Datenquellen heran, zum Beispiel Lebensläufe, Interviews und soziale Medien, sodass sie vorhersagen können, welche Bewerber:innen die offene Stelle am besten besetzen könnten (Agrawal et al., 2019); dadurch werden die Einblicke von Fachexpert:innen häufig übertroffen (van den Broek et al., 2021).

Die Eignung von KI für den Einstellungsprozess spiegelt sich in der Zunahme von Anbieter:innen von KI-Lösungen zur Personalauswahl wieder (Raghavan et al., 2020). KI-basierte Systeme können in verschiedenen Phasen des Auswahlprozesses unterschiedliche Funktionen haben, z.B. Entfernung voreingenommener Formulierungen aus Stellenanzeigen, Screening von Lebensläufen, Analyse von Video-Interviews oder Bewertung von jemandes Eignung für einen Job (Gonzalez et al., 2019; Raghavan et al., 2020). Die zahlreichen Möglichkeiten der Anwendung von KI in den frühen Phasen des Auswahlprozesses unterstreichen die Tatsache, dass Algorithmen das Potenzial haben, substantiell zu beeinflussen, welche Bewerber:innentypen in der engeren Auswahl landen (Bogen & Rieke, 2018).


Derzeit sprechen hauptsächlich die Effizienzsteigerung innerhalb des Auswahlverfahrens und die Optimierung entsprechender Kennzahlen wie Einstellungszeit, Einstellungskosten, Einstellungsqualität und Mitarbeiter:innenbindung für den Einsatz algorithmischer Auswahlverfahren in Unternehmen (Bogen & Rieke, 2018; Gonzalez et al., 2019). Darüber hinaus kann KI zur Reduzierung des interpersonellen Bias (van den Broek et al., 2021) und somit zu Unternehmenszielen in Hinsicht auf Viel-

falt und Inklusion beitragen (Bogen & Rieke, 2018). Praktiker:innenzeitschriften wie Forbes (Windley, 2021), Harvard Business Review (Chamorro-Premuzic & Akhtar, 2019) und Business Insider (Garfield, 2017) berichten über Vorteile der Anwendung von KI bei der Personalauswahl, während sie Praktiker:innen einen vorsichtigen Ansatz nahelegen, sich nicht einzig und allein auf Algorithmen zu verlassen. Dies impliziert, dass es neben dem Potenzial der KI auch eine Reihe von Gefahren gibt. Auch die akademische Forschung hat damit begonnen, zu untersuchen, ob KI Diskriminierungen in der Personalauswahl tatsächlich reduziert oder möglicherweise sogar ausschließt. In den nächsten Abschnitten werden identifizierte Möglichkeiten und Gefahren der Anwendung von KI in Auswahlverfahren ausgearbeitet.

2.1. KI vs. Menschen: Wer erzielt bessere Ergebnisse in der Personalauswahl?

Innerhalb des Auswahlverfahrens ist das Screening von Lebensläufen ein Bereich, der als Quelle für Diskriminierung identifiziert wurde (z.B. Bertrand & Mullinathan, 2004). Zwei Studien (Cowgill, 2020; Li et al., 2020) entwickelten KI-Algorithmen für das Screening von Lebensläufen und verglichen sie mit der Leistung von Recruiter:innen. Li et al. (2020) betrachten die Personalauswahl als eine Herausforderung für Unternehmen, sog. Exploitation und Exploration im Prozess abzuwägen. Exploitation bedeutet, dass im Auswahlverfahren primär Bewerber:innen mit einer aufgrund ihres Werdegangs hoch eingeschätzten Passgenauigkeit zur Position ausgewählt werden. Exploration hingegen erlaube mehr Spielraum in der Vorhersage von Passgenauigkeit und würde mehr Bewerber:innen mitunter eher unkonventionellen Hintergründen in die engere Auswahl lassen. Aktuelle, auf Trainingsdaten basierende Einstellungsalgorithmen sagen eher positive Ergebnisse für Gruppen mit passgenauen Werdegängen voraus, wobei Bewerber:innen mit unkonventionellen Hintergründen unberücksichtigt bleiben, obwohl Sie aufgrund anderer Merkmale im Job dennoch hervorragend abschneiden könnten. Der Einbau von Exploration in Einstellungsalgorithmen könnte somit ein Schritt in Richtung eines inklusiveren Ansatzes im Recruiting sein (Li et al., 2020), was den gezielten Aufbau eines diverseren Bewerber:innen-Pools (Bogen & Rieke, 2018) begünstigen könnte.

Dementsprechend vergleichen Li et al. (2020) menschliche Recruiter:innen mit drei Algorithmen: zwei, die durch überwachtes Lernen auf möglichst hohe Passgenauigkeit trainiert wurden, und einen mit



erhöhtem Spielraum, einen sog. „Explorationsbonus“. Die Studienergebnisse zeigten, dass der Explorationsalgorithmus wesentlich höhere Anteile von unterrepräsentierten Gruppen am weiteren Auswahlverfahren teilnehmen lies. Dieser Algorithmus schnitt nicht nur signifikant besser ab als der Mensch, er übertraf auch die beiden überwachten Algorithmen. Als wichtiges Ergebnis ist zu vermerken, dass die beiden in dieser Studie entwickelten Algorithmen für überwachtes Lernen in der Auswahl einer möglichst diversen Bewerber:innengruppe sogar schlechter abschnitten als menschliche Recruiter:innen. Dies deutet darauf hin, dass Algorithmen Vorurteile auch übernehmen könnten. Die Autor:innen betonen die Relevanz des Algorithmus-Designs für seine Leistungsfähigkeit: Nicht alle Algorithmen seien gleich leistungsfähig und dies sei ein wichtiger Faktor, der vor allem bei einer großen Auswahl zu berücksichtigen sei.

In einem experimentellen Setting der Personalauswahl stellt Cowgill (2020) fest, dass ein Algorithmus die Lebensläufe von Bewerber:innen auswählte, die mit höherer Wahrscheinlichkeit entweder in die nächste Runde des Auswahlverfahrens kamen oder das entsprechende Jobangebot bekamen. Ferner erhöhte der Algorithmus die Anzahl „unkonventioneller“ Einstellungen, unter anderem von Frauen, Minderheiten, Bewerber:innen ohne Empfehlungen oder praktischer Erfahrung, oder Bewerber:innen von weniger renommierten Universitäten. Oberflächlich betrachtet, scheint es, dass der Algorithmus mit ähnlichen Verfahren (d.h. LASSO und anderen Standard-Machine-Learning-Techniken) wie den überwachten Algorithmen gemäß Li et al. (2020) trainiert wurde, die feststellten, dass diese Algorithmen zu weiteren Vorurteilen führen könnten. Diese gegensätzlichen Ergebnisse weisen auf die Notwendigkeit zukünftiger Forschung zur Herbeiführung eines Konsenses bezüglich der Leistungsfähigkeit von Algorithmen hin. Die Einbeziehung unterschiedlicher Algorithmus-Designs, wie des Explorationsalgorithmus von Li et al. (2020), könnte zu tiefergehenden Analysen und Vergleichen führen und weitere Erkenntnisse darüber erschließen, wie Diskriminierung zu reduzieren wäre. Wie außerdem von Cowgill (2020) unterstrichen, könnte die Zusammensetzung des Trainingsdatensatzes zu unterschiedlichen Ergebnissen führen. In seinen Modellen zeigt der Autor, wie sich mehr oder weniger statistisches Rauschen auf die Ergebnisse auswirken könnte. Deshalb ist es von entscheidender Bedeutung, die Ergebnisse solcher Studien mithilfe unterschiedlich zusammengesetzter Trainingsdaten zu reproduzieren.

Neben dem Lebenslauf-Screening könnte KI auch für die Analyse von Job-Eignungstests eingesetzt werden. Dabei handelt es sich um von Bewerber:innen ausgefüllte Fragebögen, die im Auswahlverfahren berücksichtigt werden können. Hoffman et al. (2018) testeten die Ergebnisse einer solchen KI und finden im Vergleich zur menschlichen Auswahl positive Auswirkungen auf die Beschäftigungsdauer der eingestellten Arbeitskräfte. Sie untersuchen außerdem, ob die Qualität der Neueinstellungen höher ist, wenn Recruiter:innen der Empfehlung des Algorithmus folgen oder diese aktiv verwerfen. Sie zeigen, dass Recruiter:innen, die die Empfehlungen des Algorithmus verwerfen, schlechtere Entscheidungen treffen: Die von ihnen ausgewählten Arbeitskräfte wiesen durchschnittlich kürzere Beschäftigungsdauern und mangelnde Produktivitätssteigerung auf. Die Studie zeigt somit, dass KI gegenüber Recruiter:innen eine Verbesserung darstellen kann: nicht nur bei der Identifizierung leistungsfähigerer Bewerber:innen, sondern auch bei dem Aufbau eines diverseren Bewerber:innen-Pools. Der angewandte Algorithmus war jedoch urheberrechtlich geschützt (Hoffman et al., 2018) und sein Design oder die verwendete Lernmethode wurden nicht offengelegt. Für eine abschließende, objektive Bewertung der KI-Technologie, wie oben erörtert, wäre eine Offenlegung der Modelltechnologie und der zum Trainieren des Modells verwendeten Daten erforderlich.

2.2. Vorsicht: Nicht jede KI ist unvoreingenommen

Im vorangehenden Abschnitt wurden einige Fälle skizziert, in denen KI menschliche Entscheidungen verbessern und diskriminierende Ergebnisse im Personalauswahlverfahren reduzieren könnte. Jedoch wurde auch gezeigt, dass unterschiedliche Algorithmen zu unterschiedlichen Ergebnissen führen können. Wie bei den Algorithmen für überwachtes Lernen gemäß Li et al. (2020) betont, besteht die Gefahr, dass ein Algorithmus weiteren Bias mit sich bringen könnte. In Verbindung damit existieren zahlreiche Medien-Artikel und Studien, die die Kehrseite von KI-Algorithmen zur potenziellen Verstärkung von Vorurteilen hervorheben, z.B. auch in anderen Bereichen wie der Gesichtserkennung (Leslie, 2020). Deshalb untersucht dieser Abschnitt die zugrundeliegenden Quellen und entsprechende Entstehung von KI-Bias anhand von Beispielen.

Derzeitige KI-Algorithmen hängen stark von Trainingsdaten ab und lernen hauptsächlich durch Anwendung einer Kombination mehrerer statistischer




Ansätze, die auf Trainingsdaten angewandt werden. Dieses Lernverfahren wird allgemein als ‚Machine Learning‘ oder ‚Deep Learning‘ bezeichnet (OECD, 2019). Eine kürzlich durchgeführte Literaturstudie (Cecere et al., 2021) identifizierte die Grundursachen des KI-Bias als hauptsächliche Folge des Einsatzes von (I) *nichtrepräsentativen Daten* und *Trainingsmustern*, (II) *falsch gelabelten Ergebnissen* und (III) *voreingenommenen Programmierer:innen* (Cecere et al., 2021; Cowgill, 2020; Cowgill & Stevenson, 2020; Cowgill & Tucker, 2020). Algorithmen werden nicht intelligent geschaffen; sie müssen an Daten trainiert werden, bevor sie auf die reale Welt Anwendung finden. Verzerrte oder nicht repräsentative Trainingsdaten können diskriminierende Auswahlprozesse reproduzieren, die ursprünglich bereits zu der nicht repräsentativen Stichprobe geführt haben. Somit kann der Bias weiter etabliert werden (Cowgill & Tucker, 2020; Kleinberg et al., 2018). Da KI-Algorithmen erst nach dem Training aus früheren Verhaltensweisen und Merkmalen von Personen lernen, kann es zu einem Teufelskreis von Diskriminierung kommen (Cecere et al., 2021). Im Kontext der Personalauswahl kann der Algorithmus, wenn die Trainingsdaten auf vergangenen Beschäftigungsdaten basieren und z.B. ein implizites Bias beinhalten, welches Weiße gegenüber afroamerikanischen und lateinamerikanischen Mitarbeiter:innen bevorzugt, solche Muster erkennen und bestimmte Gruppen von Bewerber:innen unberücksichtigt lassen. Die Entscheidungsfindung könnte dann zugunsten der überrepräsentierten Gruppe stattfinden, die in Vergangenheit eher zu einem Bewerbungsgespräch eingeladen wurde (Köchling & Wehner, 2020; Tambe et al., 2019). Der globale Tech-Gigant Amazon zum Beispiel entdeckte 2018, dass sein KI-basiertes Rekrutierungssystem auf vergangenen Arbeitsleistungsdaten beruhte, die stark männlich dominiert waren und höhere Leistungsbewertungen für weiße Männer enthielten. Infolgedessen ergab der an dieser Stichprobe trainierte Algorithmus höhere Punktwerte für weiße männliche Bewerber, während Frauen und Bewerber:innen, deren Attribute mit Frauen assoziiert wurden, vom Algorithmus aussortiert wurden (Tambe et al., 2019).

Algorithmische Verzerrungen können auch auf falsche Labels in den Trainingsdaten zurückgeführt werden. Dies könnte infolge subjektiver Präferenzen von Entscheidungsträger:innen oder unzureichenden statistischen Modellen, die sensible Faktoren wie demographische Informationen unberücksichtigt lassen, entstehen (Cowgill & Tucker, 2020). In einer Studie betonen Obermeyer et al. (2019) die Bedeutung der Wahl relevanter Labels für das Training des Algorith-

mus durch Dokumentation wichtiger algorithmischer Unterschiede, zu denen es im Gesundheitswesen kam. Sie zeigten, dass bei gleichem algorithmisch vorhergesagten Risiko zur Angabe der Bedürfnisse im Gesundheitswesen, weniger gesunde schwarze Patient:innen ähnliche Risikobewertungen erhielten wie gesündere weiße Patient:innen, während sie tatsächlich schwerere Fälle von Diabetes, Bluthochdruck, Cholesterin und Anämie zeigten. Der Grund für ein solches Rassenvorurteil war in diesem Fall, dass die Vorhersage des Algorithmus eher auf Gesundheitskosten als auf tatsächlichen Krankheiten basierte. Da weniger Geld für schwarze Patient:innen ausgegeben werde, die den gleichen Bedürftigkeitsgrad aufweisen, kommt der Algorithmus zu dem falschen Schluss, schwarze Patient:innen seien genauso gesund wie weiße Patient:innen (Obermeyer et al., 2019). Dieses Beispiel zeigt eine potenzielle Konsequenz, zu der es infolge falscher Labels kommen könnte; die Trainingsdaten labeln niedrige Kosten verursachende Patient:innen als gesund, während diese geringen Kosten in einigen Fällen auf ungleichen Zugang zum Gesundheitswesen und auf sozioökonomischen Status zurückzuführen sind (Cecere et al., 2021; Cowgill & Tucker, 2020; Obermeyer et al., 2019).

Alle Entwickler:innen die das Programmieren, Labeln und Training von Algorithmen beaufsichtigen, könnten ebenfalls einen Einfluss auf das Entstehen algorithmischer Vorurteile haben, besonders wenn ihnen technisches Wissen über KI-Ethik fehlt oder wenn sie selbst signifikante Vorurteile haben. Die Arbeit solcher voreingenommener Programmierer:innen kann zu nicht repräsentativen Trainingsmustern und falsch gelabelten Ergebnissen führen. Außerdem könnte es sein, dass Programmierer:innen während des Entwicklungsprozesses verfügbaren Trainingsbeispielen und Genauigkeitsmessungen nicht genügend Aufmerksamkeit schenken (Cowgill & Tucker, 2020).

In dem Versuch, die Entstehung und Verbreitung der Hauptursachen von KI-Bias zu erforschen, untersuchen Cowgill et al. (2020), ob KI-Bias eher durch verzerrte Trainingsdaten oder voreingenommene Programmierer:innen entstehen. Die Ergebnisse ihres Feldversuchs zeigen, dass verzerrte Trainingsdaten die Hauptursache für voreingenommene Vorhersagen sind. Entwickler:innen hingegen seien nicht systematisch voreingenommen und übertragen Vorurteile nicht systematisch auf die von ihnen entwickelten Algorithmen (Cecere et al., 2021; Cowgill et al., 2020). Kurz gesagt: Während jeder Ursprung von KI-Bias fall-spezifisch ist, resultiert der zugrunde liegende Mechanismus aus der Tatsache, dass sich menschliche Vorurteile auf KI-basierte Ansätze übertragen. KI-



Algorithmen produzieren diskriminierende und verzerrte Ergebnisse, wenn ihre antrainierten Daten und Muster entweder ungenau oder verzerrt sind (Köchling & Wehner, 2020; Mujtaba & Mahapatra, 2019). Beinhalten Trainingsdatensätze zu einem bestimmten Grad kulturelle oder persönliche Vorurteile, kann ein KI-Algorithmus diese übernehmen und reproduzieren, woraus sich unfaire Einstellungsentscheidungen in Personalauswahlverfahren ergeben, die wiederum zu Unter- oder Überrepräsentation bestimmter Arbeitnehmer:innengruppen führen können. Die damit einhergehende Diskriminierung wirft dementsprechend Fragen über die Vertrauenswürdigkeit von KI-Systemen auf (Lee, 2018; OECD, 2019).

Zusammenfassend ist ein Verständnis der Trainingsdaten, die der KI zugeführt werden, relevant, um zu beurteilen, ob KI Diskriminierung in der Personalauswahl tendenziell reduziert oder verschärft. Wird eine KI mithilfe verzerrter Trainingsdaten angelernet, könnte Diskriminierung verstärkt werden. Dies legt den Gedanken nahe, dass die bestehenden Datensätze neben den gewählten Algorithmus-Designs bei der Auswahl eines KI-Tools zunächst kritisch beurteilt werden sollten, da verschiedene Ursachen von Bias innerhalb der Trainingsdaten unterschiedliche Algorithmusansätze erfordern.


3. Wahrnehmung algorithmischer Entscheidungen: vertrauenswürdig oder zweifelhaft?

Während KI bereits als ein potenzielles Tool zur Reduzierung von Diskriminierung ausgearbeitet wurde, verursacht die Anwendung von KI in der Personalauswahl andere Bedenken jenseits von Diskriminierung. Wenngleich KI Diskriminierung reduzieren könnte, könnte sie für die Anwendung in der Personalauswahl dennoch ungeeignet sein. Eine primäre Sorge ist, ob Personen, die algorithmischen Entscheidungen ausgesetzt sind, diese im Vergleich zu menschlichen Entscheidungen als fairer ansehen, unabhängig davon, ob sie tatsächlich fairer sind oder nicht (Lee, 2018). Diese menschliche Wahrnehmung ist aus zwei Gründen wichtig: Erstens, wenn Bewerber:innen einen Algorithmus und somit den Rekrutierungsprozess als unfair wahrnehmen, könnten sie sich aktiv dazu entscheiden, sich nicht auf eine Position zu bewerben oder eine solche anzunehmen. Eine solche Selbstselektion von Bewerber:innen kann sich negativ auf die Organisation auswirken: Wenn sich Bewerber:innen mit hoher Passgenauigkeit zu der ausgeschriebenen Position dazu entscheiden, sich nicht auf diese zu bewerben, entgeht der Organisation ein Produktivitäts-

zuwachs, der potenziell nicht durch weniger passgenaue Bewerber:innen kompensiert werden kann. Die stattdessen eingestellten Bewerber:innen könnten so wenig passgenau sein, dass sie ersetzt werden müssen, sodass durch erhöhte Fluktuation und ineffiziente Verteilung von Kompetenznachfrage und -angebot Mehrkosten für die Organisation entstehen.

Zweitens, wenn Bewerber:innen trotz Zweifel an der Fairness des Einstellungsprozesses beschließen, in ein Unternehmen einzutreten, könnte dies zu einem eingeschränkten Vertrauensverhältnis zwischen Arbeitnehmer:in und Arbeitgeber:in führen. Dieser Mangel an Vertrauen könnte sich wiederum auf die Dauer des Beschäftigungsverhältnisses auswirken und somit erneut zu Fluktuation führen, die in der Regel mit hohen Kosten verbunden ist. Zum Beispiel werden durch Fluktuation verursachte Kosten in amerikanischen Unternehmen mit über 16 Milliarden US-Dollar veranschlagt (Scott, 2017). Demnach sind Zufriedenheit von Bewerber:innen mit dem Auswahlprozess und ihr Vertrauen in die Fairness der Personalentscheidungen ein wichtiger Faktor für einstellende Organisationen.

Dementsprechend lässt sich diskutieren, ob intelligente Algorithmen in der Personalauswahl von betroffenen Bewerber:innen als positiv oder negativ wahrgenommen werden. Newman et al. (2020) finden, dass durch KI getroffene Einstellungsentscheidungen weniger fair wahrgenommen werden als identische, menschliche Entscheidungen. Ähnlich zeigen Acikgoz et al. (2020), dass von einer KI durchgeführte Vorstellungsgespräche weniger fair erscheinen als solche, die von Menschen durchgeführt werden. Zusätzlich finden Gonzalez et al. (2019), dass Arbeitskräfte tendenziell weniger Vertrauen in Unternehmen setzen, die Algorithmen als Entscheidungsgrundlage in der Personalauswahl verwenden. Gonzalez et al. (2022) skizzieren potenzielle Mechanismen für das mangelnde Vertrauen in Entscheidungen durch Algorithmen im Vergleich zu Entscheidungen durch Menschen. Bewerber:innen sind häufig nur beschränkt darüber informiert, wie eine KI in einem bestimmten Entscheidungsprozess Anwendung findet, oder sind selbst bei verfügbarer Information vielleicht nicht in der Lage, sie zu verstehen oder zu verarbeiten. Wenn sie stattdessen Heuristiken anwenden oder sich Meinungen basierend auf weniger fundierten Quellen bilden, kann dies ein negatives Image von KI-Anwendungen in Auswahlverfahren zusätzlich verstärken. Bewerber:innen können KI-Entscheidungen auch als weniger fair wahrnehmen, weil sie mit der KI selbst und dem Entscheidungsprozess, in den die KI eingebettet ist, nicht vertraut sind. Es ist naheliegend, dass



sie einem menschlichen Entscheidungsprozess positiver gegenüberstehen, indem sie glauben, mehr Informationen über Beurteilungskriterien und Fairness zu erhalten.

Diese potenziellen Mechanismen könnten mithilfe erhöhter Transparenz bei der Anwendung von KI abgeschwächt werden. Während eine Organisation in der Lage dazu sein kann, zu erklären, wo innerhalb eines Prozesses eine KI implementiert wird und ob sie die endgültige Entscheidungsgewalt hat, ist es für Bewerber:innen aufgrund mangelnder Informationen wahrscheinlich schwieriger zu durchschauen, wie eine KI zu ihren Entscheidungen gelangt. Moderne KI-Algorithmen werden mit neuronalen Netzen und Deep-Learning-Techniken trainiert; dies sind undurchsichtige Techniken, da noch nicht genau bekannt ist, wie eine KI zu ihren Entscheidungen gelangt. Forscher:innen arbeiten auf erklärbare KI hin (siehe z.B. Arrieta et al., 2020 für eine Diskussion), um ein tiefergehendes Verständnis ihrer jeweiligen Entscheidungsprozesse zu ermöglichen. Dies ist ein notwendiger Schritt, der dabei helfen soll, KI „vertrauenswürdig“ zu gestalten.

4. Politische Auswirkungen

Welche regulatorischen Anstrengungen werden derzeit unternommen, um sicherzustellen, dass KI humanzentriert, vertrauenswürdig und sicher ist? Der Vorschlag der Europäischen Kommission (2021) für eine KI-Verordnung stuft die Anwendung von KI in der Personalauswahl als hohes Risiko ein. Im Rahmen dieser Einstufung wird KI in der Personalauswahl nicht verboten, sondern wird vielmehr reguliert und muss eine Reihe von Bedingungen erfüllen, wie zum Beispiel die Einrichtung eines Risikomanagementsystems, Transparenzbedingungen und menschliche Überwachung. Das Risikomanagementsystem erfordert die Identifizierung und Analyse von bekannten und vorhersehbaren Risiken sowie deren Bewertung (Europäische Kommission, 2021). Ein weiterer wichtiger Aspekt, der auch in diesem Kurzdossier beleuchtet wird, ist die potentielle Reproduktion menschlicher Vorurteile durch KI. Mit Blick auf weitere, in diesem Kurzdossier angesprochenen Risiken können Transparenzanforderungen, die verlangen, dass Benutzer:innen in der Lage sein müssen, den System-Output zu interpretieren und seinen beabsichtigten Zweck zu begreifen, dazu beitragen, einige der Bedenken hinsichtlich der wahrgenommenen Vertrauenswürdigkeit von KI zu entkräften.


Eine im Zusammenhang mit der KI-Verordnung in der Literatur aufgeworfene Frage ist, ob die Kosten

die Innovation und Anwendung von KI-Algorithmen infolge eines übermäßig hohen Belastungsaufwands einschränken (Czarnocki, 2021). Ergänzend dazu debattieren Ebers et al. (2021), dass eine zu breite Definition von KI-Systemen in der Verordnung die Gefahr einer Überregulierung birge. Sie argumentieren, dass die verwendete Definition nahezu jedes Computerprogramm erfassen könnte und Programmierer:innen, Betreiber:innen und Nutzer:innen von KI somit in Unsicherheit hinsichtlich der Rechtslage geraten könnten. Umgekehrt finden Ebers et al. (2021) jedoch, dass die Verordnung in anderen Bereichen nicht umfangreich genug sei. Für Hochrisikosysteme, wie z.B. Anwendungen in der Personalauswahl, könnten detailliertere Risikoklassifizierungen erforderlich sein, damit von Nutzer:innen eine angemessene Risikobewertung durchgeführt werden könne. Schließlich birge der weitergefasste Ansatz der Verordnung die Gefahr, dass nationale Bemühungen zur Regulierung von KI-Nutzung untergraben würden, ein Teil des bestehenden nationalen Schutzes digitaler Grundrechte abgeschwächt und künftige Regulierungsversuche beeinträchtigt würden (Veale & Zuiderveen Borgesius, 2021).

Die genannten Positionen bzw. Argumente zeigen schlaglichtartig einige der Aspekte auf, die gegenwärtig bei der Regulierung von KI diskutiert werden. Da die Verhandlungen zur KI-Verordnung derzeit noch andauern, ist es noch zu früh, seine Auswirkungen adäquat zu beurteilen; es ist jedoch eine interessante und wichtige Entwicklung, die sich erheblich auf die künftigen Entwicklungen von KI in der Europäischen Union auswirken wird.

5. Zusammenfassung

Vorangegangene Erörterungen in diesem Kurzdossier verdeutlichen das Potenzial der Nutzung von KI als Tool zur Reduzierung von Diskriminierung in der Personalauswahl. Es gibt aussagekräftige Forschungserkenntnisse, die dafür sprechen, dass eine KI die Vielfalt unter erfolgreichen Bewerber:innen erhöhen kann, während sie gleichzeitig insgesamt bessere Leistungen der Bewerber:innen durch hohe Passgenauigkeit zur Position bewirken kann. Allerdings unterstreichen vergangene Anwendungsfälle, wie der diskriminierende Rekrutierungsalgorithmus bei Amazon die Notwendigkeit eines umsichtigen Vorgehens. Wie beschrieben, beinhaltet derzeitiges KI-Training ein starkes Element menschlicher Interaktion. Ob es nun um die Pflege der durch KI angewandten Trainingsdaten oder das Labelling dieser Daten geht – Bias kann vom Menschen auf die KI übertragen wer-



den. Die KI kann infolgedessen Vorurteile verankern oder gar verstärken, wodurch Diskriminierung fortgesetzt oder sogar verschärft wird. Die Ergebnisse der überwachten Algorithmen gemäß Li et al. (2020) deuten auf diese Möglichkeit hin, wobei der leistungsstärkere Explorationsalgorithmus spezifisch auf die Auswahl einer weiteren Bewerber:innengruppe ausgerichtet ist. Dies unterstreicht die Relevanz von Designentscheidungen für die Leistungsfähigkeit von Algorithmen: Bei einem unzureichenden Design könnte eine KI bei Menschen unbewusst vorhandene Vorurteile reproduzieren und damit ihrem ursprünglichen Zweck objektiver Einstellungsentscheidungen entgegenwirken.

Wie die Forschung zu individueller Wahrnehmung von KI in der Personalauswahl zeigt, haben Personen zu Entscheidungen einer KI weniger Vertrauen, selbst wenn diese Entscheidungen ebenso fair oder noch fairer sind als menschliche Entscheidungen. Dieses Ergebnis betont die Notwendigkeit für Transparenz überall dort, wo KI Anwendung findet. Menschen neigen dazu, dem Unbekannten zu misstrauen, sodass wir zu dem Schluss gelangen, dass zur Erreichung größerer Akzeptanz für und Beteiligung an algorithmische(n) Auswahlverfahren, Transparenz

von entscheidender Bedeutung ist. Während die Erklärbarkeit von KI noch in den Kinderschuhen steckt, würde Transparenz auf Seiten der Benutzer:innen und Anbieter:innen von KI-Rekrutierungstools in Bezug auf die Designentscheidungen und Trainingsdaten auch dabei helfen, Vertrauen zwischen Menschen und Maschinen aufzubauen. Technologien entwickeln sich nicht deterministisch; KI wird sich nicht instinktiv auf mehr oder weniger Diskriminierung hin entwickeln. Innerhalb jedes technologischen Rahmens gibt es zahlreiche Entscheidungen, die in Bezug auf Design und Anwendung eines Artefakts getroffen werden müssen. KI-Technologien werden daher durch sorgfältig getroffene Entscheidungen transparenter und somit besser erklärbar, indem die Kriterien im Detail ausdrücklich dokumentiert und den Beteiligten zugänglich gemacht werden.

Dieses Kurzdossier beleuchtet die Rolle von KI in der Personalauswahl und einige ausschlaggebende Entscheidungen, die mit dieser Technologie getroffen werden können. Beim Entwerfen oder bei der Anwendung einer KI ist es von entscheidender Bedeutung, die Trainingsdaten, das auf die KI angewandte Trainingsverfahren und die angewandten Auswahlkriterien zu berücksichtigen.

Quellenangaben

- Acikgoz, Y., Davison, K.H., Compagnone, M., & Laske, M. (2020). Justice perceptions of artificial intelligence in selection. *International Journal of Selection and Assessment*, 28(4), 399–416. <https://doi.org/10.1111/ijsa.12306>
- Agrawal, A., Gans, J.S., & Goldfarb, A. (2019). Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction. *Journal of Economic Perspectives*, 33(2), 31–50. <https://doi.org/10.1257/jep.33.2.31>
- Allgemeines Gleichbehandlungsgesetz (AGG) idF vom 14.08.2006 (BGBl. I S. 1897) zuletzt geändert durch Artikel 1 des Gesetzes vom 23. Mai 2022 (BGBl. I S. 768). Abgerufen am 22. September 2022
- Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., & Benjamins, R. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities, and challenges toward responsible AI. *Information Fusion*, 58, 82–115.
- Banks, R.R., & Ford, R.T. (2008). (How) does unconscious bias matter: Law, politics, and racial inequality. *Emory LJ*, 58, 1053.
- Becker, G.S. (2010). *The economics of discrimination*. University of Chicago press.
- Bendick Jr, M., Jackson, C.W., & Reinoso, V.A. (1994). Measuring employment discrimination through controlled experiments. *The Review of Black Political Economy*, 23(1), 25–48. <https://doi.org/10.1007%2FBF02895739>
- Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, 94(4), 991–1013. <https://www.doi.org/10.1257/0002828042002561>
- Bogen, M., & Rieke, A. (2018). *Help wanted: An examination of hiring algorithms, equity, and bias*. Upturn. Abgerufen am 2. April 2022 von <https://www.upturn.org/work/help-wanted/>
- Cecere, G., Corrocher, N., & Jean, C. (2021). *Fair or Unbiased Algorithmic Decision-Making? A Review of the Literature on Digital Economics. A Review of the Literature on Digital Economics*. SSRN. <https://ssrn.com/abstract=3943389>
- Chamorro-Premuzic, T., & Akhtar, R. (2019). *Should Companies Use AI to Assess Job Candidates?* Harvard Business Review. Abgerufen am 26. April 2022 von <https://hbr.org/2019/05/should-companies-use-ai-to-assess-job-candidates>
- Cowgill, B. (2020). Bias and productivity in humans and algorithms: Theory and evidence from resume screening. *Columbia Business School, Columbia University*, 29. https://conference.iza.org/conference_files/MacroEcon_2017/cowgill_b8981.pdf
- Cowgill, B., Dell'Acqua, F., Deng, S., Hsu, D., Verma, N., & Chaintreau, A. (2020). Biased programmers? Or biased data? A field experiment in operationalizing ai ethics. *Proceedings of the 21st ACM Conference on Economics and Computation*, 679–681. <https://doi.org/10.1145/3391403.3399545>
- Cowgill, B., & Stevenson, M.T. (2020). Algorithmic social engineering. *AEA Papers and Proceedings*, 110, 96–100. <https://www.doi.org/10.1257/pandp.20201037>
- Cowgill, B., & Tucker, C.E. (2020). Algorithmic fairness and economics. *Columbia Business School Research Paper*. SSRN. <https://dx.doi.org/10.2139/ssrn.3361280>
- Czarnocki, J. (2021). *Good intentions, unintended consequences? How the Proposal for the AI Act might kick innovation out of the EU*. CITIP Blog. Abgerufen am 3. Mai 2022 von <https://www.law.kuleuven.be/citip/blog/good-intentions-unintended-consequences/>
- Ebers, M., Hoch, V. R. S., Rosenkranz, F., Ruschmeier, H., & Steinrötter, B. (2021). The European Commission's Proposal for an Artificial Intelligence Act-A Critical Assessment by Members of the Robotics and AI Law Society (RAILS). *J 2021, Vol. 4, Pages 589-603*, 4(4), 589–603. <https://doi.org/10.3390/J4040043>
- Europäische Kommission (2021). *Proposal for a Regulation of the European Parliament and of the Council. Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*. <https://artificialintelligenceact.eu/the-act/>
- Garfield, L. (2017). *A startup claims to have finally figured out how to get rid of bias in hiring with artificial intelligence*. Business Insider. Abgerufen am 28. April 2022 von <https://www.businessinsider.com/hiring-diversity-brain-games-artificial-intelligence-automation-2017-9>
- Gonzalez, M., Capman, J., Oswald, F., Theys, E., & Tomczak, D. (2019). “Where’s the I-O?” Artificial Intelligence and Machine Learning in Talent Management Systems. *Personnel Assessment and Decisions*, 5(3). <https://doi.org/10.25035/pad.2019.03.005>
- Gonzalez, M.F., Liu, W., Shirase, L., Tomczak, D.L., Lobbe, C.E., Justenhoven, R., & Martin, N.R. (2022). Allying with AI? Reactions toward human-based, AI/ML-based, and augmented hiring processes. *Computers in Human Behavior*, 130. <https://doi.org/10.1016/j.chb.2022.107179>
- Hoffman, M., Kahn, L.B., & Li, D. (2018). Discretion in hiring. *The Quarterly Journal of Economics*, 133(2), 765–800. <https://doi.org/10.1093/qje/qjx042>

- Internationale Arbeitsorganisation (1958) Übereinkommen 111: Übereinkommen über die Diskriminierung in Beschäftigung und Beruf idF vom 04.06.1958. Abgerufen am 22. November 2022
- Kaas, L., & Manger, C. (2012). Ethnic discrimination in Germany's labour market: A field experiment. *German Economic Review*, 13(1), 1–20. <https://doi.org/10.1111/j.1468-0475.2011.00538.x>
- Kleinberg, J., Ludwig, J., Mullainathan, S., & Rambachan, A. (2018). Algorithmic fairness. *AEA Papers and Proceedings*, 108, 22–27. <https://www.doi.org/10.1257/pandp.20181018>
- Köchling, A., & Wehner, M.C. (2020). Discriminated by an algorithm: A systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research*, 13(3), 795–848.
- Lee, M.K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1). <https://doi.org/10.1177%2F2053951718756684>
- Leslie, D. (2020). *Understanding Bias in Facial Recognition Technologies: An Explainer*. The Alan Turing Institute. <https://doi.org/10.5281/zenodo.4050457>
- Li, D., Raymond, L.R., & Bergman, P. (2020). *Hiring as exploration*. (NBER Working Paper Series, No. w27736) National Bureau of Economic Research. <https://doi.org/10.3386/w27736>
- McGinnity, F., Nelson, J., Lunn, P., & Quinn, E. (2009). *Discrimination in Recruitment: Evidence from a field experiment*. (Equality Research Series) Equality Authority; The Economic and Social Research Institute. Abgerufen am 15. Mai 2022 von <http://hdl.handle.net/10147/76641>
- McGowan, M.A., & Andrews, D. (2015). *Labour market mismatch and labour productivity: Evidence from PIAAC data*. OECD. <https://doi.org/10.1787/5js1pzx1r2kb-en>
- Mujtaba, D.F., & Mahapatra, N.R. (2019). Ethical considerations in ai-based recruitment. *2019 IEEE International Symposium on Technology in Society (ISTAS) Proceedings*. <https://www.doi.org/10.1109/istas48451.2019.8937920>
- Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453.
- OECD (2019). *Artificial Intelligence in Society*. <https://doi.org/10.1787/eedfee77-en>
- Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 469–481. <https://doi.org/10.1145/3351095.3372828>
- Scott, A. (2017). *Tech Leavers Study: A First-of-its-kind Analysis of why People Voluntarily Left Jobs in Tech*. Kapor Center for Social Impact.
- Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial Intelligence in Human Resources Management: Challenges and a Path Forward. *California Management Review*, 28. <https://doi.org/10.1177%2F0008125619867910>
- van den Broek, E., Sergeeva, A., & Huysman, M. (2021). When the Machine Meets the Expert: An Ethnography of Developing AI for Hiring. *MIS Quarterly*, 45(3), 1557–1580. <https://doi.org/10.25300/MISO/2021/16559>
- Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the Draft EU Artificial Intelligence Act. *Computer Law Review International*, 22(4), 97–112. <https://papers.ssrn.com/abstract=3896852>
- Windley, D. (2021). *Is AI The Answer To Recruiting Effectiveness?* Forbes. Abgerufen am 26. April 2022 von <https://www.forbes.com/sites/forbeshumanresourcescouncil/2021/06/16/is-ai-the-answer-to-recruiting-effectiveness/?sh=431b6ff2d7c7>

Diese Publikation wird im Rahmen der Öffentlichkeitsarbeit des Bundesministeriums für Arbeit und Soziales kostenlos herausgegeben. Sie darf weder von Parteien noch von Wahlbewerbern oder Wahlhelfern während des Wahlkampfes zum Zwecke der Wahlwerbung verwendet werden. Dies gilt für Europa-, Bundestags-, Landtags- und Kommunalwahlen. Missbräuchlich ist insbesondere die Verteilung auf Wahlveranstaltungen, an Informationsständen der Parteien sowie das Einlegen, Aufdrucken oder Aufkleben parteipolitischer Informationen oder Werbemittel. Untersagt ist gleichfalls die Weitergabe an Dritte zum Zwecke der Wahlwerbung. Unabhängig davon, wann, auf welchem Weg und in welcher

Anzahl diese Publikation dem Empfänger zugegangen ist, darf sie auch ohne zeitlichen Bezug zu einer bevorstehenden Wahl nicht in einer Weise verwendet werden, die als Parteinahme der Bundesregierung zugunsten einzelner politischer Gruppen verstanden werden könnte. Außerdem ist diese kostenlose Publikation – gleichzeitig wann, auf welchem Weg und in welcher Anzahl diese Publikation dem Empfänger zugegangen ist – nicht zum Weiterverkauf bestimmt.

Alle Rechte einschließlich der fotomechanischen Wiedergabe und des auszugsweisen Nachdrucks vorbehalten.

05/2022

www.denkfabrik-bmas.de



Bundesministerium
für Arbeit und Soziales