GENDER
OPEN
REPOSITORIUM

Algorithmische Entscheidungsfindung aus der Gleichstellungsperspektive – ein Balanceakt zwischen Gender Data Gap, Gender Bias, Machine Bias und Regulierung

Lütz, Fabian

2023

https://doi.org/10.25595/2961

Veröffentlichungsversion / published version
Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:

Lütz, Fabian: *Algorithmische Entscheidungsfindung aus der Gleichstellungsperspektive – ein Balanceakt zwischen Gender Data Gap, Gender Bias, Machine Bias und Regulierung*, in: Gender : Zeitschrift für Geschlecht, Kultur und Gesellschaft, Jg. 15 (2023) Nr: 1, 26–41. DOI: https://doi.org/10.25595/2961.

Erstmalig hier erschienen / Initial publication here: https://doi.org/10.3224/gender.v15i1.03

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY 4.0 Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu dieser Lizenz finden Sie hier:

https://creativecommons.org/licenses/by/4.0/deed.de

Terms of use:

This document is made available under a CC BY 4.0 License (Attribution). For more information see:

https://creativecommons.org/licenses/by/4.0/deed.en

GenderOpen – Repositorium für die Geschlechterforschung: www.genderopen.de

FREIE UNIVERSITÄT BERLIN

HUMBOLDT-UNIVERSITÄT ZU BERLIN

Fabian Lütz

Algorithmische Entscheidungsfindung aus der Gleichstellungsperspektive – ein Balanceakt zwischen Gender Data Gap, Gender Bias, Machine Bias und Regulierung

Zusammenfassung

Der Beitrag analysiert, inwieweit Algorithmen bei der Erreichung der Gleichstellungsziele hilfreich oder hinderlich sind, insbesondere unter Berücksichtigung der Phänomene des Gender Bias, Gender Data Gap und des *Machine Bias*. Am Beispiel von Rekrutierungsalgorithmen werden insbesondere die negativen Konsequenzen für die Gleichstellung von Männern und Frauen erläutert. Der Beitrag zeigt aber auch auf, inwieweit Algorithmen gezielt für die Erreichung von Gleichstellungszielen verwendet werden könnten, unter anderem zur Verfolgung positiver Maßnahmen und zur Aufdeckung von Diskriminierungen.

Schlüsselwörter
Algorithmen, Gender Bias, Gender Data Gap, Machine Bias, Rekrutierungsalgorithmen, Gleichstellung

Summary

Algorithmic decision-making from a gender equality perspective – a balancing act between the gender data gap, gender bias, machine bias and regulation

The article analyses to what extent algorithms enable or hinder the achievement of gender equality goals, especially considering the phenomena of gender bias, machine bias and the gender data gap. The, mostly negative, consequences for gender equality will be explained drawing on the example of recruitment algorithms. However, it will also be shown to what extent algorithms could be used specifically to achieve gender equality, for instance by using algorithms in positive action measures and to detect discrimination.

Keywords
algorithms, gender bias, gender data gap, machine bias, recruitment, gender equality

1 Einleitung

Auch wenn künstliche Intelligenz (KI) und Algorithmen die Welt nicht im Sinne einer Dystopie revolutionieren, wie manche AutorInnen dies prognostizieren (Boden 2016; Campbell 2022; Kissinger/Schmidt/Huttenlocher 2021; Lee 2021), steht die Gleichstellungspolitik vor neuen Herausforderungen. KI ist der Forschungszweig, der sich mit der Entwicklung von Computersystemen befasst, die Aufgaben erledigen können, die normalerweise menschliche Intelligenz erfordern (Kelleher 2019; Kleesiek et al. 2020). Ein Algorithmus ist eine „ausreichend detaillierte und systematische Handlungsanweisung, um ein mathematisches Problem zu lösen, so dass bei korrekter Implementierung (Übersetzung in Code) der Computer für jede korrekte Inputmenge den korrekten Output berechnet“ (Zweig 2019: 313). Diese Herausforderungen benötigen zwar kein vollständiges Umdenken in der Gleichstellungspolitik, denn die Probleme

Open Access © 2023 Autor*innen. Dieses Werk ist bei der Verlag Barbara Budrich GmbH erschienen und steht unter der Creative Commons Lizenz Attribution 4.0 International (CC BY 4.0).

der Offline-Welt bestehen auch in der Welt der Algorithmen (Dowek/Abitoul 2020). Jedoch haben die durch Algorithmen verursachten Gender-relevanten Probleme teilweise einen anderen Ursprung und können in veränderter Form oder Intensität bestehen (European Commission 2020, 2021a, 2021c). Der Einsatz von Algorithmen, z. B. bei der Rekrutierung oder zum algorithmischen Assessment in der Arbeitswelt (Crawford 2021), bringt neue Herausforderungen mit sich, die nicht zwingend mit den derzeitigen politischen und juristischen Maßnahmen handhabbar sind. Algorithmen sind heute in der Lage, Schwangerschaften und damit zusammenhängende Abwesenheiten vom Arbeitsplatz auf der Grundlage weniger Datenpunkte vorherzusagen (Aranda/Hagerty/Jemio 2022). Dieses Prognosepotenzial könnte problematisch sein, wenn Algorithmen die klassischen Arbeitszeitmodelle von Männern und Frauen für ihre Analyse verwenden (European Commission 2021), insbesondere Teilzeit und Abwesenheiten durch Mutterschutz und Elternzeit, wodurch das Risiko einer ungünstigeren algorithmischen Bewertung von Frauen erhöht wird. Um solche unfairen Ergebnisse des algorithmischen Entscheidungsprozesses für Frauen zu vermeiden, sollten reduzierte Arbeitszeiten oder Abwesenheiten vom Arbeitsplatz wie z. B. familienbedingte Teilzeit, Stillzeiten, Mutterschutz oder Elternzeit spezifisch gekennzeichnet werden (Crawford 2021).

Um Diskriminierungen auf dem Arbeitsmarkt entgegenzutreten, ist eine vollumfängliche Umsetzung der Work-Life-Balance-Richtlinie („WLB-RL“) in allen Mitgliedsstaaten der Europäischen Union unerlässlich, um gegen Diskriminierungen aufgrund der Inanspruchnahme von Vaterschaftsurlaub oder Elternzeit vorzugehen. Ein Diskriminierungsverbot für die in der WLB-RL verankerten Rechte für Eltern und Pflegende ist erstmals auf EU-Ebene explizit vorgesehen in Art. 11 der WLB-RL (European Commission 2019; Lütz 2021; Oliveira/de la Corte Rodríguez /Lütz 2020). Ungewiss ist, wie ein Algorithmus zur Bewertung der Arbeitsleistung mit der WLB-Problematik umgeht. Zwar hilft eine Kennzeichnung von spezifischen Abwesenheiten (Vaterschaftsurlaub, Mutterschutz, Elternzeit, Stillzeiten), damit diese zu erklärbaren Phänomenen werden. Dies beugt aber nicht automatisch Diskriminierungen vor. Insbesondere wegen der ungleichen Inanspruchnahme von Elternzeit und Teilzeitmodellen durch Männer und Frauen, bleibt ein Diskriminierungsrisiko bei der Bewertung der Arbeitsleistung und Rekrutierung durch Algorithmen bestehen. Die geltenden rechtlichen und politischen Instrumente können die Gleichstellungsproblematik von Algorithmen teilweise in den Griff bekommen, weshalb nur dort zu justieren ist, wo dies für die Reduzierung von Ungleichbehandlungen und Diskriminierungen notwendig ist.

Am Beispiel von Rekrutierungsalgorithmen wird erläutert, wo eine Anpassung des Rechtsrahmens und der politischen Instrumente notwendig ist, um die gleichstellungspolitischen Ziele zu erreichen.[1]

1 In Deutschland sind die gleichstellungspolitischen Ziele in Art. 3 Abs. 2 Grundgesetz (GG) sowie im Allgemeinen Gleichbehandlungsgesetz (AGG) vom 14. August 2006 (BGBl. I S. 1897), das zuletzt durch Artikel 8 des Gesetzes vom 3. April 2013 (BGBl. I S. 610) geändert worden ist, festgelegt. In der Schweiz regelt Art. 8 Abs. 3 der Bundesverfassung (BV) die Gleichberechtigung von Mann und Frau sowie das Bundesgesetz über die Gleichstellung von Männern und Frauen vom 24. März 1995 (Gleichstellungsgesetz, GlG; SR 151.1).

2 Hindernisse und Chancen von Algorithmen zur Erreichung der Gleichstellungsziele

Für die Frage, inwieweit Algorithmen bei der Erreichung der Gleichstellungsziele hilfreich oder hinderlich sind, müssen insbesondere die Phänomene Gender Bias, Gender Data Gap und Machine Bias untersucht werden.

2.1 Gender Bias

Gender Bias ist kein gleichstellungs- oder Algorithmus-spezifisches Phänomen, sondern für alle menschlichen Entscheidungen relevant (Perez 2019). Die Vorurteile der realen Welt spiegeln sich in den Datensätzen der Algorithmen wider und werden zur (undifferenzierten) Entscheidungsgrundlage. Inwieweit Bias die Entscheidung in der Lebens- und Arbeitswelt beeinflusst und was man dagegen unternehmen kann, wurde eingehend in der Literatur diskutiert, insbesondere aus dem Blickwinkel der Verhaltensökonomie (Bohnet 2016; Bohnet/Van Geen/Bazerman 2016). Unter einem Bias versteht man menschliche, systematische Entscheidungsfehler (Kahneman/Sibony/Sunstein 2021), wobei hier vor allem die Gender Biases relevant sind, wenngleich die KI-Problematik für andere Diskriminierungsgründe wie *race* (Eubanks 2018) und Disability (United Nations 2022) ebenfalls diskutiert wird (Achiume 2020). Unter Vorurteil versteht man gemeinhin ein Urteil über Personen, Gruppen oder eine Sachlage ohne sorgfältige Untersuchung, Abklärung und Abwägung. Vorurteile und Biases sind von Bedeutung, weil sie sich in den Datensätzen der Algorithmen und in der Programmarchitektur wiederfinden (Crawford 2013). Beispielsweise wäre denkbar, dass Algorithmen die geschützte Charakteristik „Geschlecht“ systematisch mit geringer bezahlten Jobs oder geringeren Gehältern in Verbindung bringen (Yeung/Lodge 2019). Sofern Bilder verwendet werden, um Algorithmen zu trainieren (Kay/Matuszek/Munson 2015; Russel/Selvaraj 2021; Steed/Caliskan 2021), perpetuieren sich Vorurteile im Hinblick auf Jobs und es entsteht ein Gendered Worldview (Crawford, 2021). Das von der EU geplante Gesetz über künstliche Intelligenz (GKI) sieht für Rekrutierungsalgorithmen „eine Untersuchung im Hinblick auf mögliche Verzerrungen (Bias)“ (Art. 10 Abs. 2f GKI) vor, um für gewisse Qualitätsstandards bei den Trainings-, Validierungs- und Testdatensätzen zu sorgen (European Commission 2021b). Die Problematik der Gender Biases ist folglich zentral für die Analyse und Ausgangspunkt für viele Diskriminierungen.

2.2 Gender Data Gap

Gender Biases können jedoch auch dann entstehen, wenn keine Daten vorhanden sind, weil sie nicht erhoben werden. Manchmal ist das Vorliegen von Daten entscheidend, um ein optimales und diskriminierungsfreies Funktionieren der Algorithmen zu gewährleisten. Diese Sachlage, dass für Frauen nicht genügend Daten gesammelt oder vorhanden sind, wird als Gender Data Gap bezeichnet (Buvinic/Levine 2016). Neben der Quantität ist auch die Qualität der Daten entscheidend für die Leistung von Algorithmen. Sind nicht genügend Daten einer Kategorie, die durch das Antidiskriminierungsrecht geschützt ist, vorhanden, so können Algorithmen zu ungerechten oder diskriminieren-

den Entscheidungen führen. Dem Rekrutierungsalgorithmus von Amazon beispielsweise fehlten genügend Daten von Frauen für die optimale Analyse von Bewerberinnen, was dazu führte, dass Männer für die offenen Stellen empfohlen wurden (Yam/Skorburg 2021). Die Literatur äußert sich teilweise kritisch über die Datenungleichheiten im Bereich der Algorithmusrekrutierung (Kraft-Buchman/Arian 2021). Da viele Daten dezentral produziert werden (z. B. durch soziale Medien), lassen sie sich jedoch nur schwierig regulieren, wodurch der Gender Data Gap nicht einfach zu korrigieren ist. Schließlich wird auch der Authority Gap (Sieghart 2021), also das Phänomen, dass Frauen insbesondere in der Arbeitswelt oft weniger professionell wahrgenommen werden als Männer, Einfluss auf die Zusammensetzung und den Inhalt der Daten haben. Für faire algorithmische Entscheidungsprozesse bedarf es also idealerweise repräsentativer Datensätze. Wo dies nicht möglich ist, sollte man sich der begrenzten Aussagekraft von Datensätzen mit mangelnder Diversität bewusst werden.

2.3 Machine Bias

Algorithmen sind entgegen mancher Einschätzung weder neutral noch objektiv (Kleinberg et al. 2018). Zwar kann Maschinen wohl (noch) nicht die Möglichkeit von Empfindungen, Empathie und Emotionen zugeschrieben werden, womit Algorithmen den Vorteil haben, nicht auf die gleiche Art und Weise wie Menschen beeinflusst zu werden. Da Algorithmen mithilfe von mathematisch programmierten Zielen arbeiten, sind sie zwar auf den ersten Blick *objektiver* als Menschen (Desai/Kroll 2017; Hosanagar 2020; LaValle 2006), jedoch beeinflussen Biases ebenfalls die algorithmische Entscheidung. Aufgrund von Gender Bias und Gender Data Gap tendieren Algorithmen zu sehr menschenähnlichen Entscheidungsfehlern, dem sog. Machine Bias (Fry 2018). Problematisch ist, dass nach derzeitiger Erkenntnis Gleichstellung oder Fairness nur schwierig in die Algorithmen programmiert werden können (Wachter/Mittelstadt/Russel 2021). Die Frage von menschlichen und maschinellen Biases wird auch dann relevant, wenn selbstlernende Algorithmen nicht von Menschen (beaufsichtigt) werden, sondern wenn die Maschine selbst (unbeaufsichtigt) lernt, z. B. bei Beschriftungen von Bildern im Internet. Hier entsteht also nicht etwa ein Bias des Menschen durch das „Labeln“ von Bildern, sondern der Algorithmus trägt möglicherweise die in den Daten entdeckten Biases und Vorurteile in die eigene Beschriftung mit ein (Steed/Caliskan 2021). Wie bereits dargelegt, arbeiten Algorithmen mit *verzerrten* Datensätzen. Somit ist der Entscheidungsprozess bei Algorithmen ähnlich verzerrt wie beim Menschen. Allerdings entscheiden Algorithmen bei gleichem Input stets gleich und unterliegen demnach nicht dem Entscheidungsfehler *Noise* (Kahneman/Sibony/Sunstein 2021), der bei Menschen häufig zu verschiedenen Ergebnissen bei mehrmaligen Entscheidungen führt. Dies könnte im Bereich der Rekrutierung nutzbar gemacht werden, um Einstellungsverfahren fairer zu gestalten und Diskriminierungen zu reduzieren.

3 Rekrutierungsalgorithmen und Gleichstellung

Rekrutierungsalgorithmen werden seit Jahren von verschiedenen Unternehmen angeboten (Strazzulla 2022) und von ca. 88 Prozent der großen Unternehmen verwendet (Wisenberg Brin 2019). Rekrutierungssoftware kann in verschiedenen Stufen des Auswahlverfahrens eingesetzt werden (O'Neil 2016), z. B. Sourcing, Vermittlung zukünftiger MitarbeiterInnen, Rekrutierung, CV-Parsing und automatische Vorauswahl (Textkernel 2022). Unternehmen werben mit reichhaltigen datenbasierten Arbeitsmarktinformationen und damit, dass ihre KI-Rekrutierungssoftware Unternehmen bei einer effizienteren Rekrutierung unterstützt (Textkernel 2018; Textkernel 2022). Das klingt zwar verlockend für die Personalabteilung, stellt aber gleichzeitig aufgrund der historischen Daten ein Einfallstor für Gender Bias und Diskriminierung dar. Trotz Effizienz dürfen Grundrechte wie die Geschlechtergleichstellung nicht vernachlässigt werden. KI-Unternehmen setzen auch vermehrt Gesichtserkennungssoftware (*Facial Recognition*) oder *Affect and emotion recognition*-Software ein, wie z. B. HireVue, Human, Emotient (Apple) and Affectiva (Buijsman/Jänicke 2021; Crawford 2021), was gleichstellungspolitisch problematisch ist, da diese Software häufig nicht gleich gut bei Frauen und Männern funktioniert.

Das GKI (European Commission 2021b) plädiert für Regulierung, da algorithmische Rekrutierungssysteme Nachteile insbesondere für Grundrechte haben könnten. Auch die Resolution der UN (United Nations 2021b) und der zugrunde liegende Bericht erwähnen das Risiko, welches mit algorithmischen Entscheidungen verbunden ist und welche Bedeutung eine Diskriminierung für das Leben der Menschen haben kann: „These systems [...] decide who has a chance to be recruited for a job" (United Nations 2021a: 57).

Biases und Stereotype in den Daten können diskriminierende Wirkungen auf die Entscheidungen des Rekrutierungsalgorithmus haben. Wenn z. B. eine KI-Rekrutierung für eine CEO-Position durchgeführt wird, so wird der Algorithmus alle verfügbaren Daten verwenden, um eine Entscheidung zu treffen. Es wäre dann nicht auszuschließen, dass z. B. eine Bildersuchmaschinenanfrage aufgrund des Gender Data Gap bei der Eingabe von „CEO" keine oder nur eine sehr geringe Anzahl von weiblichen CEO als Ergebnis anzeigt, was suggeriert, dass Frauen keine CEO sein können oder sind, obwohl die Statistiken mindestens 8,5 Prozent Frauen in Europa als CEO ermitteln (European Commission 2020). Für Unternehmen stehen meist die Präzision der KI bei der Personalauswahl und die Optimierung des Rekrutierungsprozesses durch KI im Vordergrund. Diskriminierungsfreie Entscheidungen haben oft nicht die Priorität, die sie haben sollten: „Machine learning systems are [...] constructing […] gender: they are defining the world within the terms they have set, and this has long-lasting ramifications for the people who are classified" (Crawford 2021: 146).

Aus gleichstellungsrechtlicher Perspektive ist die Verwendung von historischen Daten vor dem Hintergrund der bislang (noch) nicht erreichten und nur sehr langsam verwirklichten Gleichstellungsziele und der bestehenden Ungleichheiten zwischen Männern und Frauen problematisch (Berghahn 2004).

3.1 Das Problem der Algorithmus-basierten Personalauswahl

Die automatisierte Personalauswahl ist in der Unternehmenswelt etabliert (Hangartner/Kopp/Siegenthaler 2021; Knobloch/Hustedt 2019; Köchling/Wehner 2020) und mit ihr die Rolle und Macht der Algorithmen, welche tiefgreifende Entscheidungen mit möglicherweise diskriminierenden Wirkungen für Männer und Frauen treffen können (Lütz 2022).[2]

Problematisch bei Algorithmen kann insbesondere die Word2Vec-Technik sein (Buijsman/Jänicke 2021; Russel/Norvig 2021). Word2Vec stellt ein Eingangstor für Diskriminierungen dar, weil der Algorithmus Wahrscheinlichkeiten berechnet, wie eng verwandt ein Wort mit einem anderen ist und wie häufig dies im gleichen Kontext genannt wird. Dies kann sich negativ auf die Gleichstellung auswirken, weil sich in den Ergebnissen der Wortpaare Vorurteile und potenzielle Diskriminierungen wiederfinden. Ein viel diskutiertes Beispiel ist „Man is to computer programmer as woman is to homemaker?" (Bolukbasi et al. 2016: 1). Die historischen Daten, welche die Algorithmen unter anderem in digitalisierten Büchern, Zeitschriften und im Internet finden, enthalten Vorurteile und Biases. Zwar bringen Rekrutierungsalgorithmen auch Vorteile, weil Ansichten von ArbeitgeberInnen mit Vorurteilen mithilfe der KI teilweise objektiviert werden können. Algorithmen, insbesondere „deep learning"-Algorithmen (Kelleher 2019) oder „neural networks" (Schmidhuber 2015) können jedoch dazu führen, dass Vorurteile verstärkt werden. Ein besonderes Diskriminierungspotenzial besteht aus der Perspektive der Intersektionalität, also einer multiplen Diskriminierung, z. B. aufgrund des Geschlechts und der Religion oder Herkunft. Die Gefahr einer Diskriminierung ist deshalb so groß, weil die im Verborgenen operierenden Algorithmen auf der Basis der Datensätze mehrere potenzielle Ansatzpunkte für Diskriminierung haben und oft nicht klar ist, welches Element zu einer Diskriminierung führt. Die Funktionsweise von Algorithmen zu verstehen, ist daher zentral, um die Risiken für die Gleichstellung zu erkennen und eine adäquate Regulierung zu entwerfen. Während menschliche Entscheidungen erklärt werden können, besteht diese Möglichkeit bei Algorithmen nicht. Nachvollziehbarkeit und Erklärbarkeit von algorithmischen Entscheidungen sind unerlässliche Ziele, die unter dem Stichwort Transparenz (Esposito 2021; Hacker/Petkova 2017; Waltl/Vogl 2018) oder Erklärbarkeit (Hacker/Grundmann/Naumann 2020; Kamiran/Žliobaitė/Calders 2013; Rudin/Radin 2019; Waltl/Vogl 2018) diskutiert werden. Regulierungen wie der EU-GKI-Vorschlag zielen in diese Richtung (vgl. Kap. 4.1.2). Sie sind nicht nur sinnvoll, um algorithmische Entscheidungen in der Gesellschaft akzeptabel zu machen, sondern helfen auch potenziellen Opfern einer Diskriminierung, Beschwerde oder Klage einzureichen.

2 Auf europarechtlicher Ebene ist Richtlinie 2006/54/EC relevant, die für algorithmische Diskriminierung im Bereich von Rekrutierungsalgorithmen Anwendung findet (European Union 2006). Zu der Problematik, inwieweit algorithmische Diskriminierung vom geltenden Recht erfasst ist und welcher Anpassungsbedarf im EU-Recht besteht, vgl. Lütz (2022).

3.2 Möglichkeiten für *Positive Action* („positive Maßnahmen") bei der KI-Rekrutierung

Im EU-Recht erlaubt Art. 157 Abs. 4 AEUV[3] unter gewissen Voraussetzungen positive Maßnahmen im Bereich des Arbeitsmarktes, beispielsweise „zur Erleichterung der Berufstätigkeit des unterrepräsentierten Geschlechts oder zur Verhinderung bzw. zum Ausgleich von Benachteiligungen in der beruflichen Laufbahn" (Art. 157 Abs. 4 AEUV; vgl. dazu McCrudden 2019). Diese auf den Arbeitsmarkt abzielende Vorschrift könnte z. B. im Hinblick auf die oben aufgezeigten Risiken von Rekrutierungsalgorithmen positive Effekte für die Gleichstellung entfalten. Es wäre denkbar, im Sinne von *positive action through algorithms* Probleme der automatisierten Personalauswahl für die Gleichstellung zu lösen.

Erstens wäre es möglich, per Gesetz oder auf Initiative von Unternehmen dafür zu sorgen, dass automatisch eine bestimmte Anzahl von Bewerbungen des unterrepräsentierten Geschlechts bei der automatisierten CV-Auswahl oder Vorauswahl in die nächste Stufe des Verfahrens gelangt (Celis et al. 2021). Diese aus der Informatik bekannte „Rooney Rule" könnte insbesondere bei Rekrutierungsverfahren einen Implicit Bias verhindern, indem durch eine Art positive Maßnahme automatisch mindestens eine unterrepräsentierte Person in die Shortlist aufgenommen werden soll (Celis et al. 2021).

Dies könnte für ein KI-Unternehmen bedeuten, dass automatisch mehr Frauen bei der Vor- oder CV-Auswahl berücksichtigt werden und bei einem Kindergarten mehr Männer, um deren Chancen zu erhöhen. Dies ließe sich nicht nur aus gleichstellungspolitischen Gründen rechtfertigen, sondern auch aufgrund der nicht repräsentativen und verzerrten Datensätze, die häufig zu nachteiligen Entscheidungen gegenüber Frauen führen, wenn die Personalauswahl automatisiert ist (Dastin 2018). Positive Maßnahmen könnten demnach als Korrektiv für eine ungleiche bzw. ungerechte Datenlage dienen. Allerdings müssten die positiven Maßnahmen sowohl die Kriterien des EuGH als auch eventuell bestehende grundrechtliche Grenzen (McCrudden 2019) beachten, wie z. B. das im deutschen Grundgesetz verankerte Prinzip der Bestenauslese für den öffentlichen Dienst (Art. 33 Abs. 2 GG) oder die Unternehmensfreiheit (Art. 12 GG; Schiek 2000). Auch wenn die Ziele der Gleichstellung europarechtlich und grundrechtlich verankert sind, darf die Handlungsfreiheit der Unternehmen bei der KI-Rekrutierung nicht mehr als erforderlich und nicht unangemessen beschränkt werden. Im Einklang mit der EuGH-Rechtsprechung Badeck (C-158/97, 28.03.2000) und den Schlussanträgen von Generalanwalt Saggio (C-158/97, 10.06.1999) wird eine Regelung, die dem unterrepräsentierten Geschlecht Zugang zu Interviews erleichtert, als europarechtlich zulässig erachtet. Im Badeck-Fall enthielt das Hessische Landesrecht eine Regelung, die Frauen Zugang zu Jobinterviews ermöglichte, weil die Regelung den Zugang zu einem Interview als unerlässliche Maßnahme ansah, um erfolgreich zu sein (Schiek 2000). Solange eine positive Maßnahme beim Einsatz von Algorithmen nur darauf abzielt, das Errei-

3 „Im Hinblick auf die effektive Gewährleistung der vollen Gleichstellung von Männern und Frauen im Arbeitsleben hindert der Grundsatz der Gleichbehandlung die Mitgliedstaaten nicht daran, zur Erleichterung der Berufstätigkeit des unterrepräsentierten Geschlechts oder zur Verhinderung bzw. zum Ausgleich von Benachteiligungen in der beruflichen Laufbahn spezifische Vergünstigungen beizubehalten oder zu beschließen." (Art. 157 Abs. 4 AEUV)

chen der nächsten Stufe des Auswahlverfahrens, z. B. ein Interview, zu ermöglichen und gleichzeitig die Kompetenz der KandidatInnen geprüft wird, dürfte dies im Einklang mit der Rechtsprechung stehen (Schiek 2000).

Zweitens könnten Algorithmen gezielt eingesetzt werden, um mögliche Muster von Diskriminierungen aufzudecken (Kleinberg et al. 2020), indem sich Algorithmen mit bestimmten Profilen bewerben, um durch Modifizierung der Gender-relevanten Parameter herauszufinden, ob eine unterschiedliche Behandlung stattfindet. Damit stünde potenziellen Opfern einer Diskriminierung ein praktisches Tool zur Verfügung, das auch von NGOs oder Antidiskriminierungsstellen genutzt werden könnte.

Drittens hat die deutsche Bundesregierung kürzlich darauf hingewiesen, dass sie für Rekrutierungsalgorithmen eine Risikoanalyse empfiehlt (BMFSFJ 2021). Da aufgrund der stets gleichen Verarbeitung von Daten bei gleichem Input mit gleichem Output zu rechnen ist, könnten einige menschliche diskriminierende Verhaltensweisen ausgeschlossen werden, auch wenn sich Biases und Vorurteile aus den Datensätzen nicht vollständig eliminieren lassen. Es gibt gewichtige Stimmen in der Literatur, die von einer Minimierung der diskriminierenden Wirkung von Algorithmen bei der Rekrutierung überzeugt sind (Kleinberg et al. 2020).

4 Zwischen Regulierung und Nutzung von Algorithmen für Gleichstellung

Ausgangspunkt jeder Regulierung muss nicht nur der aktuelle Stand der Technik, sondern auch das geltende Recht sein. Die Flexibilität des geltenden Rechts entscheidet über die notwendige Art der Regulierung. Es gibt verschiedene Regulierungsansätze (z. B. eine Regulierungsmatrix (Zweig 2019)), die in Wissenschaft und Praxis diskutiert werden. Ein ambitionierter Vorschlag ist der Verordnungsentwurf GKI der Europäischen Kommission (European Commission 2021b), welcher eine Regulierung nach Risiken der KI vorsieht.

4.1 Regulierungsvorschläge

Es muss zwischen rechtlichen und politischen Maßnahmen unterschieden werden. Da rechtliche Vorschriften gerichtlich durchgesetzt werden und mehr Rechtssicherheit bieten, ist der Gesetzesform Vorrang zu geben. Während rein politische Maßnahmen häufig nicht die gewünschte Wirkung erzielen können, sind sie ideal im Zusammenspiel mit rechtlichen Regeln. Es gibt verschiedene Ideen und Elemente der Regulierung, die entweder in Gesetzesform gegossen oder aber als Empfehlungen für Unternehmen ausgesprochen werden können.

4.1.1 Nichtbindende Maßnahmen

Erstens ist an KI-Standards und Best Practices zu denken, die von Unternehmen selbst etabliert werden. Diese sind allerdings nur selbstbindend und im Falle einer Diskriminierung nicht gerichtlich geltend zu machen. Standards von nationalen oder internati-

onalen Standardsetzungsorganisationen (wie DIN, CEN-CENELEC, ISO) haben eine besondere Stellung, weil sie zunächst als nichtstaatliche Standards keine Rechtswirkung entfalten, jedoch rechtliche Wirkungen entfalten können, wenn nationale oder europäische Normen auf sie verweisen.

Zweitens wirkt sich das Erreichen von insgesamt mehr Gleichstellung positiv auf die Gesellschaft und die Datenwelt aus, was automatisch zu einer Verbesserung der algorithmischen Entscheidungen führt. Dies betrifft sowohl die Verabschiedung von neuen Gesetzen zur Gleichstellung, die stets auf positive Effekte abzielen, als auch politische Maßnahmen gegen Gender Bias und Diskriminierung.

Drittens wird mehr Diversität in KI-relevanten und MINT-Berufen nicht nur im Rahmen der europäischen Gleichstellungspolitik (European Commission 2020) gefordert, sondern findet auch Zustimmung in den UN-Berichten (United Nations 2021a, 2021b) und der Literatur (Leavy 2018).

Viertens werden Weiterbildung und andere werbende Maßnahmen in KI-Unternehmen von Akteuren wie dem Europarat gefordert, damit eine Vorstellung von Diversität und Gleichstellung herrscht, die sich nicht negativ auf das Programmieren auswirkt (Council of Europe 2020).

4.1.2 Rechtliche Maßnahmen

Das GKI (European Commission 2021b) hat zwar nicht explizit gleichstellungsspezifische Ziele, ist jedoch als komplementär zum geltenden Antidiskriminierungsrecht zu verstehen. Zudem wird die Thematik der „Nichtdiskriminierung" an mehreren Stellen des Vorschlags erwähnt und das GKI klassifiziert Rekrutierungsalgorithmen als Hochrisiko-KI-System, womit sie der Regulierung unterfallen. Randziffer 36 des GKI-Vorschlags hebt die Bedeutung für die Gleichstellung hervor:

> „KI-Systeme, die in den Bereichen Beschäftigung […] insbesondere für die Einstellung und Auswahl von Personen, für Entscheidungen über Beförderung und Kündigung sowie für die Zuweisung, Überwachung oder Bewertung von Personen in Arbeitsvertragsverhältnissen, sollten ebenfalls als hochriskant eingestuft werden, da diese Systeme die künftigen Karriereaussichten und die Lebensgrundlagen dieser Personen spürbar beeinflussen können. […] Solche Systeme können während des gesamten Einstellungsverfahrens und bei der Bewertung, Beförderung oder Nichtbeförderung von Personen in Arbeitsvertragsverhältnissen historische Diskriminierungsmuster fortschreiben, beispielsweise gegenüber Frauen." (European Commission 2021b: Rz. 36)

Solche Hochrisiko-KI-Systeme müssen gemäß Art. 6 und 8 GKI bestimmte Anforderungen erfüllen, wie z. B. die Einrichtung einer Daten-Governance (Art. 10), technische Dokumentations- (Art. 11) und Aufzeichnungspflichten (Art. 12), Transparenzpflichten (Art. 13) sowie menschliche Aufsicht (Art. 14). Aus gleichstellungsrechtlicher Perspektive können diese Vorschriften eine mögliche Beschwerde oder Klage erleichtern, weil z. B. der Zugang zu Informationen und Beweismitteln erleichtert wird und somit indirekt ein Beitrag für die Gleichstellung geleistet werden kann. Hervorzuheben ist hier die Rolle von technischen Normen von Organisationen wie CEN-CENELEC oder ISO, die im Rahmen der Standardisierung im Bereich von KI eine Rolle spielen können. Sofern das Prinzip der Diskriminierung hierdurch nicht umgangen und angetastet wird und es um rein technische Standardisierung und nicht um eine Konkretisierung von algorithmi-

scher Diskriminierung geht, um die Problematik der algorithmischen Diskriminierung handhabbar zu machen, ist dies nicht problematisch, solange ein Verweis auf solche technischen Standards gerichtlich überprüft werden kann.

Der Europarat hat eine Empfehlung zu den menschenrechtlichen Auswirkungen algorithmischer Systeme verabschiedet, um Grundrechtsverletzungen durch Algorithmen zu vermeiden oder abzumildern (Council of Europe 2020). Die Präambel fasst das Ziel prägnant zusammen: „Need to ensure that […] gender and other societal and labour force imbalances that have not yet been eliminated from our societies are not deliberately or accidentally perpetuated through algorithmic systems". In § 2.2 dieser Empfehlung wird die Bedeutung der erwähnten Datenqualität angesprochen: „States should carefully assess what human rights may be affected as a result of the nature of the data, which may stand in as a proxy for classifiers such as gender". Durch öffentlichkeitswirksame Maßnahmen zielt die Empfehlung in § 7.4 auch auf private Akteure ab, die KI-Systeme entwickeln und verwenden: „private actors […] should engage in and promote targeted […] gender-sensitive efforts to promote public awareness of relevant rights, including in particular, information about applicable complaint mechanisms and procedures."

Die OECD hat ebenfalls eine Empfehlung zu KI und Grundrechten vorgelegt (OECD 2019). Sie fordert in § 1.1: „Stakeholders should proactively engage in responsible stewardship of trustworthy AI in pursuit of beneficial outcomes for people […], such as […] advancing inclusion of underrepresented populations, reducing economic, social, gender and other inequalities". § 1.2 ruft Menschenrechte und Werte in Erinnerung: „AI actors should respect the rule of law, human rights and democratic values, throughout the AI system lifecycle. These include […] non-discrimination and equality, diversity, fairness, social justice, and internationally recognised labour rights". Obwohl die OECD-Empfehlung als rechtliches Dokument klassifiziert ist, handelt es sich nicht um zwingendes Recht. Dennoch ist die potenzielle Wirkung nicht zu unterschätzen, weil sich die Empfehlung an OECD-Staaten und Unternehmen richtet, die maßgeblich im Bereich der KI aktiv sind.

Auf UN-Ebene schließlich gibt es sowohl im Bereich der Gleichstellung (United Nations 2021a, 2021b) als auch im Bereich Behinderung (United Nations 2022) Berichte und Entschließungen, die das Zusammenspiel von KI und Diskriminierung ausführlich diskutieren. Der Bericht der UN-Hochkommissarin für Menschenrechte A/HRC/48/31 z. B. warnt in § 4: „advances in new technologies must not be used to erode human rights, deepen inequality or exacerbate existing discrimination" (United Nations 2021a). Auch deutet § 36 darauf hin, dass „research has demonstrated a disturbing degree of gender […] bias in Google's search results" und erinnert in § 49: „Assessing human rights impacts is an essential element of human rights due diligence processes. […] Particular attention should be paid to disproportionate impacts on women and girls".

Alle genannten rechtlichen und quasi-rechtlichen Vorschläge der verschiedenen Institutionen beinhalten sinnvolle Regulierungsideen für Gleichstellungspolitik, welche vom nationalen Recht übernommen und konkretisiert werden könnten. Um einen einheitlichen Schutz vor Diskriminierung zu gewährleisten, sollte Regulierung für Algorithmen als globales Problem auf europäischer oder internationaler Ebene verstanden werden.

4.1.3 *Discrimination aware Data-Mining* (DADM) und Folgenabschätzung für Algorithmen

Die Bedeutung der Datengrundlage für algorithmische Entscheidungen und das mögliche Risiko einer Diskriminierung lassen sich teilweise auf der Ebene des Datensammelns vermindern, indem *Discrimination-aware-Data-Mining*-(DADM-)Methoden verwendet werden (Berendt/Preibusch 2014, 2017; Pedreshi/Ruggieri/Turini 2008). Problematisch ist allerdings, dass durch das Weglassen einer geschützten Charakteristik (Geschlecht) nicht unbedingt eine Diskriminierung verhindert wird, weil diese auf Umwegen durch Korrelation von verschiedenen Datenpunkten entstehen kann. Man spricht von Proxies, die Rückschlüsse auf geschützte Charakteristika wie Geschlecht zulassen (Veale/Binns 2017). Deshalb wird teilweise in der Literatur vorgeschlagen, DADM-Methoden für die Entdeckung und Kennzeichnung von möglichen Diskriminierungsmustern zu verwenden (sog. exploratory DADM) und nicht für das gezielte Sammeln von Daten (Berendt/Preibusch 2014).

Viele Autoren plädieren für Folgenabschätzung (sog. impact assessments) für Algorithmen, die bereits in vielen Bereichen mit Auswirkungen auf Menschenrechte (Kemp/Vanclay 2013) Standard sind (Dawson/Nonnecke 2021; Mantelero 2018; Yam/Skorburg 2021). Das Auditing von Algorithmen[4] stellt eine weitere Möglichkeit dar, um Biases und Diskriminierungspotenziale zu entdecken, bevor diese Algorithmen eingesetzt werden. Auditing kann entweder freiwillig von Unternehmen durchgeführt oder gesetzlich vorgeschrieben werden. Auch sog. Gender Impact Assessments werden positiv diskutiert (Peletz/Hanna 2019; Verloo/Roggeband 1996). Folgenabschätzung, exploratory DADM-Methoden und Auditing für Algorithmen scheinen sinnvoll, wenn diese gesetzlich festgeschrieben und nicht nur fakultativ von Unternehmen durchzuführen sind.

4.2 Algorithmen und bessere Durchsetzung des Antidiskriminierungsrechts

Ein effizienterer Ressourceneinsatz auf staatlicher Seite kann mithilfe von Algorithmen erreicht werden. Hierdurch könnten mehr Fälle bearbeitet werden, Muster und Parallelen zwischen ähnlich gelagerten Fällen gefunden werden, um schneller Lösungen zu finden. Aufgrund der hohen Anzahl an Beschwerden und begrenzter Ressourcen der Behörden (was sich vermutlich im Bereich des Gleichstellungsrechts auf absehbare Zeit nicht ändern wird, auch aufgrund vermehrter Regelungen, die umgesetzt und durchgesetzt werden müssen), könnte KI nicht nur Unternehmen, sondern auch Behörden sinnvoll unterstützen, um eine effektivere Rechtsdurchsetzung im Bereich der Gleichstellung zu gewährleisten (Kleinberg et al. 2020). Hierbei darf allerdings die Expertise von Menschen nicht ersetzt werden. Algorithmen sollten unterstützend eingesetzt werden und Menschen sollten die Kontrolle behalten, wie dies in Art. 14 EU GKI formuliert wird (Pasquale 2020).

4 Vgl. https://orcaarisk.com, ein von Cathy O'Neil gegründetes Auditing-Unternehmen für Algorithmen.

5 Schlussfolgerungen und Ausblick

Abschließend lässt sich festhalten, dass KI und Algorithmen die Gleichstellungspolitik vor neue Herausforderungen stellen. Mit den richtigen Instrumenten lassen sich die aufgezeigten Probleme und Risiken der Diskriminierung teilweise lösen oder minimieren. Auch birgt die KI nicht nur Gefahren für Gleichstellung, sondern es ergeben sich auch Chancen, die es zu nutzen gilt, z. B. im Rahmen positiver Maßnahmen, die gezielt zur Förderung der Gleichstellung eingesetzt werden können. Auch ließe sich im Sinne von „hybrid human-AI decisions" (Kissinger et al. 2021) für eine sinnvolle Kooperation zwischen Mensch und Maschine argumentieren, die im Bereich der Rekrutierung beim Einsatz von Algorithmen mehr Gleichstellung erreichen kann, sofern man sich der Unzulänglichkeiten beider Entscheidungssysteme bewusst ist. Durch die zunehmende Verbreitung von KI, z. B. bei der automatisierten Rekrutierung und den Risiken für Diskriminierungen aufgrund des Geschlechts, sind eine breite politische Diskussion und die Etablierung von Regulierung unerlässlich.

Literaturverzeichnis

Achiume, Tendayi (2020). *Racial discrimination and emerging digital technologies: Report of the Special Rapporteur on Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance.* Report of the Special Procedure of the Human Rights Council A/HRC/44/57. Zugriff am 08. Dezember 2022 unter https://digitallibrary.un.org/record/3879751.

Aranda, Florencia; Hagerty, Alexia & Jemio, Diego (2022). The Case of the Creepy Algorithm That 'Predicted' Teen Pregnancy. *Wired*, 16.02.2022. Zugriff am 08. Dezember 2022 unter https://www.wired.com/story/argentina-algorithms-pregnancy-prediction/.

Berendt, Bettina & Preibusch, Sören (2014). Better decision support through exploratory discrimination-aware data mining: foundations and empirical evidence. *Artificial Intelligence and Law*, *22*(2), 175−209. https://doi.org/10.1007/s10506-013-9152-0

Berendt, Bettina & Preibusch, Sören (2017). Toward accountable discrimination-aware data mining: the Importance of keeping the human in the loop—and under the looking glass. *Big data, 5*(2), 135−152. https://doi.org/10.1089/big.2016.0055

Berghahn, Sabine (2004). Der Ritt auf der Schnecke – Rechtliche Gleichstellung in der Bundesrepublik Deutschland. In Mechtild Koreuber & Ute Mager (Hrsg.), *Recht und Geschlecht: Zwischen Gleichberechtigung, Gleichstellung und Differenz* (S. 60–79). Nomos. https://doi.org/10.5771/9783845258645_60

Boden, Margaret A. (2016). *AI: Its nature and future*. Oxford: Oxford University Press.

Bohnet, Iris (2016). *What works*. Boston: Harvard University Press.

Bohnet, Iris; Van Geen, Alexandra & Bazerman, Max (2016). When performance trumps gender bias: Joint vs. separate evaluation. *Management Science, 62*(5), 1225−1234.

Bolukbasi, Tolga; Chang, Kai-Wei; Zou, James; Saligrama, Venkatesh & Kalai, Adam (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *arXiv,* 1607.06520. https://doi.org/10.48550/arXiv.1607.06520

Buijsman, Stefan & Jänicke, Bärbel (2021). *Ada und die Algorithmen: wahre Geschichten aus der Welt der künstlichen Intelligenz*. München: C. H. Beck.

Bundesministerium für Familie, Senioren, Frauen und Jugend (BMFSFJ) (2021). *Dritter Gleichstellungsbericht. Stellungnahme der Bundesregierung*. Zugriff am 08. Dezember 2022 unter

https://www.bmfsfj.de/resource/blob/182020/a9013da06f712b5d431b36aaaea39359/stellung- nahme-bundesregierung-gleichstellungsbericht-data.pdf.
Buvinic, Mayra & Levine, Ruth (2016). Closing the gender data gap. *Significance, 13*(2), 34−37.
Campbell, Catriona (2022). *AI by Design: A Plan for Living with Artificial Intelligence*. Boca Raton: CRC Press.
Celis, L. Elisa; Hays, Chris; Mehrotra, Anay & Vishnoi, Nisheeth K. (2021). The Effect of the Rooney Rule on Implicit Bias in the Long Term. In *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (S. 678–689). https://doi.org/10.1145/3442188.3445930
European Commission (2019). *Richtlinie (EU) 2019/1158 des Europäischen Parlaments und des Rates vom 20. Juni 2019 zur Vereinbarkeit von Beruf und Privatleben für Eltern und pflegende Angehörige und zur Aufhebung der Richtlinie 2010/18/EU des Rates.*
European Commission (2020). *Striving for a Union of equality: Gender equality strategy 2020–2025*. Publications Office of the European Union. https://doi.org/10.2775/671326
European Commission (2021a). *Algorithmic discrimination in Europe – Challenges and opportunities for gender equality and non-discrimination law.* Publications Office of the European Union. https://doi.org/10.2838/544956
European Commission (2021b). *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*. Brüssel, COM(2021) 206 final.
European Commission (2021c). *2021 report on gender equality in the EU*. Directorate-General for Justice and Consumers, Publications Office of the European Union. https://doi.org/doi/10.2838/57887
European Union (2006). Richtlinie 2006/54/EG des Europäischen Parlaments und des Rates vom 5. Juli 2006 zur Verwirklichung des Grundsatzes der Chancengleichheit und Gleichbehandlung von Männern und Frauen in Arbeits- und Beschäftigungsfragen (Neufassung), *OJ L 204*, 26.7.2006, 23–36.
Council of Europe (2020). *Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems.* Zugriff am 08. Dezember 2022 unter https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154.
Crawford, Kate (2013). The hidden biases in big data. *Harvard business review*, 01.04.2013. Zugriff am 08. Dezember 2022 unter https://hbr.org/2013/04/the-hidden-biases-in-big-data.
Crawford, Kate (2021). *The Atlas of AI.* New Haven: Yale University Press.
Dastin, Jeffrey (2018). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*, 11.10.2018. Zugriff am 08. Dezember 2022 unter https://reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G.
Dawson, Philip & Nonnecke, Brandie (2021). *Human Rights Implications of Algorithmic Impact Assessments – Priority Considerations to Guide Effective Development and Use.* Carr Center Discussion Paper Series. Carr Center for Human Rights Policy, Harvard Kennedy School: Harvard University.
Desai, Deven R. & Kroll, Joshua A. (2017). Trust but verify: A guide to algorithms and the law. *Harvard Journal of Law & Technology, 31*(1), 2–64.
Dowek, Serge & Abitoul, Gilles (2020). *The Age of Algorithms.* Cambridge: Cambridge University Press.
Esposito, Elena (2021). Transparency versus explanation: The role of ambiguity in legal AI. *Journal of Cross-disciplinary Research in Computational Law, 1*(1), 1−17.
Eubanks, Virginia (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. New York: St. Martin's Press.
Fry, Hannah (2018). *Hello World: How to be Human in the Age of the Machine*. London: Random House.

Hacker, Philipp; Krestel, Ralf; Grundmann, Stefan & Naumann, Felix (2020). Explainable AI under contract and tort law: legal incentives and technical challenges. *Artificial Intelligence and Law, 28*, 415−439. https://doi.org/10.1007/s10506-020-09260-6

Hacker, Philipp & Petkova, Bilyana (2017). Reining in the big promise of big data: Transparency, inequality, and new regulatory frontiers. *Northwestern Journal of Technology & Intellectual Property, 15*(1), 6–42.

Hangartner, Dominik; Kopp, Daniel & Siegenthaler, Michael (2021). Monitoring hiring discrimination through online recruitment platforms. *Nature, 589*(7843), 572−576.

Hosanagar, Kartik (2020). *A human's guide to machine intelligence: how algorithms are shaping our lives and how we can stay in control.* London: Penguin Books.

Kahneman, Daniel; Sibony, Olivier & Sunstein, Cass R. (2021). *Noise: a flaw in human judgment*. Boston: Little Brown.

Kamiran, Faisal; Žliobaitė, Iindre & Calders, Toon (2013). Quantifying explainable discrimination and removing illegal discrimination in automated decision making. *Knowledge and information systems, 35*(3), 613−644.

Kay, Matthew; Matuszek, Cynthia & Munson, Sean A. (2015). *Unequal representation and gender stereotypes in image search results for occupations.* Proceedings of the 33rd annual acm conference on human factors in computing systems. https://doi.org/10.1145/2702123.2702520

Kelleher, John D. (2019). *Deep learning*. Boston: MIT Press.

Kemp, Deana & Vanclay, Frank (2013). Human rights and impact assessment: clarifying the connections in practice. *Impact Assessment and Project Appraisal, 31*(2), 86−96.

Kissinger, Henry A; Schmidt, Eric & Huttenlocher, Daniel (2021). *The Age of AI: And Our Human Future*. Boston: Little Brown.

Kleesiek, Jens; Murray, Jacob; Strack, Christian; Kaissis, Georgios & Braren, Rickmer (2020). Wie funktioniert maschinelles Lernen? *Der Radiologe, 60*(1), 24−31. https://doi.org/10.1007/s00117-019-00616-x

Kleinberg, Jon; Lakkaraju, Himabindu; Leskovec, Jure; Ludwig, Jens & Mullainathan, Sendhil (2018). Human decisions and machine predictions. *The quarterly journal of economics, 133*(1), 237−293.

Kleinberg, Jon; Ludwig, Jens, Mullainathan, Sendhil & Sunstein, Cass R. (2020). Algorithms as discrimination detectors. *Proceedings of the National Academy of Sciences, 117*(48), 30096−30100. https://doi.org/10.1073/pnas.1912790117

Knobloch, Tobias & Hustedt, Carla (2019). *Der maschinelle Weg zum passenden Personal.* Gütersloh: Bertelsmann Stiftung.

Köchling, Alina & Wehner, Marius Claus (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. https://doi.org/10.1007/s40685-020-00134-w

Kraft-Buchman, Caitlin & Arian, Renée (2021). *Artificial Intelligence Recruitment: Digital Dream or Dystopia of Bias?* Zugriff am 08. Dezember 2022 unter https://womenatthetable.net.

LaValle, Steven M. (2006). *Planning algorithms*. Cambridge: Cambridge University Press.

Leavy, Susan (2018). *Gender bias in artificial intelligence: The need for diversity and gender theory in machine learning.* Proceedings of the 1st international workshop on gender equality in software engineering. https://doi.org/10.1145/3195570.3195580

Lee, Kai-Fu (2021). *AI 2041 – Ten Visions for the Future.* New York City: Currency.

Lütz, Fabian (2021). Die Väterbeteiligung in Europa und der Schweiz – Die Rolle der Väter für mehr Gleichberechtigung. *sui generis*, 147–156. https://doi.org/10.21257/sg.178

Lütz, Fabian (2022). Gender equality and artificial intelligence in Europe. Addressing direct and indirect impacts of algorithms on gender-based discrimination. *ERA Forum, 23,* 33–52. https://doi.org/10.1007/s12027-022-00709-6

Mantelero, Alessandro (2018). AI and Big Data: A blueprint for a human rights, social and ethical impact assessment. *Computer Law & Security Review*, *34*(4), 754−772. https://doi.org/10.1016/j.clsr.2018.05.017

McCrudden, Christopher (2019). *Gender-based positive action in employment in Europe: a comparative analysis of legal and policy approaches in the EU and EEA.* Publications Office of the European Union. https://doi.org/10.2838/770680

O'Neil, Cathy (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York: Crown.

Organisation for Economic Cooperation and Development (OECD) (2019). *OECD/LEGAL/0449*, Recommendation of the Council on Artificial Intelligence 2019.

Oliveira, Álvaro; de la Corte Rodríguez, Miguel & Lütz, Fabian (2020). The New Directive on Work-Life Balance: Towards a New Paradigm of Family Care and Equality? *European law review,* (3), 295−323.

Pasquale, Frank (2020). *New Laws of Robotics: Defending Human Expertise in the Age of AI*. Cambridge: Belknap Press.

Pedreshi, Dino; Ruggieri, Salvatore & Turini, Franco (2008). *Discrimination-aware data mining.* Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. https://doi.org/10.1145/1401890.1401959

Peletz, Nicole & Hanna, Kevin (2019). *Gender Analysis and Impact Assessment: Canadian and International Experiences*. Zugriff am 08. Dezember 2022 unter https://ok-cear.sites.olt.ubc.ca/files/2019/07/FINAL_WEB_Gender_Analysis_-Impact_Assessment.pdf.

Perez, Caroline Criado (2019). *Invisible women: Exposing data bias in a world designed for men*. London: Random House.

Rudin, Cynthia & Radin, Joanna (2019). Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition. *Harvard Data Science Review*, *1*(2), 1–10.

Russel, Newlin Shebiah & Selvaraj, Arivazhagan (2021). Gender discrimination, age group classification and carried object recognition from gait energy image using fusion of parallel convolutional neural network. *IET IMAGE PROCESS*, *15*(1), 239–251. https://doi.org/10.1049/ipr2.12024

Russel, Stuart & Norvig, Peter (2021). *Artificial intelligence: a modern approach* (4. überarb. Aufl.). London: Pearson Education.

Schiek, Dagmar (2000). Positive Action before the European Court of Justice – New Conceptions of Equality in Community Law? From *Kalanke* and *Marschall* to *Badeck*. *International Journal of Comparative Labour Law and Industrial Relations, 16*(3), 251−275. https://doi.org/10.54648/271016

Schmidhuber, Jürgen (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Sieghart, Mary Ann (2021). *The Authority Gap: Why women are still taken less seriously than men, and what we can do about it*. London: Random House.

Steed, Ryan & Caliskan, Aylin (2021). *Image Representations Learned With Unsupervised Pre-Training Contain Human-like Biases.* Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event, Canada. https://doi.org/10.1145/3442188.3445932

Strazzulla, Phil (2022). The Top 12 Best AI Recruiting Tools – 2022. An unbiased look at the top AI recruiting software along with key use cases, pitfalls and more. *Select Software Reviews*. Zugriff am 08. Dezember 2022 unter https://www.selectsoftwarereviews.com/buyer-guide/ai-recruiting.

Textkernel (2018). Extract! 4.0 – Textkernel Launches the First Fully Deep Learning Powered Resume Parsing Solution. *Cision PR Newswire,* 08.02.2018. Zugriff am 08. Dezember 2022

unter https://prnewswire.com/news-releases/extract-40---textkernel-launches-the-first-fully-deep-learning-powered-resume-parsing-solution-673303343.html.
Textkernel (2022). *Textkernel staffing*. Zugriff am 08. Dezember 2022 unter https://www.textkernel.com/staffing/.
United Nations (2021a). *The right to privacy in the digital age*. Report of the United Nations High Commissioner for Human Rights, A/HRC/48/31.
United Nations (2021b). Resolution A/HRC/48/31.
United Nations (2022). *Rights of persons with disabilities.* Report of the Special Rapporteur on the rights of persons with disabilities, A/HRC/49/52, U.N.H.R.C. *A/HRC/49/52.* Zugriff am 08. Dezember 2022 unter https://documents-dds-ny.un.org/doc/UNDOC/GEN/G21/397/00/PDF/G2139700.pdf.
Veale, Michael & Binns, Reuben (2017). Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society*, *4*(2). https://doi.org/10.1177/2053951717743530
Verloo, Mieke & Roggeband, Connie (1996). Gender impact assessment: The development of a new instrument in the Netherlands. *Impact Assessment*, *14*(1), 3−20.
Wachter, Sandra; Mittelstadt, Bernd & Russell, Chris (2021). Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. *Computer Law & Security Review*, *41*, Article 105567. https://doi.org/10.1016/j.clsr.2021.105567
Waltl, Bernhard & Vogl, Roland (2018). Increasing Transparency in Algorithmic Decision-Making with Explainable AI. *Datenschutz und Datensicherheit*, *42*(10), 613−617. https://doi.org/10.1007/s11623-018-1011-4
Wisenberg Brin, Dinah (2019). Employers Embrace Artificial Intelligence for HR. *SHRM.org*, 22.03.2019. Zugriff am 08. Dezember 2022 unter https://shrm.org/resourcesandtools/hr-topics/global-hr/pages/employers-embrace-artificial-intelligence-for-hr.aspx.
Yam, Josephine & Skorburg, Joshua A. (2021). From human resources to human rights: Impact assessments for hiring algorithms. *Ethics and Information Technology*, *23*(4), 611−623. https://doi.org/10.1007/s10676-021-09599-7
Yeung, Karen & Lodge, Martin (2019). *Algorithmic regulation*. Oxford: Oxford University Press.
Zweig, Katharina (2019). *Ein Algorithmus hat kein Taktgefühl: Wo künstliche Intelligenz sich irrt, warum uns das betrifft und was wir dagegen tun können*. München: Heyne.

Zur Person

Fabian Lütz, Ass.iur., Maître en droit (Paris), LL.M. (Bruges) – Universität Lausanne (UNIL). Arbeitsschwerpunkte: EU-Recht, Gleichstellungs- und Antidiskriminierungsrecht, Work-Life-Balance, Regulierung von Algorithmen.
E-Mail: fabian.luetz@unil.ch