

Where bias can creep in

Gendersensibilität beim Einsatz von KI-Technologien im Kontext beruflicher Weiterbildung im Forschungsprojekt KIRA

Kerstin Raudonat,¹ Monika Pröbster,¹ Kai Schmieder,² Simone Martinetz,² Nicola Marsden¹

Abstract: KI-Technologien können im Kontext beruflicher Weiterbildung den gezielten Kompetenzaufbau für künftige Berufsfelder unterstützen. Gleichwohl besteht die Gefahr, dass existierende Vergeschlechtlichungen im beruflichem Kompetenzaufbau reproduziert werden. Im Forschungsprojekt KIRA sollen berufliche Lernangebote KI-unterstützt auf individuelle Bedürfnisse zugeschnitten und passgenau vermittelt werden. Dabei soll schon während des Entwicklungsprozesses der Fortschreibung existierender Stereotypen entgegengewirkt werden. Hierzu werden mögliche Diskriminierungen in einem zirkulären Modell der Ko-Konstruktion konzeptionalisiert und so Ansatzpunkte für Reflexion und Intervention identifiziert. Im Beitrag werden vier Felder entlang dieses Kreislaufes möglicher Diskriminierung beleuchtet und es wird aufgezeigt, dass KI-gestützte Empfehlungen zur Kompetenzentwicklung nur mit einer ganzheitlichen Betrachtung wirklich fair sein können.

Keywords: Geschlechtergerechtigkeit; berufliche Weiterbildung; Kompetenzen; KI

1 Einleitung

Ethik und Gleichstellung spielen eine wesentliche Rolle beim Einsatz von künstlicher Intelligenz (KI): Gerade wenn es darum geht, auf Basis von KI Empfehlungen für grundlegende Lebensentscheidungen zu geben, sollte sichergestellt sein, dass sich keine Verzerrungen (Biases) auf Basis von Stereotypen in KI-Technologien materialisieren. Im Rahmen des Forschungsprojekts „KI-gestütztes Matching individueller und arbeitsmarktbezogener Anforderungen für die berufliche Weiterbildung“ (KIRA) sollen Empfehlungen für den Bereich der beruflichen Weiterbildung mittels KI ermöglicht werden. Es gilt, Gendersensibilität im Forschungs- und Entwicklungsprozess sicherzustellen. Dies betrifft einerseits das Thema Datensets, das für KI-Technologien von wesentlicher Bedeutung ist, und andererseits Themen aus dem Bereich beruflicher Weiterbildung wie beispielsweise vergeschlechtlichte Kompetenzzuschreibungen, Verzerrungen bei der Berufs- und Personalwahl sowie der Berufs- und Weiterbildungsberatung. Im Folgenden wird das Projekt KIRA skizziert und es

¹ Hochschule Heilbronn, Sozioinformatik, Max-Planck-Str. 39, 74081 Heilbronn vorname.nachname@hs-heilbronn.de

² Forschungs- und Innovationszentrum Kognitive Dienstleistungssysteme (KODIS), Fraunhofer IAO, Bildungscampus 9, 74076 Heilbronn vorname.nachname@iao.fraunhofer.de

werden kritische Stellen im Forschungs- und Entwicklungsprozess identifiziert, an denen Verzerrungen ihren Weg in eine KI-basierte Lösung finden können (zur Bedeutung von Datensets und den Konsequenzen von Verzerrungen s. z.B. [BS16; Hu21]).

2 Das Projekt KIRA

Die digitale Transformation bringt nicht nur technologische Veränderungen mit sich, sondern fordert auch berufliche Weiterbildung und Qualifizierung von Mitarbeiter*innen für neue Tätigkeitsfelder. Im Verbundprojekt KIRA untersucht ein Konsortium an Forschungs- und Praxispartner*innen, wie die berufliche Weiterbildung datengestützt und mit künstlicher Intelligenz (KI) neu gestaltet werden kann. Dabei werden die datentechnischen Voraussetzungen für eine Individualisierung von beruflichen Lernangeboten mit Bezug auf unternehmens- bzw. arbeitsmarktpolitische Erfolgsmerkmale in den Fokus gerückt. Berufliche Lernangebote sollen auf individuelle Bedürfnisse zugeschnitten und passgenau vermittelt werden, so dass sie den individuellen Lernbedürfnissen entsprechen, gleichzeitig aber auch die arbeitsmarktpolitischen Anforderungen für den Erfolg von Lerninhalten berücksichtigen. Diese beziehen sich auf die konkreten Bedarfe von Unternehmen, aber auch auf übergeordnete Zielstellungen wie die Erschließung des Fachkräftepotenzials von Frauen für die Wirtschaft. Der Einbezug dieser übergeordneten Themen hat weitreichende Implikationen für die Einordnung und Gewichtung der individuellen Datenpunkte oder die Möglichkeit, aus vergangenen Daten zu extrapolieren. Gerade mit den in KIRA fokussierten neuen IT-bezogenen Qualifikationsbedarfen in transformationsintensiven Branchen und Unternehmen wird ein Bereich adressiert, der einerseits eine hohe Dynamik aufweist, andererseits von massiven Vergeschlechtlichungen in der Zuschreibung und in der Selbstwahrnehmung dieser Kompetenzen geprägt ist (vgl. Kapitel 3). Ziel der Modellbildung und des Matchings im Rahmen von KIRA ist es – nicht zuletzt im Sinne eines ethischen Einsatzes von KI – diesen Biases Rechnung zu tragen und so für die digitale Transformation der Arbeitswelt eine gleichberechtigte Teilhabe von Frauen in den zunehmend digitalisierten Arbeitsprozessen zu ermöglichen. Dies impliziert: KI-gestütztes Lernen und Qualifizieren muss fair sein.

3 Identifikation kritischer Stellen im Entwicklungsprozess

Im Projekt KIRA wird das Modell „Kreislauf der Diskriminierung“ [MRPht] verwendet, um frühzeitig – und auch im Verlauf - Faktoren im Entwicklungsprozess zu identifizieren, die kritisch im Hinblick auf Verzerrungen bezüglich Gender- und Diversitätsaspekten sind. Das Modell bildet den Zusammenhang zwischen gesellschaftlichen Biases und Diskriminierung in Anwendungen als einen wiederkehrenden Kreislauf der Produktion und Reproduktion sozialer Ungleichheiten im Kontext der Gestaltung von IT-Systemen ab. Abb. 1 visualisiert das zirkuläre Zusammenspiel von Produkt, Welt, Personen und Prozess, um eine differenzierte Betrachtung und Analyse von Diskriminierung und deren

Herstellung zu ermöglichen. Das Modell dient einerseits zur Erklärung von Diskriminierung und andererseits zum Identifizieren (segment)spezifischer Ansatzpunkte für Reflexion und Intervention.



Abb. 1: Kreislauf der Diskriminierung [MRPhT]

Im Folgenden werden anhand des Modells in den vier Feldern exemplarisch verschiedene kritische Faktoren beleuchtet, die speziell für die gendersensible Entwicklung von KI-Technologien im Kontext beruflicher Weiterbildung mit Fokus auf digitale Kompetenzen relevant sind. Auf dieser Basis werden Punkte identifiziert, an denen Interventionen für Geschlechtergerechtigkeit ansetzen können.

3.1 Produkt: Diskriminierung in Anwendungen

KI-Technologien sind nicht neutral und führen – wie sowohl wissenschaftlich als auch zunehmend öffentlich kritisiert – oftmals zu Verzerrungen und Diskriminierungen. Drei Ebenen, auf denen Biases auftreten können, können unterschieden werden: Biases in den Daten zum Training von KI-Anwendungen, Biases in den Algorithmen selbst und Biases im Kontext der Nutzung [Dr19; Me21].

Fehlende Diversität steht bei der Erhebung von Daten und Entwicklung von Datensets mehr und mehr in der Kritik. Unausgewogenes, fehlendes, korrumpiertes oder systematisch-verzerrtes Datenmaterial führt regelmäßig zur Diskriminierung von Mitgliedern gesellschaftlicher Gruppen wie z.B. Frauen und People of Color [BS16]. Unausgewogene Datensets zum Training von KI-Anwendungen führen dazu, dass Diskriminierungen via maschinellem Lernen in Anwendungen verankert werden - sie werden damit sozusagen unsichtbar und im Zuge nur bedingt erklärbarer Modelle eingeschränkt zugänglich. Zudem werden Datensets, die die Basis für maschinelles Lernen darstellen, häufig „used, shared, and reused with little visibility into the processes of deliberation that led to their creation“ [Hu21, S. 560]. Auch die Kategorienbildung (data labeling) kann solchen Verzerrungen unterworfen sein bzw. Verzerrungen einbringen oder fortsetzen. Somit ist es eine so essenzielle wie schwierige

Anforderung an den Einsatz von KI-Technologien, sicherzustellen, dass Diskriminierungen und Biases nicht in die Algorithmen der KI miteinfließen und infolgedessen diese Stereotype und Biases unbemerkt durch die KI repliziert werden.

Systeme, die KI-Technologien integrieren, werden zunehmend u.a. bei der Personalauswahl eingesetzt und bieten anschauliche Beispiele für Diskriminierung (s. Review bei [KW20]). So hat z.B. das von Amazon eingesetzte KI-System zur Beurteilung von internen Bewerber*innen Frauen systematisch herausgefiltert und somit auch keine Frauen zur Besetzung von internen Stellen vorgeschlagen. Diese „Empfehlung“ der KI-Software basierte auf dem zuvor trainierten Algorithmus, der die Einstellungen der vergangenen Jahre analysierte und zu dem Ergebnis kam, dass die Eigenschaft „Frau“ im Unternehmen keine Aussicht auf Erfolg hat [Da18]. Auch mit dem Versuch, Biases zu vermeiden, indem die Information zu bestimmten, potenziellen Quellen von Diskriminierung wie Geschlecht, Alter oder ethnische Zugehörigkeit als Variable aus dem Daten eliminiert wird, lässt sich das Problem nicht einfach lösen. Im Gegenteil: Diskriminierende Entscheidungsmechanismen können immer noch auftreten, sind aber schwerer zu entdecken [WBS18]. Teilweise führen auch andere Faktoren, die nichts mit der konkreten Datenbasis zu tun haben, zu unerwünschten Effekten [SG19]. Beispielsweise werden jungen Frauen bei Google seltener Werbeanzeigen für MINT-Stellen angezeigt, wenn der Algorithmus auf Kostenoptimierung ausgelegt ist, da Werbung für diese Zielgruppe vergleichsweise teuer ist [LT19].

3.2 Welt: Gesellschaftliche Biases

Das Feld Welt bezieht sich auf die Lebenswelten, in denen Personen leben und sozialisiert werden, die u.a. als Akteur*innen in Beratungs- und Entwicklungsprozessen zu berücksichtigen sind. Diese Lebenswelten sind geprägt von gesellschaftlichen Biases und sozialer Ungleichheit entlang unterschiedlicher Kategorien bzw. kategorialer Zuschreibungen. Eine wesentliche Kategorie ist dabei "Geschlecht". Stereotype und Vorurteile, die sich z.B. auf Geschlecht beziehen, beeinflussen Verhalten und können entsprechend auf aktionaler Ebene zu Diskriminierung von Personengruppen führen. Die moderne Stereotypenforschung zeigt, dass auch ohne die Existenz von Vorurteilen, nur auf Basis der Kenntnis von Stereotypen diskriminiert wird [KB21], also auch entgegen der Überzeugung einer Person.

Gerade im Hinblick auf MINT-Fächer und -Berufe und Kompetenzen, die im Zuge der Digitalisierung immer relevanter werden und als digitale Kompetenzen im Projekt KIRA im Fokus stehen, ergibt sich aus dem Zusammenwirken von Assoziationen ein Paket an Zuschreibungen und Abschreibungen, das sich sowohl im Kontext von Ausbildungs- und Berufswahl als auch in späteren Arbeitsfeldern und beruflicher Weiterbildung auswirkt. Noch immer gängige Assoziationen sind, Männlichkeit mit „technisch“ und „kompetent“ und Weiblichkeit mit „nicht-technisch“ und „sozial“ zu verknüpfen [EI18; KW21]. Bereits Grundschulkindern schreiben Mathematik und mathematisch-logische Fähigkeiten eher Jungen und Männern als Mädchen und Frauen zu [Cv21]. Geschlechtsspezifische Fähigkeitszuschreibungen zeigen sich u.a. in vergeschlechtlichten Interessen [BLC17] und in den

MINT-bezogenen Fähigkeitsselbstkonzepten und Leistungsattributionen [ELP17]. Derart wirksame Geschlechterstereotype lassen sich auch bei Lehrkräften und Eltern nachweisen: Bei vergleichbaren Leistungen und Fähigkeiten in MINT-Fächern werden Jungen besser eingeschätzt als Mädchen [HS05]. Entsprechendes gilt für Männer und Frauen im Beruf [SE16], speziell bezogen auf Umgang mit digitalen Technologien [Fr20]. Tatsächlich zeigen Studien im Hinblick auf die Zuschreibung von Kompetenz generell, dass Frauen eher unterbewertet und Männer eher überbewertet werden [E118]. Dieser Unterschied in der Kompetenzwahrnehmung spiegelt sich auch auf Ebene der Selbsteinschätzung wieder – jedoch ohne entsprechendes geringeres Leistungskorrelat auf Seiten der Frauen; Männer hingegen überschätzen sich eher [Si03].

Im Bereich Berufswahl ist zudem der Einfluss von Eltern zu berücksichtigen: So zeigt z.B. die Langzeitstudie von Jacobs, Chhin und Bleeker [JCB06], dass die (u.a. geschlechtsstereotypen) Erwartungen von Eltern ein signifikanter Faktor sowohl im Hinblick auf die beruflichen Erwartungen der Kinder als auch deren Berufswahl sind. Bezüglich des Einschlagens mathematischer Karrieren zeigten Bleeker und Jacobs [BJ04] zudem, dass gerade die Erwartungen der Mütter im frühen Kindesalter einen deutlichen Einfluss auf späteres Wahlverhalten unabhängig von den tatsächlichen mathematischen Fähigkeiten der Kinder haben. Auch die Verknüpfung von Führungsrollen mit insbesondere dem männlichen Geschlecht [Gi17; Sc96] sind ein wichtiger Bias im Kontext beruflicher Weiterbildung, da Weiterqualifizierung häufig auf das zukünftige Übernehmen einer anderen Rolle – beispielsweise einer Führungsrolle – ausgerichtet ist.

Vor diesem Hintergrund ist im Projekt KIRA stets zu bedenken, dass bisherige Berufswege zumeist schon Biases aufweisen und entsprechend nicht unbedingt die individuellen Fähigkeiten und Interessen angemessen widerspiegeln. Eine Analyse von diesbezüglichen Dokumenten kann also immer nur ein Teil des KI-gestützten Matching-Prozesses im Kontext beruflicher Weiterbildung sein.

3.3 Personen: Individuelle und gruppenbezogene Praktiken

Die dargestellten Mechanismen auf Basis von Stereotypen finden ihren Niederschlag in individuellen und gruppenbezogenen Praktiken, die sich im Bereich beruflicher (Weiter)Bildung und im beruflichen Alltag zeigen, z.B. bei der Berufs- und Bildungsberatung, bei der Ausschreibung und Besetzung von Stellen in Unternehmen und in Teams, die Software für diese Einsatzbereiche entwickeln (z.B. KI-basierte Recruitment-Tools, siehe oben).

Im Feld der Berufsberatung tragen Geschlechterstereotype dazu bei, bestehende Ungleichheiten im Arbeitsmarkt zu perpetuieren. So zeigt beispielsweise die Studie von Ostendorf [Os05] zur Berufsberatung der Bundesagentur für Arbeit, dass diese der beruflichen Segmentation der Geschlechter nicht nur nicht entgegenwirkt, sondern diese mitverursacht. U.a. würden viele Berater*innen Mädchen nur dann gewerblich-technische Berufe anraten, wenn die Mädchen in Mathematik und Naturwissenschaften „besonders gute Schulnoten

vorweisen können sowie eine starke Persönlichkeit mitbringen, die sich in einer Männergesellschaft wird behaupten können“ [Os05, S. 126]. Auch im Rahmen von Beratungen eingesetzte Interessenstests weisen Gender-Biases auf (s. dazu z.B. [WH13]). Im Hinblick auf das Fach Informatik zeigt eine aktuelle Studie, dass – entgegen der häufigen Assoziation, dass gerade mathematisches Wissen ein guter Prädiktor für den Erwerb von Coding Skills sei – andere mögliche Prädiktoren diskutiert werden [RPJ17], z.B. sprachliche Begabung [Pr20]. In der Beratungspraxis scheint dies allerdings bisher wenig Beachtung zu finden, wobei gerade bei Schülerinnen, die sowohl in Mathematik als auch in Sprachen gut sind, zumeist eher die Sprachkompetenz wahrgenommen wird [SG18].

Zahlreiche Ansatzpunkte für das Wirken von Stereotypen finden sich bei Stellenausschreibungen, Stellenbesetzungen oder auch Vorstellungen und Plänen dazu, wer z.B. intern im Hinblick auf welche Position weiterqualifiziert wird. Bei Stellenbesetzungen werden bereits bei der Bewertung beruflicher Unterlagen Verzerrungen deutlich. Moss-Racusin et al. [Mo12] zeigten z.B., dass im Rahmen der Besetzung einer Stelle als Laborleitung identische Bewerbungsunterlagen unter dem Label „Mann“ als signifikant kompetenter und einstellungstauglicher eingestuft wurden als unter dem Label „Frau“. Zudem wurden ein höheres Einstiegsgehalt und mehr Karriere-Mentoring angeboten. Antragstellerinnen für ein Stipendium müssen deutlich höhere wissenschaftliche Leistung erbringen, um genauso kompetent wahrgenommen zu werden wie Antragsteller [WW08]. Krawczyk und Smyk [KS16] zeigten zudem, dass akademische Artikel von Autorinnen schlechter bewertet wurden als die von Autoren. Auch im Zuge der Gestaltung von Softwareprodukten für die Einsatzbereiche Bildung, Beratung und Personal-Recruitment werden geschlechtsbezogene Verzerrungen wirksam. Eine ausführliche Darstellung zu Biases und Diskriminierung bei der Entwicklung von IT-Systemen findet sich bei Marsden, Raudonat und Pröbster [MRPht].

3.4 Prozess: Verzerrungen durch Design

Für Gestaltungsprozesse gibt es unterschiedliche Methoden und Werkzeuge, die Entwickler*innen im Hinblick auf eine angemessene Konzeption und Berücksichtigung der Nutzer*innen unterstützen können bzw. sollen. Diese können durch ihre Konzeption oder in der Praxis ihrer Nutzung entweder hilfreich im Hinblick auf Gleichstellung sein oder wiederum Verzerrungen in den Prozess einbringen. Insbesondere der unsachgemäße oder unreflektierte Einsatz von Methoden und Tools birgt das Risiko, Stereotype in die Produktgestaltung zu integrieren, die in diskriminierenden Produkten resultieren. So führen beispielsweise die in der UX- und HCI-Praxis eingesetzten Empathy Maps häufig zu einer Explikation eigener Stereotype statt zu einer empathischen Konfiguration der Nutzer*innen [SD19]. Dies ist umso problematischer, da bei dem Einsatz von Tools und Methoden oftmals eine gewisse Standardisierung und Ergebnisqualität qua Tool/Methode unterstellt werden, die so aber nicht existieren.

Für den Bereich KI-Technologien sollen hier beispielhaft zwei Tools benannt werden, die zum Monitoring im Entwicklungsprozess von Anfang an eingesetzt werden können. So

kann zur Einschätzung von (geplanten) KI-Systemen die Kritikalitätsmatrix von Kraft und Zweck [Bu20; Zw22] genutzt werden. Entlang der Dimensionen Schadenspotenzial und Grad der Abhängigkeit werden hier KI-Systeme bewertet, die Entscheidungen unterstützen. Es werden also der konkrete Anwendungskontext und die damit verbundenen sozialen Prozesse berücksichtigt. Das zweite Tool zielt auf eine verantwortliche Gestaltung und Nutzung von Datensets: Das Framework for Dataset Development Transparency [Hu21] dient der Unterstützung von Entscheidungsfindung und Verantwortlichkeit entlang des Software Development Lifecycle. Es lenkt die Aufmerksamkeit unter anderem auf die häufig unbeachtete Arbeit und auf Entscheidungen, die zur Erstellung eines Datensets geführt haben - und macht damit auch an dieser Stelle Verantwortung adressierbar [Hu21].

4 Fazit und Ausblick

Die vorangegangenen Ausführungen zeigen anhand von verschiedenen Beispielen kritische Stellen im Hinblick auf Gendersensibilität und Biases im Kontext von KI-Technologien im Feld beruflicher Bildung auf, die im Projekt KIRA zu berücksichtigen sind. Ein vor diesem Hintergrund gendersensibler Einsatz von KI-Technologien in KIRA kann (und soll) die Chance bieten, Biases aktiv zu begegnen und die Erschließung digitaler Kompetenzen und digitalisierungsrelevanter Berufs- und Tätigkeitsfelder für z.B. Frauen zu fördern. Dies kann geschehen, indem z.B. Stereotype überwindende Empfehlungen gegeben werden, die an Interessen und Fähigkeiten orientiert sind, die aus der einfachen Betrachtung bisheriger Erwerbsbiographien so zunächst nicht offensichtlich erscheinen (z. B. Relevanz sprachlicher Begabung für Coding Skills, siehe oben). Auch die Rahmensetzung von Empfehlungen ist wichtig: Wenn zum Beispiel im Rahmen von Studien zu 'Future Skills' die Kompetenz 'Scrum' als eine technologische Fähigkeit kontextualisiert wird, ist davon auszugehen, dass eine Empfehlung, sich die Methode Scrum anzueignen, anders aufgenommen wird, als wenn diese eingebettet ist in den Kontext von Organisations- und Moderationsfähigkeiten.

Es besteht weiterer Forschungsbedarf, wie geeignete Interventionen im oben dargestellten Kreislauf der Diskriminierung im Zusammenhang mit Empfehlungen zum individuellen Aufbau digitaler Kompetenzen und digitalisierungsrelevanter Berufs- und Tätigkeitsfelder aussehen könnten. Wie dargestellt spielen Biases und Stereotype der beratenden Personen eine wesentliche Rolle und fließen in den Beratungsprozess ein – ein KI-gestütztes System läuft also Gefahr, vergeschlechtlichte Empfehlungen zu reproduzieren und so den Status quo zu reifizieren.

Wesentlich ist nicht zuletzt die Frage, wie es gelingen kann, dass gute Empfehlungen auch auf Akzeptanz stoßen bei den Personen, die sich weiterqualifizieren möchten. Schließlich ist die beste Empfehlung nur dann von Nutzen, wenn die beratene Person sie auch annehmen und für sich umsetzen kann. Selbststereotype können hier verhindern, dass den Empfehlungen gefolgt wird. Denn aufgrund der Allgegenwart von Stereotypen ist es – folgerichtig, aber paradoxerweise – so, dass im beruflichen Beratungsprozess genau jene Empfehlungen als

besonders plausibel und non-biased wahrgenommen werden, die gängigen Geschlechterstereotypen entsprechen [Wa21]. Interventionen in nur einem der vier Felder des Kreislaufes der Diskriminierung scheinen somit wenig erfolgsversprechend: Nur wenn alle vier Felder ganzheitlich betrachtet werden, können Empfehlungen zur Kompetenzentwicklung im Kontext von KIRA nachhaltig wirksam werden.

Literatur

- [BJ04] Bleeker, M. M.; Jacobs, J. E.: Achievement in math and science: Do mothers' beliefs matter 12 years later? *Journal of Educational Psychology* 96/1, S. 97, 2004, ISSN: 1939-2176.
- [BLC17] Bian, L.; Leslie, S.-J.; Cimpian, A.: Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science* 355/6323, S. 389–391, 2017, ISSN: 0036-8075.
- [BS16] Barocas, S.; Selbst, A. D.: Big data's disparate impact. *California Law Review* 104/, S. 671–732, 2016.
- [Bu20] Bundesministerium für Arbeit und Soziales: Begleitband des Bundesministeriums für Arbeit und Soziales, 2020, URL: <https://www.bmas.de/DE/Service/Publikationen/a811-eu2020-begleitband.html>.
- [Cv21] Cvencek, D.; Brečić, R.; Gaćeša, D.; Meltzoff, A. N.: Development of Math Attitudes and Math Self-Concepts: Gender Differences, Implicit–Explicit Dissociations, and Relations to Math Achievement. *Child Development* 92/5, e940–e956, 2021, ISSN: 0009-3920.
- [Da18] Dastin, J.: Amazon scraps secret AI recruiting tool that showed bias against women. In: *Ethics of Data and Analytics*. Auerbach Publications, S. 296–299, 2018, ISBN: 1003278299.
- [Dr19] Draude, C.; Klumbyte, G.; Lücking, P.; Treusch, P.: Situated algorithms: a sociotechnical systemic approach to bias. *Online Information Review* 44/2, S. 325–342, 2019, ISSN: 1468-4527.
- [El18] Ellemers, N.: Gender stereotypes. *Annual review of psychology* 69/Jan, 69:275–298, 2018.
- [ELP17] Ertl, B.; Luttenberger, S.; Paechter, M.: The Impact of Gender Stereotypes on the Self-Concept of Female Students in STEM Subjects with an Under-Representation of Females. *Frontiers in Psychology* 8/, 2017, ISSN: 1664-1078, URL: <https://www.frontiersin.org/article/10.3389/fpsyg.2017.00703>.
- [Fr20] Franken, S.; Herr, V.; Kempf, L.; Mauritz, N.; Wattenberg, M.: Women Resource 4.0 - Potenziale von qualifizierten Frauen, 2020, URL: https://www.fh-bielefeld.de/multimedia/Fachbereiche/Wirtschaft+und+Gesundheit/Forschung/Denkfabrik+Digitalisierte+Arbeitswelt/WomenRessource4_0_Abschlussbericht.pdf.

- [Gi17] Gipson, A. N.; Pfaff, D. L.; Mendelsohn, D. B.; Catenacci, L. T.; Burke, W. W.: Women and Leadership. *The Journal of Applied Behavioral Science* 53/1, S. 32–65, 2017, ISSN: 0021-8863, URL: <https://doi.org/10.1177/0021886316687247>.
- [HS05] Herbert, J.; Stipek, D.: The emergence of gender differences in children's perceptions of their academic competence. *Journal of applied developmental Psychology* 26/3, S. 276–295, 2005, ISSN: 0193-3973.
- [Hu21] Hutchinson, B.; Smart, A.; Hanna, A.; Denton, E.; Greer, C.; Kjartansson, O.; Barnes, P.; Mitchell, M.: Towards accountability for machine learning datasets: Practices from software engineering and infrastructure. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, S. 560–575, 2021.
- [JCB06] Jacobs, J. E.; Chhin, C. S.; Bleeker, M. M.: Enduring links: Parents' expectations and their young adult children's gender-typed occupational choices. *Educational Research and Evaluation* 12/4, S. 395–407, 2006, ISSN: 1380-3611.
- [KB21] Kurdi, B.; Banaji, M. R.: Implicit social cognition: A brief (and gentle) introduction. Preprint on *PsyArXiv*, 2021.
- [KS16] Krawczyk, M.; Smyk, M.: Author's gender affects rating of academic articles: Evidence from an incentivized, deception-free laboratory experiment. *European Economic Review* 90/, S. 326–335, 2016, ISSN: 0014-2921.
- [KW20] Köchling, A.; Wehner, M. C.: 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, S. 795–848, 2020, ISSN: 2198-2627.
- [KW21] Kohlrausch, B.; Weber, L.: Gender Relations at the Digitalised Workplace: The Interrelation Between Digitalisation, Gender, and Work. *Gender a výzkum/-Gender and Research* 21/2, S. 13–31, 2021, ISSN: 2570-6578.
- [LT19] Lambrecht, A.; Tucker, C.: Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. *Management science* 65/7, S. 2966–2981, 2019, ISSN: 0025-1909.
- [Me21] Mehrabi, N.; Morstatter, F.; Saxena, N.; Lerman, K.; Galstyan, A.: A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)* 54/6, S. 1–35, 2021, ISSN: 0360-0300.
- [Mo12] Moss-Racusin, C. A.; Dovidio, J. F.; Brescoll, V. L.; Graham, M. J.; Handelsman, J.: Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences* 109/41, S. 16474–16479, 2012, URL: <http://www.pnas.org/content/109/41/16474.abstract>.
- [MRPht] Marsden, N.; Raudonat, K.; Pröbster, M.: Kreislauf der Diskriminierung - Gesellschaftliche Biases und Diskriminierung in IT-Systemen als wiederkehrender Kreislauf der (Re)Produktion sozialer Ungleichheiten./, eingereicht.

- [Os05] Ostendorf, H.: Politische Steuerung durch Symbole und Verfahrensweisen—Die Mädchenpolitik der Berufsberatung. In (Henninger, A.; Ostendorf, H., Hrsg.): Die politische Steuerung des Geschlechterregimes. Springer, S. 115–137, 2005.
- [Pr20] Prat, C. S.; Madhyastha, T. M.; Mottarella, M. J.; Kuo, C.-H.: Relating natural language aptitude to individual differences in learning programming languages. *Scientific reports* 10/1, S. 1–10, 2020, ISSN: 2045-2322.
- [RPJ17] Román-González, M.; Pérez-González, J.-C.; Jiménez-Fernández, C.: Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test. *Computers in Human Behavior* 72/, S. 678–691, 2017, ISSN: 0747-5632.
- [Sc96] Schein, V. E.; Mueller, R.; Lituchy, T.; Liu, J.: Think manager—think male: A global phenomenon? *Journal of organizational behavior* 17/1, S. 33–41, 1996, ISSN: 0894-3796.
- [SD19] Siegel, D.; Dray, S.: The map is not the territory: empathy in design. *Interactions* 26/2, S. 82–85, 2019, ISSN: 1072-5520.
- [SE16] Steffens, M.; Ebert, I. D.: Frauen–Männer–Karrieren: Eine sozialpsychologische Perspektive auf Frauen in männlich geprägten Arbeitskontexten. Springer-Verlag, 2016, ISBN: 3658107502.
- [SG18] Stoet, G.; Geary, D. C.: The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological science* 29/4, S. 581–593, 2018, ISSN: 0956-7976.
- [SG19] Suresh, H.; Gutttag, J. V.: A framework for understanding unintended consequences of machine learning. *arXiv preprint arXiv:1901.10002* 2/, 2019.
- [Si03] Sieverding, M.: Frauen unterschätzen sich: Selbstbeurteilungs-Biases in einer simulierten Bewerbungssituation. *Zeitschrift für Sozialpsychologie: ZFSP* 34/3, S. 147–160, 2003, ISSN: 0044-3514.
- [Wa21] Wang, C.; Wang, K.; Bian, A.; Islam, R.; Keya, K. N.; Foulde, J.; Pan, S.: Bias: Friend or Foe? User Acceptance of Gender Stereotypes in Automated Career Recommendations. *arXiv preprint arXiv:2106.07112*/, 2021.
- [WBS18] Williams, B. A.; Brooks, C. F.; Shmargad, Y.: How algorithms discriminate based on data they lack: Challenges, solutions, and policy implications. *Journal of Information Policy* 8/1, S. 78–115, 2018, ISSN: 2381-5892.
- [WH13] Wetzel, E.; Hell, B.: Gender-related differential item functioning in vocational interest measurement. *Journal of Individual Differences* 34/3, S. 170–183, 2013.
- [WW08] Wenneras, C.; Wold, A.: Nepotism and sexism in peer-review. In (Wyer, M.; Barbercheck, M.; Giesman, D.; Özturk, H.; Wayne, M., Hrsg.): *Women, Science, and Technology*. Routledge, S. 64–70, 2008, ISBN: 0203895657.
- [Zw22] Zweig, K. A.: Risk Matrix, 2022, URL: <http://aalab.cs.uni-kl.de/resources/>.