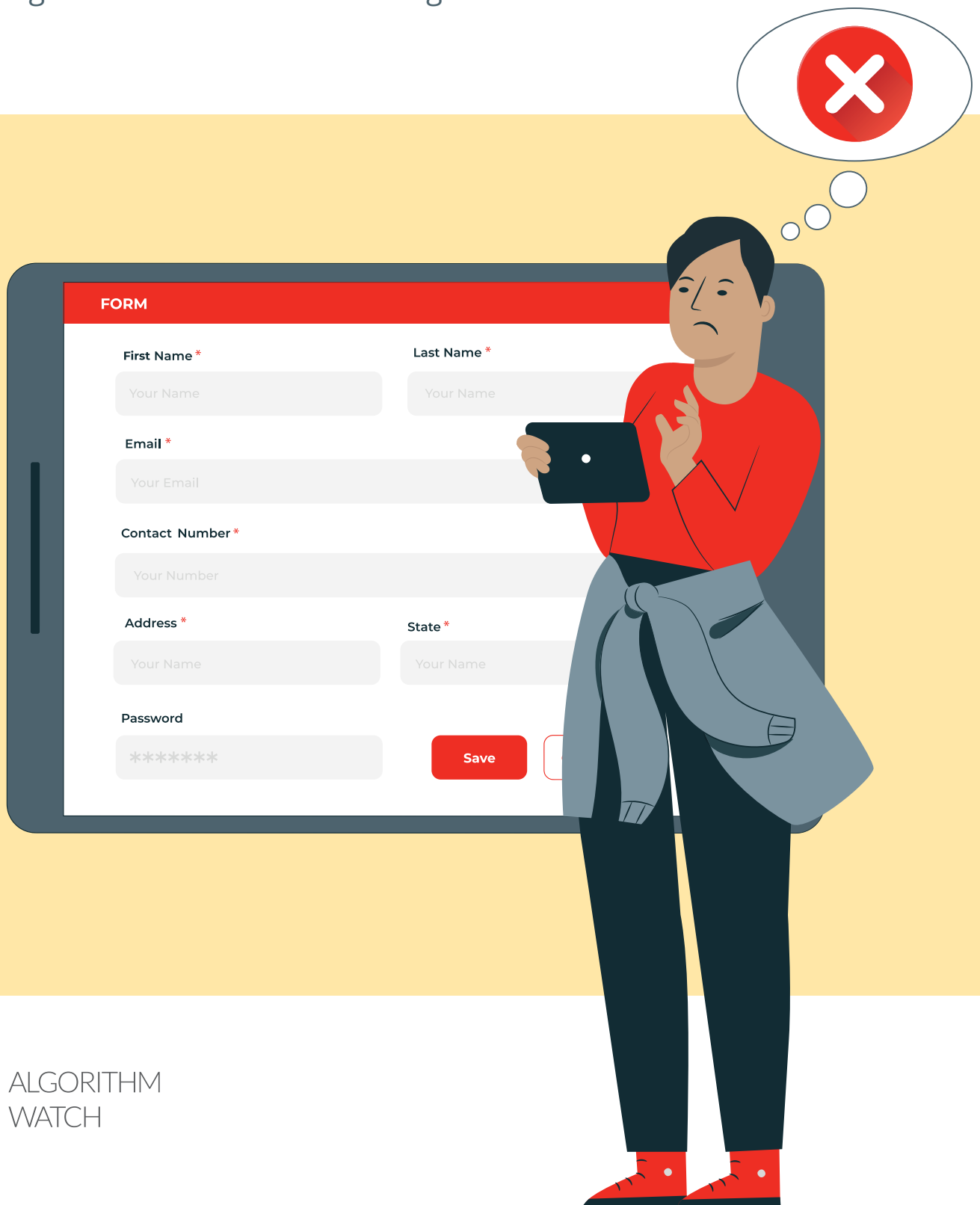


Automatisierte Entscheidungssysteme und Diskriminierung

Ursachen verstehen, Fälle erkennen,
Betroffene unterstützen

Ein Ratgeber für Antidiskriminierungsstellen



Impressum

**Automatisierte Entscheidungssysteme und Diskriminierung:
Ursachen verstehen, Fälle erkennen, Betroffene unterstützen.**
Ein Ratgeber für Antidiskriminierungsstellen

Juni 2022

Online verfügbar auf: <https://algorithmwatch.org/de/autocheck/>

Herausgeber:

Matthias Spielkamp
AW AlgorithmWatch gGmbH
Linienstraße 13
10178 Berlin
Deutschland

Kontakt: info@algorithmwatch.org

Autorin und Projektmanagerin von „AutoCheck“

Jessica Wulf

Layout und Illustrationen

Beate Autering

Lektorat

Karola Klatt

Gefördert durch die



Antidiskriminierungsstelle
des Bundes



Diese Publikation steht unter der Lizenz Creative Commons Namensnennung 3.0
Deutschland

<https://creativecommons.org/licenses/by/3.0/de/legalcode>

► Einleitung Seite 4

1 Grundlagen und zentrale Begriffe

Hier werden zentrale Begriffe geklärt und in die Funktionsweise automatisierter Entscheidungssysteme eingeführt. Seite 5

2 Fallbeispiele

Die Fallbeispiele erzählen von ungerechten oder diskriminierenden Ergebnissen durch die Anwendung von automatisierten Entscheidungssystemen. Seite 6

3 Ursachen

Schaubilder veranschaulichen, wie ungerechte oder diskriminierende Ergebnisse durch automatisierte Entscheidungssysteme entstehen. Seite 11

4 Anwendungsfelder

Mit einem Fokus auf privaten Dienstleistungen und dem Zugang zu Gütern werden relevante Anwendungsfelder automatisierter Entscheidungssysteme übersichtlich dargestellt. Seite 14

5 Erkennen

Ein Fragenkatalog soll Beratende in der Fallklärung dabei unterstützen, Hinweise auf Algorithmen oder automatisierte Entscheidungssysteme zu erkennen. Seite 16

6 Umgang

Anknüpfend an gängige Schritte in der Antidiskriminierungsberatung wird aufgezeigt, wie Beratende mit dem Verdacht auf Benachteiligung durch automatisierte Entscheidungssysteme umgehen können. Seite 18

7 Glossar

Im Glossar werden Begriffe, die für diese Publikation relevant sind, verständlich erläutert. Seite 22

8 Quellen zum Thema

Eine Liste mit Empfehlungen für Bücher, Filme, Initiativen und vielem mehr möchte zur weiteren Beschäftigung mit automatisierten Entscheidungssystemen anregen. Seite 24

SYMBOLE IN DIESER PUBLIKATION



Ethnische Herkunft



Geschlechtsidentität



Alter



Finanzsektor



Versicherungssektor



Online-Handel



Öffentliche Verwaltung



Arbeitsmarkt



Wohnungssektor



Gesundheitssektor



Bildungssektor

► Einleitung

„Es steht für mich außer Frage, dass das Diskriminierungsrisiko durch automatisierte Entscheidungssysteme alle Lebensbereiche umfassen und in der Antidiskriminierungsarbeit eines der relevantesten Themen sein wird.“ (Interviewpartnerin 02¹)

Die Einschätzung unserer Kreditwürdigkeit, die Anzeige individualisierter Preise im Online-Handel oder ein Chatbot, der das (hoffentlich) richtige Formular anbietet – das sind Anwendungsfälle von automatisierten Entscheidungssystemen². Diese Systeme begegnen uns in immer mehr Bereichen unseres Lebens. Ihr Einsatz führt jedoch nicht unbedingt zu objektiveren oder neutraleren Entscheidungen, sondern immer wieder zu ungerechten, sogar diskriminierenden Ergebnissen.

Die oben zitierte Antidiskriminierungsexpertin schätzt deshalb Diskriminierungsrisiken durch automatisierte Entscheidungssysteme als sehr relevant für die Zukunft der Antidiskriminierungsarbeit ein. Zu ähnlichen Schlüssen kommen auch eine von der Antidiskriminierungsstelle des Bundes (ADS) in Auftrag gegebene Studie³ und ein Bericht des Europäischen Netzwerks der Gleichstellungsgremien (Equinet)⁴.

Doch was genau sind automatisierte Entscheidungssysteme? Wie kommt es zu der Diskriminierung und was kann in der Beratung dagegen getan werden? Diese und weiterführende Fragen beantworten wir in konkreten und verständlichen Anleitungen, Schaubildern und Texten in dieser Handreichung.

Diese Publikation ist im Rahmen des Projekts *AutoCheck – Handlungsanleitung für den Umgang mit automatisierten Entscheidungssystemen für Antidiskriminierungsstellen* entstanden. Sie ist eine erste Einführung in die Diskriminierungsrisiken beim Einsatz von automatisierten Entscheidungssystemen. Für ein tiefergehendes Verständnis der Thematik werden im Rahmen des Projekts Weiterbildungskonzepte entwickelt, die als Workshops für Multiplikator*innen veröffentlicht und angeboten werden.

Wir wenden uns mit dieser Handreichung in erster Linie an Mitarbeitende von Antidiskriminierungsstellen in Deutschland. Durch Kompetenzaufbau in diesem zunehmend bedeutsamen Themenfeld sollen sie in die Lage versetzt werden, Risiken besser zu erkennen, einzuschätzen und dadurch in konkreten Diskriminierungsfällen Betroffene besser unterstützen zu können. Darüber hinaus soll diese Publikation auch den von Diskriminierung Betroffenen die Möglichkeit geben, sich über das Thema zu informieren.

Der Fokus liegt auf dem Zugang zu Gütern und privaten Dienstleistungen, beispielsweise Benachteiligungen im Online-Handel oder beim Abschließen einer Versicherung. Nur am Rande gehen wir auf andere Anwendungsbereiche des Allgemeinen Gleichbehandlungsgesetzes (AGG) ein.

Das Projekt wurde von der ADS gefördert und von AlgorithmWatch umgesetzt. Wir möchten uns an dieser Stelle herzlich für die Förderung und Unterstützung bedanken sowie für die angenehme Zusammenarbeit mit Nathalie Schlenzka, Niklas Hofmann und Gudula Fritz.

1 Im Rahmen des Projekts *AutoCheck* wurden 15 Mitarbeitende von Antidiskriminierungsstellen und Antidiskriminierungsexpert*innen interviewt. Themenschwerpunkte der Interviews waren: Vorwissen und Berührungspunkte mit automatisierten Entscheidungssystemen in der Beratung, Einschätzung der Relevanz des Themas im Antidiskriminierungskontext und alltägliche Aufgaben und Themen in der Beratungsarbeit. Alle Zitate in dieser Publikation, sofern nicht anders gekennzeichnet, sind Äußerungen aus den Interviews.

2 Für diese Publikation relevante Begriffe finden Sie im Glossar ab S. 22.

3 Orwat, Carsten (2019). Diskriminierungsrisiken durch Verwendung von Algorithmen: eine Studie, erstellt mit einer Zuwendung der Antidiskriminierungsstelle des Bundes. *Nomos*.

4 Allen QC, Robin & Masters, Dee (2020). Regulating For An Equal AI: A New Role For Equality Bodies. *An Equinet Publication*. https://equineteurope.org/wp-content/uploads/2020/06/ai_report_digital.pdf (14.04.2022).

1 Grundlagen und zentrale Begriffe

„Ich hätte gerne eine grundlegende Einführung, sodass man weiß, worüber man spricht.“ (Interviewpartner 13)

Vielleicht haben Sie sich beim Lesen des Titels „Automatisierte Entscheidungssysteme und Diskriminierung“ schon gefragt, was automatisierte Entscheidungssysteme eigentlich sind. Diesen und weitere zentrale Begriffe dieser Publikation werden wir in diesem Kapitel definieren und in Zusammenhang setzen. Weitere relevante Begriffe finden Sie im Glossar ab S. 22. Diese Begriffe sind im Text der Publikation Grau und unterstrichen.

BEGRIFFSKLÄRUNG

Warum verwenden wir in dieser Publikation den Begriff „automatisiertes Entscheidungssystem“ statt von einer „Künstliche Intelligenz“ zu sprechen, die Entscheidungen über Menschen trifft?

Künstliche Intelligenz ist zu allererst ein Forschungsfeld der Informatik, das sich mit maschinellern Lernen und der Automatisierung kognitiver Aufgaben befasst. Darüber hinaus suggeriert der Begriff die Nähe zu menschlicher Intelligenz und damit auch zu so etwas wie Verantwortung für Entscheidungen. Das ist irreführend [siehe Kapitel Ursachen, S. 11]. Passender ist der Begriff „automatisiertes Entscheidungssystem“, den wir stattdessen verwenden.

Systeme zum automatisierten Entscheiden (englisch: **A**utomated **D**ecision-**M**aking, kurz: ADM) führen Entscheidungsmodelle oder -wege mithilfe von Algorithmen aus. ADM-Systeme können Entscheidungen unterstützen oder vorbereiten, indem sie Empfehlungen aussprechen oder Daten aufbereiten. Zum Beispiel könnte ein ADM-System verwendet werden, um aus einer großen Anzahl Bewerber*innen für eine Wohnung, eine Liste mit Kandidat*innen zu filtern, die bestimmten Kriterien entsprechen. Basierend auf dieser Liste entscheidet ein Mensch, wer zu

einer Wohnungsbesichtigung eingeladen wird. Die Ausführung einer Entscheidung kann auch vollständig an ADM-Systeme delegiert werden, dann ist im Anwendungsfall kein Mensch mehr beteiligt.

ADM-Systeme basieren auf Algorithmen, einer bestimmten Form der Handlungsanleitung, die zur Lösung eines mathematischen Problems führt. Algorithmen beschreiben „einen für den Computer korrekt interpretierbaren Lösungsweg, der für jede durch das mathematische Problem definierte mögliche Eingabe die korrekte Lösung in endlicher Zeit berechnet.“⁵

Wir möchten hier zwei Arten von Algorithmen unterscheiden, die die Funktionsweise von ADM-Systemen beeinflussen: regelbasierte und lernende Algorithmen. Bei einem regelbasierten Algorithmus sind die Handlungsanleitungen, die der Algorithmus ausführt, explizit programmiert worden. Ein Beispiel dafür sind Online-Formulare, die man oft nur dann abschicken kann, wenn man sich einer der binären Geschlechterkategorien „männlich“ oder „weiblich“ zuordnet. Diese Bedingung für ein korrekt ausgefülltes Formular ist von einem Menschen so programmiert worden.

Lernende Algorithmen sollen dagegen in einer vorgelagerten Phase Regeln aus bereits vorhandenen Daten lernen, die zu einem gewünschten Ergebnis führen. Das können Menschen unterstützen, indem sie die für das Training zur Verfügung stehenden Daten annotieren, also zum Beispiel Daten, die das gewünschte Ergebnis beinhalten, markieren. In der nächsten Phase werden die erlernten Regeln auf ähnliche, aber unbekannte Daten angewendet, um ein richtiges Ergebnis zu erhalten. Das bedeutet, dass bereits vorhandene Daten hier einen großen Einfluss auf die gelernten Regeln und damit das Ergebnis haben.

Gleich welche Art von Algorithmen sie verwenden, der Kontext muss mitbetrachtet werden, um ADM-Systeme besser verstehen zu können. Deswegen

5 AlgorithmWatch (2016). 1. Arbeitspapier: Was ist ein Algorithmus? <https://algorithmwatch.org/de/arbeitspapier-was-ist-ein-algorithmus/> (28.02.2022).

ist es sinnvoll, sie als sozio-technische Systeme zu betrachten, das heißt neben den technischen auch ihre sozialen Komponenten in die Untersuchung einzubeziehen. ADM-Systeme entstehen aus einem spezifischen historischen und gesellschaftlichen Kontext heraus, werden durch diesen beeinflusst und haben wiederum Einfluss auf spezifische Gruppen und Individuen in dieser Gesellschaft und die gesellschaftliche Entwicklung insgesamt. Dieser Einfluss kann auch negativ sein und zu Benachteiligung oder Diskriminierung führen, wie die Fallbeispiele im folgenden Kapitel zeigen.

2 Fallbeispiele

*„Fallbeispiele sind immer super.
Man kann viel Theorie lesen, aber
man merkt sie sich erst an Beispielen.“
(Interviewpartnerin 06)*

In diesem Teil der Publikation finden Sie fünf Fallbeispiele von ungerechten oder diskriminierenden Ergebnissen durch die Anwendung von ADM-Systemen in Deutschland und Italien. Sie dienen als Einstieg in das Thema und zur Veranschaulichung der Problematik. Sie sind nah am Beratungsalltag und sollen Hinweise darauf geben, bei welchen Beratungsfällen ADM-Systeme eine Rolle spielen



könnten. Der Fokus liegt auf privaten Dienstleistungen und dem Zugang zu Gütern in Deutschland. Wir verwenden auch ein Beispiel aus Italien und Beispiele aus anderen Anwendungsbereichen, da es auf die Logik ankommt, die hinter ADM-Systemen und ihrer Anwendung steckt. Diese Logik ist auf andere Anwendungskontexte übertragbar.



a) Nicht kreditwürdig aufgrund von Alter und Geschlecht?

Der Kauf auf Rechnung wird verweigert, mutmaßlich wegen einer für die Kreditwürdigkeit ungünstigen Kombination von Alter und Geschlecht.⁶

Im Jahr 2021 möchte eine Frau in Hessen in einem Online-Shop Kleidung kaufen. Als Zahlungsart gibt sie „Kauf auf Rechnung“ an. Die damit verbundene Bonitätsprüfung, auch „Scoring“ genannt, führt dazu, dass ihr der Kauf auf Rechnung verwehrt wird. Die Frau ist irritiert, sie übt einen guten Job aus, ist um die 40 Jahre alt und hat ihres Wissens nach keine negativen Einträge bei Auskunfteien.⁷ Sie fragt beim Online-Händler nach. Dieser schickt ihr das Ergebnis der Bonitätsprüfung der externen Auskunftei. Hier schien ein automatisiertes System ausgewertet zu haben, dass Frauen um die 40 nicht kreditwürdig genug sind. Diese Hintergründe zu der automatisierten Entscheidung erfährt die Betroffene nur, als sie auf Eigeninitiative beim externen Kreditprüfungsunternehmen anruft und ein Mitarbeiter Vermutungen dazu äußert, warum sie abgelehnt wurde:

⁶ Das Fallbeispiel basiert auf einem Beratungsfall, der uns in einem Interview mit der verantwortlichen Beraterin einer Antidiskriminierungsstelle geschildert wurde. In Finnland gab es einen ähnlichen Fall, der vom Nicht-Diskriminierungs- und Gleichbehandlungstribunal rechtsverbindlich entschieden wurde: <https://algorithmwatch.org/de/kreditscoring-urteil-aus-finnland-wirft-fragen-zur-diskriminierung-auf/>.

⁷ Die größten Auskunfteien in Deutschland sind Schufa Holding AG (mit dem größten Marktanteil), Creditreform Boniversum, Crif Bürgel und Infoscore Consumer Data. Weiterführende Informationen zu Auskunfteien: <https://www.finanztip.de/schufa/auskunfteien/> (24.03.2022).

Weil „Frauen in ihrem Alter geschieden und damit mittellos“ seien – es also am Zusammenhang zwischen Alter und Geschlecht liege. Ein Testing-Verfahren [S. 19, Testing-Verfahren] bietet sich an, ist aber zu aufwendig. Auch für ein Klageverfahren [S. 20, Klageverfahren] nach AGG werden zu niedrige Erfolgschancen im Verhältnis zu den Kosten gesehen.⁸ Erfreulicherweise führt die Beschwerde der Frau zumindest dazu, dass der Online-Händler die Auskunftfei wechselt.

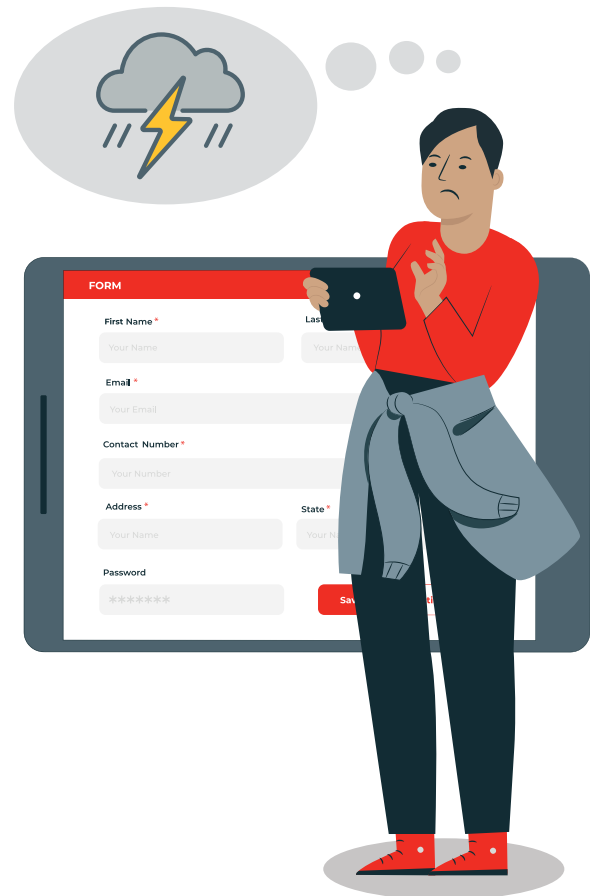
In diesem Fallbeispiel könnte zum Beispiel das verwendete Modell die Ursache für das Ergebnis der ungenügenden Kreditwürdigkeit sein.⁹ Dann wäre hier die Annahme, dass Geschlecht und Alter Einfluss auf die Zahlungsfähigkeit einer Person haben, in das Modell eingegangen. Weitere Informationen finden Sie im Kapitel Ursachen [S. 11].



b) Online-Formulare

Online-Formulare können aufgrund der Kürze des Nachnamens nicht ausgefüllt oder abgeschickt werden.¹⁰

2013 möchte Clemens Li auf der Webseite Bitcoin.de die Kryptowährung Bitcoin kaufen. Das gelingt ihm jedoch nicht, weil das Online-Anmeldeformular der



Seite seinen Nachnamen nicht akzeptiert. Diese Erfahrung macht Herr Li nicht nur auf Bitcoin.de, sondern auch auf einigen anderen Webseiten, die Online-Formulare verwenden. Immer wieder kann er sich nicht anmelden oder etwas nicht kaufen, weil beim Ausfüllen des Namensfelds eine Fehlermeldung erscheint: Der Name „Li“ ist zu kurz. Oft sind Formulare so programmiert, dass Namen mindestens drei Buchstaben enthalten müssen, damit sie akzeptiert und der Vorgang abgeschlossen werden kann. Menschen mit weniger als drei Buchstaben im Namen scheinen bei der Programmierung nicht mitgedacht worden zu sein. So auch im Falle von Bitcoin.de. Herr Li beschwert sich bei den Betreibern der Webseite und bekommt sechs Monate später eine Antwort vom Support per E-Mail: Der Kundenbetreuer bittet ihn, für die Registrierung einen Punkt an seinen Namen anzuhängen, „da unser System zwingend drei Buchstaben vorschreibt.“ Herr Li lehnt das ab und kann infolge auf der Webseite keine Kryptowährung kaufen.

8 Folgende Studie thematisiert u. a. Herausforderungen für Klageverfahren beim Zugang zu Waren und Dienstleistungen im Zusammenhang mit dem AGG: Beigang, Steffen et al. (2021). Möglichkeiten der Rechtsdurchsetzung des Diskriminierungsschutzes bei der Begründung, Durchführung und Beendigung zivilrechtlicher Schuldverhältnisse: Bestandsaufnahme, Alternativen und Weiterentwicklung. *Nomos*. https://www.antidiskriminierungsstelle.de/SharedDocs/downloads/DE/publikationen/Expertisen/rechtsdurchsetzung_zivilrecht.pdf?__blob=publicationFile&v=2 (03.03.2022).

9 Die tatsächlichen Ursachen für das Ergebnis wissen wir nicht, da wir keine Informationen von Seiten der Auskunftfei zu dem Fall oder dem verwendeten Modell haben.

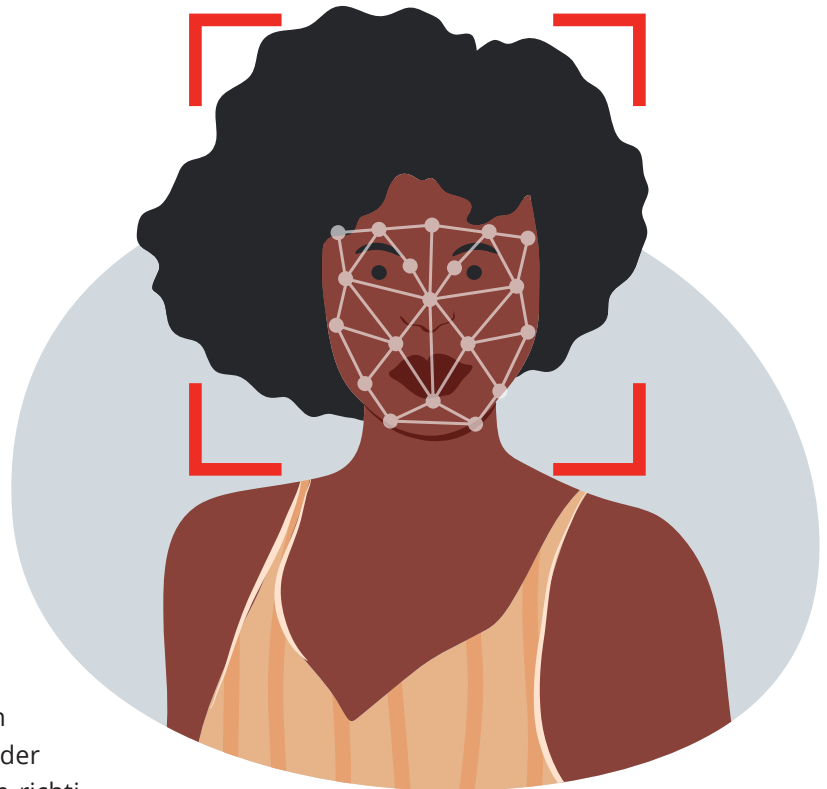
10 Lulamae, Josephine (2022). Name Zu Kurz, Registrierung Ausgeschlossen: Diskriminierende Online-Formulare. *AlgorithmWatch*: <https://algorithmwatch.org/de/fehlerhafte-online-formulare/> (03.03.2022).



c) Fotoautomat Hamburg

Gesichtserkennungssoftware erkennt das Gesicht einer Schwarzen Frau nicht als Motiv.¹¹

Audrey K. möchte 2020 beim Landesbetrieb Verkehr in Hamburg einen internationalen Führerschein beantragen. Sie hat alle erforderlichen Unterlagen für die Beantragung zu ihrem Termin mitgebracht, abgesehen von einem biometrischen Foto, das sie in dem Fotoautomaten im Amt machen möchte. Um ein biometrisches Bild aufzunehmen, muss man sein Gesicht in einen bestimmten Ausschnitt der Kamera halten. Erst wenn das Gesicht im richtigen Ausschnitt erkannt wird, wird ein Foto gemacht. Der Fotoautomat des Landesbetrieb Verkehr erkennt das Gesicht von Audrey K. jedoch nicht als Motiv. Die Gesichtserkennungssoftware, die in diesem Fotoautomaten verwendet wird, erkennt offenbar nur Gesichter mit hellen Hauttönen. Das Verhalten der Mitarbeiterin im Landesbetrieb Verkehr lässt Audrey K. darauf schließen, dass das Problem vor ihrem Besuch schon länger bestand und bekannt war. Sie erinnert sich daran, dass die Mitarbeiterin gesagt habe, „es könnte ein Problem mit Ihrer Hautfarbe geben.“ Der Fotoautomat gehört der Bundesdruckerei. Dort sieht man das Problem nicht im Softwaresystem, sondern meint, es läge an der Ausleuchtung der Kabine. Die Bundesdruckerei beruft



sich darauf, dass die neueste Technologie verwendet werde. Dass aber auch neueste Technologie rassistische und sexistische Ergebnisse liefern kann, zeigt die Studie der Forscherinnen Joy Buolamwini und Timnit Gebru.¹² Die Forscherinnen untersuchen in dieser Studie die Genauigkeit von KI-Anwendungen zur Geschlechtsklassifizierung von Gesichtern, die zum Beispiel von Microsoft oder IBM angeboten wurden. Sie kommen zu dem Ergebnis, dass die untersuchten KI-Anwendungen eine hohe Fehlerquote für Frauen mit dunkleren Hauttönen aufzeigen im Vergleich zu Männern mit helleren Hauttönen. Als einer der Gründe für dieses Ergebnis wird die fehlende Diversität in Daten benannt. Auch in diesem Fallbeispiel von Audrey K. könnte die Ursache des Problems an den Daten liegen, die für die Entwicklung der Software verwendet wurden.¹³

11 Schipkowski, Katharina (2020). Rassismus im Bild: Blind für schwarze Menschen. *Taz.de*: <https://taz.de/Rassismus-im-Bild/!5700872/> (03.03.2022). Ähnliche Fälle: Achenbach, Cornelia (2019). Junge kann Fotoautomaten im Osnabrücker Bürgeramt nicht nutzen. *Neue Osnabrücker Zeitung*: <https://www.noz.de/lokales/osnabrueck/artikel/-20523365> (03.03.2022). Karangwa, Sandra (2015). Entlarvt: Schwarze zu dunkel für biometrische Passfotos. *MiGAZIN.de*: <https://www.migazin.de/2015/07/13/schwarze-sind-zu-dunkel-fuer-das-biometrische-passfoto/> (03.03.2022).

12 Buolamwini, Joy & Gebru, Timnit (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceeds of Machine Learning Research*: 81, 1-15. <http://gendershades.org/> (03.03.2022).

13 Die tatsächlichen Ursachen für das Ergebnis wissen wir nicht.



d) Genderstereotypes Anzeigen von Stellenausschreibungen

Stellenanzeigen werden automatisiert genderstereotyp verbreitet.¹⁴

Im Jahr 2020 schreibt AlgorithmWatch in einem Versuch fiktive Stellen aus. Es handelt sich um Ausschreibungen für eine*n LKW-Fahrer*in und eine*n Erzieher*in. Die Stellenausschreibungen werden auf der Plattform Facebook angezeigt und sind mit Anzeigen auf einer Jobplattform verlinkt. Jobsuchende können bei Facebook nach Jobs suchen und werden dann zu den Anzeigen auf der verlinkten Jobplattform weitergeleitet. Es ist bei Plattformen wie Facebook möglich, bei der Veröffentlichung einer Anzeige eine bestimmte Zielgruppe auszuwählen, die man bevorzugt erreichen möchte. Diesen Vorgang nennt man „Targeting“. Im Versuch wird darauf bewusst verzichtet. Sobald die Anzeigen veröffentlicht sind, werden sie vor allem den Menschen angezeigt, die, basierend auf einer Berechnung durch einen Algorithmus, am wahrscheinlichsten die Anzeigen anklicken. Die Ausschreibung für eine*n LKW-Fahrer*in wird 4.864 Männern gezeigt, aber nur 386 Frauen. Die Anzeige für eine*n Erzieher*in, die genau zur selben Zeit geschaltet wurde, erreicht 6.456 Frauen, aber nur 258 Männer. Das bedeutet: Facebook führt beim Veröffentlichen der Anzeigen ein Targeting durch, ohne eine Erlaubnis dafür einzuholen oder darüber zu informieren. Diese Form von Targeting ist übertragbar auf andere Anzeigen, etwa Wohnungsanzeigen. Die besondere Problematik in diesem und ähnlichen Fällen:

und auf welchen Kriterien das basiert.¹⁵ Weil die Algorithmen der Plattformen meist bestehende Strukturen abbilden, kann bestehende Ungleichheit reproduziert werden.

In diesem Fallbeispiel könnte das Ziel der Anwendung die Ursache dafür sein, dass Stellenanzeigen nach Genderstereotypen verbreitet werden.¹⁶ Denn das Ziel könnte hier sein, die Klicks auf Stellenanzeigen für Anzeigekunden zu maximieren. Weitere Informationen zu Ursachen von Diskriminierung durch ADM-Systeme finden Sie im nächsten Kapitel [Ursachen, S. 11].

*„Die Algorithmen an sich sind nicht das Problem, sondern dass sie Diskriminierung und strukturelle Ungleichheit verdecken. Bei Diskriminierung kommt es nicht auf die Absicht an, sondern auf die Wirkung. Diskriminierung ist in die Struktur und auch in uns eingeschrieben: Wir haben das alle gelernt.“ (Interviewpartner*in 04)*

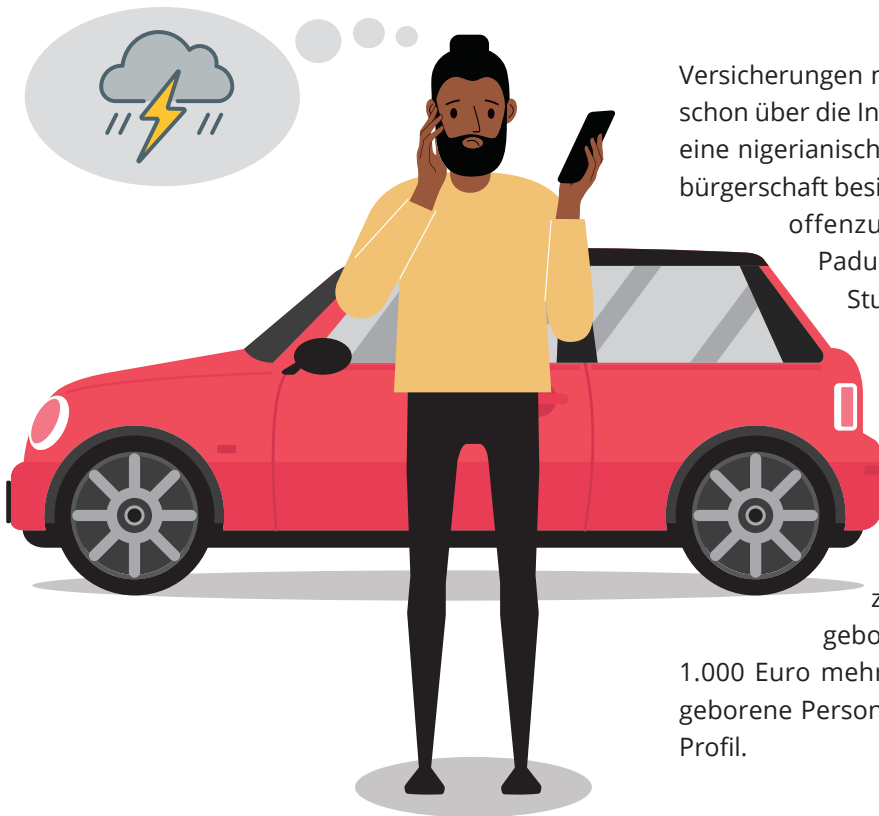


Betroffene Personen wissen nicht, dass ihnen etwas anderes angezeigt wird als anderen Nutzer*innen

¹⁴ Kayser-Bril, Nicolas (2020). Automatisierte Diskriminierung: Facebook verwendet grobe Stereotypen, um die Anzeigenschaltung zu optimieren. *AlgorithmWatch*: <https://algorithmwatch.org/de/automatisierte-diskriminierung-facebook-verwendet-grobe-stereotypen-um-die-anzeigenschaltung-zu-optimieren/> (03.03.2022).

¹⁵ Eine Einschätzung zur Rechtsdurchsetzung dieses Beispiels: Fröhlich, Wiebke (2020). Männer fahren LKW, Frauen erziehen Kinder – Diskriminierendes Gendertargeting durch Facebook. *Verfassungsblog*: <https://verfassungsblog.de/diskriminierende-facebook-algorithmen/> (03.03.2022).

¹⁶ Die tatsächlichen Ursachen für das Ergebnis wissen wir nicht, da wir von Facebook keine Informationen erhalten haben, z. B. welche Ziele, Modelle oder Daten verwendet wurden.



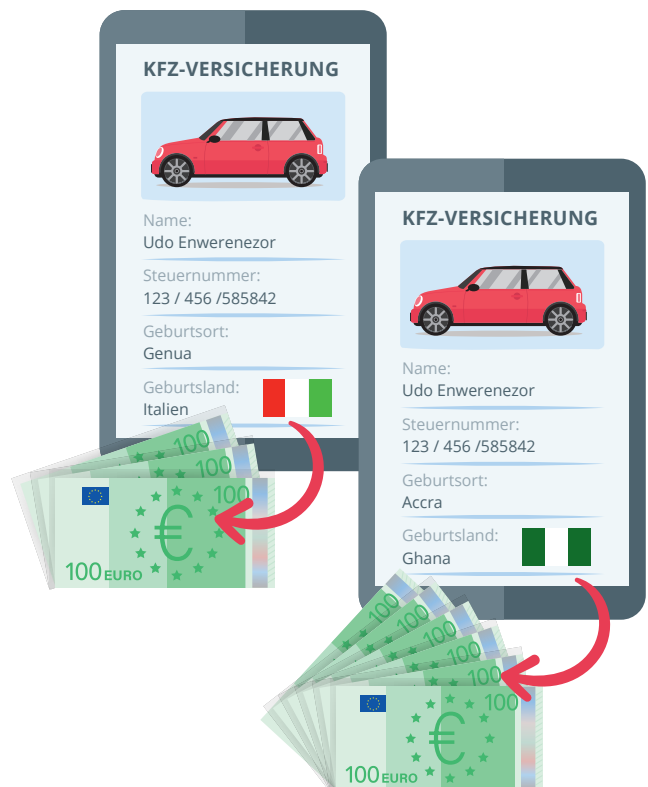
Versicherungen mit der Abfrage der Steuernummer schon über die Information verfügen, dass er sowohl eine nigerianische als auch eine italienische Staatsbürgerschaft besitzt. Um potenzielle Benachteiligung offenzulegen, führen die Universitäten Padua, Udine und Carnegie Mellon eine Studie mithilfe von Vergleichsportalen und fiktiven Profilen durch. Die Erfahrung von Udo Enwerenezor bestätigt sich: Abhängig vom Geburtsort und der Staatsbürgerschaft oder beidem variieren die Preise für Kfz-Versicherungen zum Teil stark.¹⁸ Die Studie zeigt, dass eine Person, die in Ghana geboren wurde, unter Umständen über 1.000 Euro mehr zahlen muss als eine in Mailand geborene Person mit einem ansonsten identischen Profil.



e) Preise für Autoversicherungen werden beeinflusst durch die Nationalität der Kund*innen

Kfz-Versicherungstarife werden durch die Staatsbürgerschaft beeinflusst.¹⁷

Udo Enwerenezor lebt in Italien und hat eine doppelte Staatsbürgerschaft: sowohl die Italienische als auch die nigerianische. Seine Erfahrung aus der Vergangenheit ist, dass die Preise für eine Kfz-Versicherung sich unterscheiden je nachdem, welche der Staatsbürgerschaften er angibt. Gibt er an, nigerianischer Staatsbürger zu sein, liegt der Preis höher. Seit die Kfz-Versicherung nicht mehr telefonisch abgeschlossen wird, sondern ausschließlich online, kann er nicht mehr nachprüfen, ob seine Staatsbürgerschaften sich unterschiedlich auf den Preis auswirken würden. Das liegt daran, dass die



¹⁷ Boscarol, Francesco (2022). Costly Birthplace: discriminating insurance practice. *AlgorithmWatch*: <https://algorithmwatch.org/en/discriminating-insurance/> (03.03.2022).

¹⁸ Fabris, Alessandro et al. (2021). Algorithmic Audit of Italian Car Insurance: Evidence of Unfairness in Access and Pricing. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 458-468). http://www.dei.unipd.it/~silvello/papers/2021_aies2021.pdf (03.03.2022).

3 Ursachen

*„Alles was programmiert wird, ist nur so gut wie die Menschen, die es programmieren.“ (Interviewpartner*in 07)*

Die Verwendung von ADM-Systemen kann zu ungerechten, sogar zu diskriminierenden Ergebnissen führen, die weitreichende, negative Effekte auf das Leben von Menschen haben können. ADM-Systeme sind weder neutral noch objektiv. Sie müssen als sozio-technische Systeme verstanden werden – das heißt als Systeme, die in einem spezifischen gesellschaftlichen Kontext entstehen und von den darin herrschenden Vorstellungen und Verhältnissen beeinflusst sind.

„Automatisierte Entscheidungssysteme werden von Menschen gemacht, die den gleichen Wahrnehmungen unterliegen, denen sie auch im Alltag unterliegen.“ (Interviewpartnerin 11)

Unbewusste Vorurteile, strukturelle Ungleichheit und Diskriminierung, die in Gesellschaften verbreitet sind, finden sich auch hier wieder. Denn sozio-technische Systeme spiegeln Annahmen, Werte, Perspektiven und Voreingenommenheit der Menschen, die sie entwickeln und anwenden. Das liegt daran, dass Menschen an verschiedenen Punkten in der Entwicklung und Anwendung dieser Systeme Entscheidungen treffen – im schlimmsten Fall kommt es so zu unerwünschter Diskriminierung. Wie falsche Vorstellungen und Voreingenommenheit die Funktionsweise von ADM-Systemen negativ beeinflussen können, erklären wir in den folgenden Abschnitten.

*„Dann die Frage: Wer hat das programmiert? Menschen, die keine Diskriminierungserfahrungen machen, sich damit nicht auseinandersetzen und nicht sensibel dafür sind?“ (Interviewpartner*in 07)*

Die folgenden Schaubilder in diesem Kapitel zeigen auf, an welchen Stellen Menschen Entscheidungen treffen, wie sie ihre Annahmen und Perspektiven dadurch einbringen und welche Fragen wir zu den ADM-Systemen stellen sollten.

Daten

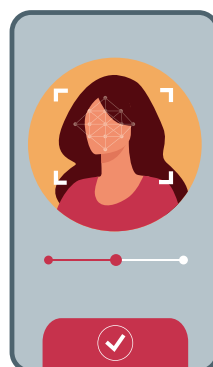
Die Realität lässt sich über Daten nie eins zu eins darstellen. Daten sind nicht neutral und spiegeln (Macht-)Verhältnisse der Gesellschaft wider. Menschen entscheiden, welche Daten wann und wie erhoben werden. Daten unterscheiden sich in ihrer Qualität zum Beispiel darin, ob sie aktuell oder veraltet sind und welche Gruppen in ihnen über- oder unterrepräsentiert werden. Daten sind in der Vergangenheit erfasst worden und spiegeln diese wider. Werden sie benutzt, um zum Beispiel die Zukunft vorherzusagen, beruht dies auf der Annahme, dass die Zukunft der Vergangenheit gleicht oder ähnelt. Vor allem bei lernenden Algorithmen beeinflussen die verwendeten Daten die Funktionsweise der ADM-Systeme stark. Menschen bereinigen und annotieren Daten, damit sie für lernende Algorithmen verwendbar sind.

Die verwendeten Daten könnten zum Beispiel der Grund dafür gewesen sein, dass das Gesicht von

DATEN

Sind die Daten repräsentativ für die Gruppen, auf die das ADM-System einen Effekt haben wird?

Wer wählt die Daten aus und warum wählt die Person diese Daten (z. B. Qualität der Daten, Verfügbarkeit oder Kosten)?



Sind die Daten aktuell und ausreichend? Wer prüft das?

Welche Fragen können mit den Daten beantwortet werden?

Wer hat die Daten annotiert und nach welchen Kriterien?



Audrey K. [Fallbeispiel c), S. 8] von der Gesichtserkennungssoftware des Fotoautomaten nicht erkannt wurde und sie deshalb kein biometrisches Foto machen konnte. Wenn der verwendete Algorithmus in der Phase des Lernens überwiegend mit Bildern von Gesichtern mit hellen Hauttönen trainiert wurde, diese also überrepräsentiert waren und damit Gesichter mit dunklen Hauttönen unterrepräsentiert, funktioniert das ADM-System schlechter oder gar nicht für die unterrepräsentierte Gruppe.

Ziel

Menschen entscheiden nicht nur über die verwendeten Daten, sie entscheiden auch über das Ziel der Verwendung eines ADM-Systems. Zunächst wird ein Problem identifiziert, das gelöst werden soll. Das kann zum Beispiel die Aufgabe sein, eine geeignete Person aus einem Bewerbungspool auszuwählen. Die Lösung könnte in einem von Menschen durchgeführten Recruiting-Prozess bestehen. Es kann jedoch auch entschieden werden, dass ein ADM-System für die Lösung des Problems geeignet scheint. Des Weiteren wird entschieden, welche Ziele das System verfolgen soll. Ziele können zum Beispiel Kostensenkung und Effizienzsteigerung sein. Ein Ziel kann aber auch sein, den Einfluss von Voreingenommenheit auf Entscheidungsprozesse zu reduzieren und Diskriminierung zu vermeiden.

Im [Fallbeispiel d), S. 9], zeigt Facebook Stellenanzeigen genderstereotyp den Nutzer*innen an. Das Ziel könnte hier sein, die Klicks auf Stellenanzeigen

für Anzeigenkunden zu maximieren. Basierend auf historischen Daten könnte die Wahrscheinlichkeit berechnet worden sein, dass Nutzerinnen eher auf eine Stellenanzeige „Erzieher*in“ klicken, während sich Nutzer eher für „LKW-Fahrer*in“ interessieren. Hätte Facebook zum Beispiel die Förderung von Geschlechtergerechtigkeit in der Gesellschaft als Ziel definiert, wäre die Anzeigenverteilung womöglich anders ausgefallen.

Modell

Ein Modell beschreibt einen Ausschnitt der Wirklichkeit; die Realität in ihrer Komplexität wird reduziert. Eine Situation wird so vereinfacht, dass sie in ein mathematisches Problem übersetzt werden kann. Menschen entscheiden, welche Kriterien und Aspekte relevant für ein Problem sind, das mithilfe des Modells gelöst werden soll – beispielsweise durch Anwendung eines Algorithmus. Wenn zum Beispiel geeignete Bewerber*innen automatisiert anhand ihrer Lebensläufe ausgewählt werden sollen und definiert wird, was „geeignet“ bedeutet, müssen die Kriterien festgelegt werden, an denen sich „geeignet“ festmachen lässt. Solche Kriterien könnten sein: Berufserfahrung, höchster Abschluss, Sprachkenntnisse usw. Dahinter steckt dann zum Beispiel die Annahme, dass eine Person mit mehreren Jahren Berufserfahrung besser für eine bestimmte Stelle geeignet ist als eine Person mit wenig Berufserfahrung. Diese Annahme könnte jedoch auch infrage gestellt werden, wenn bessere Kriterien für eine Eignung bestehen.





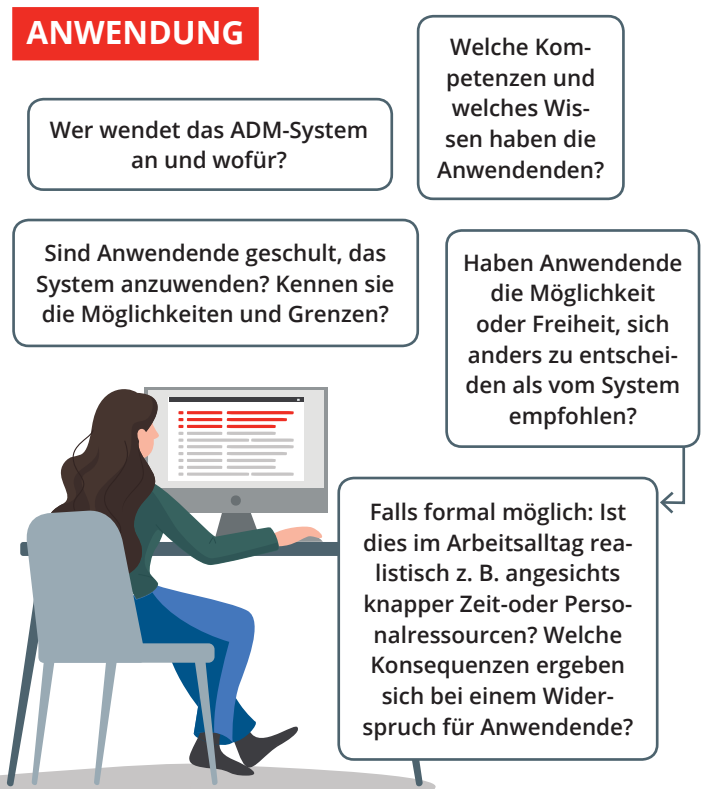
Bei Scoring-Verfahren der Auskunfteien werden Individuen einer Vergleichsgruppe mit ähnlichen Merkmalen zugeordnet. So kann basierend auf der Zahlungsmoral der Vergleichsgruppe berechnet werden, mit welcher Wahrscheinlichkeit eine Person eine Rechnung zahlen wird. Im [Fallbeispiel a\)](#) [S. 6] könnte das Modell die Ursache für das Ergebnis der ungenügenden Kreditwürdigkeit sein.¹⁹ Dann wäre hier die Annahme, dass Geschlecht und Alter einen Einfluss auf die Zahlungsfähigkeit einer Person haben, in das Modell eingegangen.

Anwendung

Die Entscheidung, ein ADM-System für eine Aufgabe einzusetzen, basiert auf Annahmen über die Möglichkeiten und Fähigkeiten dieser Systeme. Menschen tendieren dazu, ADM-Systeme zu überschätzen und sich auf Empfehlungen durch ADM-Systeme zu verlassen. Diese Tendenz wird in der englischsprachigen Fachliteratur „Automation Bias“ genannt.²⁰ Sie wird umso stärker, je mehr Zeitdruck Anwendende haben oder je mehr negative Konsequenzen sie fürchten müssen, wenn sie der Empfehlung des ADM-Systems widersprechen oder nicht folgen. Deswegen ist es sehr wichtig, dass Anwendende geschult werden

und lernen, wie diese Systeme funktionieren, damit sie die Empfehlungen angemessen hinterfragen können.

Abschließend wollen wir noch darauf hinweisen, dass selbst ein auf Diskriminierung und Voreingenommenheit geprüftes System diskriminierend eingesetzt werden kann. Wenn zum Beispiel eine Gesichtserkennungssoftware entwickelt wird, die für alle Menschen sehr gut funktioniert, diese jedoch vor allem benutzt wird, um Schwarze Menschen gezielt zu kontrollieren, dann ist die Anwendung des ADM-Systems rassistisch.



¹⁹ Da wir keine Informationen von Seiten der Auskunftei zu dem Fall oder dem verwendeten Modell haben, kennen wir die tatsächlichen Ursachen für das Ergebnis nicht.

²⁰ Automation Bias untersucht beispielsweise Cummings, Mary (2004). Automation Bias in Time Critical Decision Support Systems. *AIAA 1st Intelligent Systems Technical Conference*. <https://web.archive.org/web/20141101113133/http://web.mit.edu/aeroastro/labs/halab/papers/CummingsAIAAbias.pdf> (14.04.2022).

4 Anwendungsfelder²¹

„Wir wissen ja oft nicht, wo ein Algorithmus dahinterstecken kann und wo nicht.“ (Interviewpartnerin 07)

In diesem Schaubild sind Bereiche dargestellt, in denen ADM-Systeme angewendet werden können und Beispiele für deren Anwendung. Hier wird aus Gründen der Übersichtlichkeit nicht darauf eingegangen, wie die abgebildeten ADM-Systeme zu diskriminierenden Ergebnissen führen können. Das Schaubild soll in Beratungsfällen herangezogen werden und helfen, einen begründeten Verdacht für die Verwendung von ADM-Systemen zu entwickeln. Der

21 Bitte beachten: Dieses Kapitel ist als Doppelseite gestaltet.

Fokus liegt auf privaten Dienstleistungen und dem Zugang zu Gütern, Bereiche darüber hinaus werden nur grob dargestellt. Es ist schwierig, mit Sicherheit zu sagen, wo ADM-Systeme derzeit in Deutschland tatsächlich eingesetzt werden, denn es fehlen Transparenzpflichten und Auskunftsrechte – weder muss angegeben werden, dass ein ADM-System zur Unterstützung angewendet wird, noch auf welchen Annahmen und Daten dieses System basiert. Ein Hinweis darauf, dass automatisierte Systeme angewendet werden, ist zum Beispiel, dass schon lange statistische Berechnungen durchgeführt werden oder Datensätze vorhanden sind, wie bei Versicherungen oder Bonitätsprüfungen. Oft soll mithilfe automatisierter Entscheidungssysteme etwas sortiert, kategorisiert, vorhergesagt oder effizienter gemacht werden.



Angebote Zahlungs-möglichkeiten:

Zahlungsmöglichkeiten könnten sich aufgrund der angegebenen Wohnadresse verändern – bei einer Adresse wird die Option „Kauf auf Rechnung“ angezeigt, bei einer anderen Adresse nicht.



Kundenkonten:

Beim Erstellen eines Kundenkontos teilen Kund*innen persönliche Informationen, wie Geburtsdatum, Geschlecht oder Staatsbürgerschaft. Diese Informationen könnten von einem ADM-System verarbeitet werden.



Versicherungssektor

- Online-Formulare und Kundenkonten
- Chatbot-Interaktionen
- Gruppenspezifische Versicherungstarife (Studierende, Senior*innen)
- Verhaltensbasierte Versicherungstarife (z. B. Telematik)
- Automatisierte Überprüfung eingereichter Unterlagen
- Dauerschuldverhältnisse: Bonitätsprüfung
- Online abgeschlossene Versicherungen, z. B. Reisrücktritts-, Haftpflicht-, Kfz-, Rechtsschutzversicherung
- Automatisierte Abwicklung von Schadensmeldungen („digitale*r Sachbearbeiter*in“)



Finanzsektor

- Online-Formulare und Kundenkonten
- Chatbot-Interaktionen
- Kreditvergabe: Bonitätsprüfung
- Automatisierte Überprüfung eingereichter Unterlagen
- Online abgeschlossene Finanzprodukte, z. B. Girokonto, Kreditkarte, Depot, Online-Kredit
- Online-Banken und Fin-Techs



Online-Handel

- Online-Formulare und Kundenkonten
- Chatbot-Interaktionen
- Gruppenspezifische Preise, z. B. Studierende, Senior*innen)
- Prämien basierend auf Kundenprofilen
- Preisdifferenzierung basierend auf Uhrzeit oder Ort des Kaufs
- Datenbasierte, gezielte Werbung (z. B. basierend auf vorherigem Suchverlauf)
- Dauerschuldverhältnisse bei Mobilfunk-, Internetverträgen, Abonnements (z. B. Spotify, Netflix): Bonitätsprüfung
- Angebotene Zahlungsmöglichkeiten, z. B. basierend auf Adresse
- Ratenzahlung oder Zahlung auf Rechnung: Bonitätsprüfung



Chatbot-Interaktionen: Über ein Dialogfenster können Fragen an den Chatbot gestellt werden, die automatisiert beantwortet werden. Teilweise können Chatbots durch Anwendungen maschinellen Lernens aus Eingaben lernen und verändern ihre Antworten mit der Zeit.



Vergabe von Kitaplätzen: Basierend auf vorab definierten Vergabekriterien könnte ein ADM-System empfehlen, welche Kinder in welcher Kita einen Platz bekommen sollten.



Bewerbungsverfahren: Eine Spracherkennungssoftware könnte eingesetzt werden, um anhand von Stimmproben Rückschlüsse auf die sprechende Person zu ziehen.



Mustererkennung: Ein ADM-System könnte zur Mustererkennung eingesetzt werden, mit dem Ziel durch Bildanalyse Melanome frühzeitig anhand von Bildern der Haut von Menschen zu erkennen.



Wohnungssektor

- Online-Formulare und Kundenkonten
- Chatbot-Interaktionen
- Mietverhältnis: Bonitätsprüfung
- Vorsortierung Wohnungsbewerbungen (bei großer Anzahl an Mietwohnungen oder Bewerbungen)
- Wohnungsanzeigen auf Online-Plattformen: Wem werden welche Inserate angezeigt?



Gesundheitssektor

- Private Krankenversicherung und Zusatzversicherungen, z. B. Zahn- oder Krankenhauszusatzversicherung (siehe Versicherungssektor)
- Online-Terminvergabe, z. B. für Covid-Impfungen
- Prävention und Diagnostik: Mithilfe von Mustererkennung sollen Erkrankungen frühzeitig erkannt werden.



Bildungssektor

- Automatisierte Kitaplatzvergabe
- Automatisierte Auswahl der Bewerbungen für private Schulen, Hochschulen, Universitäten



Arbeitsmarkt

- Bewerbungsverfahren (z. B. Sprachanalyse, Videoanalyse)
- Automatisierte Auswahl der Bewerbungen anhand von Noten, Lebenslauf usw.
- Stellenausschreibungen über Online-Portale
- Automatisierte Platzierung
- People Analytics, z. B. Leistungsevaluation, Personalentwicklung, Berechnen der Wechselbereitschaft)








Gruppenspezifische Versicherungstarife und Formulare: Basierend auf persönlichen Angaben, wie Geburtsdatum oder Beschäftigungsverhältnis, könnten automatisiert unterschiedliche Tarife berechnet werden.



Zahnzusatzversicherung: Angaben aus Online-Formularen oder dem Kundenkonto könnten verwendet werden, um automatisiert einen Versicherungstarif zu berechnen.



-  ■ Wohnungssuche über Online-Wohnungsportale
-  ■ Online abgeschlossene Verträge wie Versicherungsverträge
-  ■ Interaktion mit einem Chatbot, z. B. im Online-Handel
-  ■ ABER: Bonitätsprüfung bei der Kreditvergabe oder Video-/Sprachaufnahmen bei Bewerbungsverfahren sind ebenfalls Beispiele, in denen ADM-Systeme zum Einsatz kommen können, auch wenn Interaktionen nicht online stattgefunden haben.
- 

5 Erkennen

„Wie lerne ich zu erkennen, ob möglicherweise verborgene digitale Prozesse dahinterliegen?“ (Interviewpartner 10)

Der folgende Fragenkatalog soll helfen, in den Fallbeschreibungen von Ratsuchenden oder Betroffenen Hinweise auf [ADM-Systeme](#) zu erkennen und bei Verdacht passende Rückfragen zu stellen. Dafür sind keine technischen Vorkenntnisse erforderlich.

Wo fand die Interaktion statt?

- Wenn die Interaktion online stattfand, kann das ein Hinweis auf den Einsatz von automatisierten Systemen sein. Beispiele für solche Online-Interaktionen:

Wurde ein Online-Formular ausgefüllt oder über eine Online-Maske (Kontaktformular, Chatbot etc.) interagiert?

- Beispiele hierfür sind: Das Eröffnen eines Depots mithilfe eines Online-Kontos oder das Abschließen einer Reiseversicherung über eine Webseite.
- Wurden persönliche Angaben gemacht oder personenbezogene Daten geteilt? Falls ja, welche? (Solche Daten können auch in Online-Konten gespeichert sein.)
- Wurden Fragen nach Merkmalen gestellt, die durch das AGG geschützt sind (unmittelbare

PROXY-VARIABLEN

Proxy-Variablen werden stellvertretend für andere Variablen eingesetzt, zum Beispiel wenn diese nicht direkt beobachtet- oder messbar sind. Das englische Wort „proxy“ bedeutet „Stellvertreter*in“. Mit Proxy-Variablen kann auf bestimmte – auch geschützte – Merkmale geschlossen werden, ohne dass sie direkt abgefragt oder erfasst werden. Aus der Angabe „30 Jahre Arbeitserfahrung“ lässt sich beispielsweise ablesen, dass die Person mindestens Mitte 40 sein muss. So können Proxy-Variablen einen entscheidenden Einfluss auf die Entscheidungen eines ADM-Systems haben. ([Hier finden Sie ein Erklärvideo zu Proxy-Variablen](#)).

„Ein wichtiger Punkt: Oft wird davon ausgegangen, dass es nicht zu einer Diskriminierung kommt, wenn keine personenbezogenen Daten verwendet werden. Das stimmt so nicht.“ (Interviewpartnerin 02)

Diskriminierung)? Wurde nach Angaben gefragt, die als Proxy-Variablen wirken könnten (mittelbare Diskriminierung)?

- Wurden Dokumente hochgeladen? Falls ja, welche? Beinhalteten diese Hinweise auf oder Informationen zu geschützten Merkmalen?

Wie sah die Rückmeldung der Firma oder Organisation auf die Interaktion aus?

- Wie schnell kam die Rückmeldung? Folgende Hinweise deuten auf ADM-Systeme:
 - Fällt sofort beim Eingeben von Daten in ein Online-Formular eine Entscheidung oder wird sie unmittelbar im Anschluss per E-Mail mitgeteilt?
 - Wurden viele Dokumente hochgeladen und die Rückmeldung kam dennoch sehr schnell?
- Kommen in der Rückmeldung bestimmte Begriffe oder Phrasen vor, die auf eine automatisierte Entscheidung hindeuten?
 - Eine gängige Phrase ist zum Beispiel: „Dieses Schreiben wurde maschinell erstellt und ist ohne Unterschrift gültig.“
 - Häufige Begriffe sind: „automatisiert“, „maschinell“, „computergestützt“, „automatisch“, „elektronisch“.

Falls online etwas gekauft wurde, welche Zahlungsmöglichkeiten gab es und welche wurde ausgewählt?

- Bei den Optionen „Ratenzahlung“ oder „Bezahlung auf Rechnung“ ist es wahrscheinlich, dass eine Bonitätsprüfung vorgenommen wird.
- Handelt es sich um ein Dauerschuldverhältnis, wie Miete, Telekommunikationsvertrag, Abonnements bei Netflix oder Spotify usw.? Auch hier ist es möglich, dass eine Bonitätsprüfung vorgenommen wird.

- Auch bei der Teilnahme an Lastschriftverfahren kann eine Bonitätsprüfung erfolgen.

Einer Bonitätsprüfung muss in der Regel explizit zugestimmt werden. Diese Zustimmung wird meist im Rahmen der Zustimmung zum Kauf oder der Zahlung eingeholt. Die Information dazu kann in den Vertragsunterlagen, den Allgemeinen Geschäftsbedingungen (AGB) oder ähnlichen Dokumenten stehen.

Wurde der automatisierten Verarbeitung von Daten explizit zugestimmt?

- Damit vollautomatisch entschieden werden darf, muss die betroffene Person nach Artikel 22 des europäischen Datenschutzgesetzes (DSGVO) oder Paragraph 31 des Bundesdatenschutzgesetzes (BDSG) explizit zustimmen. Die Information dazu kann in den Vertragsunterlagen, den AGB, der Datenschutzvereinbarung oder ähnlichen Dokumenten stehen. Schauen Sie auch nach Wörtern und Phrasen wie: „automatisiert“, „maschinell“, „computergestützt“, „automatisch“, „mithilfe von Algorithmen“.
- Ein Beispiel aus der Datenschutzerklärung einer Online-Buchhandlung:
 - „Auf Basis dieser Daten berechnet unser Dienstleister anhand mathematisch-statistischer Verfahren Wahrscheinlichkeitswerte zur Beurteilung des Zahlungsausfallrisikos und zur Verifizierung Ihrer Adresse (Prüfung auf Zustellbarkeit).“

Ich würde vermuten, dass Diskriminierungsrisiken durch automatisierte Entscheidungssysteme sehr, sehr große Relevanz haben, aber nicht bei uns in der Beratung ankommen oder ich sie als Beraterin nicht erkenne.“ (Interviewpartnerin 00)

6 Umgang

*„Was mache ich denn dann, wenn ich Automatisierung vermute? Ist das ein Indiz für Diskriminierung? Das ist natürlich eine große Frage.“
(Interviewpartnerin 00)*

Wenn Sie im Rahmen eines Beratungsfalls vermuten, dass ein automatisiertes Entscheidungssystem verwendet wurde, ist die naheliegende Frage: Welche Konsequenzen hat das für den weiteren Beratungsprozess? In diesem Teil der Publikation beschreiben wir Ideen für den Umgang mit einem solchen Verdacht, anknüpfend an gängige Interventionen der Antidiskriminierungsberatung. Wie in allen anderen Beratungsfällen geht es zunächst darum, herauszufinden und umzusetzen, was die ratsuchende Person möchte.²⁴ Darüber hinaus sollten Sie sinnvollerweise einen Hinweis auf ein ADM-System in die Dokumentation des Beratungsfalls aufnehmen. Dadurch lassen sich die Fälle besser bündeln und eventuelle Muster erkennen.

24 Antidiskriminierungsstelle des Bundes (2012). Schritt für Schritt durch die rechtliche Antidiskriminierungsberatung – Leitfaden für Beraterinnen und Berater: S. 30-59. Antidiskriminierungsverband Deutschland (2015). Standards für eine qualifizierte Antidiskriminierungsberatung: S. 12f.

Wie lässt sich ein Verdacht auf negative Auswirkungen eines ADM-Systems in gängige Vorgehensweisen und Interventionen der Antidiskriminierungsberatung integrieren?

Kontaktaufnahme, Stellungnahme, Beschwerdeschreiben

Eine Kontaktaufnahme mit einer verantwortlichen Person, Organisation oder Stelle, in Form einer Bitte um Stellungnahme (etwa mit dem Ziel der Einigung) oder ein Beschwerdeschreiben bieten sich an, wenn dies im Sinne der ratsuchenden Person ist.

Besteht in einem konkreten Fall ein Verdacht, dass ein Algorithmus oder ADM-System zu der diskriminierenden oder ungerechten Entscheidung beigetragen hat, beschreiben Sie diesen Verdacht bei der Kontaktaufnahme. Sie können dafür folgende Formulierungshilfen und Musterfragen verwenden:

„Aufgrund von Hinweisen gehen wir davon aus, dass in diesem Fall ein automatisiertes Entscheidungssystem verwendet wurde, das heißt, dass Entscheidungsmodelle automatisiert durch einen Algorithmus oder ein Computerprogramm angewendet wurden. Haben Sie ein solches System verwendet?“

WAS KÖNNEN SIE TUN?

Das Diskriminierungsrisiko durch ADM-Systeme ist ein relativ neues Problem in der Antidiskriminierungsberatung. Fälle können sich in Abhängigkeit vom verwendeten ADM-System stark unterscheiden. Derzeit existieren erst wenige Fallbeispiele aus Deutschland. Dementsprechend gibt es wenig Erfahrungswissen, auf das Sie zurückgreifen können.

Deswegen empfehlen wir, bei einem Verdacht auf Auswirkungen eines ADM-Systems als erstes eine der folgenden Stellen zu kontaktieren, die sich eingehend mit dieser Problematik beschäftigen. Sie können zum einen direkt bei der Fallbearbeitung unterstützen und zum anderen an passende Stellen (etwa Verbraucher- oder Datenschutzstellen) weiterverweisen:

AlgorithmWatch

✉ info@algorithmwatch.org
☎ 030-99 40 49 000

Antidiskriminierungsstelle des Bundes

✉ beratung@ads.bund.de
(im Betreff: „Bitte um kollegiale Beratung zu KI-Fall“)

Antidiskriminierungsverband Deutschland (advd)

✉ info@antidiskriminierung.org

Falls ja, um welche Software handelt es sich? Fand in diesem Fall eine vollautomatisierte Entscheidung statt oder wurde die Entscheidung automatisiert vorbereitet und dann durch einen Menschen getroffen oder kontrolliert? Wie wurde das automatisierte System gegen Diskriminierung und andere ungewollte Einflüsse gesichert? Wie wurde sichergestellt, dass keine personenbezogenen Merkmale verwendet wurden?“

Testing-Verfahren²⁵

Eine Methode, um Diskriminierung zu überprüfen oder sichtbar zu machen, ist das Testing-Verfahren. Bei einem Verdacht auf Diskriminierung bei der Wohnungssuche zum Beispiel funktioniert es folgendermaßen: „Bei Testings bewerben sich jeweils eine Test- und eine Kontrollperson um eine Wohnung. Dieses Tester_innenpaar weist möglichst viele Ähnlichkeiten auf, unterscheidet sich aber in dem zu testenden Merkmal. Personen, die den Verdacht haben, auf dem Wohnungsmarkt oder in anderen Bereichen ungerechtfertigt benachteiligt [worden] zu sein, können Einzelfalltestings anwenden. Die durch das Testing gewonnenen Indizien für eine Diskriminierung können dann zur Anspruchs begründung gerichtlich verwendet werden.“²⁶ Die Methode ist auf andere Anwendungsbereiche des AGG übertragbar, zum Beispiel auf die Kreditvergabe.

Die Testing-Methode bietet sich im Zusammenhang mit automatisierten Entscheidungen an, da die Entscheidungen immer nach dem gleichen grundlegenden Muster – also systematisch – entschieden

werden. Testing-Verfahren sind jedoch mitunter sehr aufwendig und anspruchsvoll, zum Beispiel bei Bonitätsprüfungen.



FALLBEISPIEL TESTING-VERFAHREN

Die Frau aus [Fallbeispiel a), S. 6] fühlt sich diskriminiert und wendet sich an eine Antidiskriminierungsstelle. Der Fall bietet sich nach Ansicht der Beraterin für ein Testing an, da das Scoring-Verfahren immer der gleichen Logik folgt. Um aussagekräftig zu sein, müsste das Testing jedoch groß angelegt werden mit einer sehr hohen Anzahl von Vergleichspersonen. Für die Beratungsstelle ist das zum damaligen Zeitpunkt zu aufwendig und ressourcenintensiv.

Es muss unterschieden werden, ob die Ergebnisse des Testing-Verfahrens als Indiz in ein Klageverfahren oder unterstützend für eine Stellungnahme verwendet werden sollen. Leider gibt es außerhalb von Diskriminierung am Wohnungsmarkt kaum Rechtsprechung, die Hinweise darauf gibt, ob die Ergebnisse eines solchen Testing-Verfahrens vor Gericht als Indiz anerkannt würden. Die Hürden vor Gericht sind hoch.

Großes Potenzial könnten die Ergebnisse eines Testing-Verfahrens dagegen im Rahmen einer Aufforderung zur Stellungnahme haben und für die Öffentlichkeitsarbeit rund um einen solchen Fall. Es muss jedoch immer entschieden werden, ob genügend Ressourcen für ein solches Verfahren vorhanden sind.

Bei Online-Formularen lässt sich ein Testing relativ einfach durchführen, da Sie Angaben leicht ändern können. Sollte zum Beispiel bei einer Wohnungsanfrage über ein Wohnungsportal der Verdacht bestehen, dass aufgrund des Geschlechts diskriminiert wird, kann das Geschlecht testweise geändert werden. So kann überprüft werden, ob die Rückmeldung beziehungsweise Entscheidung sich je nach angegebenem Geschlecht ändert.

25 Weiterführende Literatur zur Testing-Methode: Senatsverwaltung für Arbeit, Integration und Frauen. Landesstelle für Gleichbehandlung – gegen Diskriminierung. Diskriminierung sichtbar machen (2012). Dokumentation der LADS-Fachrunde vom 8.11.2012 zu fachlichen und methodologischen Anforderungen an Testing-Verfahren. Antidiskriminierungsstelle des Bundes (2020). Fair mieten – fair wohnen. Leitfaden für Mieterinnen und Mieter und Beratungsstellen.

26 Senatsverwaltung für Arbeit, Integration und Frauen. Landesstelle für Gleichbehandlung – gegen Diskriminierung. Diskriminierung sichtbar machen (2012). Dokumentation der LADS-Fachrunde vom 8.11.2012 zu fachlichen und methodologischen Anforderungen an Testing-Verfahren, S. 10.

**HINWEIS**

Testings können sehr aufwendig sein und binden viele Ressourcen. Es ist sinnvoll, Verdachtsfälle mit Bezug zu automatisierten Entscheidungen zu bündeln und etwa bei Vernetzungstreffen zu besprechen. Dafür ist es hilfreich, ADM-Systeme explizit in die Falldokumentation mit aufzunehmen, zum Beispiel in der Fallbeschreibung oder als eigene Kategorie in der Falldatenbank.


Möchten Sie sich mit Kolleg*innen austauschen, die Erfahrungen mit Testing-Verfahren haben? Folgende Beratungsstellen haben in der Vergangenheit Testing-Verfahren durchgeführt:

Planerladen gGmbH,

Schwerpunkt Wohnungsmarkt

 www.integrationsprojekt.net

 integration@planerladen.de

 0231-833 225

Antidiskriminierungsbüro Sachsen e. V.

 <https://www.adb-sachsen.de/>

 beratung@adb-sachsen.de  0341-306 907 77

**Antidiskriminierungsnetzwerk Berlin des
Türkischen Bundes in Berlin-Brandenburg**

 <https://www.adnb.de/>

 adnb@tbb-berlin.de  030-61 30 53 28

Sichtbarmachung: Öffentlichkeits- und Pressearbeit

Da es sich um ein relativ neues Thema handelt und bisher wenige Fallbeispiele aus Deutschland bekannt sind, ist Presse- und Öffentlichkeitsarbeit bei einem auftretenden Fall wichtig. Durch das Bekanntmachen der Fälle werden Menschen für das Thema und die Problematik sensibilisiert und eine öffentliche Diskussion unterstützt. Auch für die Aufmerksamkeit auf politischer Ebene ist eine Bekanntmachung von Diskriminierungsfällen durch automatisierte Entscheidungsfindungsprozesse wichtig. Sammlungen von Fallbeispielen tragen zu einem klareren Lagebild über den Einsatz von ADM-Systemen und den damit verbundenen Risiken bei.

Wenn die Antidiskriminierungsberatungsstelle, für die Sie arbeiten, selbst keine Öffentlichkeitsarbeit macht, wenden Sie sich an AlgorithmWatch, die Antidiskriminierungsstelle des Bundes oder den Antidiskriminierungsverband Deutschland [[Kontakt Daten](#), S. 18] für Unterstützung bzw. Weitervermittlung.

Klageverfahren und strategische Prozessführung

Sollten Sie einen Fall im Bereich der privaten Dienstleistungen und dem Zugang zu Gütern haben, der an ein AGG-geschütztes Merkmal geknüpft ist, gibt es auch die Option eines Klageverfahrens²⁷ durch die betroffene Person. Zu bedenken ist jedoch, dass Klageverfahren mit Kosten und Belastung für die betroffene Person einhergehen und von Seiten der Antidiskriminierungsberatung Ressourcen für die Beratungsarbeit benötigt werden. Die benötigten Ressourcen sind vergleichbar mit anderen Bereichen, in denen es noch wenig oder keine Rechtsprechung gibt.

Die Hauptschwierigkeit ist, die Diskriminierung nachzuweisen oder so viele Indizien zu sammeln, dass sich die Beweislast umkehrt.²⁸ Sollte sich die betroffene Person für eine Klage entscheiden und Sie den Fall beratend begleiten, gilt es mit der anwaltlichen Vertretung zu besprechen, inwiefern die Aufmerksamkeit darauf gerichtet werden kann, dass (teil-) automatisiert entschieden wurde. Das könnte die Presse- und Öffentlichkeitsarbeit rund um den Fall begünstigen.

27 Hinweis: Eine Klage nach dem AGG setzt die fristgerechte Geltendmachung der Ansprüche voraus. Andernfalls ist die Klage unzulässig. Gemäß § 21 AGG muss die Geltendmachung innerhalb von zwei Monaten erfolgen, um sich die Option einer Klage offenzuhalten.

28 Ein Policy Brief, in dem AlgorithmWatch diskutiert, wie das AGG angepasst werden könnte, um Betroffene vor Diskriminierung durch ADM-Systeme rechtlich besser zu schützen: Michot, Sarah et al. (2022). Algorithm-basierte Diskriminierung – Warum Antidiskriminierungsgesetze jetzt angepasst werden müssen. *AlgorithmWatch/Digital Autonomy Hub*. https://algorithmwatch.org/de/wp-content/uploads/2022/02/DAH_Policy_Brief_5.pdf (14.04.2022).

Darüber hinaus könnte sich je nach Fall und verfügbaren Ressourcen eine strategische Prozessführung anbieten. Strategische Prozessführung im Antidiskriminierungsrecht „beschreibt eine Methode, bei der Gerichtsverfahren systematisch genutzt werden, um einen gesellschaftlichen Wandel herbeizuführen. Dabei werden taktisch juristische Mittel und flankierende Maßnahmen kombiniert, um eine über den Einzelfall hinausgehende Wirkung zu erzielen.“²⁹ Strategische Prozessführung kann angewendet werden, um gegen diskriminierende Strukturen vorzugehen. Abhängig vom jeweiligen Fall einer automatisierten Entscheidung mit diskriminierendem Ergebnis könnte sich eine strategische Prozessführung eignen, da automatisierte Entscheidungen systematisch sind und oft bestehende diskriminierende Strukturen reproduzieren.

DIE KI-VERORDNUNG DER EU

Derzeit wird auf EU-Ebene die Verordnung zu Künstlicher Intelligenz (Artificial Intelligence Act) verhandelt.³⁰ Das Gesetz soll die Verwendung von KI-Systemen regulieren, insbesondere jene, die mit einem hohen Risiko einhergehen. Der Vorschlag legt einen Rechtsrahmen fest, der den Schutz der Grundrechte gewährleisten und gleichzeitig Innovation fördern soll.

Da der breite Einsatz von ADM-Systemen ein relativ neues Phänomen ist, sind im Moment noch viele rechtliche Fragen offen. Derzeit wird diskutiert, welche Rolle der Daten- oder Verbraucherschutz zusätzlich zum oder im Zusammenspiel mit dem AGG haben könnte. Darüber hinaus verhandelt und

beschließt die Europäische Union derzeit verschiedene Gesetzesentwürfe, die einen Einfluss auf den Einsatz von ADM-Systemen im Zusammenhang mit Antidiskriminierungsrecht haben werden, vor allem die KI-Verordnung (Artificial Intelligence Act) und das Digitale-Dienste-Gesetz (Digital Services Act).

DIGITALE-DIENSTE-GESETZ DER EU

Die EU-Institutionen haben sich kürzlich auf die finale Fassung des Digitale-Dienste-Gesetzes (Digital Services Act, kurz DSA) verständigt. Es legt neue Verpflichtungen für Anbieter*innen digitaler Vermittlungsdienste fest, insbesondere für sehr große Online-Plattformen. Diese Vorschriften sollen ein sichereres Online-Umfeld fördern, in dem es Menschen möglich ist, ihre Grundrechte online wahrzunehmen, einschließlich des Rechts auf Nichtdiskriminierung. Zu den spezifischen Bestimmungen im Zusammenhang mit Nichtdiskriminierung gehören Vorschriften, die das gezielte Werben durch Targeting eindämmen, überarbeitete Meldemechanismen sowie neue Maßnahmen zur Transparenz und Rechenschaftspflicht von Plattformen. Plattformen wie Facebook, Google oder Amazon sollen bewerten, ob von ihnen „systemische Risiken“ ausgehen, sich Audits unterziehen und für Forschung im öffentlichen Interesse Daten zur Verfügung stellen.

HINWEIS

Um rechtliche Fragen im Hinblick auf den Diskriminierungsschutz durch das AGG bei Diskriminierung durch Algorithmen in den Blick zu nehmen, hat die ADS das Rechtsgutachten „Das Allgemeine Gleichbehandlungsgesetz und der Schutz vor Diskriminierung durch algorithmische Entscheidungssysteme“ (Ausschreibungstitel) in Auftrag gegeben. Das Gutachten wird von Prof. Dr. iur. Indra Spiecker gen. Döhm und Prof. Dr. Emanuel V. Towfigh erstellt und soll Ende 2022/Anfang 2023 auf der [Publikationsseite der ADS](#) veröffentlicht werden.

29 Siehe Beigang, Steffen et al. (2021). Möglichkeiten der Rechtsdurchsetzung des Diskriminierungsschutzes bei der Begründung, Durchführung und Beendigung zivilrechtlicher Schuldverhältnisse: Bestandsaufnahme, Alternativen und Weiterentwicklung. *Nomos*: S. 159. https://www.antidiskriminierungsstelle.de/SharedDocs/downloads/DE/publikationen/Expertisen/rechtsdurchsetzung_zivilrecht.pdf?__blob=publicationFile&v=2 (03.03.2022).

30 Link zum Gesetzesentwurf: <https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:52021PC0206> (14.04.2022).

7 Glossar

In diesem Glossar werden grundlegende Begriffe erläutert. Es ist uns wichtig, dass die gesamte Publikation auch ohne technisches Wissen verständlich ist. Mit den Begriffsklärungen wollen wir zu einem grundlegenden Verständnis der Funktionsweise von ADM-Systemen beitragen, das im Kontext der Antidiskriminierungsarbeit sinnvoll ist.

Algorithmus – ist eine bestimmte Form der Handlungsanleitung, die in definierten Einzelschritten zur Lösung eines mathematischen Problems führt. Algorithmen wurden lange vor der Entwicklung von Computern verwendet. Im Kontext dieser Publikation beziehen wir uns jedoch auf Algorithmen, die in verschiedenen sozio-technischen Systemen verwendet werden und deren Funktionsweise beeinflussen. Algorithmen beschreiben „einen für den Computer korrekt interpretierbaren Lösungsweg, der für jede durch das mathematische Problem definierte mögliche Eingabe die korrekte Lösung in endlicher Zeit berechnet.“³¹

ADM-Systeme / Automatisierte Entscheidungssysteme (Automated Decision-Making Systems, kurz: ADM-Systeme) – Entscheidungsmodelle oder Entscheidungswege werden automatisiert durch Algorithmen ausgeführt. ADM-Systeme können Menschen bei Entscheidungen unterstützen, indem sie Empfehlungen aussprechen oder durch das Aufbereiten von Daten Entscheidungen vorbereiten. Die Ausführung einer Entscheidung kann auch vollständig an ein ADM-System delegiert werden, dann ist bei der Anwendung der Systeme kein Mensch beteiligt.

Chatbot – ist ein Dialogsystem, das auf Algorithmen basiert. Chatbots werden zunehmend im Kundenservice und für die Interaktion im Internet eingesetzt. Fragen, die über ein Dialogfenster auf einer Webseite gestellt werden, können durch einen Chatbot automatisiert beantwortet werden. Die Kommunikation zwischen Chatbot und Menschen passiert in natürlicher Sprache. Teilweise verarbeiten Chatbots sogar gesprochene Sprache. Chatbots, die auf Methoden des maschinellen Lernens basieren, können aus Eingaben in der Anwendung dazulernen.

Daten – sind Zahlenwerte und Angaben, die durch Messungen, statistische Erhebungen, Beobachtungen oder Ähnliches erhoben wurden. Die Realität lässt sich über Daten nie eins zu eins darstellen. Menschen entscheiden, welche Daten wie erhoben und verwendet werden, und damit auch, was durch die Daten herausgefunden werden kann.³² Daten spiegeln (Macht-) Verhältnisse einer Gesellschaft wider und sind nicht neutral.

Daten annotieren – ist der Prozess, Daten zur Verwendung für lernende Algorithmen vorzubereiten, indem Daten zum Beispiel einer bestimmten Kategorie oder einem Merkmal zugeordnet werden. Annotierte Daten markieren das richtige Ergebnis für eine Aufgabe. Die Annotationen werden oft von Menschen vergeben. Wenn zum Beispiel ein Algorithmus trainiert wird, der am Ende Katzenbilder identifizieren soll, sind Bilder notwendig, die als Katzenbilder annotiert sind. Daraus „lernt“ der Algorithmus die richtige Antwort.

Erklärbare KI – ist ein Teilgebiet der Künstlichen Intelligenz (KI). Bei vielen KI-Systemen ist für Menschen nicht nachvollziehbar, aufgrund welcher Logik oder Kriterien entschieden, empfohlen oder vorhergesagt wird. Im Forschungsfeld der erklärbaren KI wird versucht, diese Logik und einzelne Ergebnisse besser nachvollziehbar zu machen. Nachvollziehbarkeit ist notwendig, um verzerrte und ungerechte Ergebnisse besser zu erkennen und zu vermeiden.

31 AlgorithmWatch (2016). 1. Arbeitspapier: Was ist ein Algorithmus? <https://algorithmwatch.org/de/arbeitspapier-was-ist-ein-algorithmus/> (28.02.2022). Eine ausführliche Definition und Beschreibung im Kontext der Antidiskriminierungsarbeit finden Sie bei: Orwat, Carsten (2019). Diskriminierungsrisiken durch Verwendung von Algorithmen: eine Studie, erstellt mit einer Zuwendung der Antidiskriminierungsstelle des Bundes. *Nomos*. S. 3-5.

32 Für mehr Informationen und Studien zu diesem Thema: Criado-Perez, Caroline (2020). Unsichtbare Frauen – Wie eine von Daten beherrschte Welt die Hälfte der Bevölkerung ignoriert. *btb*.

FinTech – ist die Abkürzung des Begriffs „Financial Technology“ und umfasst neue Produkte oder Dienstleistungen im Finanzsektor, die auf Algorithmen und automatisierten Entscheidungssystemen basieren.

Künstliche Intelligenz (KI) – ist ein Forschungsfeld der Informatik, das sich mit maschinellem Lernen und der Automatisierung kognitiver Aufgaben befasst. Das allgemeine Ziel dieses Forschungsfelds ist es, Maschinen zu befähigen, Aufgaben auszuführen, die irgendwie ‚intelligent‘ erscheinen. Was ‚Intelligenz‘ dabei bedeutet und mit welchen Techniken man sie erreichen kann, ist nicht definiert.³³ KI wird im allgemeinen Sprachgebrauch für sehr unterschiedliche automatisierte Entscheidungssysteme verwendet, zum Beispiel Spracherkennung, Schachspielen oder Übersetzungen in verschiedene Sprachen. Problematisch an dem Begriff ist, dass menschenähnliche Intelligenz suggeriert wird und damit auch Verantwortung für Entscheidungen. KI-Systeme tragen jedoch keine Verantwortung, das können nur Menschen.

Lernende Algorithmen – lösen eine Aufgabe mithilfe von zwei verschiedenen Phasen: In der ersten Phase versucht der Algorithmus, Regeln aus vorhandenen Daten zu „lernen“, die auf das gewünschte Ergebnis hindeuten. Dies geschieht oft „unter Aufsicht“, das heißt dem Algorithmus wird kenntlich gemacht, welche Daten einer Datensammlung der Vergangenheit gewünschte Ergebnisse waren und welche nicht. Die Regeln, die ein lernender Algorithmus daraus ableitet, bleiben jedoch meist im Verborgenen und sind oft nicht nachvollziehbar. In der zweiten Phase wendet der Algorithmus diese erlernten Regeln an, um Entscheidungen über unbekannte Daten derselben Art zu treffen.³⁴ Lernende Algorithmen werden für maschinelles Lernen verwendet.

Maschinelles Lernen (ML) – ein Forschungsfeld der Informatik, in dem bestimmte Methoden der Künstlichen Intelligenz erforscht werden. Ohne den Lösungsweg direkt zu programmieren, soll dieser automatisiert in den Daten gefunden werden. ML basiert auf lernenden Algorithmen, die Muster in den vorhandenen Daten analysieren und für Modelle verwenden. Diese ML-Modelle werden dann auf neue Daten angewendet.³⁵

(Mathematisches) Modell – In einem Modell wird ein Ausschnitt der Wirklichkeit beschrieben. Es ist eine Abstraktion und kann nie alle Aspekte der Wirklichkeit darstellen. Das Modell soll so genau sein, dass es Prognosen oder Analysen über eine bestimmte Thematik oder Situation ermöglicht.

Online-Plattformen – sind Webseiten, über die sich verschiedene Akteur*innen austauschen können. Online-Plattformen dienen unterschiedlichen Zwecken: Soziale Online-Netzwerke bieten Kommunikationsmöglichkeiten, Jobplattformen sollen bei der Suche nach passenden Stellenausschreibungen unterstützen und Handelsplattformen Online-Kauf und -Verkauf erleichtern. Algorithmen können hier beispielsweise die Reihenfolge bei Suchanfragen beeinflussen und damit, wer welche Inhalte wie prominent zu sehen bekommt.

Proxy-Variable – wird stellvertretend für andere Variablen eingesetzt, zum Beispiel wenn diese nicht direkt beobacht- oder messbar sind. Das englische Wort „proxy“ bedeutet „Stellvertreter*in“. Mit Proxy-Variablen kann auf bestimmte – auch geschützte – Merkmale geschlossen werden, ohne dass sie direkt abgefragt oder erfasst werden. Aus der Angabe „30 Jahre Arbeitserfahrung“ kann man beispielsweise ablesen, dass die Person mindestens Mitte 40 sein muss. So können Proxy-Variablen einen entscheidenden Einfluss auf die Entscheidungen eines automatisierten Systems haben.

33 Döbel, Inga et al. (2018). Maschinelles Lernen – eine Analyse zu Kompetenzen, Forschung und Anwendung. *Fraunhofer*: S. 44. <https://www.bigdata-ai.fraunhofer.de/de/publikationen/ml-studie.html> (09.03.2022).

34 AlgorithmWatch (2016). 2. Arbeitspapier: Überprüfbarkeit von Algorithmen. <https://algorithmwatch.org/de/zweites-arbeitspapier-ueberpruefbarkeit-algorithmen/> (14.04.2022).

35 Vgl. Plattform Lernende Systeme (2022). Glossar: Maschinelles Lernen. <https://www.plattform-lernende-systeme.de/glossar.html> (14.04.2022).

Regelbasierter Algorithmus – Die Regeln und Handlungsanleitungen, die der Algorithmus ausführt, sind explizit programmiert worden. Das ist bei Online-Formularen oft der Fall: Kann man ein Formular nur dann abschicken, wenn man sich einer der binären Geschlechterkategorien „männlich“ oder „weiblich“ zugeordnet hat, dann hat ein Mensch diese Bedingung so programmiert. Regelbasierte unterscheiden sich von lernenden Algorithmen dadurch, dass ihre Regeln/Kriterien/Bedingungen explizit programmiert wurden.

Scoring – Auskunftsteien berechnen Score-Werte basierend auf Daten. „Diese drücken aus, mit welcher Wahrscheinlichkeit ein Mensch nach Einschätzung der Auskunftstei seine Rechnungen bezahlen wird. Dabei bewerten sie die betreffende Person nicht individuell, sondern ordnen sie mit statistischen Verfahren einer Vergleichsgruppe mit ähnlichen Merkmalen zu. Die Zahlungsmoral, die diese Vergleichsgruppe in der Vergangenheit gezeigt hat, wird dann auch für die jeweilige Person angenommen.“³⁶

Sozio-technische Systeme – Automatisierte Entscheidungssysteme können besser verstanden werden, wenn sie als sozio-technische Systeme betrachtet werden, das heißt neben den technischen auch ihre sozialen Komponenten mitbetrachtet werden. Solche Systeme entstehen aus einem spezifischen historischen und gesellschaftlichen Kontext heraus, werden durch diesen beeinflusst und haben wiederum Einfluss auf die gesellschaftliche Entwicklung insgesamt und spezifische Gruppen und Individuen in dieser Gesellschaft.

Targeting – beschreibt die gezielte Verbreitung von Anzeigen. Das englische Wort „target“ bedeutet auf Deutsch „Ziel“. Dabei sollen die Menschen entsprechend ihrer Interessen angesprochen werden und die Wahrscheinlichkeit erhöht werden, dass sie die Anzeigen anklicken. Menschen, die zum Beispiel ein neues Sofa kaufen möchten, sollen vermehrt Werbeanzeigen von Sofas sehen. Dafür werden Daten verwendet, wie zum Beispiel vorheriges Suchverhalten.

8 Quellen zum Thema

„Ich würde mir wünschen, dass man einen Expertenpool, einen Wissenspool anzapfen kann.“ (Interviewpartner 10)

Hier listen wir zum einen Organisationen auf, die regelmäßig Neues über ADM-Systeme und Diskriminierungsrisiken veröffentlichen. Zum anderen empfehlen wir relevante Quellen zum Thema.

Relevante Organisationen, die regelmäßig zum Thema veröffentlichen

AlgorithmWatch

Eine gemeinnützige Forschungs- und Advocacy-Organisation mit dem Ziel, ADM-Systeme und ihre Auswirkungen auf die Gesellschaft zu untersuchen und zu erklären. Webseite: <https://algorithmwatch.org/>.

Antidiskriminierungsstelle des Bundes (ADS)

Neben der Antidiskriminierungsberatung Betroffener, gibt die ADS Studien und (Forschungs-) Projekte in Auftrag, auch über ADM-Systeme. Die Ergebnisse veröffentlicht sie hier: <https://www.antidiskriminierungsstelle.de/DE/was-wir-machen/publikationen/publikationen-node.html>.

AI + Automation Lab des Bayerischen Rundfunks

Ein interdisziplinäres Team veröffentlicht regelmäßig Artikel und Studien „an der Schnittstelle von Journalismus, Informatik und Produktentwicklung“. Webseite: <https://www.br.de/extra/ai-automation-lab/index.html>.

Bertelsmann Stiftung – Ethik der Algorithmen

In ihrem Projekt *Ethik der Algorithmen* beschäftigt sich die Bertelsmann Stiftung mit den gesellschaftlichen Auswirkungen von ADM. <https://www.bertelsmannstiftung.de/de/unsere-projekte/ethik-der-algorithmen>.

³⁶ Lietzau, Josefine (2019). Wie Schufa und Co. mit unseren Daten handeln. Finanztip. <https://www.finanztip.de/schufa/auskunftsteien/> (14.04.2022).

European Digital Rights (EDRi)

Ein Netzwerk von zivilgesellschaftlichen und Menschenrechtsorganisationen aus ganz Europa, das rund um das Thema „Digitale Rechte“ Projekte und Veranstaltungen durchführt sowie Publikationen veröffentlicht. <https://edri.org/>.

Gesellschaft für Freiheitsrechte e. V. (GFF)

Der Verein setzt sich auf deutscher und europäischer Ebene für eine Stärkung der Grund- und Menschenrechte ein, zum Beispiel über strategische Klageverfahren. Webseite: <https://freiheitsrechte.org/>.

Landesantidiskriminierungsstelle Berlin (LADS)

Zu den Aufgaben der LADS gehört es, Öffentlichkeitsarbeit zum Antidiskriminierungsrecht zu machen und Menschen zu ermutigen, ihre Rechte wahrzunehmen. Dazu gehört auch das Thema Diskriminierungsrisiken durch ADM-Systeme. Ihre Publikationen veröffentlicht die LADS hier: <https://www.berlin.de/sen/lads/ueber-uns/materialien/>.

netzforma e. V.

Der Verein setzt sich für feministische Netzpolitik ein und beschäftigt sich dabei unter anderem auch mit KI-Anwendungen und Algorithmen. Neben Publikationen bieten sie auch Workshops an. Webseite: <https://netzforma.org/aktuelles>.

Netzpolitik.org

Auf der Webseite des Online-Mediums *netzpolitik.org* erscheinen regelmäßig Artikel und Kommentare rund um das Thema digitale Freiheitsrechte – auch (Anti-)Diskriminierung wird in diesem Kontext immer wieder behandelt. Webseite: <https://netzpolitik.org/>.

Superrr Lab

Die Gruppe erforscht, wie Technologien für eine gerechtere digitale Zukunft verwendet werden können, und baut Netzwerke dafür auf. Webseite: <https://superrr.net/news/>.

Verbraucherzentrale Bundesverband (vzbv) und Verbraucherzentrale Hessen

Mit einem Fokus auf Verbraucherschutz werden jeweils leicht verständliche Informationen zu Algorithmen und KI-Anwendungen veröffentlicht.

Webseiten: <https://www.vzbv.de/algorithmen-ki> und <https://www.algo-was.de/>.

Bücher und Reports

Allen, Robin & Masters, Dee (2020). Regulating for an Equal AI: A new Role for Equality Bodies. *An Equinet Publication*. <http://ai.equineteurope.org/library/regulating-equal-ai-new-role-equality-bodies> (25.02.2022).

Benjamin, Ruha (2019). Race after Technology: Abolitionist Tools for the New Jim Code. *Polity*.

Bundesministerium für Familie, Senioren, Frauen und Jugend (2021). Dritter Gleichstellungsbericht – Digitalisierung geschlechtergerecht gestalten. <https://www.bmfsfj.de/resource/blob/184544/c0d592d2c37e7e2b-5b4612379453e9f4/dritter-gleichstellungsbericht-bundestagsdrucksache-data.pdf> (03.03.2022).

Criado Perez, Caroline (2019). Invisible Women: Exposing Data Bias in a World Designed for Men. *Random House*.

D'Ignazio, Catherine & Klein, Lauren F. (2020). Data Feminism. *MIT Press*.

Eubanks, Virginia (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. *St. Martin's Press*.

Lenzen, Manuela (2020). Künstliche Intelligenz: Fakten, Chancen, Risiken (Vol. 2904). *CH Beck*.

netzforma* e. V. (2021). Wenn KI dann feministisch. Impulse aus Wissenschaft und Aktivismus. https://netzforma.org/wp-content/uploads/2021/01/2020_wenn-ki-dann-feministisch_netzforma.pdf (25.02.2022).

O'Neil, Cathy (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. *Broadway Books*.

Orwat, Carsten (2019). Diskriminierungsrisiken durch Verwendung von Algorithmen: eine Studie, erstellt mit einer Zuwendung der Antidiskriminierungsstelle des Bundes. *Nomos*.

Xenidis, Raphaele & Gerards, Janneke (2021). Algorithmic discrimination in Europe: Challenges and opportunities for gender equality and non-discrimination law. *European Commission, Directorate-General for Justice and Consumers*. <https://op.europa.eu/en/publication-detail/-/publication/082f1dbc-821d-11eb-9ac9-01aa75ed71a1> (25.02.2022).

Zweig, Katharina (2019). Ein Algorithmus hat kein Taktgefühl: Wo künstliche Intelligenz sich irrt, warum uns das betrifft und was wir dagegen tun können. *Heyne Verlag*.

Weitere Quellen

„HR-Puzzle“ ist ein Erklärspiel über Maschinelles Lernen (ML) im Personalmangement, das verständlich macht, wie ML funktioniert: <https://algorithmwatch.org/de/hrpuzzle/> (25.02.2022).

Die Erklärvideos von AlgorithmWatch bieten einen leichten Einstieg in wichtige Grundlagen: Proxy-Variablen, Korrelation & Kausalität, Nachvollziehbarkeit algorithmischer Entscheidungssysteme: <https://algorithmwatch.org/de/auto-hr/erklaraervideos/> (25.02.2022).

Ein englischsprachiger Blog für Jurist*innen zum Thema ADM-Systeme und juristische Strategien ist „Atlas Lab“: <https://www.atlaslab.org/> (25.02.2022).

Ein Video von Nushin Yazdani und José Rojas erklärt, „Wie Maschinen diskriminieren – Facial Recognition & the Matrix of Domination“: <https://vimeo.com/476084538> (03.03.2022).

Der Comic-Essay „We need to talk, AI“ von Julia Schneider und Lena Ziyal veranschaulicht Grundlagen, Möglichkeiten und Grenzen von KI-Anwendungen: <https://weneedtotalk.ai/> (09.03.2022).

Ein Simulator, der Grundprinzipien von Empfehlungsalgorithmen erfahrbar macht und im Rahmen des Projekts *DataSkop* entstanden ist: <https://algorithmwatch.org/de/dataskop-simulator-plattform-dynamiken/> (09.03.2022).

Der Botendienst *unding.de* schickt Anliegen zu großen Internetfirmen und anderen Stellen, wenn Menschen sich durch ADM-Systeme ungerecht behandelt fühlen oder über diese ärgern. Hier finden Sie Beispiele und Umfragen zum Thema: <https://unding.de/news/>.

Die Verbraucherzentrale des Saarlandes e. V. bietet eine Online-Simulation zu datenbasierter Preisanpassung beim Online-Handel: <https://wasistdeinpreis.de/index.html> (03.03.2022).

„Coded Bias“ von Shalini Kantayya ist ein Dokumentarfilm über die Recherchen der MIT Lab-Forscherin Joy Buolamwini, die zu Diskriminierung in Algorithmen arbeitet und unter anderem in einer wissenschaftlichen Studie aufgedeckt hat, dass gängige Gesichtserkennungssoftware bei Gesichtern mit dunkler Hautfarbe schlechter funktioniert.

AI Blindspot, ein Projekt des Assembly Programms des Berkman Klein Center For Internet & Society an der Harvard Universität, bietet englischsprachige Workshop-Unterlagen zum Thema „Tools for Advancing Equity in Artificial Intelligence“: <http://aibindspot.media.mit.edu/AL.html> (25.02.2022).

Die Comic-Bücher über Verantwortlichkeit und Fairness bei Daten und KI-Anwendungen von Julia Stoyanovich und Falaah Arif Khan erklären Grundlagen anschaulich und verständlich in englischer Sprache mit einem USA-Bezug: <https://dataresponsibly.github.io/comics/> (25.02.2022).

Ein Glossar mit Erklärungen von Begriffen rund um KI-Anwendungen bietet die Plattform Lernende Systeme: <https://www.plattform-lernende-systeme.de/glossar.html>.

