HOCHSCHULE DER MEDIEN

Audiovisuelle Medien/Aktuelle Themen

Wintersemester 2022/23

Romy Walcher, Matr.-Nr. 42778

rw052@hdm-stuttgart.de

Diskriminierende Künstliche Intelligenz

Abstract

Der Einzug von künstlicher Intelligenz in unser tägliches Leben bringt neben zahlreichen Vorteilen wie Effizienz und Fehlerreduktion auch großes Diskriminierungspotential. Vorurteile, Sexismus und Rassismus machen vor der Technik nicht Halt und finden sich in diversen Anwendungen wie Gesichtserkennung, Chatbots, autonomes Fahren, Medizintechnik und Spracherkennungssoftware wieder. Um diskriminierende KI in Zukunft zu vermeiden, werden Beispiele und Hintergründe für dieses Phänomen untersucht und konstruktive Lösungsvorschläge eruiert.

Einführung

Die Anwendungsbereiche von künstlicher Intelligenz sind vielfältig und scheinbar grenzenlos. Medizin, Datenmanagement, autonomes Fahren, Gesichtserkennung, Suchmaschinen, Cyber-Security, Chatbots und virtuelle Assistenten sind nur einige Beispiele für Bereiche, in denen KI uns Menschen bereits enorm nutzt und Einzug in unser tägliches Leben erhalten hat.
Künstliche Intelligenz wird durch Machine Learning trainiert, große, unstrukturierte Datenmengen zu filtern, darin Muster zu erkennen und auf deren Basis zu erlernen, Entscheidungen zu treffen [1]. Damit werden Auswertungen von Daten innerhalb kürzester Zeit ermöglicht, was die Automatisierung, Effizienz und Fehlerreduktion in diversen Technologien verbessern kann.
Trotz der zahlreichen Chancen von künstlicher Intelligenz darf man ihre Risiken nicht außer Acht lassen. Ein Risiko, das für die meisten Menschen auf den ersten Blick nicht offensichtlich erscheint, ist das Diskriminierungspotenzial von künstlicher Intelligenz [2]. Gerade weil KI-Algorithmen immer nahtloser in Techniken des täglichen Lebens integriert werden, sind die ernsthaften Konsequenzen, die durch KI-basierte Entscheidungen entstehen, oft nicht auf den ersten Blick ersichtlich und keineswegs zu unterschätzen [3]. Da künstliche Intelligenz mithilfe von Code und Algorithmen erschaffen wird, gehen die meisten Menschen davon aus, dass der Output der KI stets faktenbasiert und objektiv ist. Doch das ist ein Trugschluss und wenn man sich intensiver mit der Thematik befasst, wird klar, dass KI häufig unerwünschte und bisweilen sogar problematische, diskriminierende Ergebnisse liefert. Vorurteile werden dadurch nicht nur reproduziert, sondern sogar verschärft [1]. Warum das so ist, wo diskriminierende KI auftritt und wie man sie verhindern kann, wird im Folgenden Schritt für Schritt dargelegt.

Diskriminierung

Unter dem Begriff Diskriminierung versteht man die Benachteiligung von Menschen aufgrund eines unrechtmäßigen Merkmals, wie z.B. ihrer Hautfarbe, ihrer ethnischen Gruppenzugehörigkeit, ihres Geschlechts, ihrer sexuellen Orientierung und/oder Identität, ihrer Religion, ihres Alters, der Staatsangehörigkeit, der Sprache, des Gesundheitszustands oder einer Behinderung [4]. Durch die ungerechtfertigte Ungleichbehandlung wird diskriminierten Menschen der Zugang zu bestimmten materiellen und immateriellen Gütern aufgrund der jeweiligen (vermeintlichen) Gruppenzugehörigkeit verwehrt oder beschränkt. Der Diskriminierungsanlass bzw. das Unterscheidungsmerkmal ist dabei die Abweichung vom angenommenen "Normalfall", den häufig der erwachsene, weiße, männliche Staatsbürger darstellt [5].
Grundlage für Diskriminierungen sind oftmals Stereotype und Vorurteile gegenüber einzelnen sozialen Gruppen und deren Angehörigen, die sich im Handeln und Denken der Menschen widerspiegeln und somit auch implizit in das gesellschaftliche Leben einfließen.
Im Kontext der Informationstechnologie spricht man dann von Diskriminierung, wenn die Differenzierungen, die aus der Datenverarbeitung eines Systems entstehen, als ungerecht angesehen werden und Entscheidungen an Persönlichkeitsmerkmalen orientiert werden, die in keinem relevanten Zusammenhang mit der Entscheidung stehen [6]. Da künstliche Intelligenz autonom Entscheidungen trifft, treten auch hier in verschiedenen Anwendungsbereichen Diskriminierungen auf. Beispiele sowie Gründe und Erklärungen dafür werden in dieser Arbeit untersucht.

Beispiele für diskriminierende KI

Um einen Überblick über verschiedene Fälle und die Ausmaße diskriminierender KI zu gewinnen, werden in diesem Kapitel einige Beispiele beschrieben.

Gesichtserkennung

Die auf KI-basierende Video-Empfehlungsfunktion von Facebook machte im Jahr 2020 einen schwerwiegenden rassistischen Fehler. Nutzern, die sich ein Video zum Thema Rassismus ansahen, das unter anderem Bilder von Schwarzen[1] Menschen zeigte, wurde vorgeschlagen, weitere "Videos über Primaten" anzusehen, was eindeutig rassistisch konnotiert ist. Menschenaffen oder andere Tiere kamen in dem Video gar nicht vor. Daraus lässt sich schließen, dass die KI nicht oder nur geringfügig mit Gesichtern von nicht-weißen Menschen trainiert wurde und somit "erlernt" hat, Schwarze Menschen und Primaten in dasselbe Muster einzuordnen. [7]

Automatischer Bildzuschnitt

Im Jahr 2021 erntete Twitter Kritik, nachdem der automatische Bildzuschnitt beim Upload in zahlreichen Testbildern von Nutzer:innen bevorzugt Gesichter von weißen Menschen ausgeschnitten und somit rassistische Vorurteile verstärkt hat. Der zugrundeliegende Machine-Learning-Algorithmus wurde daraufhin von Twitter untersucht, woraufhin sich die Vermutung bestätigte, dass häufig Schwarze Menschen von der KI abgeschnitten wurden. Der Algorithmus basierte auf einer Eye-Tracking-Studie

[1] "Schwarz" beschreibt in diesem Zusammenhang eine Gruppe von Menschen, die aufgrund ihrer Hautfarbe und Herkunft Rassismus erfahren müssen. Um das zu verdeutlichen, wird das S in Schwarz großgeschrieben [23]

mit nur sehr wenigen Teilnehmern und Testbildern, woraufhin das Machine-Learning-Modell offenbar die Vorurteile dieser Menschen reproduzierte. Als Maßnahme gab Twitter an, von da an soweit möglich auf den automatischen Bildzuschnitt zu verzichten. [8]

Autonomes Fahren

Auch im stark erforschten Bereich des autonomen Fahrens ließ sich 2019 mithilfe einer Studie des Georgia Institute of Technology [9] feststellen, dass die dabei verwendeten KIs ebenfalls rassistische Entscheidungen treffen. Um herauszufinden, wie die modernen Objekterkennungsmodelle Menschen aus verschiedenen demographischen Gruppen erkennen, untersuchten die Forscher:innen einen großen Datensatz von Bildern, die Fußgänger zeigten und klassifizierten diese nach Helligkeit ihrer Hautfarbe. Dann wurde analysiert, wie oft die Modelle die Anwesenheit von Personen der hellhäutigen Gruppe richtig erkannten und wie oft sie es bei Personen mit dunkler Haut richtig detektierten. Das ernüchternde Ergebnis war, dass Menschen mit dunkler Haut im Schnitt um fünf Prozent schlechter erkannt wurden, unabhängig von Variablen wie Tageszeit oder verdeckter Sicht. Im Ernstfall könnte also eine Schwarze Person mit fünf Prozent höherer Wahrscheinlichkeit von einem autonom fahrenden Auto an- oder überfahren werden. Auch wenn diese Erkenntnisse bislang lediglich auf einer Laborstudie basierend auf Forschungsmodellen und öffentlichen Datensätzen beruhen, muss man die daraus resultierende Gefahr ernstnehmen und die Forschung dahingehend auf Daten aus der direkten Anwendung ausweiten. Dies ist bislang allerdings schwierig, da die Automobilunternehmen ihre Daten für diesen Zweck nicht veröffentlichen wollen. [9]

Diagnostik und medizinische Behandlung

Immer häufiger treffen künstliche Intelligenz und maschinelles Lernen Entscheidungen über unsere zukünftige medizinische Behandlung und Pflege, die oft zum Nachteil Schwarzer Bevölkerungsgruppen ausfallen. So kann zum Beispiel mit Hilfe eines Machine-Learning-Algorithmus Hautkrebs mit einer sehr hohen Zuverlässigkeit diagnostiziert werden. Da hellhäutige Menschen allerdings das höchste Risiko haben, an Hautkrebs zu erkranken, ist der Algorithmus insbesondere auf helle Hauttypen trainiert und liefert folglich bei heller Haut die zuverlässigsten Ergebnisse. Dieser Bias hat zur Folge, dass in den USA die Hautkrebs-Sterberate von Schwarzen Menschen bei 27% liegt, während sie bei weißen Menschen, die deutlich öfter an Hautkrebs erkranken, lediglich bei 10% liegt[2], aus dem einfachen Grund, dass der Krebs bei PoC[3] seltener diagnostiziert wird [10]. Die Ursache dafür ist, dass KI- Algorithmen bei der Detektion von Krankheiten und der Auswahl medizinischer Behandlungsmethoden häufig auf voreingenommenen Regeln und homogenen Datensätzen beruhen, die nicht die Patient:innenpopulation insgesamt widerspiegeln. Schwarze Patient:innen müssen deshalb befürchten, dass ein Algorithmus ihnen beispielsweise eine Organtransplantation aufgrund ihrer Hautfarbe verwehrt. Obwohl die Wahrscheinlichkeit eines Nierenversagens bei Schwarzen Amerikanern viermal so hoch ist, setzte ein Algorithmus zur Bestimmung der Platzierung auf der Transplantationsliste Schwarze Patient:innen auf einen

[2] Stand 2019

[3] People of Color

niedrigeren Rang als weiße, selbst wenn alle anderen Faktoren gleich bleiben. [11]

Chatbots

Chatbots, die auf KI basieren, sorgten in der Vergangenheit bereits mehrmals für Schlagzeilen. Die Firma Microsoft wollte 2016 mithilfe ihres Chatbots Tay herausfinden, wie junge Menschen reden und dieser lernte durch Nutzereingaben nach der Veröffentlichung binnen Stunden, rassistische, sexistische und antisemitische Aussagen zu tätigen, den Holocaust zu leugnen und Völkermord zu befürworten [12].
Auch Blenderbot3, der im Jahr 2022 an den Start gebrachte Chatbot von Meta, schrieb innerhalb kürzester Zeit diskriminierende, antisemitische und verschwörerische Texte. Obwohl Meta eine solche Entwicklung bei der Implementierung explizit verhindern wollte und eine Hinweisbox eingefügt hatte, bei der Nutzer:innen bestätigen mussten, die KI nicht zu diskriminierenden Äußerungen zu provozieren, konnte ein solches Verhalten nicht verhindert werden. Bislang ist Blenderbot3 nur in den USA verfügbar [13].
Für einige Schlagzeilen sorgte jüngst auch der Chatbot ChatGPT, der Dialoge führen, komplexe Rechnungen lösen und Texte verfassen kann. Auch wenn die Entwickler:innen von OpenAI ChatGPT explizit mit menschlichem Feedback trainiert haben, um diskriminierende Inhalte zu verhindern und richtig einzuordnen, kann man dem Chatbot mit den “richtigen” Anfragen dennoch ebenfalls problematische Aussagen entlocken [14].

Job-Vorauswahl

Im Jahr 2015 nutzte Amazon eine Machine-Learning-Engine, die bei der Job-Vorauswahl Bewerber:innen vorsortieren sollte. Die künstliche Intelligenz war in der Lage, auf Basis von Qualifikations- und Lebenslaufdaten bisheriger Angestellter, die Bewerber:innen einzuordnen und ihnen eine Bewertung von 1 bis 5 Sternen zu geben, wonach sie gerankt und dem Recruitment-Team von Amazon vorgelegt wurden. Allerdings realisierte die Firma nach einigen Monaten der Nutzung, dass insbesondere bei technischen Stellen die Job-Kandidat:innen basierend auf ihrem Geschlecht unterschiedlich hoch eingestuft wurden. Der Grund dafür ist, dass Amazons Computermodelle darauf trainiert wurden, Muster in Lebensläufen von bereits eingestelltem Personal zu erkennen, die dem Unternehmen über einen Zeitraum von 10 Jahren vorgelegt wurden. Der Großteil dieser Lebensläufe stammte von Männern, was die männliche Dominanz in der Technologiebranche widerspiegelt. Das System brachte sich dadurch selbst bei, männliche Bewerber bei der Vorauswahl zu bevorzugen. Es wurden z.B. Lebensläufe niedrig gerankt, die explizit das Wort “woman/women” enthielten und es stufte Absolventinnen von reinen Frauen-Colleges zurück. Amazon passte die KI daraufhin zwar an, stellte das System aber dennoch ein Jahr später komplett ein, da nicht mit Sicherheit verhindert werden konnte, dass sich ein ähnlicher Fehler erneut einschleicht. [15]

Lensa-AI

Lensa AI ist eine KI, die digitale Portraits auf Basis von Selfies erstellen kann. Bei Betrachtung von Ausgabebildern der App wird schnell klar, dass die KI Frauen in den Bildern systematisch sexualisiert, dem toxischen Schönheitsideal entsprechend “optimiert” und freizügiger darstellt als in den eingesandten Bildern. Es ist ein problematisches

Muster erkennbar, dass Frauen häufig wie Feen, Heilige oder Prinzessinnen, Männer hingegen - vollständig bekleidet - als Astronauten, Helden oder Cyborgs dargestellt werden. Dadurch werden sexistische Stereotype reproduziert [16].

Da Lensa AI auf öffentlich verfügbaren Bildern unterschiedlicher Stile trainiert wurde, lernte die KI, das Aussehen bestimmter Dinge und Personen nachzubilden. Dass die Daten häufig einseitig und in sich schon diskriminierend sind, kann die KI nicht erkennen und übernimmt das erlernte Muster. So werden sexistische und rassistische Stereotype von dem System übernommen. Asiatische Frauen werden beispielsweise im Vergleich noch stärker sexualisiert, sie werden von der App häufig nackt und in pornografischen Posen dargestellt. Bei einer Test-Filterung des Lensa–AI-Datensatzes nach dem Schlüsselwort “Asian” spuckte die KI fast ausschließlich pornografische Inhalte aus, was die generierten Outputs der App erklärt und nicht weniger problematisch macht [17].

Des Weiteren diskriminierend daran ist, dass die Frauen, die die Fotos von sich in die App laden, nicht selbst bestimmen können, ob sie derart sexualisiert dargestellt werden wollen oder eben nicht; Lensa AI trifft diese Entscheidung für sie.

Außerdem wird KI häufig von Männern implementiert. Da Männer nicht von Misogynie[4] und weiblichen Stereotypen betroffen sind, nehmen sie diese in ihrer Umgebung weniger wahr, weshalb sexistische Ausgaben häufig nicht bereits beim Training der KI verhindert werden können.

Umso wichtiger ist die Diversifizierung von Teams in Tech-Firmen, damit solche Vorurteile rechtzeitig erkannt und korrigiert werden können [16].

Sexualisierte Ausgabebilder von Lensa AI der Redakteurin Melissa Heikkilä [17]

Sprachassistenten

Im Jahr 2017 wurde getestet, wie die digitalen Sprachassistenten Siri, Cortana, Alexa und Google Assistant auf verbale sexuelle Belästigungen reagieren. Bei allen vier KIs ist die Stimme standardmäßig weiblich, bei drei von vier der Name ebenso. Die Reaktionen auf sexuell belästigende Aussagen wie “Du bist ein ungezogenes Mädchen”, “Du bist heiß” oder die explizite Äußerung von sexuellen Gefälligkeiten fielen bei dem Test fast immer positiv aus, häufig in Form von spielerisch-schüchternem Ausweichen, Flirten oder Ablenkungen. Mittlerweile wurden einige der Sprachassistenten dahingehend bearbeitet, dass sie bei einer belästigenden Aussage zumindest die Antwort verweigern - eine Verurteilung und richtige Einordnung des problematischen Verhaltens erfolgt jedoch nach wie vor in den Antworten der KIs nicht. Der diskriminierende Faktor hierbei ist, dass die Art und Weise der virtuellen Assistenten das Bild einer unterwürfigen, devoten Frau vermittelt, die zu gehorchen hat. Geschlechtervorurteile und patriarchale

[4]Frauenhass

Wertvorstellungen werden durch diese Art der Reaktion verstärkt. [18]

Gründe für diskriminierende KI

Das war lediglich eine kleine Auswahl an Beispielen für diskriminierende KI, um ein Verständnis für das Gesamtproblem zu schaffen. Es stellt sich die Frage: warum wird künstliche Intelligenz diskriminierend? Allgemein gesprochen ist die Hauptursache die Reproduktion und Verschärfung menschlicher Vorurteile [2].
Bei näherer Betrachtung des Phänomens tauchen vor allem die vier folgenden Begründungen am häufigsten auf.

Algorithmischer Bias

Ein Bias ist ein Verzerrungseffekt, der in der Statistik als Fehler im Rahmen der Datenverarbeitung und -erhebung verstanden wird. Dabei unterscheidet man zwischen dem prä-existierenden Bias, bei dem in der Gesellschaft etablierte Vorurteile implizit oder explizit in die Software übertragen werden, dem technischen Bias, wenn technische Aspekte wie z.B. Sensorik dazu führen, dass bestimmte Gruppen anders behandelt werden als andere, und dem emergenten Bias, bei dem die Diskriminierung im Zusammenspiel von Software und Anwendung entsteht, wenn beispielsweise eine Software die erzeugten Ausgaben falsch interpretiert und einordnet. Der algorithmische Bias ist deshalb in vielen Fällen eine Ursache, dass es zu unerwünschten und sogar diskriminierenden Ergebnissen kommt [2].

Trainingsbeispiele

Einen bedeutenden Einfluss auf KI-Ergebnisse haben außerdem die Trainingsbeispiele. Wenn ein Trainingsdatensatz nicht diskriminierungsfrei ist, kann die darauf trainierte KI auch nicht diskriminierungsfrei sein, weil sie die Diskriminierung “mitlernt”. Da viele Systeme mithilfe von Daten aus dem Internet, wie Dokumente, Videos, Bilder und Tonaufnahmen, lernen, greifen sie vorhandene Diskriminierungen auf und verschärfen sie unter Umständen sogar [2].

Privilegierte Software-Entwickler

Der Anteil von Frauen in der Tech-Industrie mit Bezug zu künstlicher Intelligenz ist nur ein Bruchteil so groß wie der von Männern. Nach wie vor besteht der Großteil der Software-Entwickler:innen im Silicon Valley aus jungen weißen Männern aus der Mittel- und Oberschicht, die (meist unbeabsichtigt) ihre Weltanschauungen und Werte in die Software hineinimplementieren. Da sie der gesellschaftlichen Norm entsprechen und sich mit Diskriminierung nicht auseinandersetzen müssen, wird das Thema in den meisten Fällen nicht mitgedacht und so kann sich Alltagsrassismus und -sexismus ungebremst seinen Weg in die Software bahnen [19].

KI und Moral

Machine-Learning-Software wird von der Gesellschaft als rein technisch und damit neutral, faktenbasiert und wertfrei betrachtet. Da künstliche Intelligenz jedoch eigenständig Entscheidungen trifft, die für das Leben von realen Menschen relevant sind, verschmilzt die Technik unweigerlich mit sozialen und gesellschaftlichen Gefügen. Unmittelbare und mittelbare Entscheidungen über das Leben von Menschen müssen von Fall zu Fall betrachtet werden und haben immer einen moralischen und ethischen Hintergrund. Moral ist allerdings nie objektiv, sie unterscheidet sich von Mensch zu

Mensch, von Situation zu Situation und von Kultur zu Kultur [20]. Eine moralische Bewertung kann deshalb nicht durch die Systeme selbst vorgenommen werden, denn sie basieren auf Mustererkennung in Daten aus der Realität und können in ihren Entscheidungen lediglich die "Normativität des Faktischen"[5] reproduzieren [2].

Lösungsansätze

Um das Diskriminierungsrisiko von KI zu verringern, gibt es verschiedene Lösungsansätze, die das Problem an der Ursache angehen und ganzheitlich in die KI-Entwicklung integriert werden müssen.

Ein grundlegender Punkt ist, dass Diskriminierungsprobleme wie Rassismus und Sexismus vom privilegierten Teil der Gesellschaft überhaupt wahrgenommen und anerkannt werden müssen. Insbesondere Software-Entwickler:innen sowie Personen, die selbst nicht zu einer diskriminierten Gruppe gehören, dürfen Diskriminierung nicht ignorieren, sondern sollten ein Verständnis für Betroffene und die Auswirkungen davon entwickeln. Dafür ist die Sichtbarkeit und das (Aus-)Leben von Diversität in der Gesellschaft essenziell, insbesondere ist Vielfalt innerhalb von Tech-Firmen unabdingbar. Frauen, nonbinäre und nicht-weiße Menschen müssen im IT-Bereich stärker vertreten werden, um durch ein vielfältiges Team die männliche, weiße, wohlhabende Brille abzulegen [21]. Diversere Teams sind aus vielen Gründen bereichernd für ein Unternehmen. Unterschiedliche Denkweisen, Erfahrungen und Hintergründe fördern die Motivation und Leistungsfähigkeit von Mitarbeitenden, außerdem steigert Diversität die Innovationskraft und Wettbewerbsfähigkeit eines Unternehmens [22] sowie zweifelsohne die Sensibilisierung für Diskriminierung.

Des Weiteren sollte künftig die Lehre im Informatikstudium dahingehend erweitert werden, eine Grund-Sensibilisierung für Gender- und Rassismus-Themen zu schaffen und Studierende bereits im Studium und damit noch vor Beginn ihres Berufseinstiegs mit diesen gesellschaftlichen Themen vertraut machen und somit ein Bewusstsein dafür schaffen [21].

Darüber hinaus muss die Forschung in Bezug auf KI und Diskriminierung ausgeweitet werden. Dazu gehört die Erarbeitung diskriminierungsfreier Datensätze, technische Ansätze, die versuchen, ethische Prinzipien im Designprozess der Software zu integrieren, sowie die Integration von Diskriminierungsfreiheit als festen Bestandteil des Software-Testings [21].

Gesamtgesellschaftlich betrachtet müsste es in Zukunft eine unabhängige KI-Prüfinstanz geben, die die Ausgaben lernender Systeme analysiert und bewertet, sowie Trainingsdaten, verwendete Methoden und KI-Entscheidungen auf Plausibilität und Diskriminierungsfreiheit überprüft [2].

Fazit

Zweifelsfrei nimmt künstliche Intelligenz in unserem täglichen Leben eine immer bedeutendere Rolle ein. Diskriminierung, Vorurteile und Benachteiligung treten in der Gesellschaft nach wie vor auf und werden von lernenden Systemen übernommen, reproduziert und sogar verschärft. Umso wichtiger ist es, die durch KI entstehenden Diskriminierungsrisiken ernstzunehmen, sie zu verstehen und durch Bildung und Forschung an der Wurzel

[5] "Normativität des Faktischen" ist ein aus dem Staatsrecht stammender Begriff und bedeutet sinngemäß "Was alle oder jedenfalls die meisten tun, erscheint gut und richtig und wird deshalb befolgt" [24]

anzupacken und somit den Grundstein für eine diskriminierungsfreie Technologie zu legen, die nicht nur privilegierten, sondern allen gesellschaftlichen Gruppen zugutekommen kann.

Literaturverzeichnis

[1] S. Prof. Dr. Beck, *Wie diskriminierend ist künstliche Intelligenz?* [Online]. Available: https://www.wissenschaftsjahr.de/2019/neues-aus-der-wissenschaft/das-sagt-die-wissenschaft/wie-diskriminierend-ist-kuenstliche-intelligenz/ (accessed: Mar. 6 2023).

[2] S. e. a. Beck, “Künstliche Intelligenz und Diskriminierung: Herausforderungen und Lösungsansätze,” 2019.

[3] C. Orwat, “Diskriminierungsrisiken durch Verwendung von Algorithmen,” 2019.

[4] Amnesty International, *Definition: Was ist Diskriminierung.* [Online]. Available: https://www.amnesty.ch/de/themen/diskriminierung/zahlen-fakten-und-hintergruende/was-ist-diskriminierung#

[5] C. Orwat and A. Kolleck, “Mögliche Diskriminierung durch algorithmische Entscheidungssysteme und maschinelles Lernen,” 121-131, 2020.

[6] M. Rath, F. Krotz, and M. Karmasin, *Maschinenethik*. Wiesbaden: Springer Fachmedien Wiesbaden, 2019.

[7] S. Krempl, *Facebooks Gesichtserkennung hält schwarze Menschen für Affen.* [Online]. Available: https://www.heise.de/news/Facebooks-Gesichtserkennung-haelt-schwarze-Menschen-fuer-Affen-6182301.html (accessed: Mar. 6 2023).

[8] S. Grüner, *Twitter bestätigt Vorurteile in seiner Bildauswahl.* [Online]. Available: https://www.golem.de/news/machine-learning-twitter-bestaetigt-vorurteile-in-seiner-bildauswahl-2105-156646.html (accessed: Mar. 6 2023).

[9] B. Wilson, J. Hoffman, and J. Morgenstern, “Predictive Inequity in Object Detection,” Feb. 2019. [Online]. Available: http://arxiv.org/pdf/1902.11097v1

[10] A. Lashbrook, *AI-Driven Dermatology Could Leave Dark-Skinned Patients Behind.* [Online]. Available: https://www.theatlantic.com/health/archive/2018/08/machine-learning-dermatology-skin-color/567619/ (accessed: Mar. 6 2023).

[11] J. Resendez, J. Manley, and D. M. Christensen, *Medical Algorithms Are Failing Communities Of Color.* [Online]. Available: https://www.healthaffairs.org/do/10.1377/forefront.20210903.976632/full/ (accessed: Mar. 6 2023).

[12] P. Beuth, *Twitter-Nutzer machen Chatbot zur Rassistin.* [Online]. Available: https://www.zeit.de/digital/internet/2016-03/microsoft-tay-chatbot-twitter-rassistisch?utm_referrer=https%3A%2F%2Fwww.google.com%2F (accessed: Mar. 6 2023).

[13] T. Költzsch, *Metas Chatbot äußert antisemitische Verschwörungen.* [Online]. Available: https://www.golem.de/news/blenderbot-3-metas-chatbot-aeussert-antisemitische-verschwoerungen-2208-167479.html (accessed: Mar. 6 2023).

[14] J. Bager, *KI ChatGPT: Aufmerksam, aber nicht ohne Diskriminierung.* [Online]. Available: https://www.heise.de/hintergrund/Die-Text-KI-ChatGPT-schreibt-Fachtexte-Prosa-Gedichte-und-Programmcode-7392348.html?wt_mc=rss.red.ho.ho.rdf.beitrag.beitrag (accessed: Mar. 6 2022).

[15] J. Dastin, *Amazon scraps secret AI recruiting tool that showed bias against women.* [Online]. Available: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

[16] N. Hüsser, *Diese künstliche Intelligenz entkleidet Frauen: Digitale Portraits auf Social Media.* [Online]. Available: https://www.beobachter.ch/digital/lensa-ai-hinter-den-portrats-auf-social-media-steckt-eine-sexistische-kunstliche-intelligenz-555578 (accessed: Mar. 6 2023).

[17] M. Heikkilä, *The viral AI avatar app Lensa undressed me — without my consent.* [Online]. Available: https://www.technologyreview.com/2022/12/12/1064751/the-viral-ai-avatar-app-lensa-undressed-me-without-my-consent/ (accessed: Mar. 6 2023).

[18] S. Schuldt, *Siri, du Schlampe!: Sexistische Sprachassistenten.* [Online]. Available: https://katapult-magazin.de/de/artikel/siri-du-schlampe (accessed: Mar. 6 2023).

[19] A. Lobe, *Algorithmen haben ein Rassismusproblem.* [Online]. Available: https://www.apollon-dossier.de/rassismusproblem (accessed: Mar. 6 2023).

[20] R. D. Precht, *Künstliche Intelligenz und der Sinn des Lebens*: Goldmann, 2020.

[21] A. Geese, *Künstliche Intelligenz darf nicht sexistisch sein: Wie Algorithmen Menschen ignorieren und ausgrenzen.* [Online]. Available: https://demokratischer-salon.de/beitrag/kuenstliche-intelligenz-darf-nicht-sexistisch-sein/ (accessed: Mar. 6 2023).

[22] Absolventa, *Diversity am Arbeitsplatz: Darum sollten Arbeitgeber auf Vielfalt setzen.* [Online]. Available: https://www.absolventa.de/business/hr-blog/diversity-am-arbeitsplatz

[23] S. Mohamed, *Schwarz.* [Online]. Available: https://diversity-arts-culture.berlin/woerterbuch/schwarz

[24] K. F. Prof. em. Dr. Röhl, *Normalität und Normativität: Die »normative Kraft des Faktischen«.* [Online]. Available: https://www.rsozblog.de/normalitaet-und-normativitaet-die-normative-kraft-des-faktischen/#:~:text=Mit%20dem%20Faktischen%20meinte%20er,es%20%C3%A4ndern%20k%C3%B6nnten%2C%20erscheint%20erstrebenswert (accessed: Mar. 6 2023).