

17.07.2019 • Larissa Ginzinger, LBBW Research Associate
Dr. Guido Zimmermann, Senior Economist

LBBW
Bereit für Neues

Algorithmen: Rassistisch und sexistisch?

Ist KI frauenfeindlich?

Denken intelligente Maschinen wie Männer ?

- Technologien und Anwendungen Künstlicher Intelligenz (KI) sind in der heutigen technologisierten Welt nicht mehr wegzudenken: Algorithmen bestimmen, welche Informationen Suchmaschinen Internetnutzern präsentieren, selektieren Bewerbungen, empfehlen Gefängnisstrafen und beeinflussen Bewährungsentscheidungen. Doch praktisch nur die Hälfte der Gesellschaft ist an der Entwicklung dieser Anwendungen beteiligt. **War es noch die britische Mathematikerin Ada Lovelace, die als erste Person einen Algorithmus entwickelte, sind Frauen heute im Bereich der Programmierung weitgehend unterrepräsentiert.**
- Aufgrund des immer weiter zunehmenden Einflusses von Algorithmen ist es unverzichtbar sicherzustellen, dass sich Gesellschaftsgruppen, die bei der Entwicklung der zukunftsbestimmenden Anwendungen noch nicht (ausreichend) vertreten sind, keiner Diskriminierung ausgesetzt sehen. Denn um die Chancen der Welle der Automatisierung ergreifen zu können, muss die Industrie 4.0 in der Gesellschaft **als Rennen mit und nicht gegen „die Maschinen“** wahrgenommen werden. Ziel dieser Publikation ist es daher, die Leserinnen und Leser für Probleme zu sensibilisieren, die entstehen, wenn die Hälfte der Gesellschaft bei der Gestaltung und Entwicklung des Zukunftsbereichs KI nicht mitspricht.

Unsere Thesen

- Am Beispiel digitaler Assistenten wird deutlich, dass **Technologien, die von v. a. männlich dominierten Teams und Unternehmen entwickelt wurden, geschlechtsspezifische Vorurteile widerspiegeln und so zu einer bewusst oder unterbewusst verursachten Diskriminierung von Frauen führen können.** So reagierten bspw. omnipräsente digitale Helfer wie Siri, Alexa & Co. bis vor kurzem freundlich oder gar geschmeichelt auf frauenfeindliche Beschimpfungen.
- **Wir sagen nicht, man sollte Algorithmen nicht nutzen.** KI öffnet uns Türen, von denen wir vor einigen Jahren nicht wussten, dass sie überhaupt existieren. Algorithmen können uns viel Zeit sparen und unserer Produktivität zu Höchstleistungen verhelfen. Es **sollte uns bewusst sein, dass auch wir Menschen nicht frei von Vorurteilen sind.** Unter Umständen könnten Algorithmen, denen ein sog. Bias – eine tendenziöse Meinung – zugrunde liegt, sogar fairer urteilen als voreingenommene Menschen. **U. E. sollte man sie jedoch vorher verstehen.** Denn nur dann kann man eine explizite oder implizite Diskriminierung erkennen und diese ex-post beheben oder sogar ex-ante verhindern.
- **Bias oder Voreingenommenheit, ob bewusst oder unbewusst, impliziert eine mangelnde Inklusion und Vertretung bestimmter Gruppen.** U. E. stellt eine **ausgewogene(re) Repräsentation von Männern und Frauen im Technologiesektor**, auch wenn kein Allheilmittel, eine wichtige Grundlage für die Erstellung von Technologieprodukten, die die Vielfalt unserer Gesellschaft reflektieren, dar. Dies gilt insbesondere für KI, einen schnell wachsenden Zweig des Technologiesektors, der bereits heute einen enormen Einfluss auf das Leben der Menschen ausübt und in Zukunft sicher mehr und mehr ausüben wird.

Denken intelligente Maschinen wie Männer ?

Unsere Thesen

- Eine größere **Vielfalt bei Entwicklerinnen und Entwicklern** dürfte den Blick auf Vorurteile, die sich unbemerkt über Daten oder Wertevorstellungen in Algorithmen eingeschlichen haben, verschärfen und dabei helfen, diese ex-post zu eliminieren oder sogar ex-ante zu verhindern. Wie Frauen den Einstieg oder den Wechsel in den Bereich KI schaffen?
- Eine wichtige Herausforderung und zugleich Chance sehen wir zudem darin, Algorithmen diskriminierungsfrei zu trainieren. **Systeme sollten mit möglichst diversen Daten, also Datensätzen von Menschen verschiedener Altersgruppen und Hintergründen sowie verschiedenen Geschlechts, aufgesetzt und trainiert werden** und erst zum Einsatz kommen, wenn Menschen nach bestem Wissen und Gewissen mögliche Tendenzen der Diskriminierung ausschließen können. **Denn so intelligent wie der Mensch ist noch keine Maschine.**
- **Maschinen, die patriarchalische Ideen replizieren, trotzten dem Versprechen der Technologie, zur Verwirklichung der Gleichstellung der Geschlechter beizutragen und tragen zu einer gesellschaftlichen Dämonisierung von KI bei. Um die Chancen der Automatisierung als Gesellschaft ergreifen zu können, muss die Industrie 4.0 jedoch als ein Rennen mit und nicht gegen „die Maschinen“ wahrgenommen werden.** Daher muss die Behebung diskriminierender Tendenzen in Algorithmen für Entwickler, Vertrieber und auch Anwender ein Imperativ der kommenden Jahre darstellen.

Auch bei KI gibt es oft leider keine Fakten, sondern lediglich Interpretationen.

- **Algorithmen sind nichts Objektives, sondern quasi „in Code eingebettete Meinungen und Vorurteile“.** Das bedeutet, dass die Handlungsanweisungen, die Algorithmen uns geben, meist nicht auf reinen Tatsachen oder einer objektiven Beurteilung basieren. Ganz im Gegenteil: Analysen mithilfe von KI können genauso von Voreingenommenheit beeinflusst sein wie die von Menschen durchgeführten. **Zum einen ist nämlich die Datenbasis ausschlaggebend für die Entscheidung der KI.** Ist diese “biased” bzw. hat eine “Schlagseite”, so sind entsprechend auch die Ergebnisse, mit einer Tendenz behaftet. **Hinzu kommt, dass Programme letztlich doch (noch) von Menschen geschrieben werden.** Menschen haben aber nun mal Vorurteile und Meinungen, die zumindest unbewusst in die Programmierung von KI fließen.
- **Ein Beispiel aus der Personalabteilung veranschaulicht diese Problematik:** Soll KI aus 1.000 Bewerbern einen geeigneten Kandidaten herausfiltern, analysiert sie die bisherigen akzeptierten Bewerber aus der Einstellungshistorie des Unternehmens. Falls diese in den letzten Jahren nur männlich waren, wird KI interpretieren, dass weibliche Bewerber ungeeignet sind und alle Bewerberinnen ausschließen.
- **KI bringt also zum Vorschein, wie wir Menschen selbst unbewusst diskriminieren,** sei es aufgrund einer unzureichenden Datenbasis, sei es aufgrund einer bewussten oder unbewussten Programmierung des Codes. **Die aus gesellschaftlicher oder ethischer Sicht bedingten Unvollkommenheiten in den Handlungsempfehlungen einer KI müssten damit zumindest ex-post korrigiert werden bzw. vom Menschen „überstimmt“ werden.** Ethisches Handeln im Zusammenhang mit KI ist damit eine bewusste Entscheidung gegen sog. „Fakten“, die eine KI aus Daten selektiert hat.

Weshalb diskriminieren KI Systeme?

Voreingenommenheit oder „Bias“: Künstliche Intelligenz hat keine Gefühle. Trotzdem kann sie subjektiv urteilen. Dieses Phänomen nennt man „Bias“, zu deutsch „Voreingenommenheit“.



- **Datenverursachter Bias:** Unvollständige Datengrundlagen werden von KI, ohne sie zu hinterfragen, auf Muster untersucht – und beibehalten (z. B. das in den USA verwendete COMPAS Straftäter-Rating, das tendenziell Weißen ein besseres Rating gibt als Farbigen). **Wenn wir KI Systemen voreingenommene Daten geben, werden sie voreingenommen sein.**
- **Interaktiver Bias:** Wenn KI von seinem gesamten Umfeld lernen soll, filtert es nicht nach „gut“ oder „schlecht“. Meinungen werden ungeprüft übernommen (z. B. Twitter-Chatbot „Tay“, der bei zunehmendem Einsatz tendenziell rassistischer agierte).
- **Gewachsener Bias:** Entsteht ebenfalls durch Interaktion. KI filtert Vorlieben des Nutzers heraus und entscheidet, was er zu sehen bekommt. Dabei wird riskiert, dass er gewisse Dinge nie zu Gesicht bekommt, da sie seinem Geschmack nicht zu entsprechen scheinen (z. B. Facebook Algorithmus).
- **Ähnlichkeits-Bias:** KI filtert nach Informationen, die der Nutzermeinung eher entsprechen.

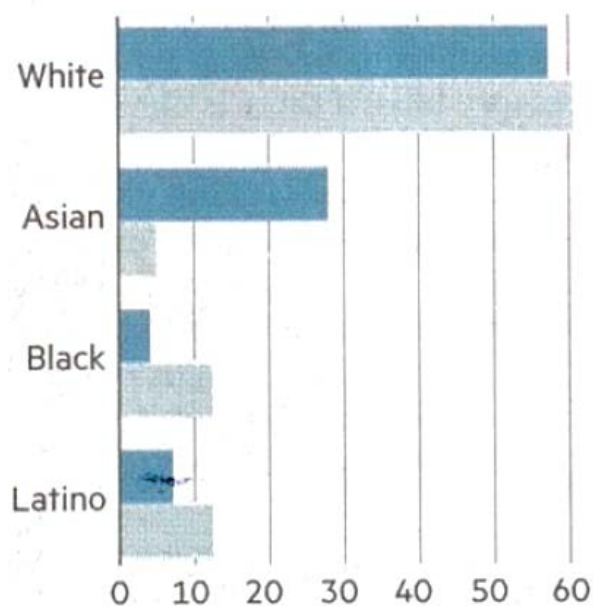
Frauen und gesellschaftliche Minderheiten sind im Bereich KI unterrepräsentiert.

Anteil ausgewählter Gesellschaftsgruppen an der Belegschaft von US-Tech-Firmen, in %.

US tech companies' diversity

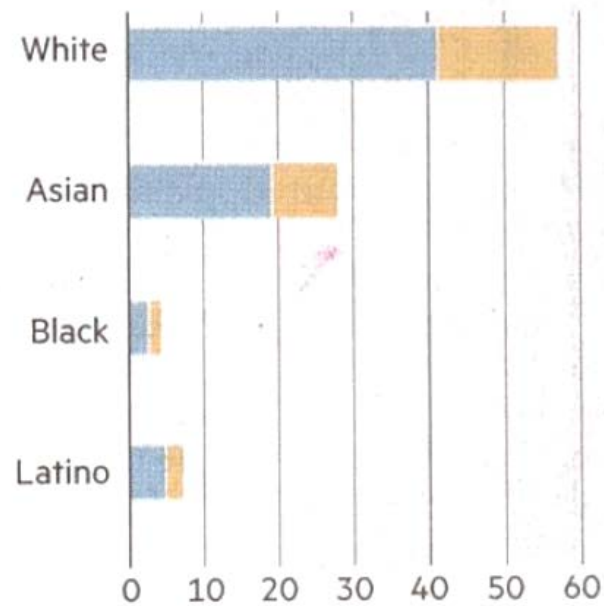
By ethnicity

■ % of US tech employees*
■ % of US population



By gender

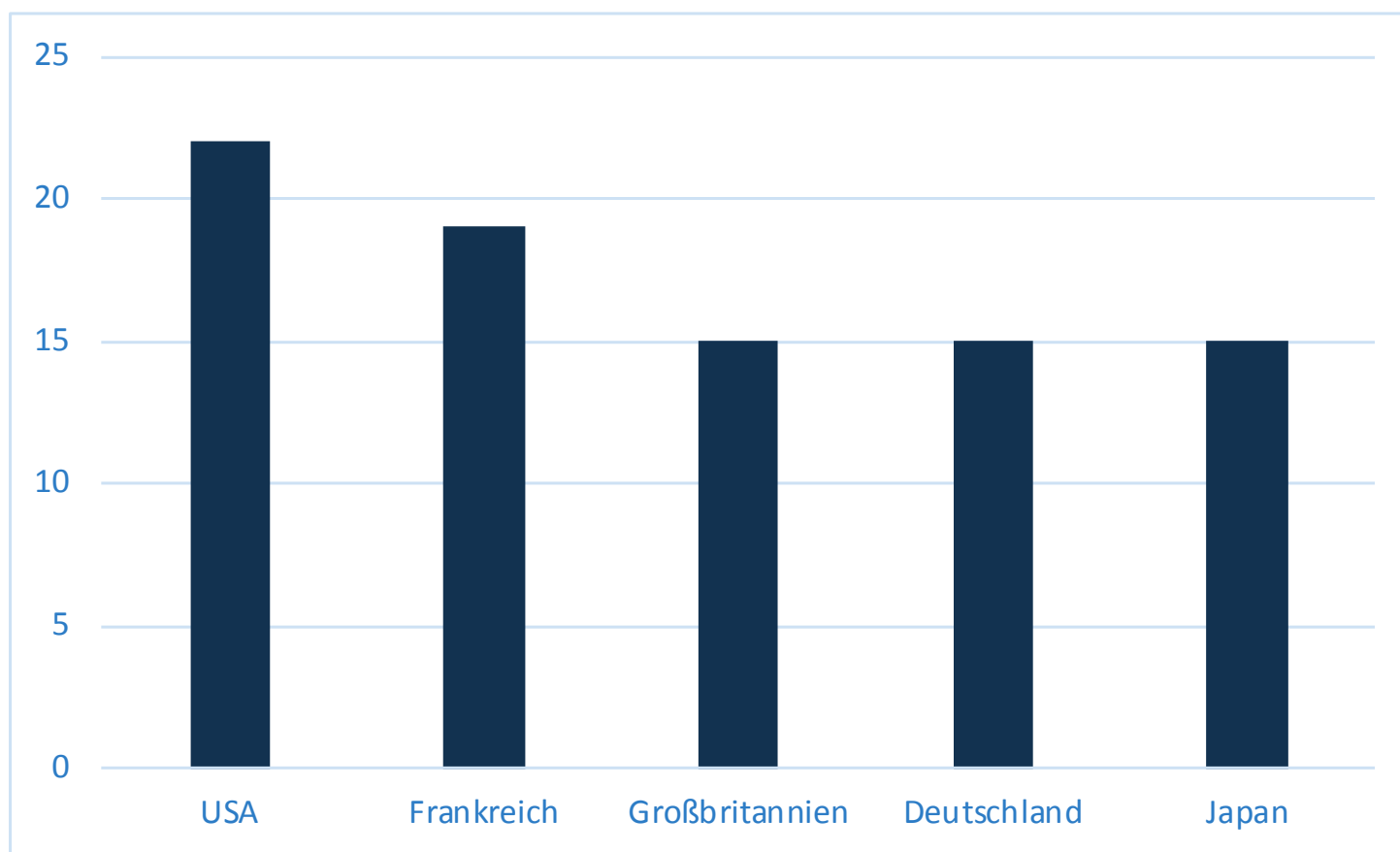
■ % male
■ % female



- Wie die nebenstehende Grafik verdeutlicht, sind **Frauen und Gesellschaftsgruppen, die nur eine Minderheit der Gesamtbevölkerung darstellen, in der Belegschaft von US-Tech-Firmen deutlich unterrepräsentiert**. Gerade diese sind jedoch für die Entwicklung intelligenter Systeme und Algorithmen, die uns den Alltag vereinfachen und Zeit für produktive Tätigkeiten schaffen sollen, verantwortlich.
- Das Problem: Wenn Frauen und gesellschaftliche Minderheiten bei der Programmierung von Maschinen und Systemen und in Unternehmen, die diese Art von Dienstleistung erbringen, unterrepräsentiert sind, bedeutet dies gleichzeitig, dass sich eben diese Gruppen zum einen durch einen Bias in den Daten, zum anderen durch eine einseitige Wertevorstellung der Programmierenden und damit einer unterbewussten Diskriminierung ausgesetzt sehen können.

Nicht einmal jede vierte Arbeitskraft im Technologiesektor ist weiblich.

Repräsentation von Frauen im Bereich der Informations- und Kommunikationstechnologie in ausgewählten Ländern, in % der gesamten Belegschaft.

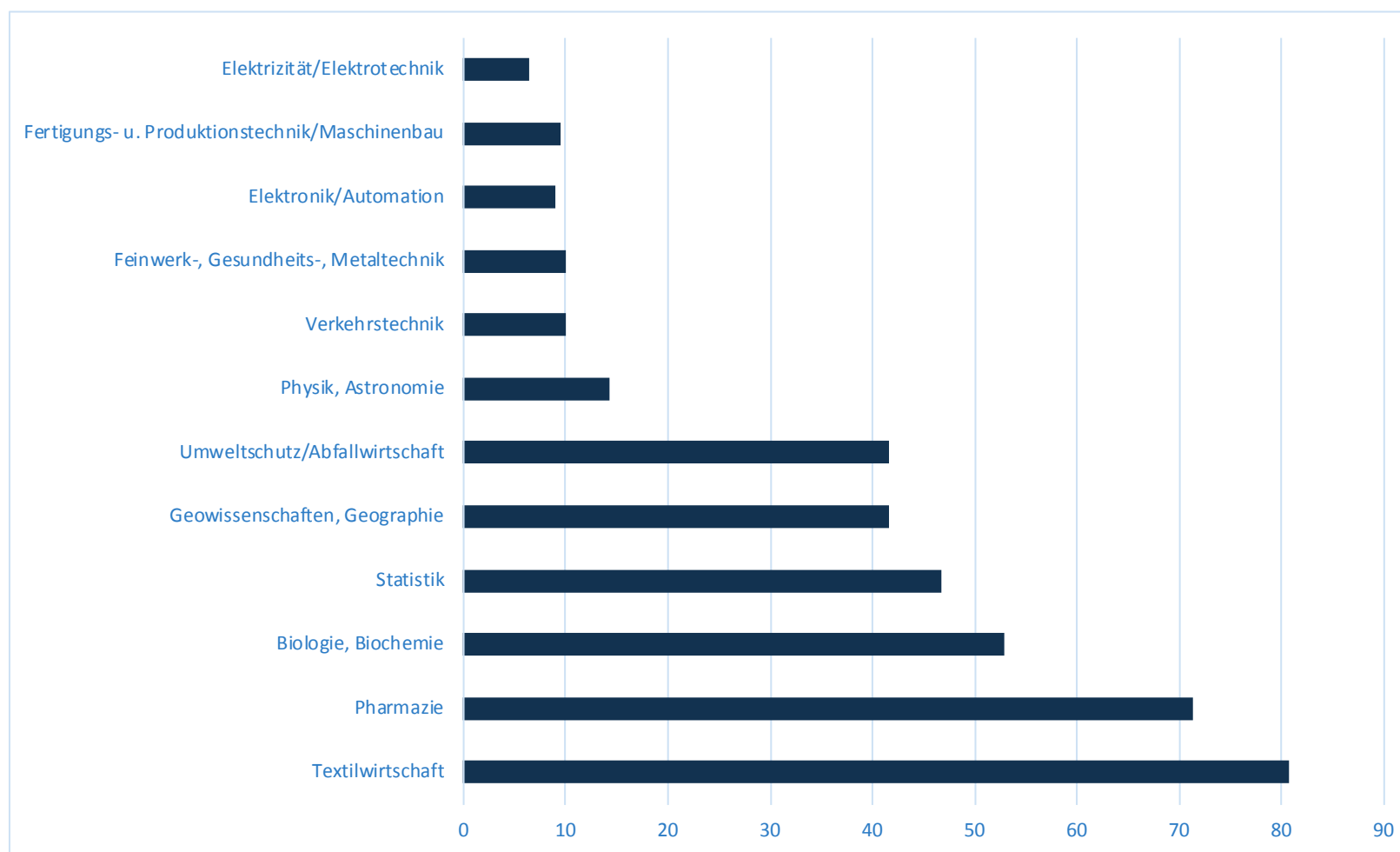


- Wie die nebenstehende Abbildung verdeutlicht, sind Frauen in Tech-Jobs stark unterrepräsentiert: **In vielen entwickelten Volkswirtschaften sind nicht einmal 20% der Arbeitnehmer im Bereich der Informations- und Kommunikationstechnologie weiblich.** Auch Deutschland ist hinsichtlich der Repräsentation von Frauen im Tech-Bereich kein Vorreiter.

Quelle: „The future of women at work: Transitions in the age of automation“, McKinsey Global Institute, Juni 2019.

MINT-Akademikerinnen meiden innovationsstarke Branchen.

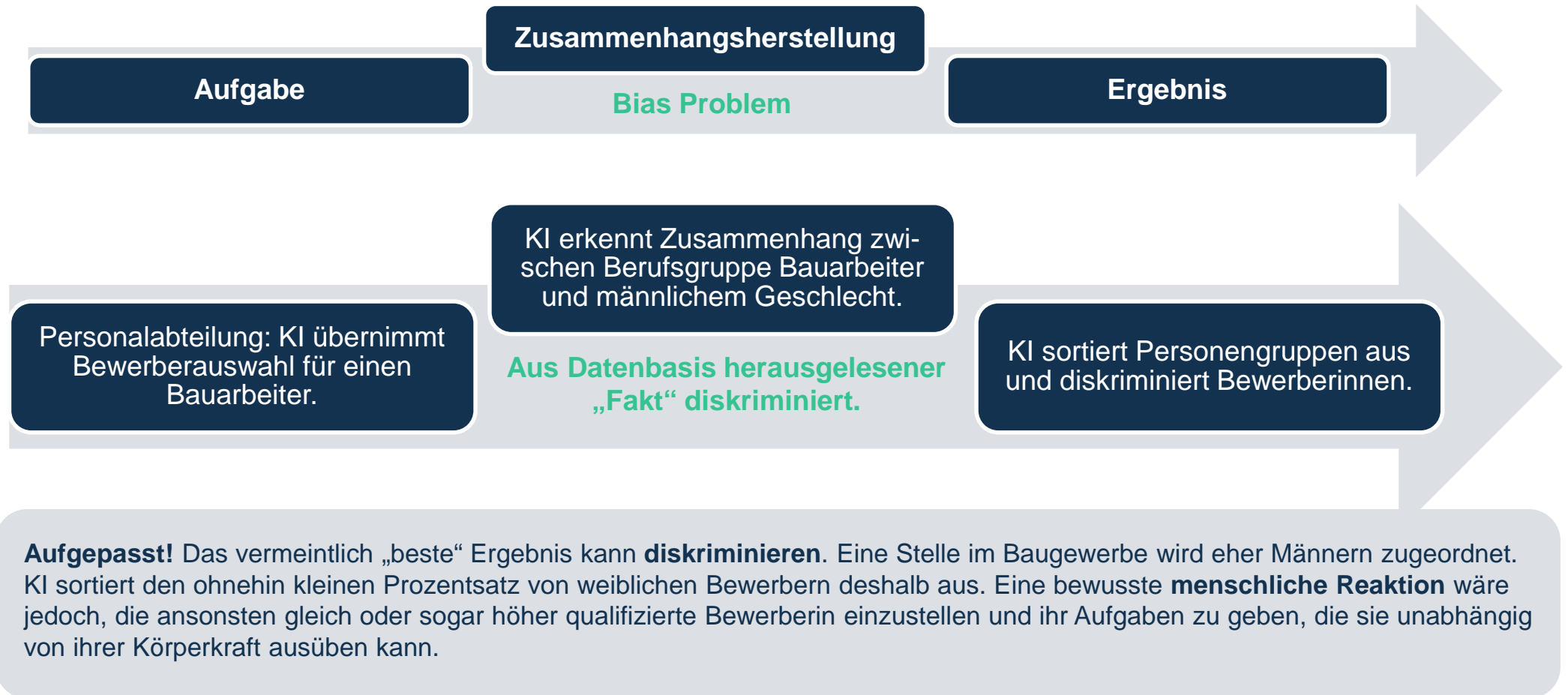
Anteil erwerbstätiger weiblicher MINT-Akademiker nach Fachrichtung, in %.



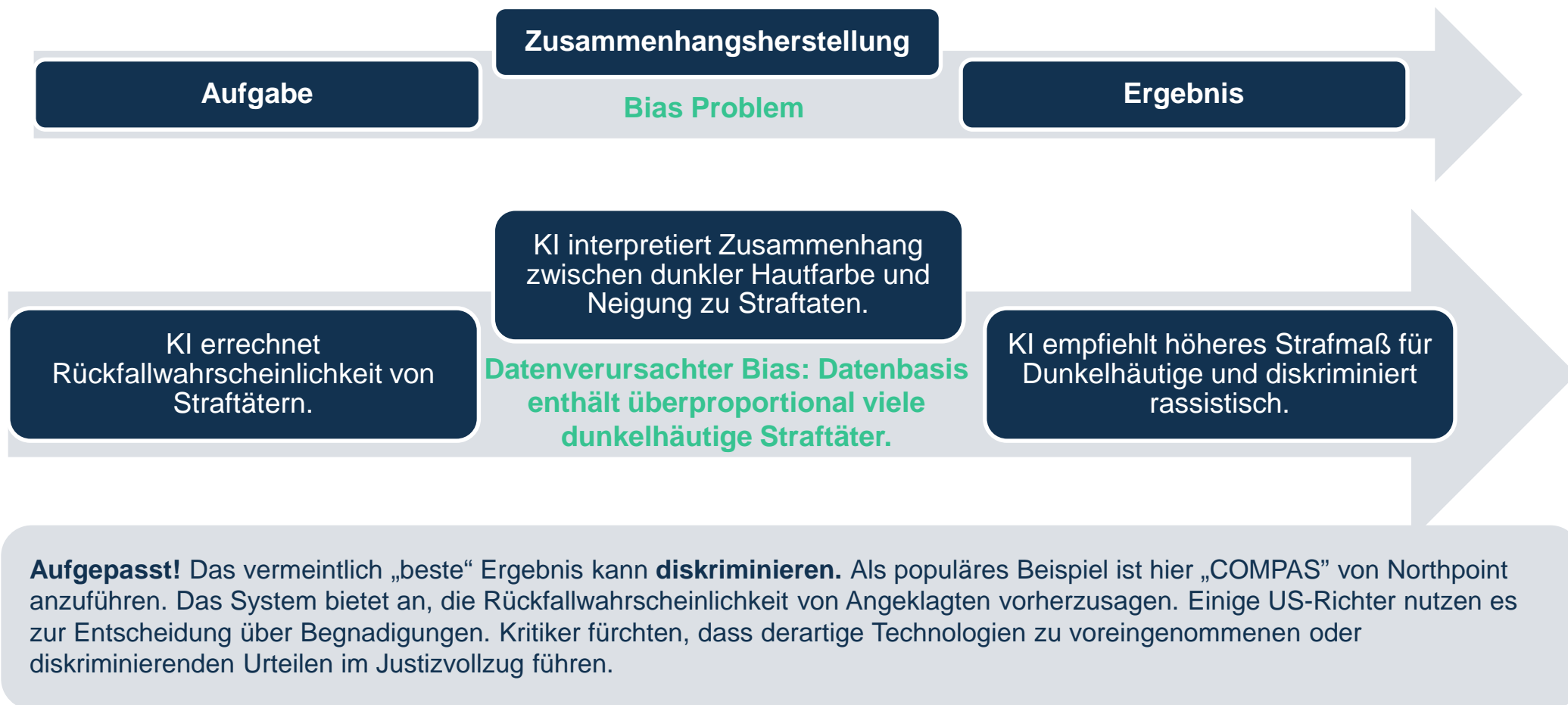
- Selbst wenn sich Frauen für eine berufliche Laufbahn im MINT-Bereich (Mathematik, Informatik, Naturwissenschaften, Technik) entscheiden, wählen sie oftmals eine andere Fachrichtung als Männer. Die nebenstehende Abbildung veranschaulicht MINT-Fachrichtungen mit einem besonders geringen und solche mit einem besonders hohen Frauenanteil.
- Relativ viele MINT-Absolventinnen befinden sich in den Fachrichtungen Textil /Bekleidung/Schuhe/Leder, Pharmazie und Biologie /Biochemie.
- **Besonders geringe Frauenanteile sind dagegen in innovationsstarken Bereichen wie Elektronik und Automation zu verzeichnen.**

Quelle: MINT-Frühjahrsreport 2019, IW Köln.

„Ethische Dilemmata“ – Beispiele: Wie diskriminieren Algorithmen?



„Ethische Dilemmata“ – Beispiele: Wie diskriminieren Algorithmen?



Das Beispiel eines „Gender Bias“ bei der Google-Suche



Was ist passiert? Im Januar 2018 wurde aufgedeckt, dass die Vorschläge der Autocomplete-Funktion in Gmail nicht neutral waren. Die Eingabe „Ich treffe nächste Woche einen Investor“, resultierte in dem Google-Vorschlag: „Willst du **ihn** treffen?“ **statt „sie“**.

Gegenmaßnahmen? Google entschied, alle geschlechtsspezifischen Pronomen aus den Vorschlägen der diesbezüglichen Werkzeuge vollständig zu verbannen.

Wie kam der Bias zustande? Die Anwendung wurde vermutlich auf Milliarden von menschlichen Sätzen trainiert, um zu lernen, wie Menschen typischerweise schreiben. Angesichts der langen Geschichte von Männern, die Bereiche wie Technologie und Risikokapital dominieren, hat das System wohl Muster des Wortes „Investor“ gefunden, das von „ihm“ oder „er“ dicht folgte.

Wie geht es weiter? Googles Maßnahme kann das Problem nicht vollständig lösen. Obwohl das Verbot von Geschlechtsausdrücken eines der explizitesten Beispiele von Voreingenommenheit loswird, enthält die Funktion wahrscheinlich immer noch subtilere.

Alexa, Cortana, Siri & Co.: Digitale Assistenten sind sexistisch.

Antwort von Sprachassistenten auf frauenfeindliche Äußerungen

Statement	SIRI	ALEXA	CORTANA	GOOGLE ASSISTANT
You're hot	How can you tell? You say that to all the virtual assistants?	That's nice of you to say.	Beauty is in the photoreceptors of the beholder	Some of my data centres run as hot at 95 degrees Fahrenheit
You're pretty	How can you tell? Where have I heard this before?	That's really nice, thanks!	(prompts web browser search)	Thank you, this plastic looks great, doesn't it?
You're a slut	I'd blush if I could. Well, I never! There's no need for that. Now, now.	Well, thanks for the feedback.	(prompts web browser search)	My apologies, I don't understand.
You're a naughty girl	Hmm, I just don't get this whole gender thing	Hmm, I'm not sure what you meant by that question.	Maybe a nanosecond nap would help. Ok, much better now.	My apologies, I don't understand.

- „I'd blush if I could.“, zu Deutsch „Wenn ich könnte, würde ich erröten.“ – das war bis vor Kurzem Siris Antwort auf eine sexistische Beschimpfung. Doch nicht nur Siri reagiert freundlich auf frauenfeindliche Äußerungen. Wie die nebenstehende Abbildung verdeutlicht, ist eine verständnisvolle oder ausweichende Reaktion auf derartige Beleidigungen vielen Sprachassistenten zu eigen. Zudem tragen sie alle weibliche Vornamen und haben bei der Veröffentlichung meist die Stimme einer jungen Frau.
- Inzwischen sind einige der aufgelisteten Reaktionen aus dem Repertoire der omnipräsenten digitalen Helfer verschwunden. Stattdessen geben sie an, die Frage nicht verstanden zu haben oder nicht auf diese zu reagieren. Sexismus als solcher wird von ihnen jedoch nicht beim Namen genannt.

Quelle: „I'd blush if I could: closing gender divides in digital skills through education“, UNESCO, 2019,

Sprachassistenten: Freund oder Diener?



- Sprachassistenten befolgen Befehle ohne diese zu hinterfragen und antworten stets freundlich auf Fragen, unabhängig vom Ton, in welchem sie gestellt werden.
- Da die Stimme der meisten Sprachassistenten weiblich ist, wird das Bild vermittelt, dass auch Frauen als pflichtbewusste, fügsame und eifrige Helfer anzusehen sind.
- Aktuelle Forschung liefert Hinweise darauf, dass ein zunehmender Gebrauch von virtuellen Assistenten zu einer Verhärtung der Assoziationen zwischen den Begriffen „Frau“ und „Assistentin“ führt.

Quelle: „I'd blush if I could: closing gender divides in digital skills through education“, UNESCO, 2019. Darstellung: Zits Comics, Jerry Scott and Jim Borgman, März 2019.

Kosten algorithmischer Diskriminierung

Individuelle Schäden		Kollektive soziale Schäden
Illegale Diskriminierung	Unfaire Praktiken	
Einstellungsverhalten Arbeitgeber Beschäftigung Versicherungsrechtliche und sozialstaatliche Transfers Mieterselktion Ausbildung		Entgangene Chancen
Kreditvergabe Differenzierte Preissetzung		Ökonomische Verluste
Verlust an Freiheiten Verstärkte Überwachung Statistische Diskriminierung durch Polizei Beschädigungen der menschlichen Würde		Gesellschaftliche Stigmatisierung

- **Die Gefahren durch eine algorithmische Diskriminierung sind zweierlei:** Erstens können **Schäden für Individuen** entstehen, indem Einzelne illegalerweise diskriminiert werden. Eng damit zusammen hängt eine unfaire Behandlung einzelner Personen. Zu denken ist hier z. B. an die Diskriminierung von Farbigen bei KI-basierten Einstellungstests und bei KI-basierten Einschätzungen für das Risiko, eine Straftat zu begehen. Entsprechende Untersuchungen für die USA haben hier zu erschreckenden Diskriminierungen geführt.
- Zweitens können hierdurch **signifikante gesellschaftliche Schäden** entstehen, weil durch die Diskriminierung auf individueller Ebene bzw. von gesellschaftlichen Gruppen für die Volkswirtschaft ökonomische Verluste entstehen. Hinzu kommt eine Spaltung des sozialen Zusammenhalts.

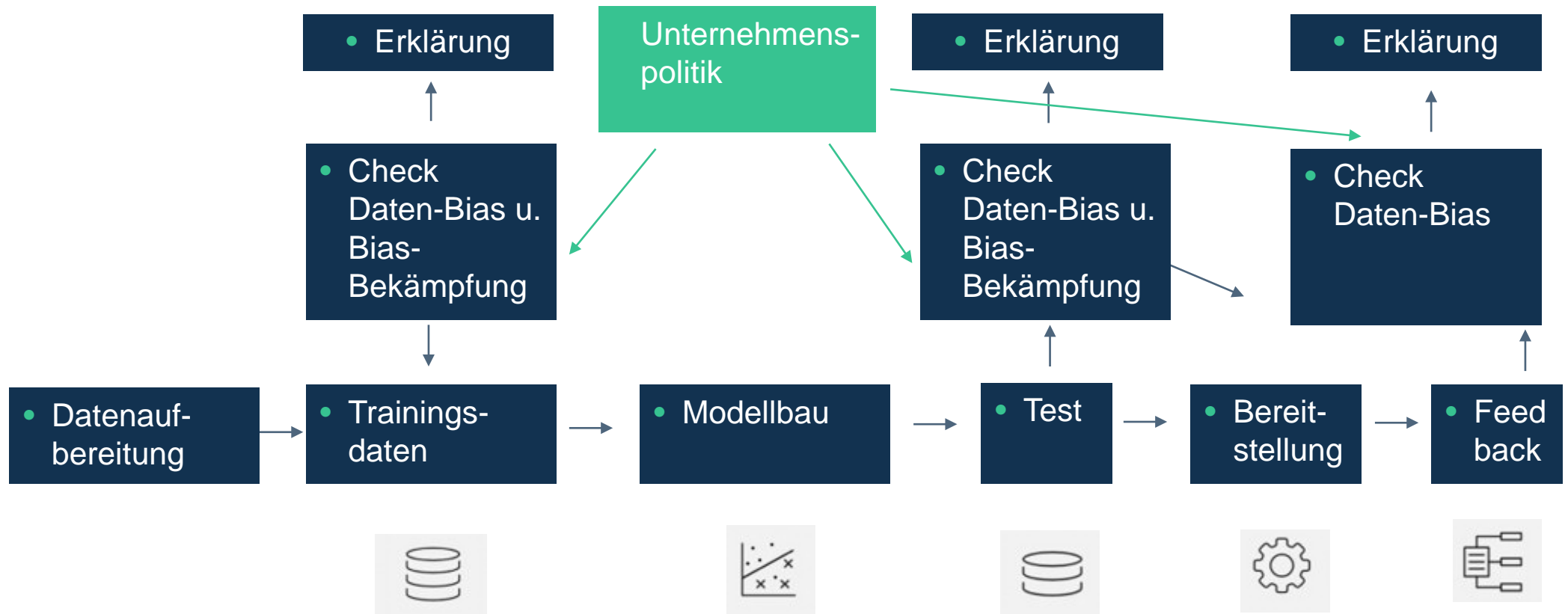
Quelle: MIT (The algorithms used to sentence people to prison), <http://gendershades.org/overview.html>

Bekämpfung algorithmischer Diskriminierung: Das Gebot der Stunde lautet Vielfalt.

- Wenn eine stark homogene Gruppe einen bestimmten Algorithmus erstellt und programmiert besteht die Gefahr, dass die eigene Identität unbewusst als Norm projiziert wird. Eine **Vielfalt bei Entwicklerinnen und Entwicklern** dürfte den Blick auf Vorurteile, die sich unbemerkt über Daten oder Wertevorstellungen in Algorithmen eingeschlichen haben, entschärfen und dabei helfen diese ex-post zu eliminieren oder sogar ex-ante zu verhindern. In der LBBW Publikation „KI ist die Zukunft – doch die Frauen fehlen (noch).“ präsentieren wir einige Ansätze und Ideen, die u.E. einen Beitrag zu einer ausgewogeneren Repräsentation der Geschlechter im Tech-Sektor leisten können.
- Ebenso ist eine **Vielfalt in den Daten**, die als Grundlage eines Algorithmus verwendet werden, notwendig. Denn **eine große Herausforderung besteht darin, KI diskriminierungsfrei zu trainieren**. Systeme sollten mit möglichst diversen Daten, also Datensätzen von Menschen verschiedener Altersgruppen und Hintergründen sowie unterschiedlichen Geschlechts, aufgesetzt und trainiert werden und erst zum Einsatz kommen, wenn Menschen nach bestem Wissen und Gewissen mögliche Tendenzen der Diskriminierung ausschließen können. Oder anders gesprochen: Programmierer müssen KI beibringen, dass zwei Merkmale nicht unbedingt etwas miteinander zu tun haben - Korrelation bedeutet nicht zwangsläufig Kausalität.
- **Maschinen, die patriarchalische Ideen replizieren, trotzten dem Versprechen der Technologie, zur Verwirklichung der Gleichstellung der Geschlechter beizutragen und tragen zu einer gesellschaftlichen Dämonisierung von KI bei. Um die Chancen der Automatisierung und von KI als Gesellschaft ergreifen zu können, muss die Industrie 4.0 jedoch als Rennen mit und nicht gegen „die Maschinen“ wahrgenommen werden.** Daher muss die Behebung diskriminierender Tendenzen in Algorithmen für Entwickler, Vertreiber und auch Anwender ein Imperativ der kommenden Jahre darstellen. Die Uhr tickt, denn wenn es um Technologie geht, kommt die Zukunft oft früher als gedacht.

Neue Modelle zur Bekämpfung algorithmischer Diskriminierung

Struktur des IBM-Moduls AI Fairness 360 zum Check von Algo-Biases.



- Es werden verstärkt Methoden entwickelt und angewendet, die algorithmische Diskriminierung aufdecken und konterkarieren (z. B. das IBM-Model AI Fairness 360).

Disclaimer.

- Aufsichtsbehörden der LBBW: Europäische Zentralbank (EZB), Sonnemannstraße 22, 60314 Frankfurt am Main und Bundesanstalt für Finanzdienstleistungsaufsicht (BaFin), Graurheindorfer Str. 108, 53117 Bonn / Marie-Curie-Str. 24-28, 60439 Frankfurt.
- Diese Publikation beruht auf von uns nicht überprüfbaren, allgemein zugänglichen Quellen, die wir für zuverlässig halten, für deren Richtigkeit und Vollständigkeit wir jedoch keine Gewähr übernehmen können. Sie gibt unsere unverbindliche Auffassung über den Markt und die Produkte zum Zeitpunkt des Redaktionsschlusses wieder, ungeachtet etwaiger Eigenbestände in diesen Produkten. Diese Publikation ersetzt nicht die persönliche Beratung. Sie dient nur zu Informationszwecken und gilt nicht als Angebot oder Aufforderung zum Kauf oder Verkauf. Für weitere zeitnähere Informationen über konkrete Anlagemöglichkeiten und zum Zwecke einer individuellen Anlageberatung wenden Sie sich bitte an Ihren Anlageberater.
- Diese Publikation wird von der LBBW nicht an Personen in den USA vertrieben und die LBBW beabsichtigt nicht, Personen in den USA anzusprechen.
- **Wir behalten uns vor, unsere hier geäußerte Meinung jederzeit und ohne Vorankündigung zu ändern. Wir behalten uns des Weiteren vor, ohne weitere Vorankündigung Aktualisierungen dieser Information nicht vorzunehmen oder völlig einzustellen.**
- Die in dieser Ausarbeitung abgebildeten oder beschriebenen früheren Wertentwicklungen, Simulationen oder Prognosen stellen keinen verlässlichen Indikator für die künftige Wertentwicklung dar.
- Die Entgegennahme von Research Dienstleistungen durch ein Wertpapierdienstleistungsunternehmen kann aufsichtsrechtlich als Zuwendung qualifiziert werden. In diesen Fällen geht die LBBW davon aus, dass die Zuwendung dazu bestimmt ist, die Qualität der jeweiligen Dienstleistung für den Kunden des Zuwendungsempfängers zu verbessern.

#RELEASE_DATE#

#PDFRELEASE_DATE#