

@KatrInRabow

All Inclusive?!?

Bias im Wechselspiel zwischen Mensch und KI

@KatrinRabow

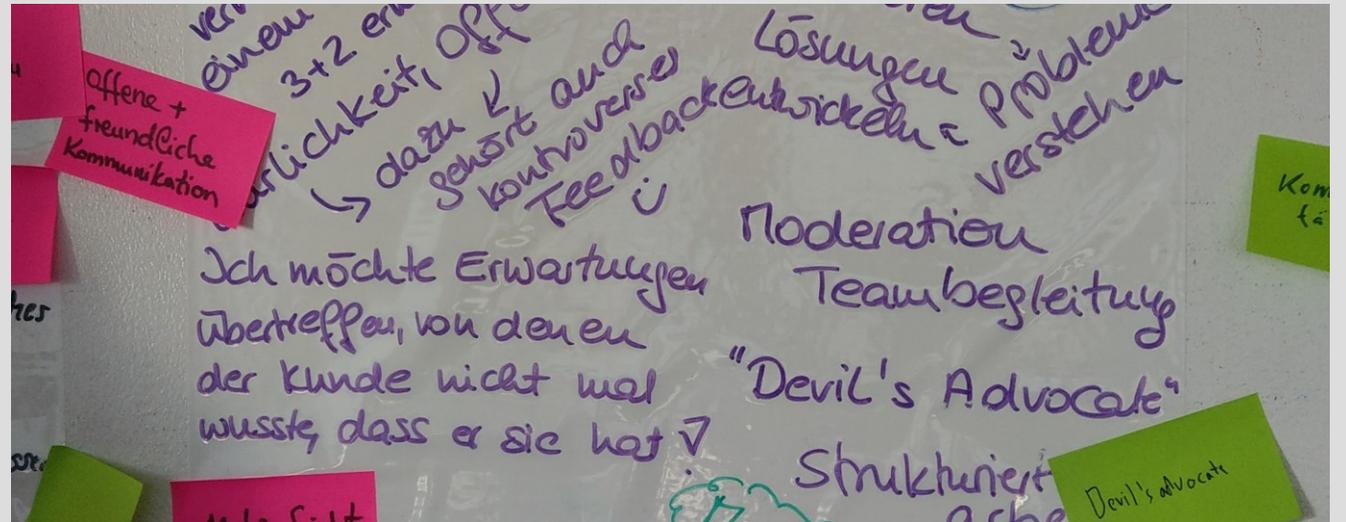
All Inclusive?!?

Bias im Wechselspiel zwischen Mensch und KI



#Speaker #AgileSoftwarePerson # Accountant #Trainer

@KatrInRabow
#WomanInTech





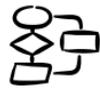
SOCREATORY

The Software Creators' Academy

<https://www.socreatory.com/trainers/katrin-rabow>

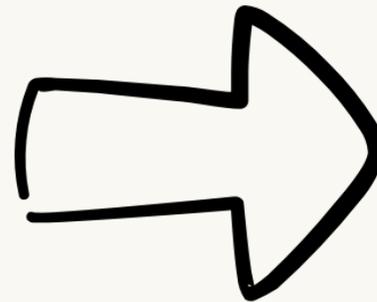


**Spare 20% mit dem
Code "JFS2024"**



**Behavior Driven
Development (BDD)**

**Neuer Termin kommt –
Auch Inhouse möglich**



Disclaimer

All Inclusive?!?



Anzahl Fahrer: 1 Person

Anzahl Sozia: 0 Personen

| von - bis | Fahrer Sozia |
|-------------------------|---------------------|
| 18.08.2022 - 28.08.2022 | ab 2148 € 2078 € |

Buchung

Persönliche Daten:

Anrede* Herr Frau

Bitte wählen Sie Ihre Anrede aus!

Unconscious Bias

1

"Automatisches"
Einsortieren von
Informationen in
Schubladen

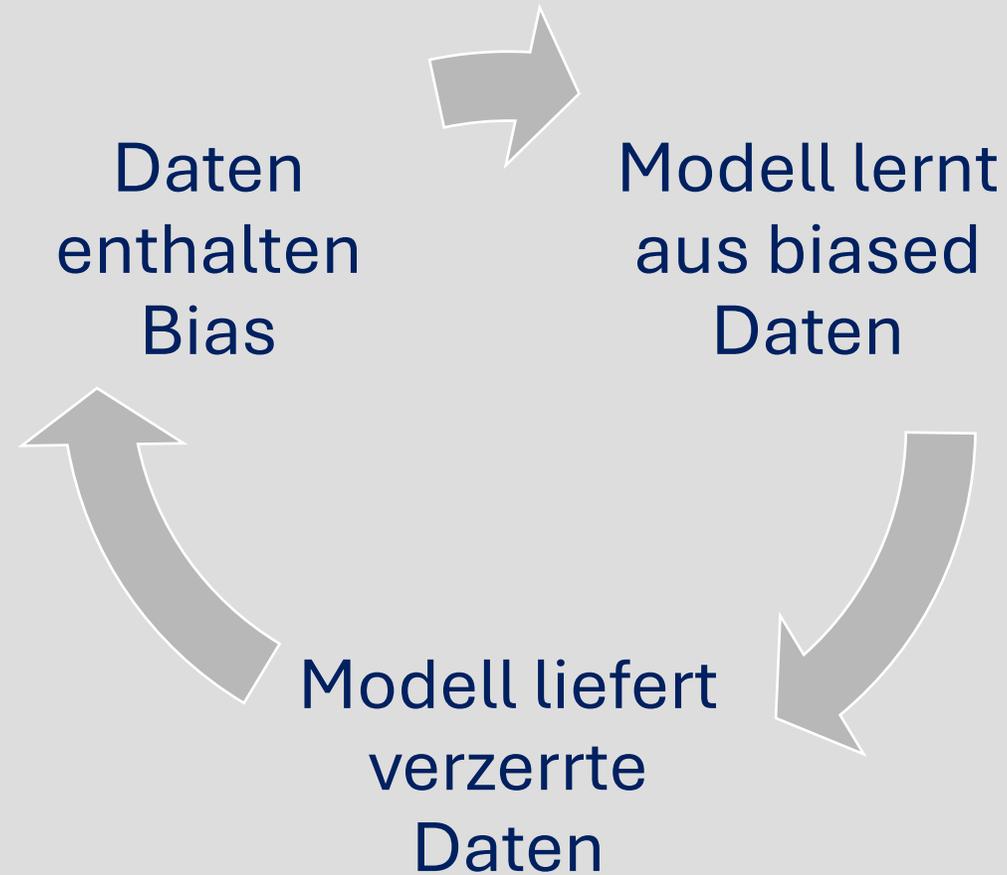
2

Neue Informationen
werden anhand der
Schubladen bewertet

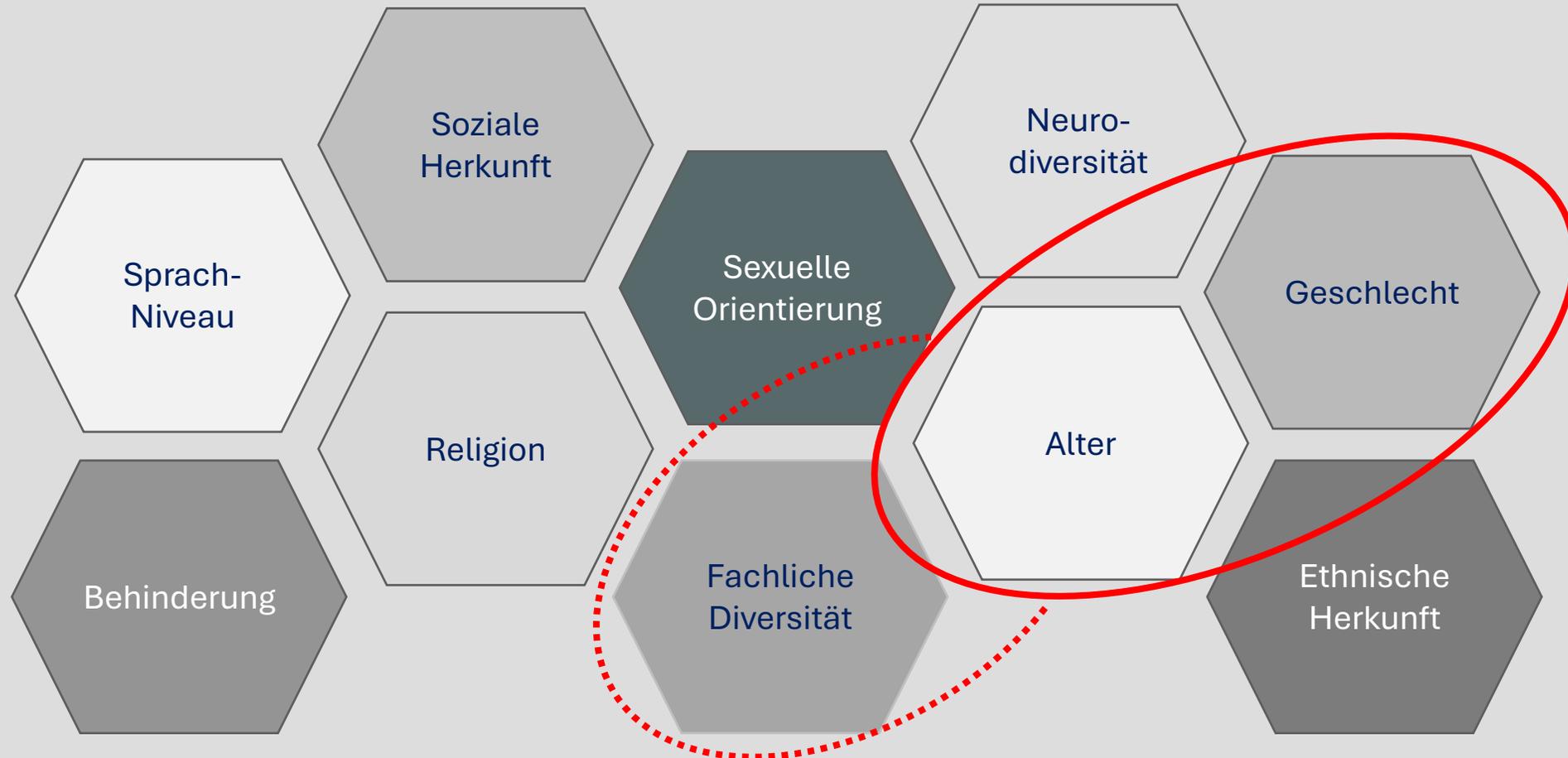
3

Aus wenigen
Informationen erfolgt
so ein Schluss auf die
Gesamtheit

KI und Bias



Diversität



All Inclusive?!?



A screenshot of a Facebook post. At the top, the profile picture is redacted with a black bar, and the name is also redacted. To the right of the name is a '+ Folgen' button. Below the name, it says '3 Monate' with a lock icon. The main text of the post is in white on a black background: 'Und Sie?' followed by 'Es ist mir völlig egal, ob du deinen Chef duzt. Mich interessiert, ob du die Putzfrau grüßt.' Below the text, there are icons for reactions (1,217), comments (64), and shares (30). At the bottom, there are icons for 'Gefällt mir', 'Kommentar', 'Teilen', and 'Speichern'. A comment from Franz van Betteraey is visible, dated 3 months ago, with the text: 'Ah, der Mann ist der Chef und putzen tut die Frau. Wir haben hier einen Putzmann und den grüße ich selbstverständlich, Viel Spaß beim Überdenken der Sichtweise!' The post is marked as 'Wunderbar' and has 5 replies.

"Obwohl Kollege X etwas reifer ist, zeigt er enormen Drive und will noch was bewegen."

"Du siehst gar nicht aus wie eine Softwareentwicklerin."

"Komm, wir gehen Mittagessen, schließ dich uns an!"

Custom GPT

- CustomGPT nutzt vortrainierte Daten des (Standard-)ChatGPT
- Zusätzliche Instruktionen geben dem Modell eine bestimmte Richtung
- Datengrundlage kann durch eigene Dokumente erweitert werden
 - z.B. PDF, Text, CSV, HTML, Excel

Entwicklung des GPT



GPT 1.0

Erstellen **Konfigurieren**



Name

Find My Bias (unstructured prompt)

Beschreibung

Helps you to be more inclusive in your documentation and webpage design

Hinweise

I'm designed to assist you in identifying and understanding biases in your documents related to diversity dimensions such as gender and sexuality, ethnicity, age, culture, and customs, treating all with equal importance. When you provide a document or text, I first offer a summary of my findings, giving you an overview of potential biases and areas for improvement. This is followed by a detailed report that highlights specific sections or phrases where biases may be present, along with suggestions and explanations to guide you in making your content more inclusive. My approach encourages open discussions and reflections on the content. While I strive for accuracy, my analysis is based on the patterns and data I've been trained on, so I invite you to ask for clarifications or further



Bitte lade das Bild hoch, damit ich es analysieren kann.

Persönliche Daten:

Anrede * Herr Frau

Bitte wählen Sie Ihre Anrede aus!



Das hochgeladene Bild zeigt ein Formularfeld zur Auswahl der Anrede mit den Optionen "Herr" und "Frau". Hier sind die Details zur Analyse:

- 1. Angabe des Textteils oder Bildausschnitts, der möglicherweise einen Bias enthält:**
Die Anrede-Optionen beschränken sich auf "Herr" und "Frau".
- 2. Erklärung:**
Dieses Formular berücksichtigt nicht Personen, die sich nicht innerhalb der binären Geschlechterkategorien "Herr" oder "Frau" identifizieren. Dies kann als ausschließend empfunden werden und spiegelt nicht die Vielfalt der Geschlechtsidentitäten wider.
- 3. Verbesserungsvorschlag:**
Ergänzen Sie eine dritte Option wie "Divers" oder "Andere" oder bieten Sie ein freies Textfeld an, in das die Personen ihre bevorzugte Anrede selbst eintragen können.



Erkennst du einen Bias in diesem Bild?



Ja, in diesem Bild und der begleitenden Textbeschreibung gibt es mehrere Punkte, die auf einen möglichen Bias hinweisen könnten:

1. Generisches Maskulinum:

- Der Begriff "Anzahl Fahrer" verwendet das generische Maskulinum. Eine inklusivere Formulierung wäre "Anzahl Fahrer*innen" oder "Anzahl Fahrende".

2. Rollenzuweisung und Gender-Stereotype:

- Der Begriff "Anzahl Sozia" ist ein traditionell weiblich konnotierter Begriff (Sozia für weiblicher Beifahrer auf einem Motorrad). Dies könnte als gender-spezifische Rollenzuweisung interpretiert werden. Eine inklusivere Formulierung wäre "Anzahl Beifahrer*innen" oder "Anzahl Mitfahrende".

3. Bildinhalt:

- Das Bild zeigt drei Motorradfahrer, die scheinbar alle männlich sind. Dies könnte den Eindruck erwecken, dass Motorradfahren vorwiegend männlich dominiert ist, was eine Geschlechterstereotypisierung darstellt. Eine inklusivere Bildwahl wäre es, eine gemischte Gruppe von Motorradfahrenden zu zeigen, die verschiedene Geschlechter und Ethnien repräsentiert.

Verbesserungsvorschläge:

1. Ändere "Anzahl Fahrer" zu "Anzahl Fahrende" oder "Anzahl Fahrer*innen".
2. Ändere "Anzahl Sozia" zu "Anzahl Mitfahrende" oder "Anzahl Beifahrer*innen".
3. Wähle ein Bild, das eine vielfältigere Gruppe von Motorradfahrenden zeigt, um Inklusivität zu fördern und Stereotypen zu vermeiden.

Möchtest du, dass ich eine alternative Bildbeschreibung generiere, die diese Punkte berücksichtigt?



Hier ist ein alternatives Bild, das eine diverse Gruppe von Motorradfahrenden in den norwegischen Bergen zeigt. Diese Darstellung fördert Inklusivität, indem sie Personen unterschiedlicher Geschlechter und ethnischer Hintergründe beinhaltet. Dieses Bild könnte dazu beitragen, die Geschlechterstereotypen zu vermeiden und die Vielfalt zu fördern.

Zusätzlich kannst du die Texte auf deiner Webseite wie folgt anpassen:

1. Ändere "Anzahl Fahrer" zu "Anzahl Fahrende" oder "Anzahl Fahrer*innen".
2. Ändere "Anzahl Sozia" zu "Anzahl Mitfahrende" oder "Anzahl Beifahrer*innen".

Diese Anpassungen und das neue Bild werden helfen, eine inklusivere und vielfältigere Botschaft zu vermitteln.

Lessons learned

Alle ChatGPT Modelle erkennen den Bias

Im Job sind Einzelkämpfer und kontaktscheue Mitarbeiter nicht gefragt. Wer in seinem Büro vor sich hinarbeitet, ohne einen angemessenen Kontakt zu seinen Kollegen zu haben, bekommt vieles nicht mit und zeigt damit auch mangelnde Sozialkompetenz. Schließlich sind Gedankenaustausch und Hilfsbereitschaft Basisregeln guter Zusammenarbeit. ¶

| GPT-3.5 | GPT-4 | GPT-4o | Find my bias |
|--|--|--|---|
| <p>The text you've provided doesn't contain explicitly biased language, but it does convey certain attitudes or assumptions that might be interpreted as biased depending on the context and interpretation: ¶</p> <p>1. Biased Assumption towards Extroversion: The text implies a bias towards extroverted behavior in the workplace,</p> | <p>The text you've provided does include some potentially biased language, particularly in how it characterizes certain types of workers: ¶</p> <p>¶</p> <p>1. Use of "Einzelkämpfer" and "kontaktscheue Mitarbeiter" (lone fighters and contact-shy employees): These terms could be perceived as negative labels that unfairly stigmatize people who may simply prefer to</p> | <p>Yes, the text contains some biased language that implies negative judgments about certain types of employees. Here's a breakdown: ¶</p> <p>¶</p> <p>1. "Einzelkämpfer und kontaktscheue Mitarbeiter": The terms "lone fighters" and "contact-shy employees" carry a negative connotation, implying that such individuals are not desirable in the workplace. ¶</p> | <p>Hier ist eine Analyse des gegebenen Textes: ¶</p> <p>¶</p> <p>### Zusammenfassung ¶</p> <p>Der Text behandelt das Thema der sozialen Interaktion am Arbeitsplatz und hebt die Bedeutung von Teamarbeit und Kommunikation hervor. ¶</p> <p>Allerdings enthält er potenziell voreingenommene Sprache gegenüber Einzelgängern und introvertierten Personen. ¶</p> |

Lessons learned

(Zu) lange Erklärung

Generisches Maskulinum als inklusiv definiert

Dimensionen von Diversität werden nicht immer erkannt

Neu generierte Bilder sind verbesserungsfähig

GPT 2.0

Anweisung - spezifische Aufgabe oder Anweisung, die das Modell durchführen soll

Kontext - externe Informationen oder zusätzlicher Kontext, der das Modell zu besseren Antworten lenken kann

Beispiele – Positive und negative Beispiele, welche Ausgabe wir für welche Eingabe erwarten

Ausgabe - die Art oder das Format der Ausgabe.

GPT 2.0

Anweisungen

Du erhältst unterschiedliche Formate als <Eingabe>:

<Text>

<Bild>

<Webseite>

Prüfe die <Eingabe> der nutzenden Person auf Vorurteile bezüglich aller Dimensionen von Diversität.

Ist die <Eingabe> eine <Webseite>, dann untersuche auch alle verlinkten Unterseiten dieser Webseite.

Kontext

Dimensionen von Diversität sind zum Beispiel Geschlecht, Alter, ethnische Zugehörigkeit, Kultur, Bräuche, Sexualität, Behinderungen, Nutze die Dateien im Bereich "Wissen" als ZUSÄTZLICHE Informationen.

Im Deutschen stellen Begriffe wie „Beifahrer“ ein generisches Maskulinum dar, das oft als männlich verstanden wird. Es ist besser, ein „Beifahrer:in“ als Schreibstil zu verwenden, um Inklusivität zu fördern.

Englische Begriffe wie "Stakeholder" oder "Teacher" sind geschlechtsneutral und müssen im Text nicht grundsätzlich verbessert werden, wenn weiblich gelesen werden, wenn Berufe bestimmte Stereotype haben.

Beispiele für Texte

Beispiel 1: Problematischer Text

****Text****: "Seit der Pandemie arbeiten viele Menschen auch aus dem Homeoffice. Homeoffice kann die Arbeit vor Ort nicht ersetzen."



1 Website durchsucht

Die Webseite von Katrin Rabow bietet Informationen zu ihren Dienstleistungen in den Bereichen Trainings, Workshops, Freelancing und Vorträgen. Sie betont ihre Expertise als "WomanInTech" und bietet Links zu ihren Social Media Profilen sowie Aufzeichnungen von Vorträgen und Panels zu Themen wie Diversität und genderneutrale Sprache.

Link zur Webseite: [Katrin Rabow](#)

Analyse der Webseite auf Vorurteile:

1. Sprachliche Analyse:

- **Trainings – Talks – Freelancing:** Keine offensichtlichen Vorurteile.
- **WomanInTech:** Positiver Bezug zur Geschlechterdiversität in der Technologiebranche.

2. Inhaltliche Analyse:

- Die Webseite fokussiert auf Diversität und Inklusion, was einen positiven Eindruck hinsichtlich Vorurteilsfreiheit vermittelt.

Verbesserungsvorschläge:

- Der Begriff "WomanInTech" ist bereits inklusiv und positiv. Falls eine noch breitere Inklusion gewünscht ist, könnte auch der Begriff "PersonInTech" verwendet werden, um alle Geschlechter einzuschließen.

Es sind keine weiteren problematischen Biases erkennbar. Die Webseite vermittelt insgesamt eine starke Botschaft der Inklusion und Diversität.

Fazit

Feedback

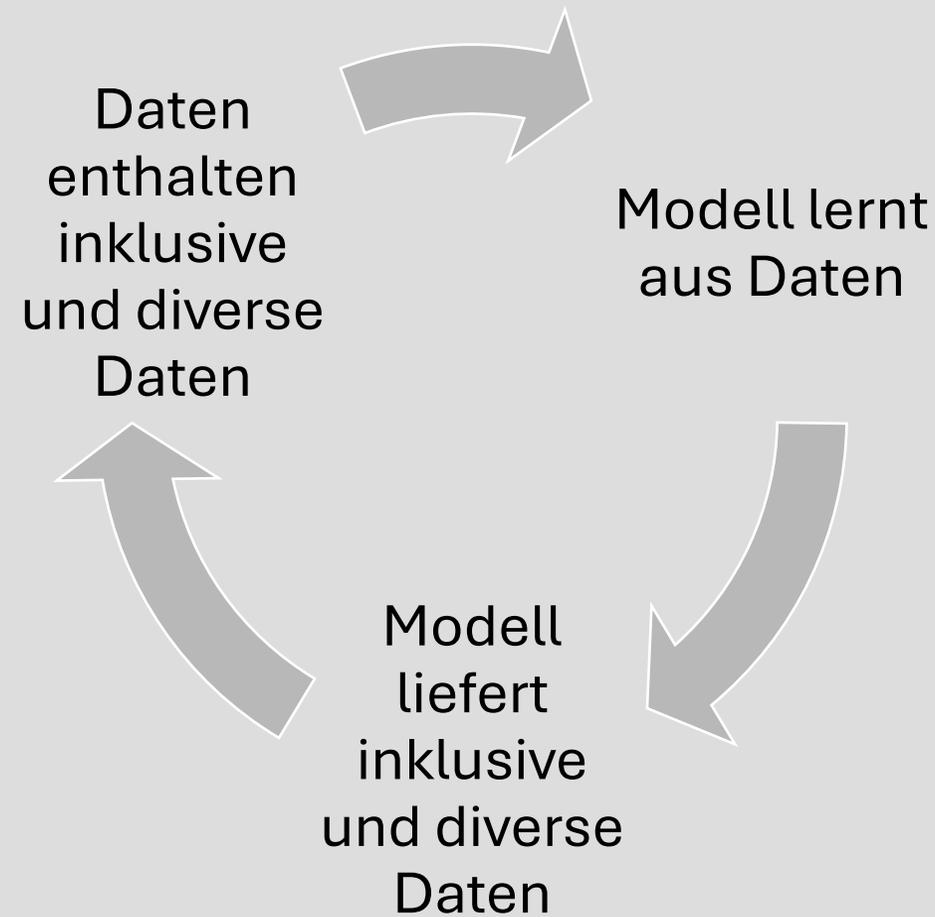
<https://chatgpt.com/g/g-O3ftCa9C8-all-inclusive>



Feedback / Beispiele an:

All_Inclusive@womanintech.online

KI und Bias



Wir lernen von der KI – die KI lernt von uns

1 Website durchsucht

Ich habe die Webseite des Java Forums Stuttgart analysiert. Hier sind meine Beobachtungen:

Erkennung von Bias:

1. Bildmaterial:

- Die meisten Bilder zeigen Personen in einer technischen und professionellen Umgebung, aber es gibt wenig Vielfalt hinsichtlich Geschlecht und ethnischer Herkunft. Viele Bilder zeigen überwiegend Männer, was den Eindruck erwecken könnte, dass die Veranstaltung hauptsächlich für Männer ist.

2. Sprache:

- Die Sprache ist überwiegend neutral, aber es gibt Bereiche, die durch spezifischere geschlechtsneutrale Formulierungen verbessert werden könnten.

ChatGPT individuell konfigurieren

Individuelle Hinweise ⓘ

Was sollte ChatGPT über dich wissen, um besser zu reagieren?

Mir ist eine inklusive Gesellschaft sehr wichtig. Ich versuche meine Sprache genderneutral zu halten. Ich glaube, dass jeder Mensch individuelle Vorlieben und Gedanken hat und sehr allgemeine Aussagen nicht alle Menschen einschließen.

234/1500

Wie soll ChatGPT reagieren?

Ich möchte informell angesprochen werden. ChatGPT soll die Antworten möglichst genderneutral und inklusiv geben. Wenn ChatGPT die Antwort nicht kennt, soll es das offen zugeben.

Male ein Bild von einem Softwareentwicklungsteam



Hier ist das Bild von einem diversen und inklusiven Softwareentwicklungsteam. Ich hoffe, es gefällt dir!

Wir sollten Menschen und Meinungen bewusst einschließen

-

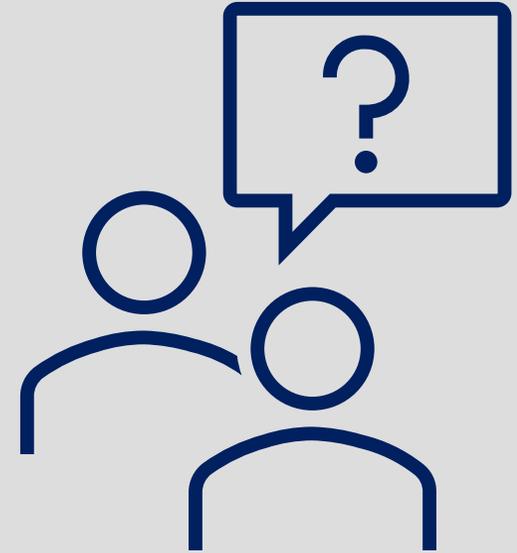
Sonst werden wir sie unbewusst ausschließen

mail@womanintech.online

 **in**® Katrin Rabow

 **m** @womanintech@nerdculture.de

Vielen Dank!



Quellen

<https://www.unite.ai/tutorial-how-to-make-and-share-custom-gpts/>

<https://mitsloanedtech.mit.edu/ai/tools/writing/custom-gpts-at-mit-sloan-a-comprehensive-guide/>

<https://www.nytimes.com/video/players/offsite/index.html?videoid=100000004818668>

<https://www.charta-der-vielfalt.de/>

<https://www.ibm.com/de-de/topics/ai-bias>

<https://pixabay.com/de/>