

Computer Says No!

Wie wir faire, inklusive Systeme bauen können

**GUTE KI.
GUTE ARBEIT.**



Team Passerelle

Das KI-Institut der IG Metall

Dr. Meike Zehlike

www.passerelle.team

KURZPROFIL

Meike Zehlike



Dr. Meike Zehlike ist mehrfach ausgezeichnete Wissenschaftlerin für künstliche Intelligenz und ihre sozialen Auswirkungen.

Als Pionierin in Deutschland beschäftigt sich Meike seit 2016 damit, inwiefern Gerechtigkeit in Mathematik übersetzt werden kann. Ihre erste Veröffentlichung wurde in die LinkedIn -Suche integriert, um fairere Ergebnisse zu garantieren. Sie ist Gründungsmitglied der europäischen Forschungsgemeinschaft für algorithmische Fairness. Sie verfügt über umfangreiche Lehr - und Beratungserfahrung zu ethischer KI.

Erfahrung und Projekte (Auszug)

Mitgründerin & Geschäftsleitung, Team Passerelle, seit 2025
Senior Researcher, Zalando Research, 2021 -2025
Beraterin ethische KI, IG Metall, Ford Motor GmbH, Dt. Bundestag, DIN u.a. seit 2020
Dozentin & Mentorin ethische KI, TU & HU Berlin, 2017 -2023
Visiting Researcher, CSH Wien, 2022
PhD Researcher, Max-Planck-Institut SWS, 2019-2021
Visiting Researcher, New York University, 2019
Visiting Researcher, UPF Barcelona, 2018
PhD Researcher, TU Berlin, 2016-2019

Bildung

Promotion in der Informatik, Humboldt -Universität zu Berlin, Abschluss Dr. rer. nat. Summa cum laude und Auszeichnung mit dem Humboldtpreis für die beste Dissertation 2022

Studium der Informatik
Technische Universität Dresden
Abschluss Diplom

Agenda

1.

Worüber wir heute
nicht sprechen

2.

Gut gemeint ist nicht
gut gemacht

3.

Faires Matching in der
Plattformarbeit

4.

Recap

Agenda

1.

Worüber wir heute
nicht sprechen

2.

Gut gemeint ist nicht
gut gemacht

3.

Faires Matching in der
Plattformarbeit

4.

Recap

Keine Zeit für...

Schade, Marmelade

1.

KI an sich

Was ist das? Wo gibt's das überall? Wie funktioniert es? Welche unterschiedlichen Technologien gibt es? Was ist ein LLM? Was ist überwachtes Lernen?

2.

Probleme mit KI

Welche Probleme gibt es beim Einsatz von KI? Warum diskriminiert KI? Wie diskriminiert KI? Wer ist besonders betroffen?

3.

Ethik

Welche Konzepte aus der Philosophie / Soziologie / youNameIt kennen? Was ist Ethik? Was ist Fairness?

4.

Recht

Wer ist verantwortlich für Diskriminierung durch KI? Wer entscheidet darüber, wie ein fairer Algorithmus aussieht? Was steht in der KI - Verordnung? Was muss man sonst noch wissen?

Agenda

1. a

Ein gaaaanz kurzer
Exkurs

2.

Gut gemeint ist nicht
gut gemacht

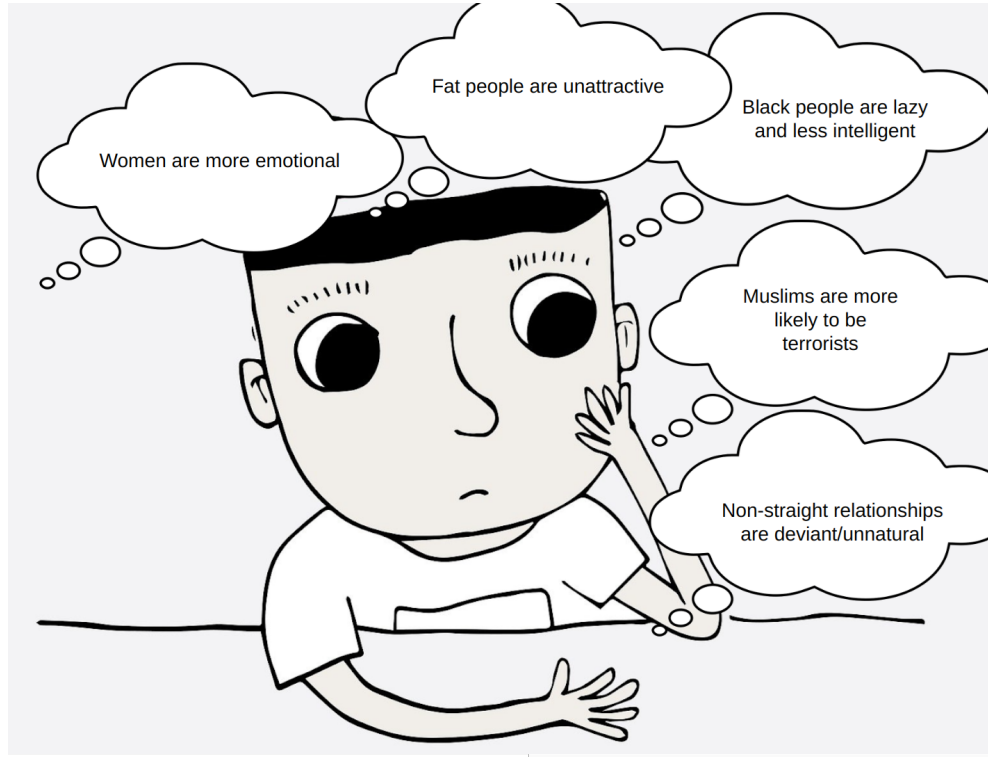
3.

Faires Matching in der
Plattformarbeit

4.

Recap

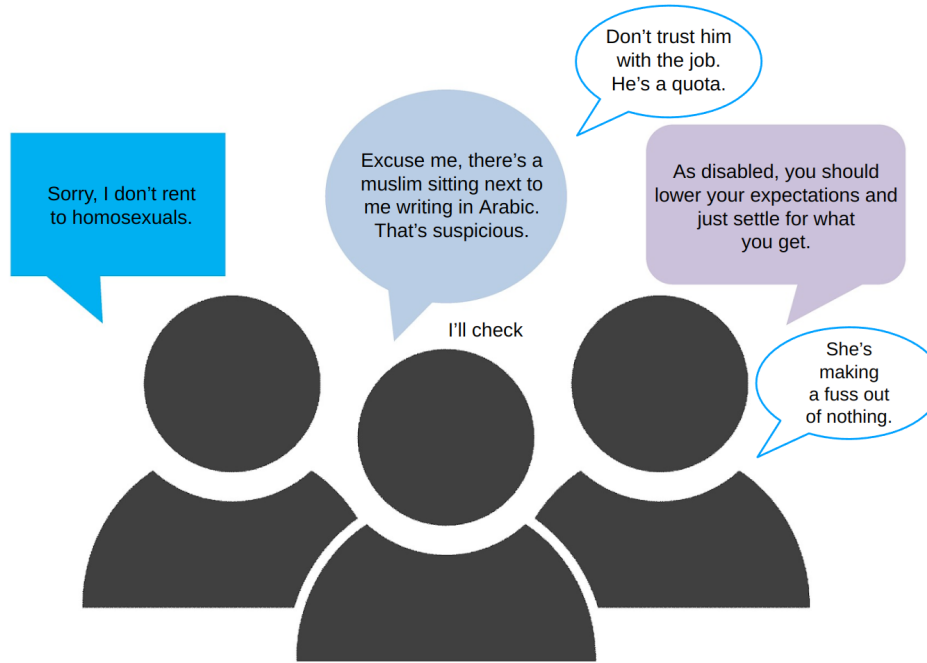
Vorurteil (Prejudice)



Ein Gefühl gegenüber einer Person, das auf deren vermeintlicher Gruppenzugehörigkeit beruht.

[Dank an Catia Faria für dieses Bild]

Diskriminierung (Discrimination)



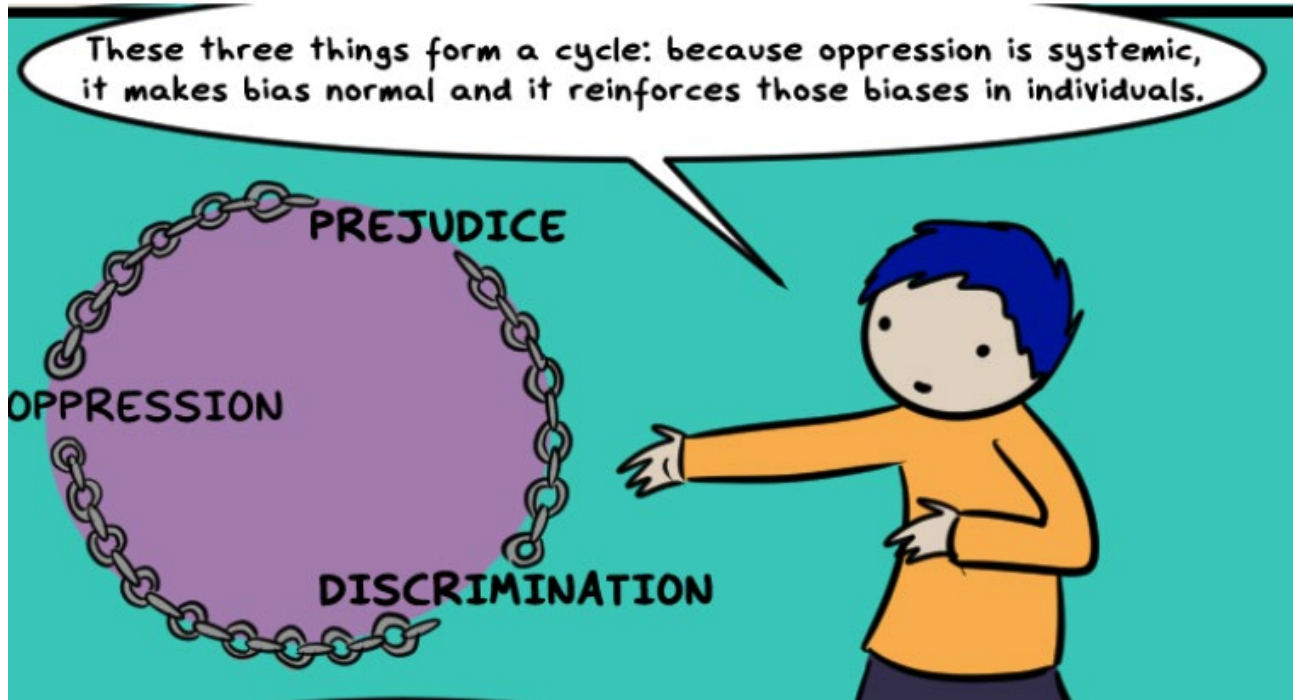
Die Tat, bei der ich mich aufgrund meiner Vorurteile anders gegenüber einer Person verhalte, der ich bestimmte Eigenschaften aufgrund vermeintlicher Gruppenzugehörigkeit zuschreibe

Unterdrückung (Oppression)



Systematische ungerechte
Behandlung oder ungerechte
Machtstrukturen, zB durch
Gesetze oder soziale Normen

Die drei bilden einen Teufelskreis



Agenda

1.

Worüber wir heute
nicht sprechen

2.

Beispiel 1 :
Gut gemeint ist nicht
gut gemacht

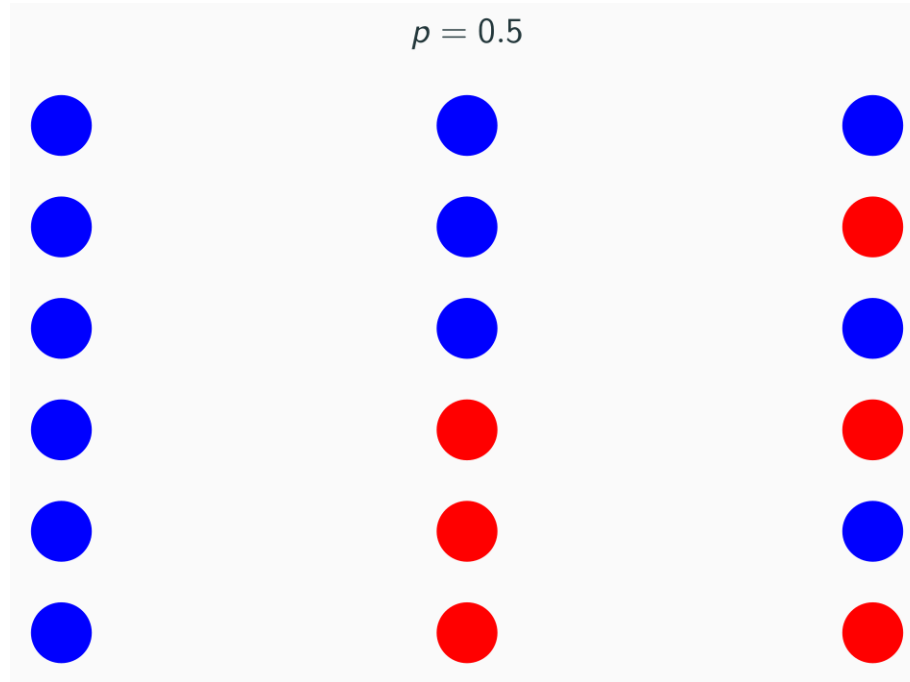
3.

Faires Matching in der
Plattformarbeit

4.

Recap

Faires Ranking



Der FA*IR Algorithmus

- Sortiere alle Gruppen separat nach einem Leistungskriterium
- Berechne die **Mindestanzahl** von geschützten Einträgen an jeder Position anhand einer Quote p_i für jede Gruppe
- Prüfe an jeder Position:
 - **Wenn es bereits ausreichend Elemente jeder Gruppe gibt:** nimm die Nächstbeste von allen Kandidaten
 - **Sonst:** wähle zufällig eine Gruppe aus von allen, die ihr Soll noch nicht erfüllt haben und nimm den besten Kandidaten aus dieser Gruppe

Der DetGreedy Algorithmus (LinkedIn)

- Sortiere alle Gruppen separat nach einem Leistungskriterium
- Berechne die Mindest- und Maximalanzahl für jede Gruppe an jeder Position
- Überprüfe an jeder Position:
 - Finde alle Gruppen, die weder ihr Minimum noch ihr Maximum erreicht haben
 - **Wenn bereits ausreichend geschützte Kandidatinnen gibt:** nimm die Nächstbeste von allen Kandidatinnen
 - **Sonst:** nimm die Nächstbeste von allen, die noch nicht ihr Maximum erreicht haben

FA*IR und DetGreedy im Vergleich

Der User-Input und das Interface sind quasi gleich
(beide bekommen eine Quote pro Gruppe)

Das Re-ranking ist auch sehr ähnlich

DetGreedy:

Vergleicht Kandidaten über
Gruppengrenzen hinweg und **ignoriert**
deshalb das Problem **intersektioneller**
Diskriminierung

FA*IR:

Vergleicht niemals Kandidaten
verschiedener Gruppen und kann
deshalb mit **intersektioneller**
Diskriminierung umgehen

Was geht mich das an?

Jede technische Entscheidung ist
immer auch eine normative

Kleine technische Unterschiede
können große Auswirkungen auf
die ethischen Werte eines Systems
haben

Die ethischen Werte unserer
Systeme sollten zu unserer
Intention passen und nicht
heimlich die Problemlösung
verhindern

Agenda

1.

Worüber wir heute
nicht sprechen

2.

Beispiel 1 :
Gut gemeint ist nicht
gut gemacht

3.

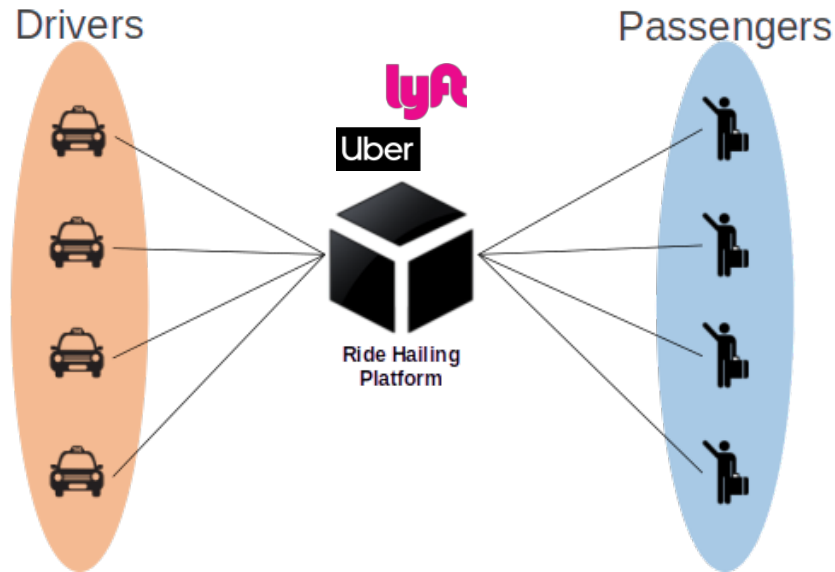
Beispiel 2:
Faires Matching in der
Plattformarbeit

4.

Recap

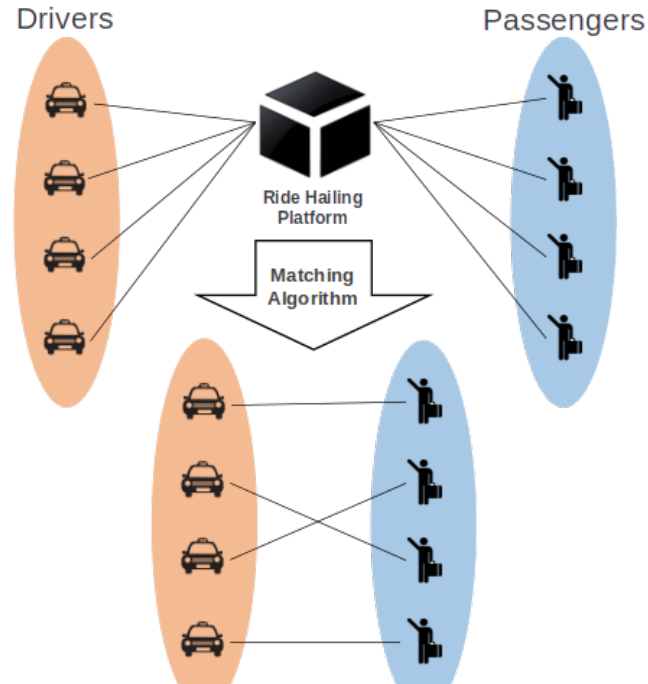
Matching Plattformen

E-Commerce (Amazon), Streaming (Spotify, Netflix), Fahrdienste (Uber, Lyft)



Sühr, T., Biega, A.J., Zehlike, M., Gummadi, K.P. and Chakraborty, A., 2019, July. *Toward fairness for repeated matchings in two-sided markets: A case study of a ride-hailing platform.* In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*

Matching Plattformen Fahrdienste



Sühr, T., Biega, A.J., Zehlike, M., Gummadi, K.P. and Chakraborty, A., 2019, July. Tied fairness for repeated matchings in two-sided markets: A case study of a ride-hailing platform. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*

Matching Plattformen Fahrdienste

Uber, Lyft Drivers Stage Strike To Draw Attention To Drivers' Plight

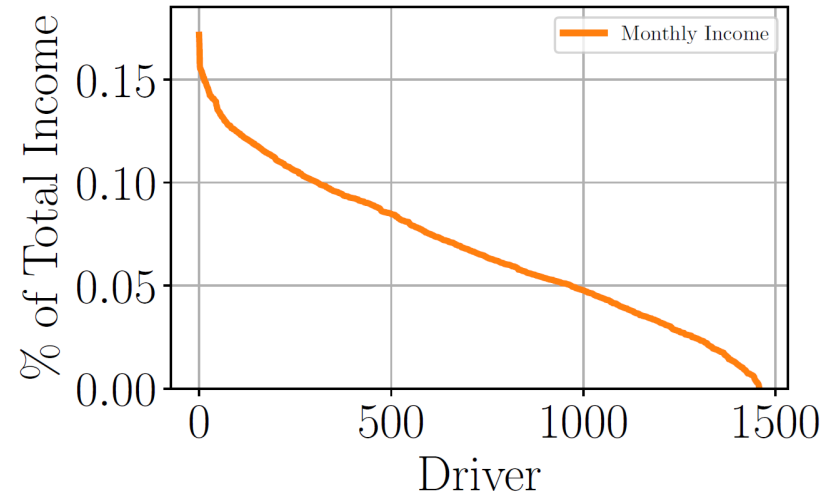
CBS

The Uber drivers forced to sleep in parking lots to make a decent living

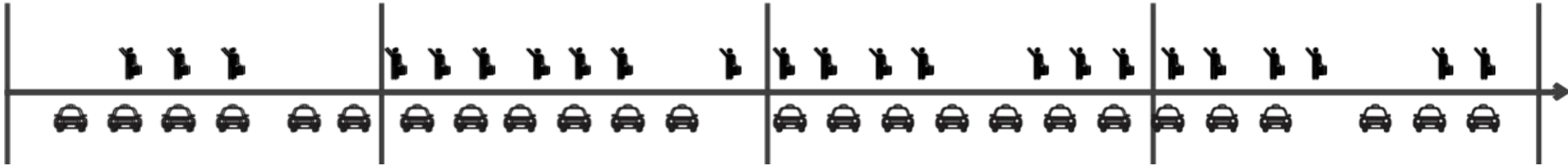
The Guardian

They were sold a fantasy of middle-class life. Now Ola and Uber drivers face a brutal reality

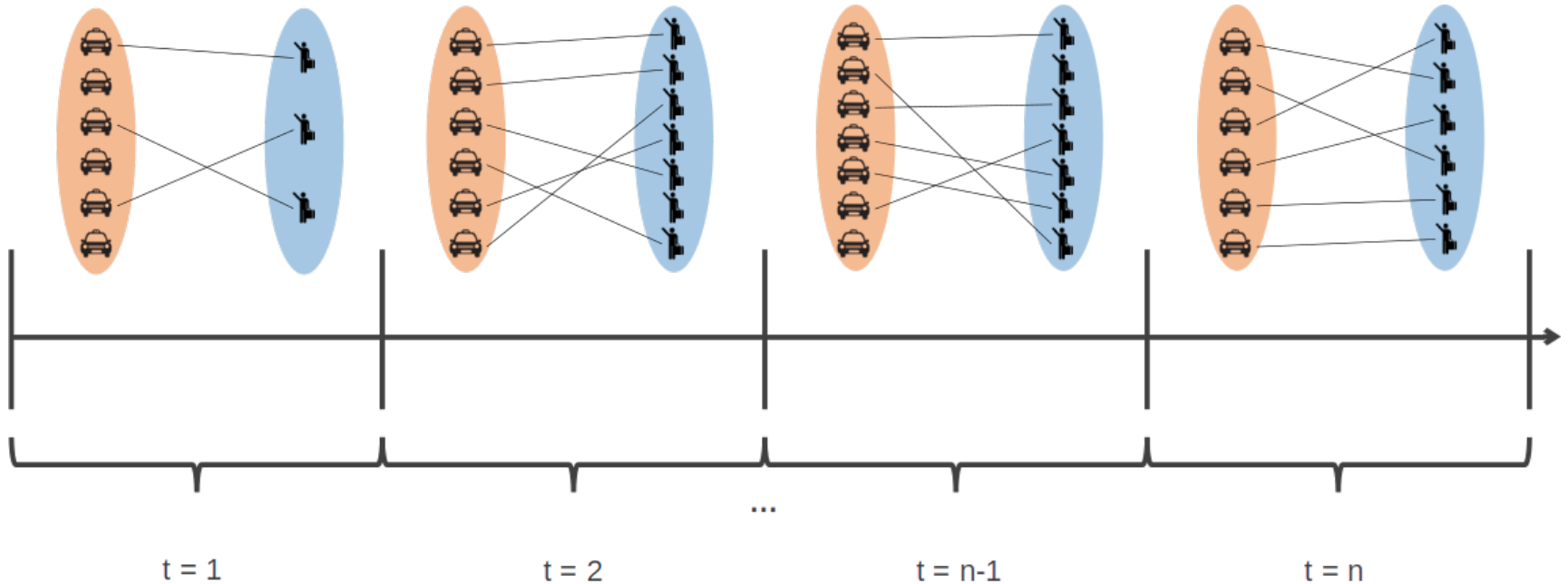
QUARTZ



Essenzielle Eigenschaften



Essenzielle Eigenschaften



Sühr, T., Biega, A.J., Zehlike, M., Gummadi, K.P. and Chakraborty, A., 2019, July. *Toward fairness for repeated matchings in two-sided markets: A case study of a ride-hailing platform*. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*

Wer ist als nächster dran?

- Optimal für die Kunden (möglichst kurze Wartezeiten) – Nearest Driver First (NDF)
- Optimal für die Fahrer (möglichst gleichmäßige Einkommensverteilung) – Worst-off Driver First (WDF)
- Balance auf beiden Seiten, d.h. beide Ziele werden gleichzeitig optimiert

Ergebnisse

- **NDF:** starke **Ungleichheit** bei den Einkommen, manche Fahrer verdienen überhaupt nichts
- **WDF:** Einkommen ist zwar **gleichverteilt**, aber das Durchschnittseinkommen **fällt um >40%** und die Wartezeit **steigt um 400%**
- **Balance -Strategie:** Die Ungleichheit **fällt um ~25%**, das Durchschnittseinkommen **steigt** leicht und die Wartezeit **fällt** leicht

Agenda

1.

Worüber wir heute
nicht sprechen

2.

Beispiel 1 :
Gut gemeint ist nicht
gut gemacht

3.

Beispiel 2:
Faires Matching in der
Plattformarbeit

4.

Recap

Was haben wir gelernt?

- Vorurteile, Diskriminierung und Unterdrückung bilden einen Teufelskreis
- Man kann Ethik in Mathematik übersetzen
- Jede technische Entscheidung hat auch eine normative Komponente
- Kleine technische Entscheidungen können große Auswirkungen auf die ethischen Werte des Systems haben
- Fairness und Effizienz sind keine Gegensätze

Danke!