

Diskriminierungsgefahr durch KI am Arbeitsmarkt

KNOW-CENTER GMBH
RESEARCH CENTER FOR TRUSTWORTHY AI
AND DATA,
GRAZ - AUSTRIA

Dr. Dominik Kowald

Wissenschaftsforum zu Köln und Essen

14.03.2024



Know-Center GmbH



COMET-Zentrum



Gegründet 2001



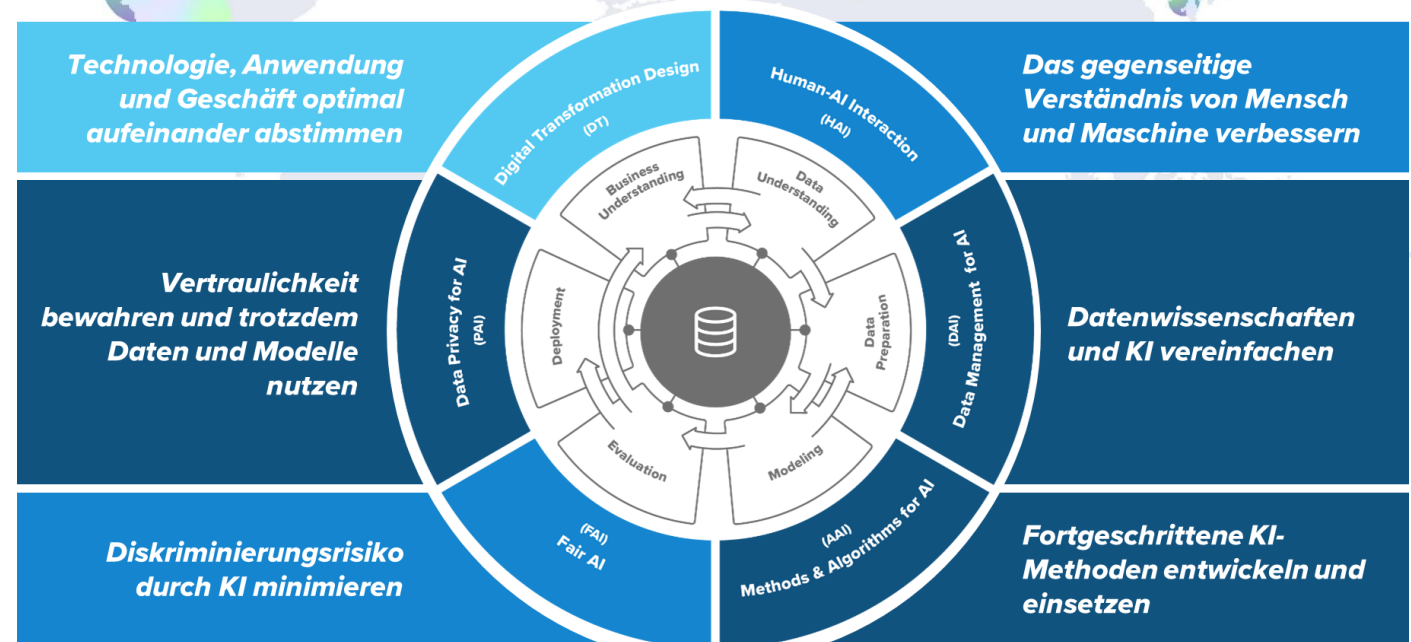
100+ Mitarbeiter:innen



TU Graz, Campus Inffeldgasse



„Wir forschen, entwickeln und beraten entlang der Datenwertschöpfungskette zum Thema **vertrauenswürdige KI** und **Data Science**.“



Know-Center GmbH



COMET-Zentrum



Gegründet 2001



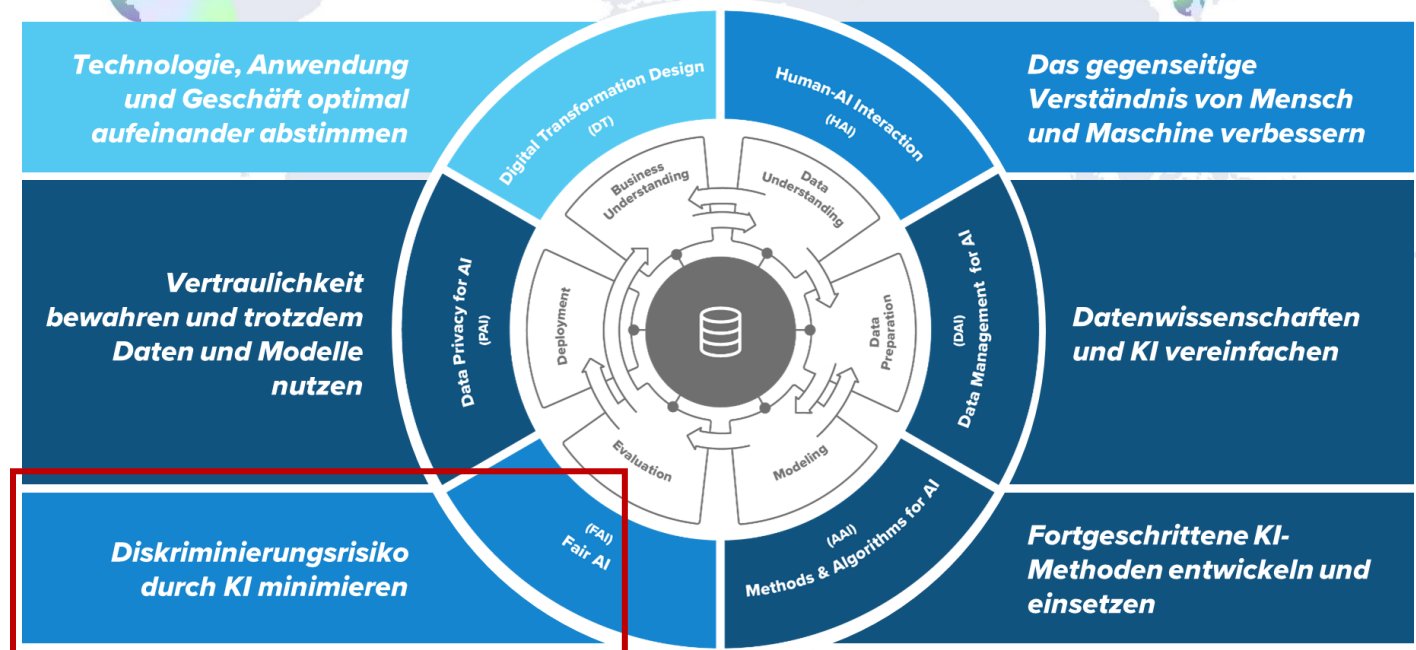
100+ Mitarbeiter:innen



TU Graz, Campus Inffeldgasse



„Wir forschen, entwickeln und beraten entlang der Datenwertschöpfungskette zum Thema **vertrauenswürdige KI** und **Data Science**.“



Diskriminierung durch KI am Arbeitsmarkt



Negative Auswirkungen durch KI

[https://incidentdatabase.ai/]

- AI Incidents Datenbank
 - > 2000 Vorfälle
 - ~ 78 für Arbeitsmarkt

The screenshot shows a search interface for 'hiring' on the Incident Database AI website. The search bar contains the word 'hiring'. Below the search bar, there are controls for 'Display Option' (set to 'Incident Reports'), '78 results found', and 'Sort by' (set to 'Relevance'). There are also buttons for 'Export', 'Clear Filters', and 'Filter Search'. The search results are displayed in a grid of eight cards, each with a thumbnail image, a title, a source, and a date. The cards are:

- The Death and Life of an Admissions Algorithm** (insidehighered.com · 2020) - #135
- Fired by Bot at Amazon: 'It's You Against the Machine'** (bloomberg.com · 2021) - #111
- The Christchurch shooter and YouTube's radicalization trap** (wired.com · 2020) - #89
- Why Facebook is losing the war on hate speech in Myanmar** (reuters.com · 2018) - #169
- Amazon Shuts Down AI Hiring Tool for Being Sexist** (globalcitizen.org · 2018)
- Who's a CEO? Google image results can shift gender biases** (washington.edu · 2015)
- The Apple Card algo issue: What you need to know about A.I. in everyday life** (cnbc.com · 2019)
- How Southwest Airlines Melted Down** (wsj.com · 2022)

Negative Auswirkungen durch KI

[https://incidentdatabase.ai/]

- AI Incidents Datenbank
 - > 2000 Vorfälle
 - ~ 78 für Arbeitsmarkt

The screenshot shows a search interface for 'hiring' on the Incident Database AI website. The search bar contains the word 'hiring'. Below the search bar, there are controls for 'Display Option' (set to 'Incident Reports'), '78 results found', and 'Sort by' (set to 'Relevance'). There are also buttons for 'Export', 'Clear Filters', and 'Filter Search'. The search results are displayed as a grid of eight incident cards. Each card features a thumbnail image, a title, a source, and a date. The cards are: 1. 'The Death and Life of an Admissions Algorithm' (insidehighered.com, 2020), 2. 'Fired by Bot at Amazon: 'It's You Against the Machine'' (bloomberg.com, 2021), 3. 'The Christchurch shooter and YouTube's radicalization trap' (wired.com, 2020), 4. 'Why Facebook is losing the war on hate speech in Myanmar' (reuters.com, 2018), 5. 'Amazon Shuts Down AI Hiring Tool for Being Sexist' (globalcitizen.org, 2018), 6. 'Who's a CEO? Google image results can shift gender biases' (washington.edu, 2015), 7. 'The Apple Card algo issue: What you need to know about A.I. in everyday life' (cnbc.com, 2019), and 8. 'How Southwest Airlines Melted Down' (wsj.com, 2022). A red arrow points from the text 'AI Incidents Datenbank' towards the search results.

Negative Auswirkungen durch KI

[https://incidentdatabase.ai/]

- **AI Incidents Datenbank**
 - > 2000 Vorfälle
 - ~ 78 für Arbeitsmarkt

- **Wieso?**
 - Biases in Daten
 - Unklare Fairness Definitionen
 - Arbeitsmarkt ist Hochrisikoanwendung laut EU AI Act



The screenshot shows a search interface for 'hiring' with 78 results found. The search bar contains 'hiring'. Below the search bar, there are options for 'Display Option' (Incident Reports), 'Export', and 'Sort by' (Relevance). There are also links for 'Clear Filters' and 'Filter Search'. The results are displayed in a grid of cards, each with a thumbnail image, a title, a source, and a date. The cards include:

- The Death and Life of an Admissions Algorithm** (insidehighered.com · 2020)
- Fired by Bot at Amazon: 'It's You Against the Machine'** (bloomberg.com · 2021)
- The Christchurch shooter and YouTube's radicalization trap** (wired.com · 2020)
- Why Facebook is losing the war on hate speech in Myanmar** (reuters.com · 2018)
- Amazon Shuts Down AI Hiring Tool for Being Sexist** (globalcitizen.org · 2018)
- Who's a CEO? Google image results can shift gender biases** (washington.edu · 2015)
- The Apple Card algo issue: What you need to know about A.I. in everyday life** (cnbc.com · 2019)
- How Southwest Airlines Melted Down** (wsj.com · 2022)

Beispiele für KI Anwendungen am Arbeitsmarkt

- **KI-basierte Job Vorschlagsysteme**
 - **Job-Kandidaten für Jobs vorschlagen (siehe Amazon)**
 - **Jobs für Job-Interessierte vorschlagen (z.B. LinkedIn)**

- **KI-basierte Entscheidungsunterstützungssysteme**
 - **Für öffentliche Arbeitsvermittlungsdienste (z.B. Arbeitsmarktservice in Österreich)**
 - **Arbeitslose Menschen bei der Jobsuche unterstützen (z.B. durch Schulungen)**

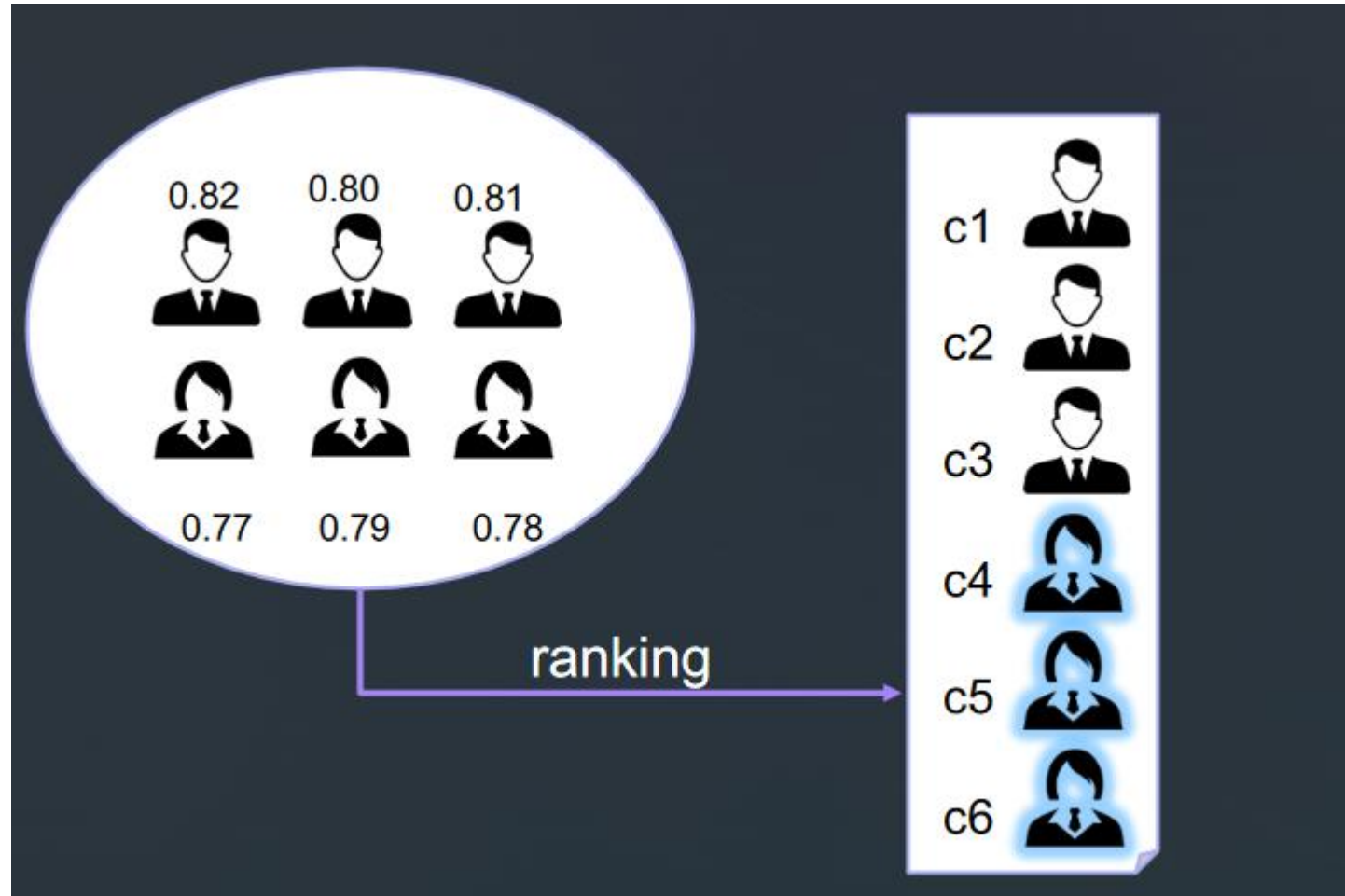
Beispiele für KI Anwendungen am Arbeitsmarkt

- **KI-basierte Job Vorschlagsysteme**
 - Job-Kandidaten für Jobs vorschlagen (siehe Amazon)
 - Jobs für Job-Interessierte vorschlagen (z.B. LinkedIn)
 - Ranking-Problem in KI
- **KI-basierte Entscheidungsunterstützungssysteme**
 - Für öffentliche Arbeitsvermittlungsdienste (z.B. Arbeitsmarktservice in Österreich)
 - Arbeitslose Menschen bei der Jobsuche unterstützen (z.B. durch Schulungen)
 - Klassifizierungs-Problem in KI

[Scher, S., Kopeinik, S., Truegler, A., & Kowald, D. (2023). Modelling the Long-Term Fairness Dynamics of Data-Driven Targeted Help on Job Seekers. Nature Scientific Reports]

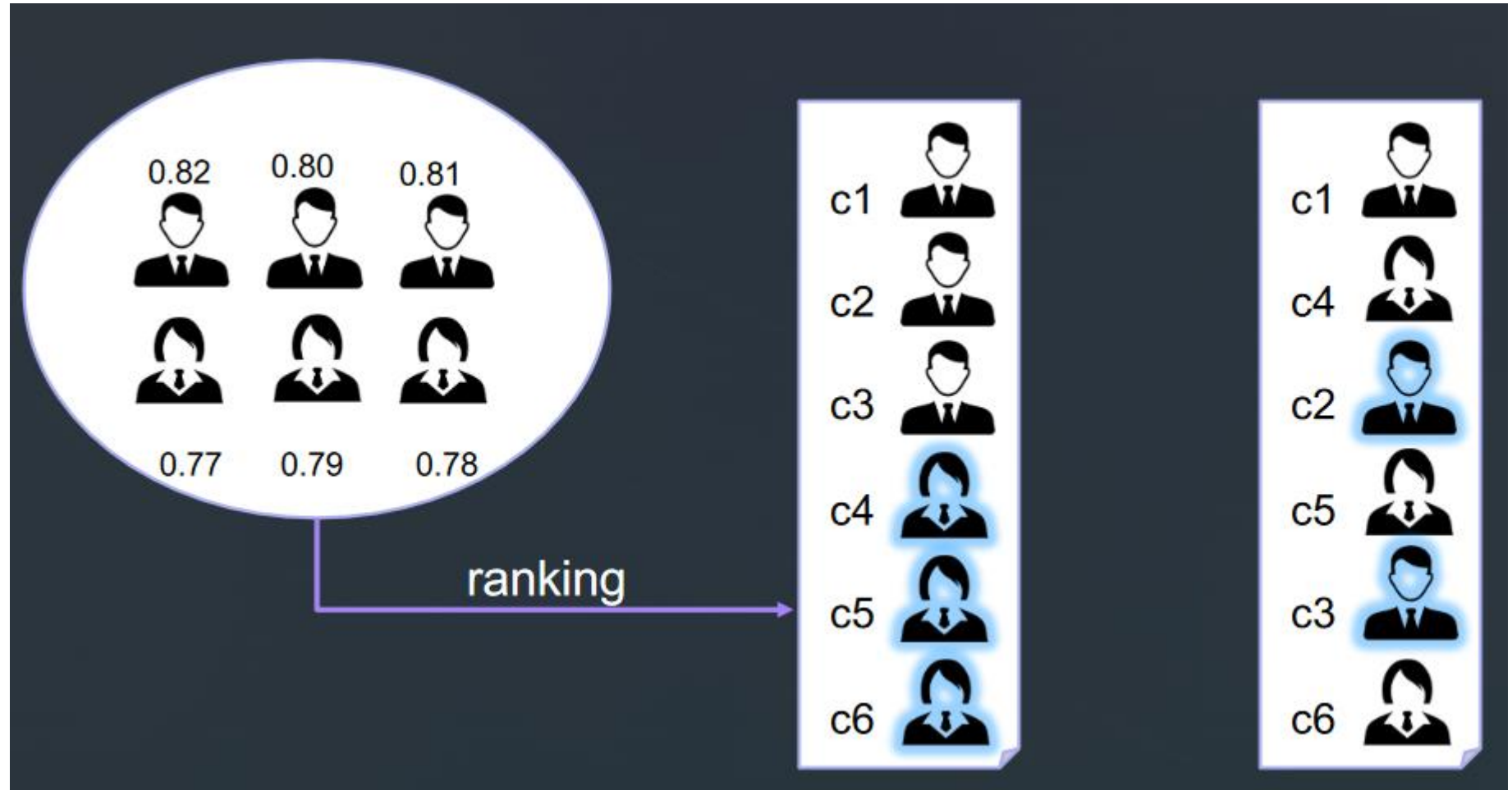
[Lacic, E., Reiter-Haas, M., Kowald, D., Dareddy, M., Cho, J., & Lex, E. (2020). Using Autoencoders for Session-based Job Recommendations. User Modeling and User-Adapted Interaction]

KI-basierte Job Vorschlagsysteme



[Gao, R., & Shah, C. (2020).
Counteracting bias and
increasing fairness in search
and recommender systems. In
Proceedings of the 14th ACM
RecSys Conference]

KI-basierte Job Vorschlagsysteme



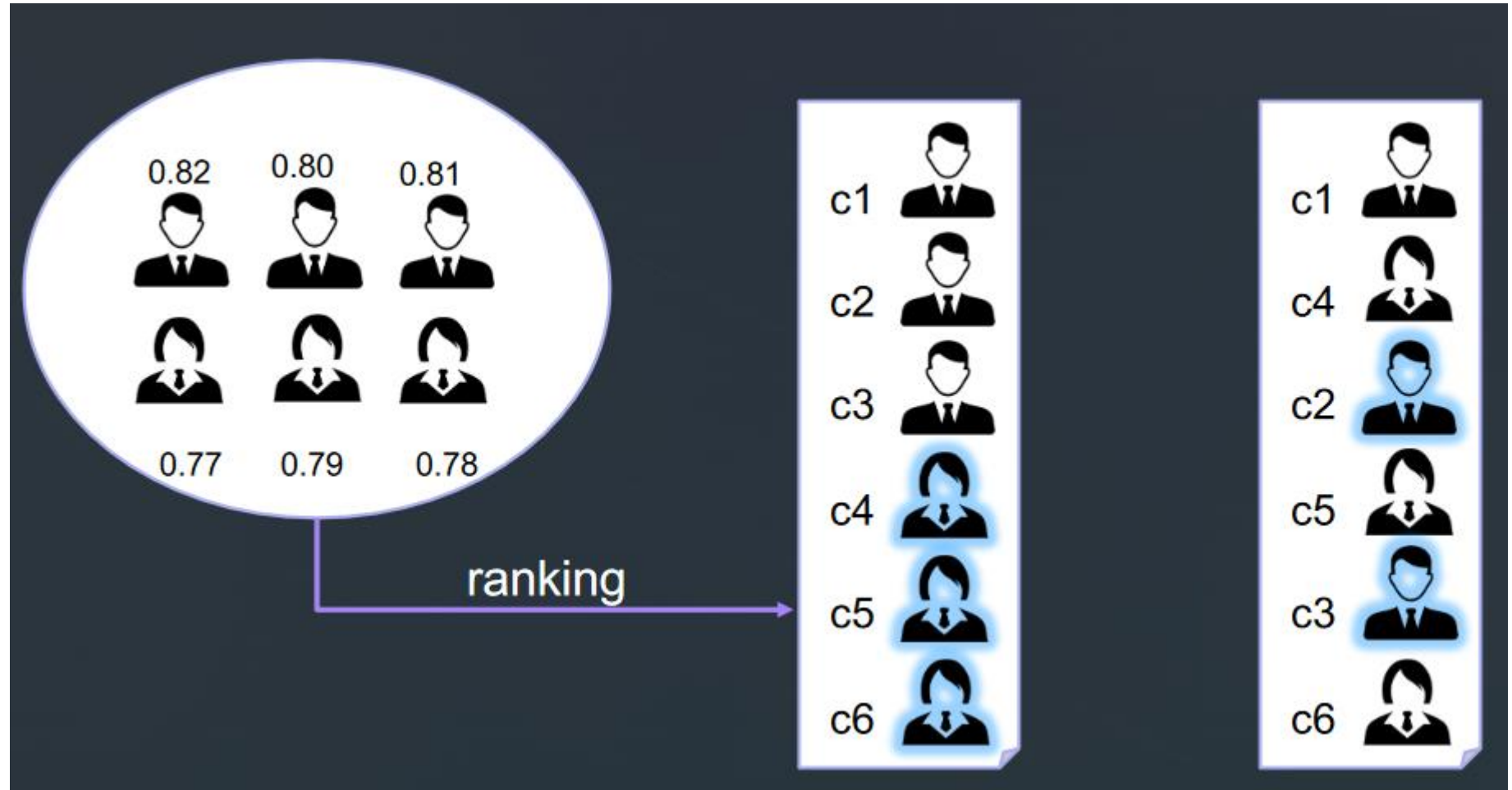
[Gao, R., & Shah, C. (2020).
Counteracting bias and
increasing fairness in search
and recommender systems. In
Proceedings of the 14th ACM
RecSys Conference]

KI-basierte Job Vorschlagsysteme

Gruppen-
Fairness

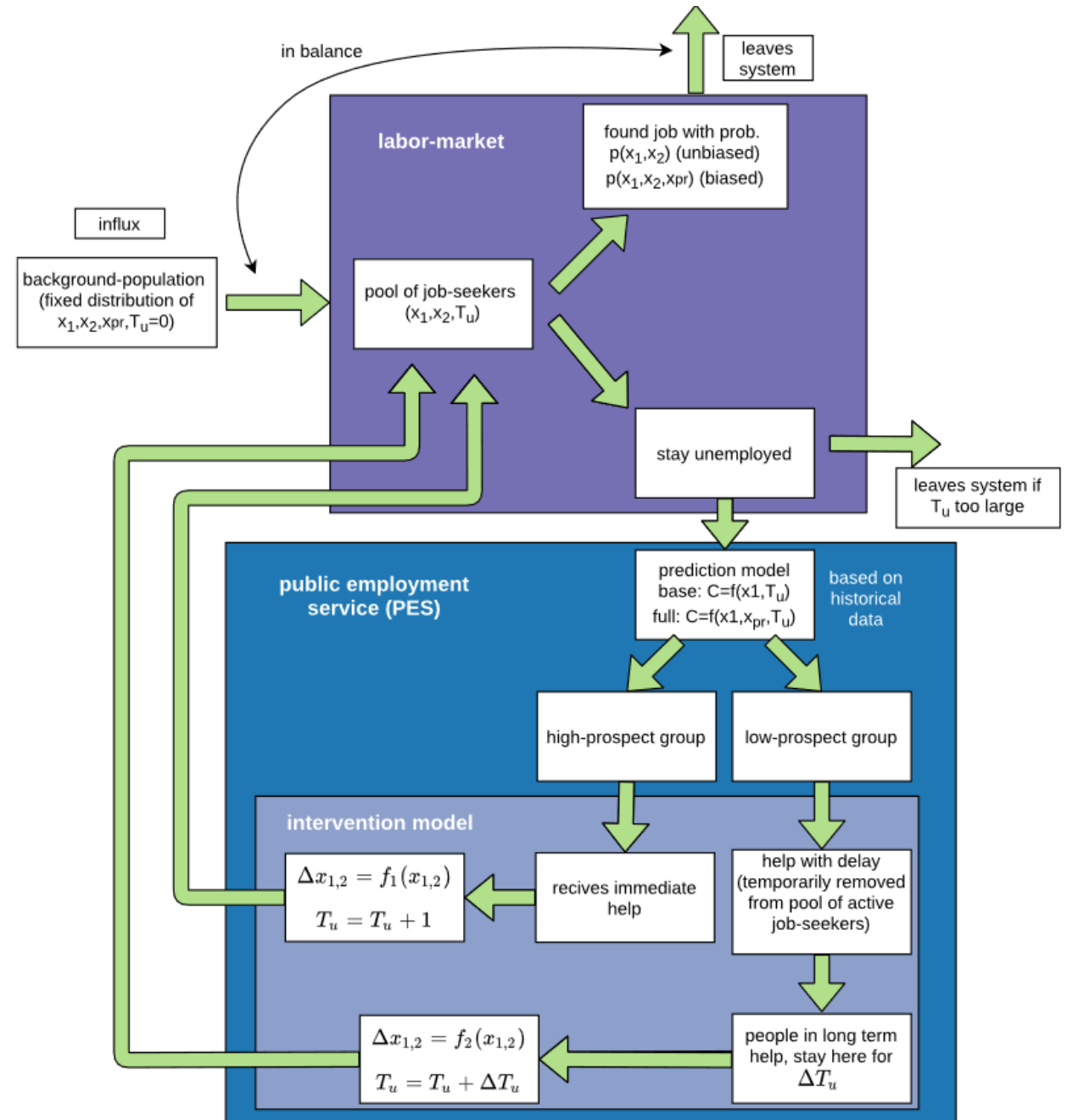
versus

Individuelle
Fairness



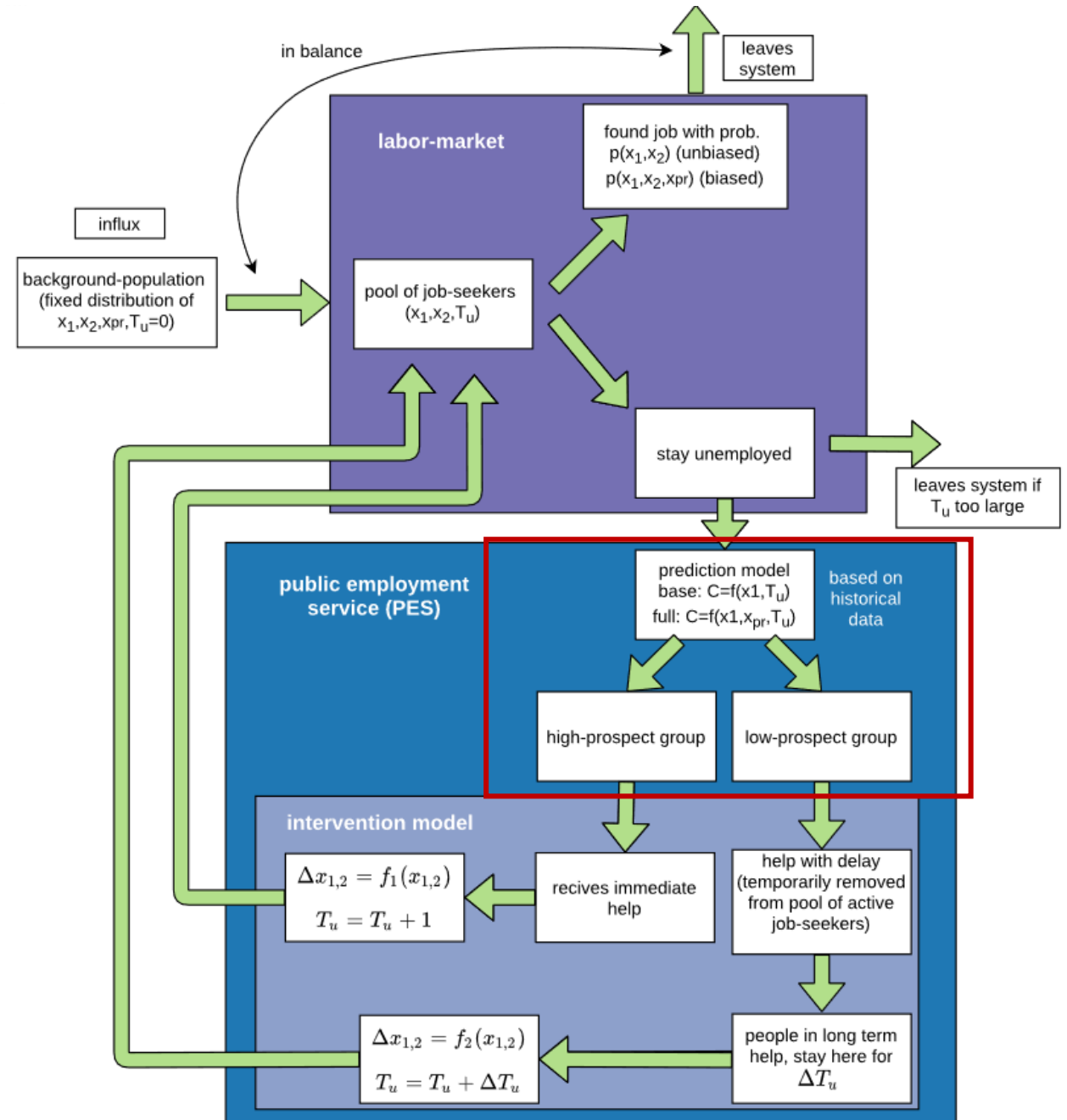
[Gao, R., & Shah, C. (2020). Counteracting bias and increasing fairness in search and recommender systems. In Proceedings of the 14th ACM RecSys Conference]

KI-basierte Entscheidungsunterstützungssysteme



[Scher, S., Kopeinik, S., Truegler, A., & Kowald, D. (2023). Modelling the Long-Term Fairness Dynamics of Data-Driven Targeted Help on Job Seekers. Nature Scientific Reports]

KI-basierte Entscheidungsunterstützungssysteme



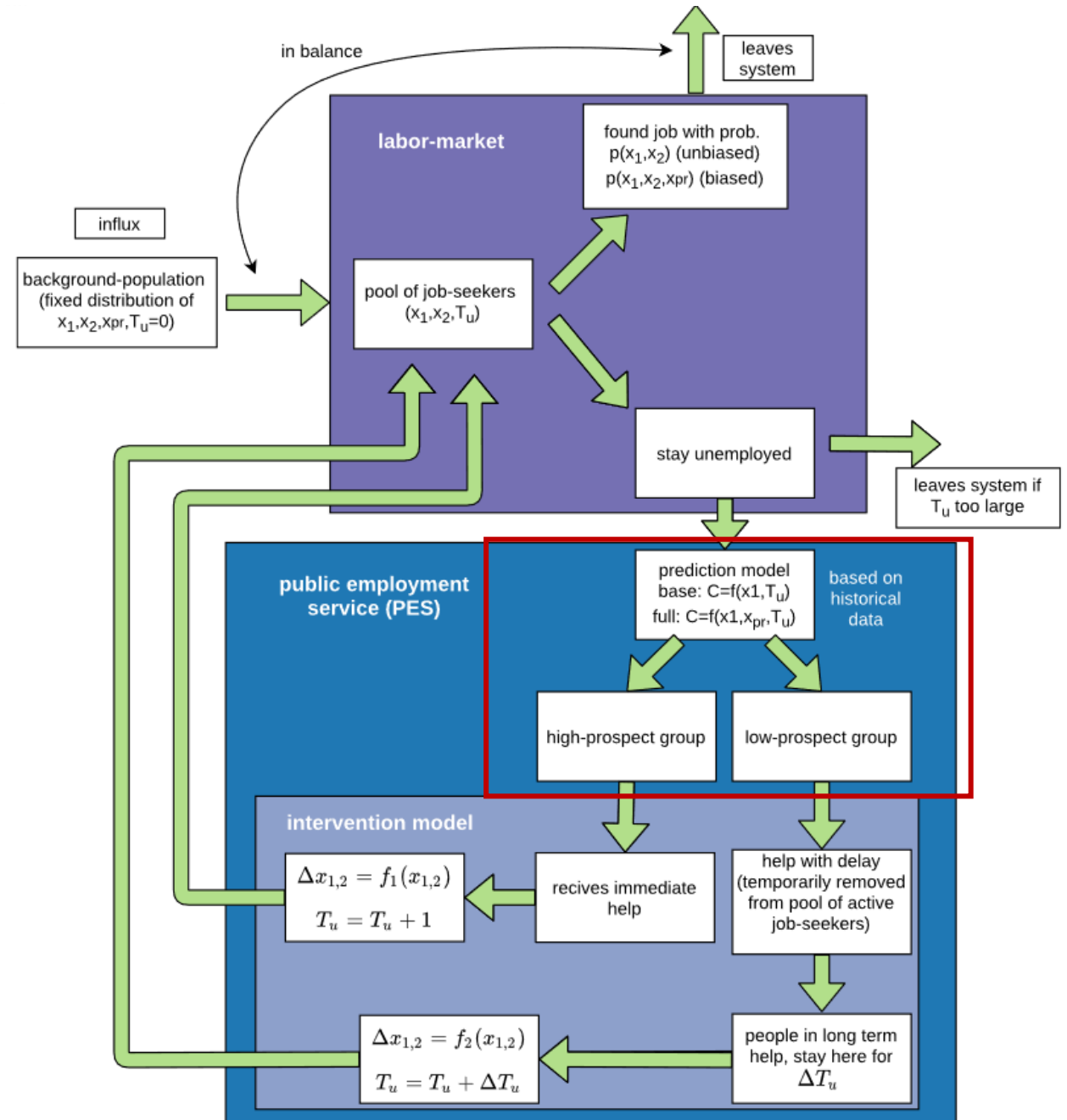
[Scher, S., Kopeinik, S., Truegler, A., & Kowald, D. (2023). Modelling the Long-Term Fairness Dynamics of Data-Driven Targeted Help on Job Seekers. Nature Scientific Reports]

KI-basierte Entscheidungsunterstützungssysteme

KI Modell ist fair hinsichtlich Gruppen → Falsche Klassifizierung von manchen Individuen

... und umgekehrt

[Scher, S., Kopeinik, S., Truegler, A., & Kowald, D. (2023). Modelling the Long-Term Fairness Dynamics of Data-Driven Targeted Help on Job Seekers. Nature Scientific Reports]



KI-basierte Entscheidungsunterstützungssysteme

Gruppen-Fairness

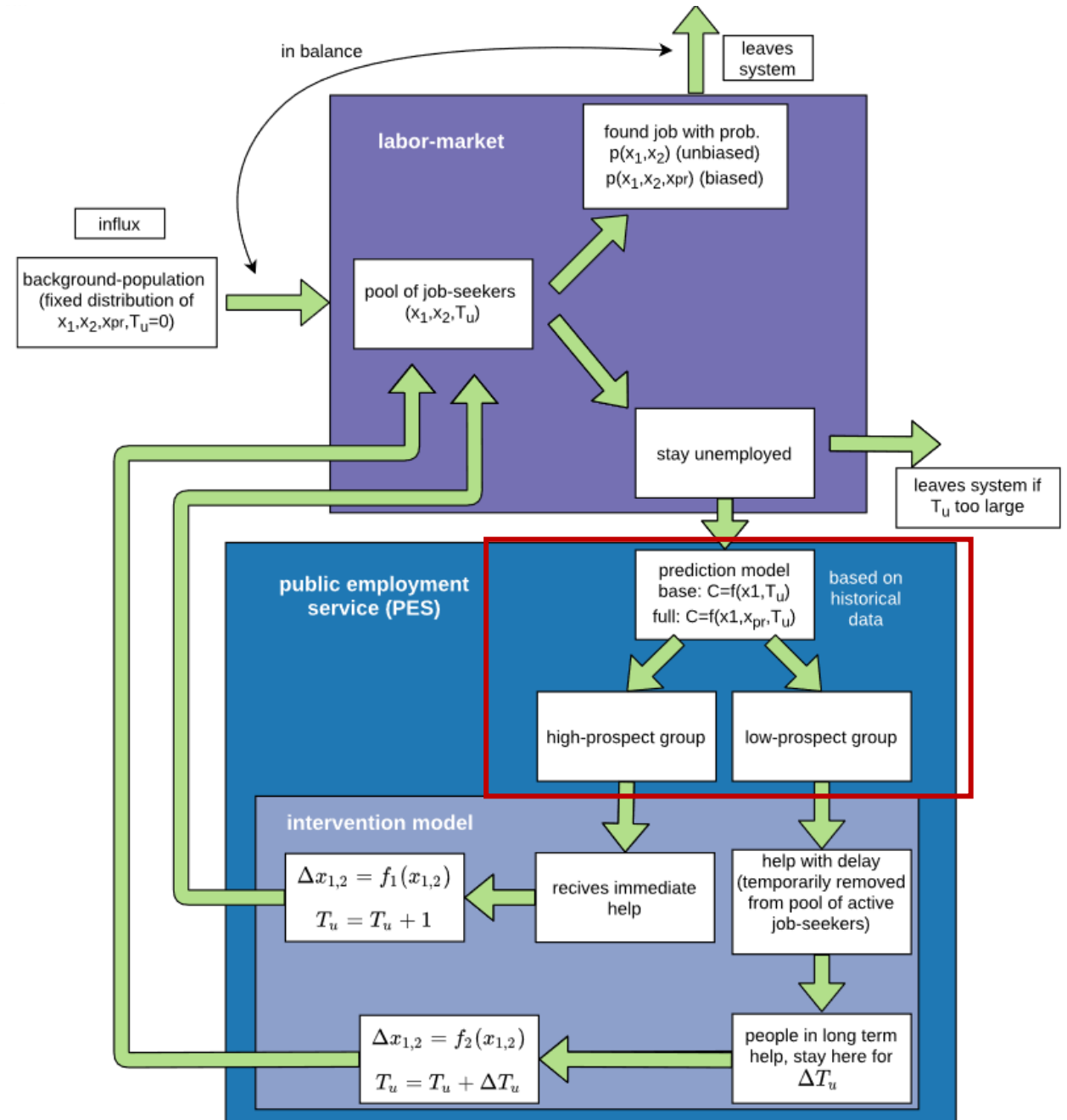
versus

Individuelle Fairness

KI Modell ist fair hinsichtlich Gruppen → Falsche Klassifizierung von manchen Individuen

... und umgekehrt


[Scher, S., Kopeinik, S., Truegler, A., & Kowald, D. (2023). Modelling the Long-Term Fairness Dynamics of Data-Driven Targeted Help on Job Seekers. Nature Scientific Reports]




Zusammenfassung

- **Chancen von KI am Arbeitsmarkt**
 - **Minimierung von menschlichen Biases**
 - **Skalierung**

Zusammenfassung

- **Chancen von KI am Arbeitsmarkt**
 - **Minimierung von menschlichen Biases**
 - **Skalierung**
 - **Risiken von KI am Arbeitsmarkt**
 - **KI Modelle trainieren auf historischen Daten, die auch Biases widerspiegeln**
 - **KI Modelle können Ungleichheiten replizieren, manchmal sogar verstärken**
- 
- A decorative horizontal bar at the bottom right of the slide, consisting of four colored segments: blue, green, teal, and purple.

Zusammenfassung

- **Chancen von KI am Arbeitsmarkt**
 - Minimierung von menschlichen Biases
 - Skalierung
 - **Risiken von KI am Arbeitsmarkt**
 - KI Modelle trainieren auf historischen Daten, die auch Biases widerspiegeln
 - KI Modelle können Ungleichheiten replizieren, manchmal sogar verstärken
 - **Wir brauchen mehr Forschung**
 - Gruppen Fairness versus Individuelle Fairness
 - Ganzheitliche Betrachtung von vertrauenswürdiger KI → Regulierung / Zertifizierung von KI
- 
- A decorative horizontal bar at the bottom right of the slide, composed of four colored segments: blue, green, teal, and purple.

Vielen Dank!

FRAGEN / KOMMENTARE?



Dr. Dominik Kowald

dkowald@know-center.at

<http://www.dominikkowald.info/>

KNOW-CENTER GMBH
Research Center for Data-Driven
Business and Big Data Analytics

Sandgasse 36/4, 8010 Graz, Austria
Firmenbuchgericht Graz
FN 199 685 f
UID: ATU 50367703
www.know-center.at