

# Diskriminierungsgefahr durch KI am Arbeitsmarkt

KNOW CENTER RESEARCH GMBH

INSTITUTE OF INTERACTIVE SYSTEMS AND  
DATASCIENCE – TU GRAZ

GRAZ - AUSTRIA

---

Priv.Doiz. Dr.techn. Dipl.Ing. Dominik Kowald

ÖGGF Graz, Session: New Work

19.09.2024



# Know Center Research GmbH



COMET-Zentrum



Gegründet 2001



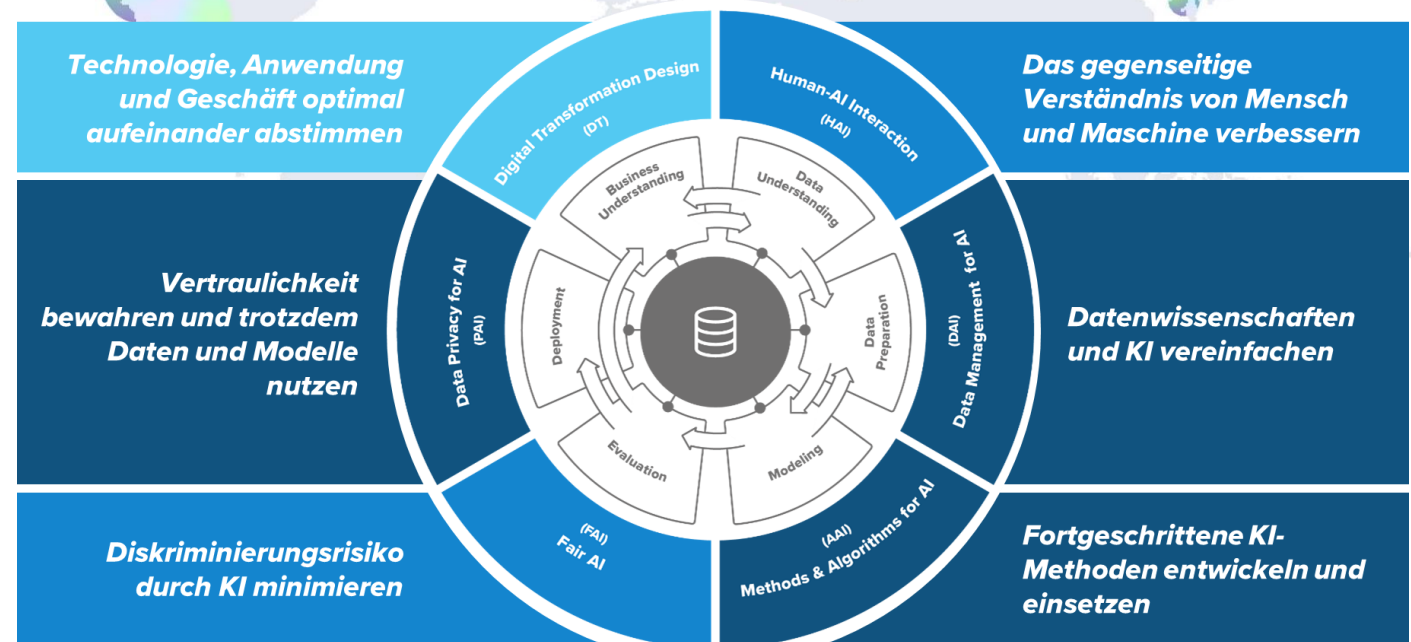
100+ Mitarbeiter:innen



TU Graz, Campus Inffeldgasse



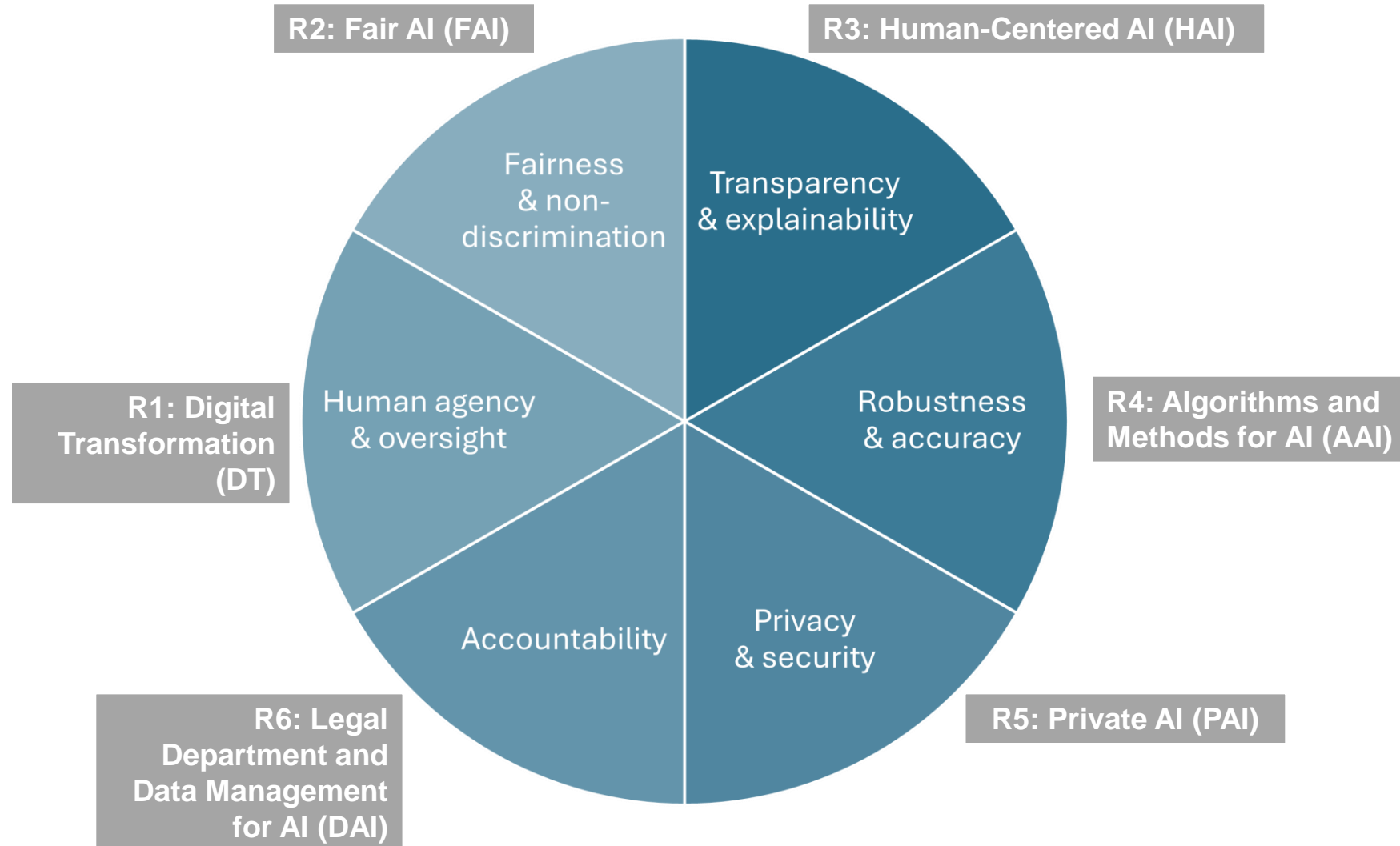
„Wir forschen, entwickeln und beraten entlang der Datenwertschöpfungskette zum Thema **vertrauenswürdige KI** und **Data Science**.“



# Trustworthy AI

Viele ähnliche Definitionen, zB:

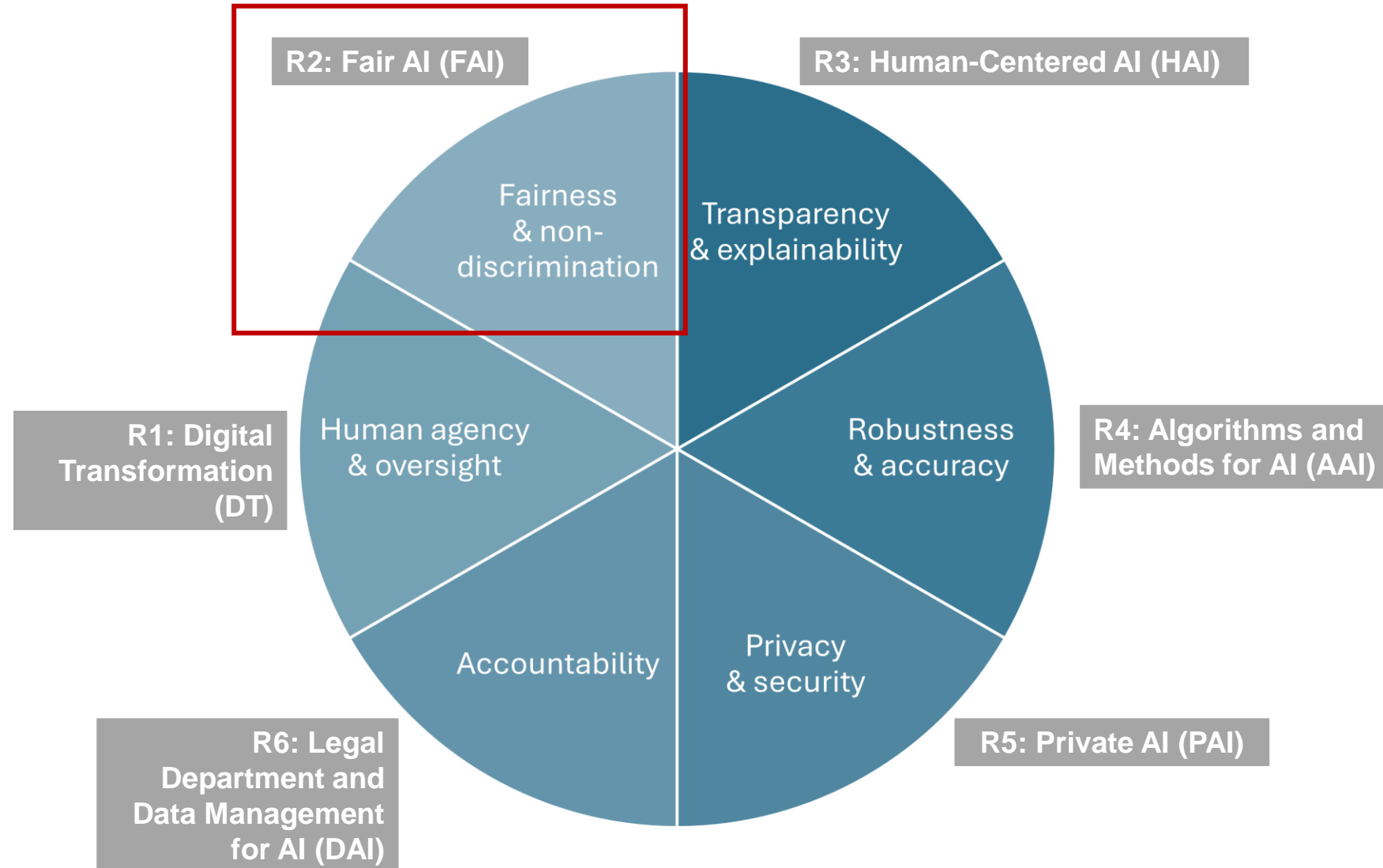
1. Kaur, D., Uslu, S., Rittichier, K. J., & Durresi, A. (2022). Trustworthy artificial intelligence: a review. *ACM computing surveys (CSUR)*
2. Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., ... & Zhou, B. (2023). Trustworthy ai: From principles to practices. *ACM Computing Surveys*, 55(9), *IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*
3. OECD AI Principles (Focuses on inclusivity, transparency, and accountability)
4. UNESCO Recommendation on the Ethics of Artificial Intelligence (Framework for ethical AI development & governance, human rights)
5. Google AI Principles (fairness, accountability, privacy, and safety in AI)
6. Microsoft AI Principles (fairness, reliability, safety, privacy, inclusiveness, transparency, and accountability)
7. European Commission: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>



# Trustworthy AI

Viele ähnliche Definitionen, zB:

1. Kaur, D., Uslu, S., Rittichier, K. J., & Durresi, A. (2022). Trustworthy artificial intelligence: a review. *ACM computing surveys (CSUR)*
2. Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., ... & Zhou, B. (2023). Trustworthy ai: From principles to practices. *ACM Computing Surveys*, 55(9), *IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*
3. OECD AI Principles (Focuses on inclusivity, transparency, and accountability)
4. UNESCO Recommendation on the Ethics of Artificial Intelligence (Framework for ethical AI development & governance, human rights)
5. Google AI Principles (fairness, accountability, privacy, and safety in AI)
6. Microsoft AI Principles (fairness, reliability, safety, privacy, inclusiveness, transparency, and accountability)
7. European Commission: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>





# Diskriminierung durch KI am Arbeitsmarkt





# Negative Auswirkungen durch KI

[<https://incidentdatabase.ai/>]

- AI Incidents Datenbank
  - > 2000 Vorfälle
  - ~ 80 für Arbeitsmarkt

The screenshot shows the Incident Database AI search results for the query 'hiring'. The search bar contains 'hiring' and shows 78 results found. The interface includes a 'Display Option' dropdown set to 'Incident Reports', an 'Export' button, and a 'Sort by' dropdown set to 'Relevance'. There are also links for 'Clear Filters' and 'Filter Search'. The results are displayed as a grid of eight cards, each with a thumbnail image, a title, a source, and a date. The cards are: 1. 'The Death and Life of an Admissions Algorithm' from insidehighered.com (2020), 135 incidents. 2. 'Fired by Bot at Amazon: 'It's You Against the Machine'' from bloomberg.com (2021), 111 incidents. 3. 'The Christchurch shooter and YouTube's radicalization trap' from wired.com (2020), 89 incidents. 4. 'Why Facebook is losing the war on hate speech in Myanmar' from reuters.com (2018), 169 incidents. 5. 'Amazon Shuts Down AI Hiring Tool for Being Sexist' from globalcitizen.org (2018), 135 incidents. 6. 'Who's a CEO? Google image results can shift gender biases' from washington.edu (2015), 111 incidents. 7. 'The Apple Card algo issue: What you need to know about A.I. in everyday life' from cnbc.com (2019), 89 incidents. 8. 'How Southwest Airlines Melted Down' from wsj.com (2022), 169 incidents.

# Negative Auswirkungen durch KI

[https://incidentdatabase.ai/]

- AI Incidents Datenbank
  - > 2000 Vorfälle
  - ~ 80 für Arbeitsmarkt

The screenshot shows a search interface for 'hiring' with 78 results found. The search bar contains the word '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. A red arrow points to the first card in the second row.

Thumbnail	Title	Source	Date
	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



# Negative Auswirkungen durch KI

[https://incidentdatabase.ai/]

- **AI Incidents Datenbank**
  - > 2000 Vorfälle
  - ~ 80 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 results are displayed as a grid of incident cards. Each card includes 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) - #111
- Who's a CEO? Google image results can shift gender biases** (washington.edu · 2015) - #111
- The Apple Card algo issue: What you need to know about A.I. in everyday life** (cnbc.com · 2019) - #89
- How Southwest Airlines Melted Down** (wsj.com · 2022) - #169



## 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)

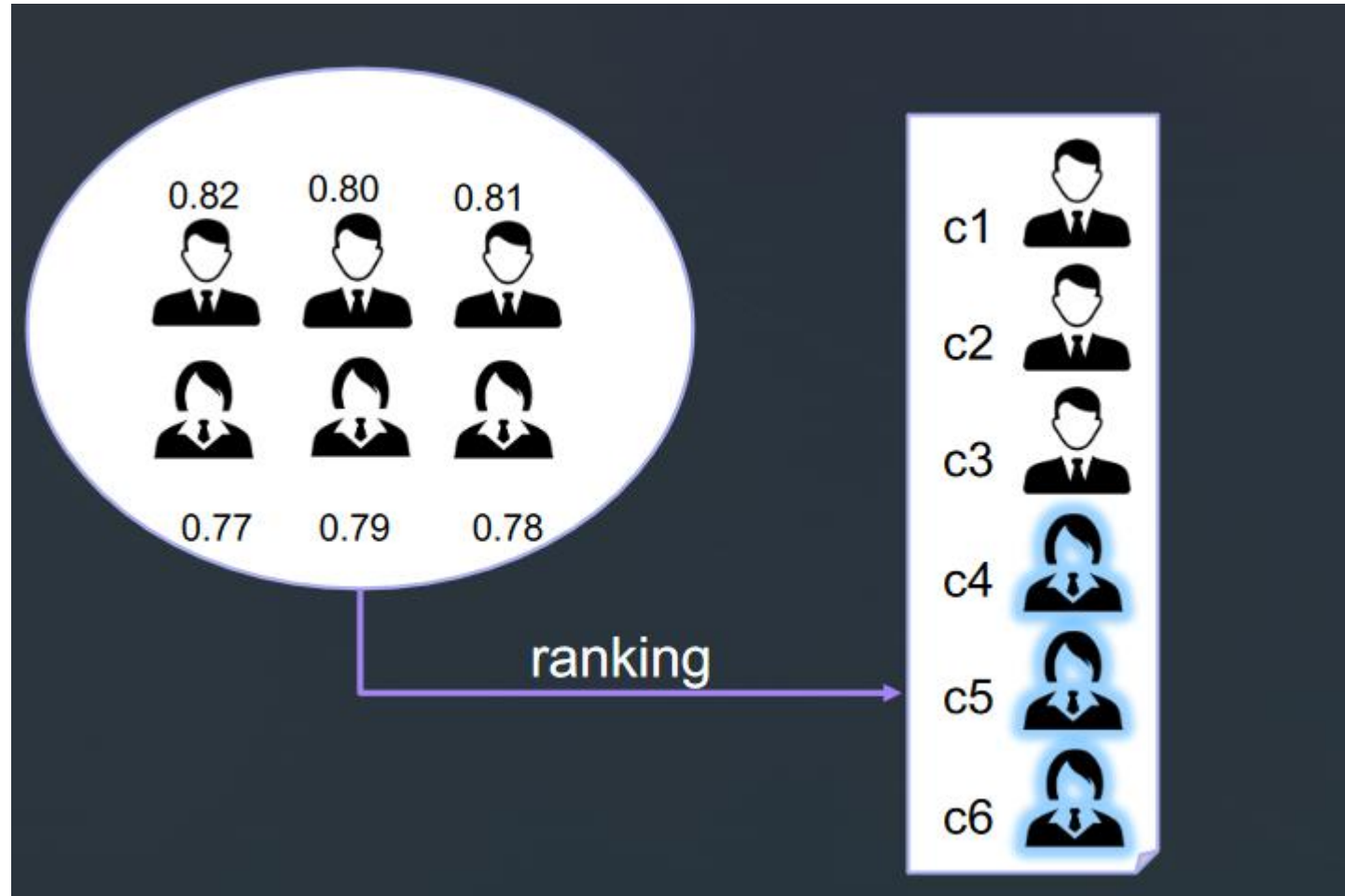
# Beispiele für KI Anwendungen am Arbeitsmarkt

- **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]

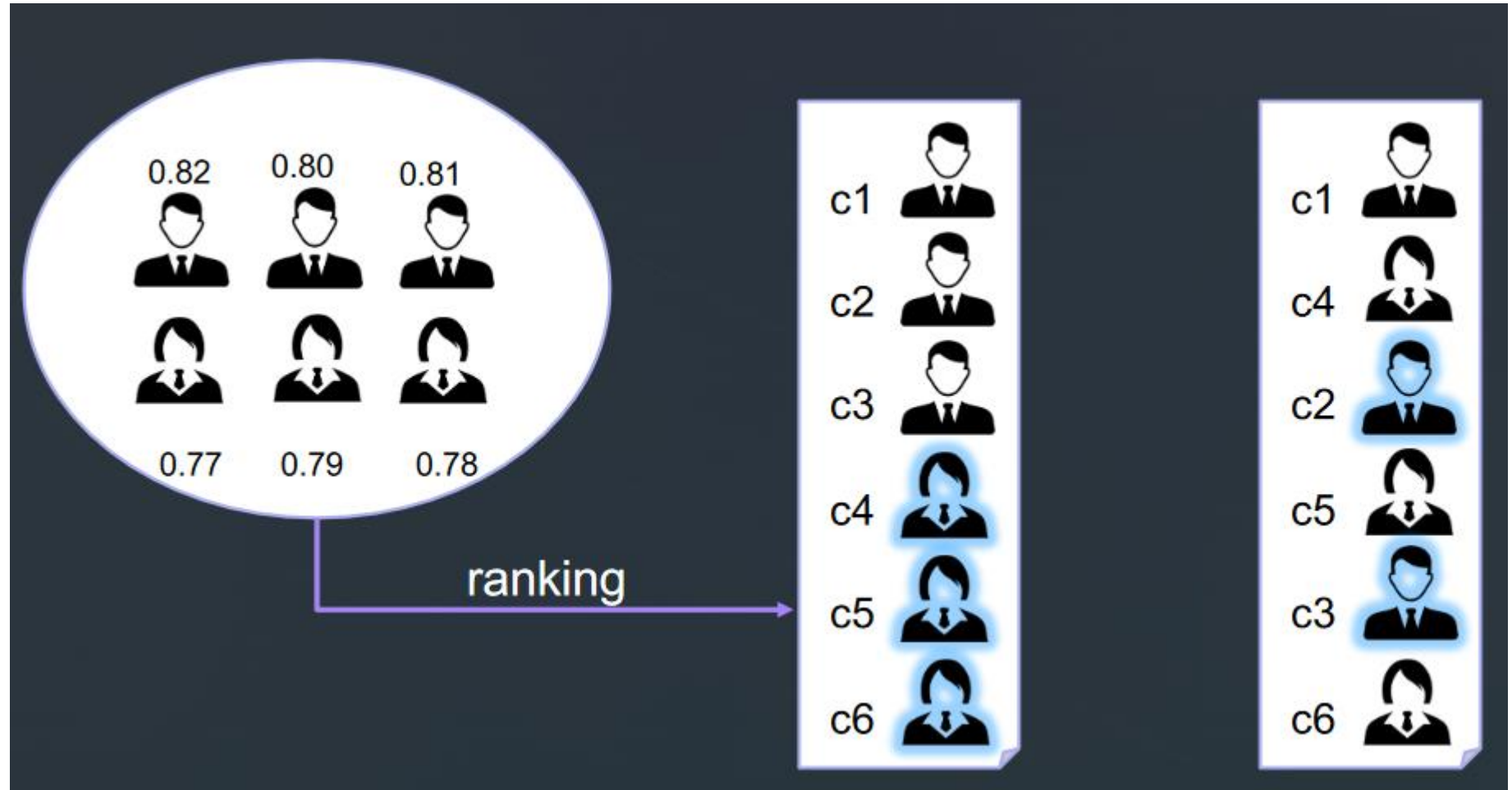
## KI-basierte Job Vorschlagsysteme



[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



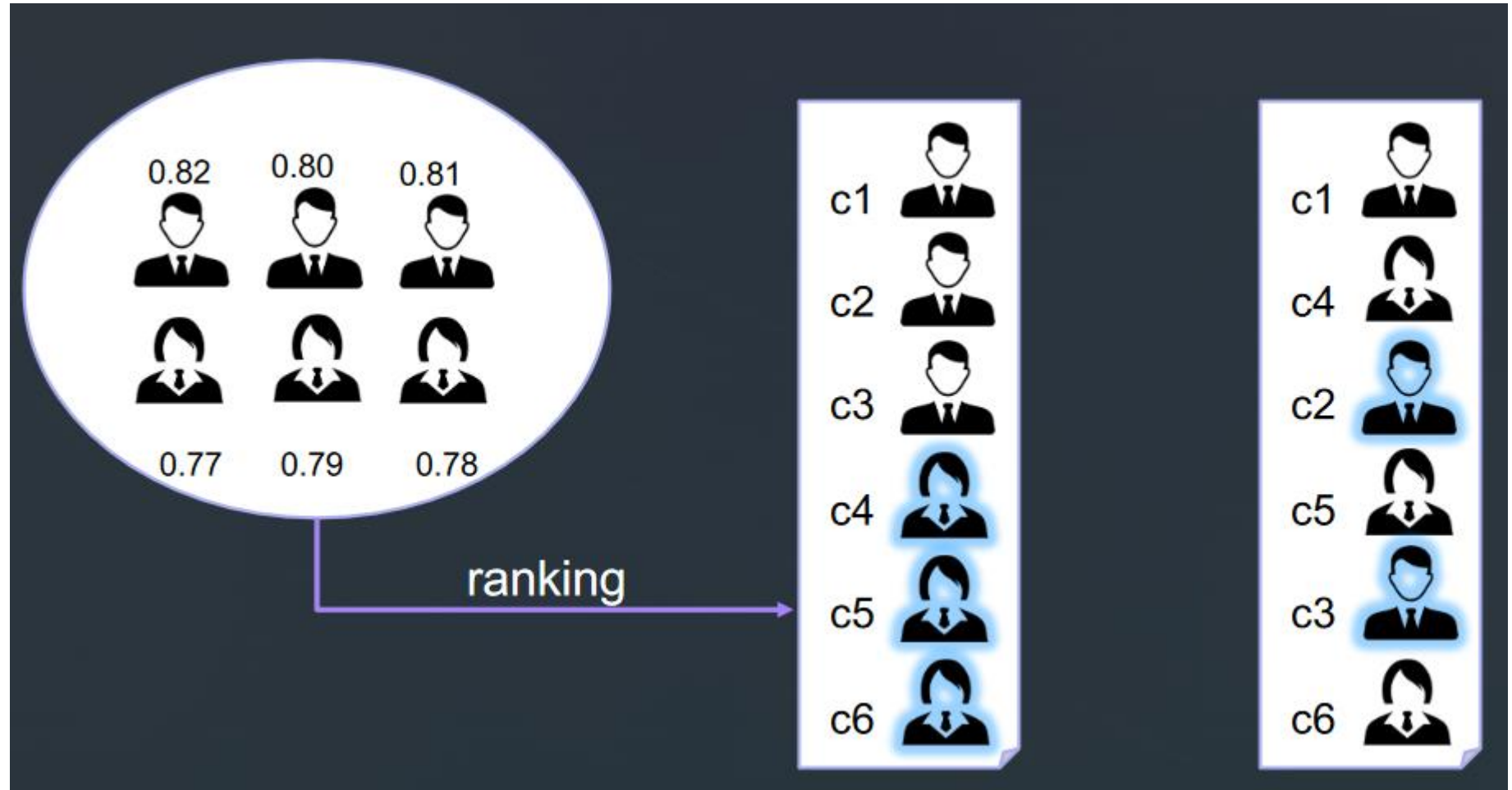
[Gao, R., & Shah, C. (2020). Counteracting bias and increasing fairness in search and recommender systems. In Proceedings of the 14th ACM RecSys Conference]

# KI-basierte Job Vorschlagsysteme

Gruppen-  
Fairness

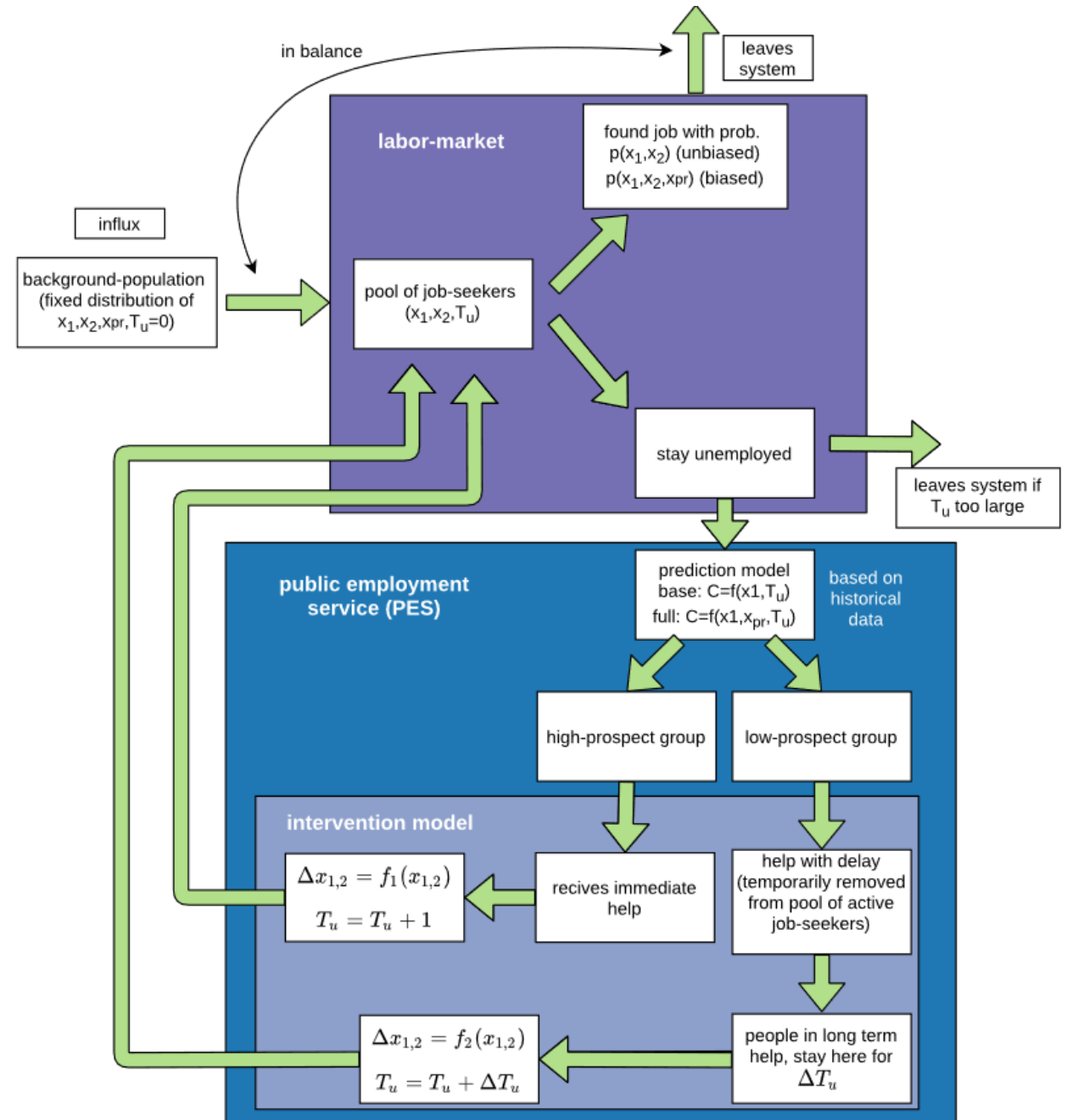
versus

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]

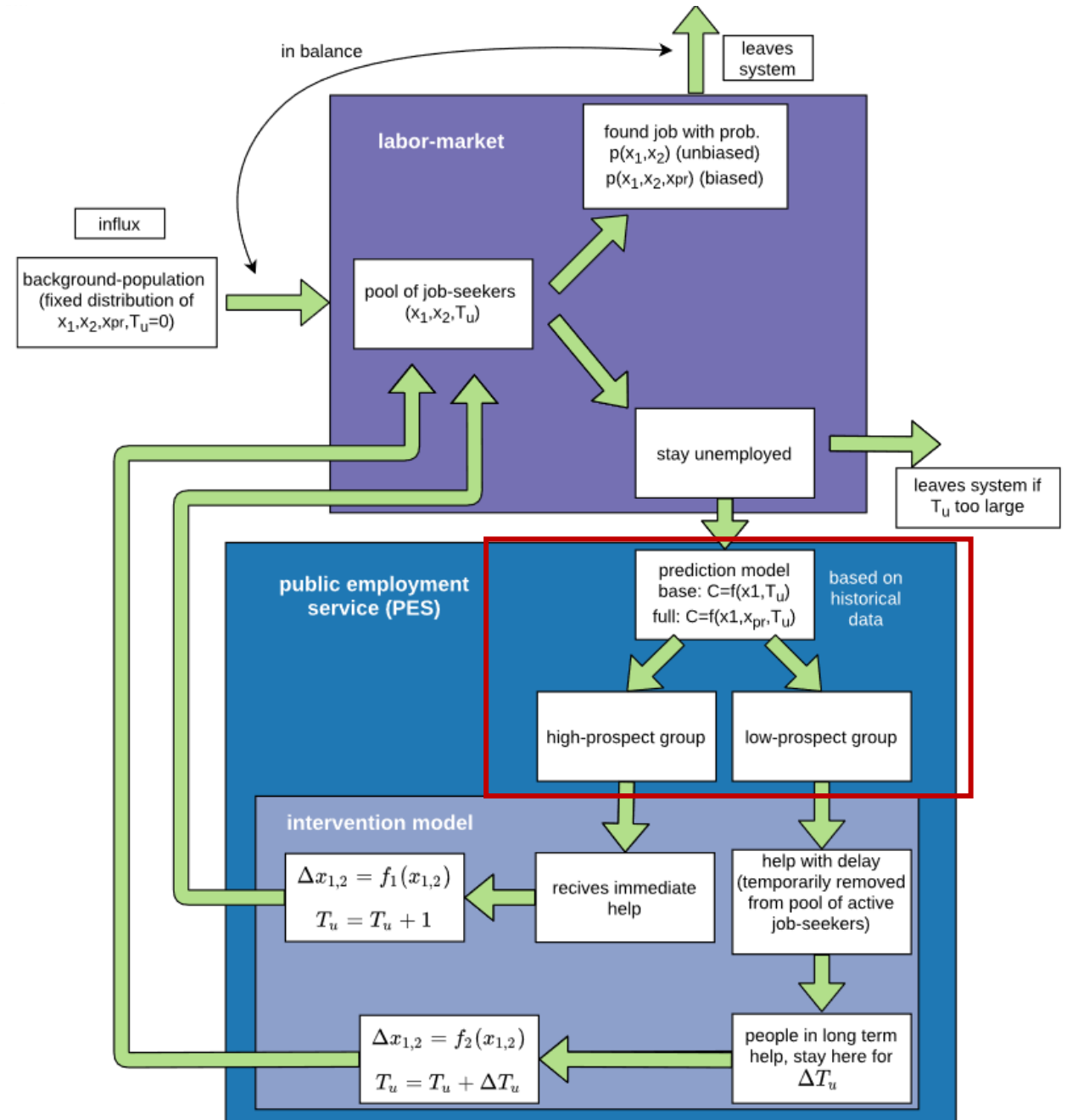
# KI-basierte Entscheidungsunterstützungssysteme



[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



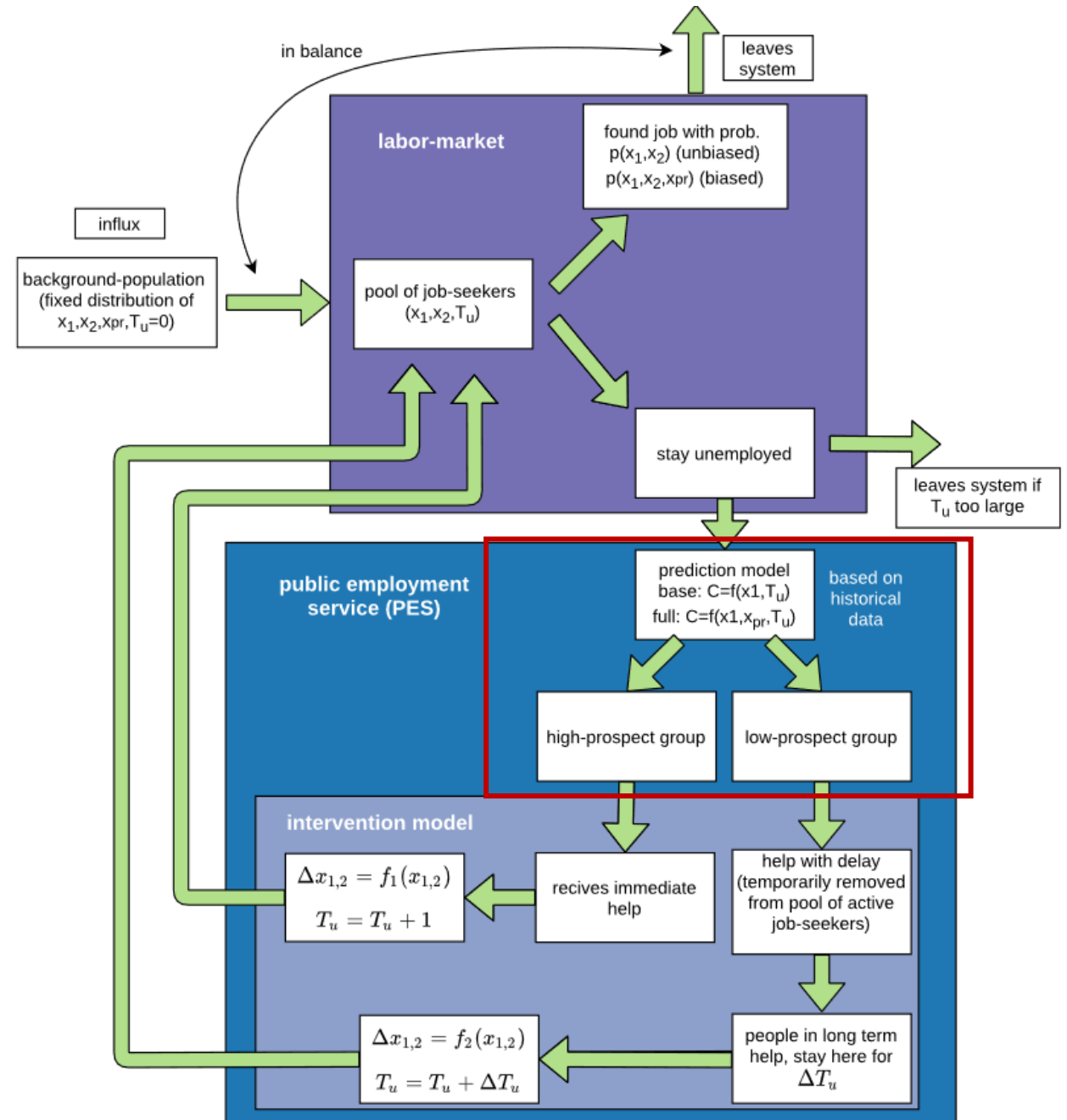
[Scher, S., Kopeinik, S., Truegler, A., & Kowald, D. (2023). Modelling the Long-Term Fairness Dynamics of Data-Driven Targeted Help on Job Seekers. Nature Scientific Reports]

# 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

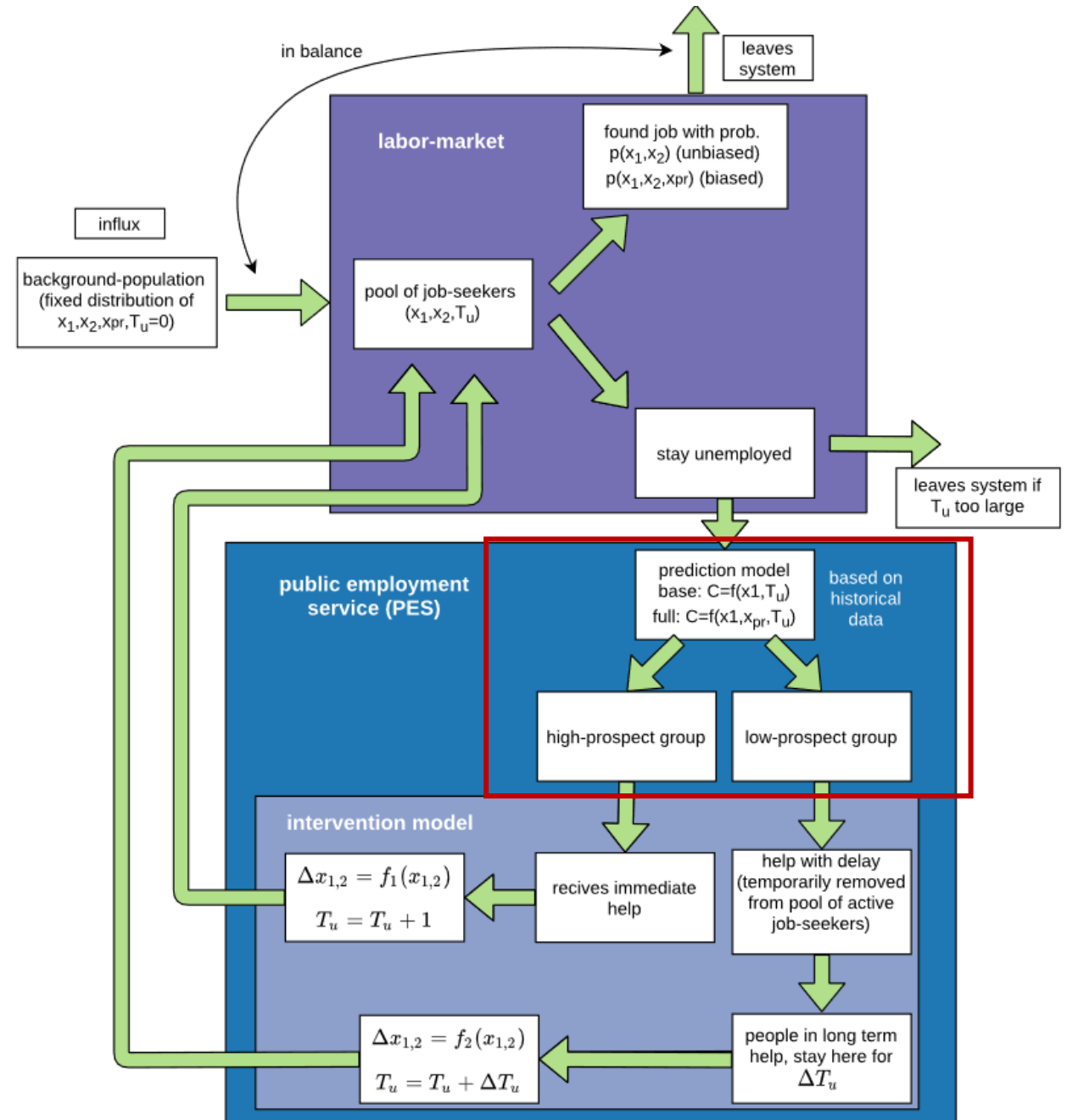
versus

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**

# 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?



Priv.Doiz. Dr.techn. Dipl.Ing. Dominik Kowald

[dkowald@know-center.at](mailto:dkowald@know-center.at)

<https://dominikkowald.info/>

---

KNOW CENTER Research GMBH

Sandgasse 34/2, 8010 Graz, Austria  
Firmenbuchgericht Graz  
FN 199 685 f  
UID: ATU 50367703  
[www.know-center.at](http://www.know-center.at)