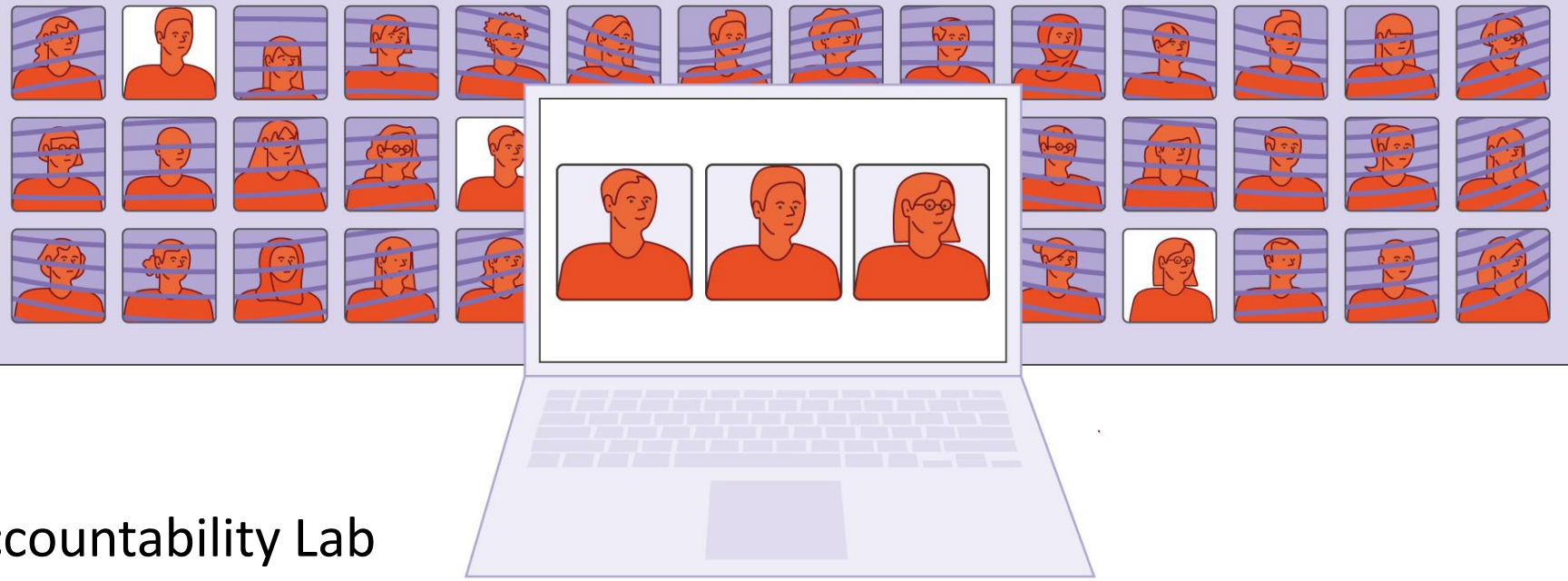
ExamAI – KI Testing & Auditing

Assurance Cases zum Prüfen von Anforderungen im HR Bereich



Marc Hauer

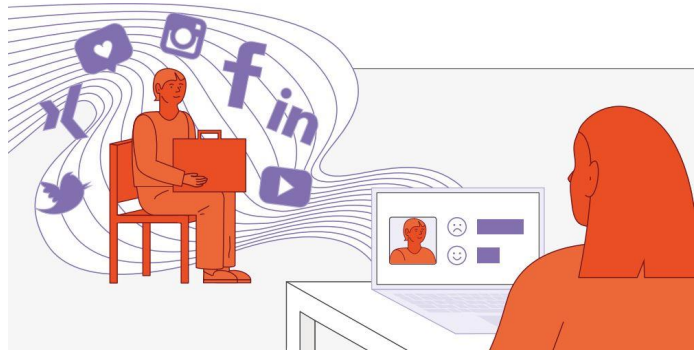
Algorithm Accountability Lab

TU Kaiserslautern

hauer@cs.uni-kl.de

ExamAI

Anwendungen von KI in HR

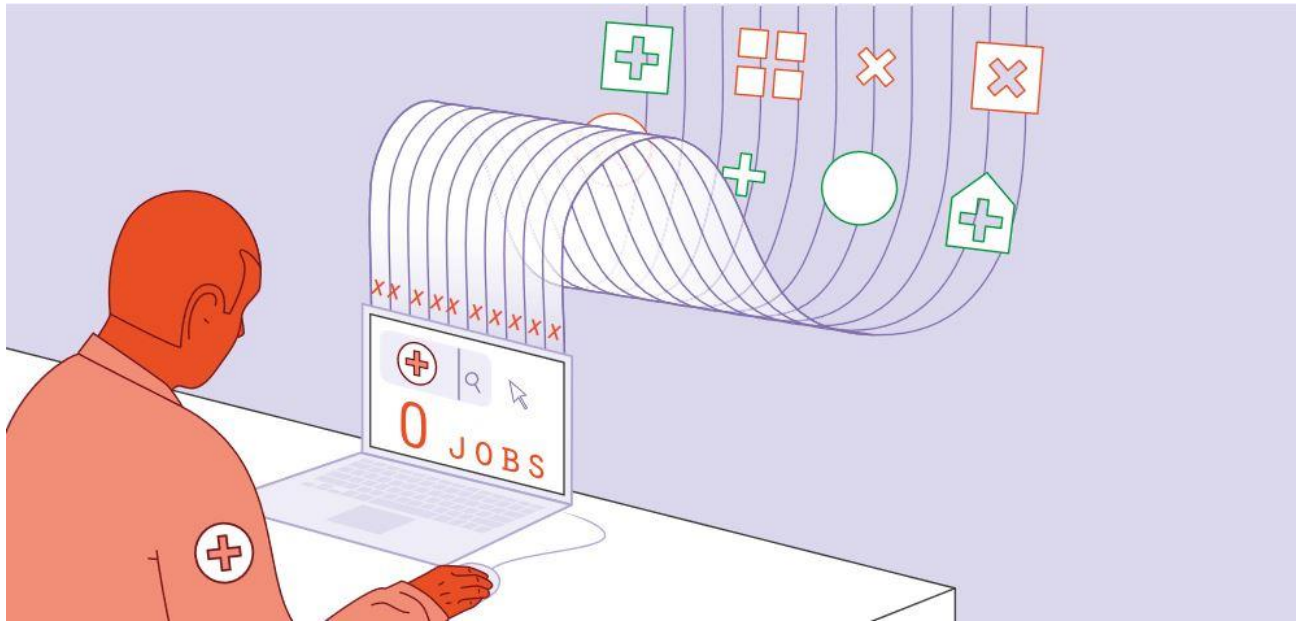


1. Automatisierte Vorschlagssysteme auf Personalplattformen
2. Persönlichkeitsbewertung per Lebenslauf/ strukturierter Eingabe oder Video
3. KI-basierte Background Checks
4. Chatbot der HR-Abteilung
5. Internes Jobprofil-Matching
6. Vorhersage der Kündigungsbereitschaft
7. Automatische Zuweisung bei Gig-Workern

https://testing-ai.gi.de/fileadmin/PR/Testing-AI/ExamAI_Publikation_Anwendungsszenarien_KI_HR.pdf

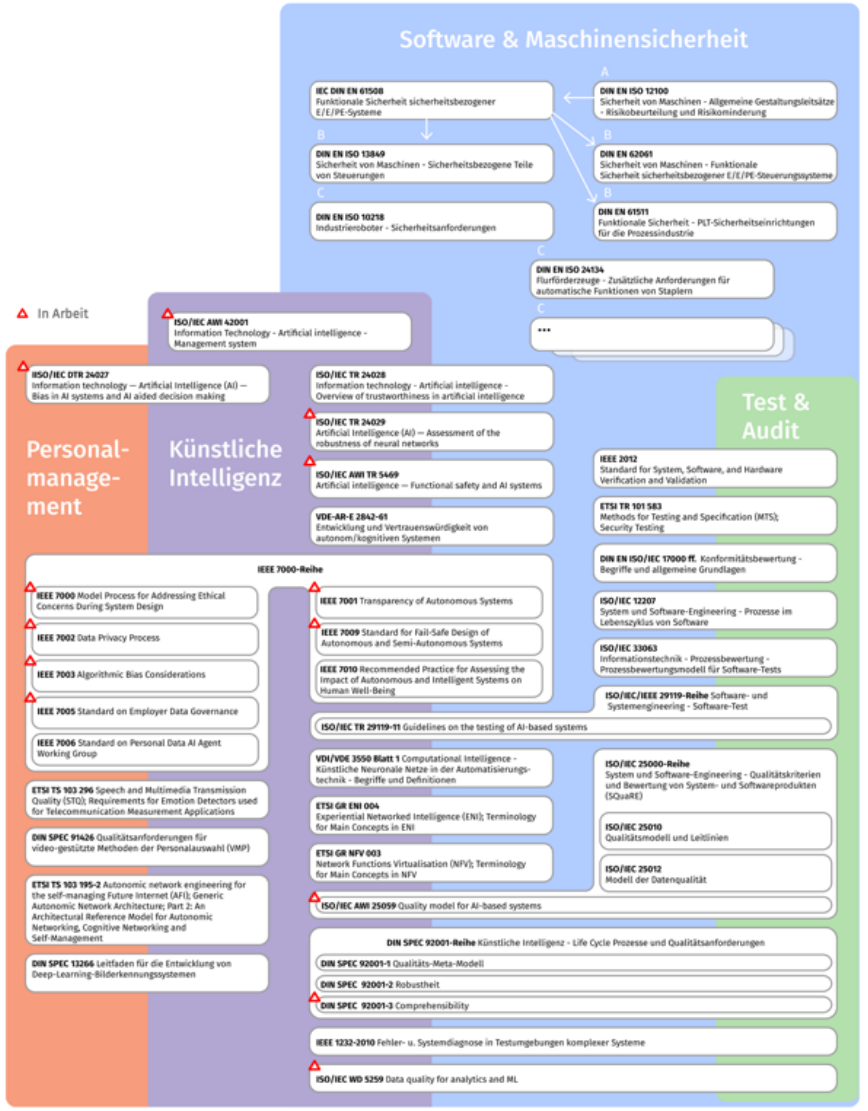
Anwendungen von KI in HR

- Art von Schädigungen bei Fehlverhalten der KI
  - Diskriminierung
- Fokus auf **Fairness** als Operationalisierung von nicht-Diskriminierung



Welche Normen zu KI in HR gibt es?

- Untersuchung passender Normen.
- Ergebnis: Kaum passende Normen zu finden.
- DIN SPEC 91426: Qualitätsanforderungen für videobasierte Methoden der Personalauswahl
- Tangential: DIN SPEC 13266: Leitfaden für die Entwicklung von Deep-Learning-Bilderkennungssystemen

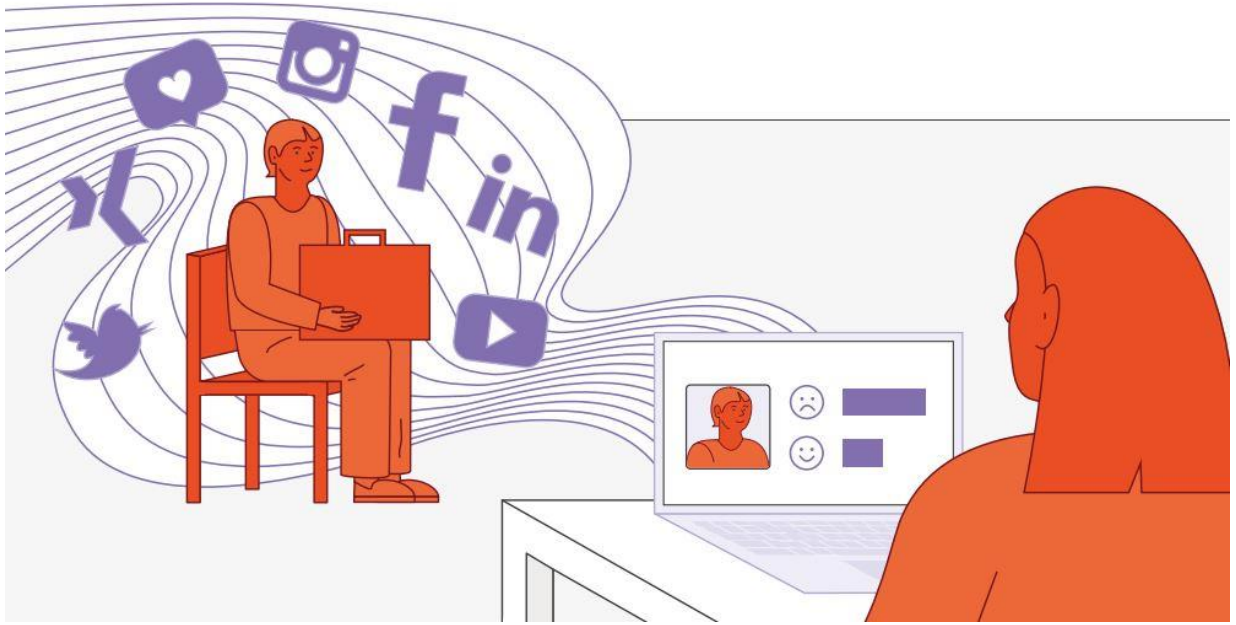


https://testing-ai.gi.de/fileadmin/PR/Testing-AI/ExamAI_Normenuebersicht_2021_04.pdf

Rechtliche Grundlagen zu Diskriminierung durch KI

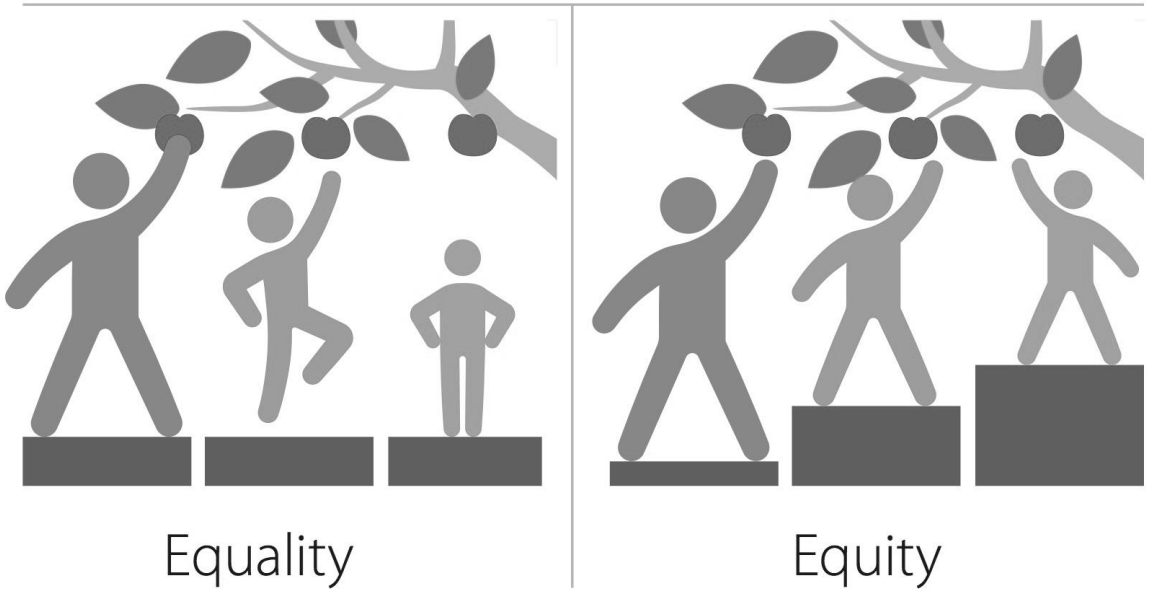
Rechtlicher Klärungsbedarf

- Wann liegt eine Diskriminierung, bzw., eine Benachteiligung vor?
- Was sind Rechtfertigungen oder sachliche Gründe?



Herausforderungen beim Testen auf Fairness

1. Was ist fair?
2. Wie adressiert man Gruppen ungleicher Größe?
3. Wie adressiert man sensible Daten und Proxy Variablen?
4. Wie löst man Konflikte zwischen verschiedenen Konzepten von Fairness?



Was ist ATDD?

Strukturierte Treffen zur Spezifikation von Softwareanforderungen

Einbeziehen aller relevanter Stakeholder

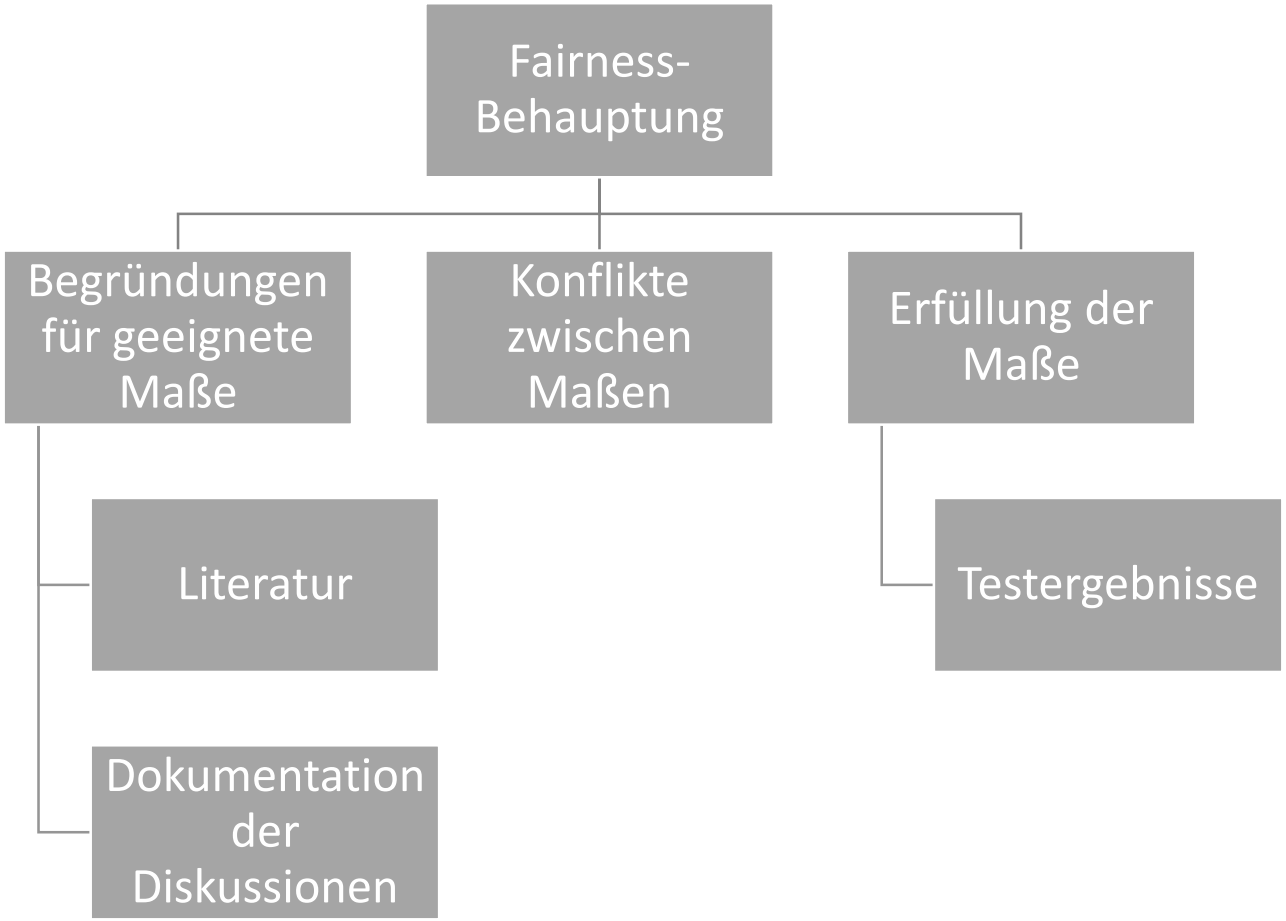
Entwicklung von Beispielsituationen für Funktionsweisen.

Darauf basierend Entwicklung formaler Akzeptanzkriterien.

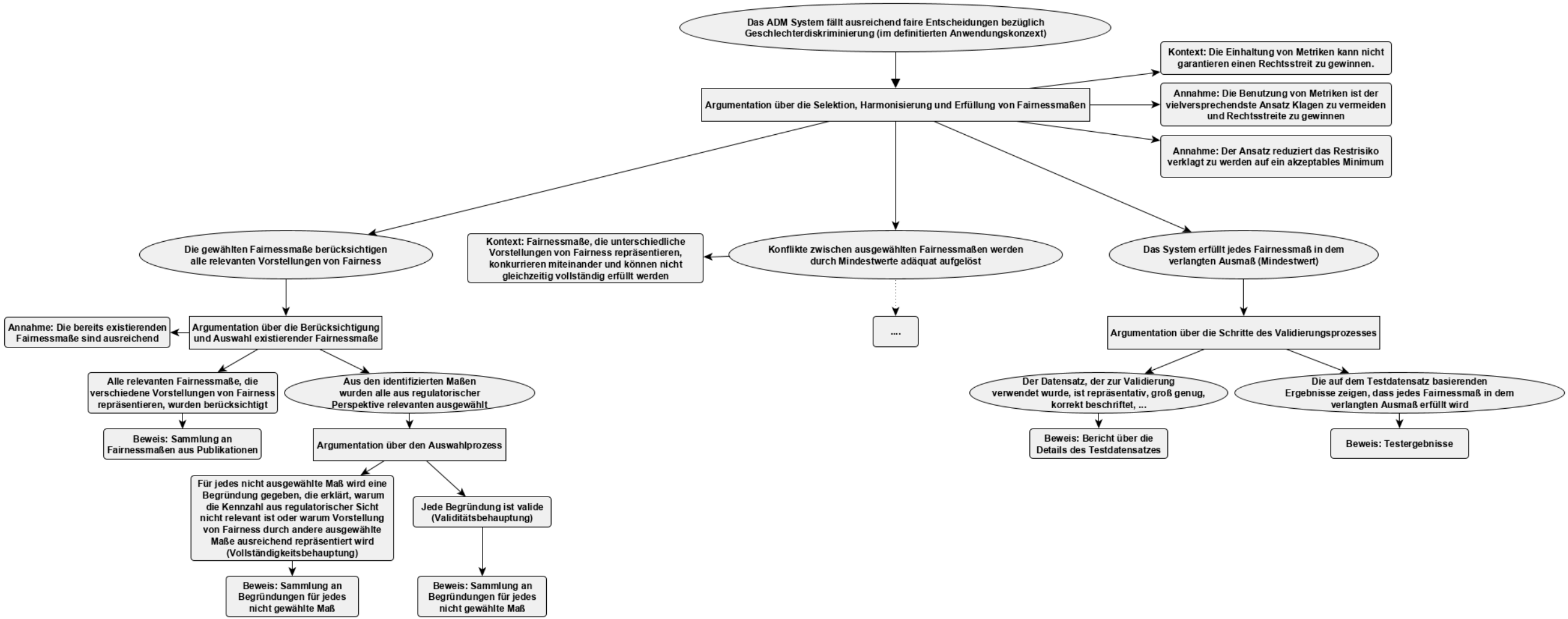
Darauf basierend Formulierung von Tests.

Teil der *test-first*-Philosophie, aber auf *high-level* Systemenebene

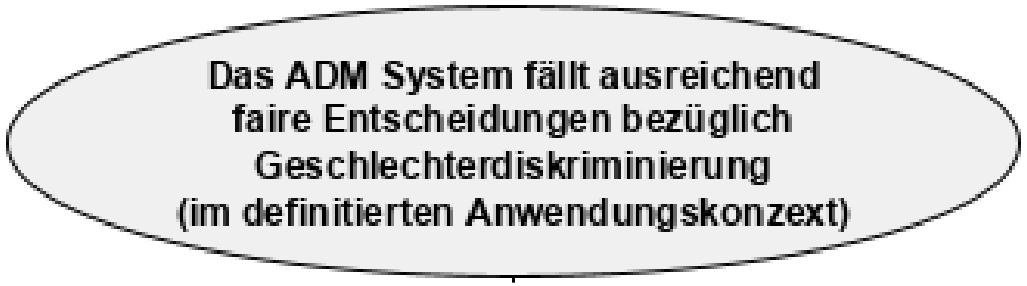
Abstrakte Ansicht



Konkretes Beispiel



https://ieeexplore.ieee.org/abstract/document/9440188


Das ADM System fällt ausreichend
faire Entscheidungen bezüglich
Geschlechterdiskriminierung
(im definitierten Anwendungskonzext)

Fazit

- Es gibt kein „Schema F“ um Fairness nachzuweisen.
- **ATDD** und **ACs** eignen sich, um systematisch über Maßnahmen, die Fairness sicherstellen, nachzudenken, sie umzusetzen und sie nachvollziehbar zu dokumentieren.
  - Vertreter aller relevanten Gruppen werden einbezogen;
  - ACs verlangen strukturierte Argumentation und Dokumentation;
  - Explizite Annahmen und automatisierbare Tests bieten Schutz vor unerwünschten Änderungen und Fehlern;
  - Auditoren können durch den AC die Angemessenheit der Maßnahmen prüfen.

Gefördert durch:
Im Rahmen des:

Ausblick

- Evaluation neuer Normen:
  - IEEE 7000 Reihe: Model Process for Addressing Ethical Concerns During system Design (kürzlich veröffentlicht)
  - ISO/IEC DTR 24027: Bias in AI systems and AI aided decision making (noch in Entwicklung)
- Wer entscheidet welche Stakeholder Gruppe angemessen ist und wie?
- Welche Fairness ist unter welchen Umständen (rechtlich) angemessen?
- ACs in der Normierung noch **zu generisch**, um einfach bau- und prüfbar zu sein;
- **Evaluierung** des Ansatzes in der Praxis notwendig;

Gefördert durch:
Im Rahmen des: