



Gleichstellung, künstliche Intelligenz, Algorithmen und automatisierte Entscheidungen: Wie lässt sich Diskriminierung im Kontext von KI erkennen und angehen?

Fabian Lütz

ERA (Trier 12-13 Juni 2023)

 This training session is funded under the 'Rights, Equality and Citizenship Programme 2014-2020' of the European Commission.

1

Geschlechtergleichstellung im Zeitalter der Algorithmen




Google (10.10.2021)

DuckDuckGo "CEO" (10/10/2021)

ERA (Trier 12-13 Juni 2023)

2

Große Sprachmodelle (LLMs): ChatGPT & Co

Das Problem der geschlechtsspezifischen Voreingenommenheit und der potenziellen Diskriminierung könnte bei der Verwendung von ChatGPT, GPT-4, Bard usw. zunehmen.

→ KI-Modelle für allgemeine Zwecke oder Grundlagenmodelle sind im KOM-Vorschlag nicht enthalten (der Rat forderte jedoch, GPAI zu berücksichtigen)

Why does a search for pictures of CEO show only male CEOs?

There could be multiple reasons why a search for pictures of CEOs predominantly shows male CEOs. One reason is that historically, men have held a majority of CEO positions in many industries. This gender disparity in leadership positions is known as the "glass ceiling" and can be attributed to various factors such as gender biases, societal expectations, and systemic discrimination.

Another reason could be the bias of the search algorithms themselves. If the search algorithm is not programmed to consider gender as a factor in its search results, it may prioritize images of male CEOs simply because they are more prevalent in the data it has been trained on.

It's important to note that this issue has been recognized and efforts are being made to improve diversity and representation in leadership positions. As more women and individuals from underrepresented groups break through the glass ceiling, we may see a shift in the search results as well.



<https://www.nytimes.com/2023/05/27/nyregion/avianca-airline-lawsuit-chatgpt.html>
A Man Sued Avianca Airline. His Lawyer Used ChatGPT. - The New York Times

6 days ago - Mr. Schwartz said that he had never used ChatGPT, and "therefore was unaware of the possibility that its content could be false." He had, he told Judge Castel, even asked the program to verify that...

ERA (TRIER 12-13 JUNI 2023)

3

3

Agenda

1. Einleitung

2. Derzeitiger und künftiger EU-Rechtsrahmen

3. KI und Geschlechtergleichstellung in der Praxis

4. Regulierung von KI und Algorithmen

5. KI als Wegbereiter für die Gleichstellung der Geschlechter?

6. Ausblick

ERA (Trier 12-13 Juni 2023)

4

4



**Ministerial Declaration
The G7 Digital and Tech
Ministers' Meeting
30 April 2023**

Responsible AI and Global AI Governance

39. In 2016, we initiated an international discussion on AI principles. This discussion helped pave the way for the 2019 [OECD AI Recommendation](#) (OECD AI Principles) and the associated work launching the [OECD AI Policy Observatory and Network of Experts](#). In 2020, we supported the launch of the [Global Partnership on Artificial Intelligence \(GPAI\)](#).
40. The OECD AI Principles provide guidance for trustworthy AI and for ensuring an open and enabling environment for AI development and deployment that is grounded in human rights and democratic values. Since adoption of these principles, the OECD continues to engage and work with the global AI community to support their implementation.
41. Rapid AI developments call for attention to, and cooperation on, emerging and medium-term policy issues including development of technical standards, developed by [international standards development organisations \(SDOs\)](#), as well as other tools to ensure the development and deployment of trustworthy AI in line with the OECD AI Principles. In this context, we welcome the contributions of existing initiatives on these topics.
42. We reaffirm our commitment to promote human-centric and trustworthy AI based on the OECD AI Principles and to foster collaboration to maximise the benefits for all brought by AI technologies. We oppose the misuse and abuse of AI to undermine democratic values, suppress freedom of expression, and threaten the enjoyment of human rights.
43. We stress the importance of [international discussions on AI governance](#) and interoperability between AI governance frameworks, while we recognise that like-minded approaches and policy instruments to achieve the common vision and goal of trustworthy AI may vary across G7 members. Tools for trustworthy AI, such as [regulatory and non-regulatory frameworks](#), [technical standards](#) and assurance techniques, can promote trustworthiness and can allow for the comparable assessment and evaluation of AI systems. We support the development of tools for trustworthy AI through multistakeholder international organisations, and encourage the development and adoption of international technical standards in SDOs through private sector-led multistakeholder processes. We commend work to date in the OECD on mapping the commonalities and differences between trustworthy AI frameworks, and we intend to work together to support such work that fosters interoperability.

1. Einleitung

- Definitionen
- Erklärung von Toronto
- Erinnerung an Gleichstellungsprobleme
- Aktualisierung Juni 2023

ERA (Trier 12-13 Juni 2023)

5

Definitionen: KI und Algorithmen

Künstliche Intelligenz

- Ein KI-System ist eine Software, die "(...) für eine gegebene Reihe von durch den Menschen definierten Zielen Ergebnisse wie Inhalte, Vorhersagen, Empfehlungen oder Entscheidungen erzeugt, die die Umgebung, mit der sie interagieren, beeinflussen;(...)".

(Art. 3 Abs. 1, EU-AI-Gesetz)

- Beispiele:
 - Maschinelles Lernen (überwachtes, unüberwachtes, verstärkendes, tiefes Lernen), statistische Ansätze, Such- und Optimierungsmethoden

Algorithmus

- "... hinreichend detaillierte und systematische Handlungsanweisungen zur Lösung eines mathematischen Problems, so dass der Computer bei korrekter Implementierung für jeden korrekten Satz von Eingaben die korrekte Ausgabe berechnet"

(Prof. Katharina Zweig)

ERA (Trier 12-13 Juni 2023)

6

6

Definitionen: Entscheidungsfehler

Vorurteil

- Kognitive Verzerrungen, Sammelbegriff für systematische Entscheidungsfehler

(Daniel Kahneman, Olivier Sibony, Cass R. Sunstein: Noise, 2021, S. 163f.)

„Noise“

- Variabilität des Irrtums, zufällige Variabilität der Urteile *oder eine* unsichtbare Steuer

(K/S/S: Noise, 2021, S. 3-5 und D. Kahneman, HBR, 2016, S. 36-43)

Algorithmische Diskriminierung = liegt vor, wenn automatisierte Systeme zu einer ungerechtfertigten Ungleichbehandlung oder Benachteiligung von Menschen, z. B. aufgrund ihres Geschlechts, beitragen.

ERA (Trier 12-13 Juni 2023)

7

7



Quelle: <https://www.torontodeclaration.org>

Erklärung von Toronto

- Pflicht der Staaten, die Menschenrechte zu schützen; dies schließt die Gewährleistung des Rechts auf Nichtdiskriminierung durch Akteure des Privatsektors ein" (Toronto, Absatz 38)
- Notwendigkeit der Überwachung des Einsatzes von maschinellem Lernen durch den privaten Sektor in Kontexten, die das Risiko diskriminierender Ergebnisse bergen (Toronto, Absatz 40)

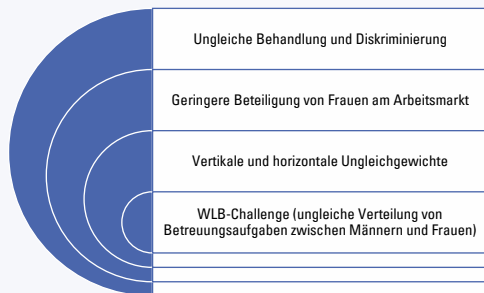
ERA (TRIER 12-13 JUNI 2023)

8

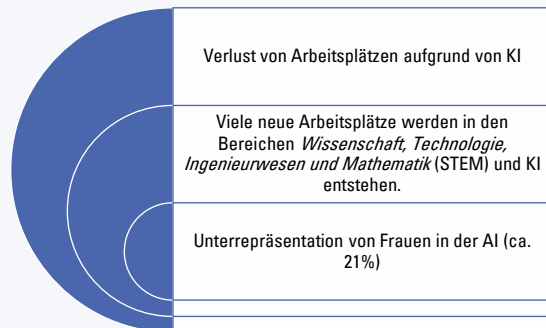
8

Erinnern an einige Probleme der Geschlechtergleichstellung (Bsp. Arbeitsmarkt)

Allgemein



AI-spezifisch



ERA (Trier 12-13 Juni 2023)

9

9

Neueste Entwicklungen: Juni 2023

- RightsCon-Gipfel Costa Rica 5-8 Juni 2023: Menschenrechte im digitalen Zeitalter
- Neue und aufkommende Technologien benötigen dringend eine Aufsicht und eine solide Transparenz: UN-Experten (2. Juni 2023) → gender & discrimination
- Policy Brief des UN-Generalsekretärs zum Global Digital Compact → Multi-Stakeholder-Ansatz (interessant für Unternehmens- und Rechtsberater)
- Konsultation UN-Menschenrechte & zu Gender, Technologie und der Rolle der Wirtschaft (15. Juni 2023)

ERA (Trier 12-13 Juni 2023)

10

10

EU Update aktuelle Entwicklungen: Juni 2023

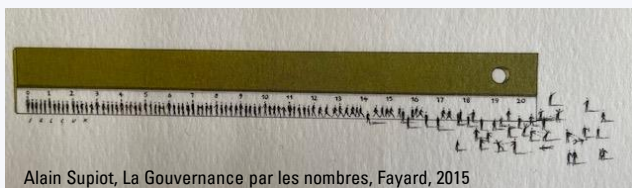
- Kompromisstext des Europäischen Parlaments (EP) (16. Mai 2023)
- Abstimmung im Plenum des EP voraussichtlich 12-15 Juni 2023
- "Nach ihrer Verabschiedung werden sie die weltweit ersten Regeln für künstliche Intelligenz sein" (EP News, 11. Mai 2023)
- Die Mitglieder des Europäischen Parlaments haben Beschwerden über KI-Systeme eingereicht und erhalten Erklärungen zu Entscheidungen, die auf KI-Systemen mit hohem Risiko basieren und ihre Rechte erheblich beeinträchtigen
- Stärkung des EU-AI-Büros, das für die Überwachung der Umsetzung des AI-Regelwerks zuständig ist

ERA (Trier 12-13 Juni 2023)

11

11

2. Gegenwärtige und zukünftige rechtliche Rahmenbedingungen

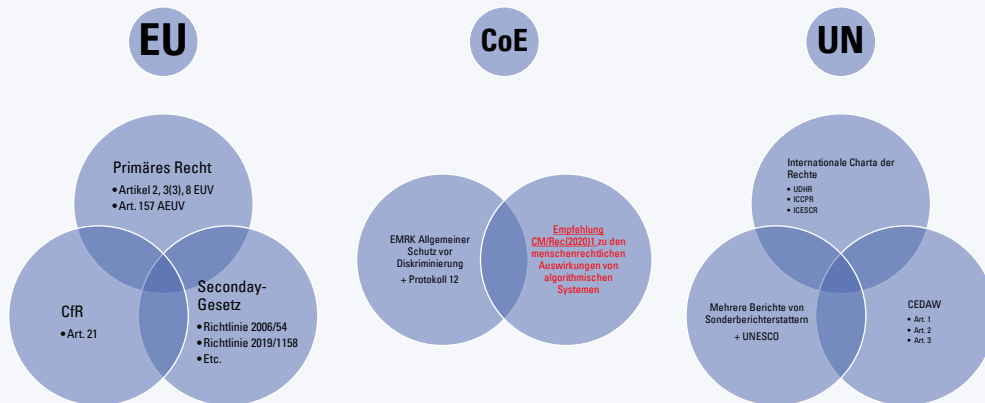


- Anwendbares Recht
- Brüsseler Effekt?
- EU-Vorschlag: AI-Gesetz

ERA (Trier 12-13 Juni 2023)

12

Derzeitige rechtliche Rahmenbedingungen



ERA (Trier 12-13 Juni 2023)

13

13

Künftiger Rechtsrahmen

EU AI-Verordnung

- DSA
 - ECAT/JRC Sevilla
- DMA
- AI-Gesetz
- EB-Normen-Richtlinie
 - Erwägungsgrund 16
- AI-Haftungsrichtlinie
- Bessere Rechtsetzung/Überprüfung des EU-Besitzstands + nächstes Mandat 2024-2029)

Europarat

- Allgemeines:
 - Rahmenkonvention des Europarats über KI, Menschenrechte, Demokratie und Rechtsstaatlichkeit (Annahme im November 2023?)
- GE/Diskriminierung:
 - Rechtsinstrument zur Bekämpfung von GE/Diskriminierung (bis 2025)

UN

- UN
 - Global Digital Compact
- OHCHR
 - Priorität, Berichte
- CEDAW
 - Allgemeine Empfehlung 40
- CSW
 - Vereinbarte Schlussfolgerungen 'CSW 67

ERA (Trier 12-13 Juni 2023)

14

14

AI & BRÜSSELER EFFEKT

John Oliver erklärt warum die EU die KI regulieren will
(Last Night Tonight Show, Februar 2023)

25:54 / 27:52

15

15

Brüsseler Effekt" zur Regulierung der algorithmischen Diskriminierung

EU an vorderster Front bei der Gestaltung der Regulierung von

KI/algorithmischer Diskriminierung

- Siehe Anu Bradford (in Allgemein+GDPR)
- Die EU eröffnete ein Büro in San Francisco, um die Entwicklungen im Bereich KI/Technologie zu verfolgen (*die USA schaffen KI, die EU reguliert KI*)
- Nach der Verabschiedung von DSA/DMA + ECAT wird die EU wahrscheinlich die erste große Wirtschaftsregion sein, die eine weitreichende Regulierungsaufsicht für KI-Systeme mit KI-Gesetz hat

PRESS RELEASE | Publication 05 September 2022

EU opens new office in San Francisco to reinforce its digital diplomacy

The European Union has opened its new office in San Francisco, California, a global centre for digital technology and innovation. The office will reinforce the EU's cooperation with the United States on digital diplomacy and strengthen the EU's capacity to reach out to key public and private stakeholders, including policy makers, the business community, and civil society in the digital technology sector.

High Representative/Vice-President Josep Borrell, said:

The opening of the office in San Francisco responds to the EU's commitment to strengthen transatlantic technological cooperation and to drive the global digital transformation based on democratic values and standards. It is a concrete step to further reinforce the EU's work on issues such as cyber and countering hybrid threats, and foreign information manipulation and interference.

The EU office in San Francisco will seek to promote EU standards and technologies, digital policies and regulations and governance models, and to strengthen cooperation with US stakeholders, including by advancing the work of the EU-US Trade and Technology Council. Find more information in the [EEAS press release](#).




Related topics

International relations

ERA (Trier 12-13 Juni 2023)

16

16



COMMISSION EUROPÉENNE

Bruxelles,
21.4.2021
COM(2021)
final
2021/0106/CX

Proposition de
RÈGLEMENT DU PARLEMENT EUROPÉEN ET DU CONSEIL

ÉTABLISSANT DES RÈGLES HARMONISÉES CONCERNANT L'INTELLIGENCE ARTIFICIELLE (LÉGISLATION SUR L'INTELLIGENCE ARTIFICIELLE) ET MODIFIANT CERTAINS ACTES LÉGISLATIFS DE L'UNION

[SEC(2021) 167 final] - (SWD(2021) 84 final) - (SWD(2021) 85 final)

ANNEXE III
SYSTÈMES D'IA À HAUT RISQUE VISÉS À L'ARTICLE 6, PARAGRAPHE 2

Les systèmes d'IA à haut risque au sens de l'article 6, paragraphe 2, sont les systèmes d'IA répertoriés dans l'un des domaines suivants:

1. Identification biométrique et catégorisation des personnes physiques:
 - (a) les systèmes d'IA destinés à être utilisés pour l'identification biométrique à distance «en temps réel» et «a posteriori» des personnes physiques;
2. Gestion et exploitation des infrastructures critiques:
 - (a) les systèmes d'IA destinés à être utilisés en tant que composants de sécurité dans la gestion et l'exploitation du trafic routier et dans la fourniture d'eau, de gaz, de chauffage et d'électricité;
3. Éducation et formation professionnelle:
 - (a) les systèmes d'IA destinés à être utilisés pour déterminer l'accès ou l'affectation de personnes physiques aux établissements d'enseignement et de formation professionnelle;
 - (b) les systèmes d'IA destinés à être utilisés pour évaluer les étudiants des établissements d'enseignement et de formation professionnelle et pour évaluer les participants aux épreuves occasionnelles requises pour intégrer les établissements d'enseignement;
4. Emploi, gestion de la main-d'œuvre et accès à l'emploi indépendant:
 - (a) les systèmes d'IA destinés à être utilisés pour le recrutement ou la sélection de personnes physiques, notamment pour la diffusion des offres d'emploi, la présélection ou le filtrage des candidatures, et l'évaluation des candidats au cours d'entretiens ou d'épreuves;
 - (b) l'IA destinée à être utilisée pour la prise de décisions de promotion et de licenciement dans le cadre de relations professionnelles contractuelles, pour l'attribution des tâches et pour le suivi et l'évaluation des performances et du comportement de personnes dans le cadre de telles relations.

EU-Gesetz über künstliche Intelligenz

COM (2021), 206 endgültig

Artikel 2

Geltungsbereich

(1) Diese Verordnung gilt für:

- (a) Anbieter, die KI-Systeme in der Union in Verkehr bringen oder in Betrieb nehmen, unabhängig davon, ob diese Anbieter in der Union oder in einem Drittland ansässig sind;
- (b) Nutzer von KI-Systemen, die sich in der Union befinden;
- (c) Anbieter und Nutzer von KI-Systemen, die in einem Drittland ansässig sind, wenn der von dem System erzeugte Output in der Union verwendet wird;

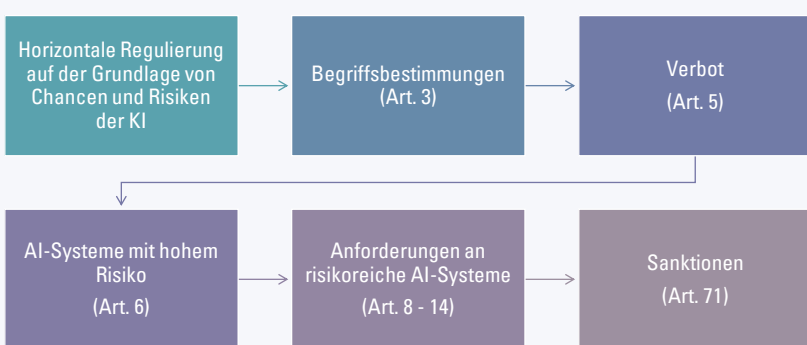
Wirtschaft und Menschenrechte (Lausanne, 1. Juni 2023)

17

17

EU KOM(2021), 206 endgültig

Gesetz über künstliche Intelligenz (AIAct)



```

graph TD
    A[Horizontale Regulierung auf der Grundlage von Chancen und Risiken der KI] --> B[Begriffsbestimmungen (Art. 3)]
    A --> C[Verbot (Art. 5)]
    B --> D[AI-Systeme mit hohem Risiko (Art. 6)]
    D --> E[Anforderungen an risikoreiche AI-Systeme (Art. 8 - 14)]
    E --> F[Sanktionen (Art. 71)]
    C --> F
    
```

ERA (Trier 12-13 Juni 2023)

18

18

EU COM AI Act & Gleichstellung

Horizontales Instrument für AI

- Der Vorschlag ergänzt das bestehende Unionsrecht zur Nichtdiskriminierung
- Konkrete Anforderungen zur Minimierung des Risikos der Diskriminierung durch Algorithmen
- Gestaltung und Qualität der für die Entwicklung von KI-Systemen verwendeten Datensätze.
- Regeln für: Tests, Risikomanagement, Dokumentation und menschliche Aufsicht während des gesamten Lebenszyklus von KI-Systemen.

Gleichstellung im AIA-Vorschlag

- "(Nicht)-Diskriminierung" erwähnt (16x)
- "Gleichstellung der Geschlechter" (1x)
- "Frauen" (x2): Art. 21 CFR + in Bezug auf die Einstellungssysteme

ERA (Trier 12-13 Juni 2023)

19

19

EU COM Entwurf eines AI-Gesetzes & Rekrutierung

Risikobasierter Ansatz

- Artikel 6 + Artikel 6 Absatz 2 + Anhang III (Nr. 4 Einstellungs- und Auswahlverfahren AI-Systeme)
 - Relevanz für die Gleichstellung der Geschlechter
 - Gefährdung der Grundrechte?

Anhang III (Einstellungssysteme)

- "Während des gesamten Einstellungsverfahrens und bei der Bewertung, Beförderung oder Weiterbeschäftigung von Personen in arbeitsbezogenen Vertragsverhältnissen können solche Systeme historische Diskriminierungsmuster, beispielsweise

Entwurf von Kompromissänderungsanträgen des Europäischen Parlaments (16/05/2023)

Artikel 4a - Allgemeine Grundsätze, die für alle KI-Systeme gelten

1.e) "**Vielfalt, Nichtdiskriminierung und Fairness**" bedeutet, dass KI-Systeme in einer Weise entwickelt und genutzt werden, die unterschiedliche Akteure einbezieht und **den gleichberechtigten Zugang, die Gleichstellung der Geschlechter und die kulturelle Vielfalt fördert, wobei diskriminierende Auswirkungen und unfaire Vorurteile, die nach Unionsrecht oder nationalem Recht verboten sind, zu vermeiden sind.**"

ERA (Trier 12-13 Juni 2023)

20

20

EU COM Entwurf AIA & Anhang III

Anhang III (Einstellungssysteme), Nr. 4

- Beschäftigung, Arbeitnehmermanagement und Zugang zur Selbstständigkeit:
 - (a) KI-Systeme, die für die Einstellung oder Auswahl natürlicher Personen verwendet werden sollen, insbesondere für die Ausschreibung freier Stellen, das Screening oder die Filterung von Bewerbungen, die Bewertung von Bewerbern im Rahmen von Vorstellungsgesprächen oder Tests;
 - (b) AI ist dazu bestimmt, Entscheidungen über die Beförderung und Beendigung von Arbeitsvertragsverhältnissen, die Aufgabenzuweisung sowie die Überwachung und Bewertung der Leistung und des Verhaltens von Personen in solchen Verhältnissen zu treffen.

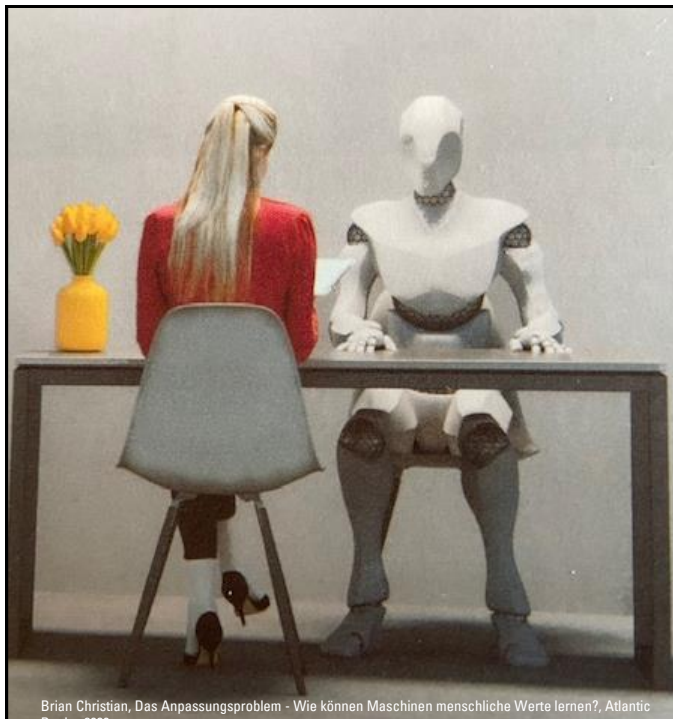
Artikel 6 Absatz 2

- Einstufung als AI-System mit hohem Risiko
- Anwendung der besonderen Bestimmungen von Kapitel 2 (Artikel 8 - 14):
 - Artikel 9: Risikomanagementsystem
 - Artikel 10: Daten und Datenverwaltung
 - Artikel 11: Technische Dokumentation
 - Artikel 12: Aufbewahrung der Aufzeichnungen
 - Artikel 13: Transparenz und Bereitstellung von Informationen für die Nutzer
 - Artikel 14: **Menschliche Aufsicht**

ERA (Trier 12-13 Juni 2023)

21

21



Brian Christian, Das Anpassungsproblem - Wie können Maschinen menschliche Werte lernen?, Atlantic Books, 2020

3. KI und Geschlechtergleichstellung in der Praxis

- KI, Vorurteile und Geschlechtergleichstellung
- Öffentliche Behörden
- Private Unternehmen
- Rechtliche Fragen
- Rechtsprechung

ERA (Trier 12-13 Juni 2023)

22

Geschlechtergleichstellung, KI und Vorurteile

Geschlechtsspezifische Vorurteile

- **Word2vec/Wortvektoren:** Sprache, maschinelles Lernen, Häufigkeit, Ranking, Metaphern, Bedeutung von Datensätzen + LLMs
- **Stereotype und Vorurteile:** "Der Mann ist für den Computerprogrammierer das, was die Frau für die Hausfrau ist?" (siehe Bolukbasi/Chang et.al. (2016), Debiasing Word Embeddings)
- Männer sollten auch mit Elternurlaub, Haushaltshilfe, Teilzeitarbeit... assoziiert werden.

Voreingenommenheit ist

menschlich

- **Daten** spiegeln die Gesellschaft wider, prägen aber auch Wahrnehmungen und Vorurteile.
- **Gender Innovations Project of Stanford (Maschinelles Lernen: Analyse des Geschlechts,** <https://genderinnovations.stanford.edu>)
- **Forschung zum Test auf implizite Assoziationen (Greenwald und Banaji) +** implicit.harvard.edu/

ERA (Trier 12-13 Juni 2023)

23

23

KI und Gleichstellung in der Praxis

Arbeitsagentur AT

- Staatliche Nutzung von AI / Monopol der Arbeitsmarktvermittlung (AMS)
- Mögliche Diskriminierung von Frauen
- Problem: WLB spiegelt sich in den Daten wider
- Quellcode des offengelegten Algorithmus

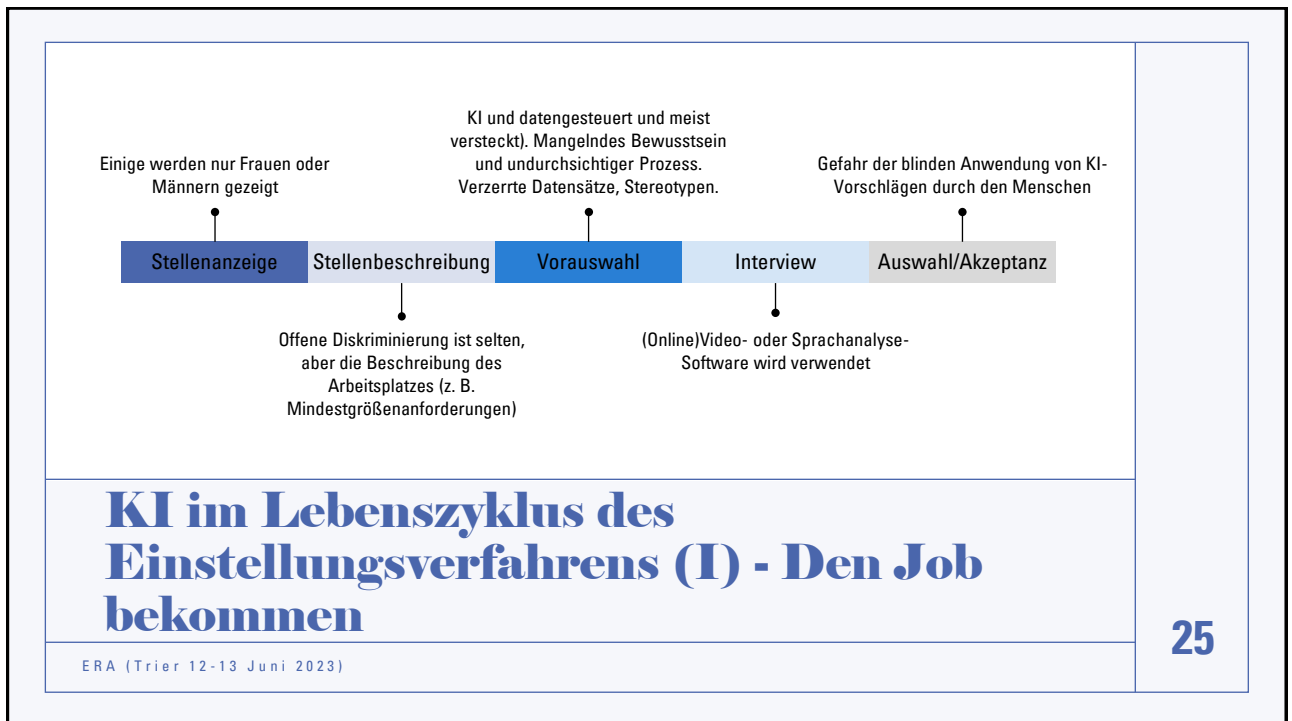
Amazon Rekrutierung KI

- Privater Schauspieler
- Mögliche Diskriminierung von Frauen
- Problem: Historische Daten (viel mehr Daten von Männern als von Frauen)
- Lösung: Neue Daten verwenden!

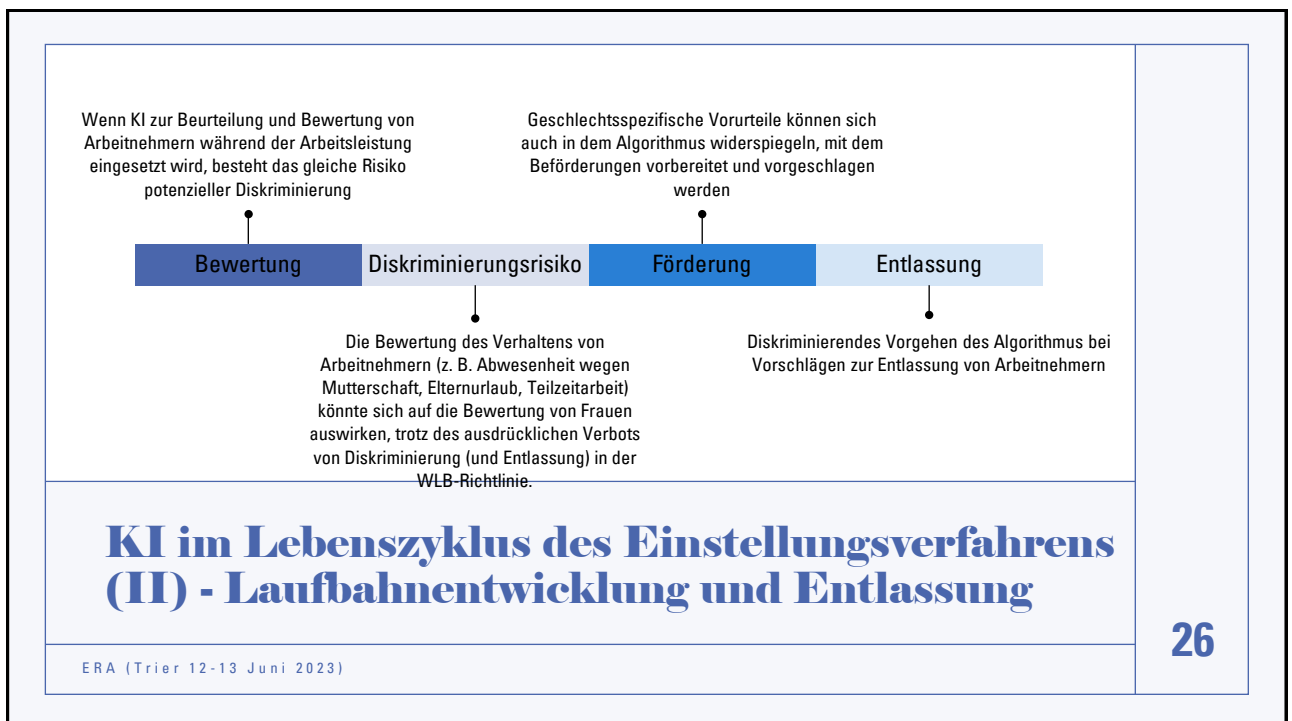
ERA (Trier 12-13 Juni 2023)

24

24



25



26

Rechtliche Fragen

Probleme der Beweisführung und des Nachweises

- Beweisführung (AI schreibt keinen Grund für die Ablehnung eines Bewerbers)
- Beweislast (wer beweist was?) + Umkehrung
- Zugang zu undurchsichtigen KI-Systemen?
- Algorithmus als Geschäftsgeheimnis/gewerblicher Rechtsschutz
 - In-Kamera-Verfahren als Lösung
- Richter als KI-Experte oder externes Fachwissen erforderlich?

Strafverfolgung

- Europäische Kommission, GD Justiz
- Nationale Antidiskriminierungsstellen (siehe neue Dir)
- Private Durchsetzung
- Verweisungsverfahren nach Art. 267 AEUV (zentrale Bedeutung für die Rechtsfortbildung)
 - Wann wird ein nationales Gericht einen Fall von Diskriminierung im Zusammenhang mit künstlicher Intelligenz dem EuGH vorlegen?
- Leitender Daten-/Computerwissenschaftler für KOM

ERA (Trier 12-13 Juni 2023)

27

27

EU-Rechtsprechung (I)

Danfoss C-109/88 (1988)

- **"Wendet ein Unternehmen ein völlig undurchsichtiges Entgeltsystem an, so obliegt es dem Arbeitgeber zu beweisen, dass seine Entgeltpraxis nicht diskriminierend ist, wenn eine Arbeitnehmerin in Bezug auf eine relativ große Zahl von Arbeitnehmern feststellt, dass das durchschnittliche Entgelt für Frauen niedriger ist als für Männer" (§ 16)**

Kelly, C-104/10 (2011)

- "gibt einem Bewerber um eine Berufsausbildung, der der Ansicht ist, dass seine Bewerbung wegen eines Verstoßes gegen den **Gleichbehandlungsgrundsatz** nicht angenommen wurde, **keinen Anspruch auf Informationen, die dem Kursträger über die Qualifikationen der anderen Bewerber für den betreffenden Kurs vorliegen, damit er 'Tatsachen glaubhaft machen kann, die das Vorliegen einer unmittelbaren oder mittelbaren Diskriminierung vermuten lassen' im Sinne** dieser Bestimmung." (§ 38)
- "Es kann jedoch nicht ausgeschlossen werden, dass eine Verweigerung der Offenlegung durch den Beklagten im Rahmen der Feststellung solcher Tatsachen die Erreichung des mit der Richtlinie verfolgten Ziels gefährden und damit insbesondere Artikel 4 Absatz 1 der Richtlinie seine Wirksamkeit nehmen könnte. Es ist Sache des nationalen Gerichts, zu prüfen, ob dies im Ausgangsverfahren der Fall ist." (§ 39)

ERA (Trier 12-13 Juni 2023)

28

28

EU-Rechtsprechung (II)

Meister C-415/10 (2012)

- Kein Recht auf Information über andere Bewerber im Falle der Nichtauswahl (§ 46)
- Es kann jedoch nicht ausgeschlossen werden, dass **die Weigerung des Beklagten, Zugang zu Informationen zu gewähren, einer der Faktoren ist, die im Rahmen der Feststellung von Tatsachen zu berücksichtigen sind, die das Vorliegen einer unmittelbaren oder mittelbaren Diskriminierung vermuten lassen**. Es ist Sache des vorlegenden Gerichts (...), zu prüfen, ob dies im Ausgangsverfahren der Fall ist." (§ 47)

Shuch-Ghannadan, C-274/18 (2019).

- **feststeht, dass** diese Rechtsvorschriften einen **wesentlich höheren Prozentsatz weiblicher Arbeitnehmer als männlicher Arbeitnehmer** benachteiligen, und wenn diese Rechtsvorschriften nicht durch ein rechtmäßiges Ziel sachlich gerechtfertigt sind oder die Mittel zur Erreichung dieses Ziels nicht angemessen und erforderlich sind. Art. 19 Abs. 1 der Richtlinie ist dahin auszulegen, dass er die Partei, die sich durch eine solche Diskriminierung für beschwert hält, **nicht verpflichtet, zur Glaubhaftmachung des Vorliegens einer Diskriminierung bestimmte statistische Daten oder Tatsachen in Bezug auf die von der streitigen nationalen Regelung betroffenen Arbeitnehmer vorzulegen, wenn diese Partei keinen oder nur einen erschwerten Zugang zu diesen Daten oder Tatsachen hat.**

ERA (Trier 12-13 Juni 2023)

29

29

EU-Rechtsprechung (III): Rechtliche Grenzen der KI?

• URTEIL DES CJEU VOM 21. 6. 2022 - RECHTSSACHE C-817/19 LIGUE DES DROITS HUMAINS ECLI:EU:C:2022:491

- Definition der KI/des maschinellen Lernens (Abs. 58, 194, 195)
- Wo liegen die Grenzen? Für den PNR-Kontext (erfordert 'vorher festgelegte' Kriterien) war die Nicht-Stabilität der Kriterien, die den Algorithmen des maschinellen Lernens eigen sind, entscheidend ("(...) diese Anforderung schließt die Verwendung von Technologien der künstlichen Intelligenz in selbstlernenden Systemen ('maschinelles Lernen') aus, die in der Lage sind, den Bewertungsprozess ohne menschliches Eingreifen oder Überprüfung zu ändern (...)", Rn. 194)
- Was wird der Gerichtshof im Bereich der Grundrechte tun? Entscheidungen, bei denen viel auf dem Spiel steht?

IT verbietet die Nutzung von ChatGPT aufgrund von Verstößen gegen die Datenschutzgrundverordnung und andere EU-Agenturen untersuchen das Problem

→ GDPR wird von Forschern und Praktikern oft als AD-Recht betrachtet (oft wegen des Fehlens spezifischer Vorschriften, die sich mit KI und Gleichstellung/Diskriminierung befassen)

→ Zusammenspiel verschiedener rechtlicher Regelungen GDPR (siehe Lütz, La pollinisation croisée entre droit de la protection des données et droit de la non-discrimination, *dennächst 2023*)

ERA

30

30

Pablo Jensen, Pourquoi la société ne se laisse pas mettre en équations, Seuil, 2018.

4. Regulierung von Algorithmen

- Regulatorische Ansätze
- Regulierungsinstrumente und Erkennung von Vorurteilen und Diskriminierung
- EU-Gerichtsbarkeit
- Blick über die Grenzen der EU hinaus

ERA (Trier 12-13 Juni 2023)

31

Regulierung

Markt

Risiko der privaten Regelsetzung?

Selbstregulierung

Verordnung

Selbstbeurteilungen oder AI-Audits?

Jurowetski et al., Die Privatisierung der KI-Forschung(-er): Ursachen und mögliche Folgen (2021)

ERA (Trier 12-13 Juni 2023)

32

Die Rolle der Wirtschaft und die Sorgfaltspflicht der Unternehmen im Bereich der Menschenrechte bei KI

Erklärung von Toronto nennt 3 Kernschritte für die Sorgfaltspflicht von Unternehmen im Bereich der Menschenrechte für maschinelle Lernsysteme:

- (a) **Ermittlung** potenziell diskriminierender Ergebnisse
- (b) **Verhinderung und Abschwächung** von Diskriminierung und Verfolgung von Reaktionen
- (c) **Transparenz** hinsichtlich der Bemühungen zur Feststellung, Verhinderung und Eindämmung von Diskriminierung

Multi-Stakeholder-Ansätze sind eine gute Praxis

Während des gesamten Lebenszyklus und der verschiedenen Phasen des KI-Systems können Due Diligence und Compliance eine Rolle spielen, um Vorurteile abzubauen und Nichtdiskriminierung zu gewährleisten

Verantwortung als Entwickler von KI-Systemen zur Vermeidung und Verringerung von Menschenrechtsverletzungen

ERA (Trier 12-13 Juni 2023)

33

33

Unternehmensebene und Rolle der Forschenden

Aequitas

An open source bias audit toolkit for machine learning developers, analysts, and policymakers to audit machine learning models for discrimination and bias, and make informed and equitable decisions around developing and deploying predictive risk-assessment tools.

© 2021 IBM

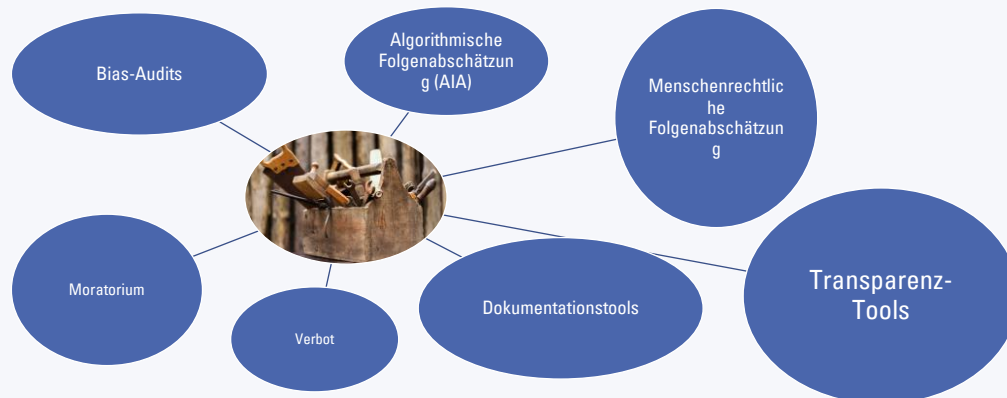
- KI-Unternehmen sollten während des gesamten KI-Lebenszyklus Tools einsetzen, um Fairness zu gewährleisten und/oder Verzerrungen in ihren Algorithmen/KI zu erkennen.
 - Spezielle Leitfäden für KI-Entwickler und Programmierer
 - Ausbildung und Schulung zur Gewährleistung fairer und unvoreingenommener Algorithmen
- (Unabhängige) **Forschende** sind unerlässlich, um den konkreten Bedarf und die Durchführbarkeit der rechtlichen Verpflichtungen zu bewerten ², z. B. um Verzerrungen im Design und in den Datensätzen zu erkennen, zu verringern und zu beseitigen, um diskriminierende Ergebnisse zu reduzieren

ERA (Trier 12-13 Juni 2023)

34

34

Instrumente zur Sicherstellung der Regulierungsziele: Abschwächung von Vorurteilen und Abbau von Diskriminierung



ERA (Trier 12-13 Juni 2023)

35

35

Globale Dimension

CoE

- Laufende Arbeiten für einen Rechtsrahmen (2025)
- Empfehlung CM/Rec(2020)1 zu den menschenrechtlichen Auswirkungen von algorithmischen Systemen
- "Es muss sichergestellt werden, dass (...) **geschlechtsspezifische und andere gesellschaftliche Ungleichgewichte und Ungleichgewichte in der Erwerbsbevölkerung, die in unserer Gesellschaft noch nicht beseitigt sind, nicht absichtlich oder versehentlich durch algorithmische Systeme aufrechterhalten werden**" (Präambel)

OECD

- OECD/LEGAL/0449, Empfehlung des Rates über künstliche Intelligenz (2019)
- "KI-Akteure sollten (...) während des gesamten Lebenszyklus von KI-Systemen die Menschenrechte und demokratischen Werte achten. Dazu gehören (...) **Nicht-Diskriminierung und Gleichheit** (...)" (1.2a)

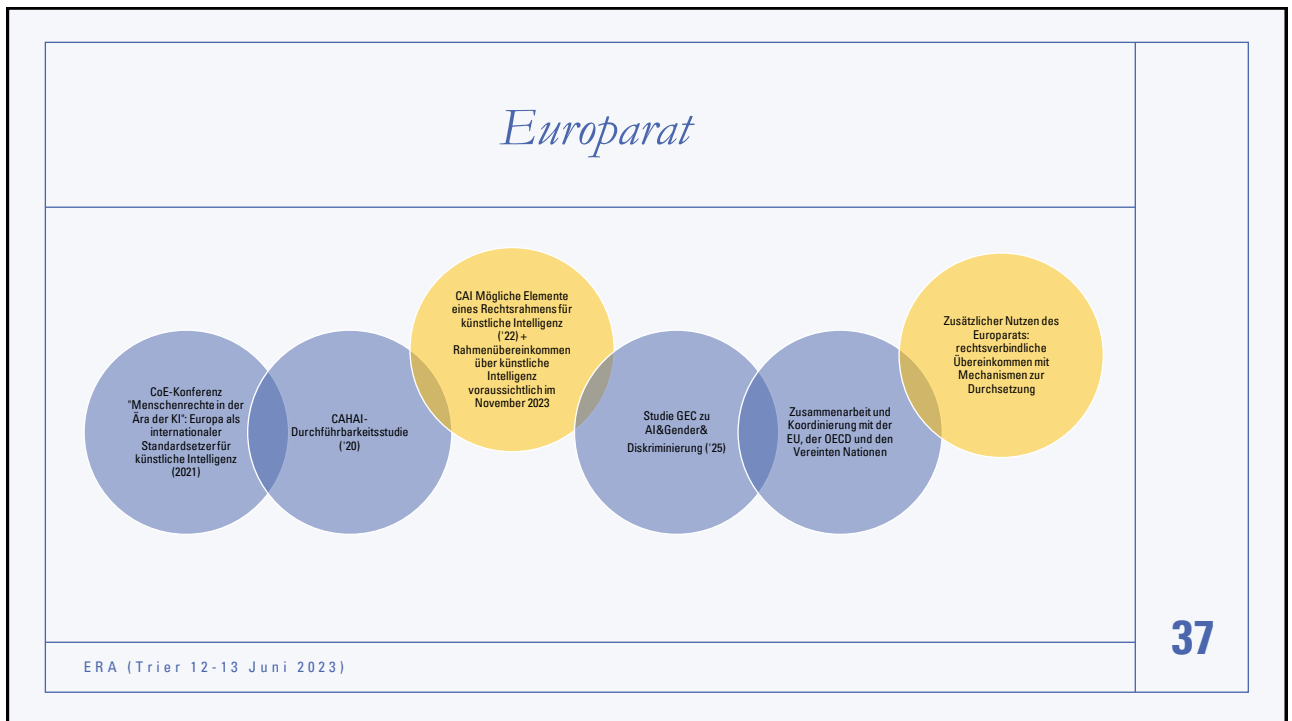
Vereinte Nationen

- UNESCO-Empfehlung zur Ethik der KI (2021), Gender als Politikbereich 6 (§§ 87-93)
- Das Recht auf Privatsphäre im digitalen Zeitalter, Bericht des UN-Hochkommissars für Menschenrechte, A/HRC/48/31:
- "Die Fortschritte der neuen Technologien dürfen nicht dazu genutzt werden, **die Menschenrechte auszuhöhlen, Ungleichheiten zu vertiefen oder bestehende Diskriminierungen zu verschärfen**" (§ 4) + "**Diese Systeme (...) entscheiden darüber, wer eine Chance hat, für eine Stelle eingestellt zu werden**" (Bericht, § 57).

ERA (Trier 12-13 Juni 2023)

36

36



37

5. KI als Wegbereiter für die Gleichstellung der Geschlechter?

- KI als Diskriminierungsdetektiv
- Vergleich Mensch/KI
- Praktische Überlegungen

ERA (Trier 12-13 Juni 2023)

Éric Sadin, La vie algorithmique - Critique de la raison numérique, L'échappée, 2015

38

KI als Diskriminierungsdetektiv

Elemente

- Vom Sichtbaren zum Unsichtbaren, vom Expliziten zum Undurchsichtigen
- Menschliche Entscheidungen sind fehleranfällig, aber auch KI ist nicht unbedingt besser
- Algorithmen können (einige) *Verzerrungen* und *Rauschen* überwinden (K/S/S, Noise, S. 334-337)

Beispiele

- Automatische Rechtsprechungsanalyse
- Gezieltere Durchsetzung, bessere Nutzung der Ressourcen
- Algorithmen verwenden, um andere Algorithmen auf mögliche Unterscheidungen zu testen
- Software zur Erkennung von Bias/Diskriminierung

ERA (Trier 12-13 Juni 2023)

39

39

Ein Vergleich zwischen Mensch und KI

Vorteile von AI

- Kohärenz, Objektivität, Geschwindigkeit?
- Bessere Entscheidungsträger?
- Ausübung des Ermessensspielraums?
- Identifizierung von Mustern
- KI als (große) Hilfe (Unterstützung), aber der Mensch muss die Kontrolle behalten

KI als Assistent

- Assistent des "kleinen Bürokraten" (Lipsky, 1980)
- Mehr Vorhersehbarkeit und Zuverlässigkeit bei Verwaltungsentscheidungen?
- Vorbereitung der Fallanalyse
- Schlussfolgerung: KI kann die Entscheidungsfindung nicht ersetzen
- *"Die beispiellose Macht der künstlichen Intelligenz, die eine **Kraft für das Gute** sein kann (...)" (Behinderung (A/HRC/49/52))*

ERA (Trier 12-13 Juni 2023)

40

40

8	5	6	5	0	21	24	11	20	23	7	0	4	5	4	7
0	0	0	0	20	0	28	26	21	24	18	4	0	0	0	0
3	0	0	11	1	125	230	0	119	13	26	9	9	0	0	2
5	0	10	10	95	239	254	229	96	21	28	15	33	0	0	3
0	0	0	8	197	251	243	253	201	58	12	12	11	0	0	0
0	0	7	18	180	183	234	248	176	79	9	8	20	0	0	0
0	0	5	62	216	168	228	235	173	175	126	6	6	0	0	0
0	0	6	145	252	241	232	231	231	237	227	41	5	30	0	0
0	0	9	143	253	242	221	225	247	246	222	186	12	17	0	0
0	0	9	40	226	225	222	222	235	225	194	160	12	14	0	0
0	0	16	6	196	230	223	217	224	228	119	6	15	25	0	0
0	0	22	6	94	244	232	232	231	228	137	6	29	11	0	0
0	0	16	10	9	210	246	238	204	241	138	5	30	6	0	0
0	0	10	11	11	162	229	227	221	250	150	7	8	13	0	0
0	0	7	8	11	170	228	238	238	243	183	159	125	6	0	0
0	0	8	5	8	195	215	225	229	228	231	241	100	4	5	0

Melanie Mitchell, Künstliche Intelligenz - Ein Leitfaden für denkende Menschen, 2021, S. 20-21

6. Schlussfolgerung und Ausblick

- KI birgt Risiken für die Geschlechtergleichstellung
- Auch Chancen
- Regulierung ist der Schlüssel
- Gemeinsame Rolle von Unternehmen und Staaten
- Globale Dimension

ERA (Trier 12-13 Juni 2023)

Ausgewählte Veröffentlichungen zu Algorithmen und Gleichstellung:

- *How the 'Brussels Effect' Could Shape the Future Regulation of Algorithmic Discrimination*, Duodecim Astra, Ausgabe 1, S. 142-163 (2021), ([online](#) verfügbar).
- *Discrimination by correlation - Towards eliminating algorithmic biases and achieving Gender Equality*, Beitrag zu einer Monographie, Transcript Verlag (2021), ([Open Access](#)).
- *Geschlechtergleichstellung und künstliche Intelligenz in Europa. Auseinandersetzung mit den direkten und indirekten Auswirkungen von Algorithmen auf geschlechtsspezifische Diskriminierung*. ERA Forum (2022), verfügbar unter: <https://doi.org/10.1007/s12027-022-00709-6> ([Open Access](#))
- *Künstliche Intelligenz und geschlechtsspezifische Diskriminierung*, in: Quintavilla Alberto, Temperman Jeroen (eds.) Human Rights and Artificial Intelligence, Oxford University Press (Juni 2023).
- *Algorithmische Entscheidungsfindung aus der Gleichstellungsperspektive - Ein Balanceakt zwischen Gender Data Gap, Gender Bias, Machine Bias und Regulierung*, GENDER 1/2 (2023), ([Open Access online](#)).
- *Geschlechtergleichstellung und künstliche Intelligenz: SDG 5 und die Rolle der UN bei der Bekämpfung von Stereotypen, Vorurteilen und geschlechtsspezifischer Diskriminierung*, in: Cristani Federica, Fornalé Elisa (eds.) Women's Empowerment and its Limits, Palgrave, Mai 2023).
- *Le rôle du droit pour contrer la discrimination algorithmique dans le recrutement automatisé*, In: Guillaume Florence (eds.) La technologie, l'humain et le droit, Stämpfli Verlag (Juni 2023)

42

ERA (Trier 12-13 Juni 2023)

DANKE
an die ERA

MERCI
für das Zuhören

DANKE
für Ihr Interesse an algorithmischer
Diskriminierung



Fabian.Luetz@unil.ch



43

ERA (Trier 12-13 Juni 2023)