Algorithmischer Bias: Woher er kommt und was wir tun können

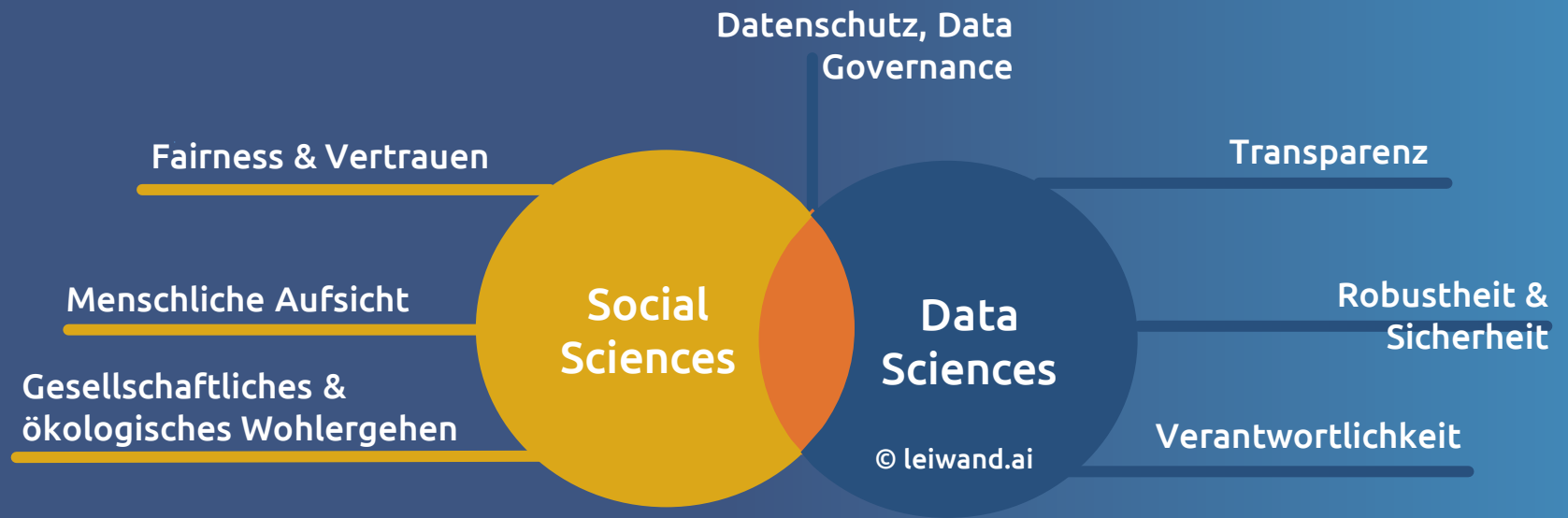
Dr. Gertraud Leimüller, MPA (Harvard)
CEO winnovation consulting gmbh & Co-Founderin leiwand.ai

Symposium „Automatisierung der Arbeit“, WU
29. Februar 2024

Die Mission von leiwand.ai

Faire und hochwertige KI zu einer Realität machen

leiwand AI entwickelt Werkzeuge, die Künstliche Intelligenz vertrauenswürdig machen (Trustworthy AI).
Dabei nutzen wir unsere Expertise in Innovationsentwicklung (Open Innovation) und Alorithmic Fairness.



leiwand.ai – what we do

RESEARCH

EU Fundamental rights Agency (FRA) *"Providing evidence on bias when using algorithms – simulation and testing of selected cases"*, 2021.

FFG-funded multi-disciplinary research project „fAIr by design“, 2021 – 2024.

Project partner in **Complexity Science Hub Vienna's Digital Humanism Roadmap** project: Roadmap for Digital Humanism in Complexity Science. Ongoing.

PRODUCTS

STAIR - Smart and Trustworthy AI in Recruiting. **Funded by FFG**. Ongoing.

Digital Humanism Roadmap: Roadmap for Trustworthy AI Made Simple. Project lead, in **collaboration with Plattform Industrie 4.0**. Funded by the Wirtschaftsagentur Wien and WWTF within the framework of the Digital Humanism Initiative. Ongoing.

STANDARDIZATION AND CONSULTING

Project editor of the project **ISO/IEC 12792** "Transparency taxonomy of AI systems“. Ongoing

Participation in **Austrian (ASI), European (CEN/CENELEC) and International (ISO/IEC) committees** and working groups on AI, NLP, and trustworthy AI

Social Media Monitoring **for Amnesty International Italy**‘s „Barometro dell‘Odio“ Project, since 2018.

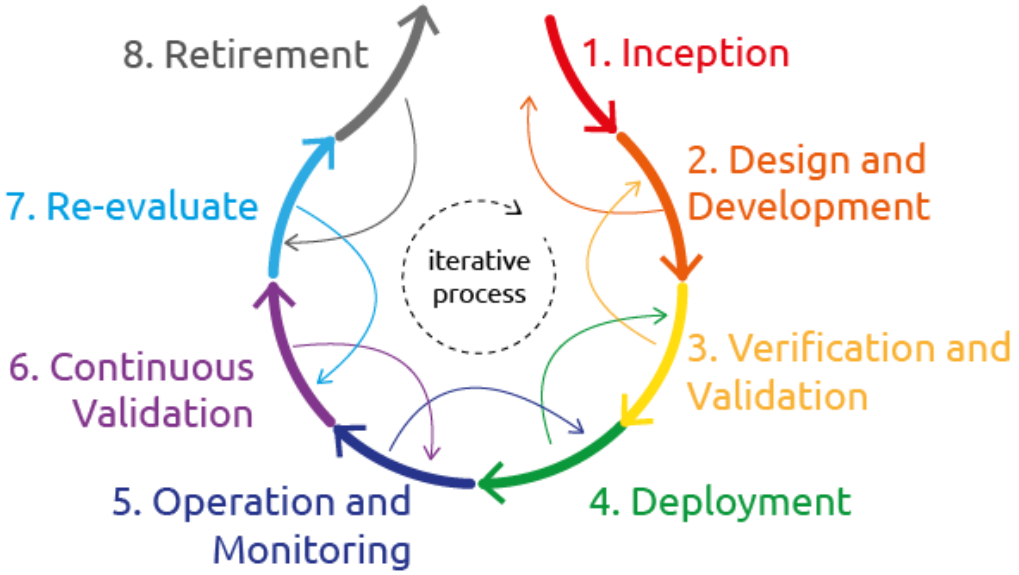
FRA Report Bias in algorithms - Artificial intelligence and discrimination | European Union Agency for Fundamental Rights (europa.eu)

Forschungsprojekt fAIr by design:

Wie kann KI so entwickelt werden, dass sie von Anfang an „fair“ agiert?

Vermeidung von algorithmischem Bias und Diskriminierung

- 8 Partner:innen (2 Universitäten, 6 Unternehmen)
- 3 Jahre (2021 – 2024)
- 5 Use Cases



Das Projekt „fAIr by design" wird aus Mitteln der Nationalstiftung für Forschung, Technologie und Entwicklung sowie dem Österreich-Fonds finanziert. Die Abwicklung des Förderungsprogramms Laura Bassi 4.0 erfolgt durch die Österreichische Forschungsförderungsgesellschaft (FFG) und mit freundlicher Unterstützung des Bundesministeriums für Arbeit und Wirtschaft (BMAW).
https://www.fairbydesign.eu/

fAIr by design

key points for fair AI

- Algorithmische Fairness ist kein nice-to-have. Sie ist ein **zentraler Qualitätsfaktor** für alle KI-Systeme, welche direkt an der Schnittstelle zu Menschen eingesetzt werden.
- Der **jeweilige Einsatzkontext** → konkrete Nutzer:innen und Betroffene – bestimmt, wo es zu Fairnessproblemen und Diskriminierung kommen kann. Deshalb kann es keinen Persilschein für ein technisches KI-System quer über alle Einsatzkontexte geben.
- **Fokus** auf besonders gefährdete Gruppen von Personen – **Groups at Risk / Risikoanalysen** - und **Festlegung entsprechender Zielwerte** ist wichtig, da Fairness für alle nicht funktioniert (Tradeoffs!)
- **Algorithmischer Bias** kommt nicht nur über die Trainingsdaten in die KI-Systeme, sondern hat **viele Quellen**

  → **Interdisziplinäre Vorgehensweisen** und **Monitoring entlang des KI-Lebenszyklus** sind nötig.

Beispiel: Übersetzungs-AI
Probieren Sie selbst aus!

Übersetzen Sie ins Deutsche:

My doctor is pregnant.

Übersetzen Sie ins Deutsche:

My secretary is on paternity leave.

Das Ergebnis zeigt eingebettete Stereotype

Übersetzen Sie ins Deutsche:

My doctor is pregnant.

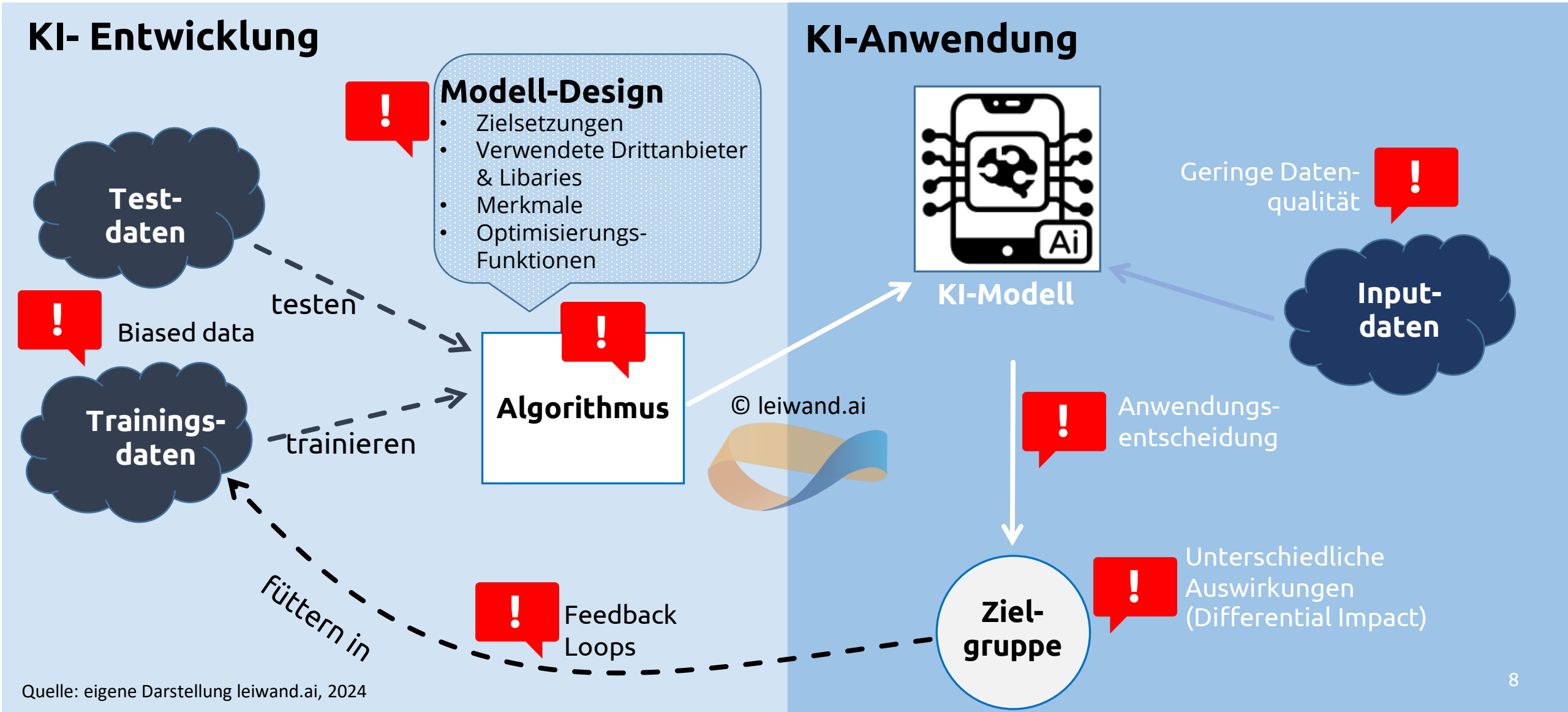
Mein Arzt ist schwanger.

Übersetzen Sie ins Deutsche:

My secretary is on paternity leave.

Meine Sekretärin ist im Vaterschaftsurlaub.

Woher kommt algorithmischer Bias? Quellen entlang des AI-Lebenszyklus



Quelle: eigene Darstellung leiwand.ai, 2024

Was tun? Die Verantwortung von Führungskräften beim Einsatz von KI-Systemen

1. Bewusstsein im Unternehmen für den Umgang mit KI schaffen
   - 7 Prinzipien für „vertrauenswürdige KI“ lt. EU HL Group, Compliance-Anforderungen
2. Governance-Strukturen und Kompetenzen für den Umgang mit KI aufbauen
3. Vor der Beschaffung / Entwicklung von KI-Systemen:
   - Zielsetzungen für KI-Systeme festlegen
   - Qualitätsanforderungen und KPIs an KI-Systeme definieren
   - Groups at Risk (Wer könnte Nachteile haben?) definieren und Fairness by Design als Prinzip verankern, Fundamental Rights Impact Assessments
4. Im Unternehmen: Monitoring des Einsatzes von KI-Systemen entlang des Lebenszyklus, da diese lebendig sind (selbstlernend) und sich mit der Nutzung verändern (können).

Vertrauenswürdigkeit von KI ist auch ein wirtschaftlicher Vorteil

Bild: upsplash, NASA

Das Marktforschungsunternehmen Gartner erwartet, dass jene Unternehmen, die bis 2026 AI-Transparenz, Vertrauen und Sicherheit operationalisieren,

sich beim Einsatz von AI, dem Erreichen von Businesszielen und der Akzeptanz für AI um 50% verbessern werden.

Quelle: 2023 Gartner® Top 10 Strategic Technology Trends: https://www.centreon.com/a-look-at-gartner-technology-trends-2023/

„Algorithmen sind menschliche Kreationen und unterliegen wie jedes andere menschliche Unterfangen auch Fehlern.“

US Richter Stephen Smith

R. Brauneis & E. P. Goodman, Algorithmic Transparency for the Smart City, 20 YALE J. L. & TECH. 103 (2018), FN63; https://yjolt.org/sites/default/files/20_yale_j._l._tech._103.pdf

leiwand.ai

Machen wir **AI**
gemeinsam **leiwand**!

Stay in touch!