

# KI

## Warum die Regulierung komplex ist

---

Prof. Dr. Katharina Zweig

Algorithm Accountability Lab  
TU Kaiserslautern



Konstituierende Sitzung der  
Enquete-Kommission  
„Künstliche Intelligenz“ am 27.9.2018

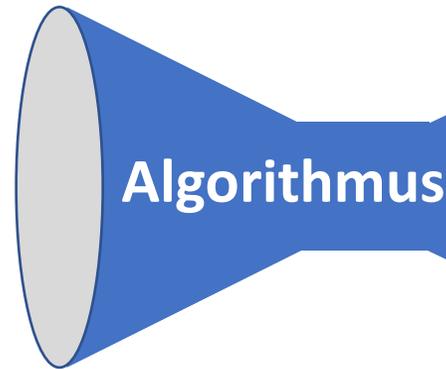
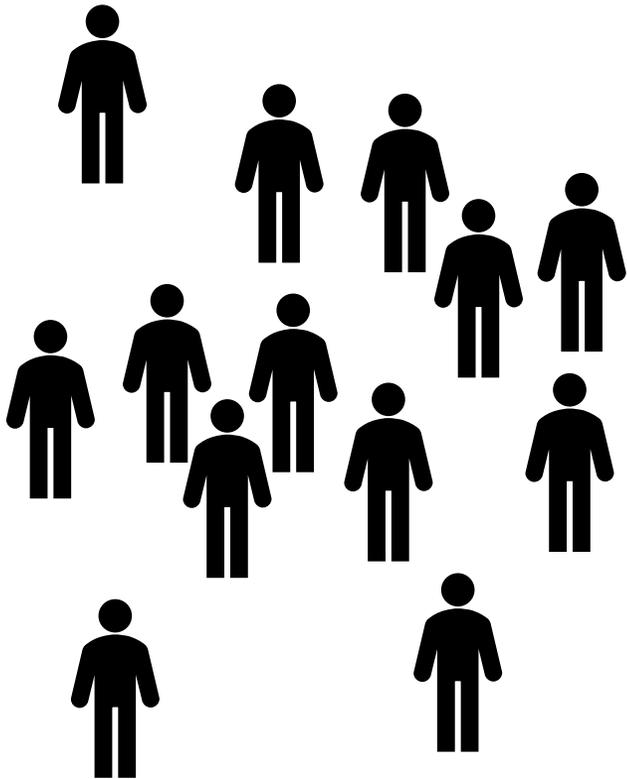
---

Aus der Rede von Bundestagspräsidenten  
Dr. Schäuble:

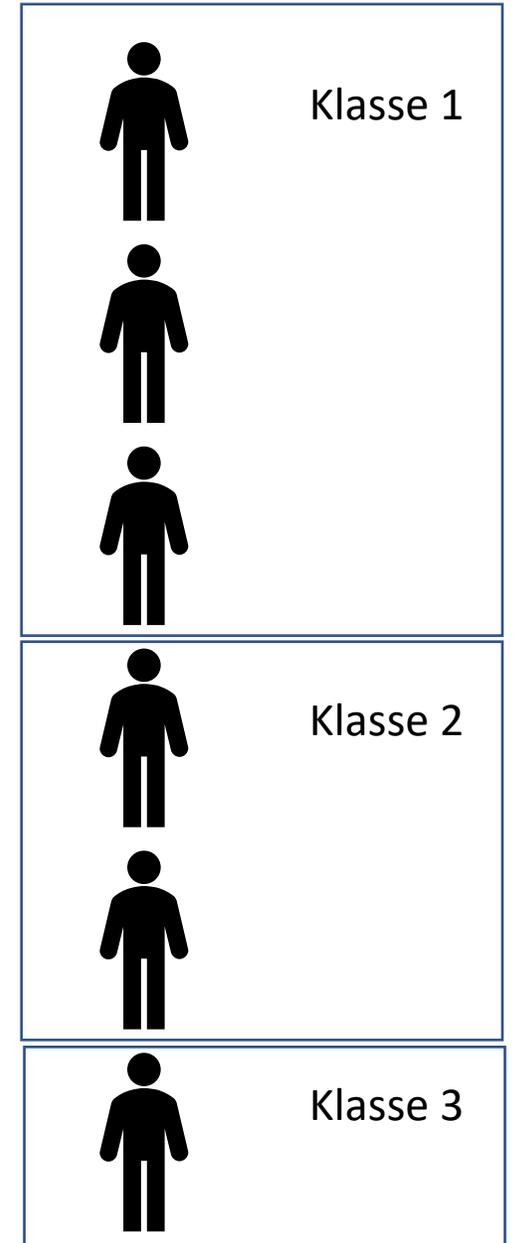
- „Die künstliche Intelligenz gilt  
Vielen als neue Zauberformel des  
technischen Fortschritts, ...
- ... sie wird dichten, ...
- ... sie wird belohnen und bestrafen ...“



# Algorithmische Entscheidungssysteme (ADM Systeme)



oder



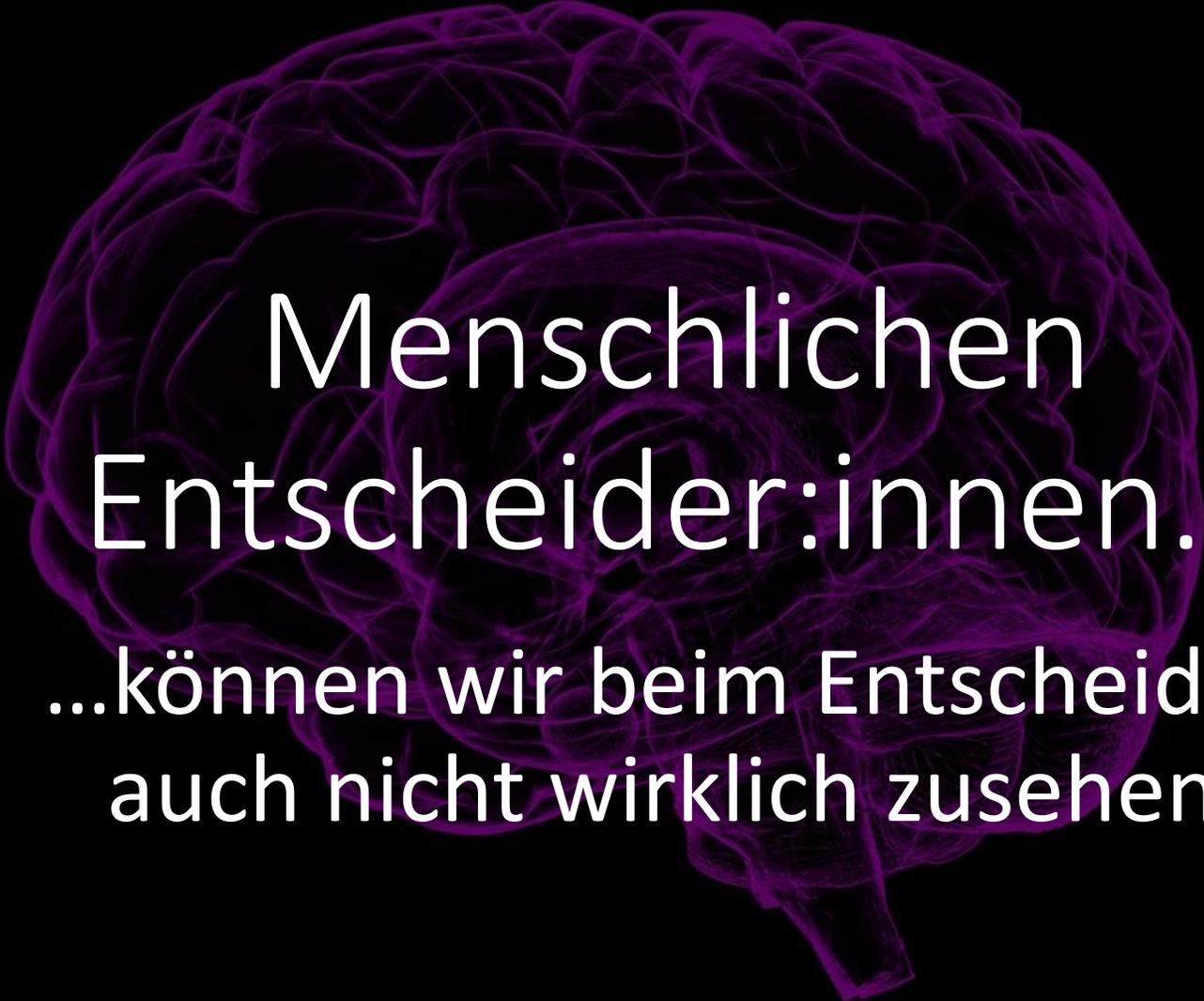
Scoring-Verfahren

Klassifikation



# Menschliche Entscheider:innen

---

A purple wireframe brain graphic, composed of many overlapping, semi-transparent lines that form the shape of a human brain. The lines are interconnected, creating a complex, mesh-like structure. The brain is positioned in the center of the image, with the text overlaid on it.

Menschlichen  
Entscheider:innen...

...können wir beim Entscheiden  
auch nicht wirklich zusehen.

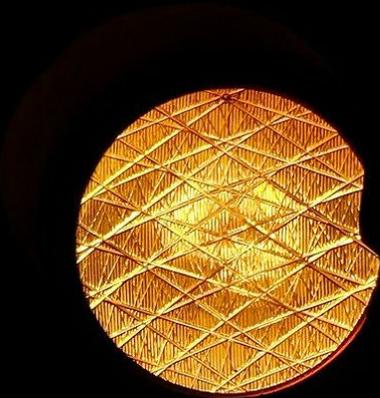
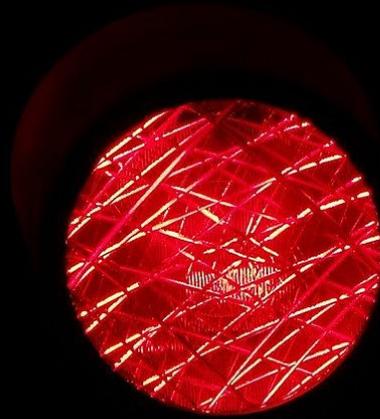
Aber wir können  
Anforderungen stellen...

- an ihre Ausbildung
- die darin erworbenen Qualifikationen
- an spezifische Inhalte
- Erfahrung

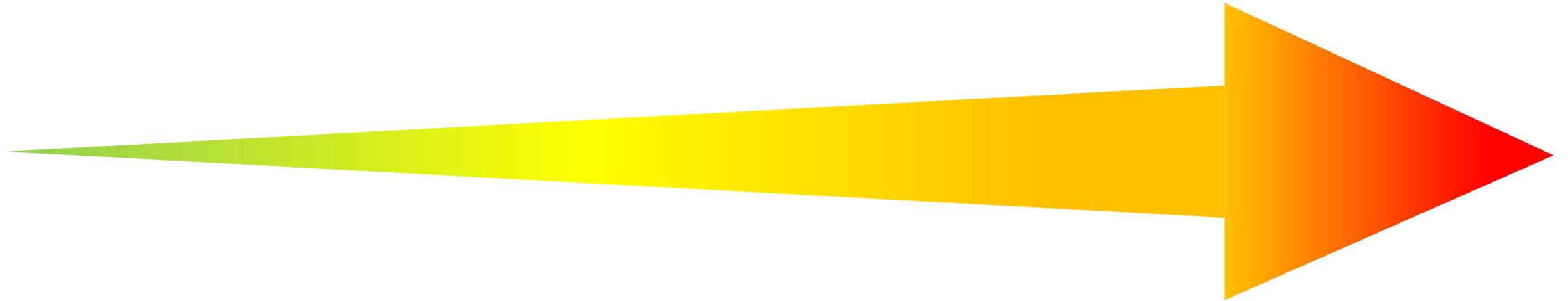


Wir können  
Entscheidungen  
absichern...

- durch Widerspruchsmöglichkeiten
- durch unabhängige Zweitgutachten
- durch Hinzuziehen mehrerer Expert:innen
- durch Hierarchien von Entscheider:innen



Mit der Wichtigkeit der Entscheidung....



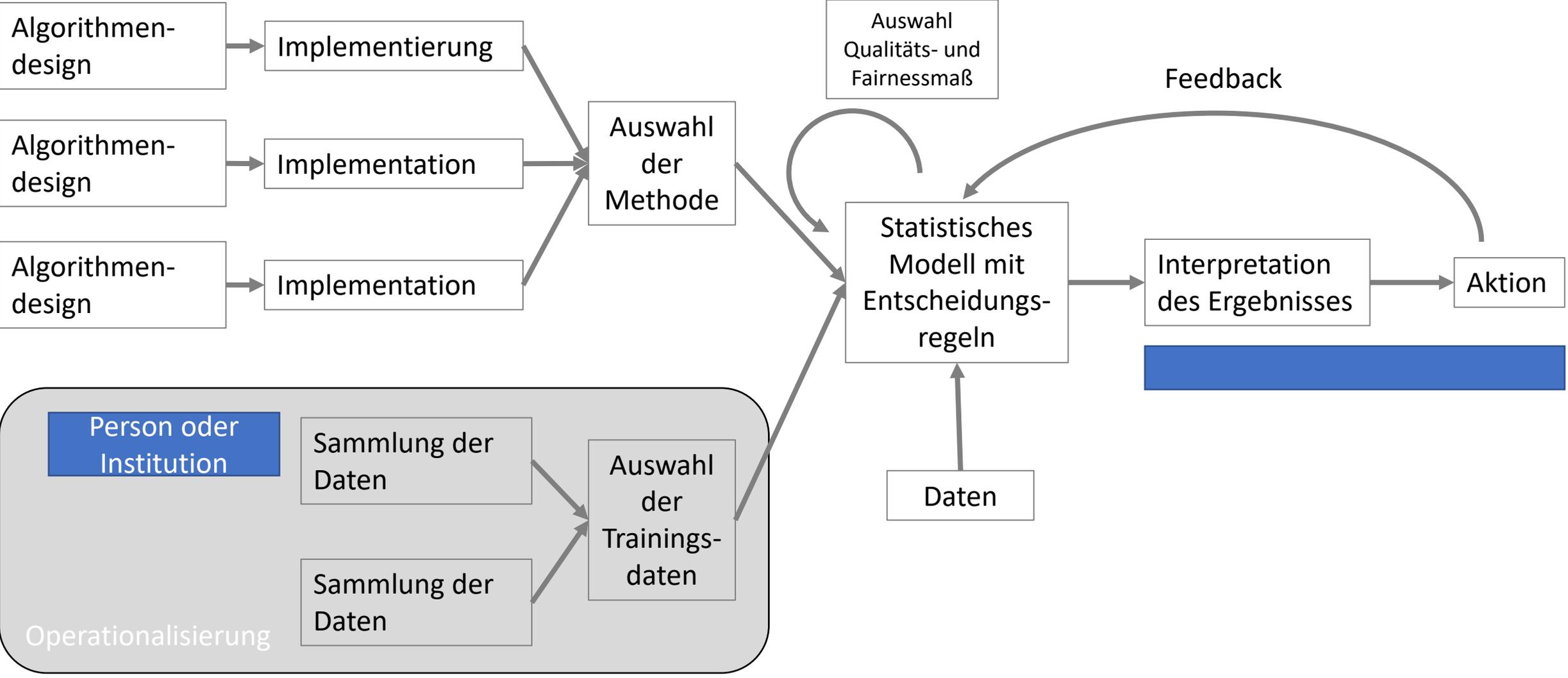
...steigen i.A. Anforderungen an die Ausbildung  
und die Möglichkeiten des Widerspruchs.



Was passiert, wenn KI entscheidet?

Data Scientist

Wissenschaftler



# Was ändert sich durch KI – wirklich?

- Unklare „Ausbildung“
- Unklare „Qualifikation“
- Fehlendes Kontextwissen
- + Dafür möglicherweise große “Erfahrung“ (Trainingsdaten)
- + Möglichkeit, nach vielen Erfahrungsmustern zu suchen
- +/- Auch kleine Muster noch verwertbar
- +/- Skalierbarkeit bis zum Monopol



# Wie bewerten bezüglich der Regulierungsnotwendigkeit? Ein Kritikalitätsmodell

## 1. Schadenspotenzial

$$\begin{aligned} & \sum \text{Schadenspotenzial für Individuen} \\ + & \sum \text{Schadenspotenzial für Gesellschaft} \end{aligned}$$

## 2. Grad der Abhängigkeit der Bewerteten von einer Entscheidung, z.B. Anbietervielzahl, Widerspruchs- und Wechselemöglichkeiten

# Wie kann die Existenz und das Ausmaß eines Schadens ermittelt werden?

## **Transparenzmechanismen**

- Transparenz über Daten
- Datenqualität, -vollständigkeit
- Methode
- Qualitäts- und Fairnessmaß
- Testszenario(s)
- Werte der Qualitäts- und Fairnessmaße
- Einsicht in Code
- ...

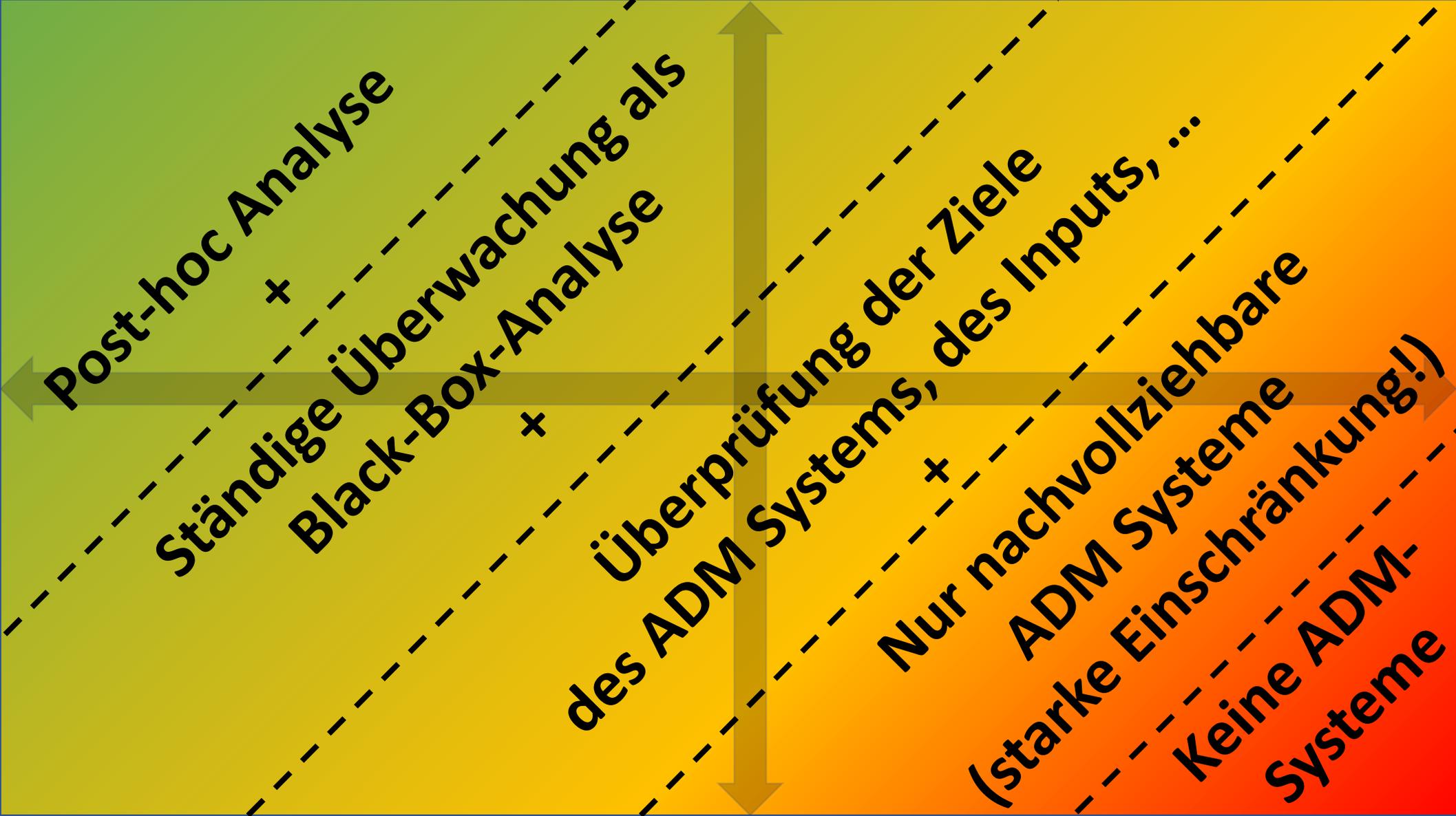
## **Nachvollziehbarkeitsmechanismen**

- Nachvollziehbarkeit der Daten
- der Datenqualität und –vollständigkeit
- der Tests
- der Möglichkeit eigener Tests
- der Bewertung von Qualität und Fairness
- ...

Geringe Abhängigkeit, z.B. viele Anbieter, einfacher Wechsel

Geringer Gesamtschaden bei Fehlurteilen

Hoher Gesamtschaden bei Fehlurteilen



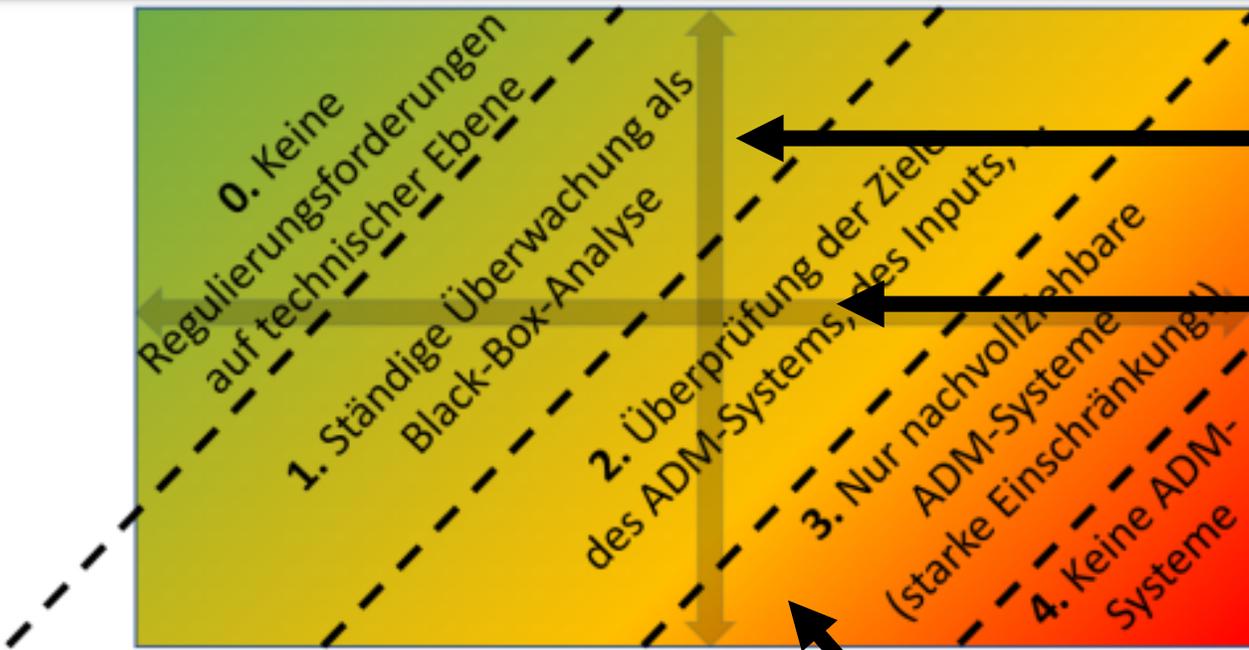
Post-hoc Analyse

Ständige Überwachung als Black-Box-Analyse

Überprüfung der Ziele des ADM Systems, des Inputs, ...

Nur nachvollziehbare ADM Systeme (starke Einschränkung!)  
Keine ADM-Systeme

Große Abhängigkeit von der Entscheidung



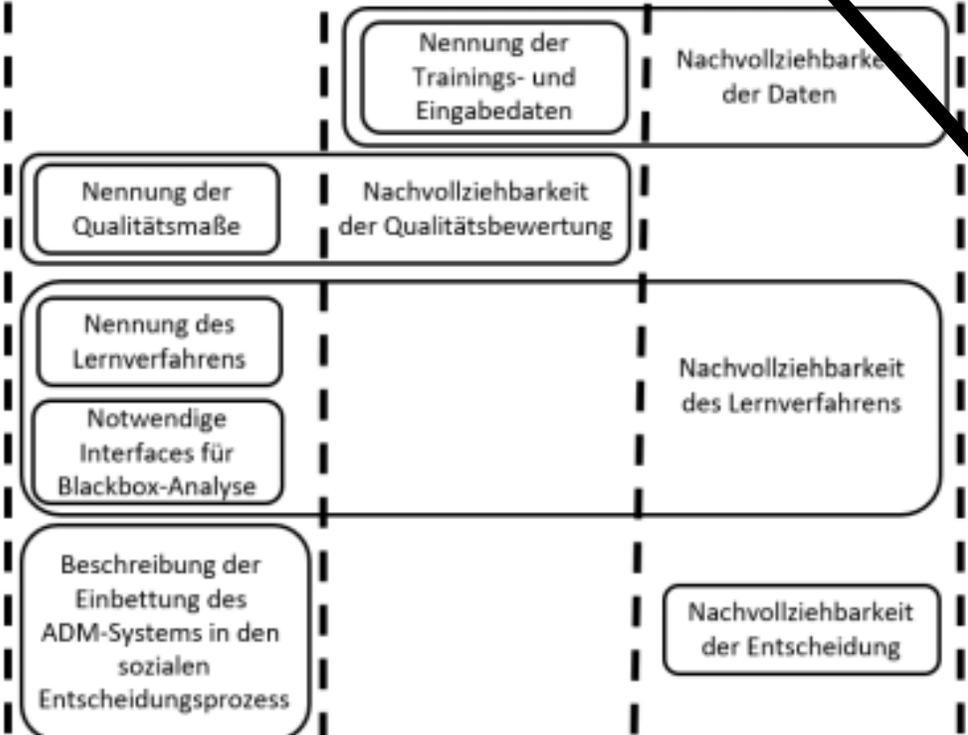
Analyse auf Diskriminierung

Krafft & Zweig (2019) fordern:

- Nennung der Art der Inputdaten
- Nachvollziehbarkeit Qualität (Fairness)

Holl, Kernbeiß & Wagner-Pinter fordern [1]:

- Nur Daten mit Kausalzusammenhang zu gewünschter Vorhersage.
- Daten müssen für Betroffene(n) leicht änderbar und löschar sein.
- Daten sind höchstens ein paar Jahre alt.



Krafft & Zweig (2019) fordern:

- Güte der Daten von außen nachvollziehbar
- Lernverfahren nachvollziehbar (Welche Methode, welche Tuningparameter?)
- Nur nachvollziehbare statistische Modelle (z.B. kein NN)

# Warum die Regulierung komplex ist

- Viele verschiedene Anwendungskontexte
- KI-Systeme ersetzen oder unterstützen Personen in unterschiedlich stark regulierten Prozessen und Positionen



# Warum sie gar nicht so komplex ist

- Weil die sozialen Prozesse mit hohem Schadenspotenzial schon heute reguliert sind.
- Weil viele Aushandlungsprozesse zur Verortung von Verantwortlichkeit schon existieren. .
- Weil die meisten Gruppen schon Schutzinstitutionen haben.





# From Principles to Practice



Möge das KI-Observatorium zu einem  
Fixpunkt in dieser Dynamik werden!

# Referenzen

- (1) Holl, Kernbeiß, Wagner-Pinter: „Personenbezogene Wahrscheinlichkeitsaussagen (»Algorithmen«) - Stichworte zur Sozialverträglichkeit, 2019
- (2) Zweig, Fischer & Lischka: „Wo Maschinen irren können“ (Serie AlgoEthik, No. 4, 2018)
- (3) Zwei Kapitel im Sammelband „(Un)Berechenbar?“ des Fraunhofer FOKUS, Kompetenzzentrum ÖFIT, 2018
  1. Zweig & Krafft: „Fairness und Qualität algorithmischer Entscheidungen“,
  2. Krafft & Zweig: „Wie Gesellschaft algorithmischen Entscheidungen auf den Zahn fühlen kann“
- (4) Krafft & Zweig: „Transparenz und Nachvollziehbarkeit algorithmenbasierter Entscheidungsprozesse“, Studie für den VZBV 2019,  
[https://www.vzbv.de/sites/default/files/downloads/2019/05/02/19-01-22\\_zweig\\_krafft\\_transparenz\\_adm-neu.pdf](https://www.vzbv.de/sites/default/files/downloads/2019/05/02/19-01-22_zweig_krafft_transparenz_adm-neu.pdf)