



# Wer soll richten: Mensch oder Maschine? Ein konstruktiver Ansatz

Prof. Dr. K.A. Zweig

TU Kaiserslautern

Algorithm Accountability Lab

@nettwwerkerin

Konstituierende Sitzung der  
Enquete-Kommission  
„Künstliche Intelligenz“ am 27.9.

---

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 ...“



Die zwei Ängste

Sie wird richten

Sie wird dichten





Aber was ist überhaupt KI?

SNV, „Eckpunkte einer nationalen Strategie für Künstliche Intelligenz“:

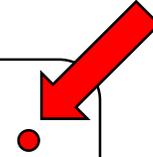
„(...) Verfahren (...),  
welche die Übertragung  
von bislang  
**menschlich getroffenen**

Entscheidungen,  
Bewertungen und  
Handlungen

**auf Computer und  
Maschinen erlauben.**

Nun, das deckt ganz schön viel ab....

Superintelligenz



Data  
Science

„KI“ im  
öffentlichen  
Diskurs

Ontologien,  
Expertensysteme

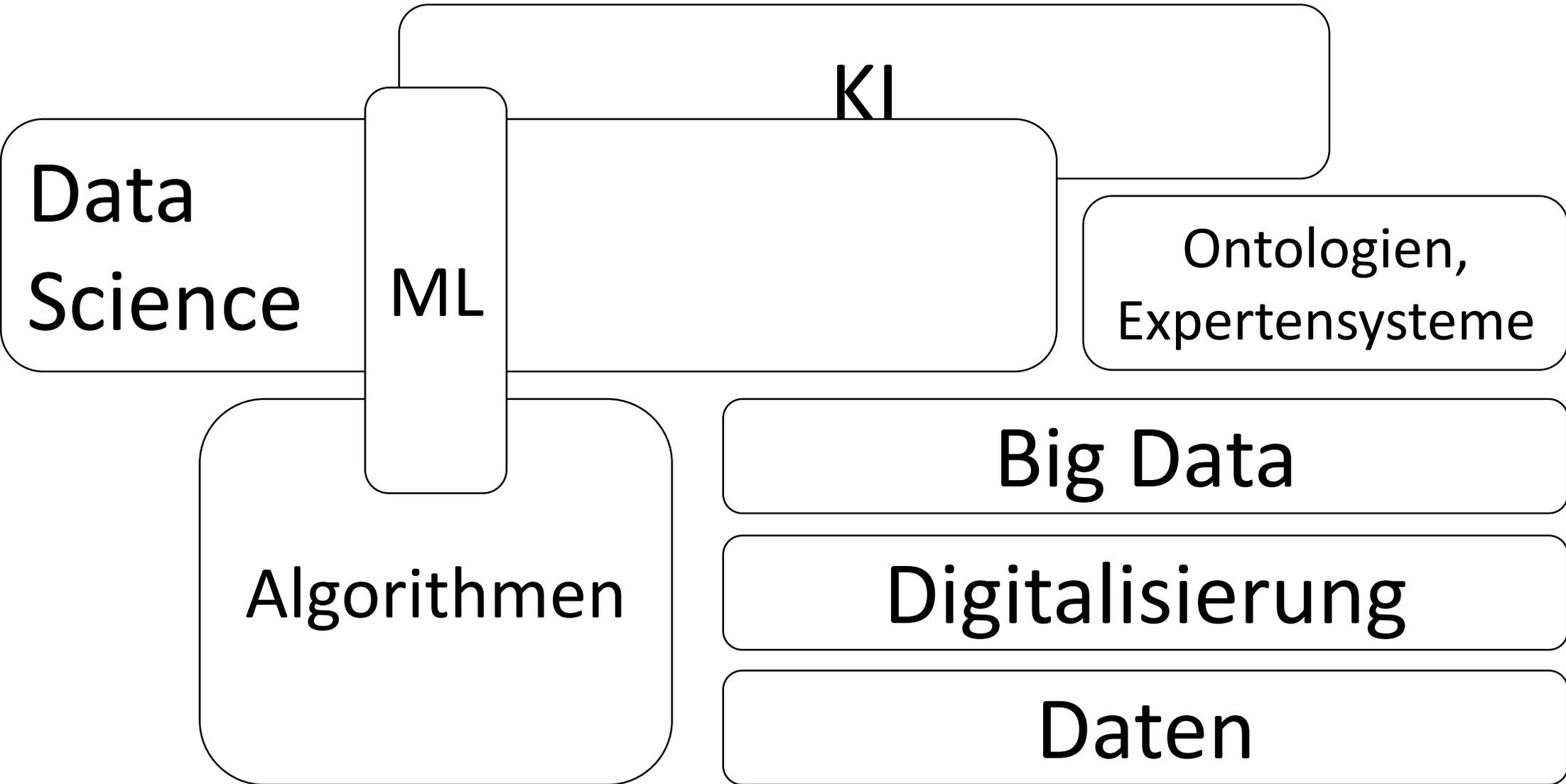
Big Data

Algorithmen

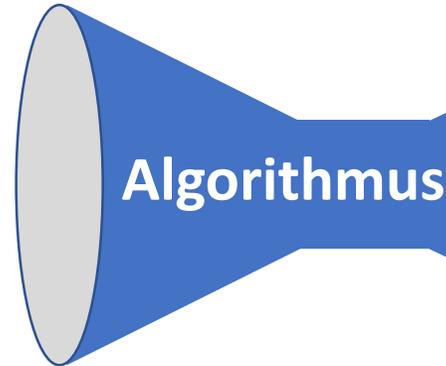
Digitalisierung

Daten

Nun, das deckt ganz schön viel ab....



# Algorithmische Entscheidungssysteme ADM Systeme



oder



Scoring-Verfahren

Klassifikation

**Forschungsschwerpunkt:  
ADM Systeme mit einer lernenden Komponente,  
die über Menschen entscheiden**

# Problemfall USA

- Zweithöchste Inhaftierungsrate weltweit.
- 6x höhere Rate von Afroamerikanern und 2x höhere Rate von Latinos als von Weißen.
- Prognose: jeder dritte afroamerikanische Junge im Alter von 10 Jahren wird eine Gefängnisstrafe absitzen müssen.



# American Civil Liberties Union



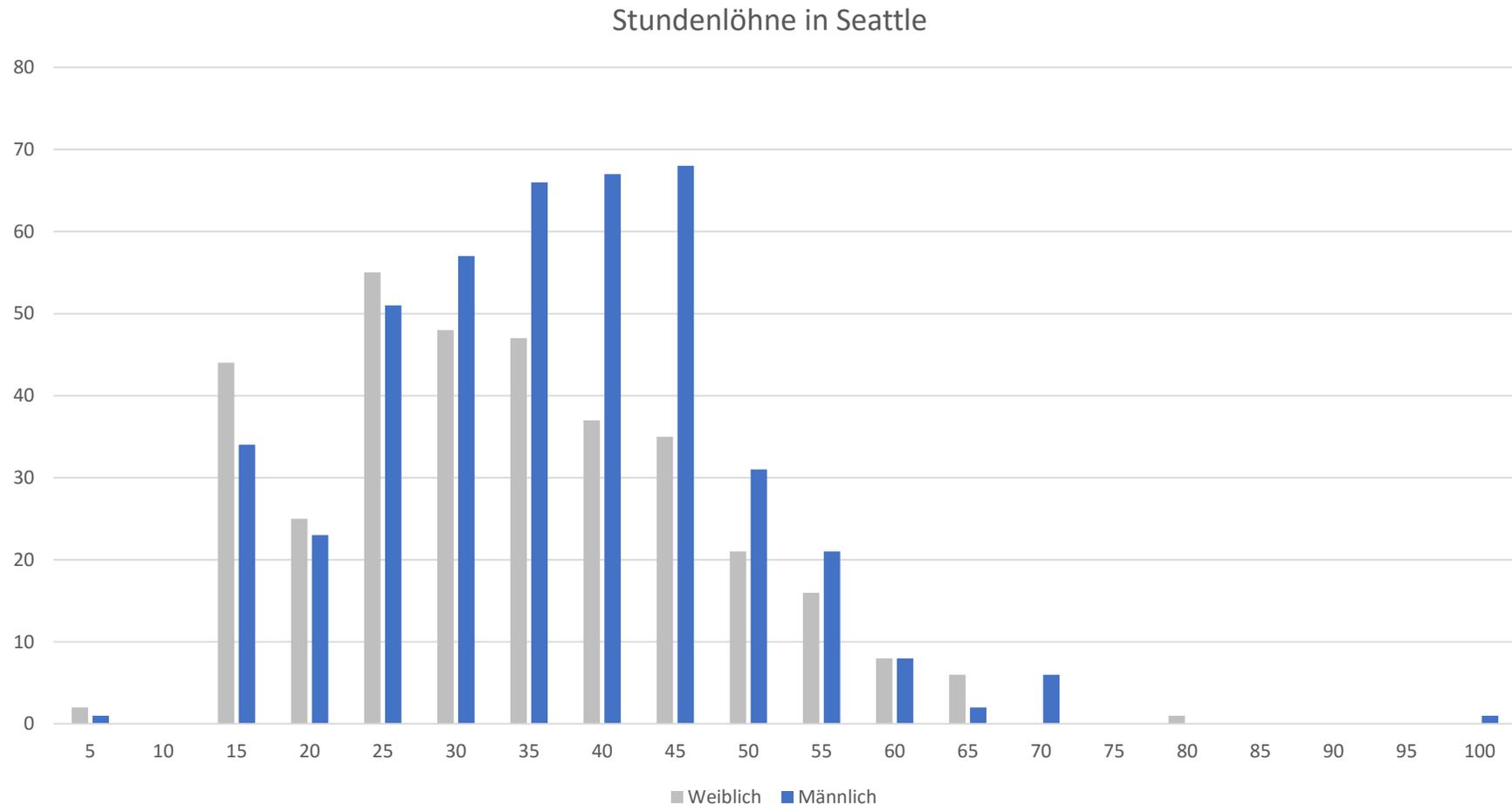
- Amerikanische Bürgerrechtsunion fordert<sup>1</sup>:
- Algorithmische Entscheidungssysteme sollten überall im Prozess eingesetzt werden, ...
- ... um Fairness und Objektivität zu sichern.
- Dazu sollen Computer aus Daten Entscheidungsregeln lernen.



Können Computer lernen?

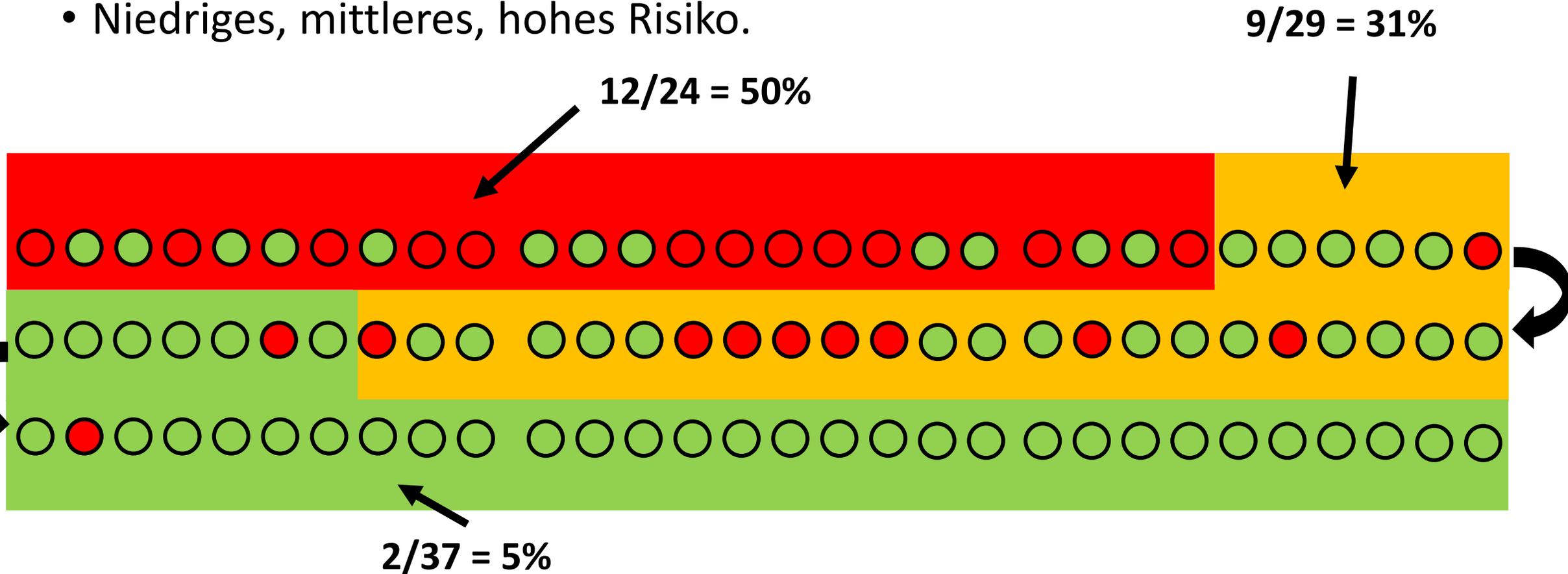
# Heißen Sie unsere(n) neue(n) Mitarbeiter(in) willkommen!

- Anteil weiblicher Angestellter?
  - 44%
- Anteil weiblicher Angestellter mit Lohn unter \$25?
  - 55%



# Vom Scoring zur Klassifikation

- ACLU fordert: Es soll drei Klassen geben.
- Niedriges, mittleres, hohes Risiko.



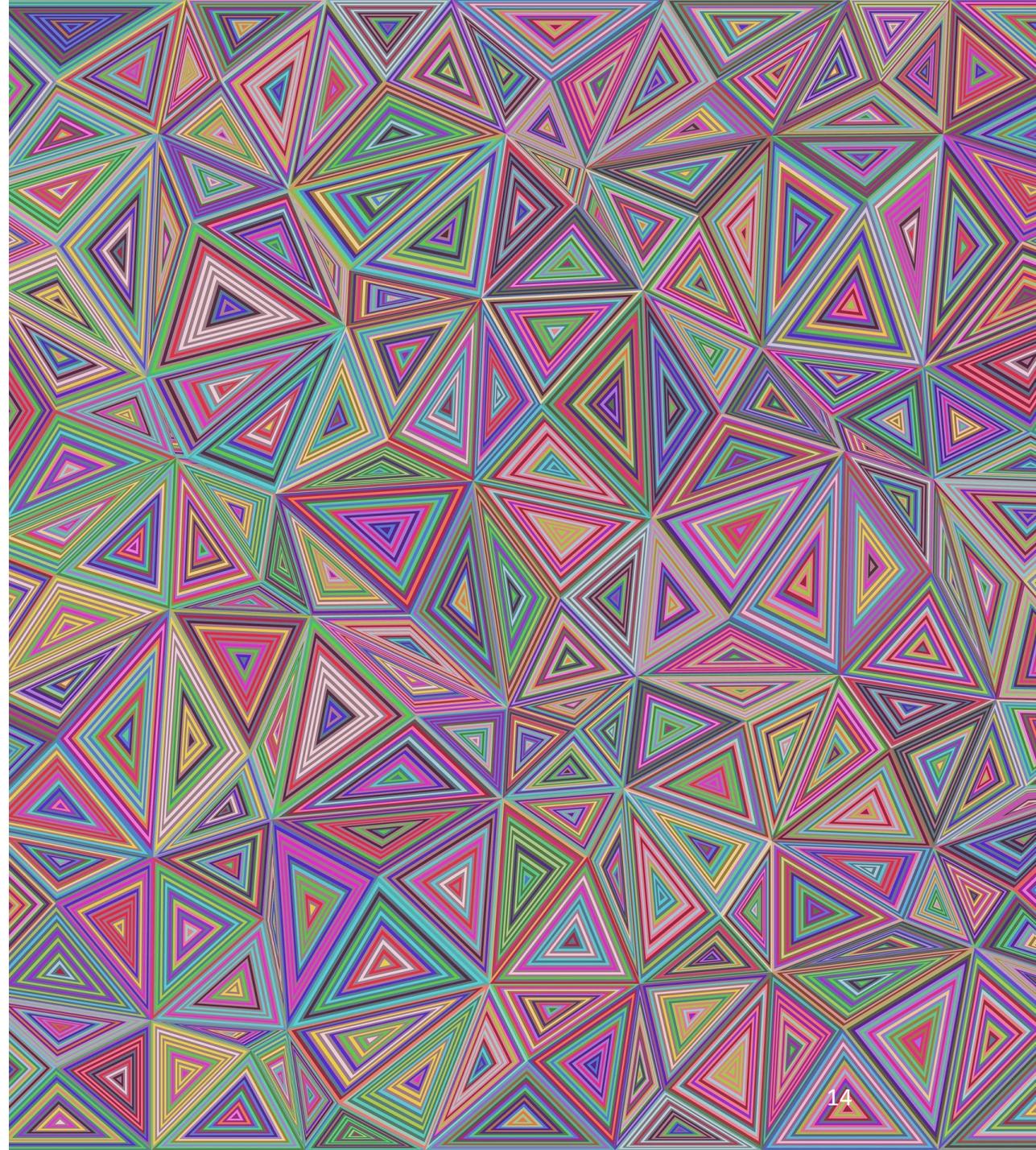
# Wie gut sind die Robo-Richter?

- Ganz schön schlecht: COMPAS
  - Hochrisiko-Kategorie:
    - Gewöhnliche Kriminaltaten: nur zu 50% richtig!
    - Schwere Straftaten: nur zu 20% richtig!
- Ein amerikanisches Terroristenidentifikationssystem tönt:
  - „Nur 0.008% falsch Positive!“
  - Bei 55 Millionen Einwohner sind das 4.400 Unschuldige, um wenige Hundert zu identifizieren.
  - Von den „Hochrisikopersonen“ also vermutlich unter 20%!



## Regel

- Algorithmen der künstlichen Intelligenz werden da eingesetzt, wo es **keine einfachen Regeln** gibt.
  - Sie sind Plan B, nicht Plan A.
- Sie suchen **Muster** in hochverrauschten Datensätzen.
- Die Muster sind daher grundsätzlich **statistischer Natur**.
- Versuchen fast immer, eine **kleine Gruppe** von Menschen zu identifizieren (Problem der **Unbalanciertheit**)



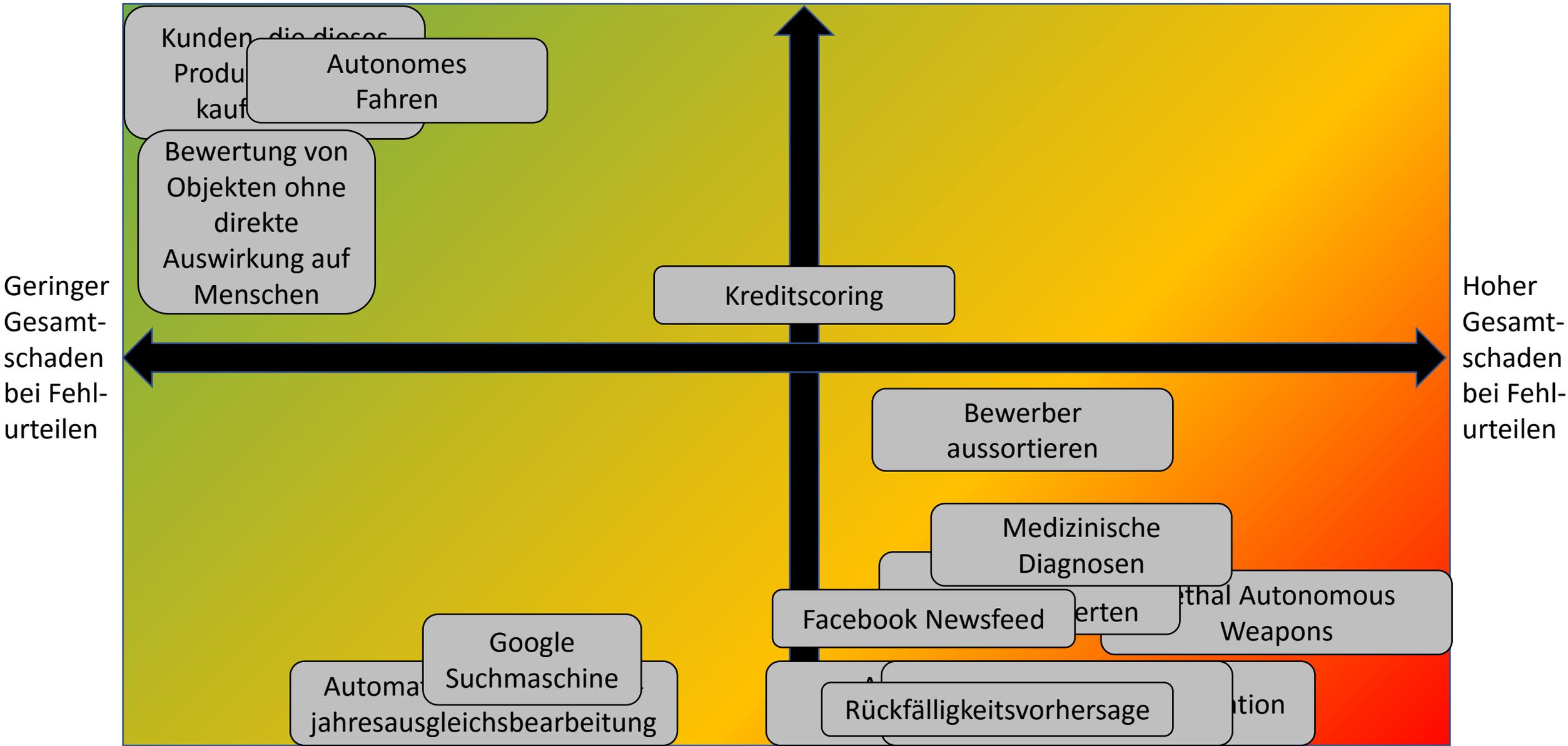
# Wie bewerten?

## 1. Schadenstiefe

$$\Sigma \quad \begin{array}{l} \text{Schaden für Individuum(Fehlurteil)} \\ +\text{Schaden für Gesellschaft(Fehlurteil)} \end{array}$$

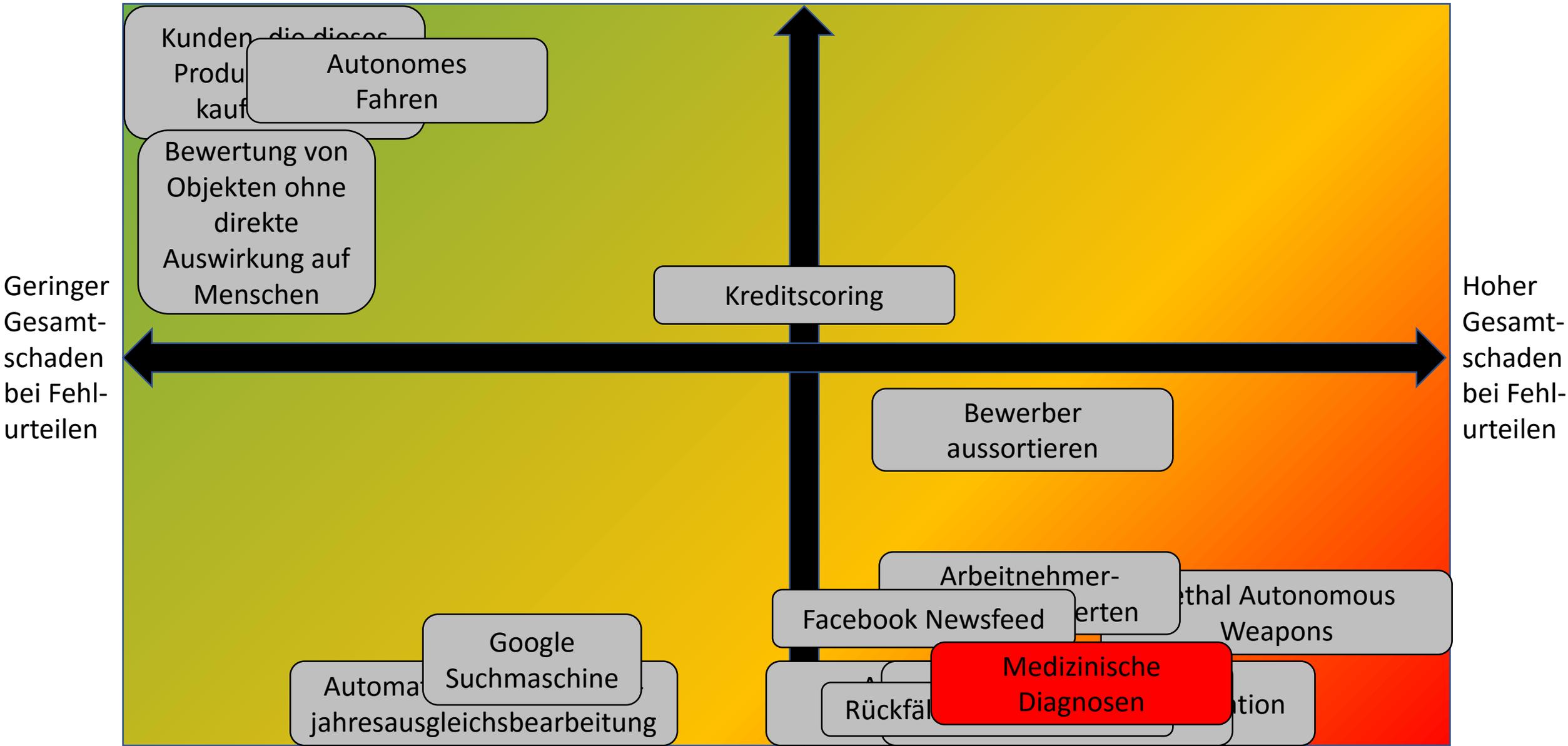
## 2. Anbietervielzahl und Wechselmöglichkeit

Viele Anbieter,  
einfacher Wechsel



Monopol

Viele Anbieter,  
einfacher Wechsel



Geringer  
Gesamt-  
schaden  
bei Fehl-  
urteilen

Hoher  
Gesamt-  
schaden  
bei Fehl-  
urteilen

Monopol

Kunden, die dieses  
Produkt  
kaufen

Autonomes  
Fahren

Bewertung von  
Objekten ohne  
direkte  
Auswirkung auf  
Menschen

Kreditscoring

Bewerter  
aussortieren

Arbeitsnehmer-  
erten

Facebook Newsfeed

lethal Autonomous  
Weapons

Automa  
jahresausgleichsbearbeitung

Google  
Suchmaschine

Rückfä

Medizinische  
Diagnosen

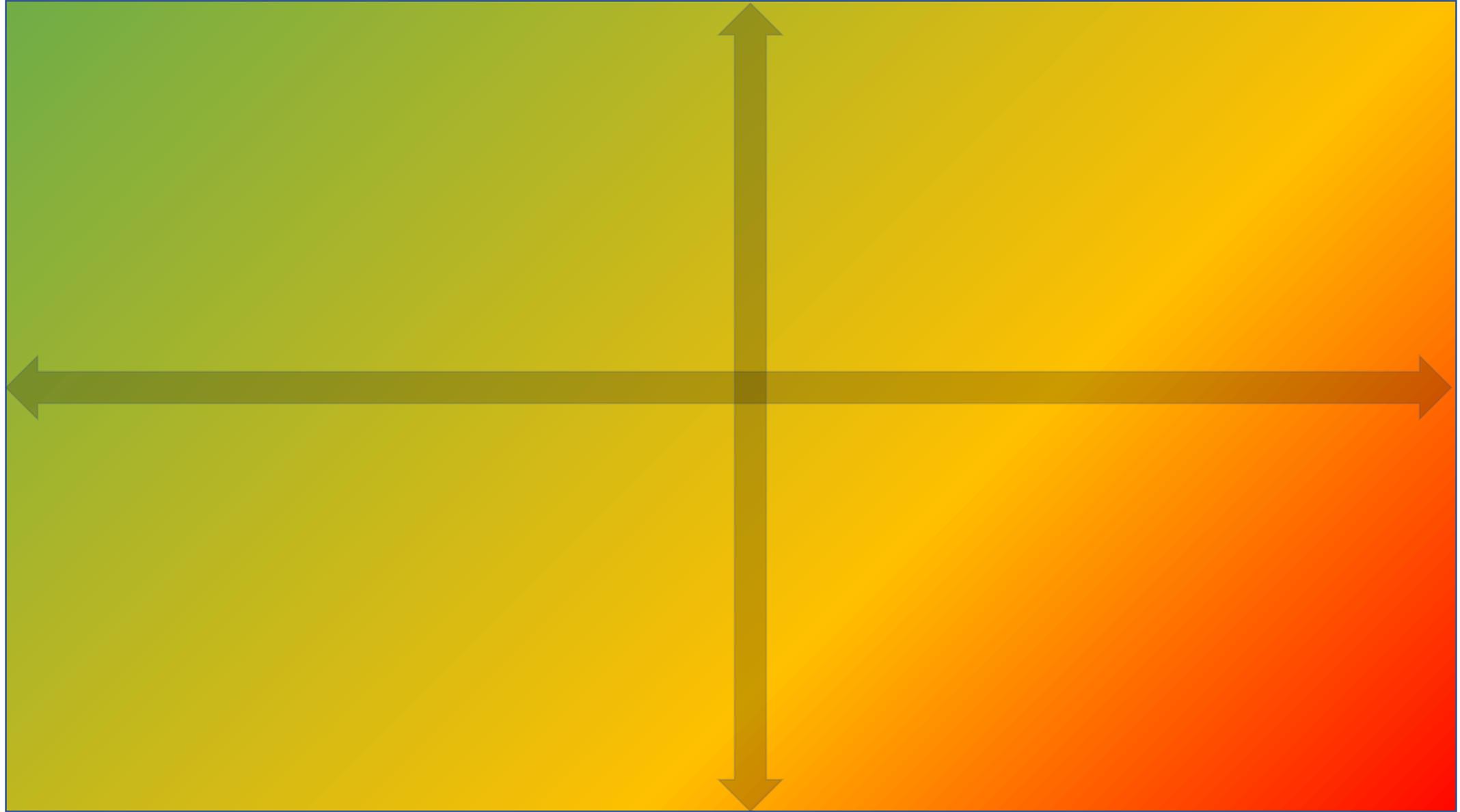
tion

Viele Anbieter,  
einfacher Wechsel

Geringer  
Gesamt-  
schaden  
bei Fehl-  
urteilen

Hoher  
Gesamt-  
schaden  
bei Fehl-  
urteilen

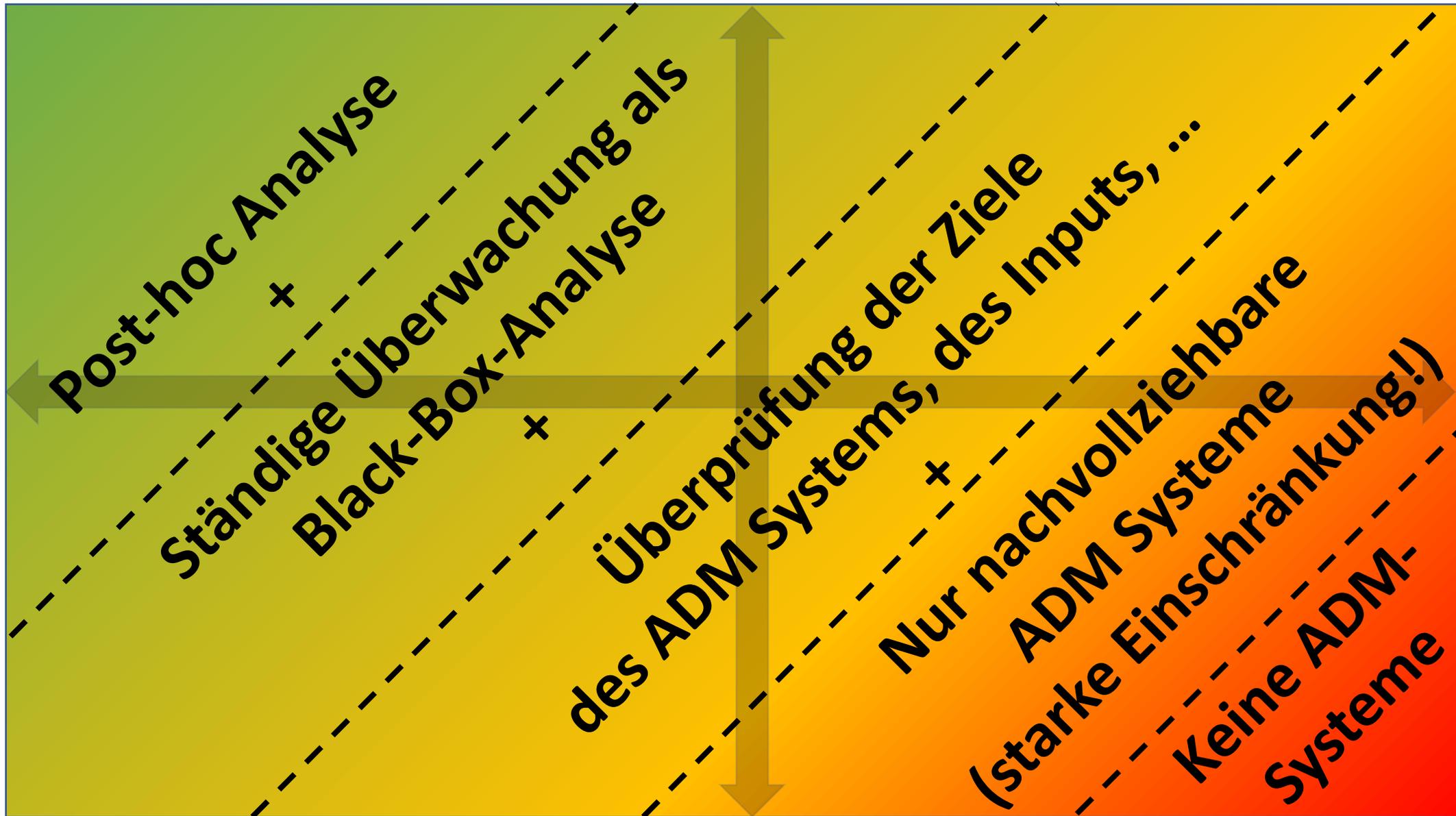
Monopol



Viele Anbieter,  
einfacher Wechsel

Geringer  
Gesamt-  
schaden  
bei Fehl-  
urteilen

Hoher  
Gesamt-  
schaden  
bei Fehl-  
urteilen



Monopol



# Zusammenfassung

---

Was sind die nächsten Schritte?

Die zwei Ängste

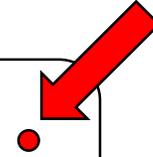
Sie wird richten

Sie wird dichten



Nun, das deckt ganz schön viel ab....

Superintelligenz



Data  
Science

„KI“ im  
öffentlichen  
Diskurs

Ontologien,  
Expertensysteme

Big Data

Algorithmen

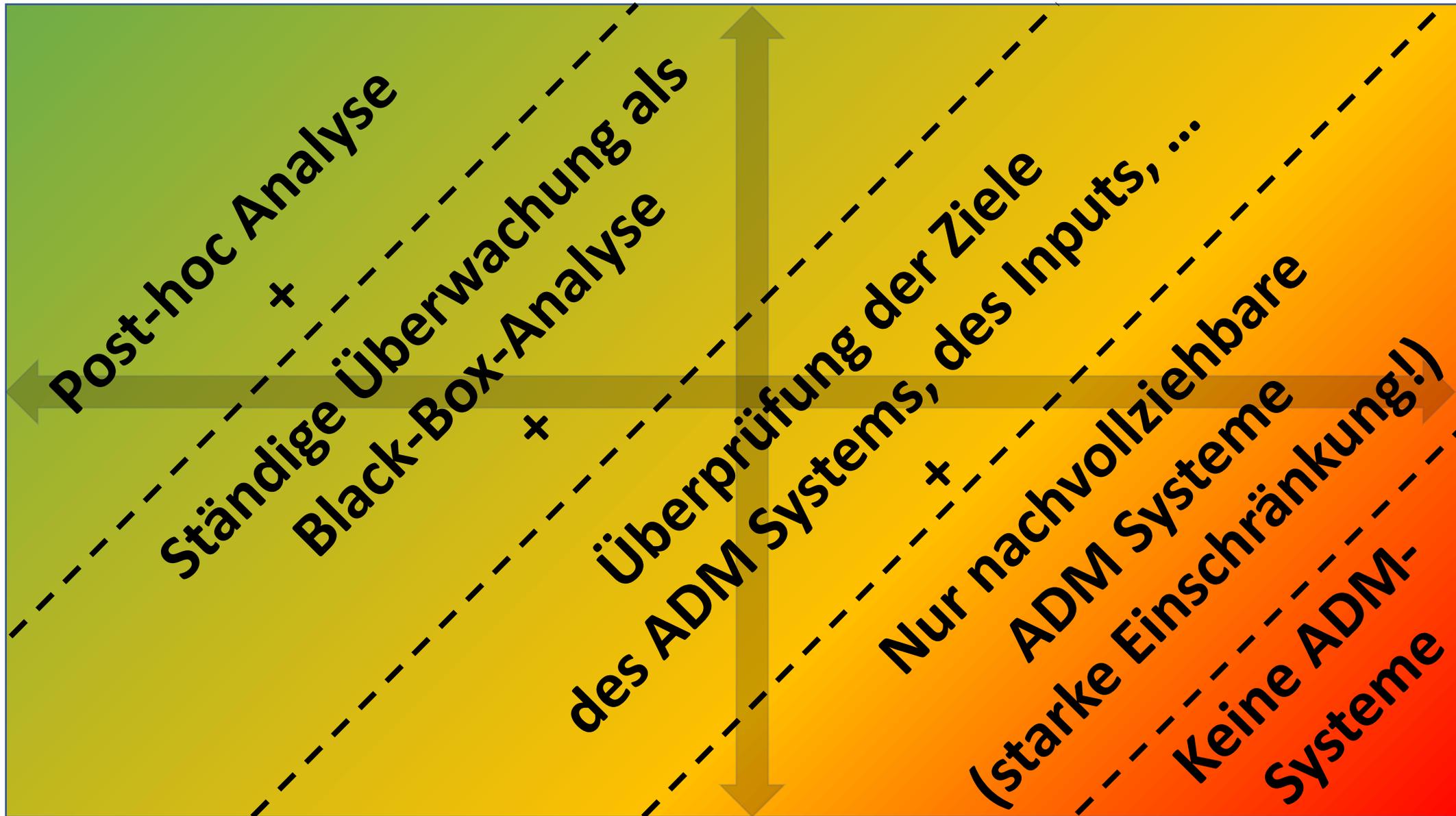
Digitalisierung

Daten

Viele Anbieter,  
einfacher Wechsel

Geringer  
Gesamt-  
schaden  
bei Fehl-  
urteilen

Hoher  
Gesamt-  
schaden  
bei Fehl-  
urteilen



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

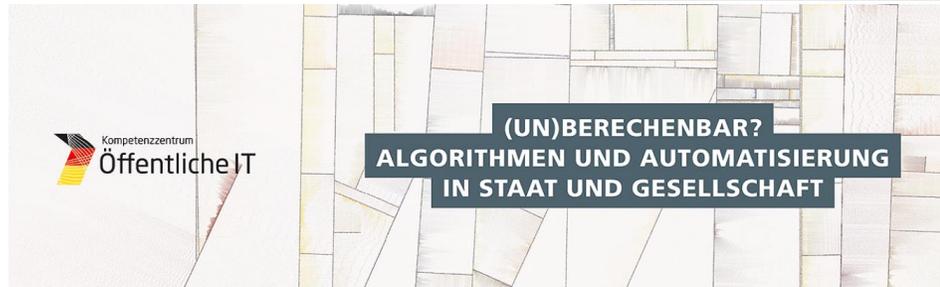
Monopol

# Weitere Informationen

Prof. Dr. Katharina A. Zweig

[zweig@cs.uni-kl.de](mailto:zweig@cs.uni-kl.de)

@nettwwerkerin bei Twitter



1. [Broschüre der Bayerischen Landesmedienanstalt](#) (Zweig, Krafft & Hauer, 2016): „Dein Algorithmus - meine Meinung“
2. [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“](#)
3. Studie für die Bertelsmann-Stiftung: Zweig, Fischer & Lischka: [„Wo Maschinen irren können“](#) (Serie AlgoEthik, No. 4, 2018)
4. Studie vom Fraunhofer FOKUS, Kompetenzzentrum Öffentliche IT (ÖFIT): Opiela, Mohabbat Kar, Thapa & Weber: [Exekutive KI 2030 – Vier Zukunftsszenarien für Künstliche Intelligenz in der öffentlichen Verwaltung](#), 2018)