



# Algorithmische Entscheidungen: Transparenz und Kontrolle

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 ...“



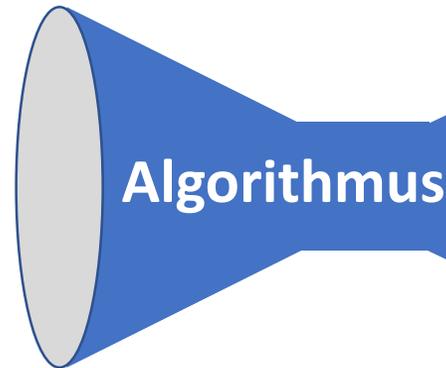
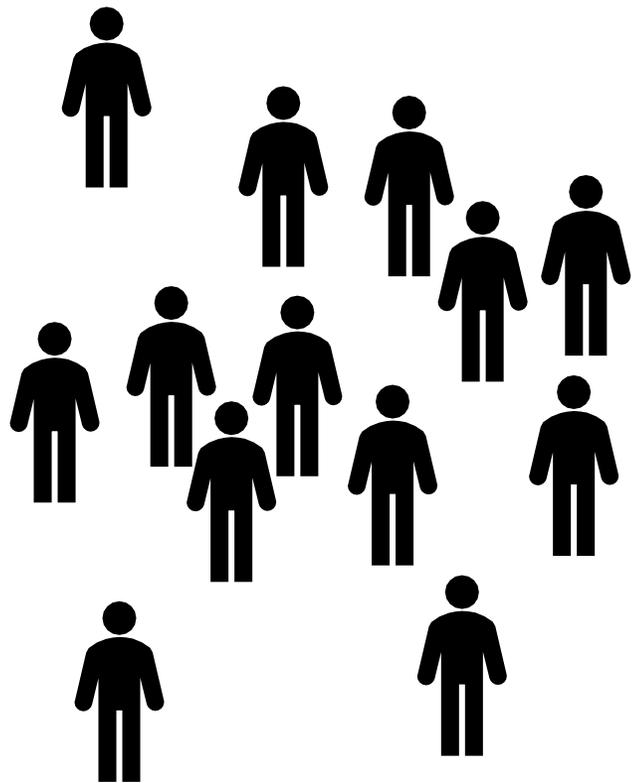
Die zwei Ängste

Sie wird dichten

Sie wird richten



# Algorithmische Entscheidungssysteme (ADM Systeme)



oder

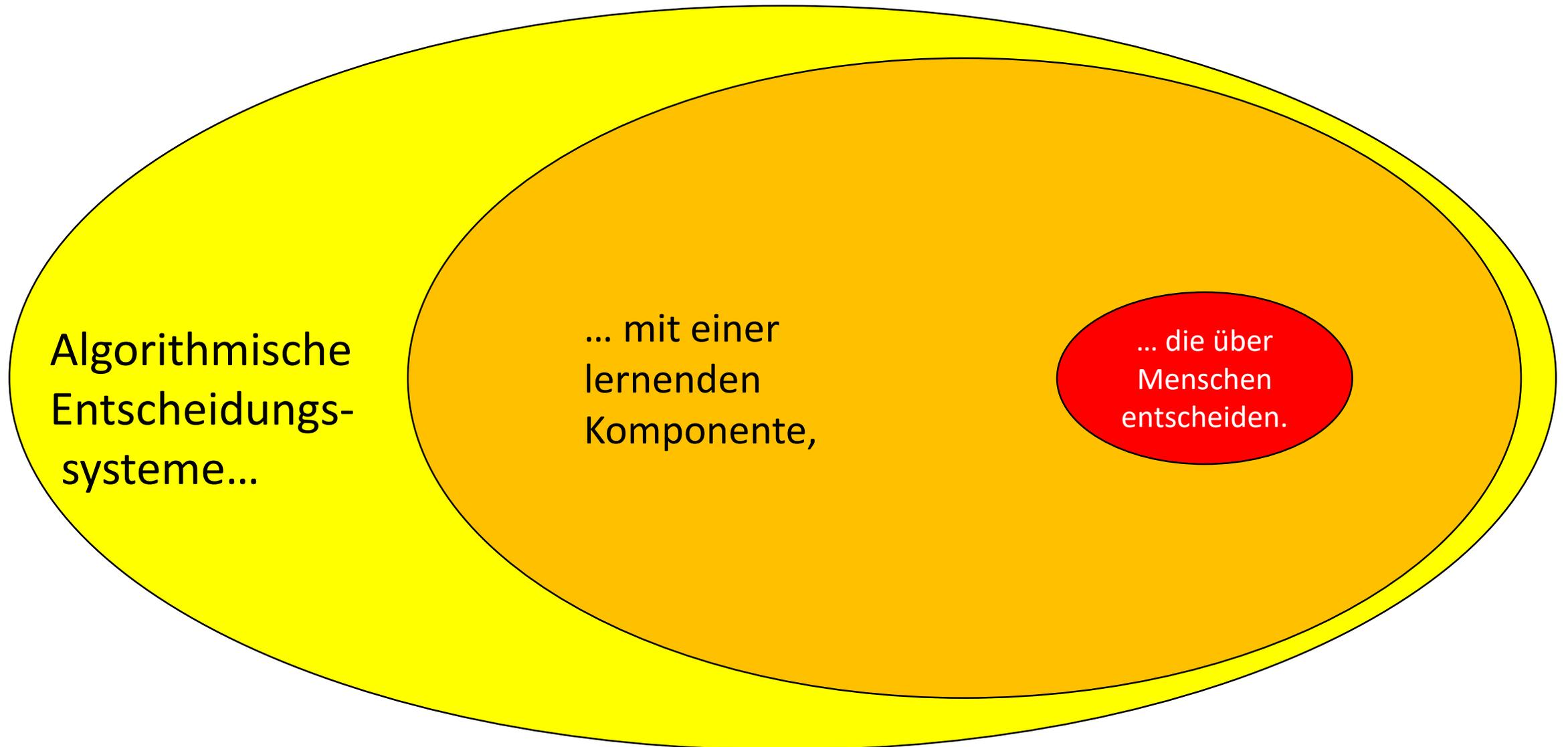


Scoring-Verfahren

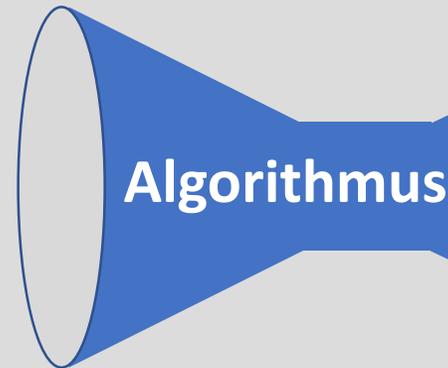


Klassifikation

# Welche ADM-Systeme sind problematisch?



# Algorithmische Entscheidungssysteme (ADM Systeme)



Nutzer  
oder

Soziales System

Scoring-Verfahren



Wie „lernt“ das System von Daten?

**DIY:**

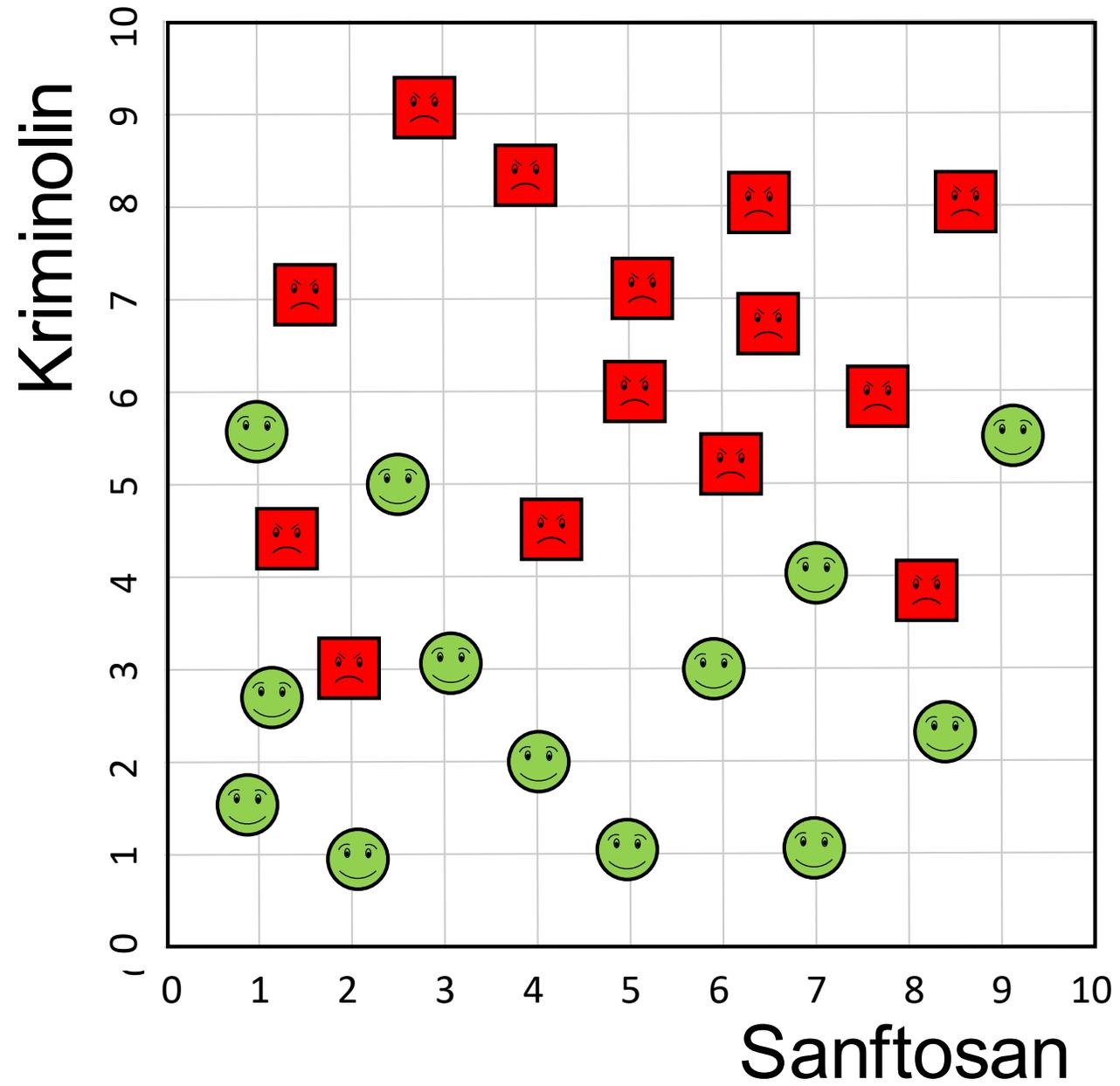
**Sie sind heute meine  
„Support Vector Machine“**



Bösartige Kriminelle



Unschuldige Bürger





Bösartige Kriminelle

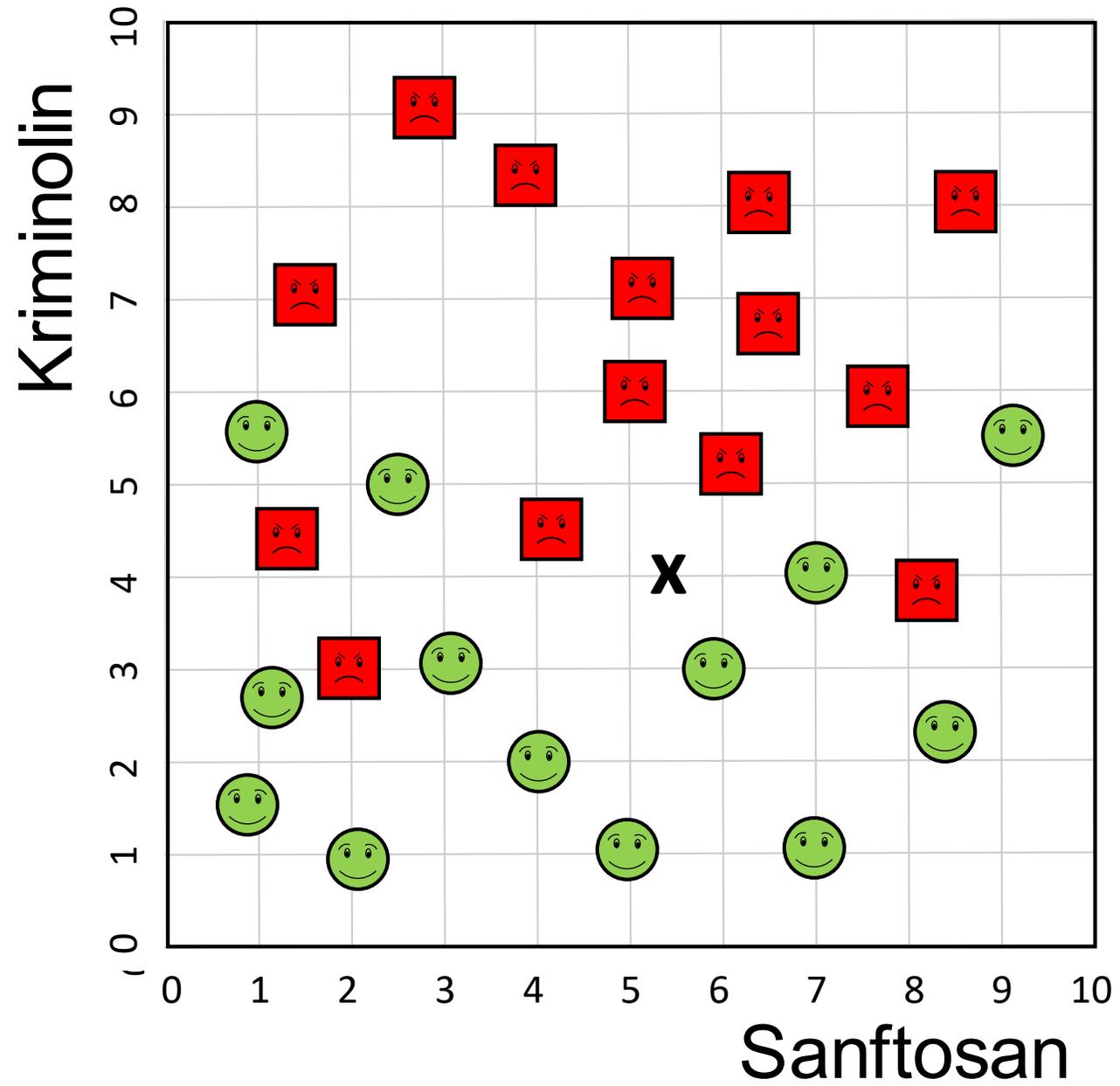


Unschuldige Bürger

Bewerten Sie Frau Müller:

5.5 Sanftosan

4.0 Kriminolin

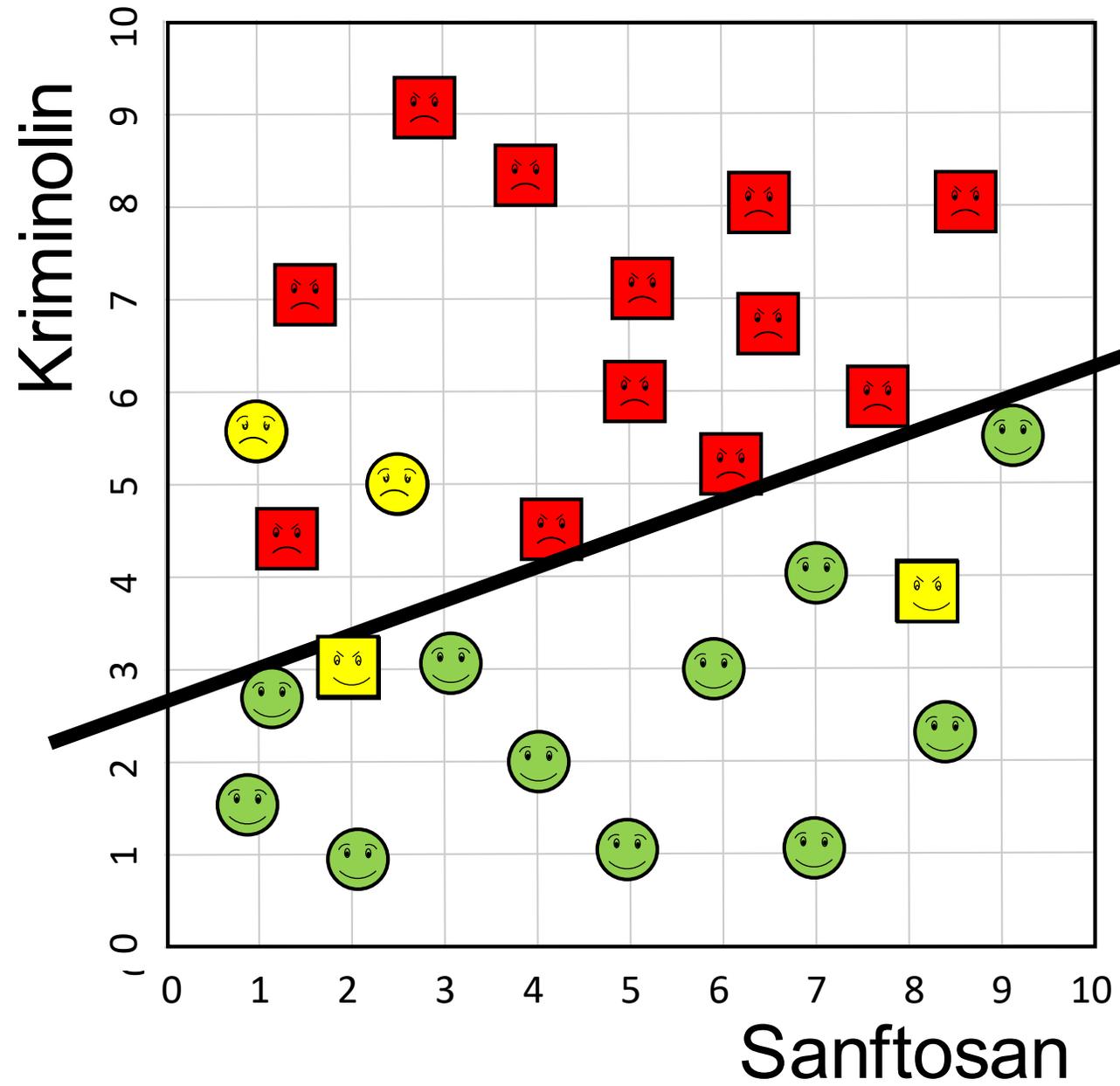




Bösartige Kriminelle

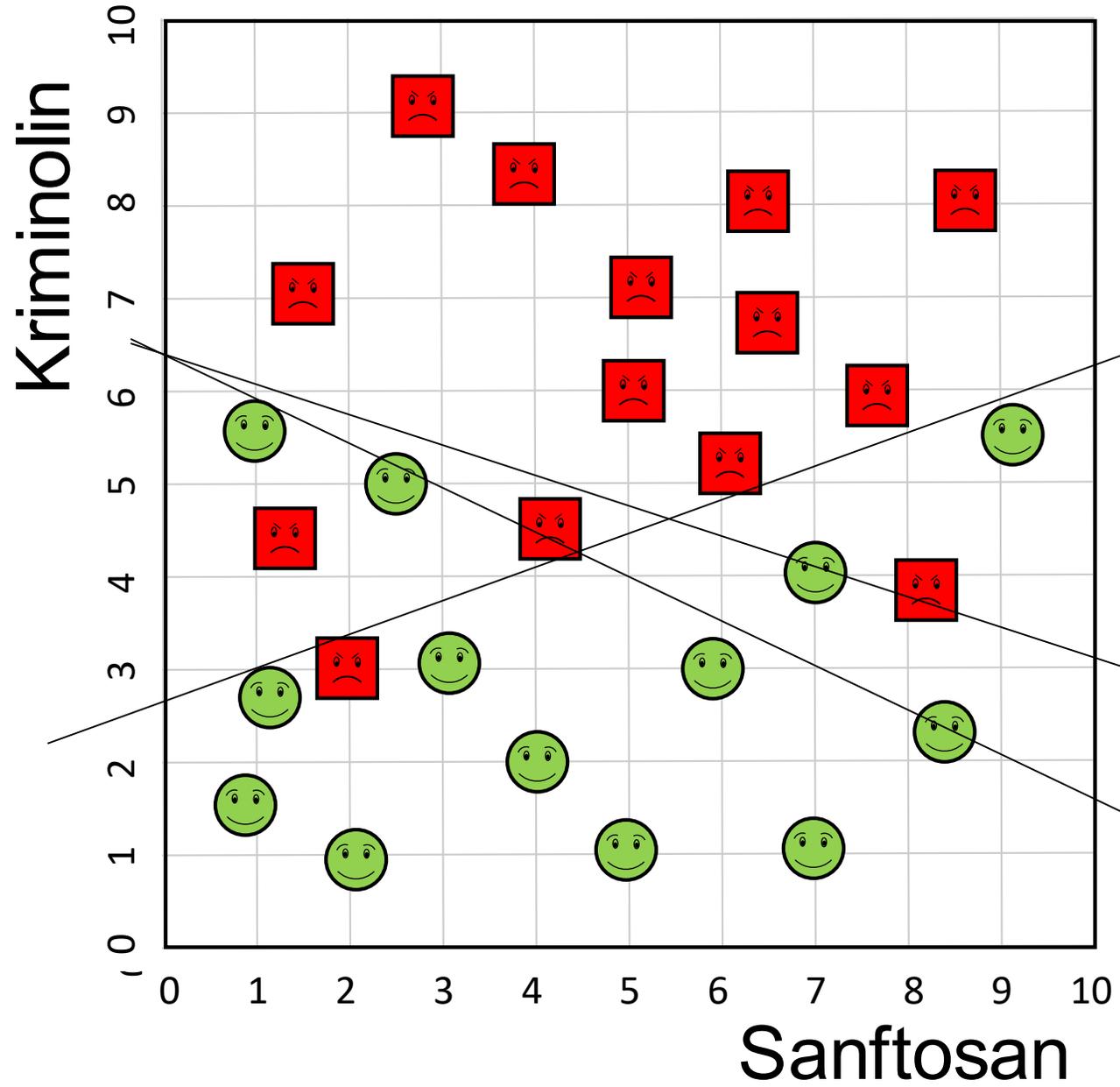
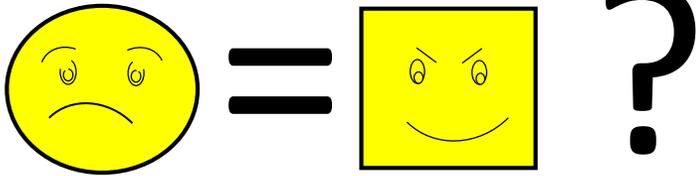


Unschuldige Bürger



 Böartige Kriminelle

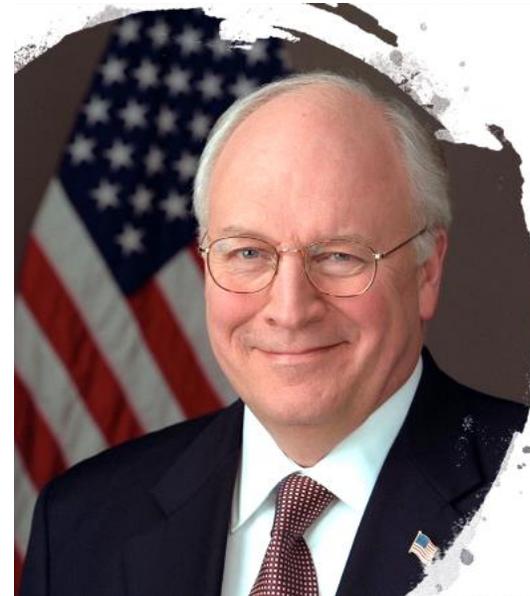
 Unschuldige Bürger





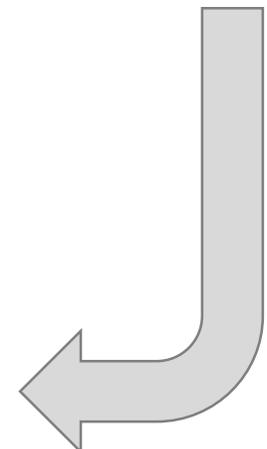
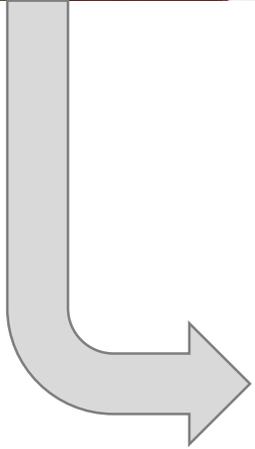
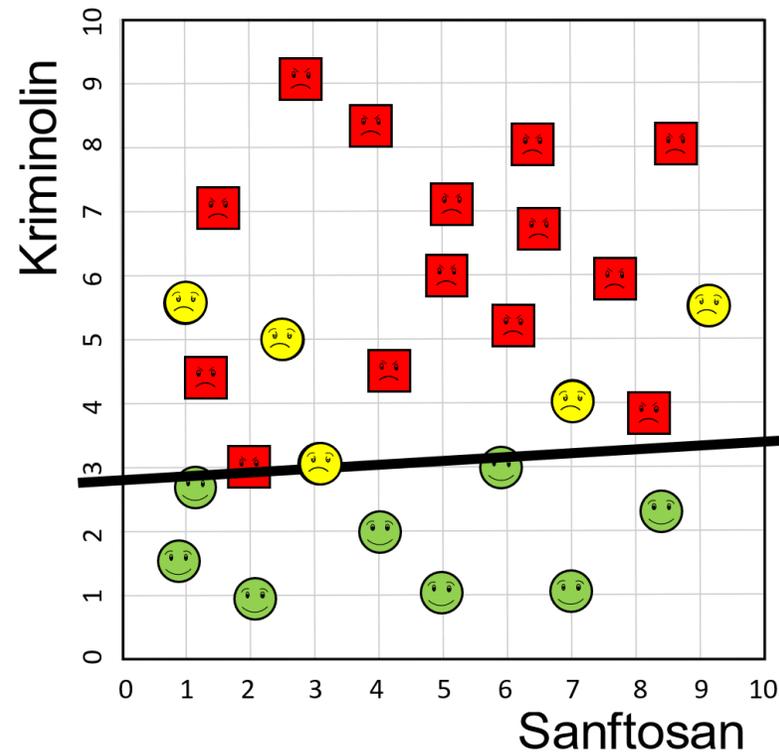
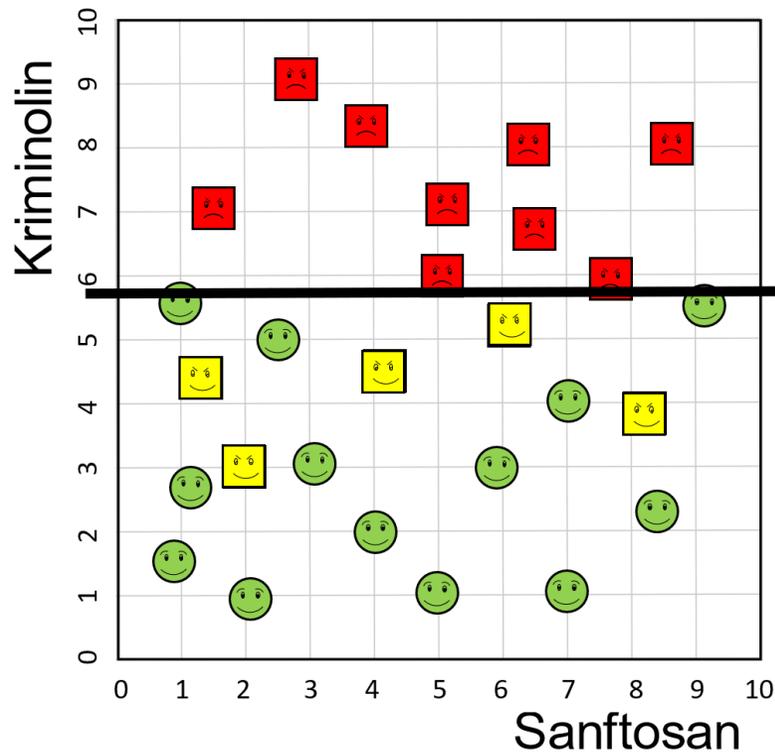
„It is better that ten guilty persons escape than that **one** innocent suffer.“

William Blackstone, Rechtsphilosoph, 1760



"I am more concerned with bad guys who got out and released than I am with a few that, in fact, were innocent."

Dick Cheney, ehemaliger Vizepräsident der USA,

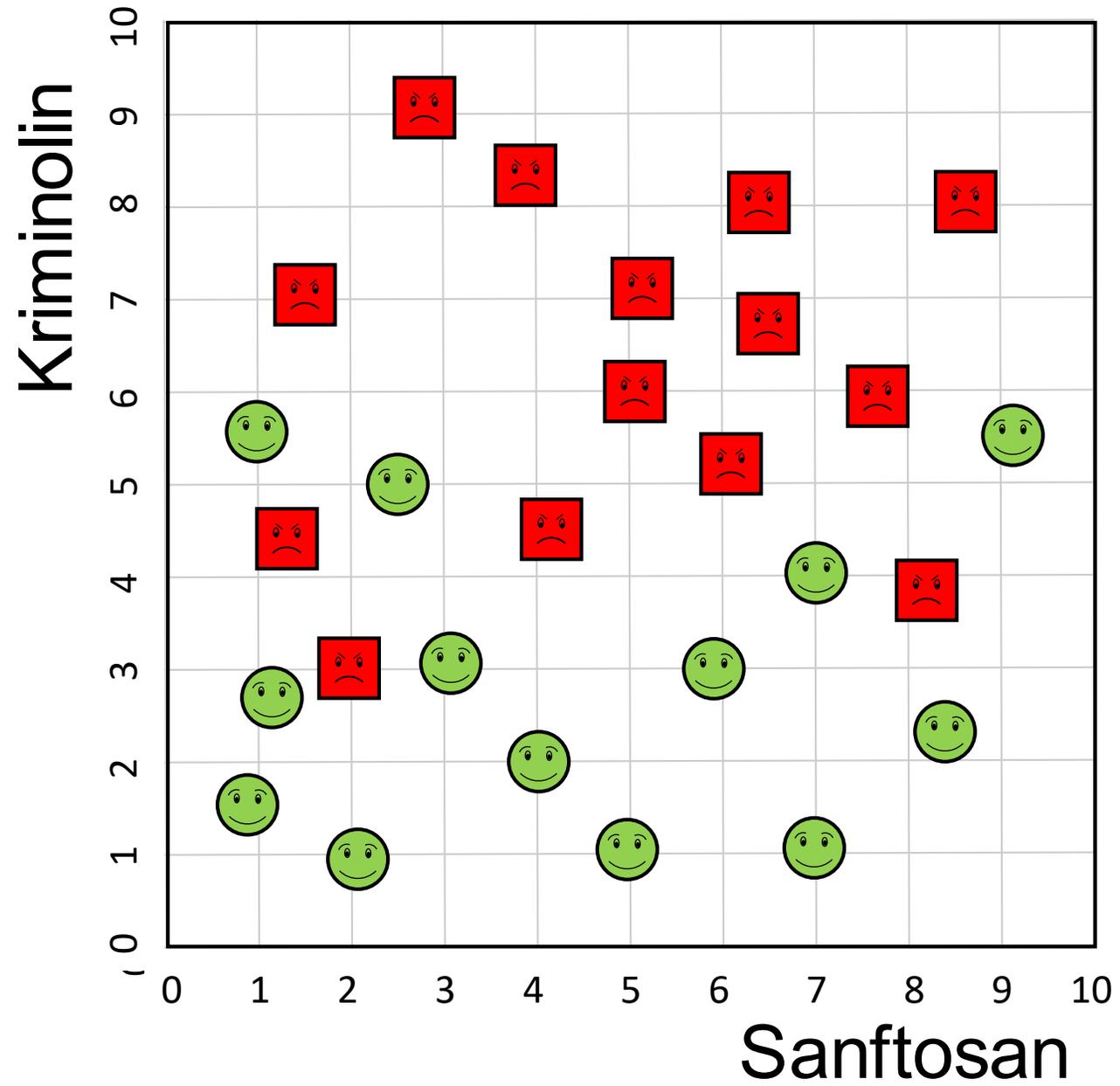




**Terroristen**



**Unschuldige Bürger**

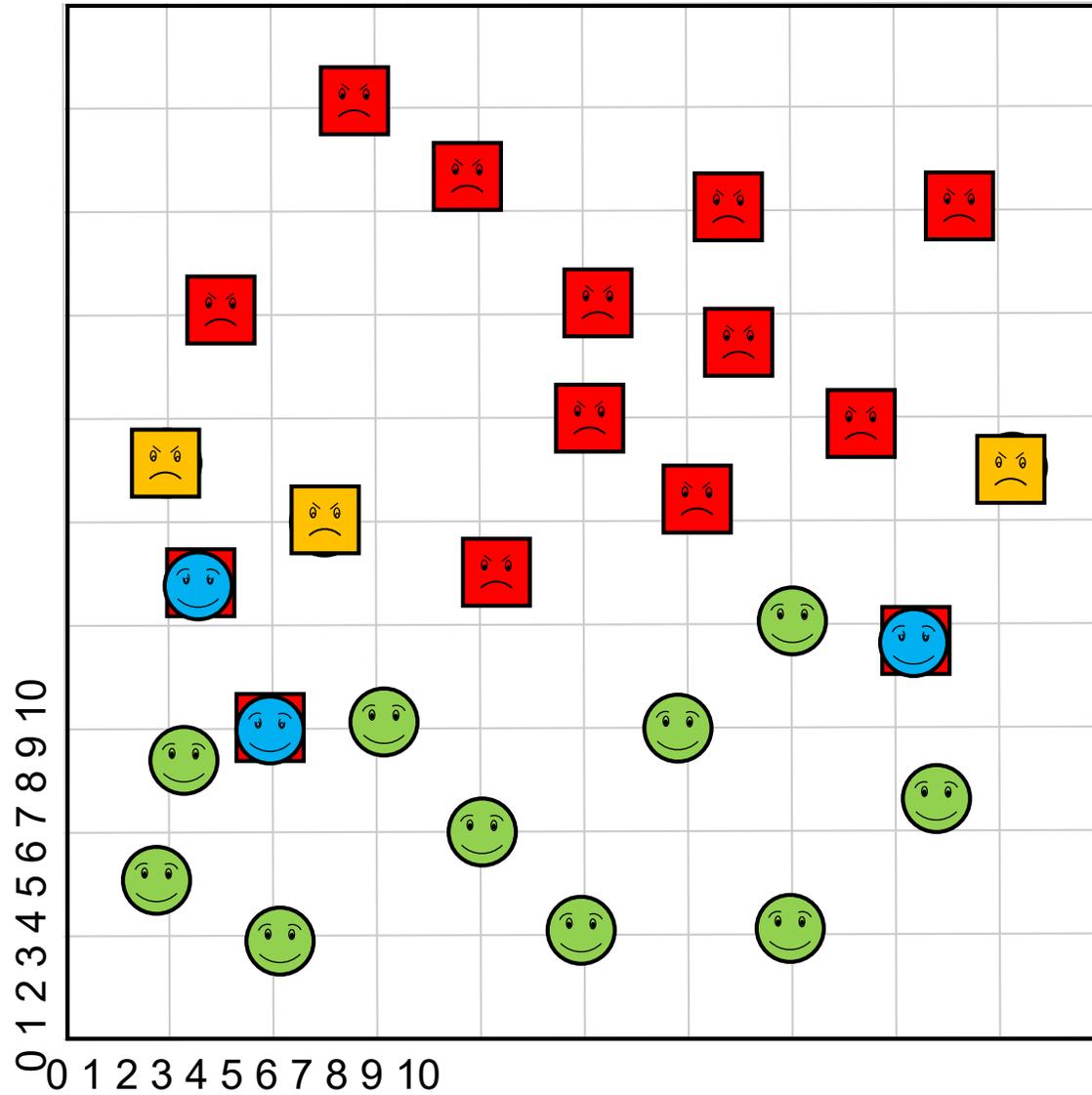


# Datenqualität

Kriminolin

 Noch nicht entdeckte Finanzbetrüger

 Unschuldig im Gefängnis



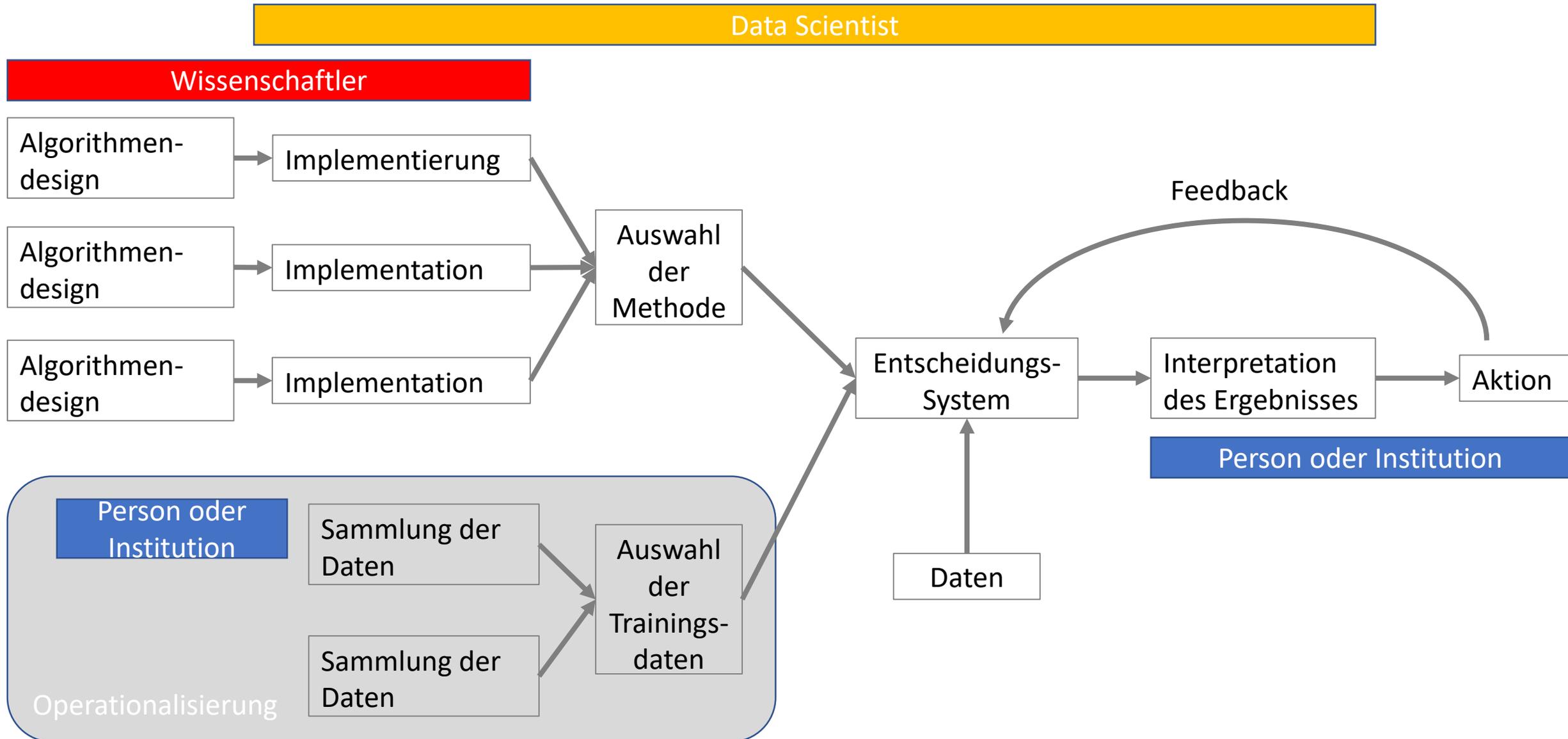
Sanftosan

# 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%!
- Im medizinischen Bereich teilweise besser als Doktoren!



# Lange Kette der Verantwortlichkeiten



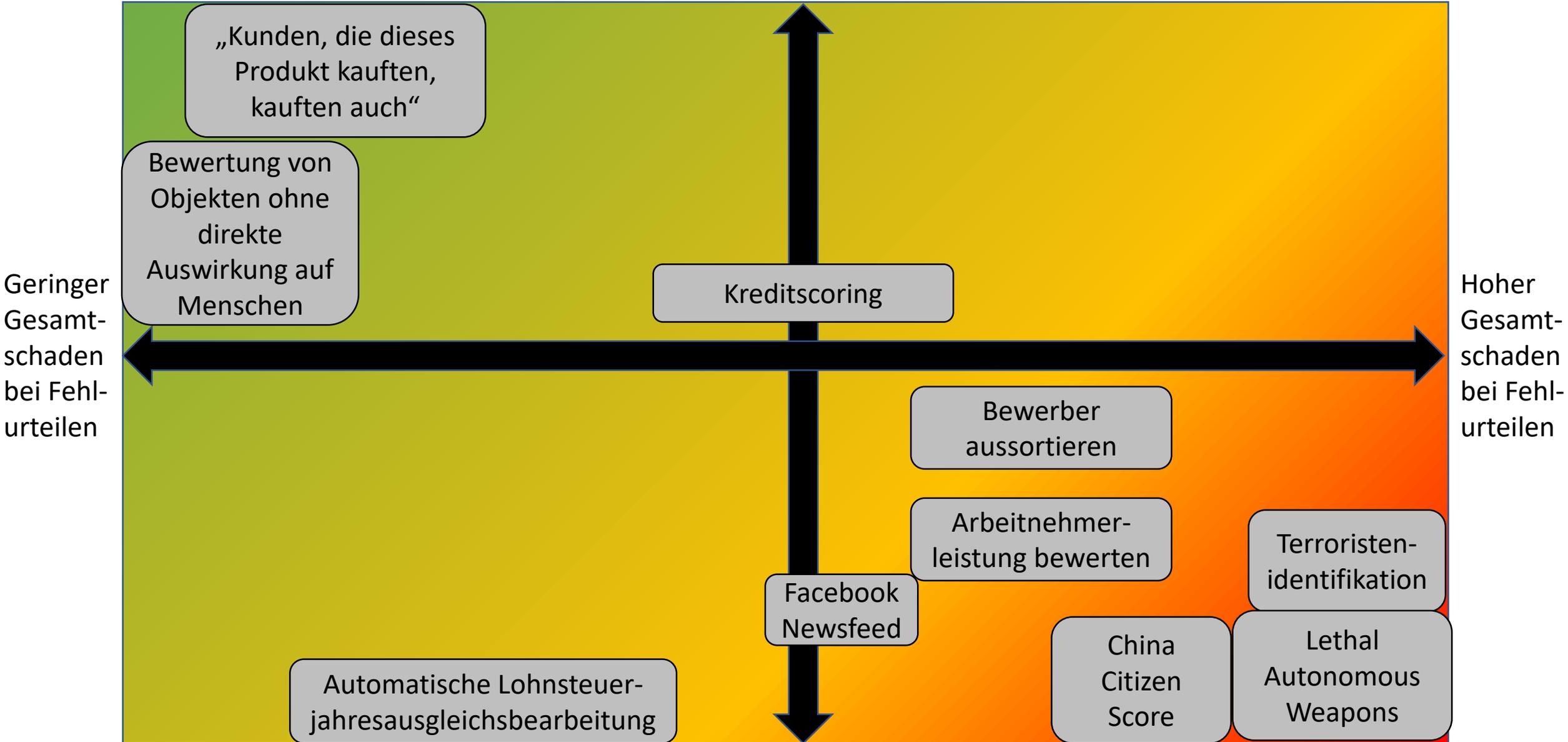
# Wie bewerten bezüglich der Regulierungsnotwendigkeit?

## 1. Schadenstiefe

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

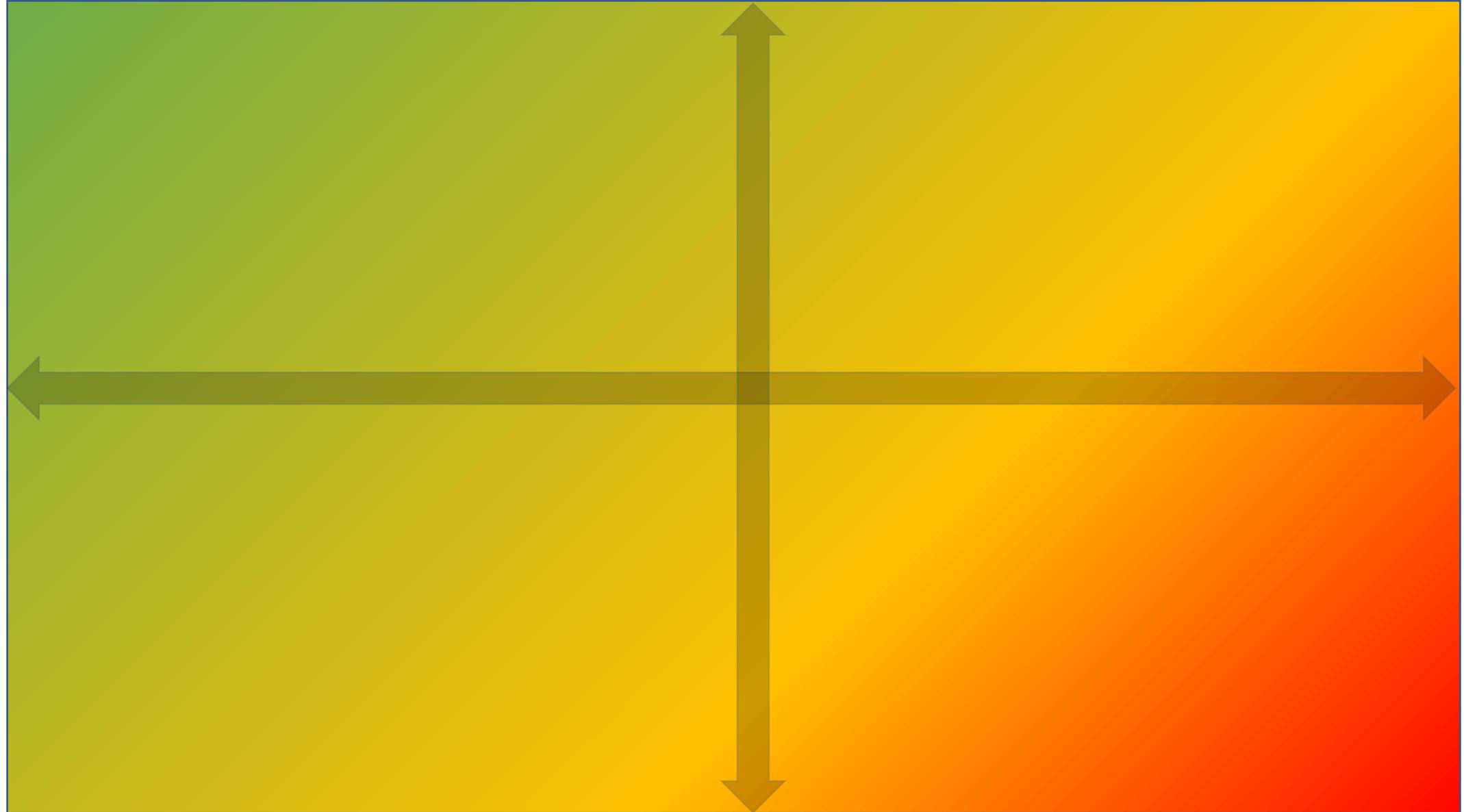
## 2. Anbietervielfzahl und Wechselmöglichkeit

Viele Anbieter,  
einfacher Wechsel



Monopol

Viele Anbieter,  
einfacher Wechsel



Monopol

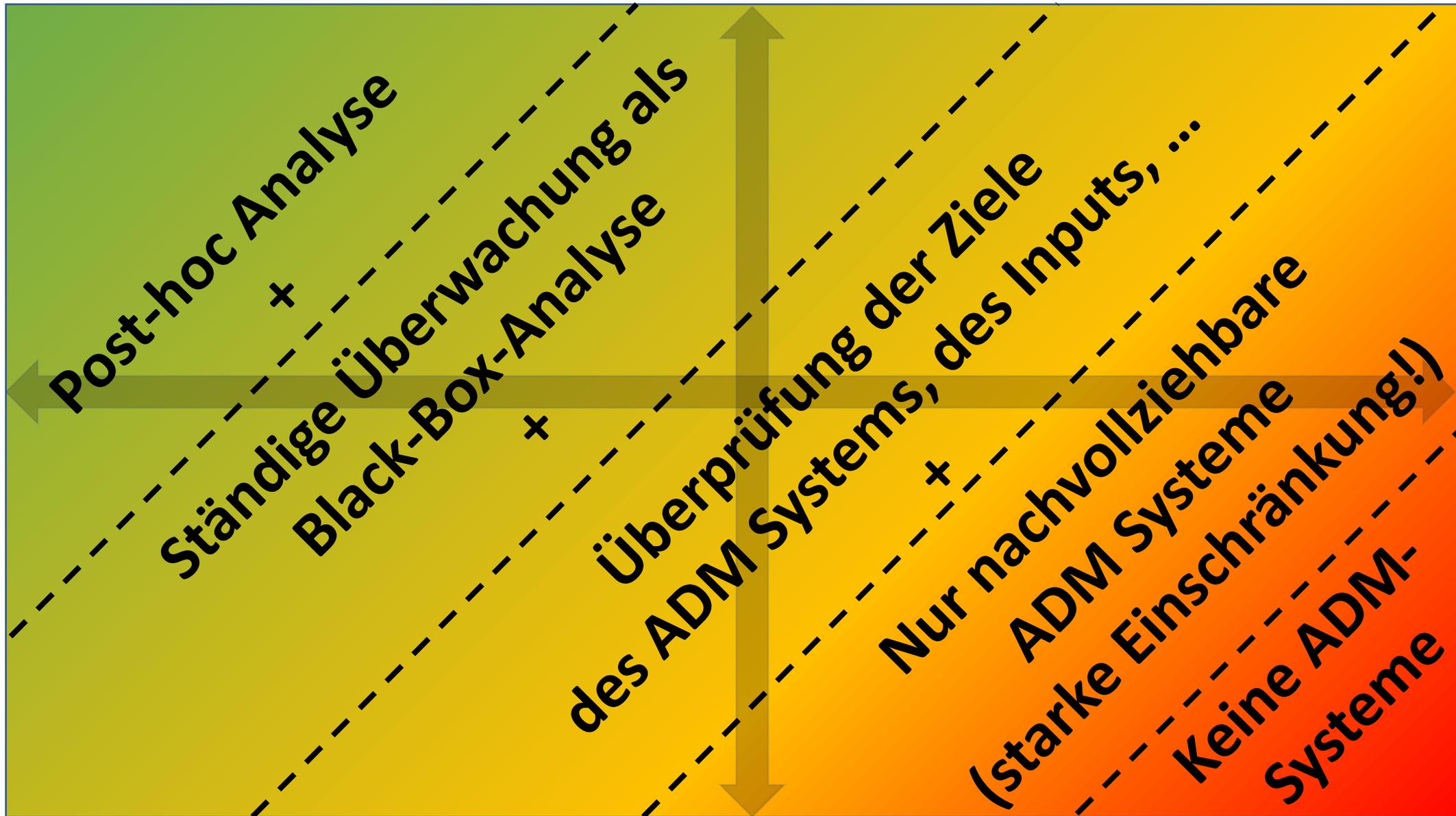
Geringer  
Gesamt-  
schaden  
bei Fehl-  
urteilen

Hoher  
Gesamt-  
schaden  
bei Fehl-  
urteilen

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



# Zusammenfassung

---

Was sind die nächsten Schritte?

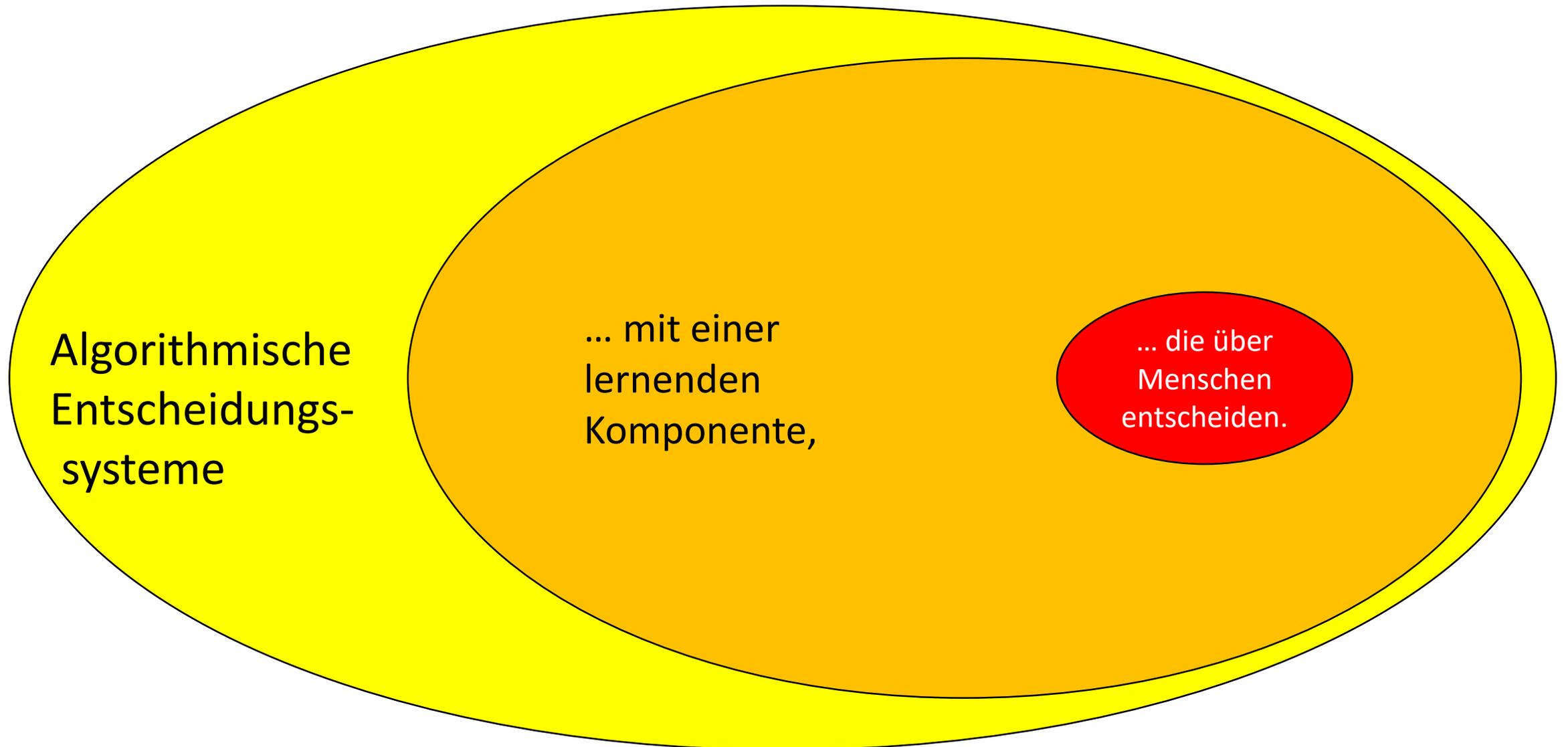
Die zwei Ängste

Sie wird richten

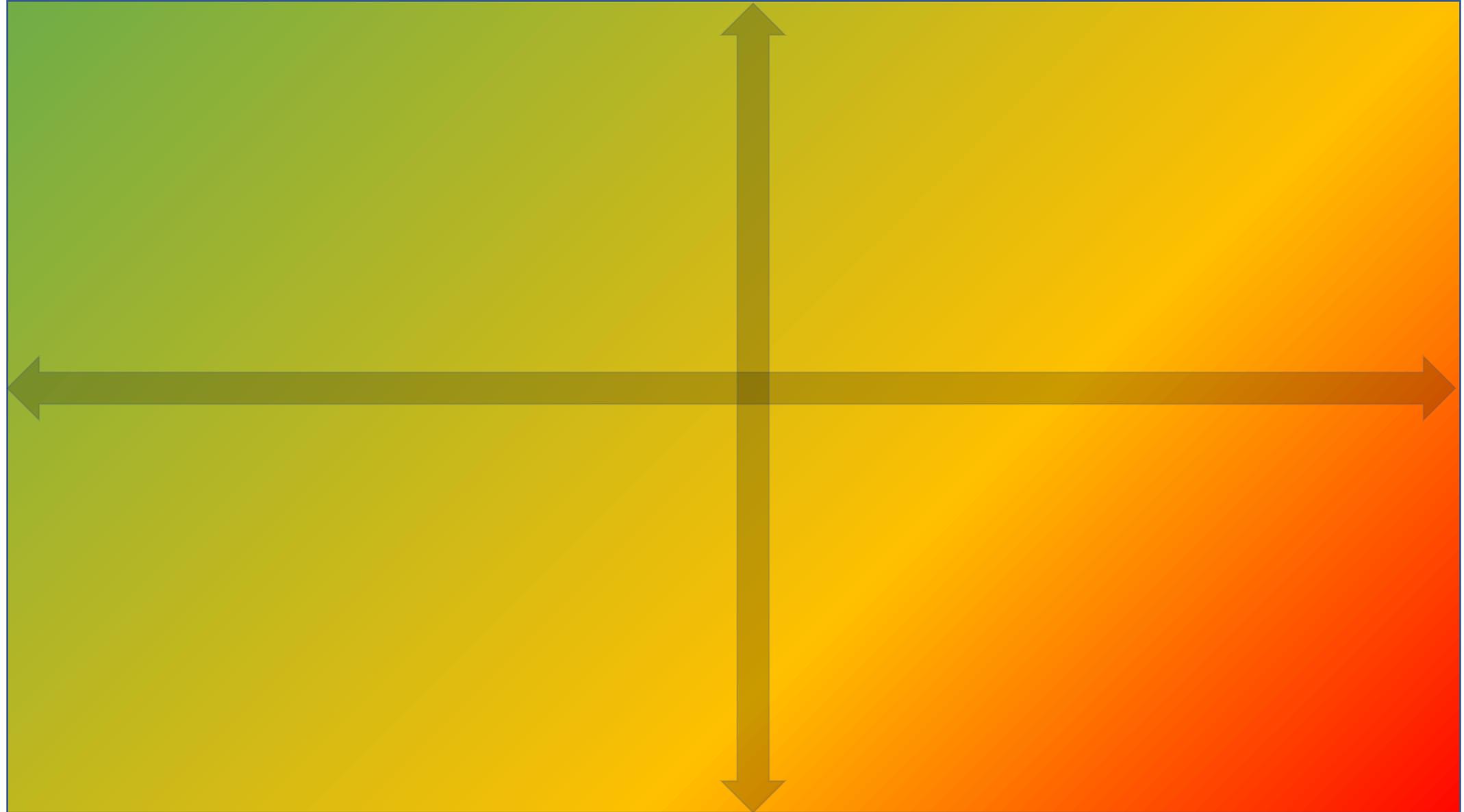
Sie wird dichten



Möglicherweise stärker zu regulieren sind:



Viele Anbieter,  
einfacher Wechsel



Monopol

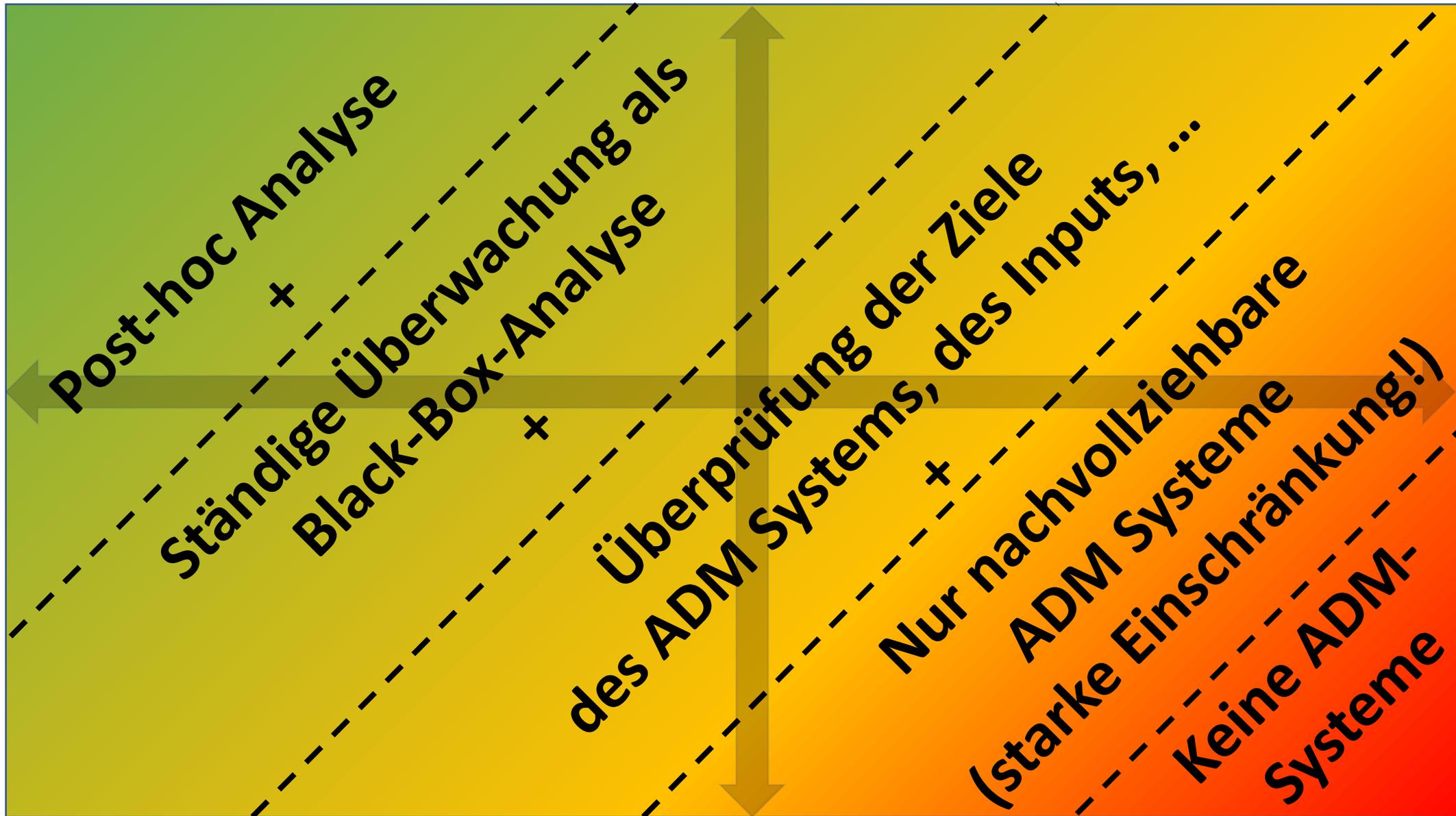
Geringer  
Gesamt-  
schaden  
bei Fehl-  
urteilen

Hoher  
Gesamt-  
schaden  
bei Fehl-  
urteilen

Viele Anbieter,  
einfacher Wechsel

Geringer  
Gesamt-  
schaden  
bei Fehl-  
urteilen

Hoher  
Gesamt-  
schaden  
bei Fehl-  
urteilen



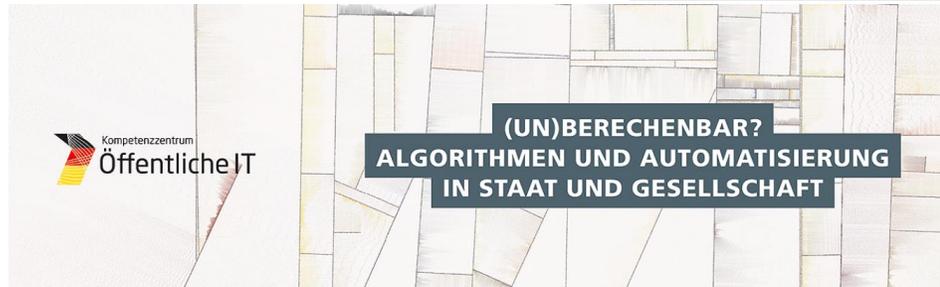
Monopol

# Weitere Informationen

Prof. Dr. Katharina A. Zweig

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

@nettwerkerin bei Twitter



1. Studie für die Bertelsmann-Stiftung:  
Zweig, Fischer & Lischka: „[Wo Maschinen irren können](#)“  
(Serie AlgoEthik, No. 4, 2018)
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 Konrad-Adenauer-Stiftung  
„Algorithmische Entscheidungen: Transparenz und Kontrolle“ (Zweig, erscheint 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)