

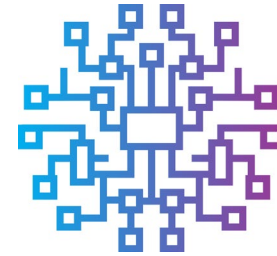


GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung



Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG



GOAL
Governance von und durch Algorithmen

AI Ethics Impact Group led by
VDE | BertelsmannStiftung



AI ETHICS – FROM PRINCIPLES TO PRACTICE

An interdisciplinary framework to operationalise AI ethics

Marc Hauer - TU Kaiserslautern & Algorithm Accountability Lab

Carla Hustedt – Projekt „Ethik der Algorithmen“ der Bertelsmann Stiftung

Wer hats gesagt?

„There is transparency regulation that I would love to see. If you look at how much regulation there is around advertising on TV and print it's just not clear why there should be less on the internet. You should have the same level of transparency required.“



AI Ethics Impact Group

| BertelsmannStiftung

VDE

H L R I S

High Performance Computing Center | Stuttgart



TECHNISCHE
UNIVERSITÄT
DARMSTADT



TECHNISCHE UNIVERSITÄT
KAISERSLAUTERN



EBERHARD-KARLS
UNIVERSITÄT
TÜBINGEN



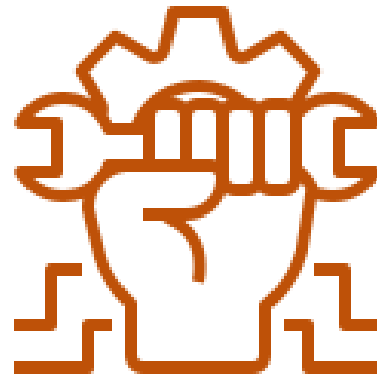
INTERNATIONAL CENTER
FOR ETHICS IN THE SCIENCES
AND HUMANITIES (IZEW)

iRights.Lab

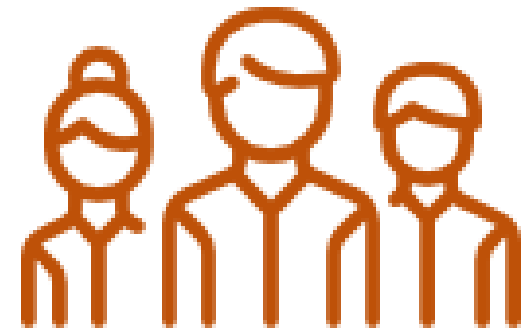
3 Herausforderungen für die praktische Implementierung von KI-Ethikprinzipien



Kontextabhängigkeit von Wertevorstellungen und –abwägungen

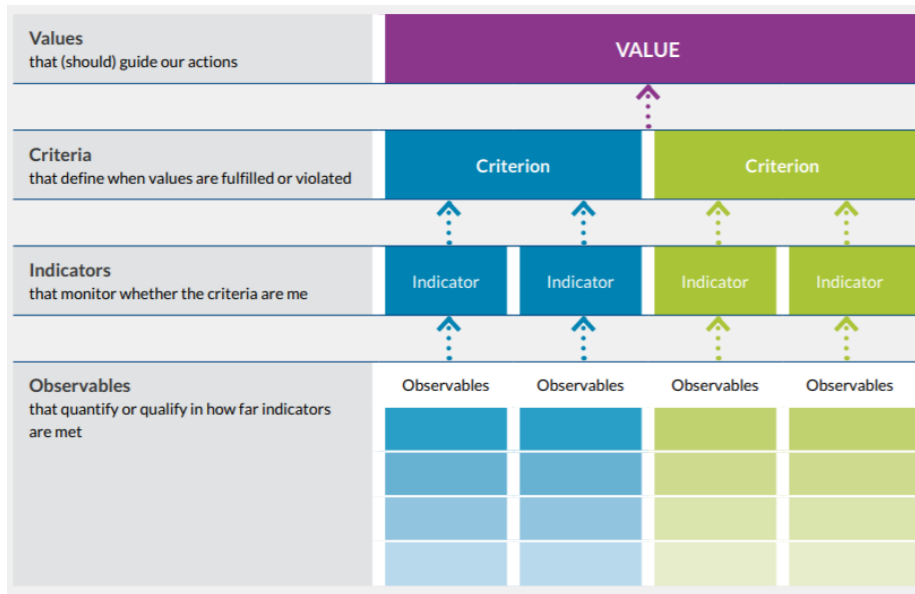


Sozio-technische Natur von ADM-Systemen



Unterschiedliche Bedürfnisse der an der Entwicklung, dem Einsatz beteiligten Akteure

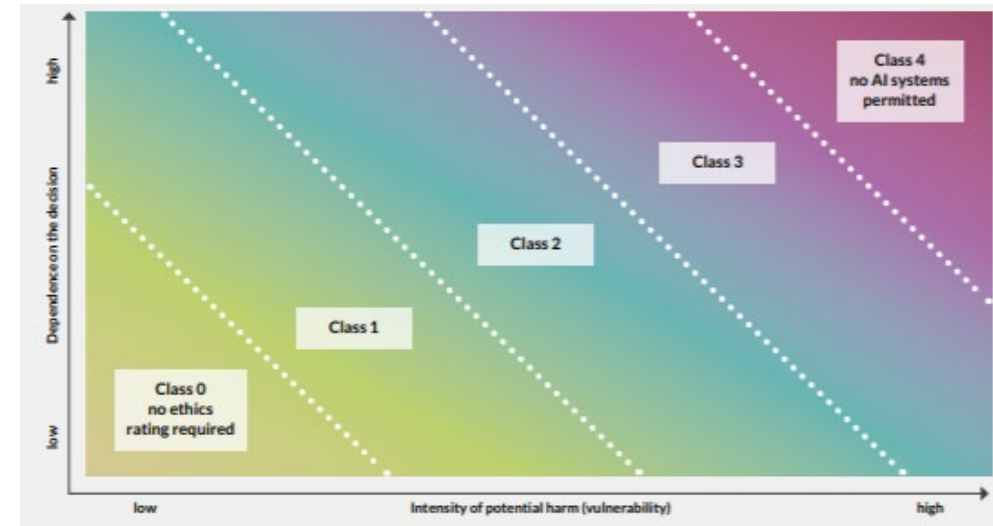
Der Ansatz der AI Ethics Impact Group



+



+



Kritikalitätsmatrix

Zwei relevante Dimensionen um potentiellen Gesamtschaden (**potential damage**) zu ermitteln.

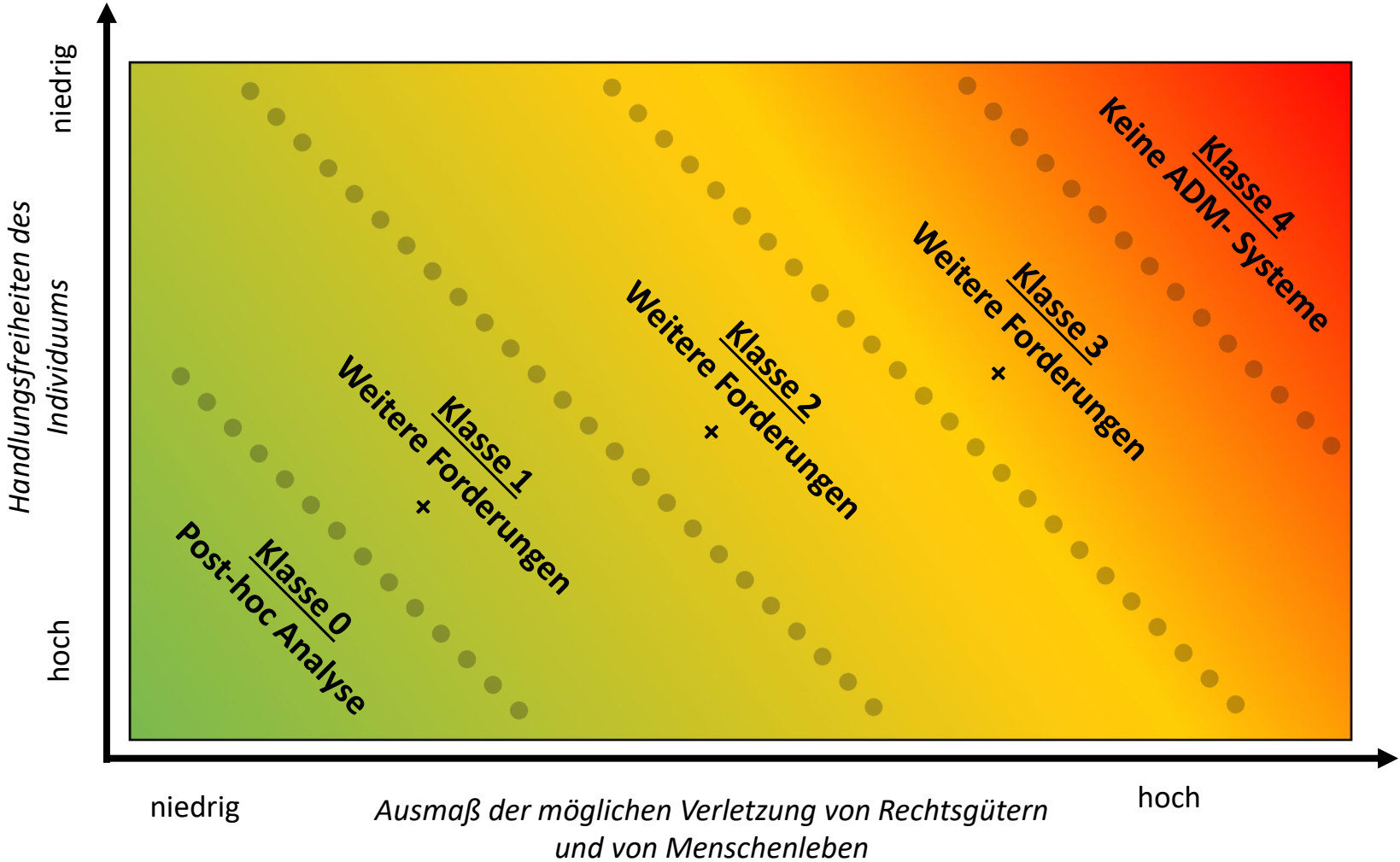
Ausmaß der möglichen Verletzung von Rechtsgütern und von Menschenleben (**harm**):

- Für einzelne Personen (**individuals**)
- Für ganze (gesellschaftliche) Gruppen von Personen (**groups**)
- Für die Gesellschaft (**society**)

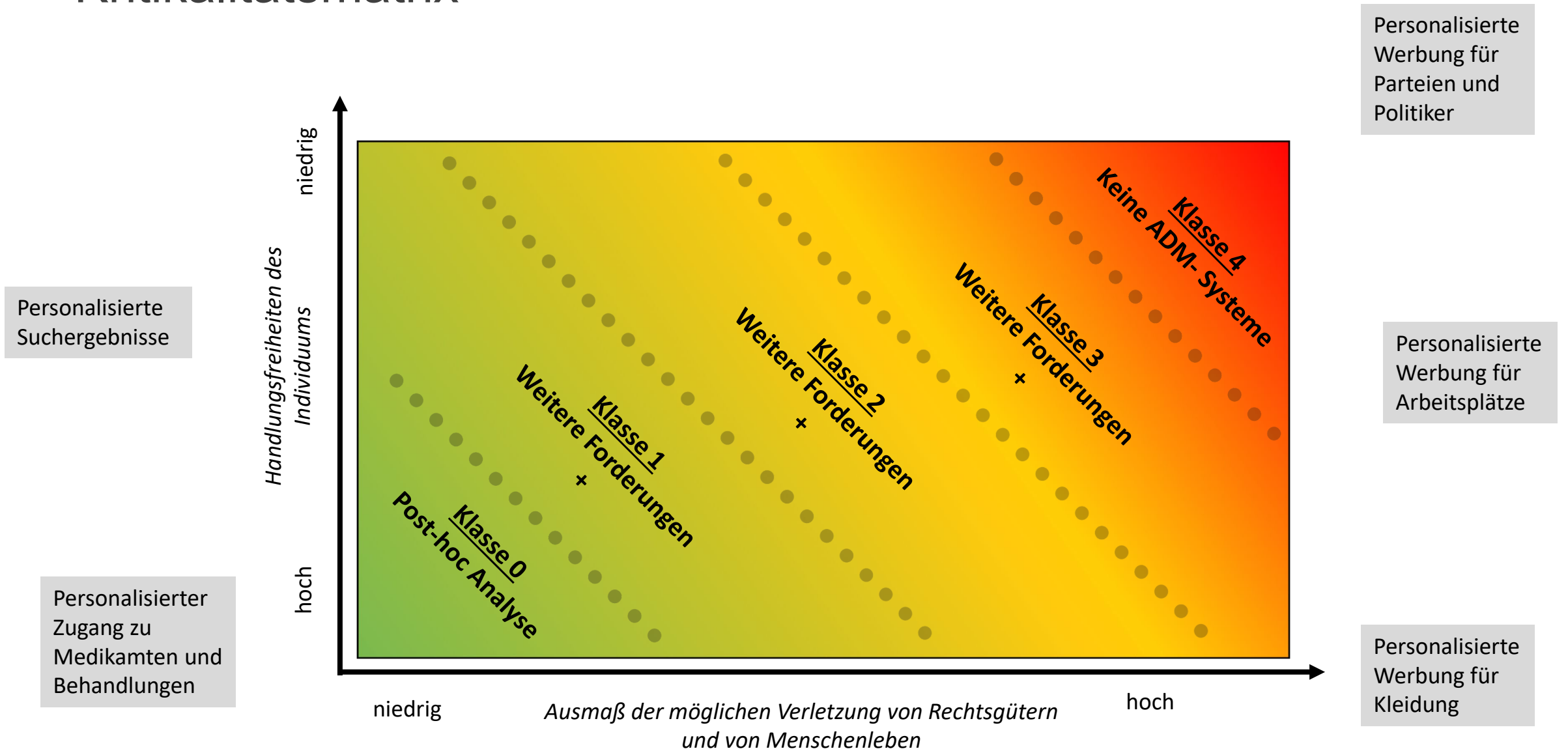
Handlungsfreiheiten des Individuums (**dependency**):

- Entscheidung durch Mensch oder Maschine (**control**)
- Wahlmöglichkeit (**switchability**)
- Möglichkeit zur Korrektur (**redress**)

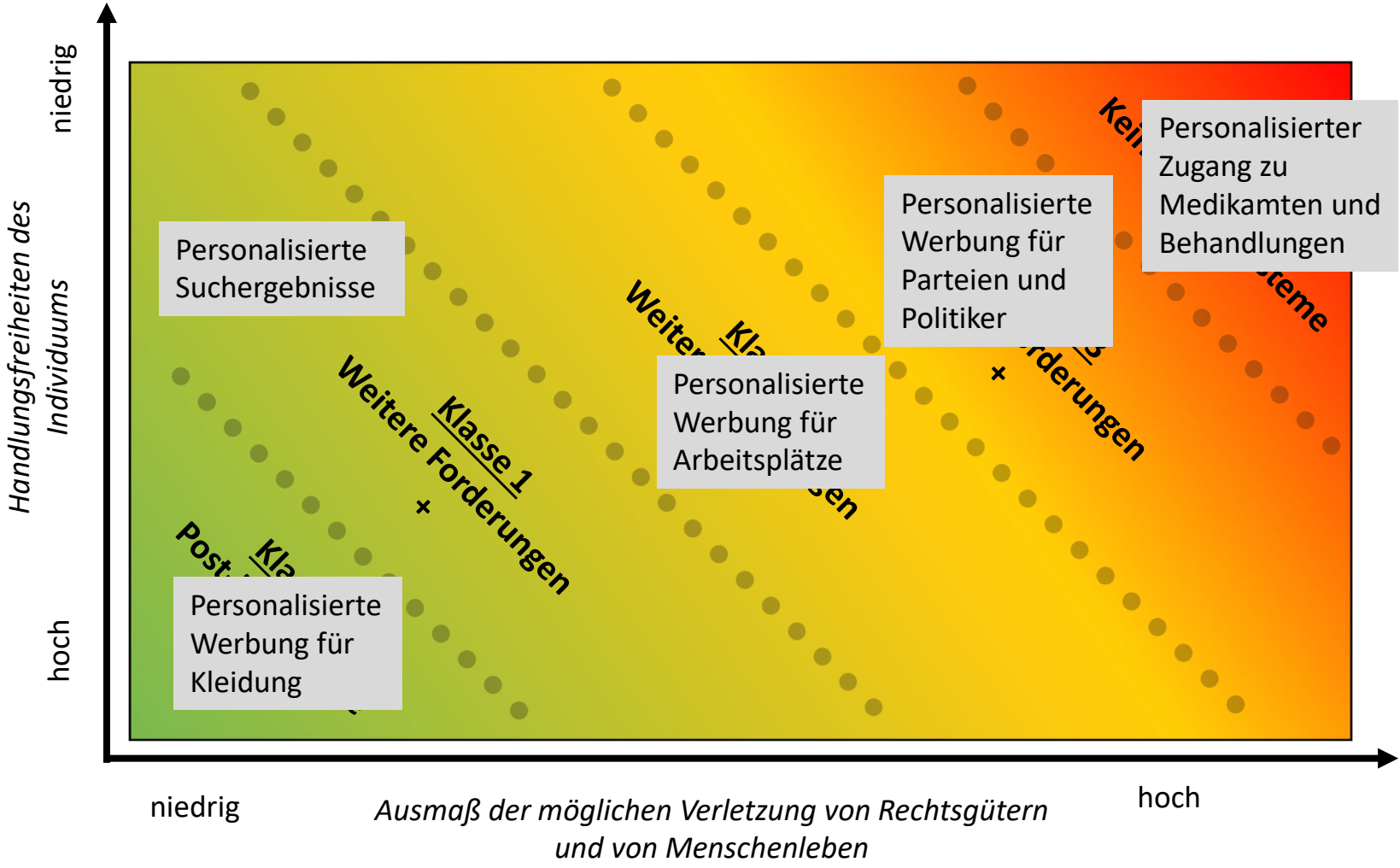
Kritikalitätsmatrix



Kritikalitätsmatrix



Kritikalitätsmatrix



Das VCIO Modell

