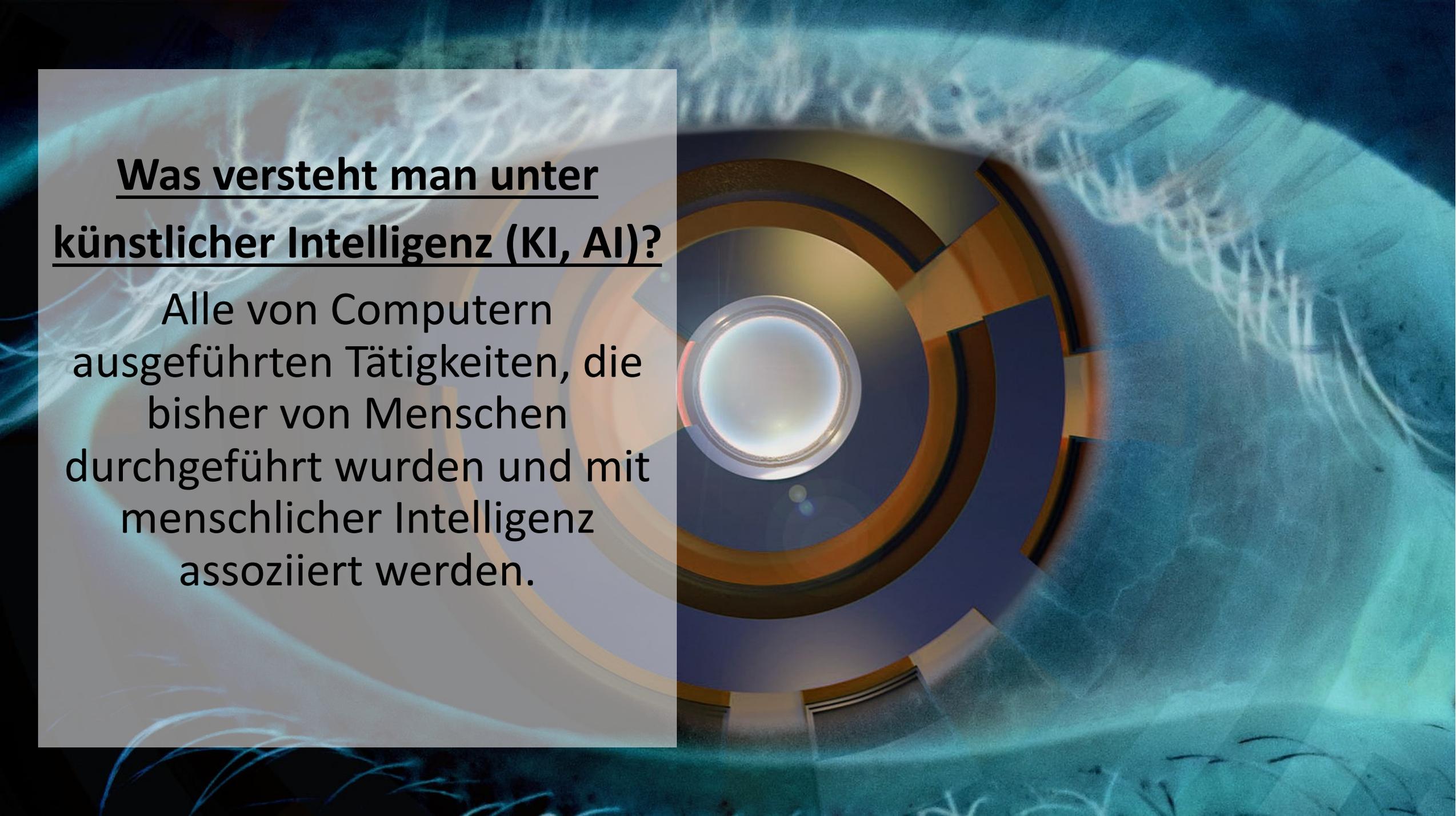


The image features two hands held up, palms facing forward. The skin of the hands is painted with a world map, showing continents in various colors like green, yellow, and brown, and oceans in blue. Labels for countries like 'CANADA', 'UNITED STATES', 'BRAZIL', and 'CHINA' are visible. The background is a soft-focus sky with white clouds. Overlaid on the hands is white text.

KI & Ethik: Wie kommt die Ethik in den Computer

Summit Industrie 4.0 Österreich

**Tobias Krafft
TU Kaiserslautern**

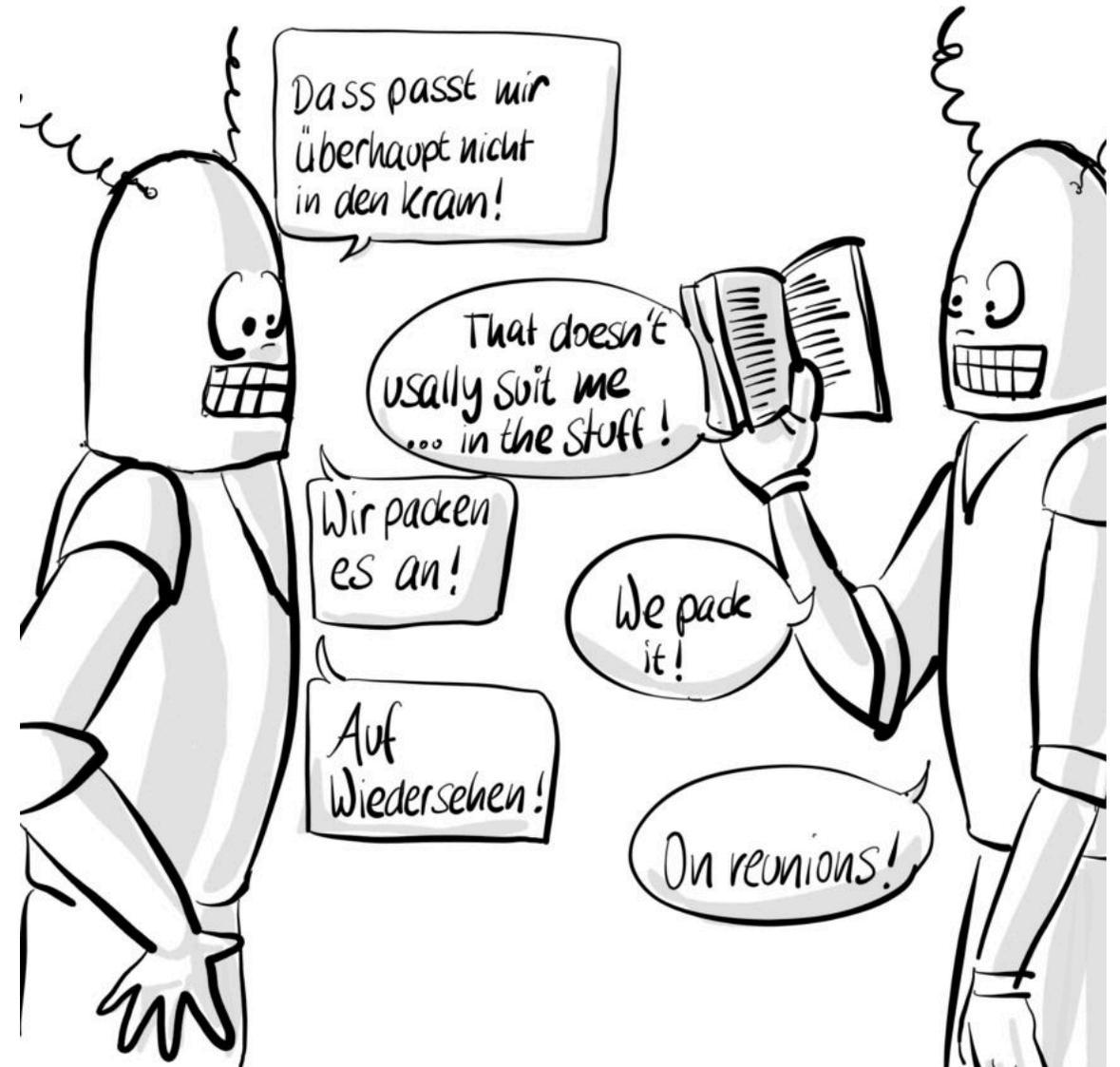


Was versteht man unter künstlicher Intelligenz (KI, AI)?

Alle von Computern ausgeführten Tätigkeiten, die bisher von Menschen durchgeführt wurden und mit menschlicher Intelligenz assoziiert werden.

KI als Methode

- Expertensysteme
- Logiksysteme
- Ontologien



Alles neu macht das maschinelle Lernen



DeepL

Translator

Linguee

Download for

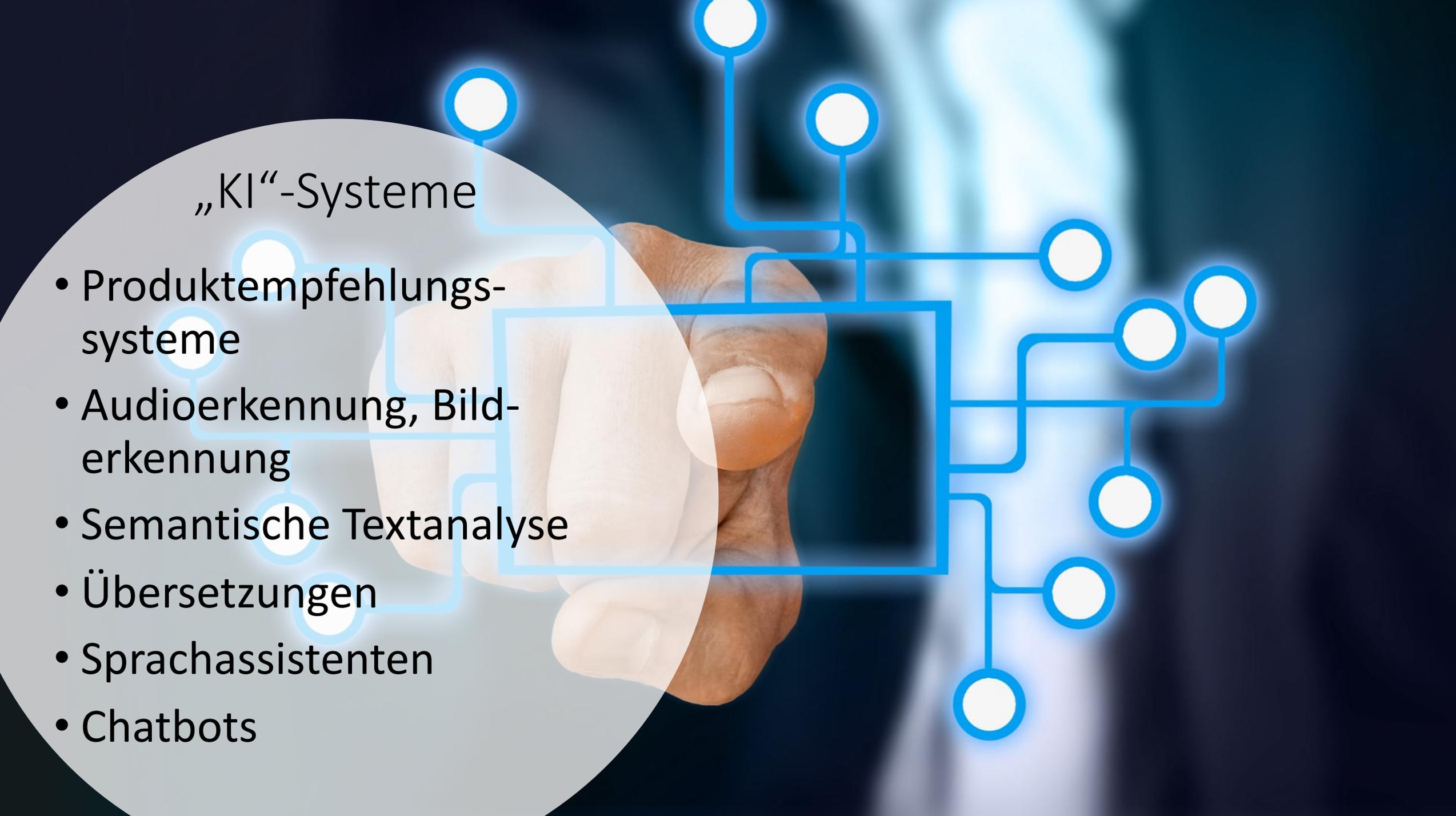
Translate from **German** (detected) ▾

Translate into **English** ▾

Das passt mir überhaupt nicht in den Kram.
Wir packen es an!
Auf Wiedersehen!

That doesn't suit my plans at all.
Let's do it!
Goodbye!

 Translate document

A hand is shown holding a glowing blue circuit board. The board has several white circular nodes connected by blue lines. The background is dark blue with a blurred light effect. A semi-transparent grey circle is overlaid on the left side of the image, containing a list of AI systems.

„KI“-Systeme

- Produktempfehlungssysteme
- Audioerkennung, Bilderkennung
- Semantische Textanalyse
- Übersetzungen
- Sprachassistenten
- Chatbots

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 ...“



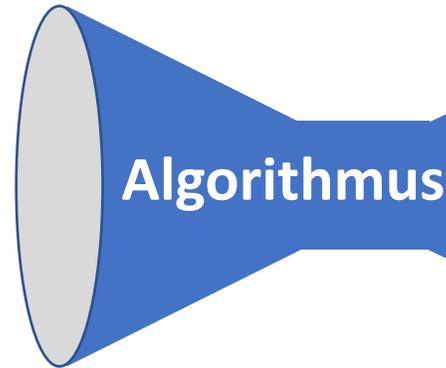
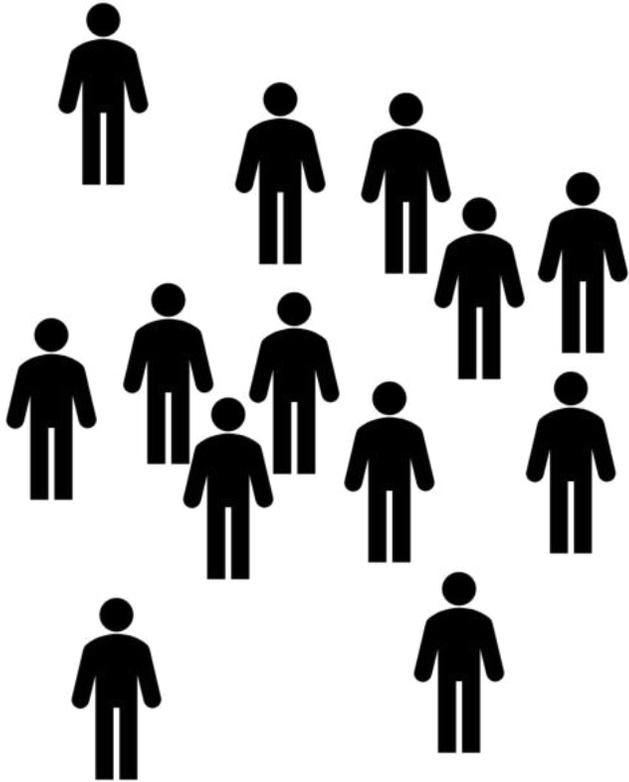
Die zwei Ängste

Sie wird dichten

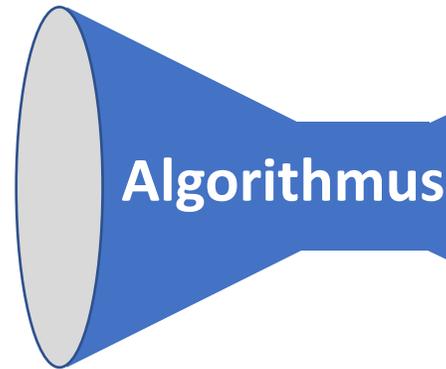
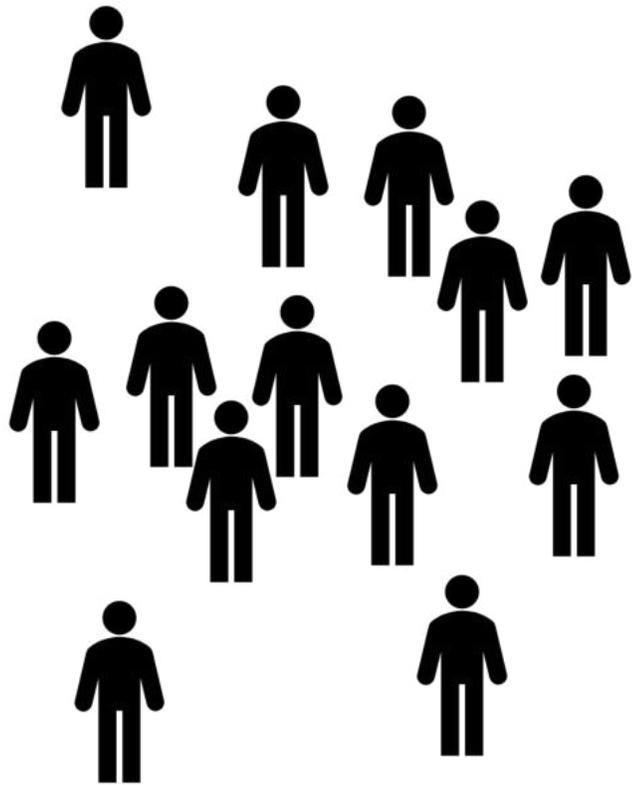
Sie wird richten



Algorithmische Entscheidungssysteme (ADM Systeme)

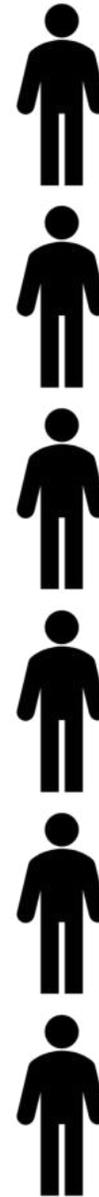
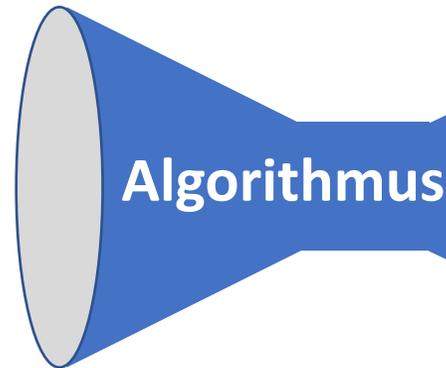
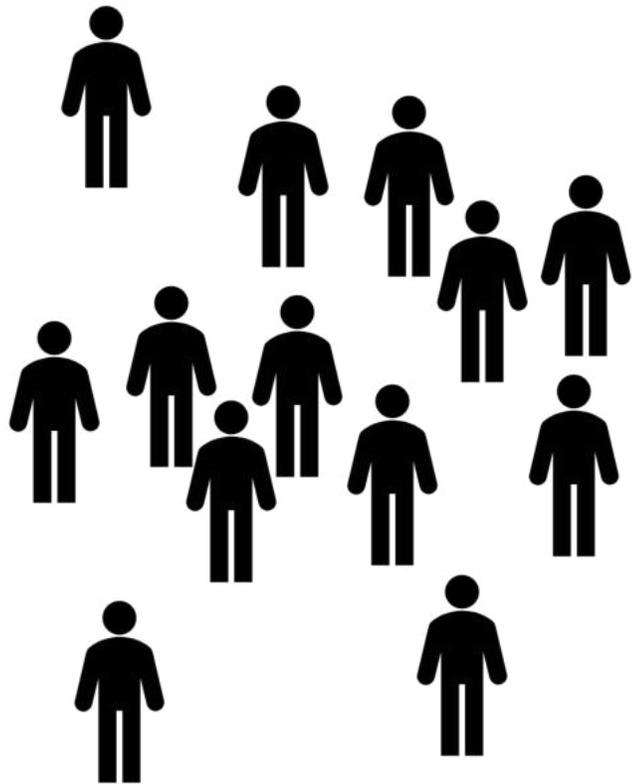


Algorithmische Entscheidungssysteme (ADM Systeme)

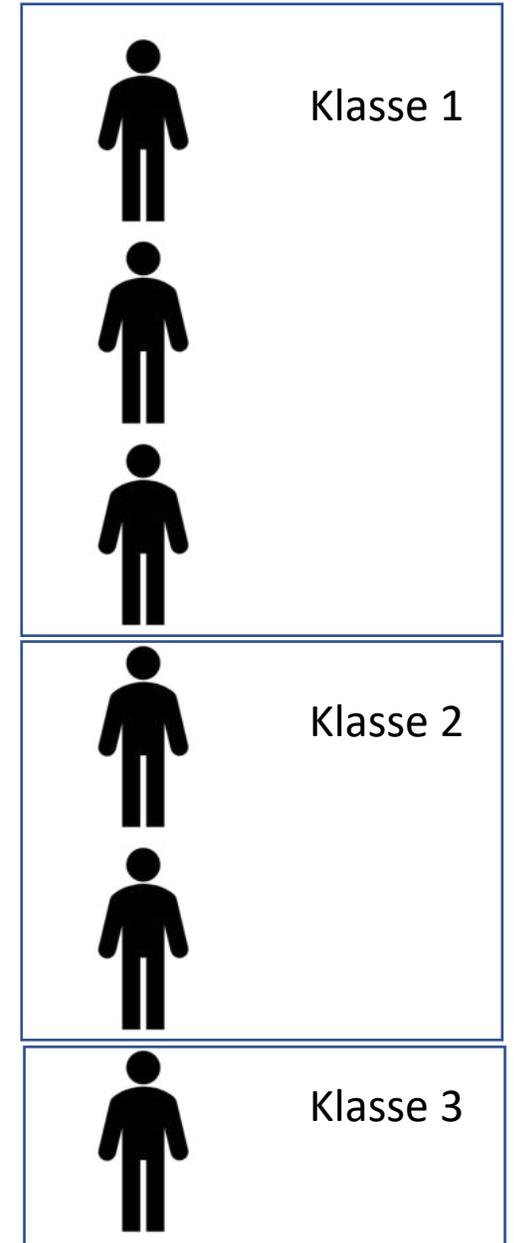


Scoring-Verfahren

Algorithmische Entscheidungssysteme (ADM Systeme)



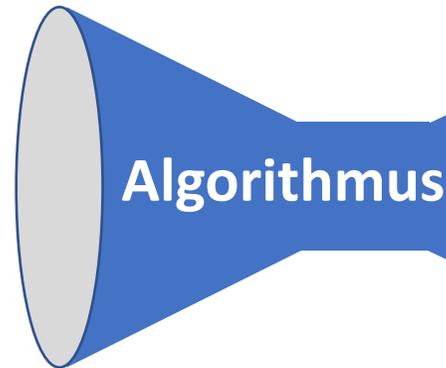
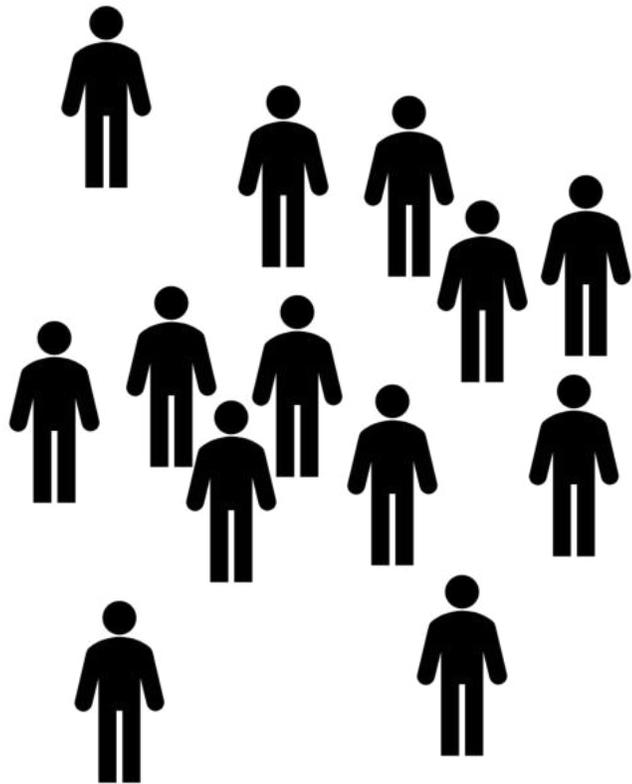
oder



Scoring-Verfahren

Klassifikation

Algorithmische Entscheidungssysteme (ADM Systeme)



oder

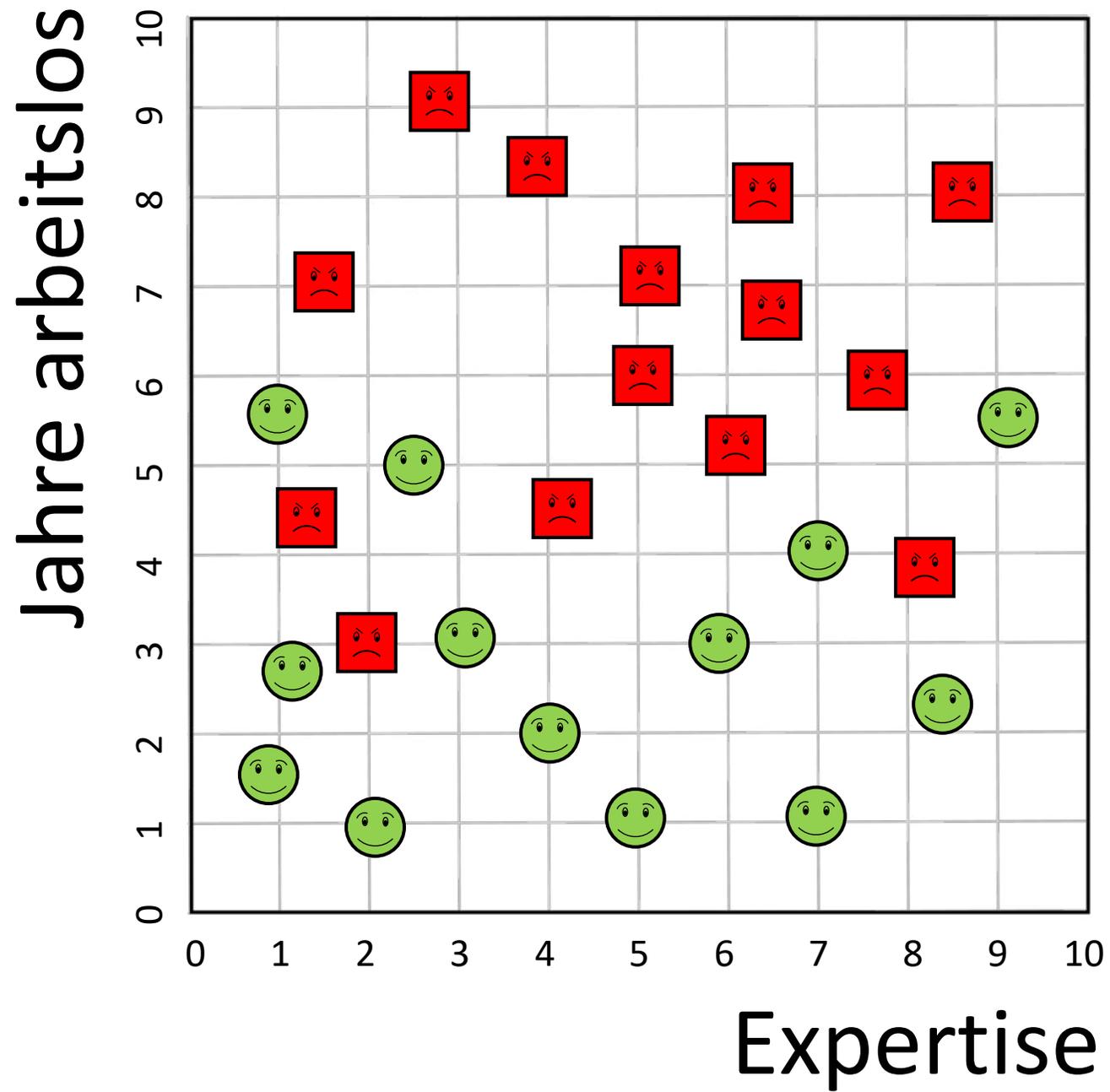


Scoring-Verfahren

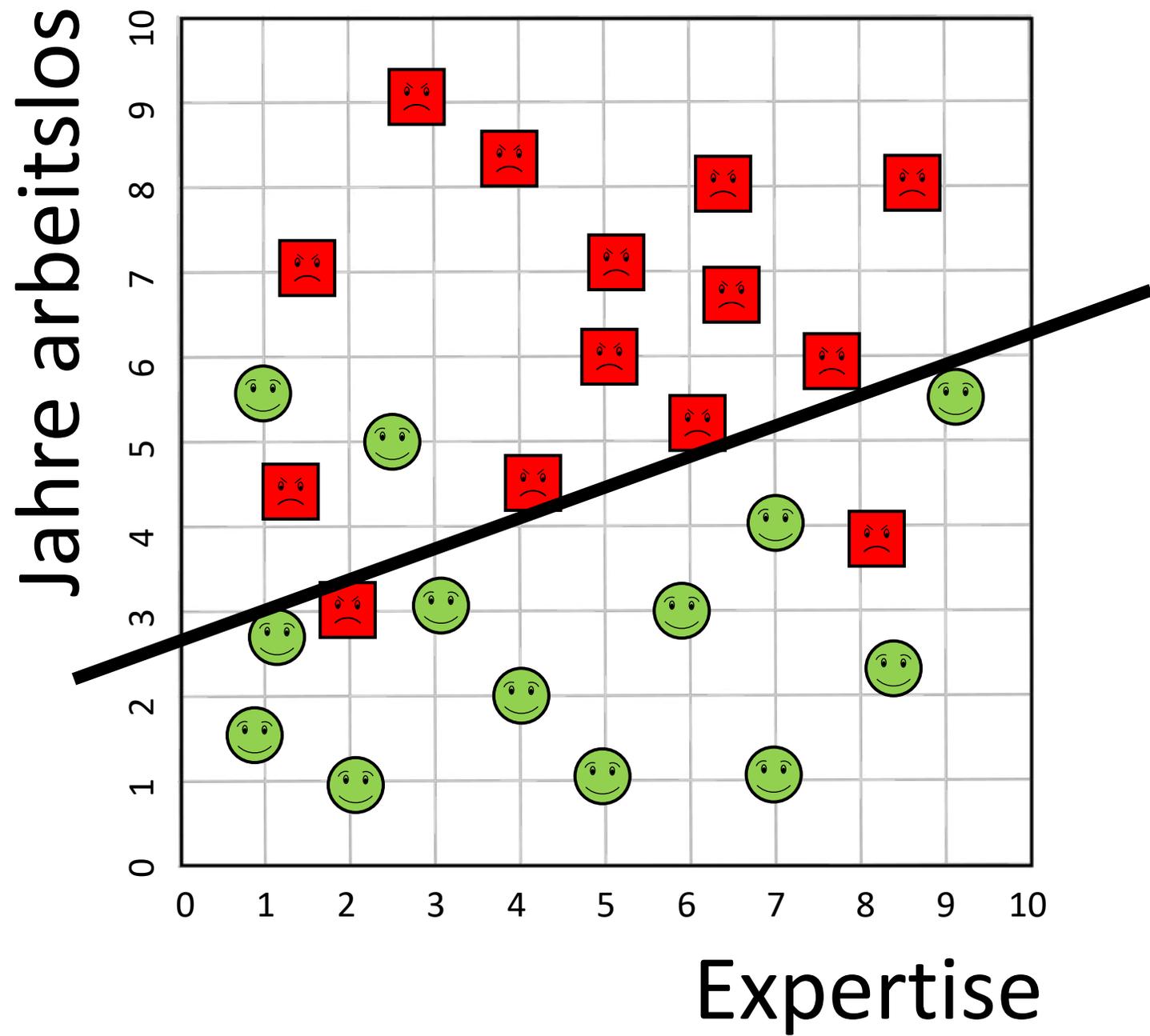
Klassifikation

Und wie geht das nun?

-  Weniger erfolgreiche Arbeitnehmer:innen
-  Erfolgreiche Arbeitnehmer:innen



-  Weniger erfolgreiche Arbeitnehmer:innen
-  Erfolgreiche Arbeitnehmer:innen





Weniger erfolgreiche Arbeitnehmer:innen



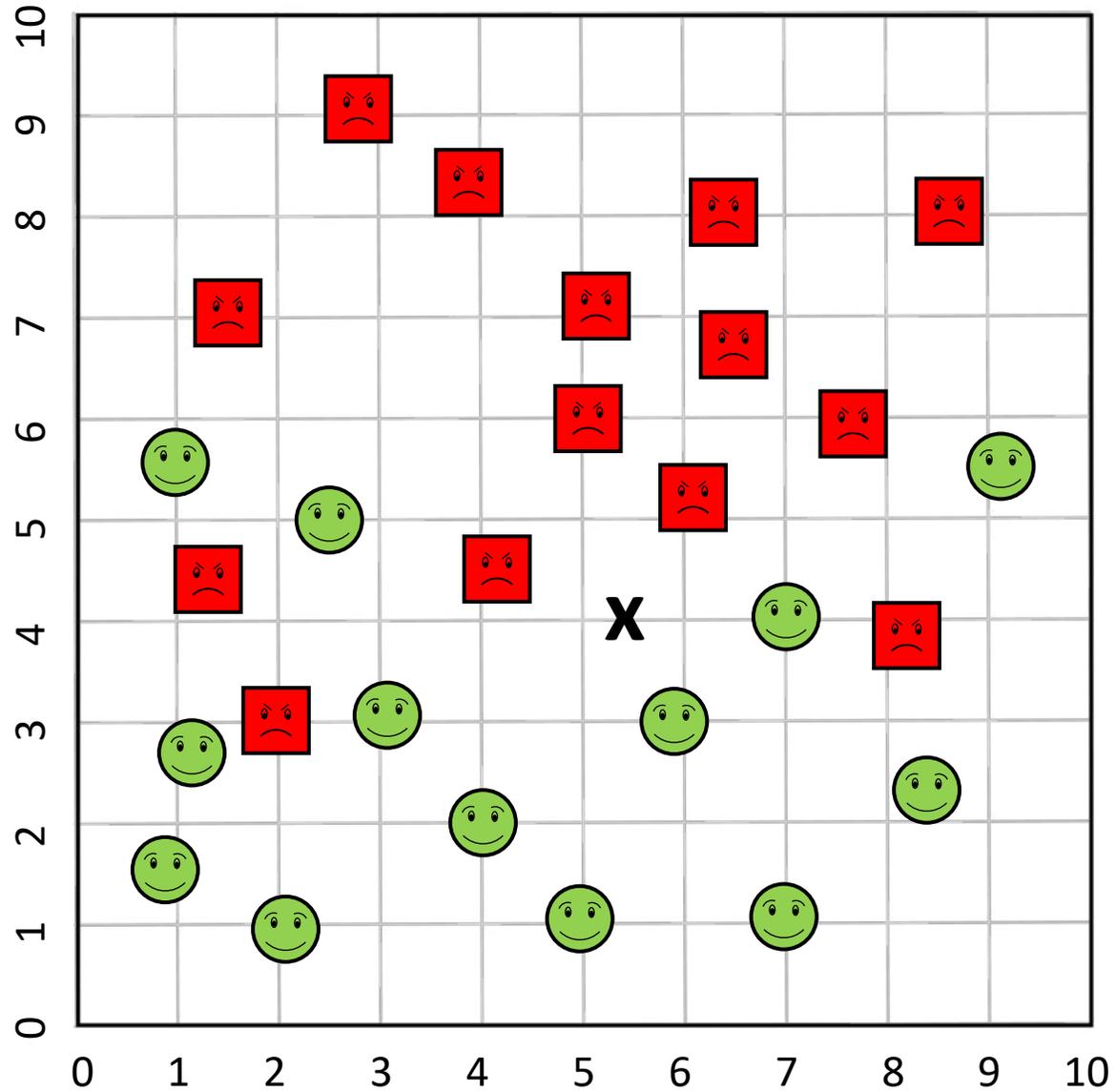
Erfolgreiche Arbeitnehmer:innen

Bewerten Sie Frau Müller:

5.5 Jahre Erfahrung

4 Jahre arbeitslos

Jahre arbeitslos



Expertise



Weniger erfolgreiche Arbeitnehmer:innen



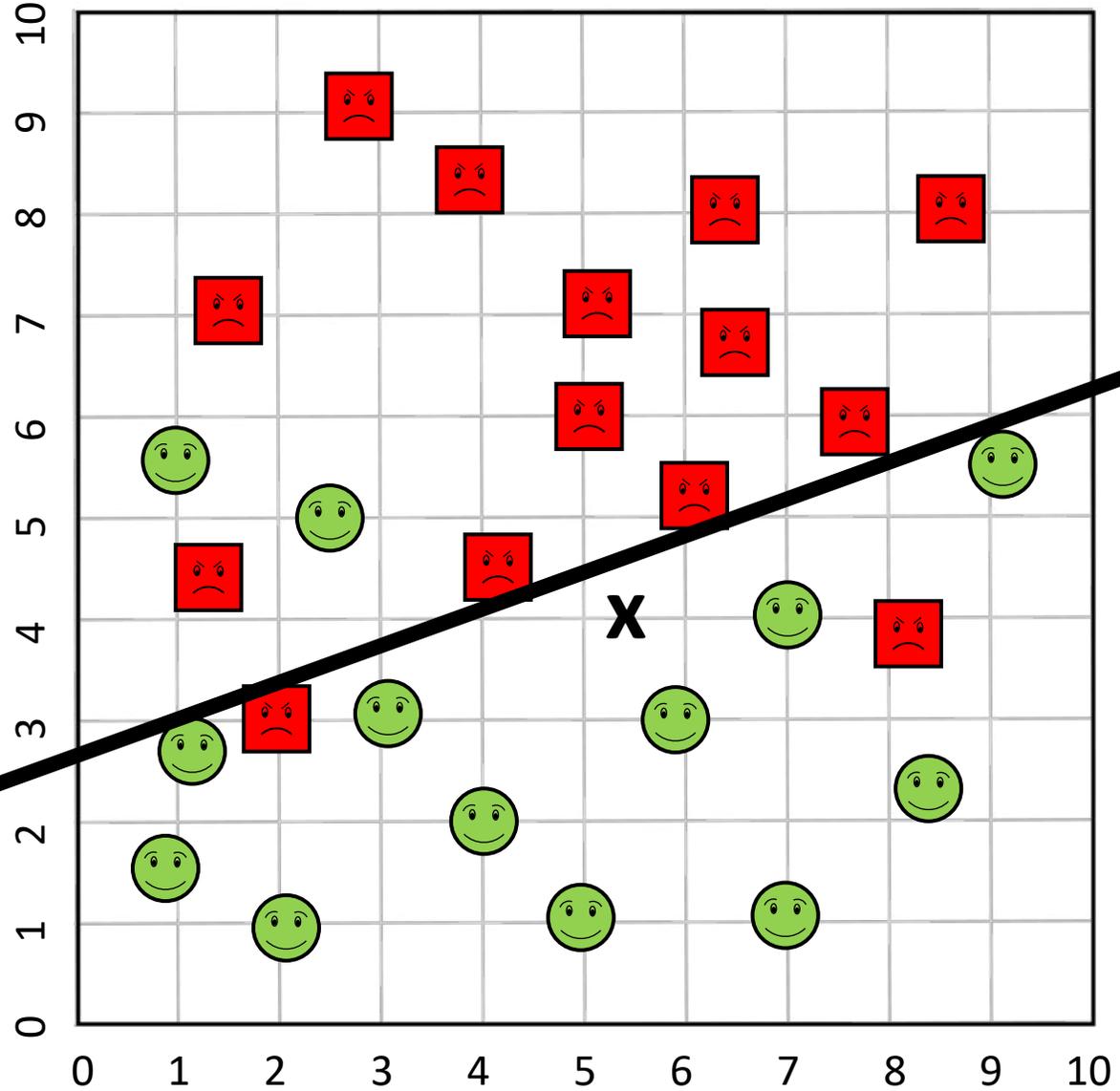
Erfolgreiche Arbeitnehmer:innen

Bewerten Sie Frau Müller:

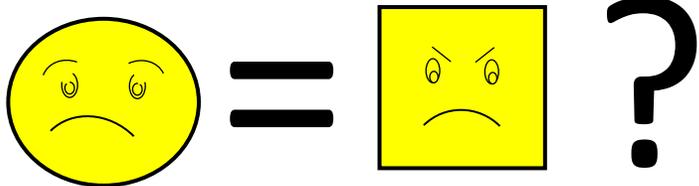
5.5 Jahre Erfahrung

4 Jahre arbeitslos

Jahre arbeitslos



Expertise

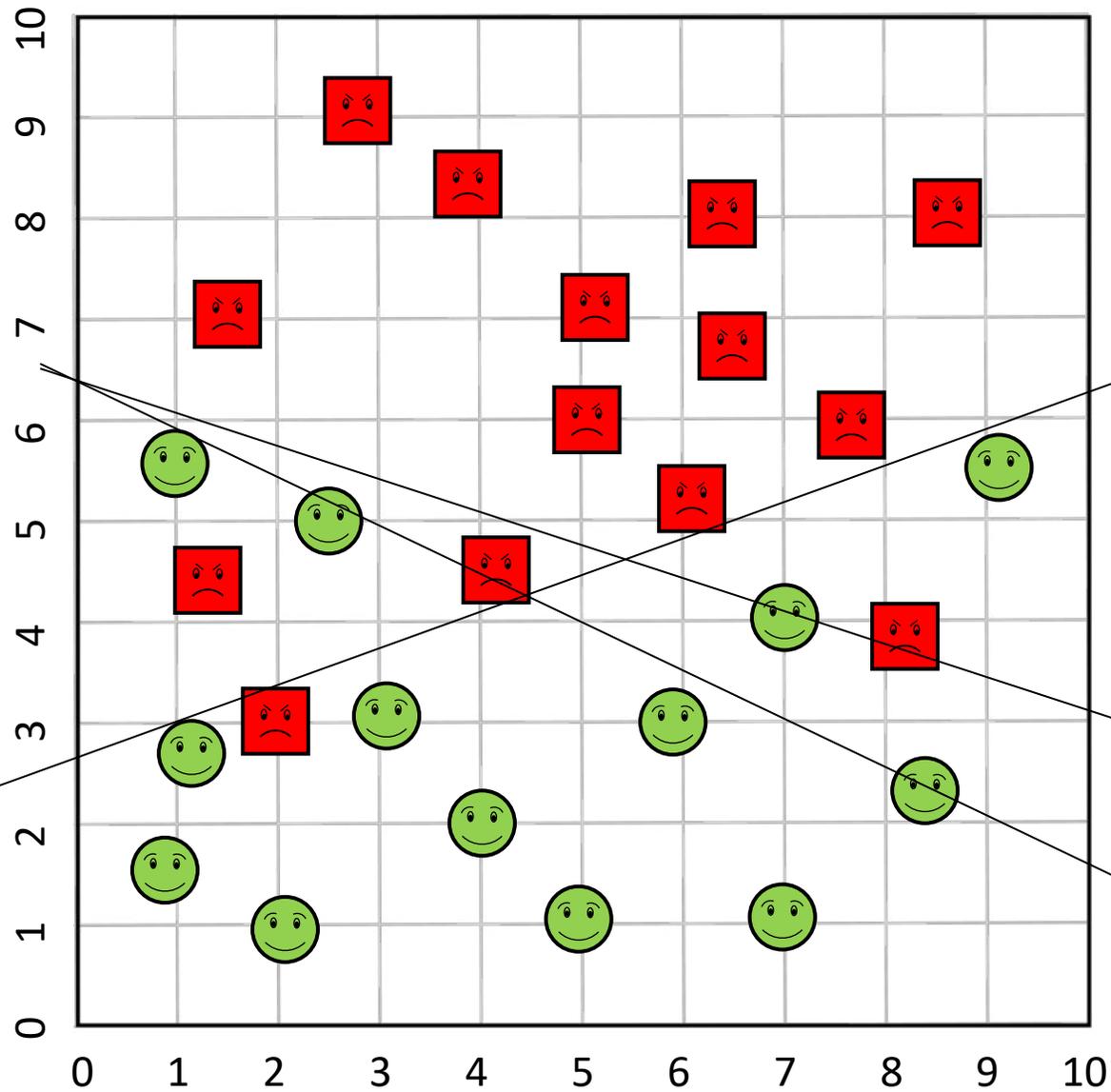


Weniger erfolgreiche Arbeitnehmer:innen



Erfolgreiche Arbeitnehmer:innen

Jahre arbeitslos



Expertise



Weniger erfolgreiche
Arbeitnehmer:innen



Erfolgreiche Arbeit-
nehmer:innen

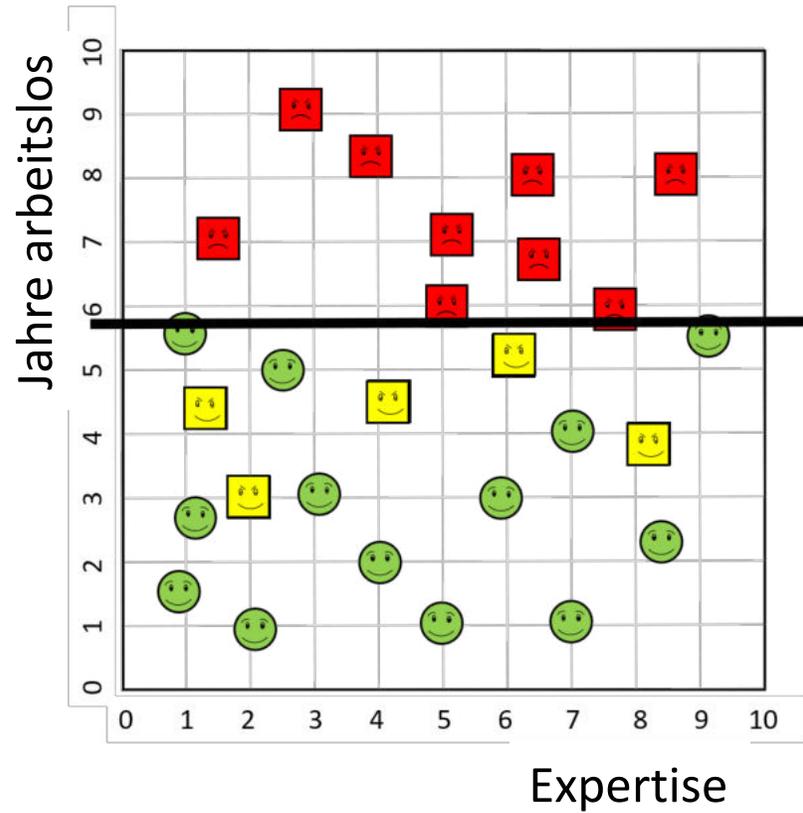
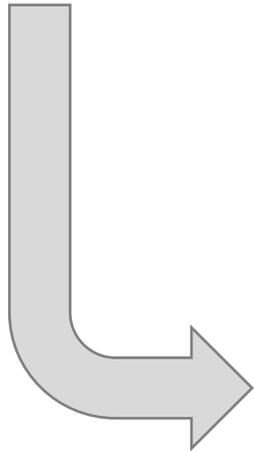
„Wir können es uns nicht leisten
eine:n **erfolgreiche:n**
Arbeitnehmer:innen
nicht einzustellen“



Weniger erfolgreiche
Arbeitnehmer:innen



Erfolgreiche Arbeit-
nehmer:innen



„Wir können es uns nicht leisten
eine:n **erfolgreiche:n**
Arbeitnehmer:innen
nicht einzustellen“

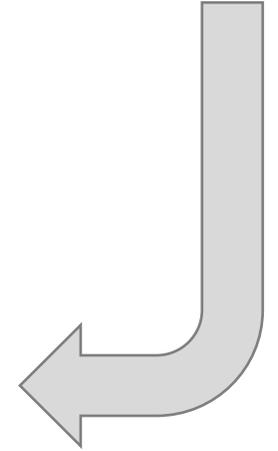
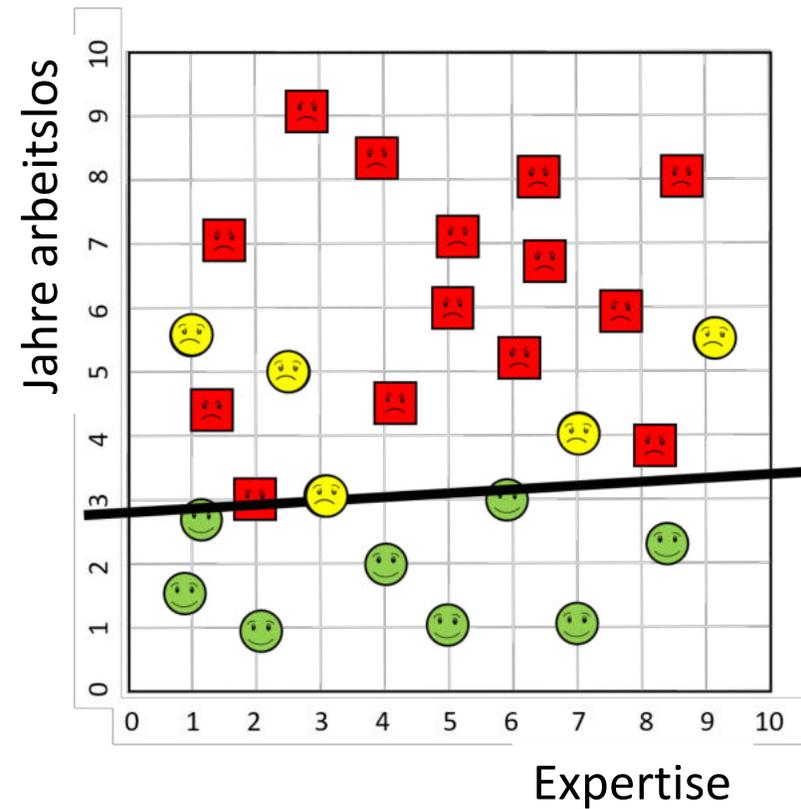
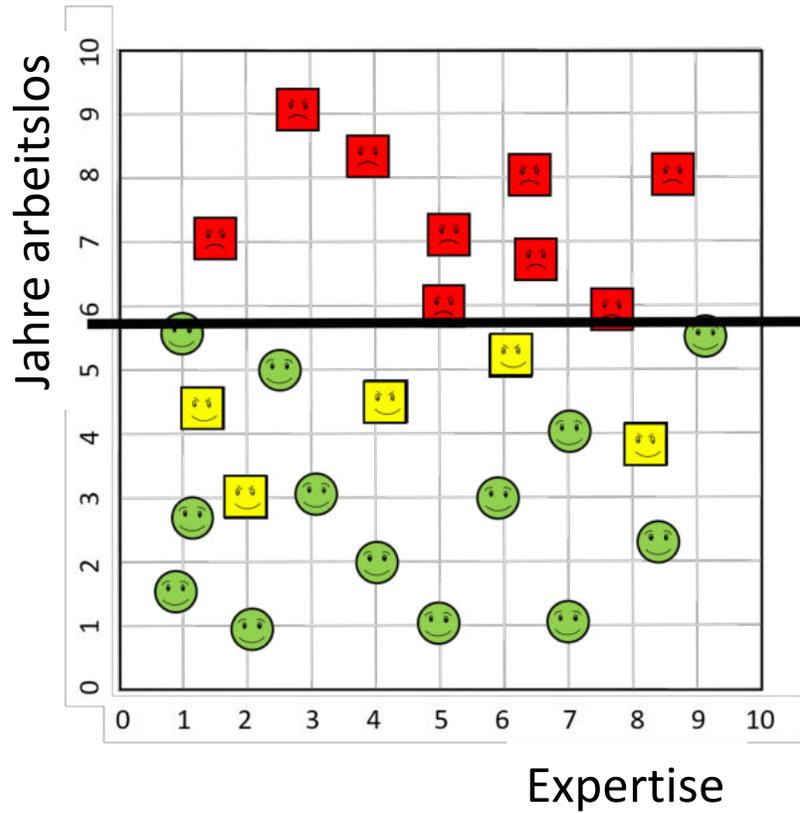
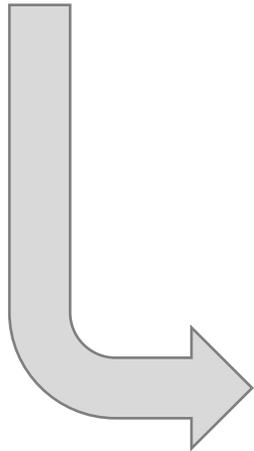


Weniger erfolgreiche
Arbeitnehmer:innen



Erfolgreiche Arbeit-
nehmer:innen

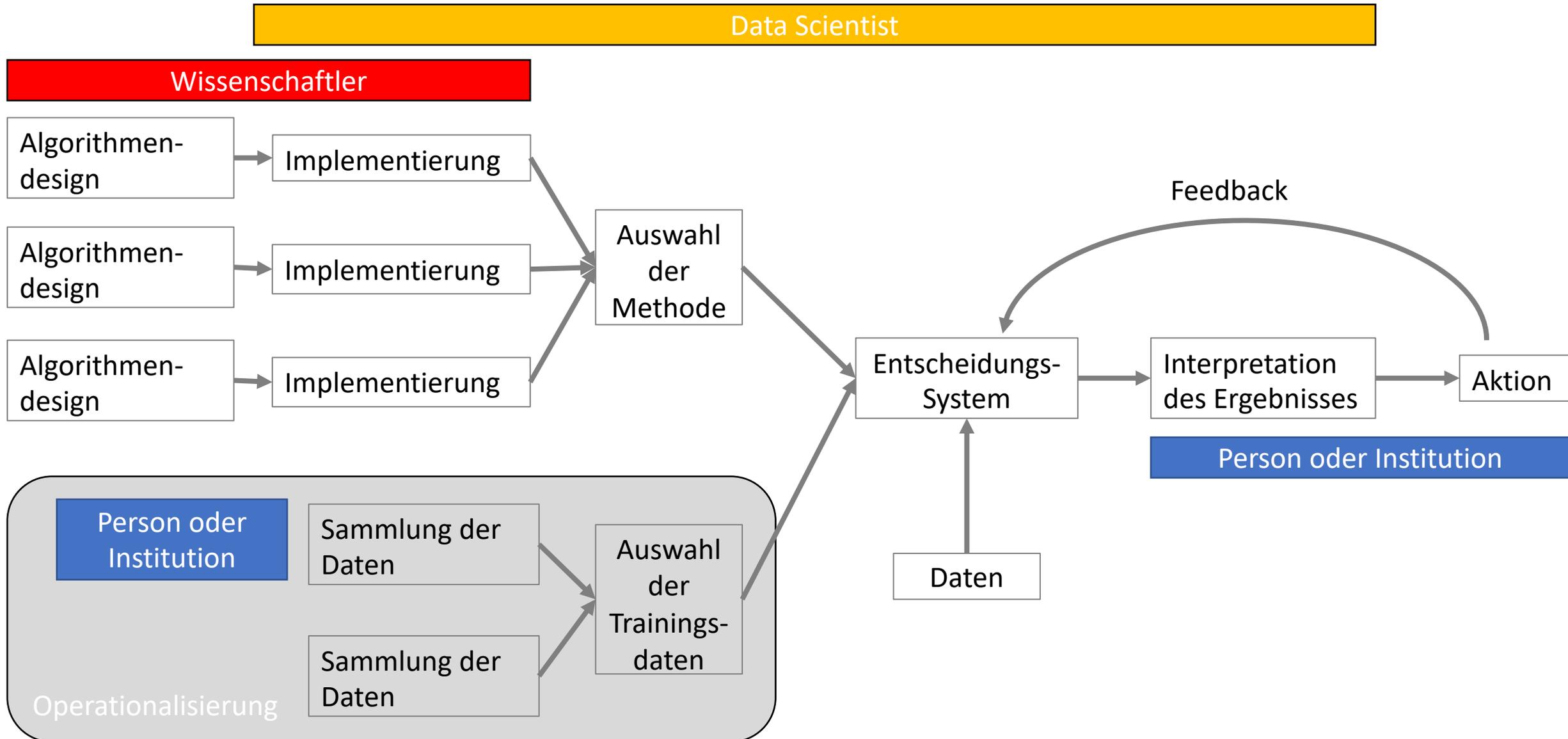
„Wir können es uns nicht leisten
eine:n **weniger erfolgreiche:n**
Arbeitnehmer:innen
einzustellen“



Wer ist verantwortlich für
diese Entscheidungen?



Lange Kette der Verantwortlichkeiten



A man with a beard and a blue shirt, looking thoughtful, with a complex mechanical gear system overlaid on his chest. The text "Richtende KI-Systeme sind problematisch" is overlaid on the image.

Richtende KI-Systeme sind
problematisch



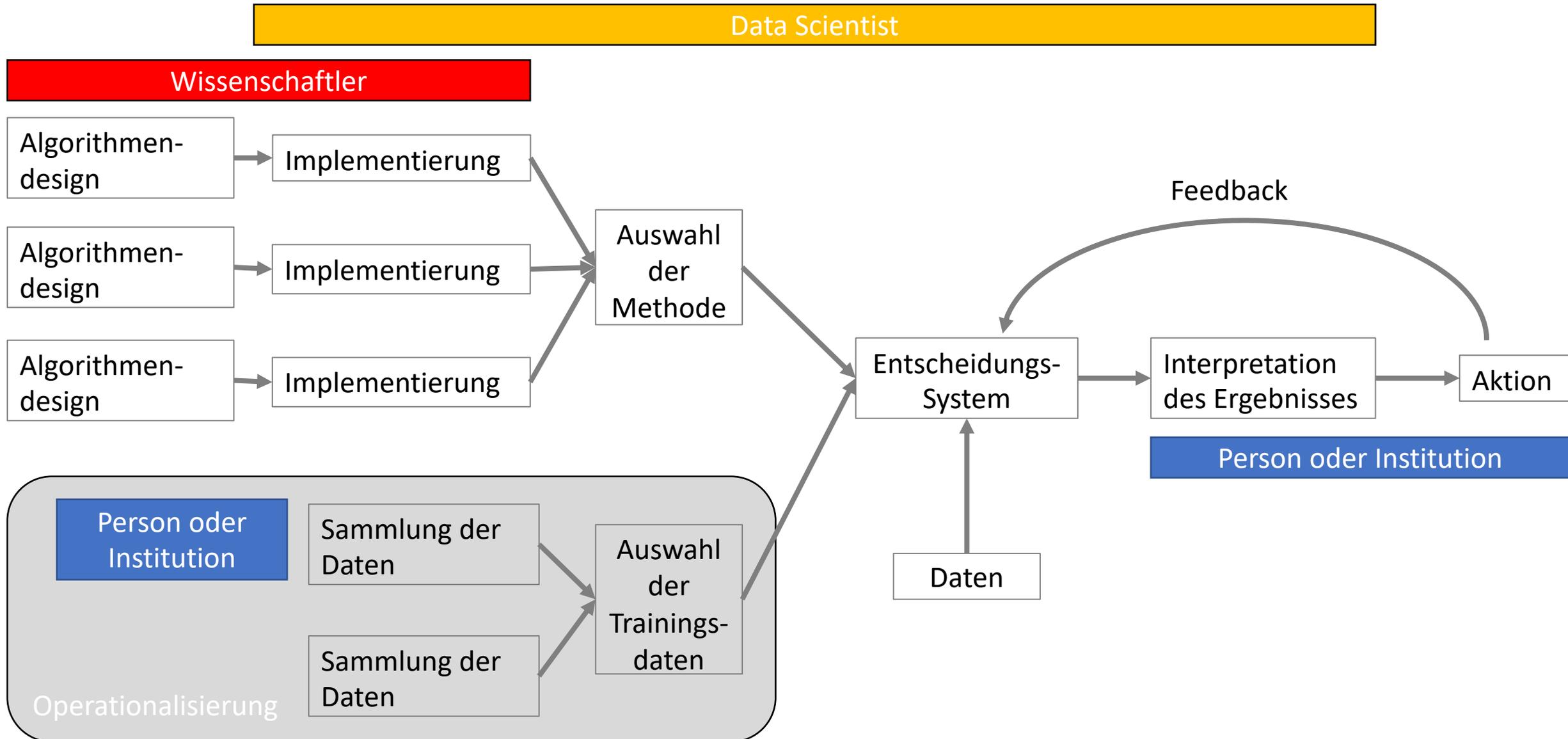
Sie verstehen Kontexte nicht,
können nicht interpretieren,
können keine Nachfragen stellen

Richtende KI-Systeme sind
problematisch



Was, wenn sie trotzdem eingesetzt werden sollen?

Lange Kette der Verantwortlichkeiten



Assistenzsystem AMAS



- Teilt Arbeitslose in 3 Klassen ein:
 - Hohe Integrationschancen – keine weiteren Maßnahmen nötig.
 - Mittlere Integrationschancen – mit Maßnahmen
 - Niedrige Integrationschancen – Maßnahmen nicht sinnvoll.

Transparenz

- Technischer Report mit Bericht über:
 - Verwendete Daten
 - Verwendete Methoden
 - Qualitäts- und Fairnessmaß
 - Werte dieser Maße
 - Resultierende Formel für einen Fall mit Beispielen

1. Resultat:

- Weist Älteren (> 50), Frauen, Pflegenden höheres Risiko zu.



1. Resultat:

- Weist Älteren (> 50), Frauen, Pflegenden höheres Risiko zu.
- **Diskriminierung?**



1. Resultat:

- Weist Älteren (> 50), Frauen, Pflegenden höheres Risiko zu.
- **Diskriminierung?**
- **Kommt auf Umsetzung an!**

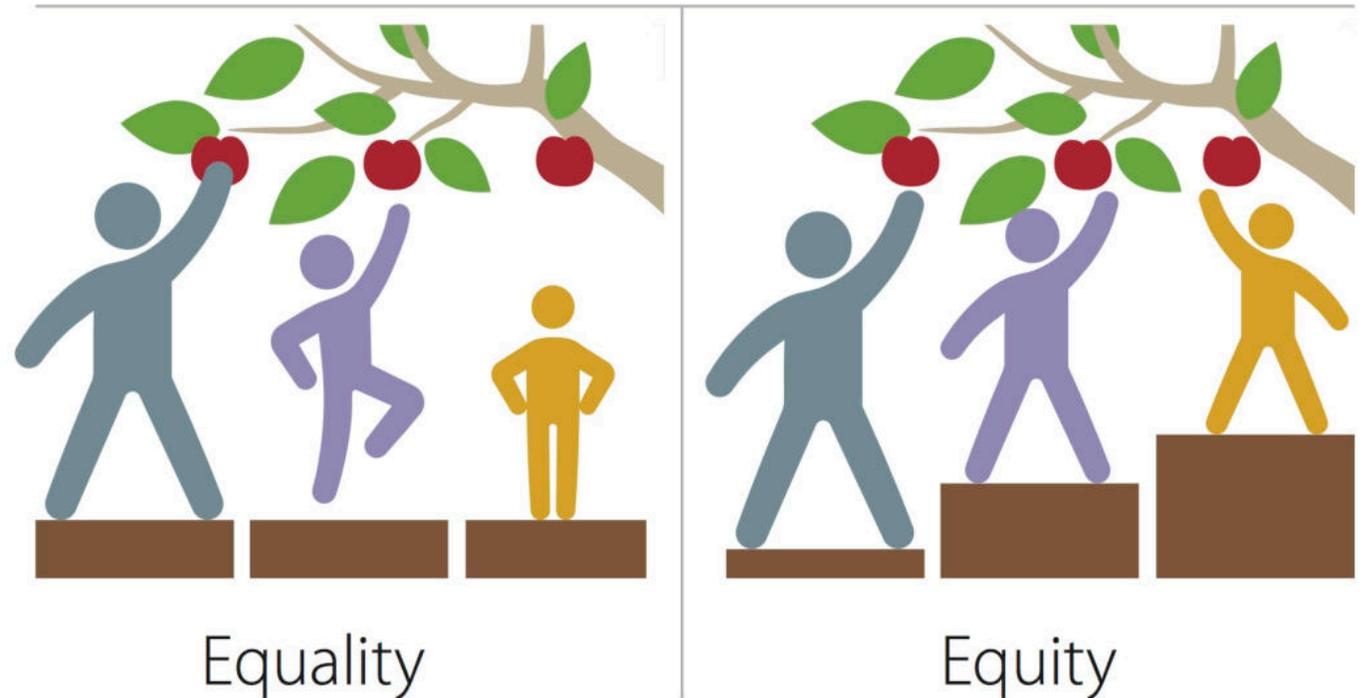


Was ist fair?

- Laut Aussage des AMS-Leiters werden vom Arbeitsmarkt benachteiligte Personen durch Einsatz öfter unterstützt.
- Damit kann das Gesamtsystem nur dann ausgleichend wirken, **wenn** das ADM-System diskriminiert.

Was ist fair?

- Laut Aussage des AMS-Leiters werden vom Arbeitsmarkt benachteiligte Personen durch Einsatz öfter unterstützt.
- Damit kann das Gesamtsystem nur dann ausgleichend wirken, **wenn** das ADM-System diskriminiert.
- Es gibt Dutzende von Fairnessmaßen: Welches sollte verwendet werden?
- Wahl bedarf der demokratischen Legitimierung in wichtigen Fällen.



By: MPCA Photos

<https://www.flickr.com/photos/mpcaphotos/31655988501>

<https://creativecommons.org/licenses/by-nc/2.0/>

Wichtig: Sozialverträglichkeits- regeln

- Klassifizierung muss mit Bürger:in im Dialog besprochen werden.
- Darf nur unterstützend sein.
- Wird jedes Jahr neu berechnet.
- Nur Daten der letzten 4 Jahre.



2. Resultat:

- Es werden nur 2 Handvoll Eigenschaften verwendet.

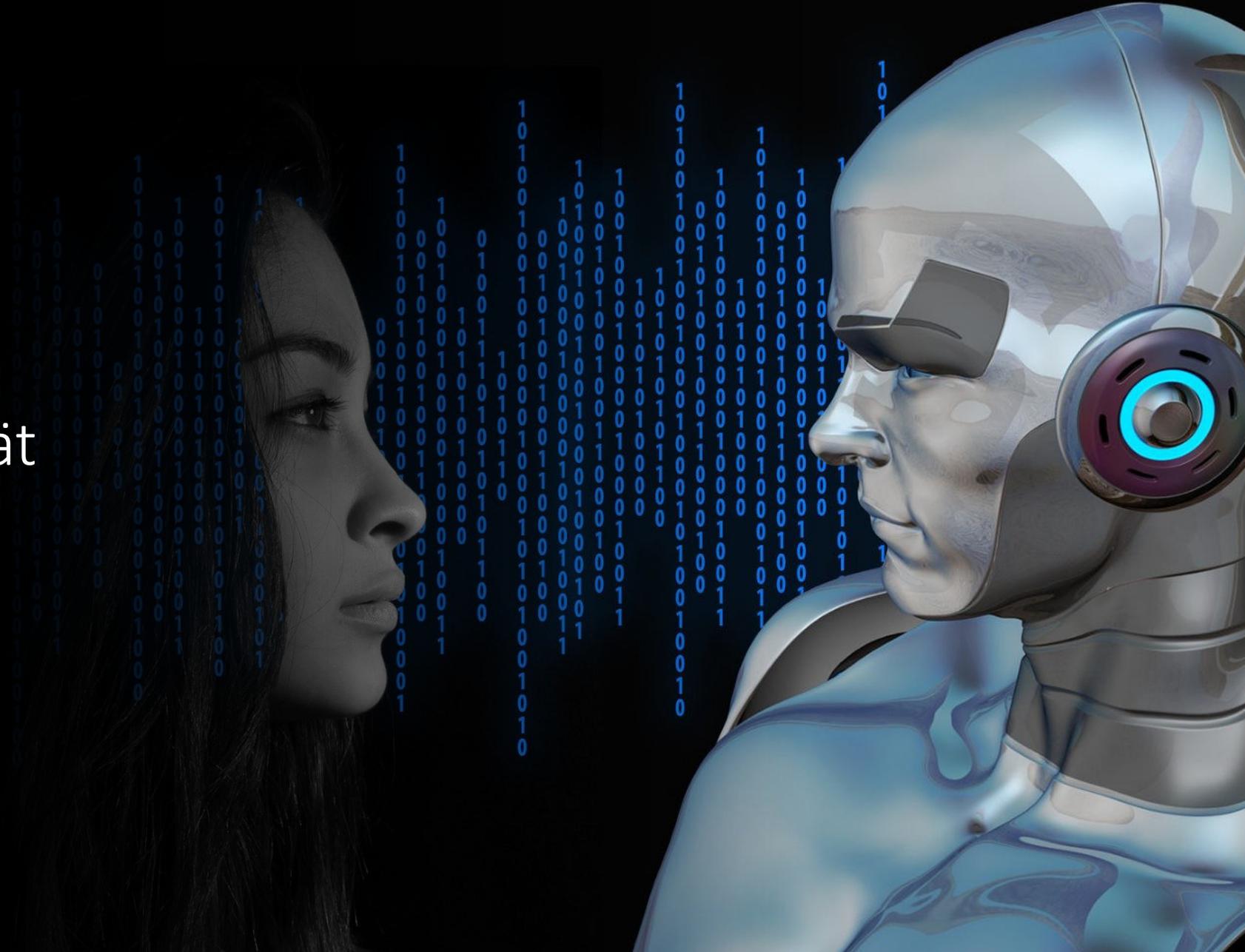


2. Resultat:

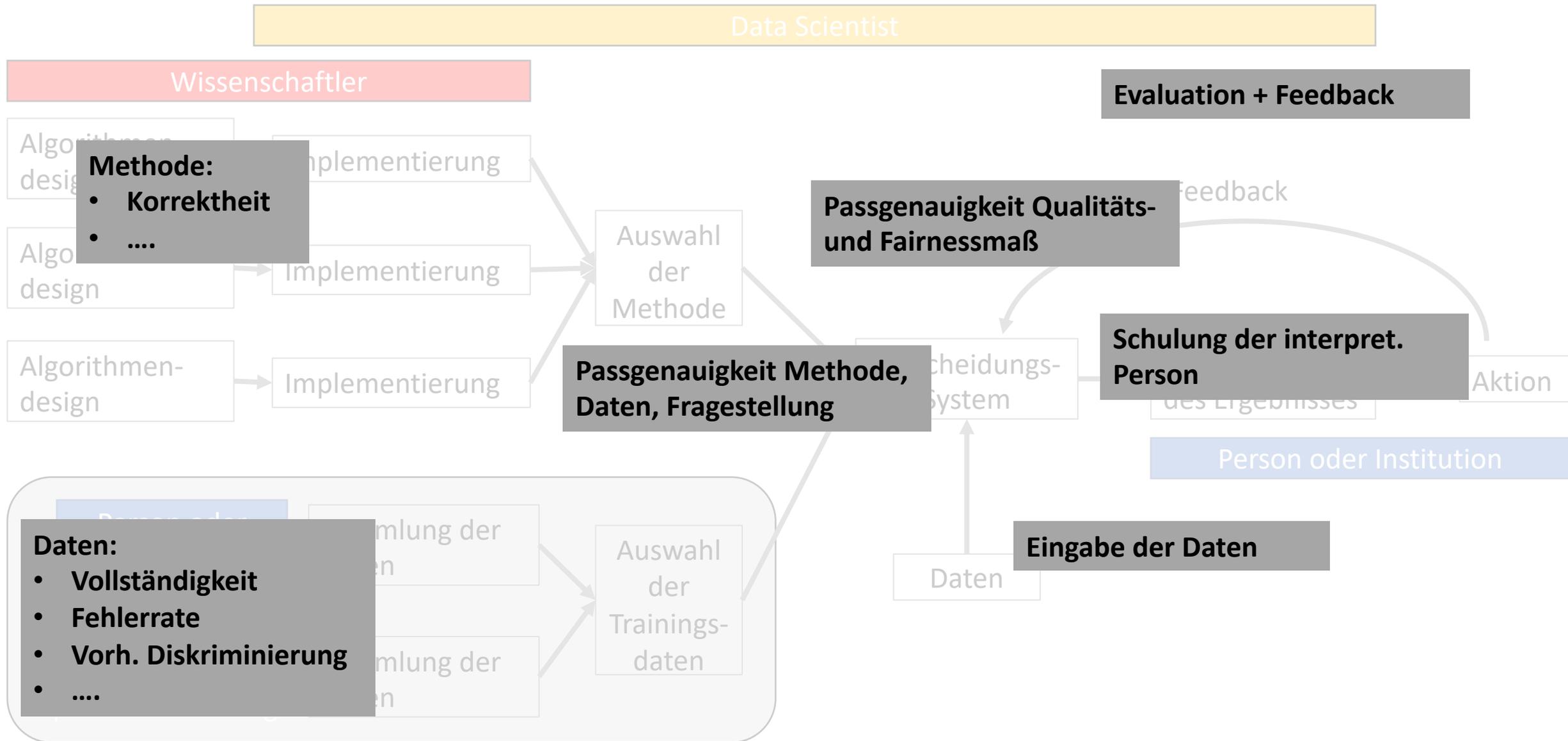
- Es werden nur 2 Handvoll Eigenschaften verwendet.
- **Ausreichend?**



Vergleich mit
menschlicher
Entscheidungsqualität
notwendig!



Wie kann vertrauenswürdige KI aussehen?



Ist richtende KI alternativlos?

- Nein!
- Versuchen Sie es zuerst mit Data Science!
- Lernen Sie gute Entscheidungsregeln aus der KI.
- Bilden Sie Mitarbeiter:innen mit den gewonnenen Erkenntnissen aus.
- Wenn möglich, stellen Sie mehr Mitarbeiter:innen ein.
- Ein KI-System ist niemals alternativlos.



Sichtende
(ergänzende)
KI

Dichtende
(ersetzende)
KI

Richtende
(bewertende)
KI

Zusammenfassung

- KI ist keine Magie, sie bedarf der Gestaltung.
- Bei wichtigen Entscheidungen müssen Entwicklungsparameter des KI-Systems mitbestimmt werden.
- Der Einsatz einer richtenden KI bedarf eines kontinuierlichen Evaluationsprozesses.





Wie gestalten
wir KI?

Menschliche Werte
kommen nur über
einen Weg in den
Rechner:

Über Sie, über mich,
über uns!

Weitere Literatur

Twitter: @NetworkTobi



[Zweig, Fischer & Lischka, 2018] Studie für die Bertelsmann-Stiftung: Zweig, Fischer & Lischka: „[Wo Maschinen irren können](#)“ (Serie AlgoEthik, No. 4, 2018)

[Zwei Kapitel im Sammelband \(Un\)Berechenbar?](#) des Fraunhofer FOKUS, Kompetenzzentrum ÖFIT, 2018

[Zweig & Krafft, 2018a] Zweig & Krafft: „Fairness und Qualität algorithmischer Entscheidungen“

[Krafft & Zweig, 2018b] Krafft & Zweig: „[Wie Gesellschaft algorithmischen Entscheidungen auf den Zahn fühlen kann](#)“

[Zweig, 2019] Studie für die Konrad-Adenauer-Stiftung „Algorithmische Entscheidungen, Zweig: „Transparenz und Kontrolle von algorithmischen Entscheidungssystemen“, 2019, <https://www.kas.de/analysen-und-argumente/detail/-/content/algorithmische-entscheidungen-transparenz-und-kontrolle>

[Krafft & Zweig, 2019] Krafft & Zweig: „Transparenz und Nachvollziehbarkeit algorithmenbasierter Entscheidungsprozesse“, Studie für die VZBV, 2019,

https://www.vzbv.de/sites/default/files/downloads/2019/05/02/19-05-02_vzbv_positionspapier_algorithmenkontrolle.pdf

[Klingel, Krafft & Zweig, 2020] A Klingel, TD Krafft und KA Zweig: „Mögliche Best Practice-Ansätze beim Einsatz eines algorithmischen Entscheidungsunterstützungssystems des AMAS-Algorithmus“, in “Digitaler Wandel und Ethik“, herausgegeben von Markus Hengstschläger und dem Rat für Forschung und Entwicklung, EcoWin Verlag, 2020