Entseelte Entscheidungen

Wenn Computer über Menschen entscheiden

IG-Metall Betriebsrätinnen Automobilbranche,
Kassel 6.6.2018

Prof. Dr. Katharina A. Zweig
Algorithm Accountability Lab
TU Kaiserslautern

@nettwerkerin

Diskriminierung bei Bewerbungen

- Lebensläufe mit „deutschen“ Namen bekommen 14% mehr Vorstellungs-angebote als solche mit „türkischen“ Namen[1].
- US-amerik. Studie: Frauen mit Kopftuch erhalten weniger Jobangebote als solche ohne[2].

1 Kaas, L. & Manger, C.: “Ethnic Discrimination in Germany's Labour Market: A Field Experiment”, German Economic Review, 2011 , 13 , 1-20
2 Ghumman, S. & Ryan, A. M.: “Not welcome here: Discrimination towards women who wear the Muslim headscarf , human relations, 2013 , 66(5) , 671-698

@nettwerkerin
Prof. KA Zweig
TU Kaiserslautern

Richter

- Richter müssen vorzeitige Haftentlassungsanträge begutachten.
- Studie: je weiter von der letzten Pause weg, desto weniger risikoreiche Entscheidungen[1].
- Eine Vielzahl solcher Studien scheint zu beweisen:

1 Danziger, S.; Levav, J. & Avnaim-Pesso, L.: “Extraneous factors in judicial decisions”, Proceedings of the National Academy of the Sciences, 2011 , 108 , 6889-6892

@nettwerkerin
Prof. KA Zweig
TU Kaiserslautern

Menschen – so irrational!

- Richter müssen vorzeitige Haftentlassungsanträge begutac
- Studie: letzten weniger Entsche
- Eine Vielzahl solcher Studien scheint zu beweisen:

Menschen sind irrational und vorurteilsbeladen.

1 Danziger, S.; Levav, J. & Avnaim-Pesso, L.: “Extraneous factors in judicial decisions”, Proceedings of the National Academy of the Sciences, 2011 , 108 , 6889-6892

@nettwerkerin
Prof. KA Zweig
TU Kaiserslautern

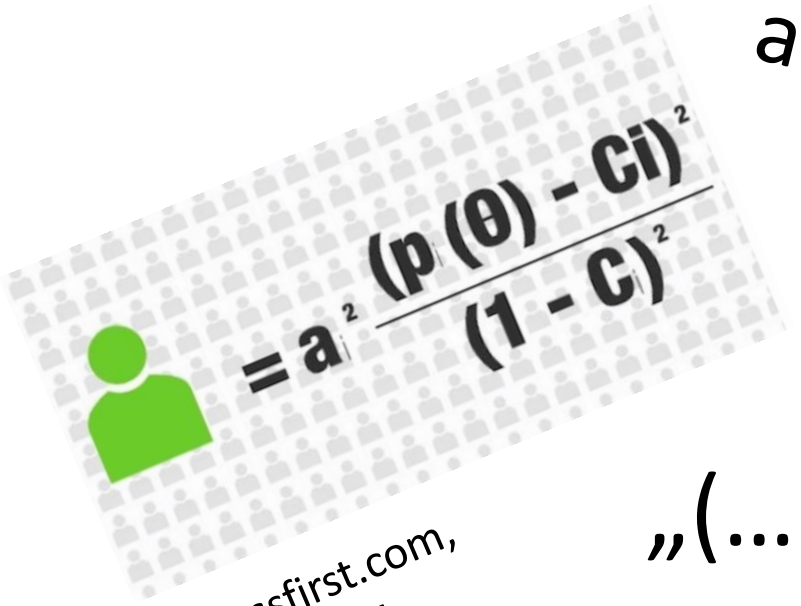
Könnten Computer das besser?

- Die ersten Firmen testen *algorithmische Entscheidungssysteme*[1].
- Eigenschaften, nach denen nicht diskriminiert werden darf, können vor ihnen besser verborgen werden.
- Sie entscheiden konsistent.

1 Claire Miller: “Can an Algorithm hire Better than a Human?”, The New York Times, June 25, 2015, https://www.nytimes.com/2015/06/26/upshot/can-an-algorithm-hire-better-than-a-human.html

@nettwerkerin
Prof. KA Zweig
TU Kaiserslautern

„Employment assessment software“



Assessfirst.com,
16.11.2017

Let‘s take the emotion out of the process and replace it with a data-driven approach…“

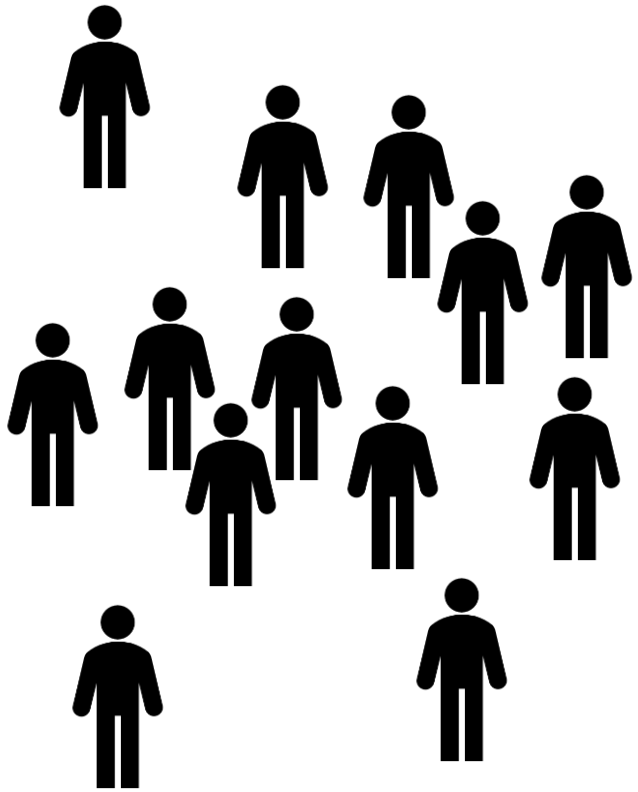
iNostix (by Deloitte),
16.11.2017

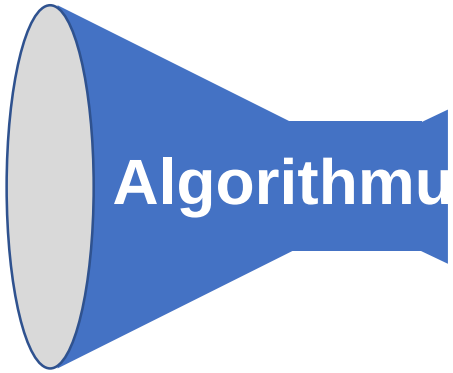
„(…) with the availability of good data, the predictive possibilities are virtually unlimited (…)”

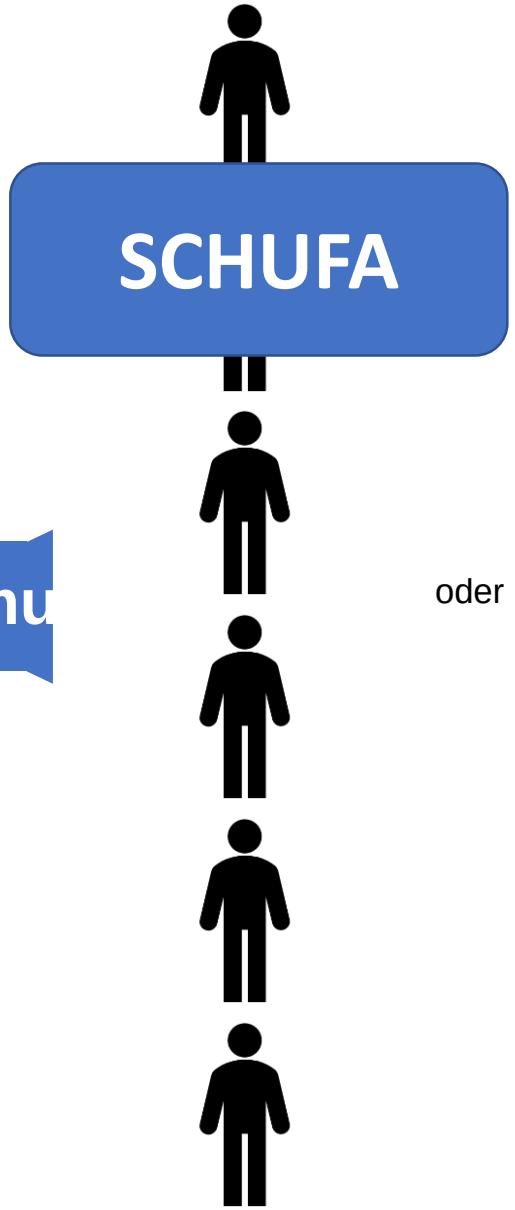
https://www.inostix.com/predict-hiring-success/
16.11.2017

@nettwerkerin
Prof. KA Zweig
TU Kaiserslautern

Algorithmische Entscheidungssysteme

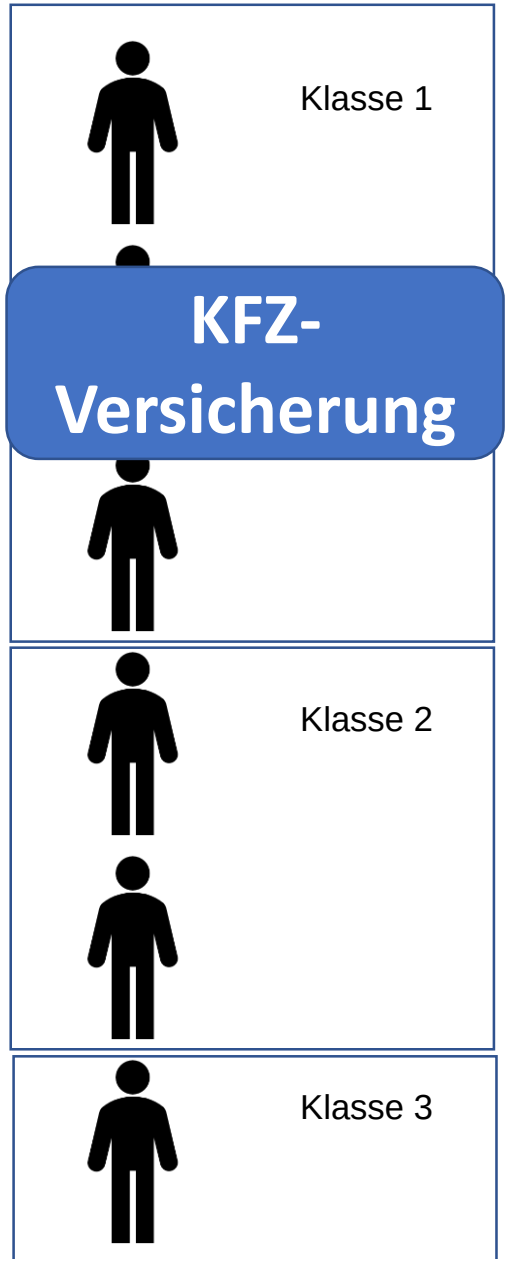






Scoring-Verfahren

oder



Klassifikation

Forschung

Vorhersage des Rückfälligkeitsrisiko von Kriminellen

Das kleine ABC der Informatik

Können

Algorithmen,

Big Data und

Computerintelligenz

Menschen besser bewerten und richten als Menschen?

A wie Algorithmus

Ein Algorithmus ist ein Problemlöser

Mathematisches Problem

Der PUTPUT der uns sagt, wie Input mit Output zusammenhängt.

INPUT

OUTPUT

Beispiel für ein Problem: Navigation

Navigation

Gegeben das Karten-material und weitere Daten, berechne die kürzeste Route zwischen Start und Ziel

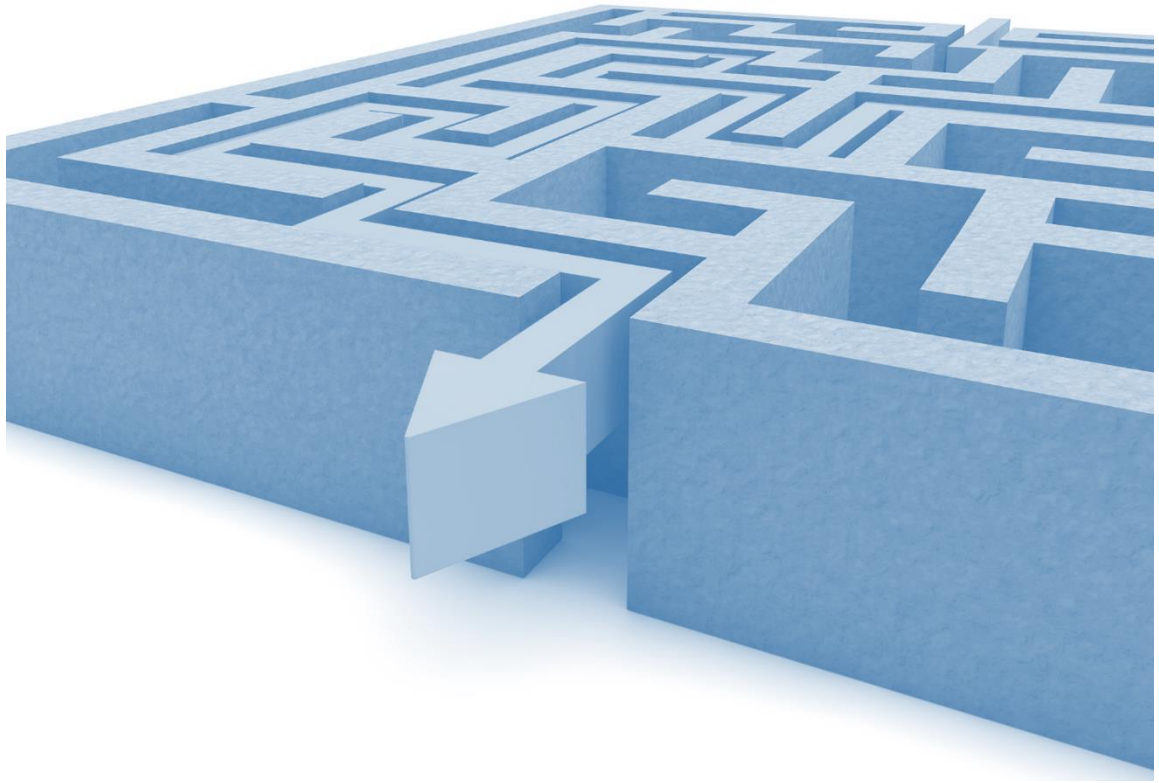
Das **Problem** sagt nicht, wie man die Lösung **findet**.

Input: Straßen, Länge, Staus, …
Start und Ziel

Output: optimale Route

Ein Algorithmus ist…

…eine für jede **erfahrene Programmiererin ausreichend detaillierte Lösungsvorschrift**, so dass bei **korrekter Implementierung** der Computer **für jede korrekte Inputmenge den korrekten Output** berechnet – in endlicher Zeit.

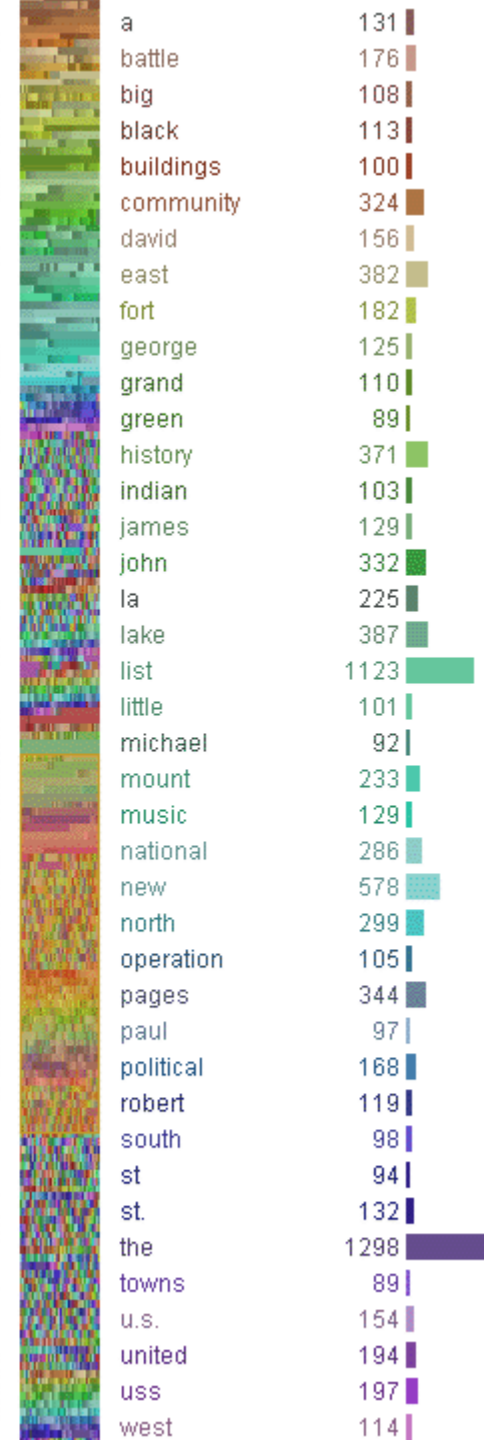


B wie Big Data

Daten als Grundlage

Big Data

- **Große Datenmengen.**
- **Außerhalb ihres spezifischen Zwecks genutzt.**
- **Daher im Einzelnen vermutlich fehlerbehaftet.**
- **Dank großer Masse und wenig individualisiertem Verhalten statistisch nutzbar.**
- **Hier werden Methoden des maschinellen Lernens benötigt.**



C wie Computerintelligenz

Sebastian lernt „heiss“ und „warm“

Juli **März** **September**

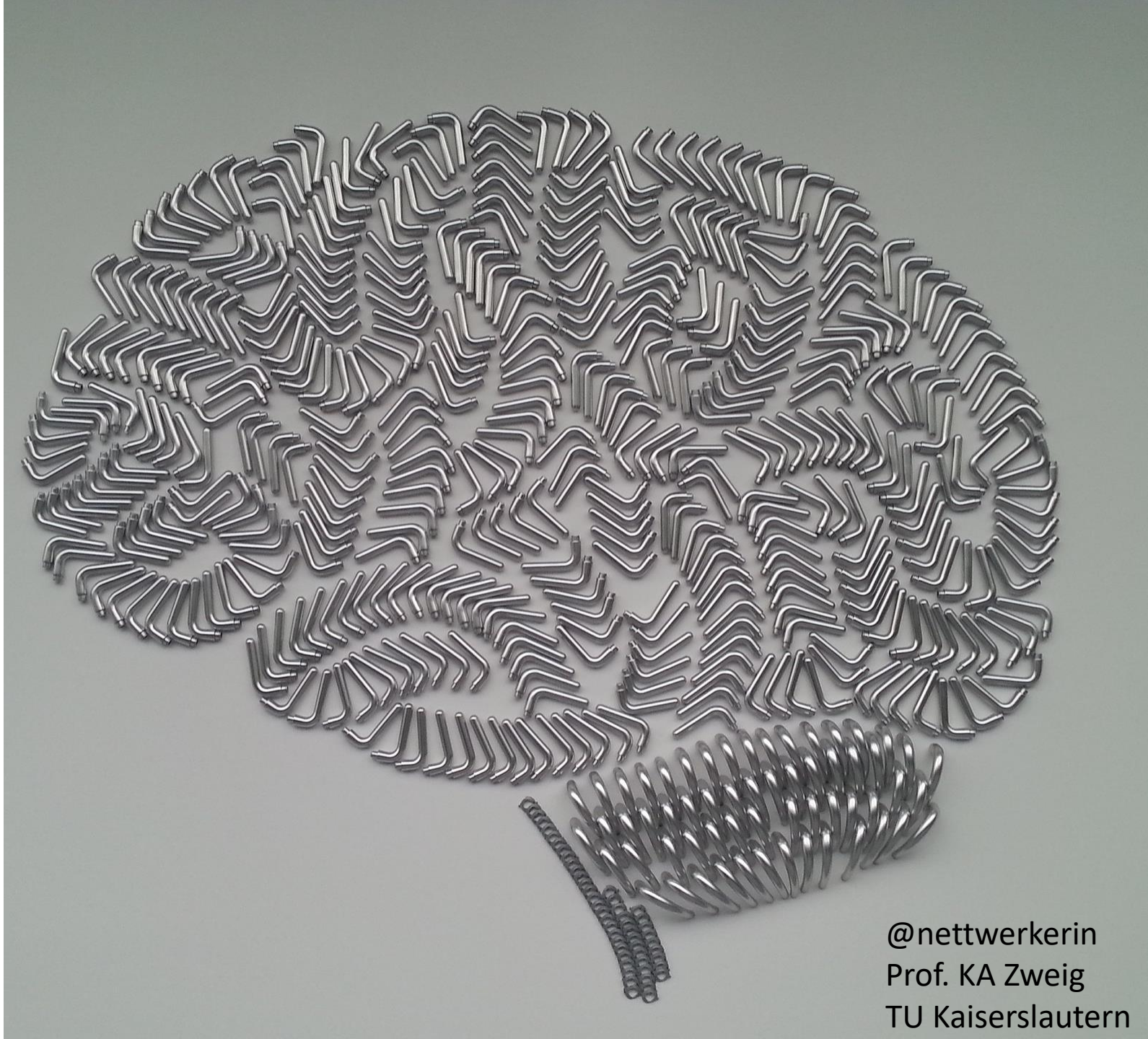
Zu vorsichtig: Darf nicht dampfen Zu mutig geworden

Alles muss kalt sein

@nettwerkerin
Prof. KA Zweig
TU Kaiserslautern

Sebastian lernt…

- Durch **Rückkopplung**: unerwartet heiß, unerwartet kalt
- Durch **Speicherung in einer Struktur:** in Neuronen und deren Verknüpfung.
- Durch **Generalisierung des Gelernten**.



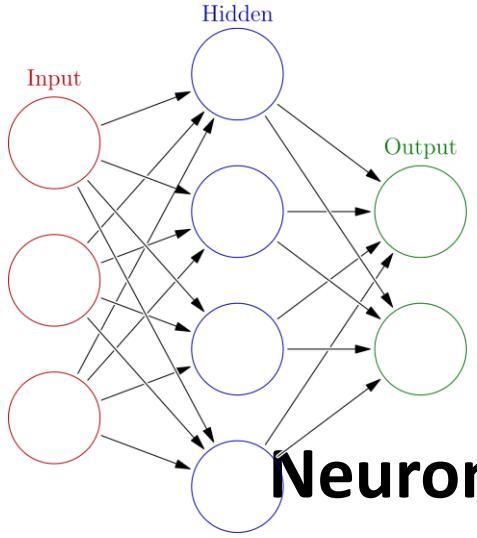
@nettwerkerin
Prof. KA Zweig
TU Kaiserslautern

Computer lernen

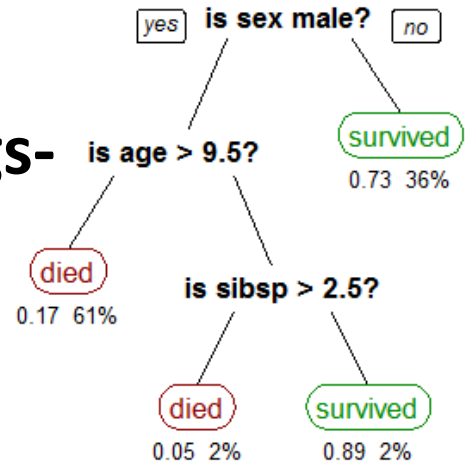
Damit ein Computer lernen kann, benötigt er ebenfalls eine **Struktur,** um Gelerntes abzuspeichern.

Optimal auch **Rückkopplung**.

Er lernt **generelle Regeln**.



Neuronales Netz

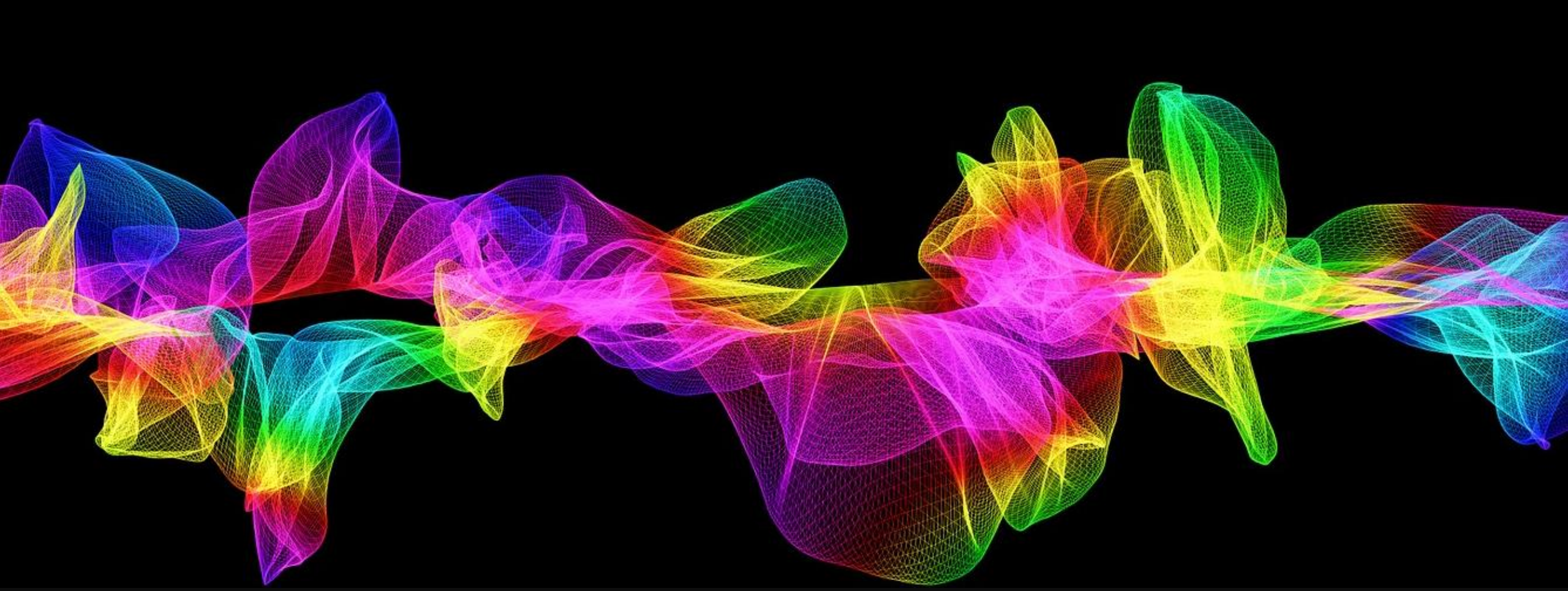


Entscheidungs-bäume

$$w_1 * \#Vh - w_2 * \#dayiVh + w_3 * I[g = male] * 1 + w_4 * I[T = R] * 1.0 + \cdots$$

Formel

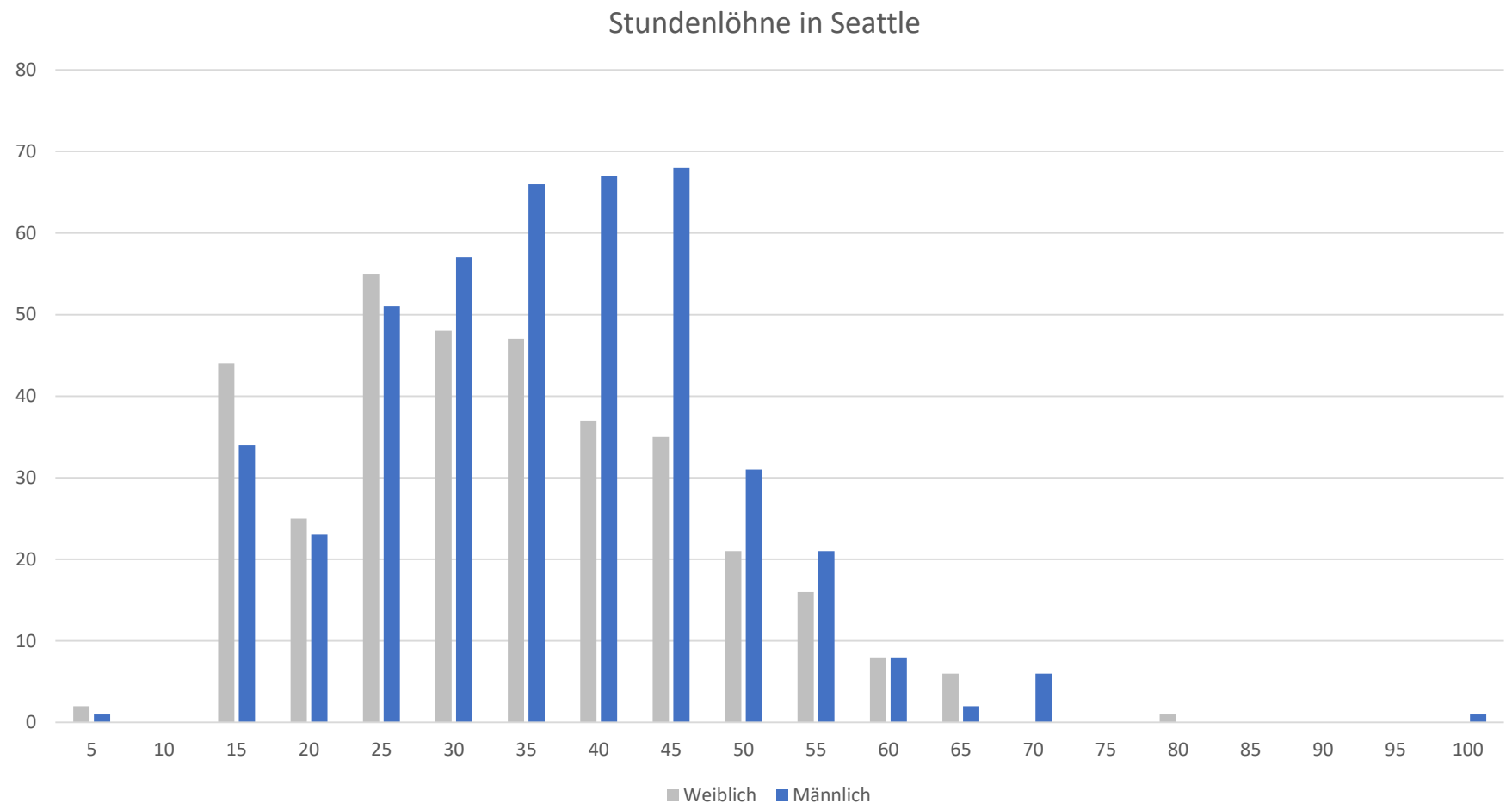
@nettwerkerin
Prof. KA Zweig
TU Kaiserslautern

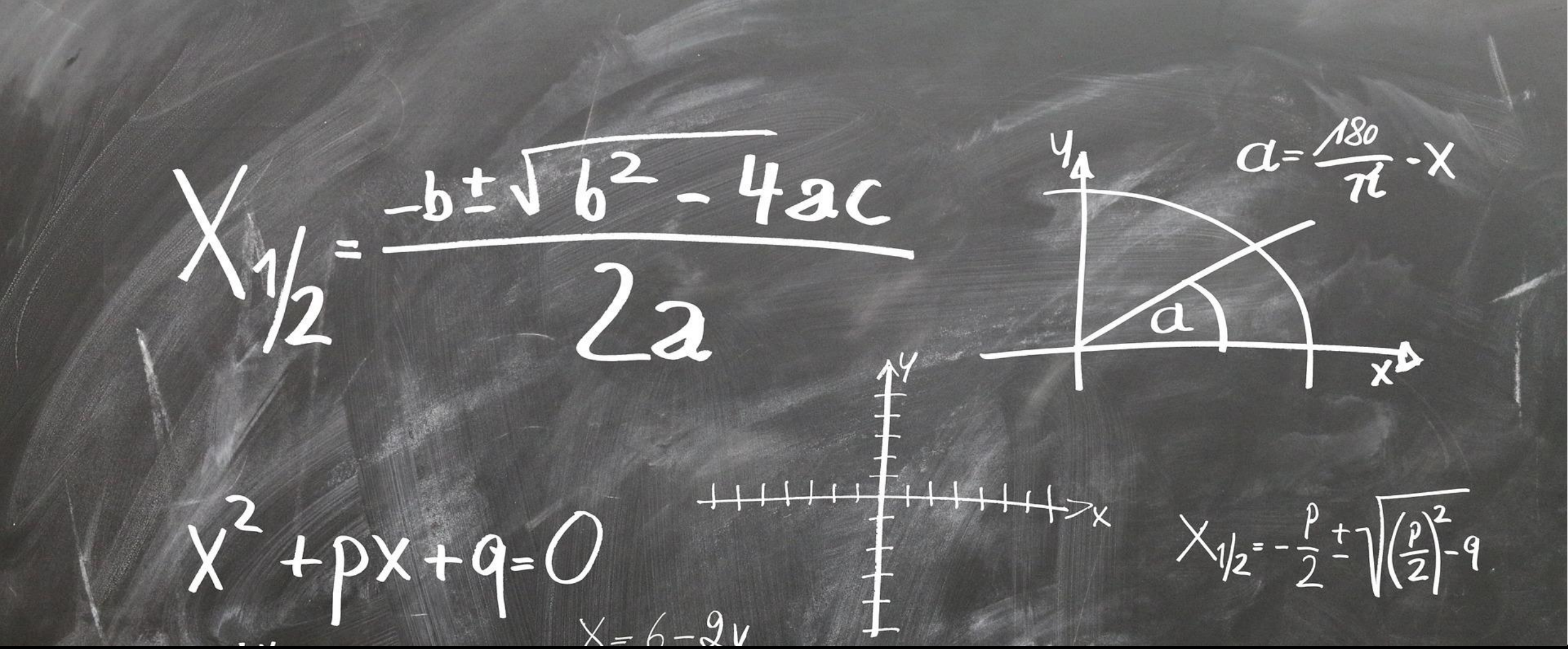


“Lernen” mit Korrelationen

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

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





Lernen mit Formeln

Datengrundlagen

- Data Mining Methoden nutzen verschiedene Informationen
- Am wichtigsten:
  - **War Einstellung erfolgreich?**

Regressionsansätze

- Die Algorithmen-designer entscheiden, welche Daten vermutlich mit „erfolgreicher Einstellung“ korrelieren.
- Die Software sollte eine einzige Zahl ausgeben.
- Je höher die Zahl, desto höher die Erfolgs-wahrscheinlichkeit.
- Beispiel Formel:

$$
\begin{aligned}
&3 * \text{ Jahre im Job} \\
&- 2 * \text{ Anzahl Arbeitgeber} \\
&+ 3 * (\text{Wenn Auslandserfahrung,} \\
&\qquad \text{dann 1, sonst 0}) \\
&+ 2,5 * (\text{Wenn Fortbildung,} \\
&\qquad \text{dann 1, sonst 0}) + \ldots
\end{aligned}
$$

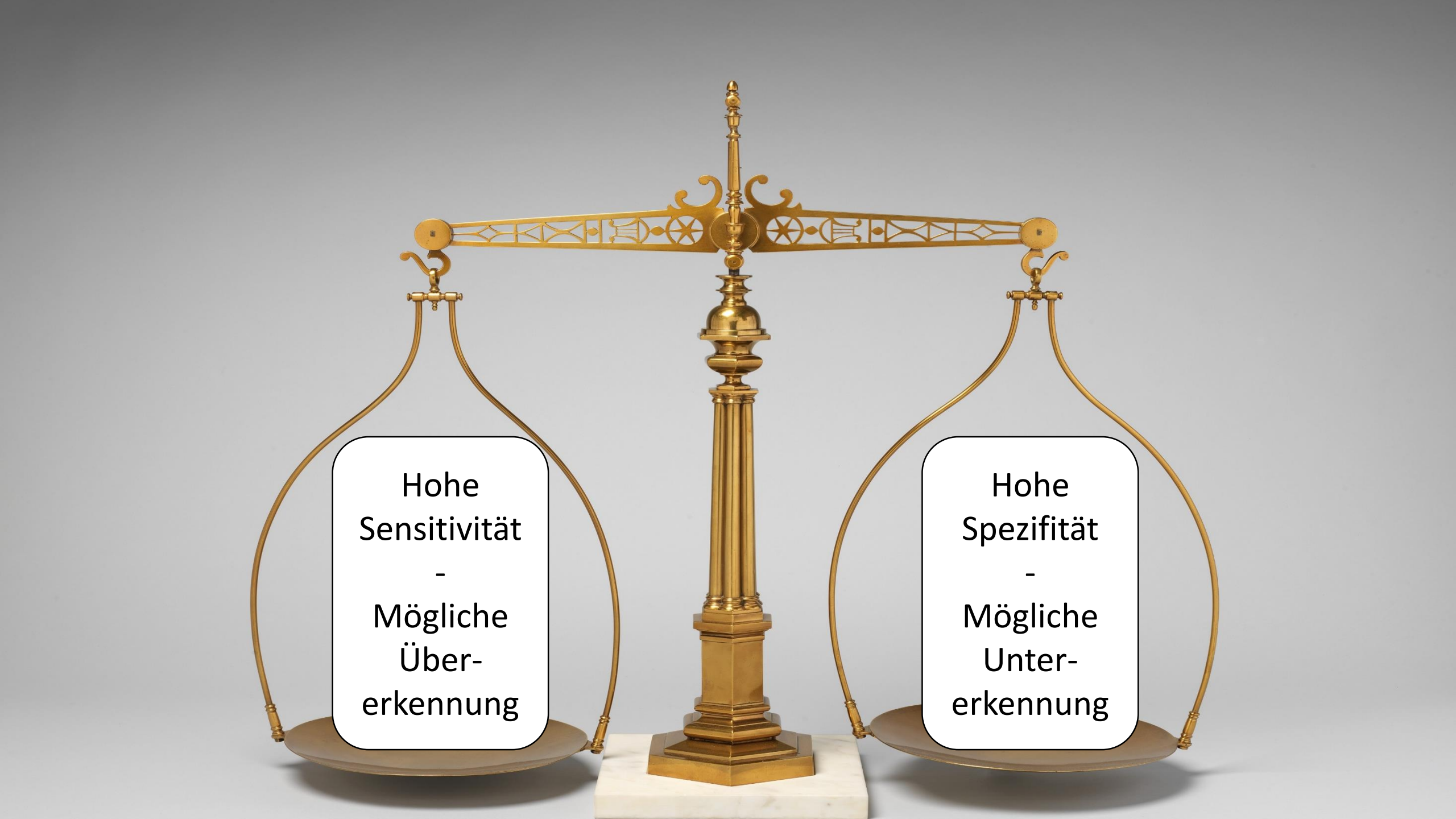
Allgemein

Der Computer bestimmt die Gewichte und bekommt ein Feedback (Rückkopplung), inwieweit die resultierende Bewertung mit dem (beobachteten) Verhalten übereinstimmt.

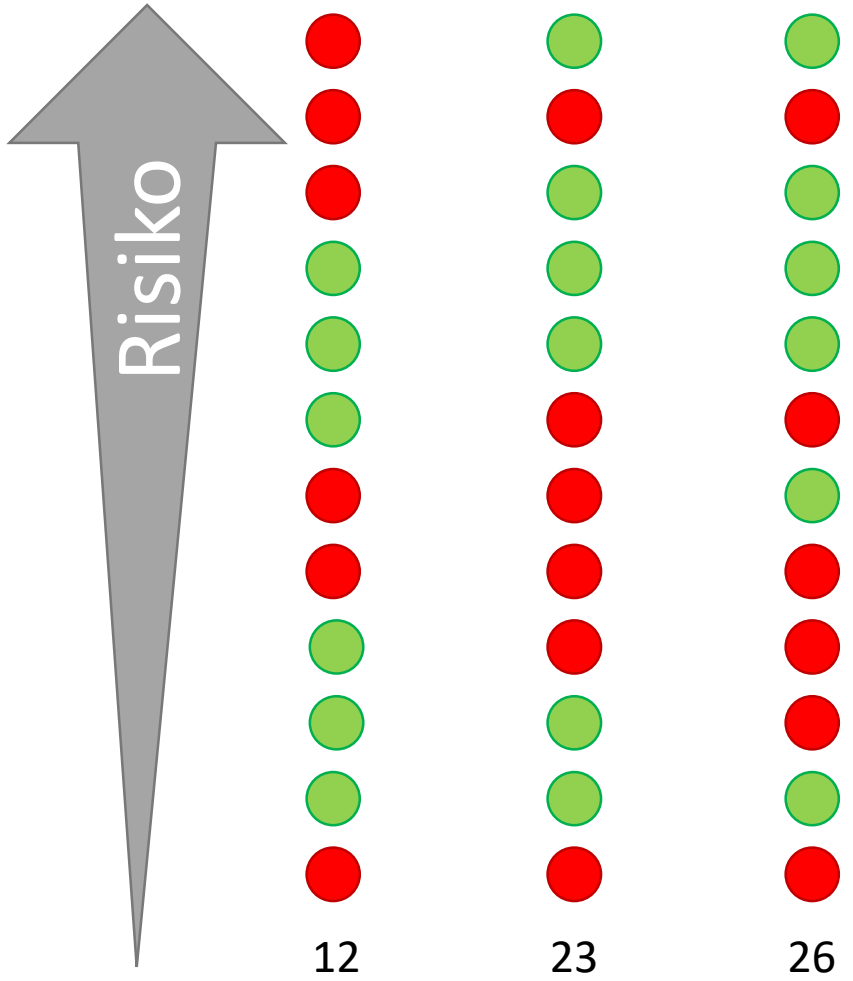
$$
\begin{aligned}
& w_1 * \text{ Jahre im Job} \\
+ & w_2 * \text{ Anzahl Arbeitgeber} \\
+ & w_3 * (\text{Wenn Auslandserfahrung,} \\
& \quad \text{dann 1, sonst 0}) \\
+ & w_4 * (\text{Wenn Fortbildung,} \\
& \quad \text{dann 1, sonst 0}) + \ldots
\end{aligned}
$$

Qualität eines Algorithmus


ROC AUC
...und 20 mehr
Positive Predictive Value
Sensitivität
Qualitäts-Maße
Accuracy


Hohe Sensitivität - Mögliche Über-erkennung
Hohe Spezifität - Mögliche Unter-erkennung

- Grüne Kugeln symbolisieren erfolgreiche, rote nicht erfolgreiche BewerberInnen.
- Optimale Sortierung: Alle grünen oben, alle roten darunter.
- Qualitätsmaß: Paare von rot und grün, bei denen die grüne Kugel über der roten einsortiert ist. (**ROC AUC**)



@nettwerkerin
Prof. KA Zweig
TU Kaiserslautern

Ist das Qualitätsmaß sinnvoll?

- Wenn die Stelle **sofort** besetzt werden muss, und nur 5 Bewerber da sind: **ja**
- Wenn es langfristig um die Identifikation der besten Talente geht: **nein**
- Hier müssen andere Qualitätsmaße benutzt werden.

@nettwerkerin
Prof. KA Zweig
TU Kaiserslautern

einen Jagdhund zu kaufen,

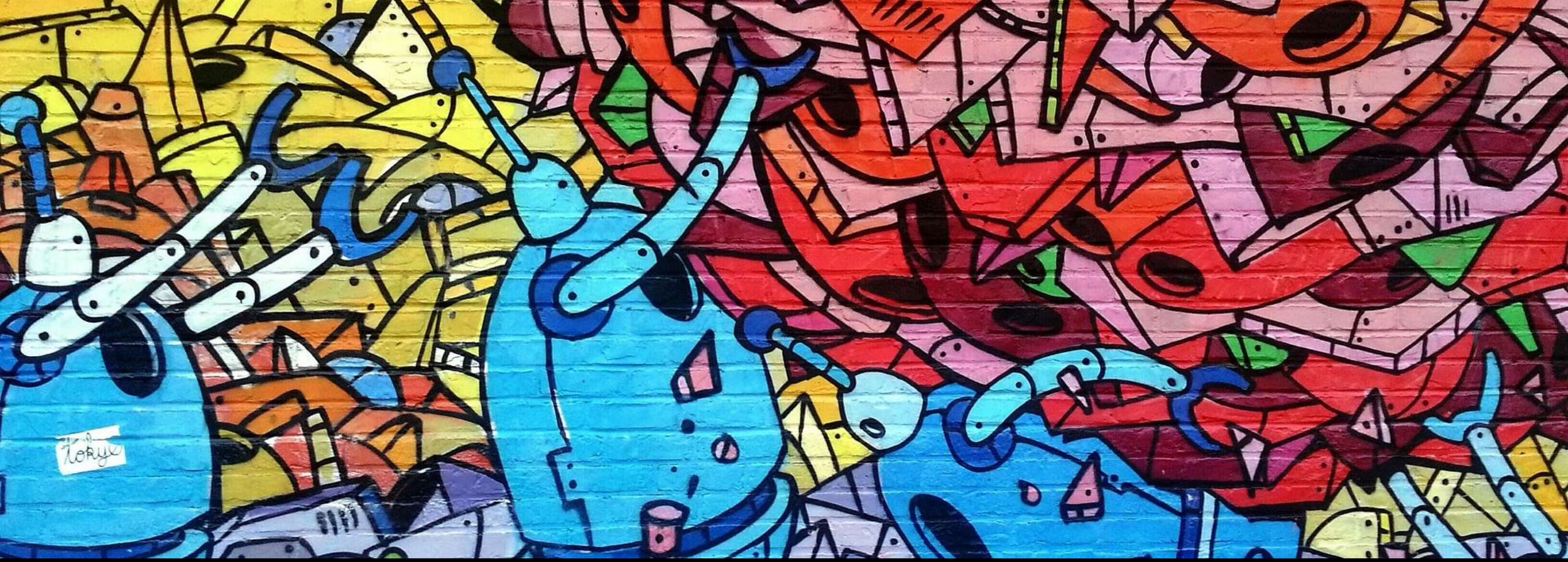
um Schafe zu hüten.

Das ist wie...

Probleme von algorithmischen Entscheidungssystemen (ADM Systemen) im People Assessment

1. **Wer entscheidet, wann ein ADM System „gut“ ist?**

@nettwerkerin
Prof. KA Zweig
TU Kaiserslautern



Wahrscheinlichkeit & Wahrheit

Regel

Algorithmen der künstlichen Intelligenz werden da eingesetzt, wo es **keine einfachen Regeln** gibt.

Sie suchen **Muster** in hoch-verrauschten Datensätzen.

Die Muster sind daher grundsätzlich **statistischer Natur.**

Versuchen fast immer, eine **kleine Gruppe** von Menschen zu identifizieren (Problem der **Unbalanciertheit**)

@nettwerkerin
Prof. KA Zweig
TU Kaiserslautern

Algorithmen…

- … basieren auf Korrelationen von Eigenschaften mit gewünschtem Verhalten.
- Quasi **algorithmisch legitimierte Vorurteile**:
  - Zu 70% erfolgreich heißt:
  - Von 100 Personen, die „genau so sind wie dieser Mensch“, sind 70 nachher erfolgreich.

Probleme von algorithmischen Entscheidungssystemen (ADM Systemen) im People Assessment

1. Wer entscheidet, wann ein ADM System „gut“ ist?
2. **ADM Systeme ergeben nur Wahrscheinlichkeiten, keine Wahrheiten.**

@nettwerkerin
Prof. KA Zweig
TU Kaiserslautern

Können Algorithmen diskriminieren?

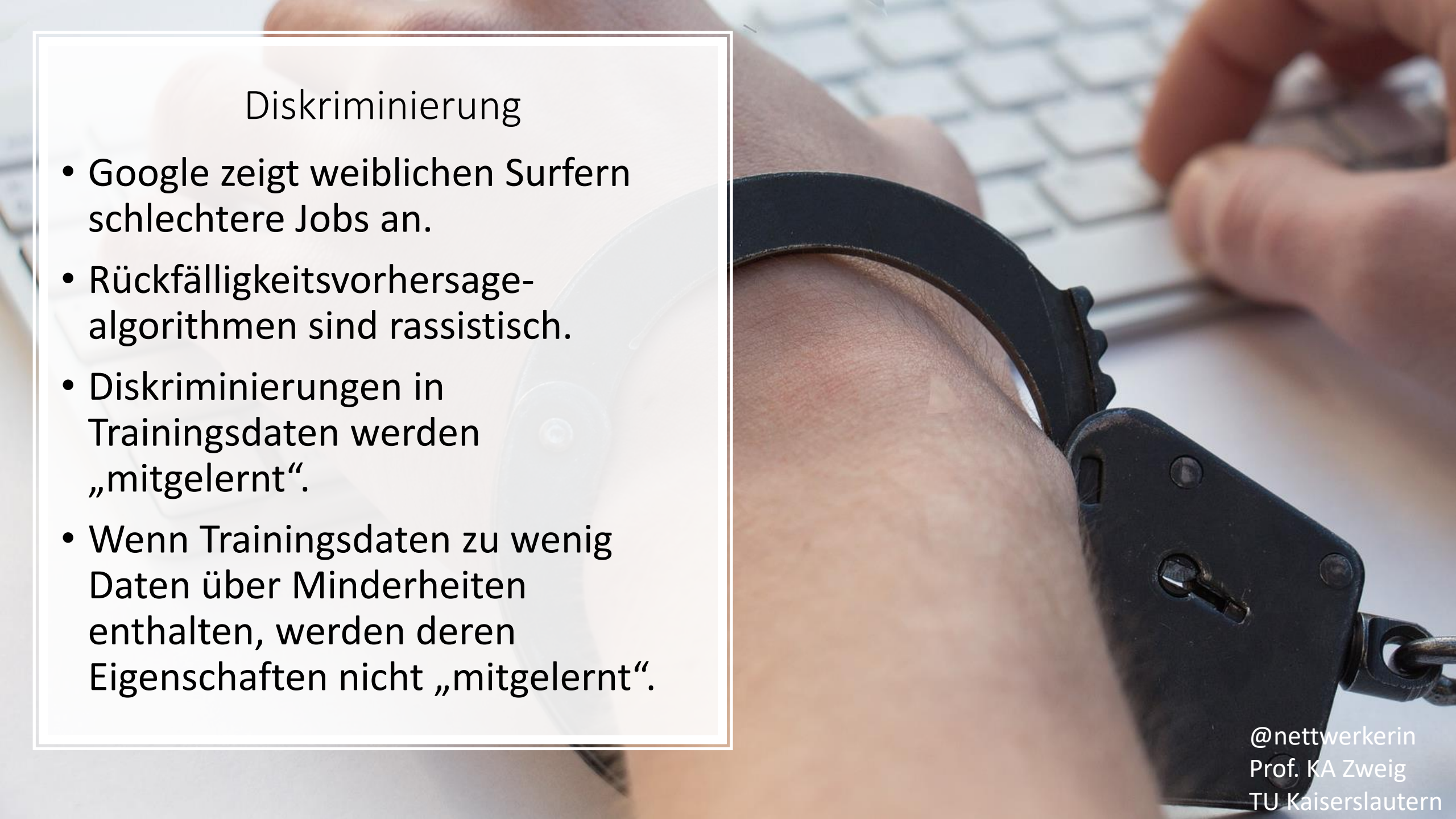
Gleichberechtigung

Wenn man auf Google nach „CEO“ sucht…

Und das, wenn ich auf Pixabay nach „Chef“ suche…

Diskriminierung

- Google zeigt weiblichen Surfern schlechtere Jobs an.
- Rückfälligkeitsvorhersage-algorithmen sind rassistisch.
- Diskriminierungen in Trainingsdaten werden „mitgelernt“.
- Wenn Trainingsdaten zu wenig Daten über Minderheiten enthalten, werden deren Eigenschaften nicht „mitgelernt“.

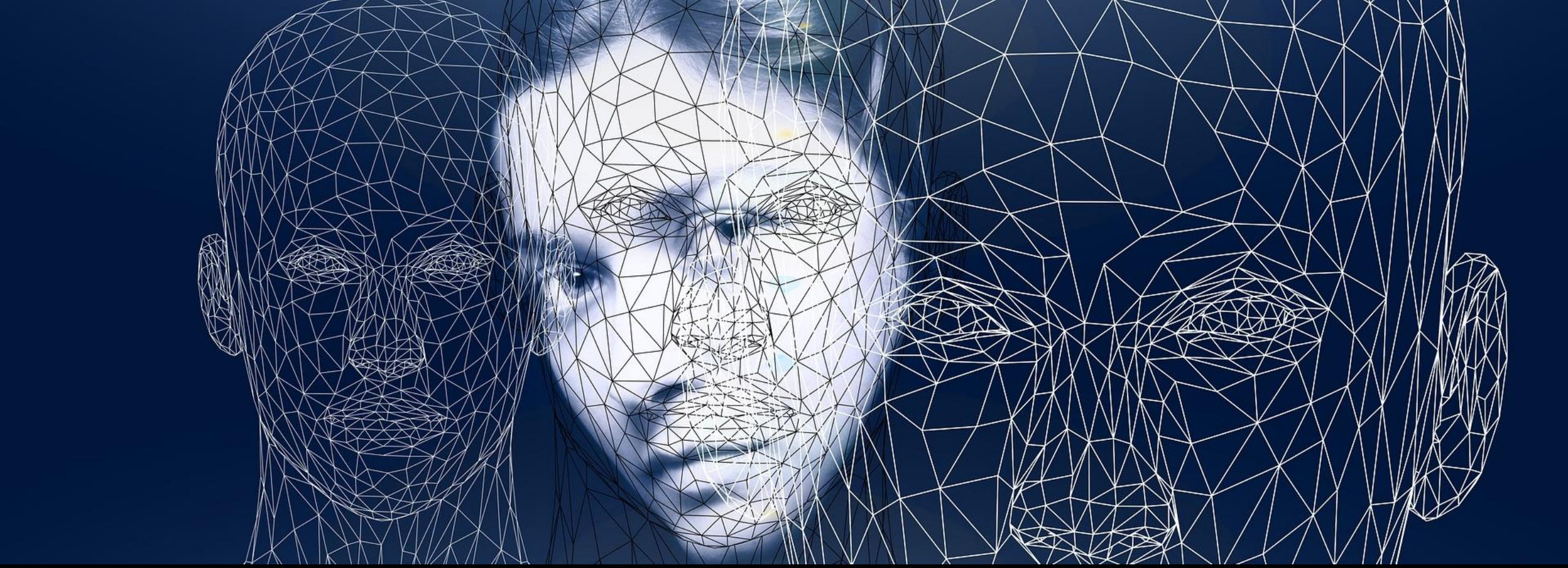


@nettwerkerin
Prof. KA Zweig
TU Kaiserslautern

Probleme von algorithmischen Entscheidungssystemen (ADM Systemen) im People Assessment

1. Wer entscheidet, wann ein ADM System „gut“ ist?
2. ADM Systeme ergeben nur Wahrscheinlichkeiten, keine Wahrheiten.
3. **ADM Systeme können diskriminieren.**

@nettwerkerin
Prof. KA Zweig
TU Kaiserslautern



Sozio-informatische Gesamtbetrachtung

Probleme der Einbettung der ADM in den sozialen Prozess

- **Aufmerksamkeitsökonomie** von Entscheiderinnen und Entscheidern.
- **„Best practice“** erfordert Nutzung der Software.
- **Delegation von Verantwortung!**
- Manchmal kann ein falsch Beurteilter **die Vorhersage prinzipiell nicht entkräften!**
  - Z.B. abgelehnte Bewerberin

@nettwerkerin
Prof. KA Zweig
TU Kaiserslautern

Probleme von algorithmischen Entscheidungssystemen (ADM Systemen) im People Assessment

1. Wer entscheidet, wann ein ADM System „gut“ ist?
2. ADM Systeme ergeben nur Wahrscheinlichkeiten, keine Wahrheiten.
3. ADM Systeme können diskriminieren.
4. **ADM Systeme können soziale Prozesse verändern.**

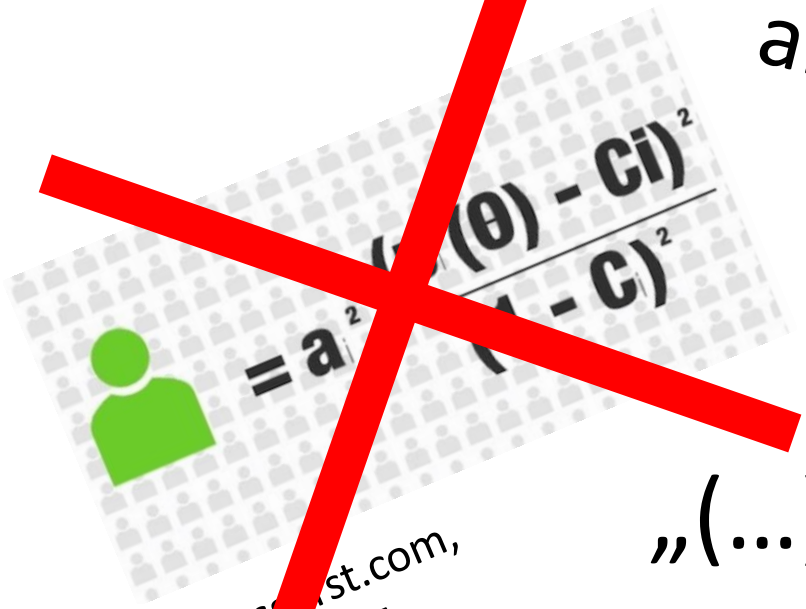
@nettwerkerin
Prof. KA Zweig
TU Kaiserslautern

Probleme von algorithmischen Entscheidungssystemen (ADM Systemen) im People Assessment

1. **Wer entscheidet, wann ein ADM System „gut“ ist?**
2. **ADM Systeme ergeben nur Wahrscheinlichkeiten, keine Wahrheiten.**
3. **ADM Systeme können diskriminieren.**
4. **ADM Systeme können soziale Prozesse verändern.**

@nettwerkerin
Prof. KA Zweig
TU Kaiserslautern

„Employment assessment software“



Assess[illegible]st.com,
16.11.2017

Let‘s take the emotion out of the process and replace it with a data-driven approach…“

iNostix (by Deloitte),
16.11.2017

„(…) with the availability of good data, the predictive possibilities are virtually unlimited (…)”

https://www.inostix.com/[illegible]dict-hiring-success/
16.11.2017

@nettwerkerin
Prof. KA Zweig
TU Kaiserslautern

Einschätzung

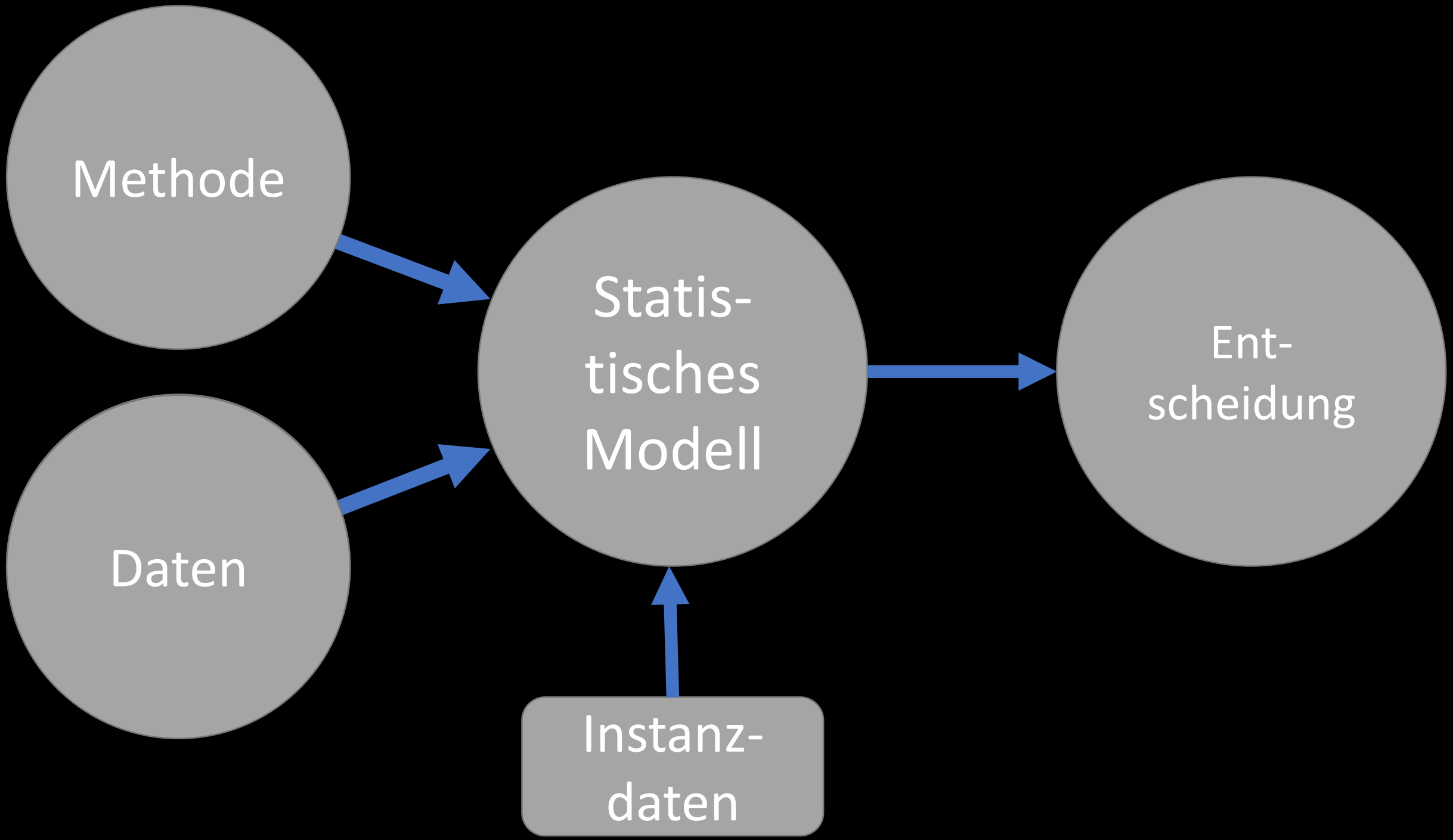
- People Assessment Systeme könnten dabei helfen, bessere Entscheidungen zu treffen.
- Allerdings ist es schwierig, sie transparent, fair und nachvollziehbar zu gestalten.

@nettwerkerin
Prof. KA Zweig
TU Kaiserslautern

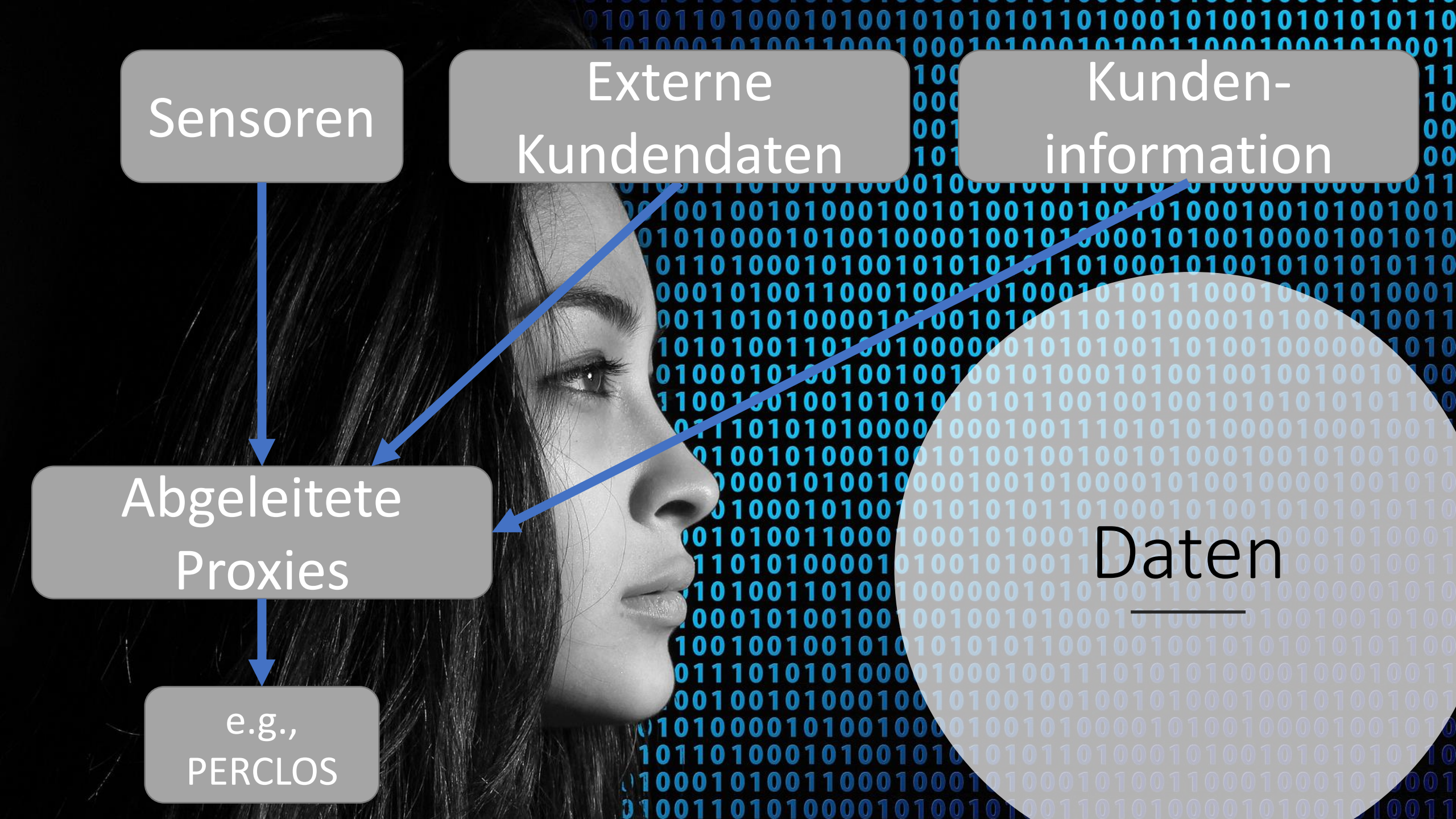
Ethik in der Softwareentwicklung

…in der Automobilbranche


Lassen Sie uns einen
Aufmerksamkeits-
Sensor bauen!


Methode
Daten
Statis-
tisches
Modell
Ent-
scheidung
Instanz-
daten

Daten ⇨ Datenauswahl


Sensoren
Externe Kundendaten
Kunden-information
Abgeleitete Proxies
e.g., PERCLOS
Daten

Google Patents

Drowsiness detection system

Abstract

This invention describes a non-intrusive system used to dete
and at risk of falling asleep at the wheel due to drowsiness.
drowsiness detection systems and a control unit. This redur
drowsiness assessment. The first subsystem consists of an
headliner and seat, which detects head movements that are
driver. The second subsystem consists of heart rate monitor
wheel. The control unit is used to analyze the sensory data a
state and therefore corresponding risk of falling asleep while
intelligent software algorithms, and the data provided by the
characteristics that may indicate a drowsy driver. If the drive
outputted which may be used to activate a response system
automobiles; this system may be used in any type of vehicle

Kopfbewegungen

Herzschlag

Augenlider (PERCLOS)

B2

OF Find Prior Art Similar

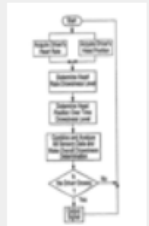
dam Basir, Jean Pierre Bhavnani, Fakhreddine Desrochers

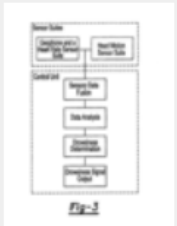
: Intelligent Mechatronic Systems Inc

: Intelligent Mechatronic Systems Inc

2-01-18

Images (3)





Classifications

G08B21/06 Alarms for ensuring the safety of persons indicating a condition of sleep, e.g. anti-dozing alarms

View 1 more classifications

Family: US (1)

Date	App/Pub Number	Status
2003-01-21	US10348037	Active
2003-08-14	US20030151516A1	Application
2004-11-23	US6822573B2	Grant

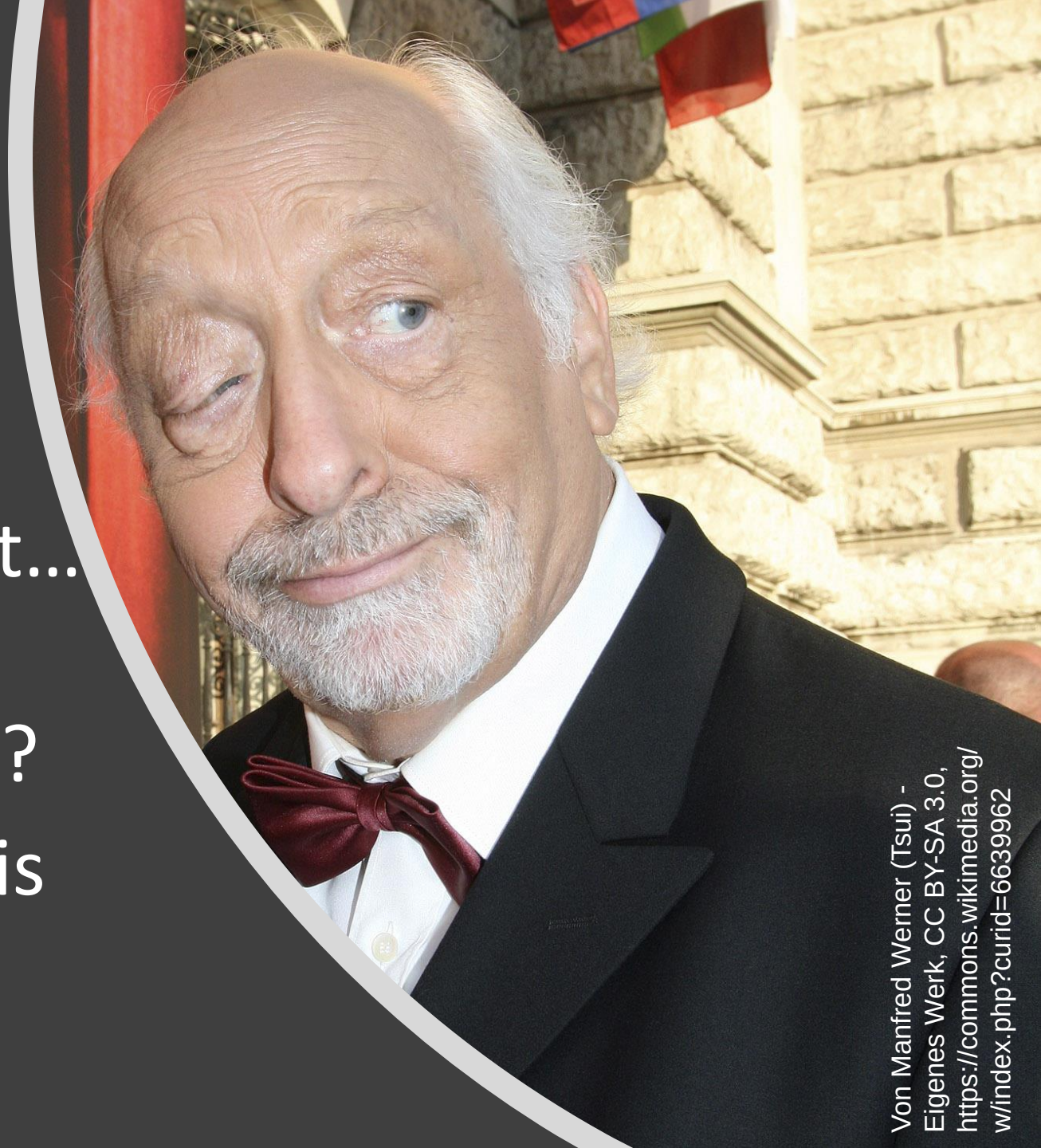
Info: Patent citations (14), Cited by (47), Legal events, Similar documents, Priority and Related Applications

External links: USPTO, USPTO Assignment, Espacenet, Global Dossier, Discuss

Ethische Überlegungen bei der Datenauswahl

Durch die Videosensorik diskriminieren wir vielleicht…

- …Personen mit trockenen Augen oder Kontaktlinsen?
- …Personen mit einer Ptosis wie Karl Dall?



Methoden-
auswahl

Daten-
auswahl

Qualitäts-
maße

Statistisches
Modell

Ent-
scheidung

Instanz-
Daten

Soziale Einbettung des Systems

- Wird sich der Fahrer oder die Fahrerin auf das System verlassen?
- Wollen Speditionen Zugriff auf Fahrerdaten?
- Könnte das System auch für Büros nützlich sein?
  - Welche Verantwortlichkeiten erwachsen daraus?

Methoden-
auswahl

Daten-
auswahl

Qualitäts-
maße

Statistisches
Modell

Ent-
scheidung

Instanz-
daten

Methoden-
auswahl

Daten-
auswahl

Qualitäts-
maße

Statistisches
Modell

Inter-
pretation

Instanz-
daten

Methoden-
auswahl

Daten-
auswahl

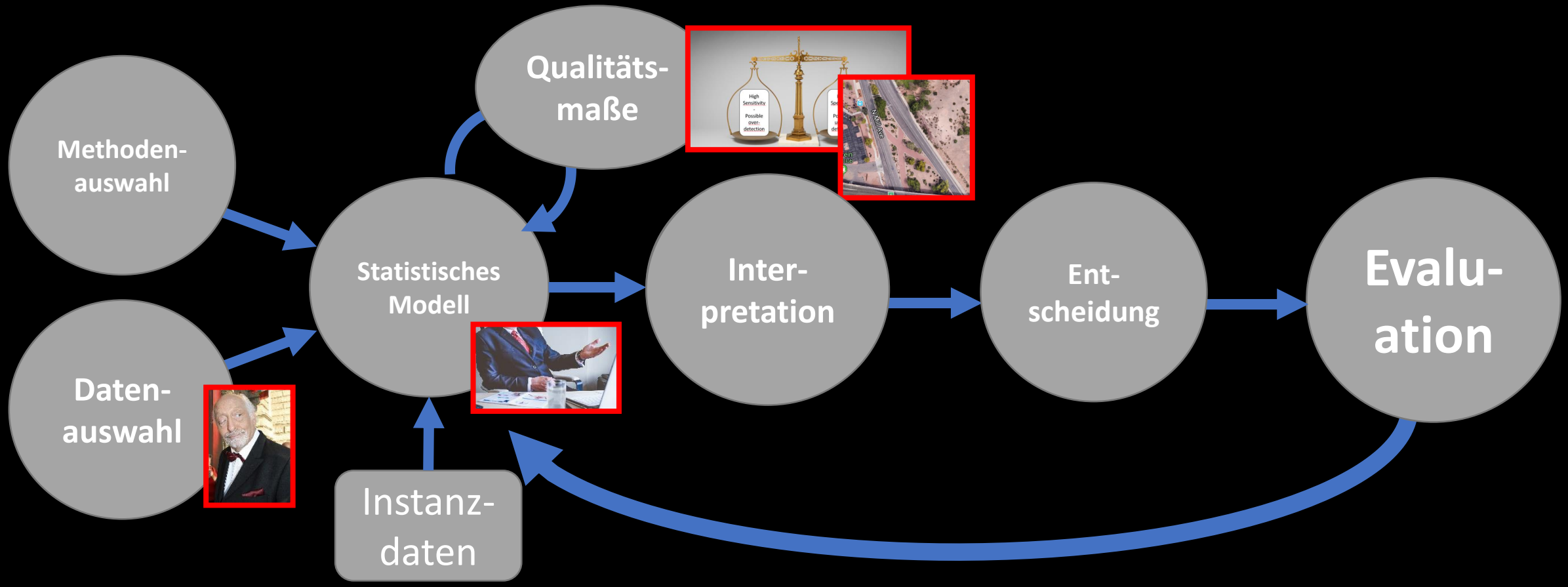
Qualitäts-
maße

Statistisch
es Modell

Instanz-
daten

Inter-
pretation

Ent-
scheidung


Methoden-
auswahl
Daten-
auswahl
Qualitäts-
maße
Statistisches
Modell
Inter-
pretation
Ent-
scheidung
Evalu-
ation
Instanz-
daten

Feedback-Probleme

Personen, die wegen der Entscheidung des Sensors nicht fahren dürfen, können nicht nachweisen, dass sie sicher gefahren wären.

Methoden-
auswahl

Daten-
auswahl

Qualitäts-
maße

Statistisches
Modell

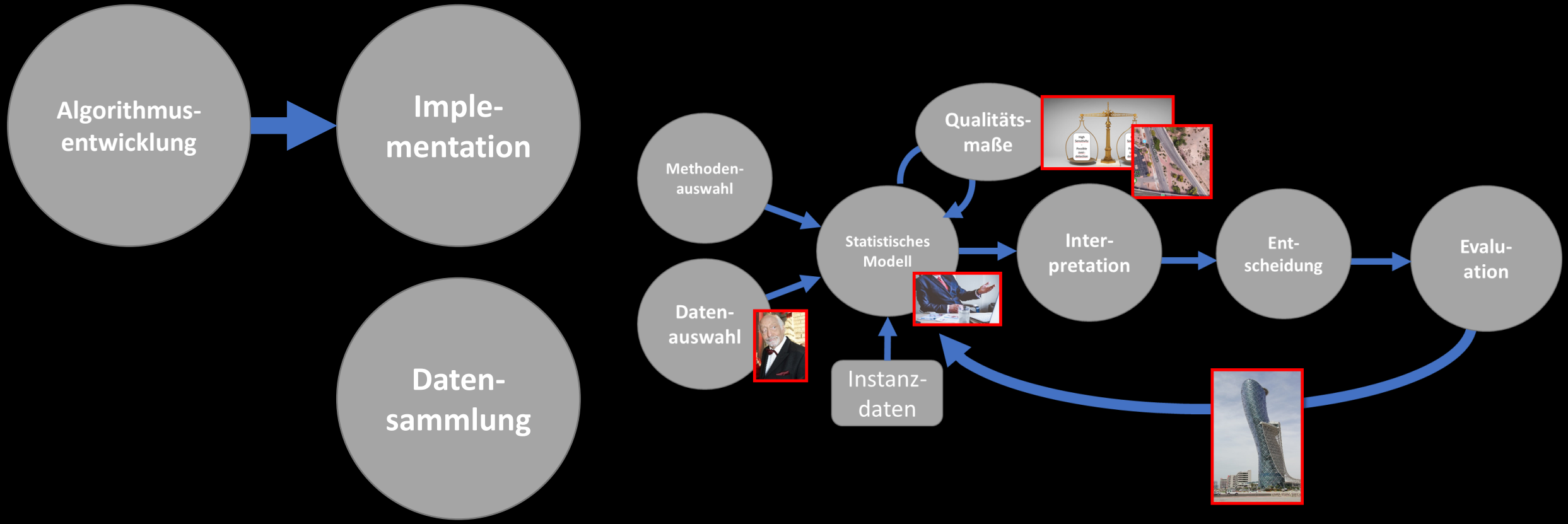
Instanz-
daten

Inter-
pretation

Ent-
scheidung

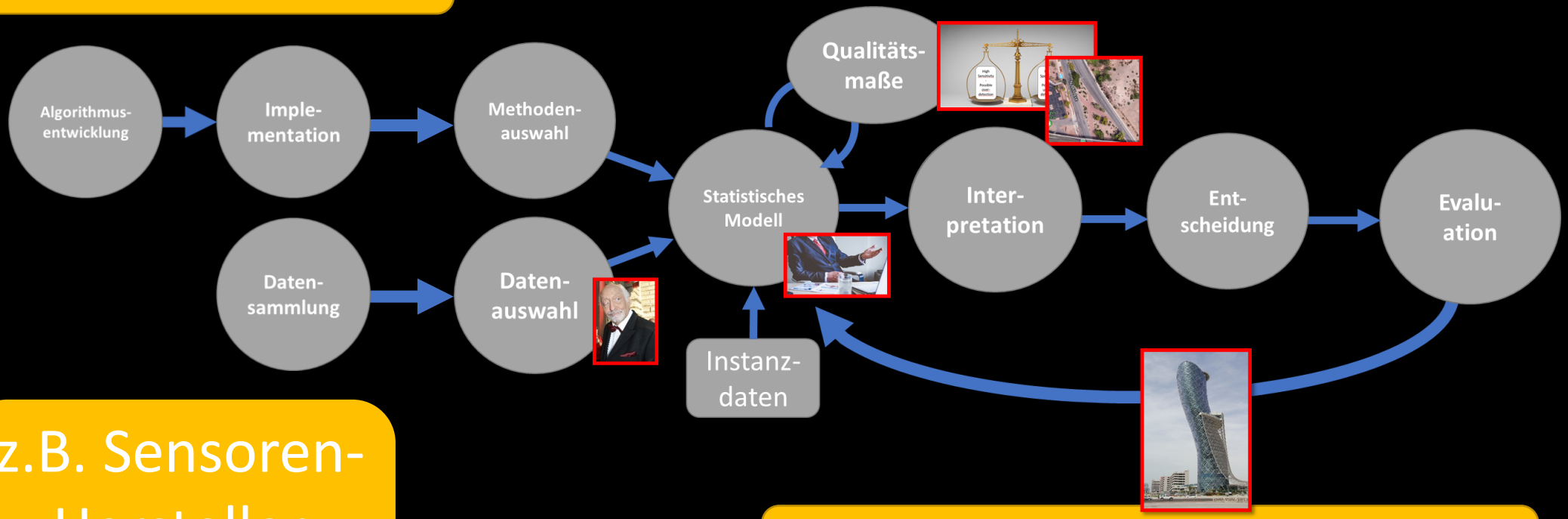
Evalu-
ation

Wer ist verantwortlich?



Wer ist verantwortlich?

Informatiker



z.B. Sensoren-Hersteller

Kunden Ihres ADM Systems

Data Scientists / Ingenieure

Kann ein fehlendes ADM System unethisch sein?

Und darüber hinaus?

Der Uber-Unfall in Tempe

Zwischen-bericht

“According to Uber, emergency braking maneuvers are not enabled while the vehicle is under computer control, to reduce the potential for erratic vehicle behavior.”

“The vehicle operator is relied on to intervene and take action.”

“The system is not designed to alert the operators.”

https://www.ntsb.gov/investigations/AccidentReports/Reports/HWY18MH010-prelim.pdf

https://www.google.de/maps/place/N+Mill+Ave,+Tempe,+AZ+85281,+USA/@33.4364084,-111.9436953,335m/data=!3m1!1e3!4m5!3m4!1s0x872b09338c84a6a3:0x4eb48ca97885c3f7!8m2!3d33.4379952!4d-111.9435544

Welche ethischen Entscheidungen traf Uber?

- Wurde die Wahrscheinlichkeit für Fußgängerübergänge von Straßenkarten und Satellitenbildern abgeleitet?
- Balance zwischen “Sicherheit (anderer)” und “Komfort der Passagiere”.
- Es fehlte ein Alarmsystem!
- Warum kein Alarmsystem, wenn der “operator” unaufmerksam ist?

Wie können wir bessere Systeme bauen?

Best Practice für die Entwicklung und Nutzung von ADM Systemen, die Menschen bewerten

- „Normale“ Algorithmen sind gut überprüfbar.
- Fokus auf **lernende Systeme,** deren Entscheidungen Menschen betreffen.
- Wichtig sind Entwicklungsteams **mit hoher Diversität**, sowohl was die Personen als auch Ausbildungen angeht.
- Weiterbildung im Bereich „Data Science Literacy“ und „Ethik für Data Scientists“.
- **Klare Kommunikationsprozesse** entlang der „langen Kette der Verantwortlichkeiten“:
  - **Z.B. Fehlerraten der Daten von Sensoren**.
  - Über alle Designentscheidungen, die getroffen wurden, und in welchem Kontext sie gültig sind.
- **Klare Abgrenzungen der jeweiligen Verantwortlichkeiten.**

„There are no simple solutions for complex problems. Whoever promises them, is fooling himself and others.“

Volkswagen at the IAA 2017