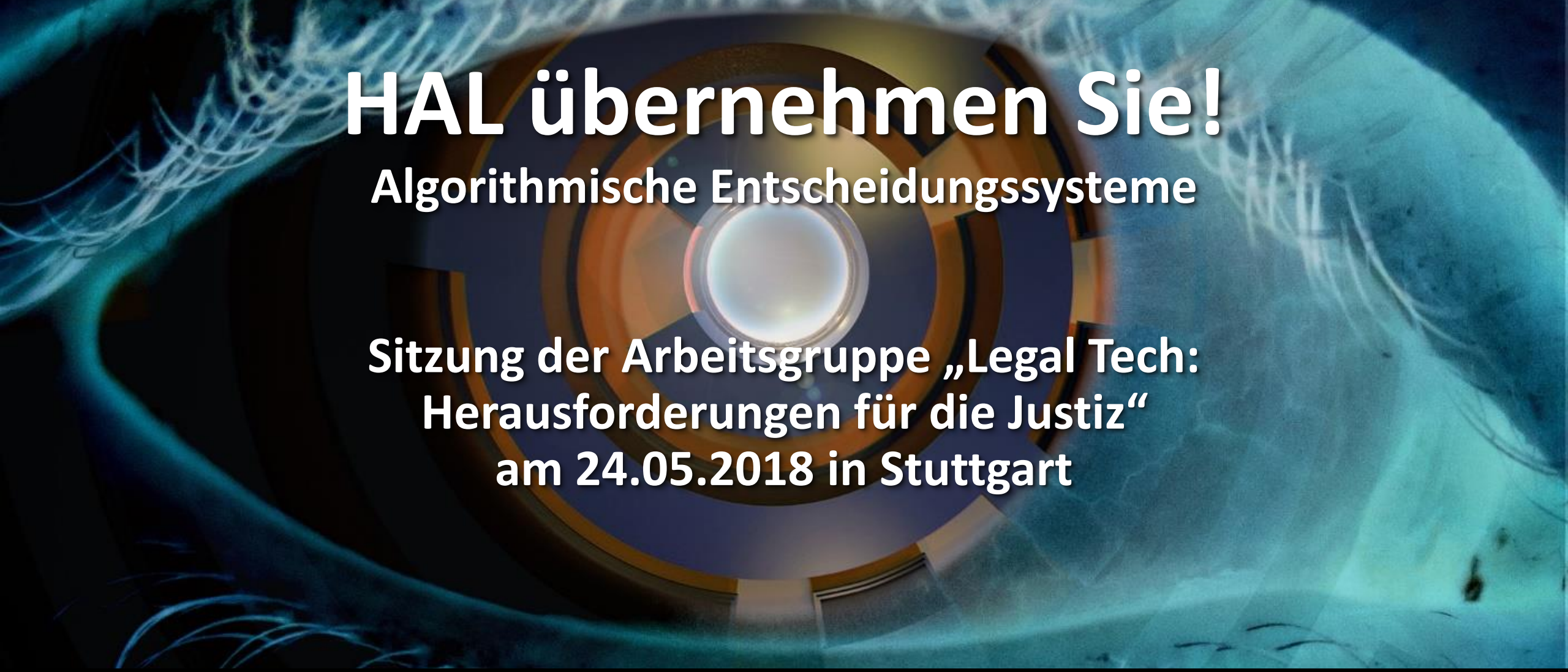


# HAL übernehmen Sie!

## Algorithmische Entscheidungssysteme

Sitzung der Arbeitsgruppe „Legal Tech: Herausforderungen für die Justiz“
am 24.05.2018 in Stuttgart


Prof. Dr. Katharina Anna Zweig

TU Kaiserslautern
Leiterin des Algorithm Accountability Labs
@nettwerkerin

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

…zwei verurteilte Kriminelle….


# Das sind Brisha und Vernon,….

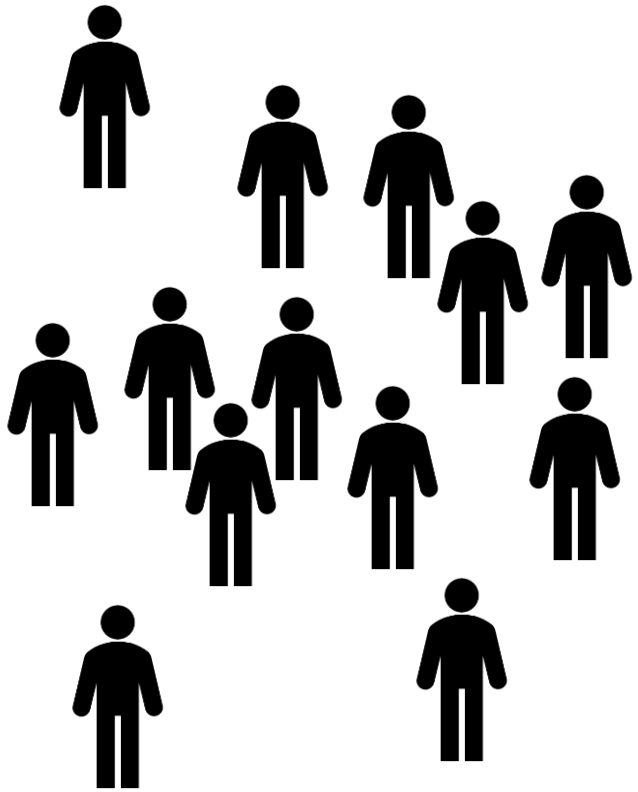

# Wer wird es wieder tun?

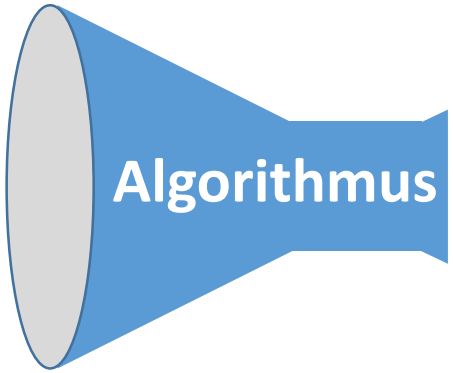
# Menschen – so irrational!

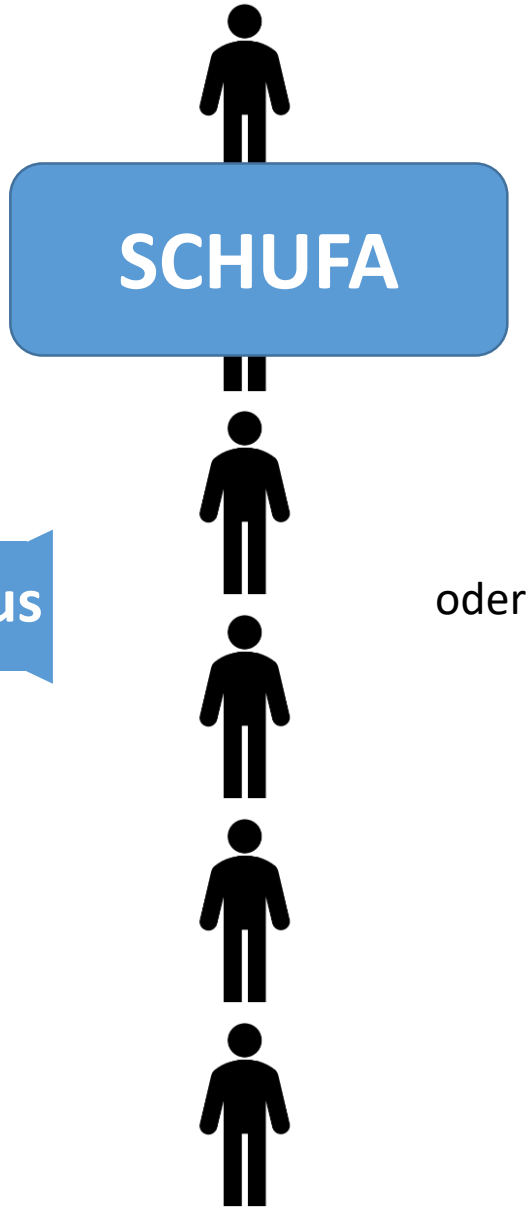
- Richter müssen vorzeitige Haftentlassungsanträge begutachten.
- Studie: je weiter von der letzten Pause weg, desto weniger risikoreiche Entscheidungen[1].
- Eine Vielzahl solcher Studien scheint zu beweisen:
  - Menschen sind irrational und vorurteilsbeladen.

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

# Algorithmische Entscheidungssysteme






Scoring-Verfahren

oder


Klassifikation

# Das kleine ABC der Informatik

Können

**A**lgorithmen,

**B**ig Data und

**C**omputerintelligenz

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

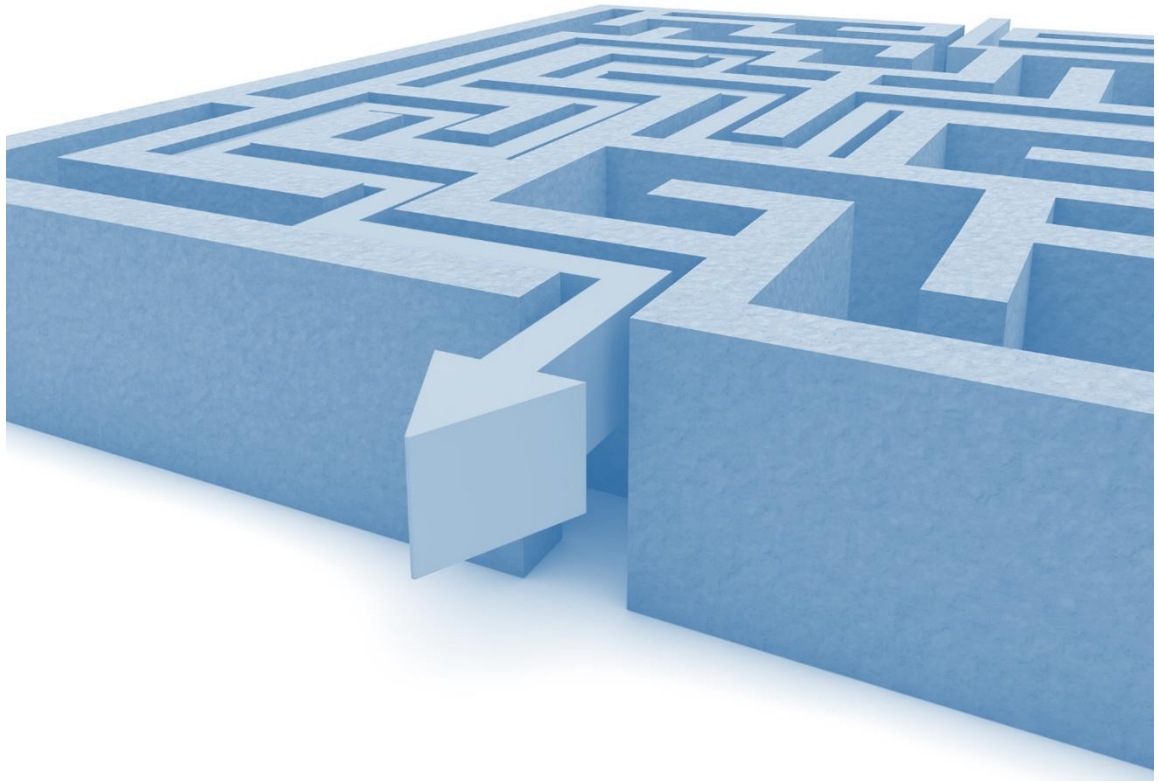
Das **Problem** sagt nicht, wie man die Lösung **findet**.


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

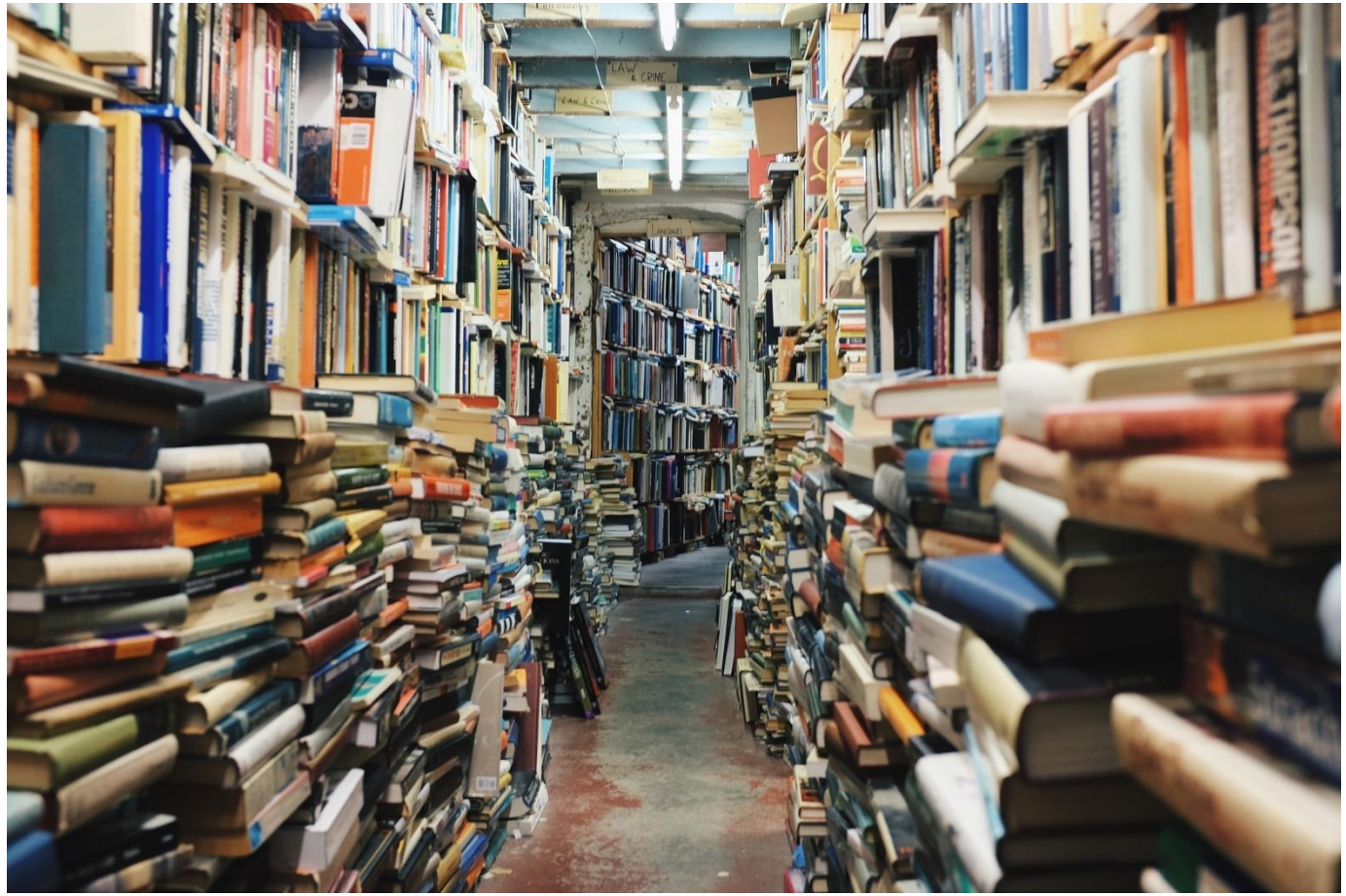
**Output: optimale Route**

# 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.



# Beispiel: Sortieren



# Sortieren 1: „Sortieren durch Einfügen“

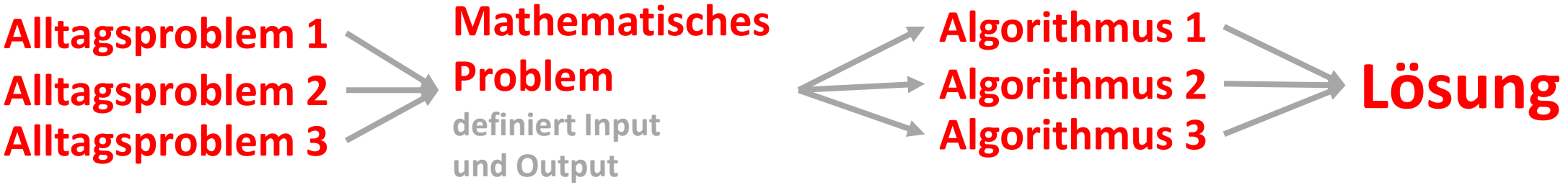
- Fange mit einem Buch an, stelle es ins Regal.
- Solange es noch Bücher gibt,
  - nimm das nächste,
  - geh am Regal entlang und sortiere es an der passenden Stelle ein.

- Alle Bücher, die schon im Regal stehen, sind in der richtigen, relativen Reihenfolge.
- Daher: wenn alle im Regal stehen, sind sie vollständig sortiert.

# Sortieren 2: Aufsteigendes Sortieren

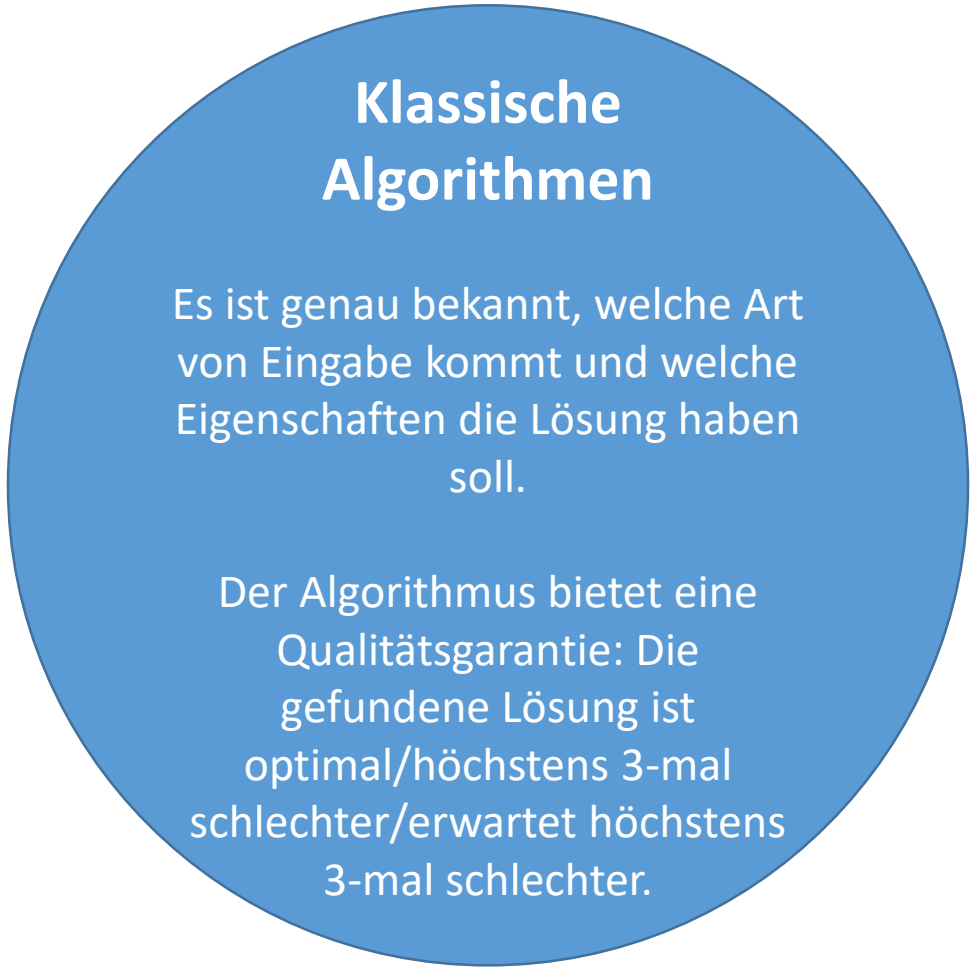
- Stelle alle Bücher irgendwie ins Regal.
- Gehe das Regal entlang – wenn dabei zwei Bücher in der falschen Reihenfolge nebeneinander stehen, vertausche sie. Tue dies bis zum Ende des Regals und gehe wieder zum Anfang.
- Laufe solange immer wieder am Regal entlang, bis im letzten Durchgang kein Tausch mehr nötig war.
- Wenn kein Tausch mehr nötig war, sind alle Bücher sortiert.

# Problem-Algorithmus-Lösung




- Ein mathematisches Problem kann also meist durch mehrere Algorithmen gelöst werden.
- Jeder Algorithmus löst nur genau ein mathematisches Problem.
- Im Sinne von „Alltagsproblemen“ löst derselbe Algorithmus sehr viele verschiedene Probleme:
  - Sortieren von Personen nach Anzahl ihrer Follower auf Twitter;
  - Anzeige von Nachrichten, sortiert nach Publikationsdatum;
  - Suchmaschineneinträge sortieren nach Bewertung durch Suchmaschinenalgorithmus;

# Algorithmen – eine Kategorisierung




- Sind oft mathematisch in ihrer Korrektheit bewiesen.
- Handwerkliche Fehler können passieren.
- Sie können auch explizit manipuliert werden und gesellschaftlich falsche / illegale Ziele verfolgen:
  - Beispiel Dieselskandal
- Für das korrekte Design, die korrekte Implementierung und die Auffindung von Fehlern/Manipulationen sind Informatikerinnen bestens ausgebildet.

# Und worüber reden dann gerade alle?

Maschinelles Lernen aus Big Data

# 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 individu-alisiertem Verhalten statistisch nutzbar.**
- **Hier werden Methoden des maschinellen Lernens benötigt.**

# C wie Computerintelligenz

# Was heißt Lernen?

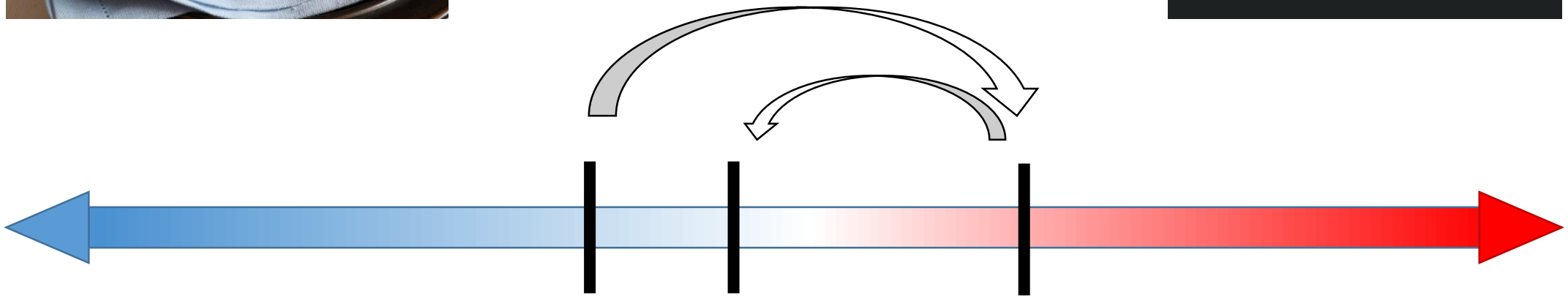
Einfach:

In derselben Situation ein vorher gezeigtes Verhalten wiederholen.

Generalisiert:

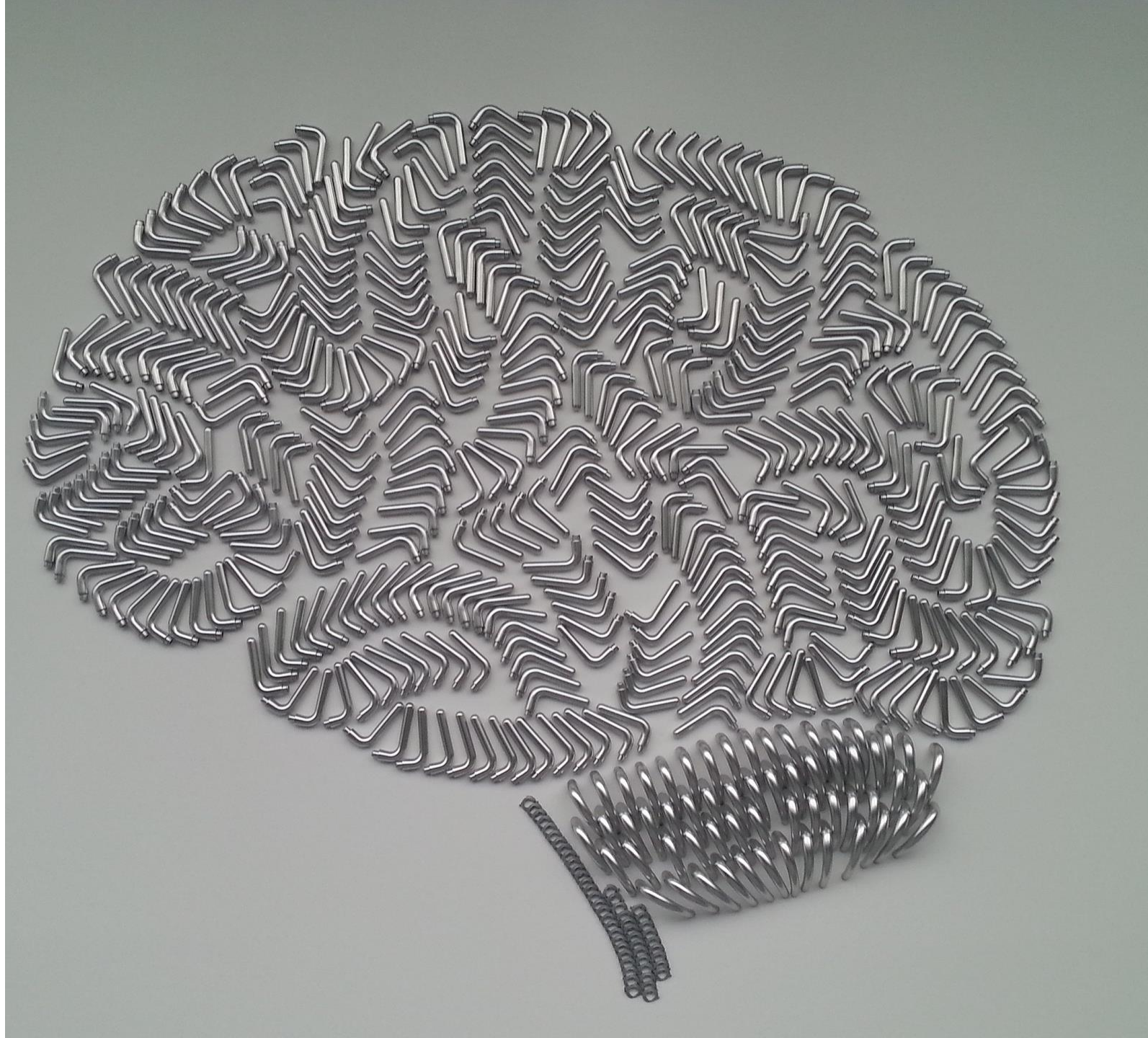
In derselben Art von Situation das richtige Verhalten aus einer Reihe von Möglichkeiten auswählen.

# Sebastian lernt „heiss“ und „warm“



# Sebastian lernt…

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

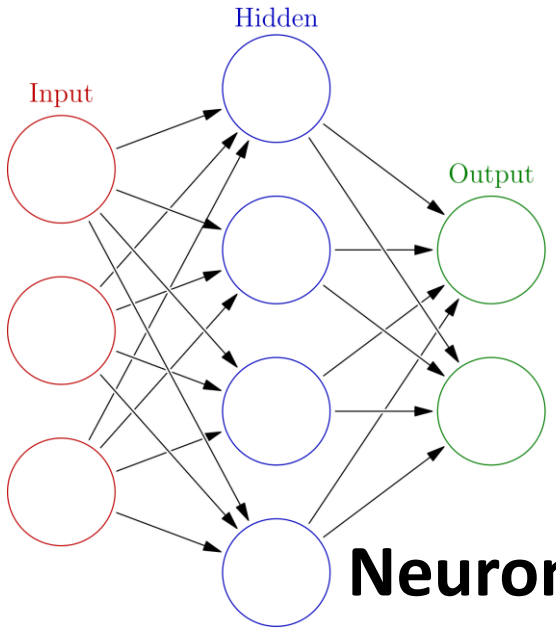


# Computer lernen

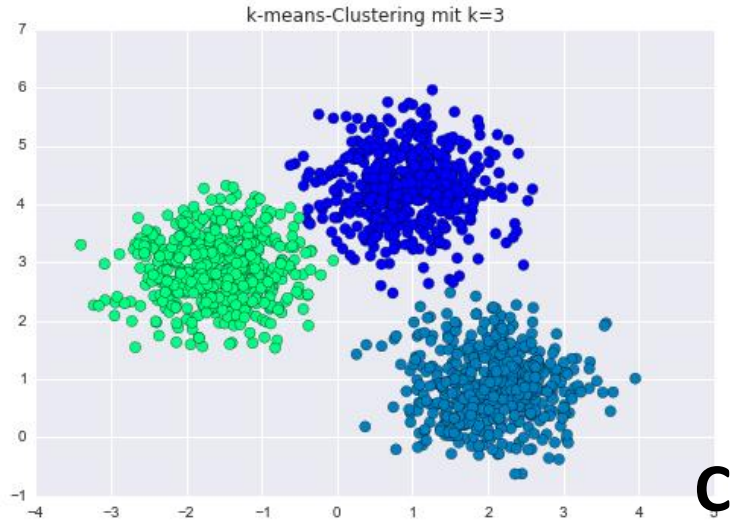
Damit ein Computer lernen kann, benötigt er ebenfalls eine **Struktur,** um Gelerntes abzuspeichern.

Optimal auch **Rückkopplung**.

Er lernt **generelle Regeln**.



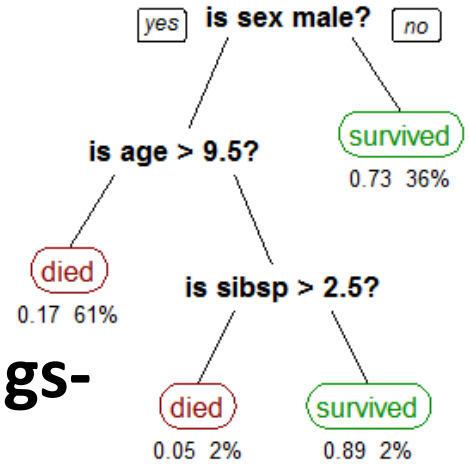

**Neuronales Netz**



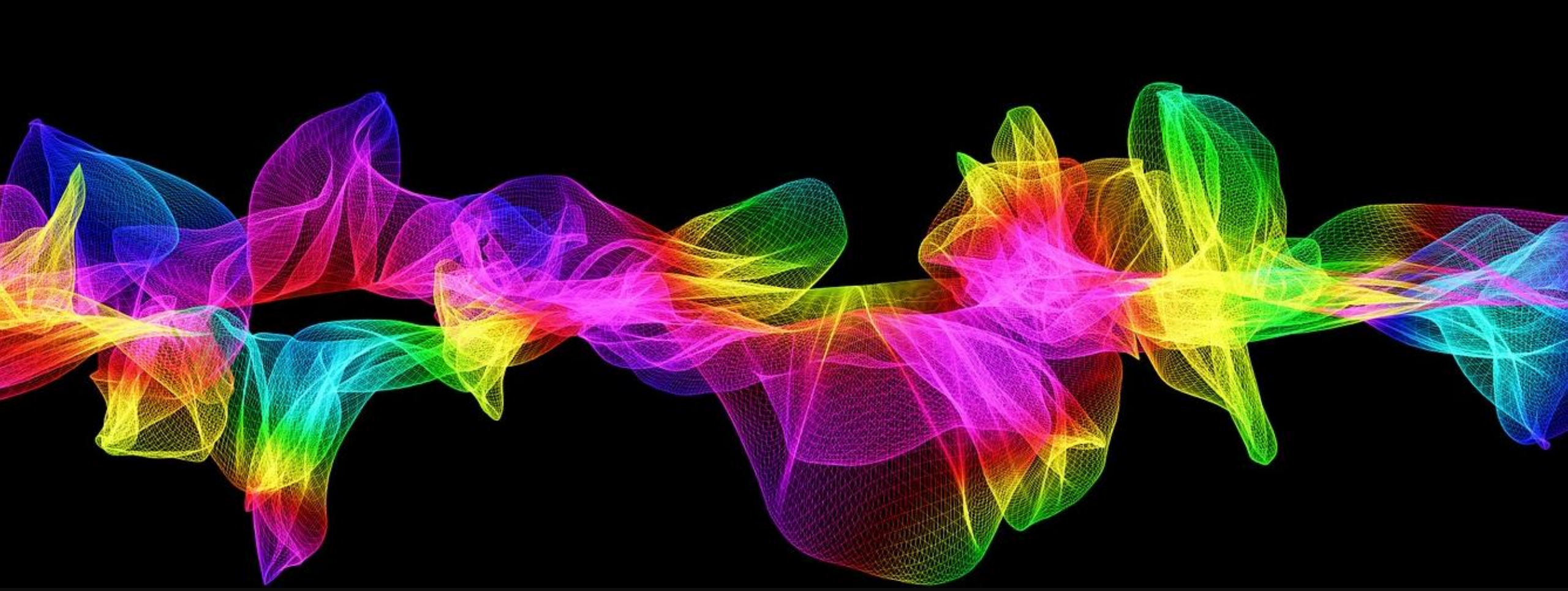

**Clustering**

$$w_1 * \#Vh - w_2 * \#dayiVh + w_3 * I[g = male] + B[\ * 1.0 + \cdots$$

**Formel**



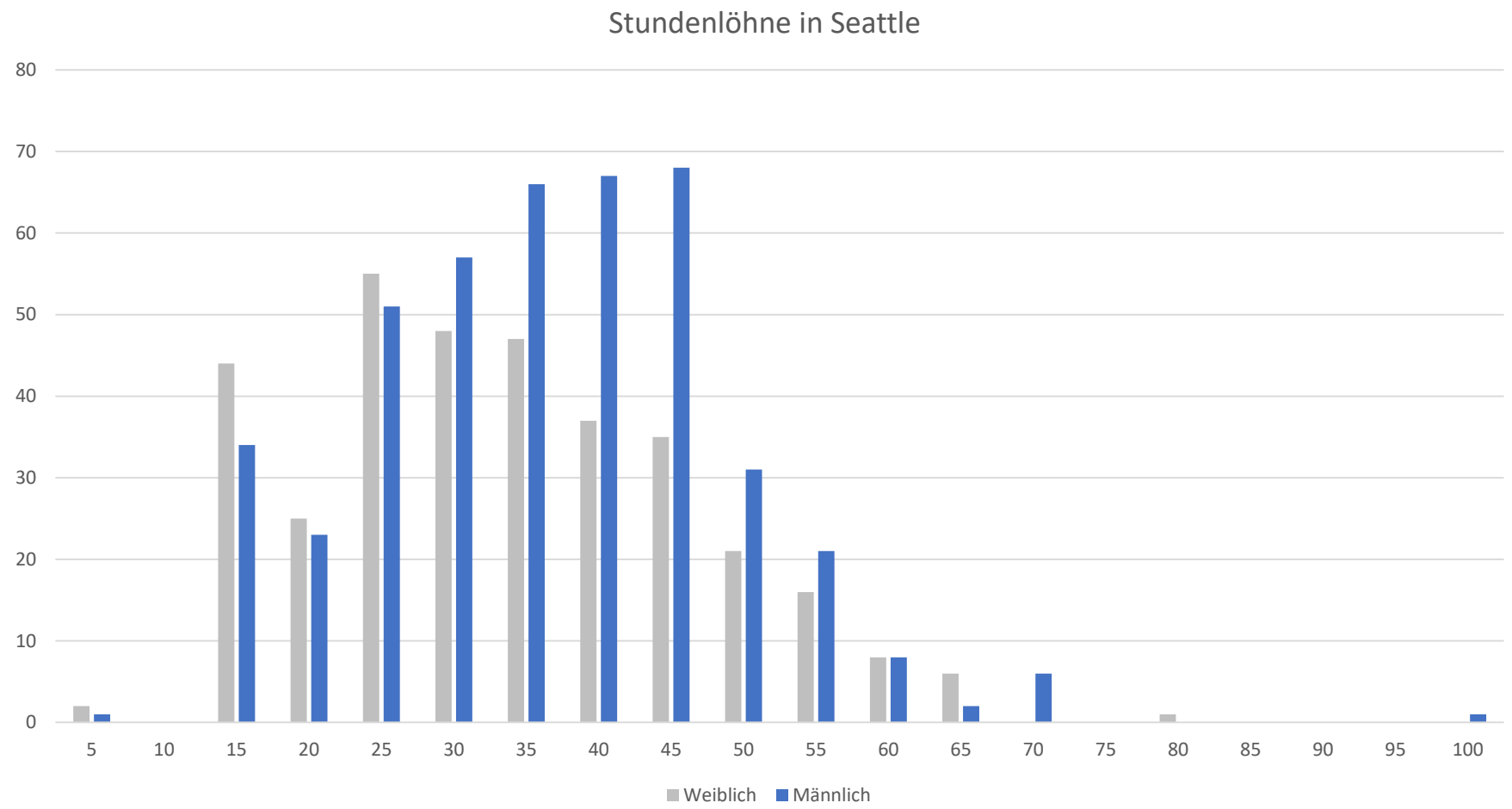

**Entscheidungs-bäume**



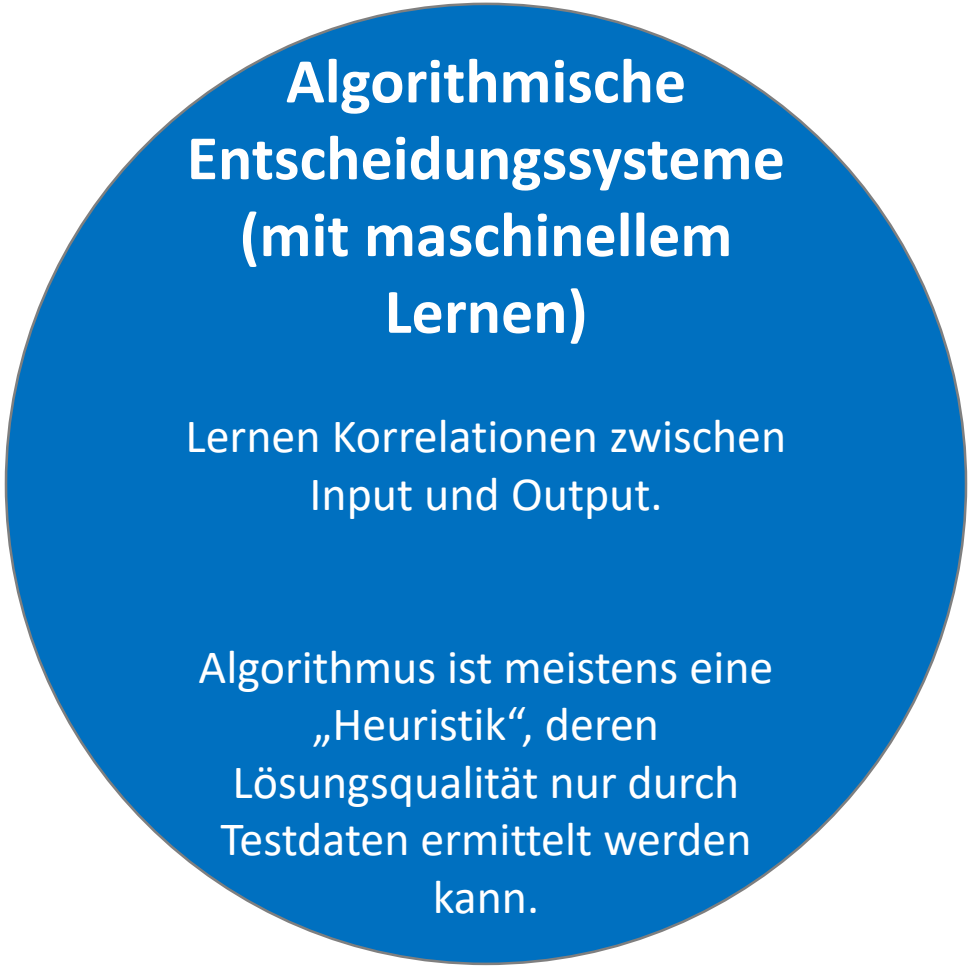
# “Lernen” mit Korrelationen

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

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



# Algorithmen – eine Kategorisierung

# Lernen mit Formeln

**Rückfälligkeitsvorhersage für (schon verurteilte) Kriminelle**

# Datengrundlagen

- Data Mining Methoden nutzen, z.B.:
  - Alter der ersten Verhaftung
  - Alter des Delinquenten (der Delinquentin!)
  - Finanzielle Lage
  - Kriminelle Verwandte
  - Geschlecht
  - Art und Anzahl der Vorstrafen
  - Zeitpunkt der letzten kriminellen Akte
  - Extra-Fragebogen
  - Aber bspw. nicht die (in den USA eindeutig zugec Zugehörigkeit.
- Wichtig: Beim Trainingsset ist bekannt, ob die Person rückfällig geworden ist oder nicht.

# Regressionsansatz

$$
\begin{aligned}
&3 * \text{ bisherige Verhaftungen} \\
&- 2 * \text{ Anzahl Tage seit letzter Verhaftung} \\
&+ 3 * (\text{Wenn Mann, dann 1, sonst 0}) \\
&+ 2,5 * (\text{ Wenn Raubüberfall, dann 1, sonst 0}) + ...
\end{aligned}
$$

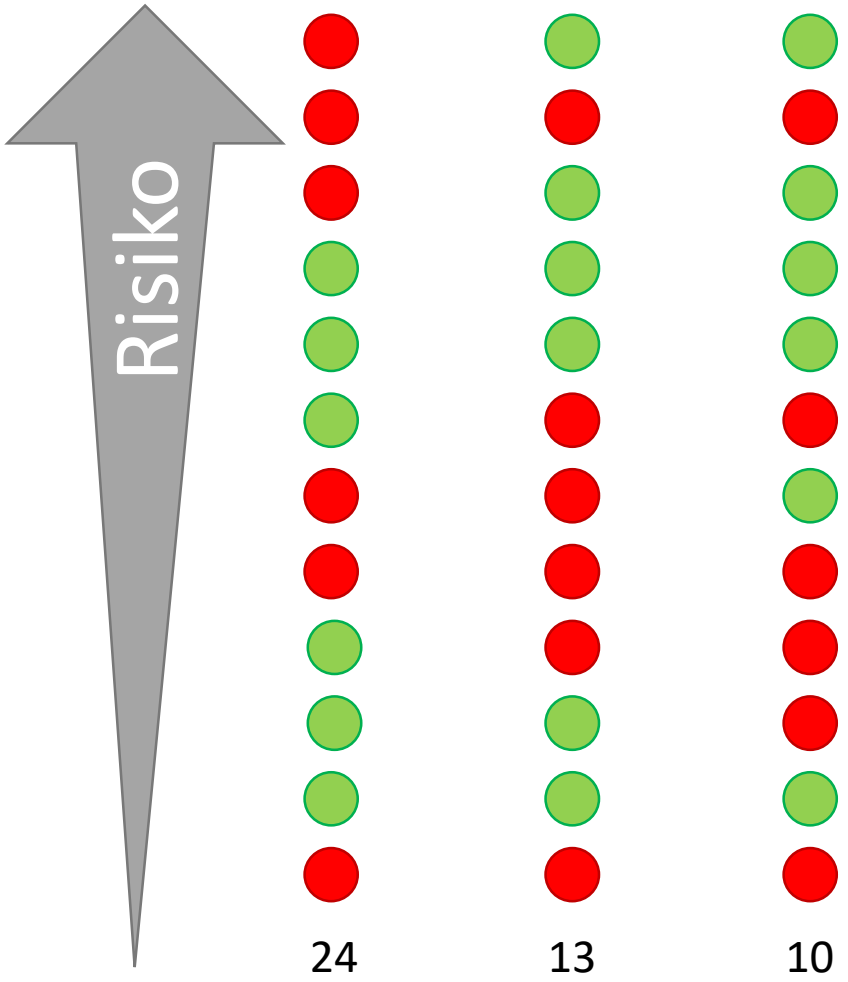
$$
\begin{aligned}
&w_1 * \text{ bisherige Verhaftungen} \\
&- w_2 * \text{ Anzahl Tage seit letzter Verhaftung} \\
&+ w_3 * (\text{Wenn Mann, dann 1, sonst 0}) \\
&+ w_4 * (\text{Wenn Raubüberfall, dann 1, sonst 0}) + ...
\end{aligned}
$$

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


# Qualität eines Algorithmus

# „Lernen“ von Gewichten

- Algorithmus probiert Gewichte und berechnet Risiko für alle Personen im Datenset.
- Bewertet jeweils, wie viele erwiesenermaßen Rückfällige möglichst weit oben stehen.
- Die Gewichtung, die das maximiert, wird für weitere Daten genommen.




Grüne Kugeln symbolisieren resozialisierte, rote rückfällige Kriminelle.

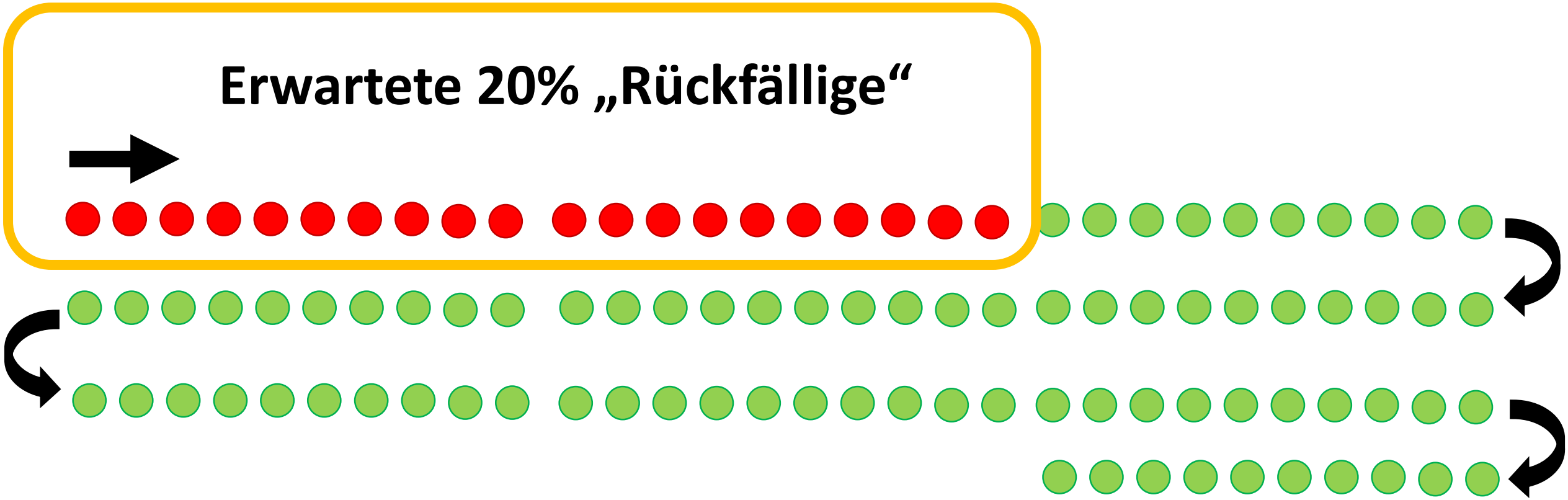
Optimale Sortierung: Alle roten oben, alle grünen darunter.

Qualitätsmaß: Paare von rot und grün, bei denen die rote Kugel über der grünen einsortiert ist.

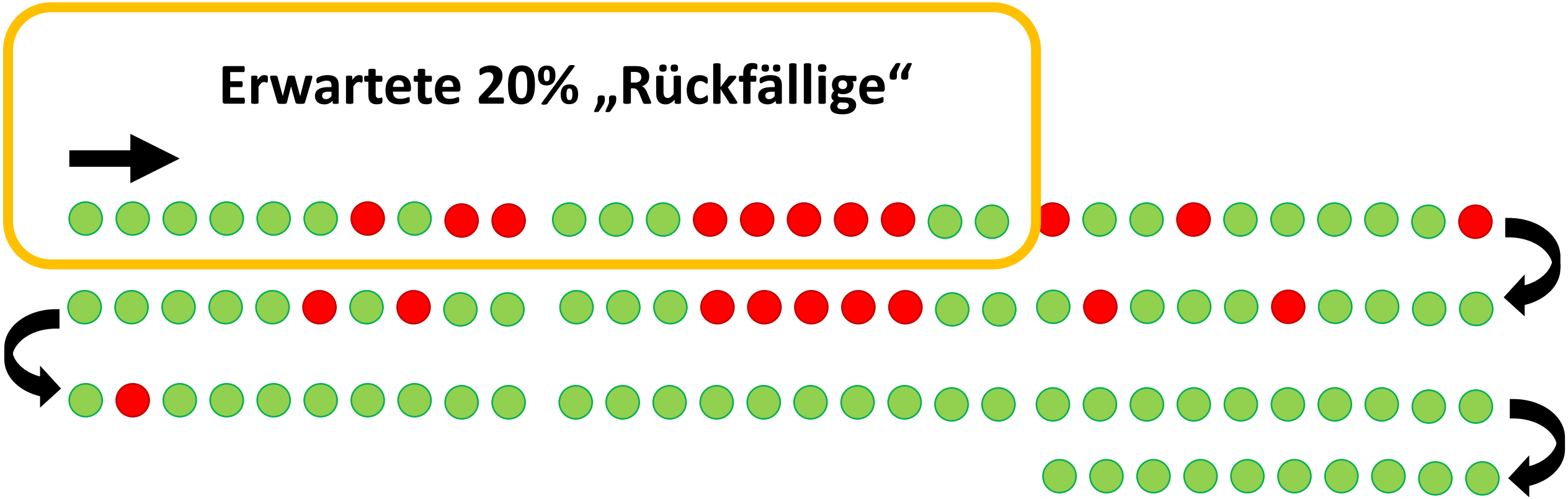
# Oregon Recidivism Rate Algorithm

- 72 von 100 Paaren werden korrekt sortiert.
- So werden aber keine Urteile gefällt!
- Sondern: Reihe von Angeklagten, von denen diejenigen mit dem höchsten Rückfallrisiko benannt werden sollen.
- Rückfallquote bei jugendlichen Kriminellen liegt z.B. bei 20%.

# Optimale Sortierung



# Mögliche Sortierung eines Algorithmus mit dieser „Güte“ (75/100 Paaren)



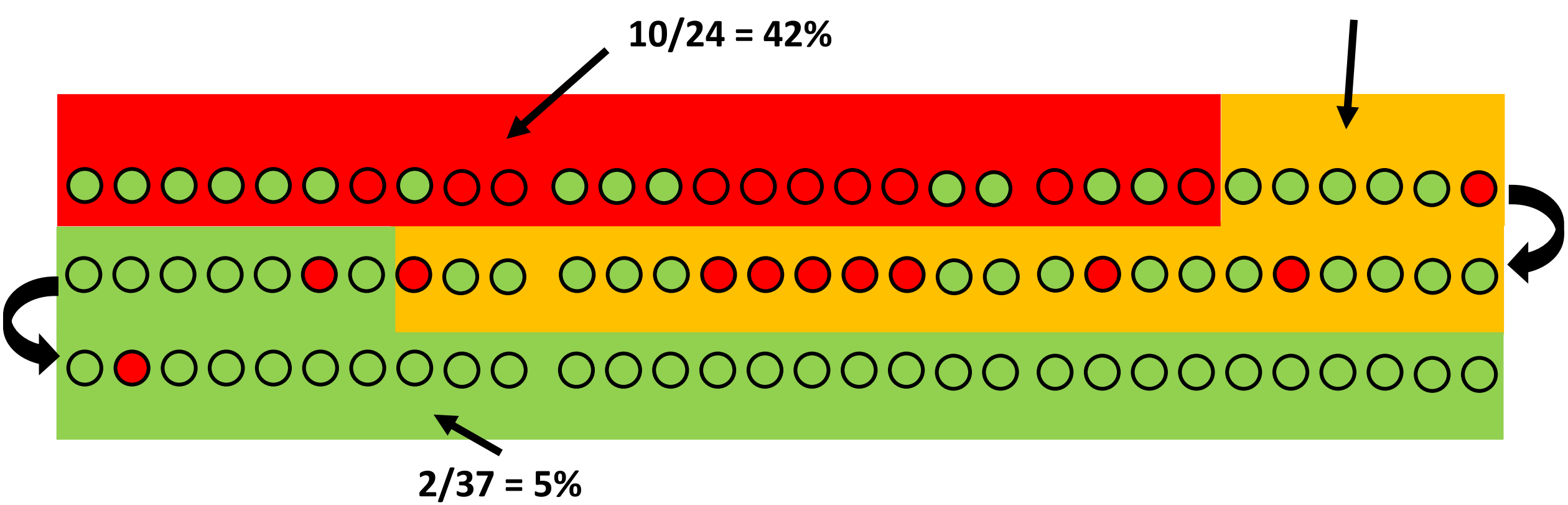
einen Jagdhund zu kaufen,

um  Schafe zu hüten.

Das ist wie...

# Vom Scoring zur Klassifikation

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

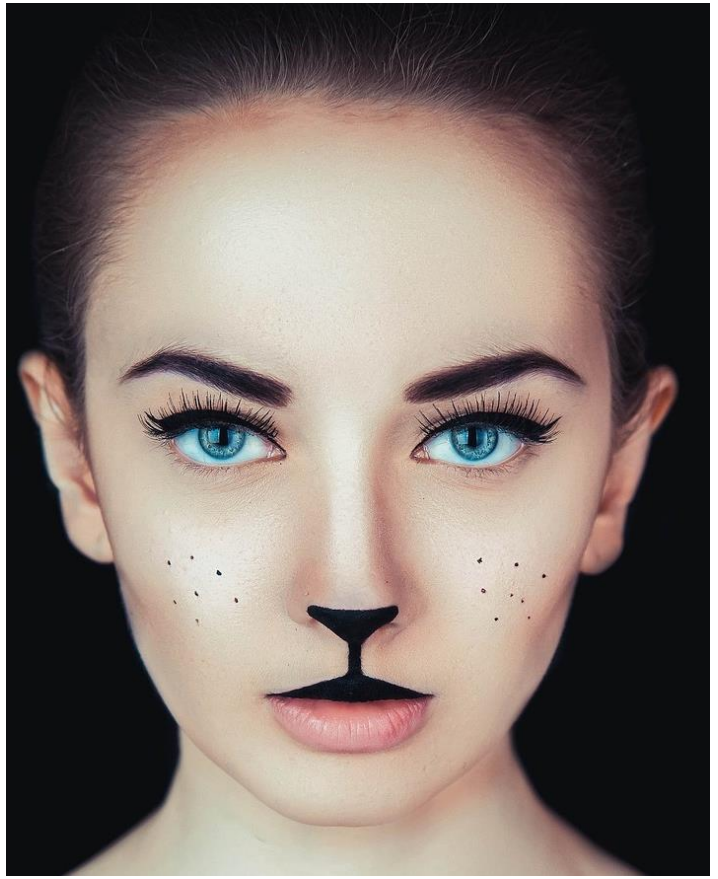



# Statistische Vorhersagen über Menschen

# Statistische Prognosen beim Wetter

# Zu 40% ein Krimineller….

- Wenn dieser Mensch eine Katze wäre und 7 Leben hätte, würde er in 3 davon wieder rückfällig werden…
- Nein!
- **Algorithmische Sippenhaftung**
  - Von 100 Personen, die „genau so sind wie dieser Mensch“, werden 40 wieder rückfällig;
  - Wir folgen einem *algorithmisch legitimierten Vorurteil*.



https://pixabay.com/de/m%C3%A4dchen-katze-k%C3%A4tzchen-mode-1048728/

# Können Algorithmen diskriminieren?

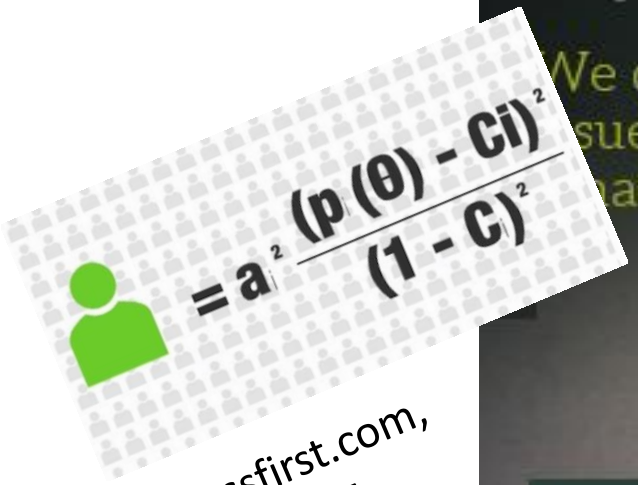
# Diskriminierung bei Bewerbungen

- Lebensläufe mit „deutschen“ Namen bekommen 14% mehr Vorstellungsangebote 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

# „Employment asse[illegible] 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

Screenshot einer Anzeige auf der Webseite von dreamhr.co.uk
16.11.2017

# Eine rosige Zukunft

„Below are a few of our solutions, but in the end, with the availability of good data, the **predictive possibilities are virtually unlimited** (emphasis by me): candidate demographics & background data, psychometric data, structured interview data, assessment data, performance data, on boarding evaluation data, training data, etc.”

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

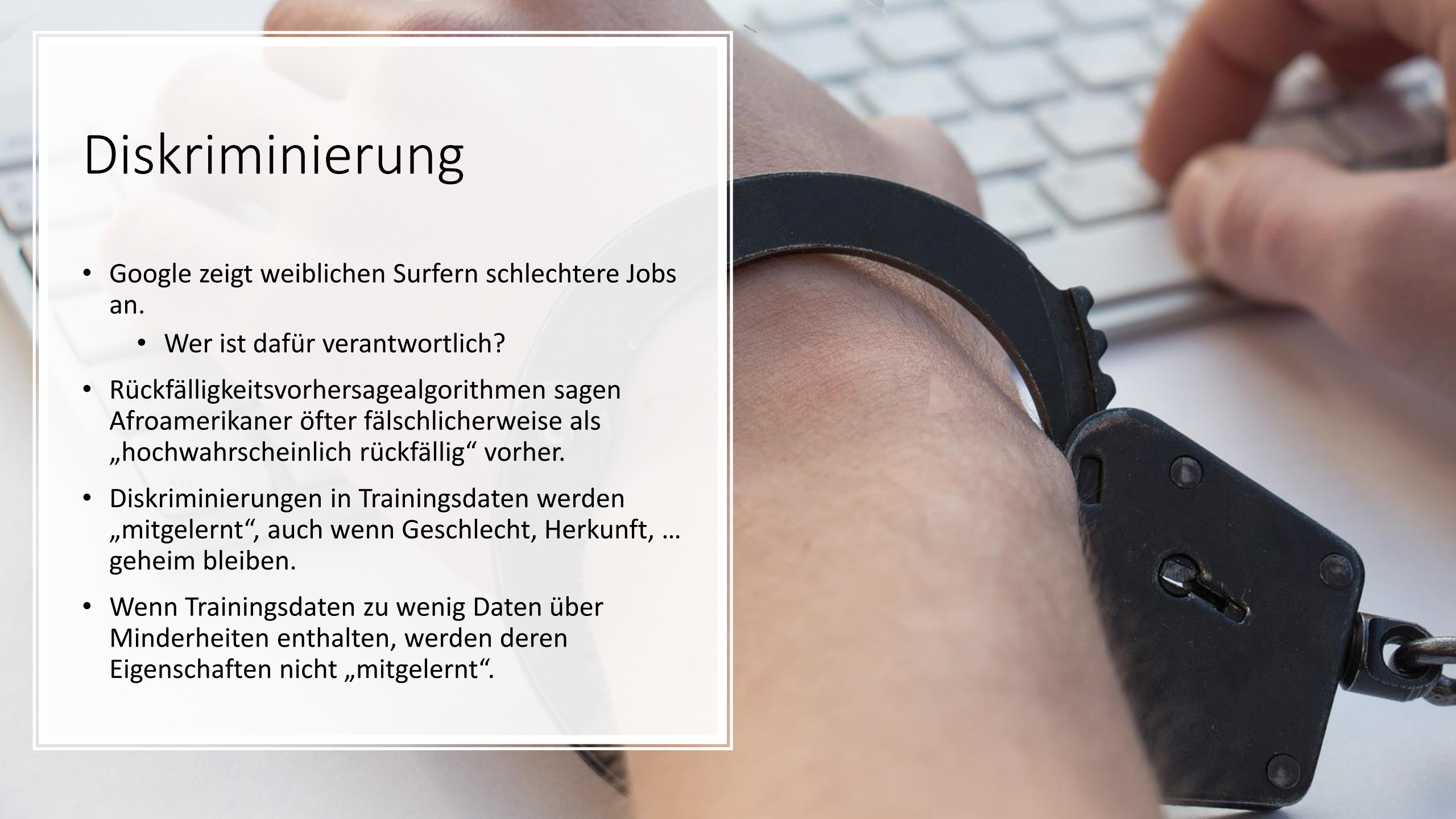
# Magie der Algorithmen?


Wenn man auf Google nach „CEO“ sucht…

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

# Diskriminierung

- Google zeigt weiblichen Surfern schlechtere Jobs an.
  - Wer ist dafür verantwortlich?
- Rückfälligkeitsvorhersagealgorithmen sagen Afroamerikaner öfter fälschlicherweise als „hochwahrscheinlich rückfällig“ vorher.
- Diskriminierungen in Trainingsdaten werden „mitgelernt“, auch wenn Geschlecht, Herkunft, … geheim bleiben.
- Wenn Trainingsdaten zu wenig Daten über Minderheiten enthalten, werden deren Eigenschaften nicht „mitgelernt“.



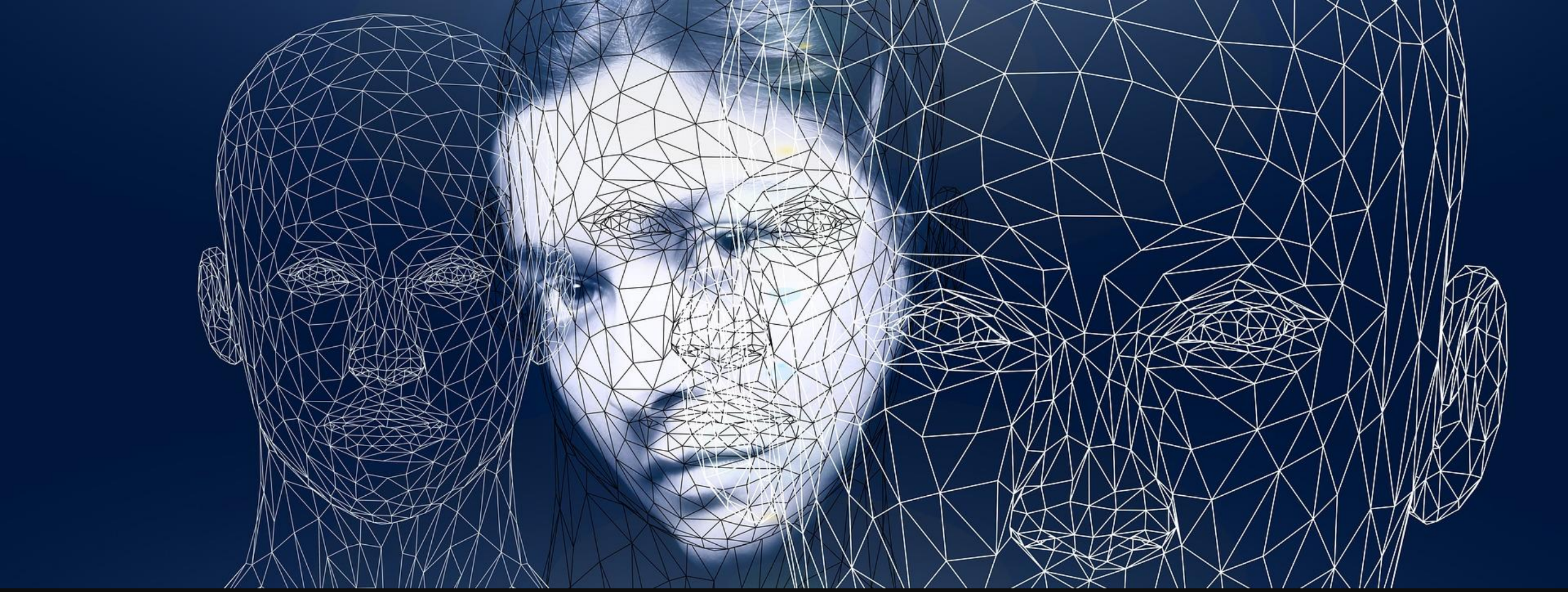
# Regel

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

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

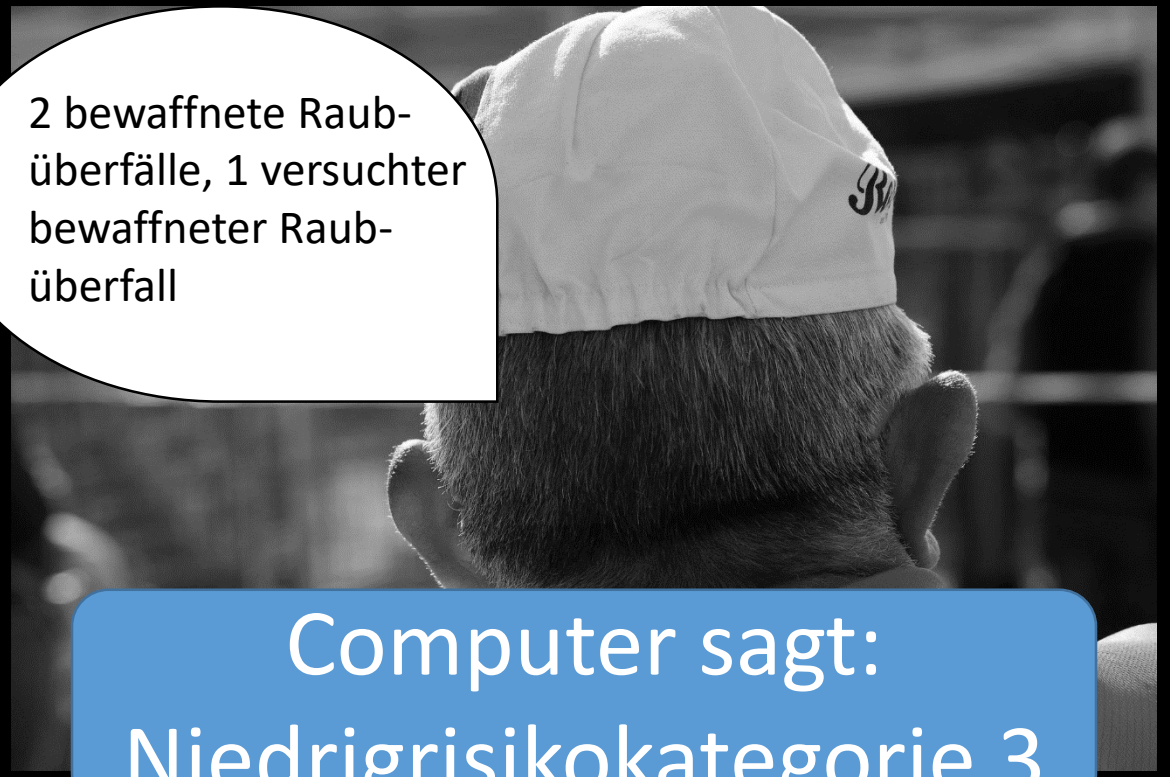
Die Muster sind daher grundsätzlich **statistischer Natur.**

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



# Sozio-informatische Gesamtbetrachtung


Computer sagt: Hochrisikokategorie 8




Computer sagt: Niedrigrisikokategorie 3

# Wer wird es wieder tun?

Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner: Machine Bias, ProPublica, 23.5.2016
https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

# Probleme der Einbettung der ADM in den sozialen Prozess

- **Aufmerksamkeitsökonomie** von Entscheiderinnen und Entscheidern.
- **„Best practice“** erfordert Nutzung der Software.
- Eine Nichtbeachtung der Empfehlung  und gleichzeitige Fehleinschätzung wirkt oft schwerer als eine Beachtung der (falschen) Empfehlung. **Delegierung von Verantwortung!**
- Manchmal kann ein(e) falsch-negativ Beurteilte(r) **die Vorhersage prinzipiell nicht entkräften!**
  - Z.B. abgelehnte Bewerberin, eingesperrte Kriminelle


# Wer hat es wieder getan?

Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner: Machine Bias, ProPublica, 23.5.2016
https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

# Algorithmen in einer demokratischen Gesellschaft

# Generell

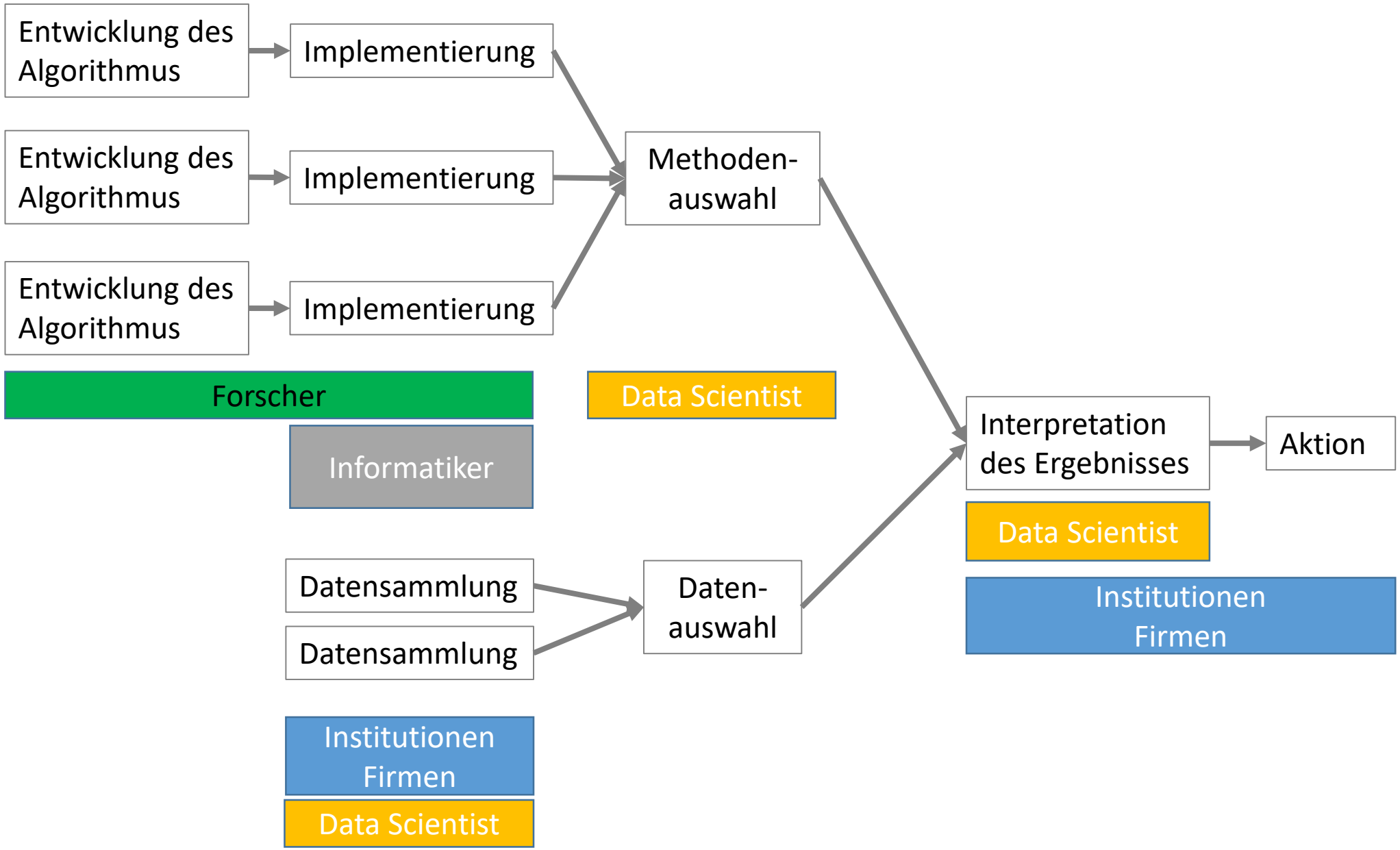
Prinzipiell können ADM Systeme vieles entscheiden:

- Automatische Leistungsbewertung
- Kreditvergabe
- Schulische und universitäre Ausbildungen, die durch algorithmische Entscheidungssysteme unterstützt werden
- Algorithmen, die das Sterberisiko von Kranken bewerten
- Gefährder-, Terroristenidentifikation
- …

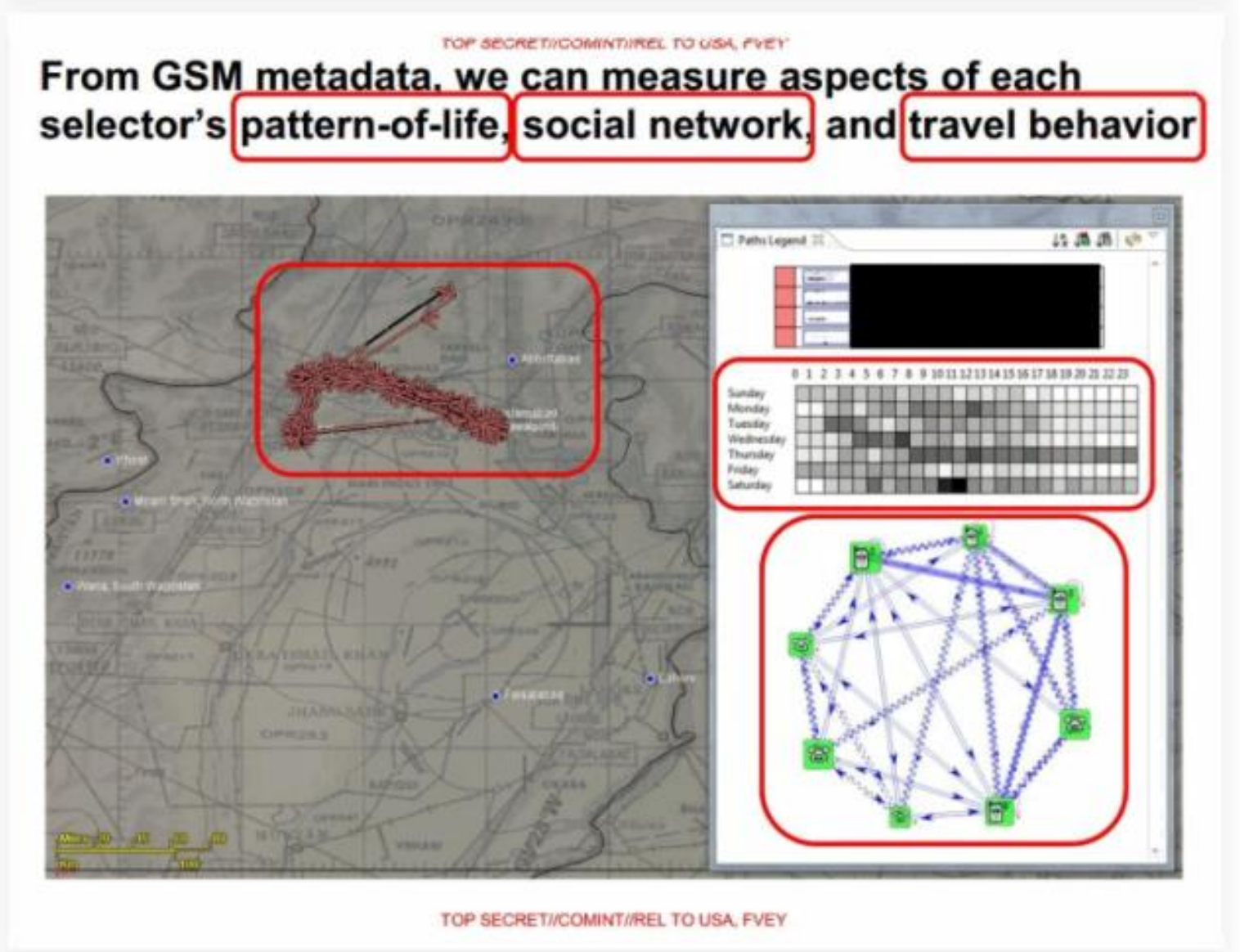
Ihre Aufgabe heute….

Entwickeln Sie ein
algorithmisches Entscheidungssystem,
dass **gewaltbereite Extremisten**
frühzeitig identifiziert!

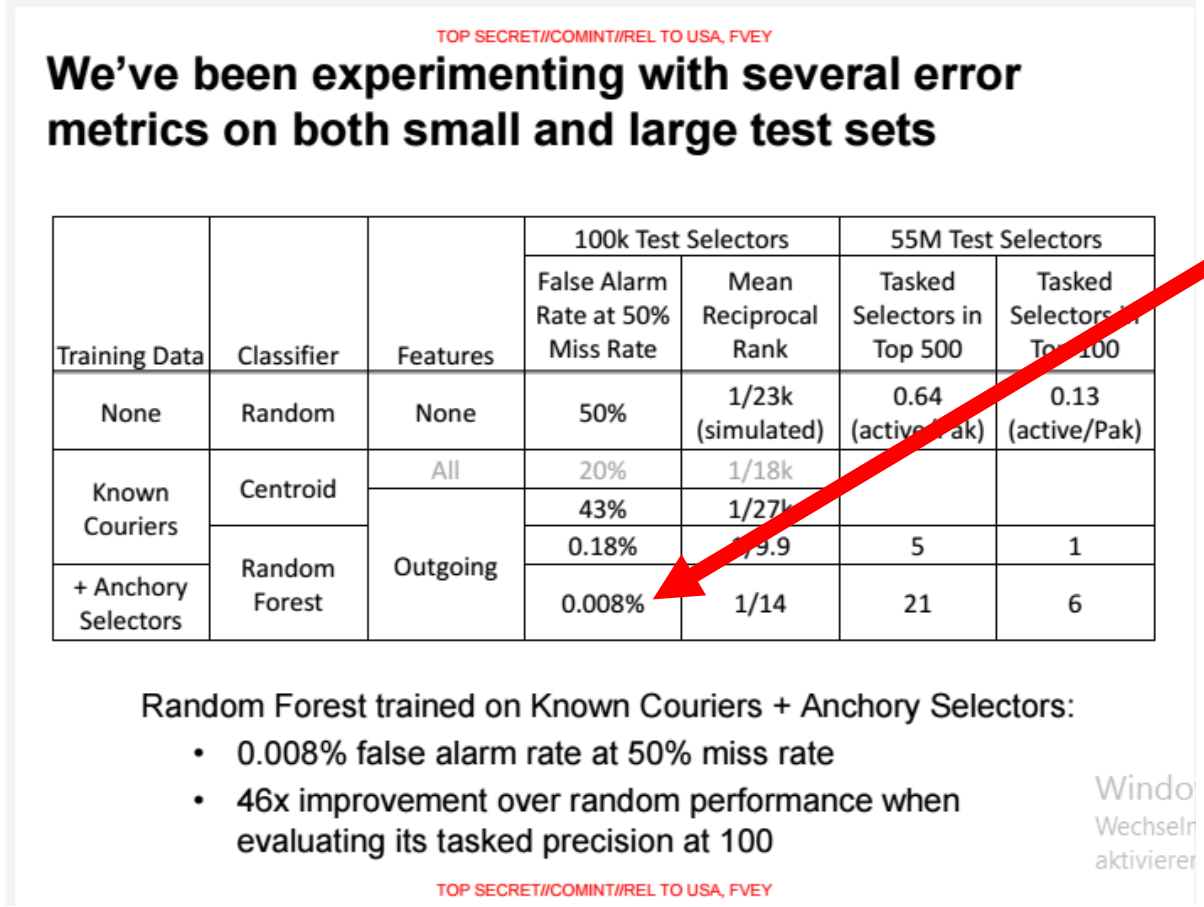
# Designprozess



# Capturing terrorists with network analysis



# Terroristenidentifikation SKYNET



TOP SECRET//COMINT//REL TO USA, FVEY

**We've been experimenting with several error metrics on both small and large test sets**

| Training Data | Classifier | Features | 100k Test Selectors | | 55M Test Selectors | |
|---|---|---|---|---|---|---|
| | | | False Alarm Rate at 50% Miss Rate | Mean Reciprocal Rank | Tasked Selectors in Top 500 | Tasked Selectors in Top 100 |
| None | Random | None | 50% | 1/23k (simulated) | 0.64 (active/Pak) | 0.13 (active/Pak) |
| Known Couriers | Centroid | All | 20% | 1/18k | | |
| | | Outgoing | 43% | 1/27k | | |
| | Random Forest | | 0.18% | 1/9.9 | 5 | 1 |
| + Anchory Selectors | | | 0.008% | 1/14 | 21 | 6 |

Random Forest trained on Known Couriers + Anchory Selectors:

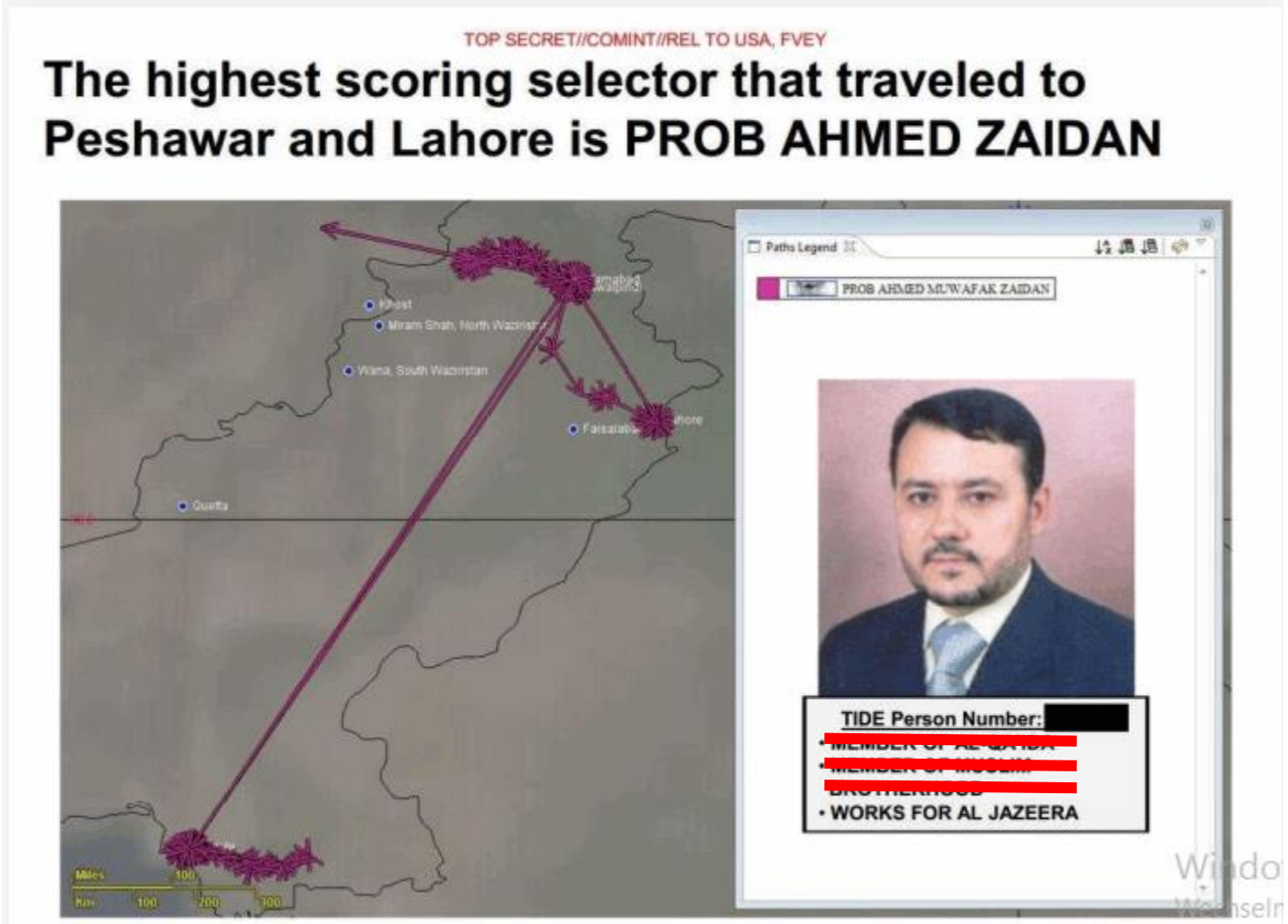
- 0.008% false alarm rate at 50% miss rate
- 46x improvement over random performance when evaluating its tasked precision at 100

TOP SECRET//COMINT//REL TO USA, FVEY

Das sind 4.400 Unschuldige, um die Hälfte der vermeintlichen Terroristen zu identifizieren!

https://theintercept.com/document/2015/05/08/skynet-courier/
https://theintercept.com/2015/05/08/u-s-government-designated-prominent-al-jazeera-journalist-al-qaeda-member-put-watch-list/

# Top-“Kurier“ der Terroristen laut Algorithmus ist…



Soziale Konzepte können hinreichend operationalisiert werden, so dass der Computer damit arbeiten kann.

Ein Mensch wird durch seine ihm zugehörigen Daten hinreichend beschrieben, um ihn zu bewerten oder zu kategorisieren.

# Zusammenfassung Menschenbild

Menschen können durch eine einzige Zahl hinreichend beschrieben werden (für jeweils eine spezifische Fragestellung).

Menschen sind irrational und vorurteilsbeladen.

Maschinen sind die besseren Entscheider.

Sie sollen Entscheidungen aus Daten ableiten.

Wenn Menschen entscheiden, müssen sie ge“nudged“werden.

Ein Mensch sollte in seinem Verhalten durch das Verhalten einer Gruppe von ihm ähnlichen Personen bewertet werden.

# Probleme von algorithmischen Entscheidungssystemen (ADM Systemen) im People und Risk 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

# Einschätzung

- Algorithmen **könnten** dabei helfen, bessere Entscheidungen zu treffen.
  - Sie sind zuverlässig.
  - Können Entscheidungs-wege transparenter machen.
  - Könnten Diskriminierung vermeiden.
- Allerdings sind sie heute oft noch nicht gut genug.

# Was muss reguliert werden?

- Nicht Algorithmen per se!
- Algorithmische Systeme, die Entscheidungen über Menschen treffen.
- Nicht ohne sozio-informatische Gesamtsicht!
- Daher: Qualitätssicherung der Entwicklung und Einbettung!

# Screenshots on first slide taken from:

1. https://www.creamhr.co.uk
2. https://www.inostix.com/predict-hiring-success/
3. http://www.assessfirst.com/predictive-recruitment-discovers-the-best-employees-through-algorithms/