Wie die Gesellschaft von künstlicher Intelligenz profitieren kann

---

und wie nicht!

Brüssel, 24.4.2018

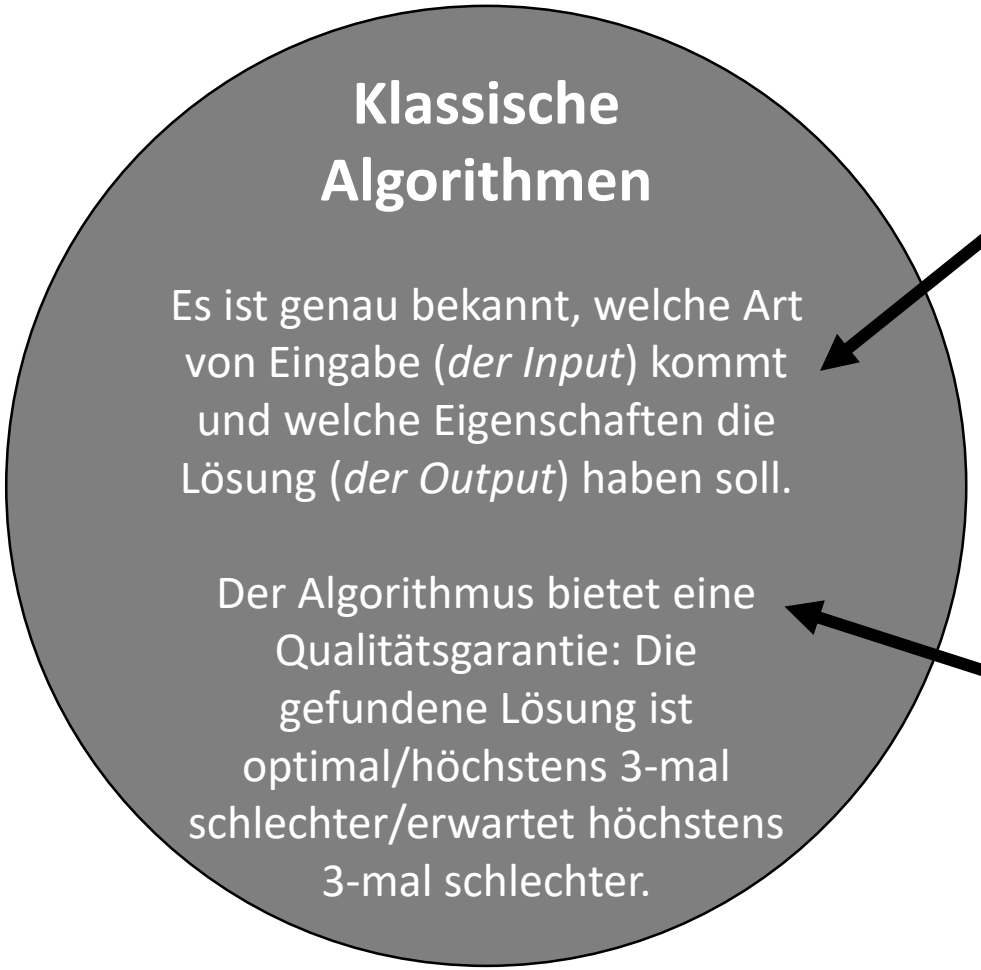
Europa-Abgeordnete in Diskussion mit

Prof. Dr. Katharina A. Zweig

@nettwerkerin
Algorithm Accountability Lab
TU Kaiserslautern

Was sind überhaupt „Algorithmen“?

Algorithmen – eine Kategorisierung



Wir nennen dies das **„mathematische Problem“**

Der Algorithmus **löst das Problem.**

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

Output: optimale Route

Ein Algorithmus ist…

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

Eine **Implementierung** ist die Übersetzung der für den Menschen verständlichen Handlungsanweisung in eine Programmiersprache.

2. Beispiel: Zinsrechnung

Sie haben 10.000 Euro und legen die für 0.5% an.

Wieviel Geld haben Sie nach 5 Jahren, wenn Sie immer reinvestieren?

Handlungsanweisung

Rechne:

10.000 €

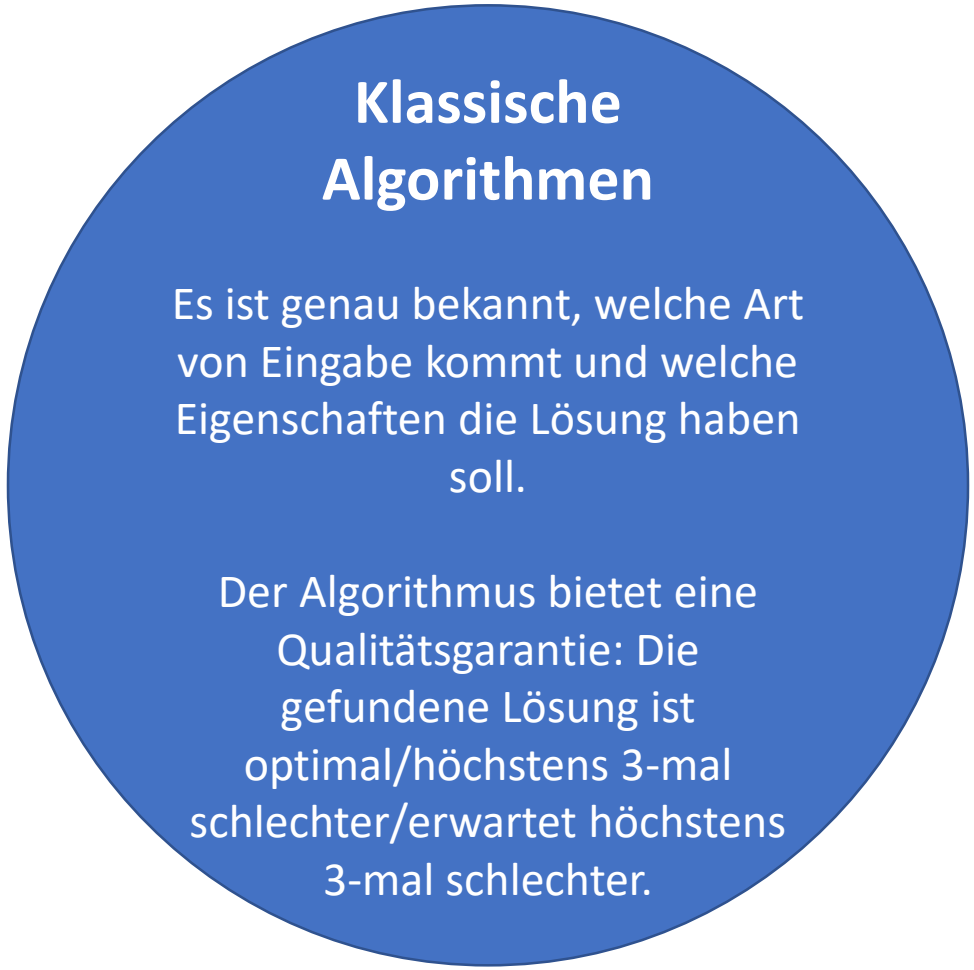
* 1.05 * 1.05

* 1.05 * 1.05

* 1.05

= 12.763 €

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?

Künstliche Intelligenz
und
maschinelles Lernen

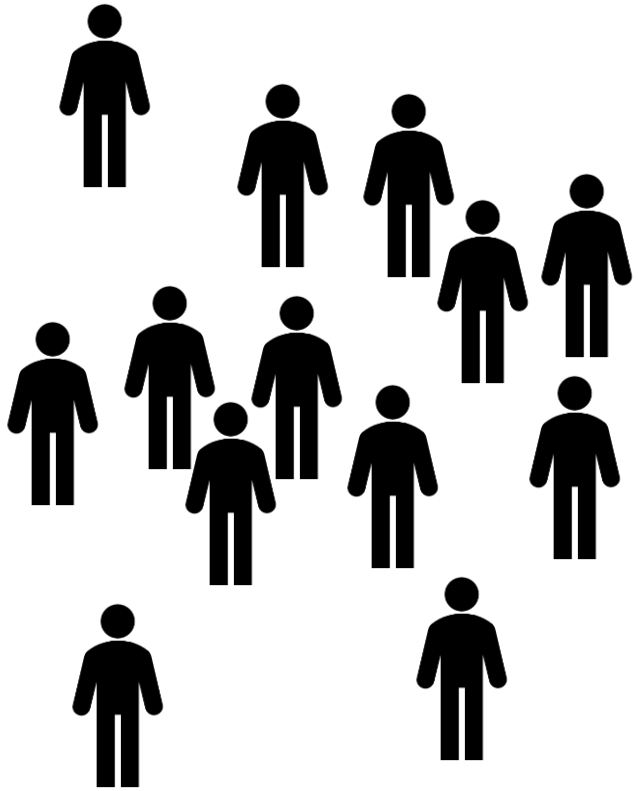
Beyond Verbal > Health news > **Your voice will guide your chores, healthcare and driving**

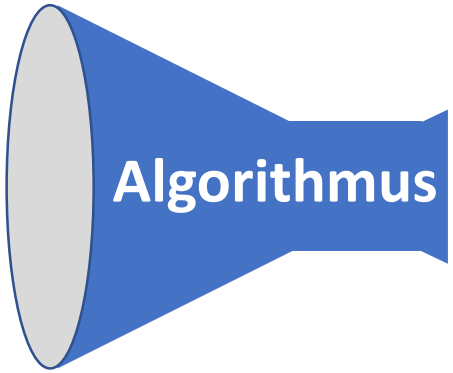
YOUR VOICE WILL GUIDE YOUR CHORES, HEALTHCARE AND DRIVING

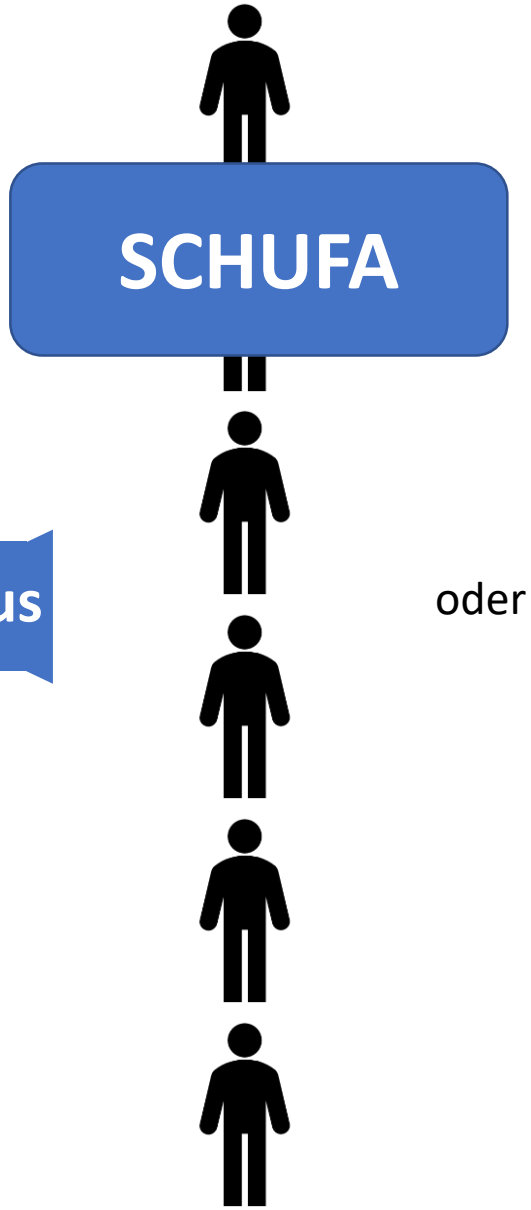
Posted on

IN 5 YEARS, VOICE TECH WILL HELP DOCTORS DIAGNOSE AND OPERATE, CARMAKERS PROVIDE CUSTOMIZED WEB CONTENT, HR PROFESSIONALS JUDGE JOB APPLICANTS AND MORE.

Algorithmische Entscheidungssysteme





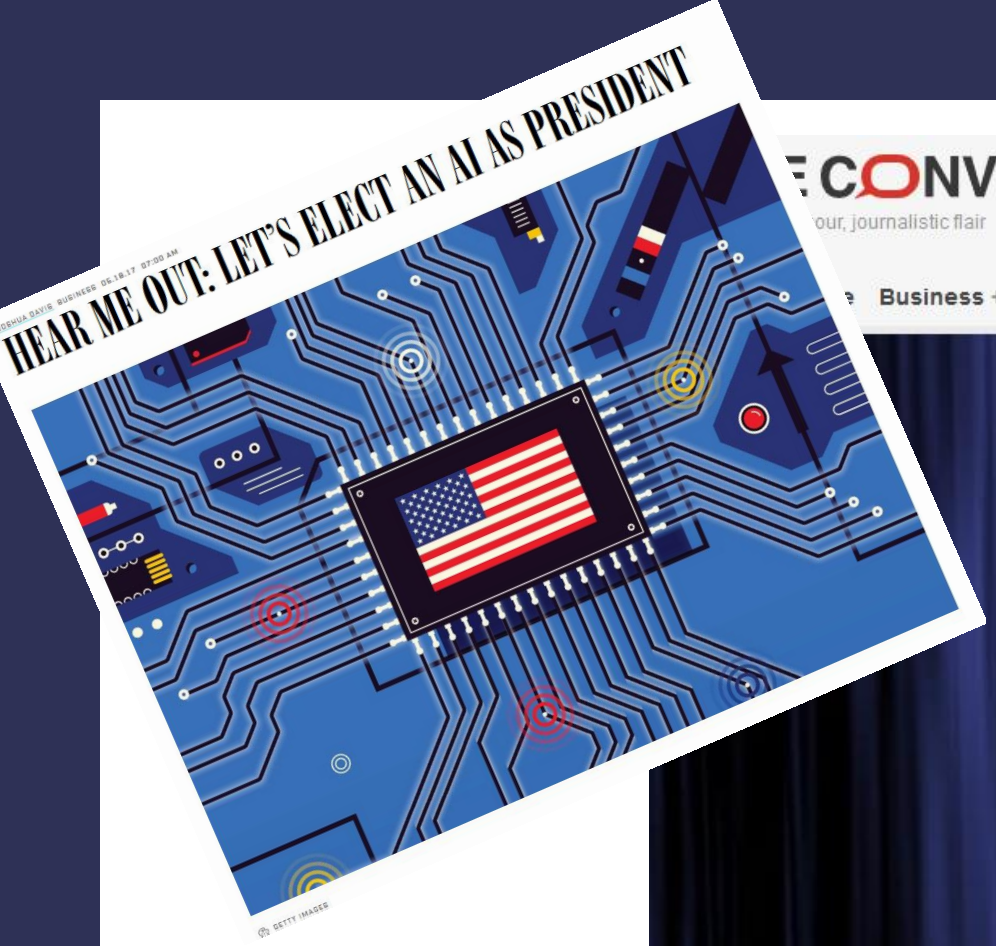


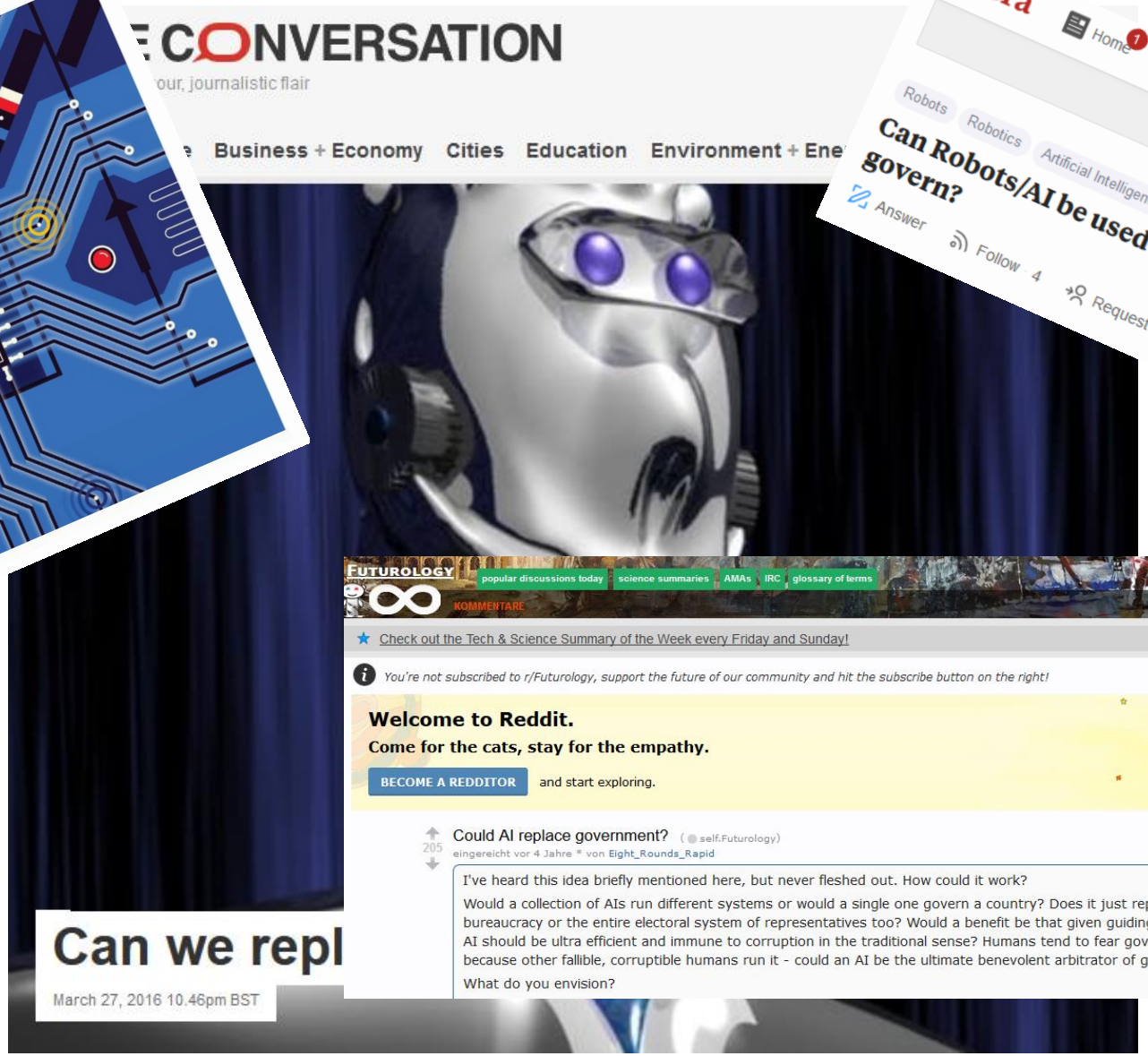
Scoring-Verfahren

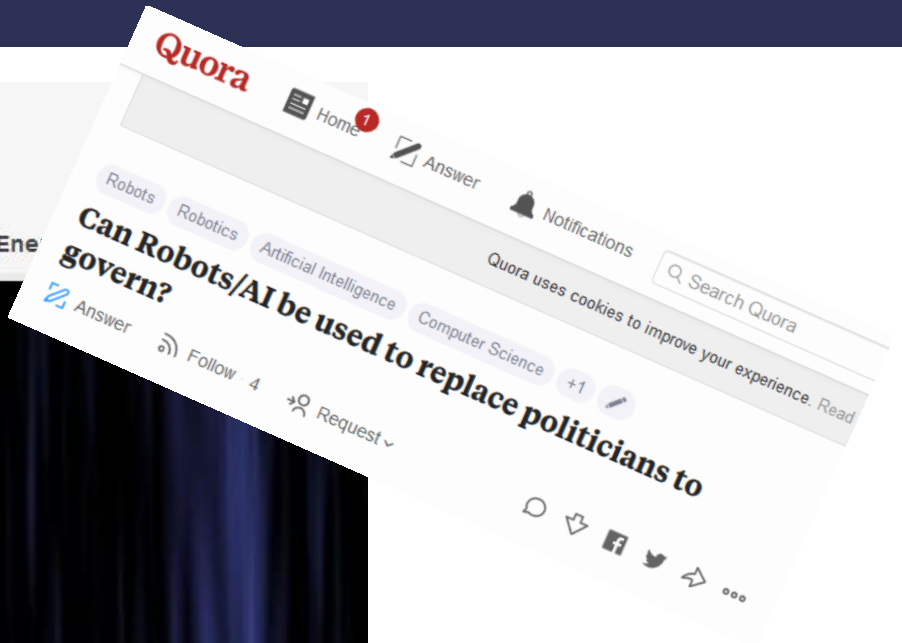
oder

Klassifikation

Wer soll politische Entscheidungen fällen?


HEAR ME OUT: LET'S ELECT AN AI AS PRESIDENT
GETTY IMAGES


E CONVERSATION
our, journalistic flair
Business + Economy
Cities
Education
Environment + Ene
Can we repl
March 27, 2016 10.46pm BST


Quora
Home
Answer
Notifications
Search Quora
Quora uses cookies to improve your experience. Read
Robots
Robotics
Artificial Intelligence
Computer Science
+1
Can Robots/AI be used to replace politicians to govern?
Answer
Follow 4
Request

FUTUROLOGY
popular discussions today
science summaries
AMAs
IRC
glossary of terms
KOMMENTARE
Check out the Tech & Science Summary of the Week every Friday and Sunday!
You're not subscribed to r/Futurology, support the future of our community and hit the subscribe button on the right!
Welcome to Reddit.
Come for the cats, stay for the empathy.
BECOME A REDDITOR
and start exploring.
Could AI replace government?
self.Futurology
eingereicht vor 4 Jahre von Eight_Rounds_Rapid
I've heard this idea briefly mentioned here, but never fleshed out. How could it work?
Would a collection of AIs run different systems or would a single one govern a country? Does it just replace the bureaucracy or the entire electoral system of representatives too? Would a benefit be that given guiding principles, an AI should be ultra efficient and immune to corruption in the traditional sense? Humans tend to fear government because other fallible, corruptible humans run it - could an AI be the ultimate benevolent arbitrator of governance?
What do you envision?

Unser Menschenbild

Sind Menschen eigentlich dazu geeignet, über andere Menschen zu entscheiden?

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.

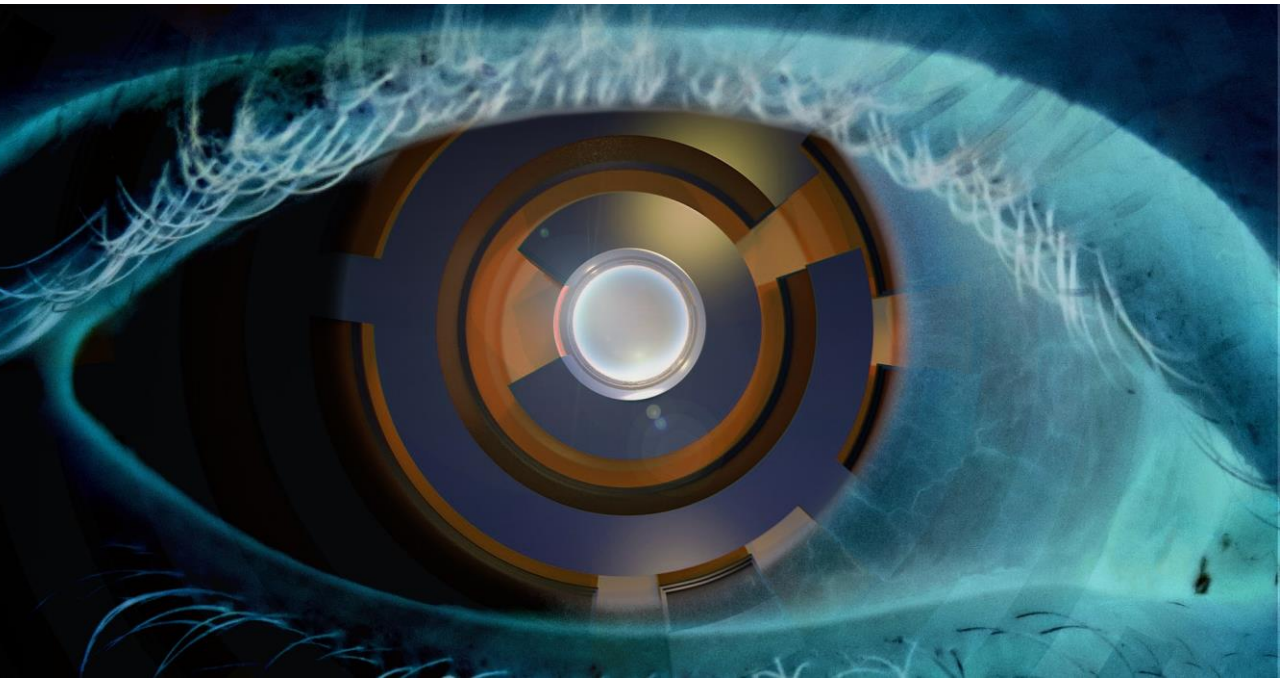
1 Danziger, S.; Levav, J. & Avnaim-Pesso, L.: “Extraneous factors in judicial decisions”, Proceedings of the National Academy of the Sciences, 2011 , 108 , 6889-6892

@nettwerkerin
Prof. KA Zweig
TU Kaiserslautern

Problemfall USA

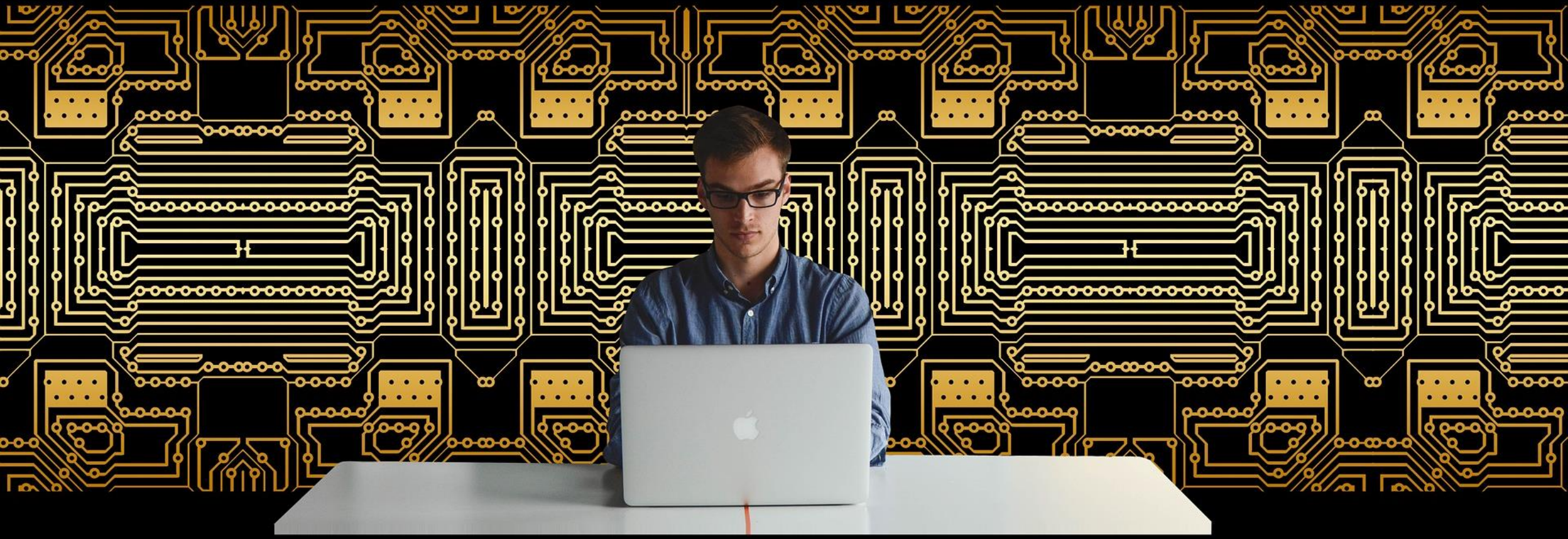
- Zweithöchste Inhaftierungssrate weltweit.
- 6x höhere Rate von Afroamerikanern und 2x höhere Rate von Latinos als von Weißen.
- Prognose: Jeder dritte afroamerikanische Junge im Alter von 10 Jahren wird eine Gefängnisstrafe absitzen müssen.

American Civil Liberties Union

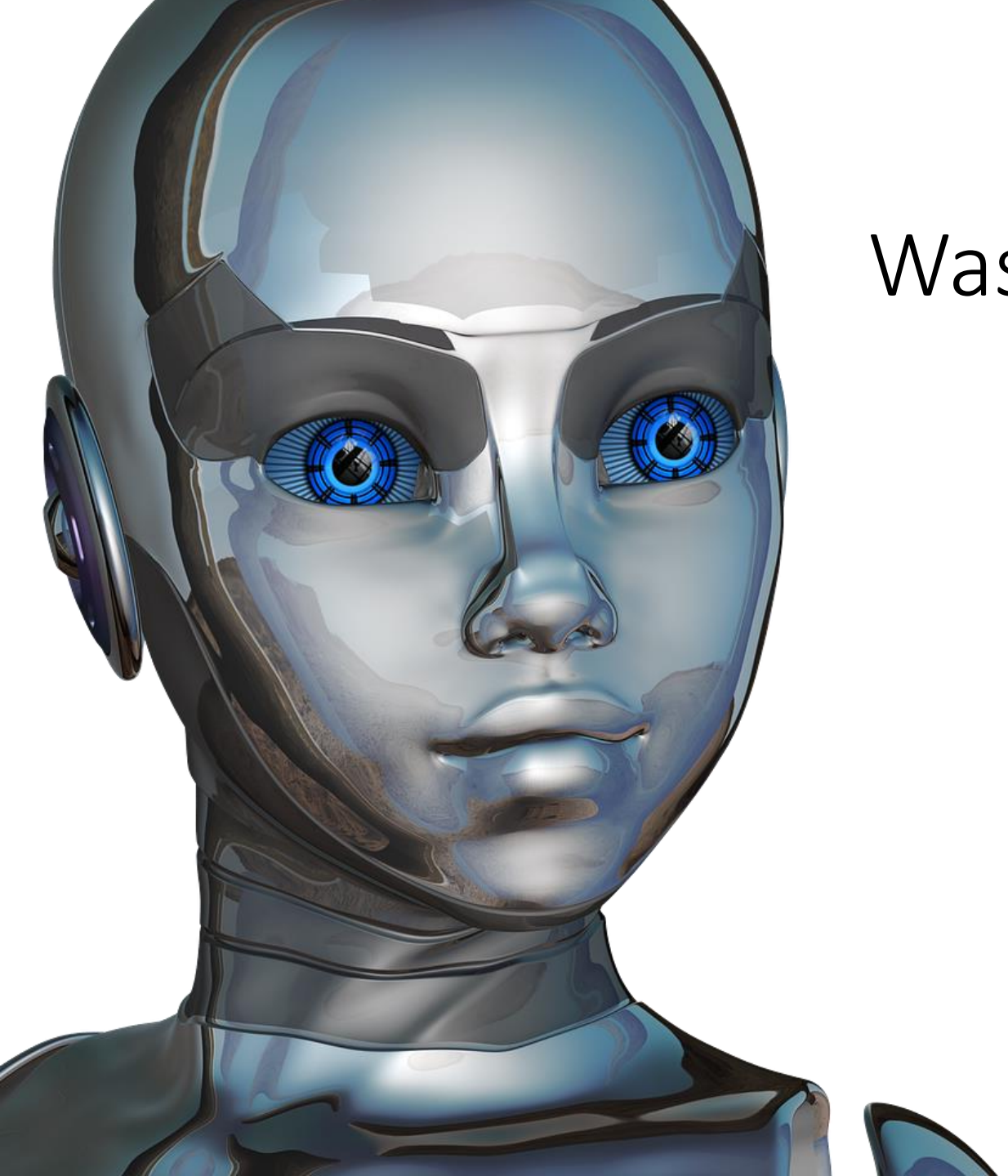


- Amerikanische Bürgerrechtsunion (seit 1920) fordert:
- Algorithmische Entscheidungssysteme sollten überall im Prozess eingesetzt werden, …
- … um Fairness und Objektivität zu sichern.
- Dazu sollen Computer aus Daten Entscheidungsregeln lernen.

SMART REFORM IS POSSIBLE - States Reducing Incarceration Rates and Costs While Protecting Communities, Report from August 2011, p. 9



Können Computer lernen?



Was heißt Lernen?

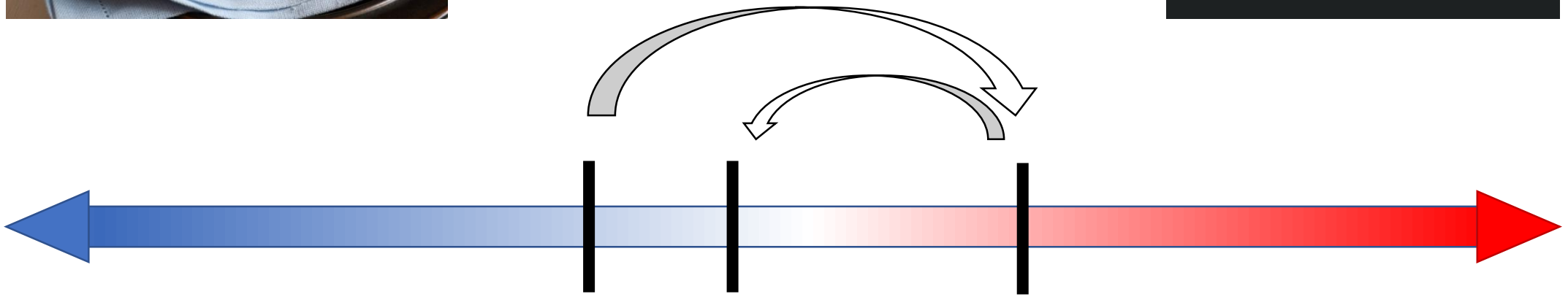
Einfach:

In derselben Situation ein vorher gezeigtes Verhalten wiederholen.

Generalisiert:

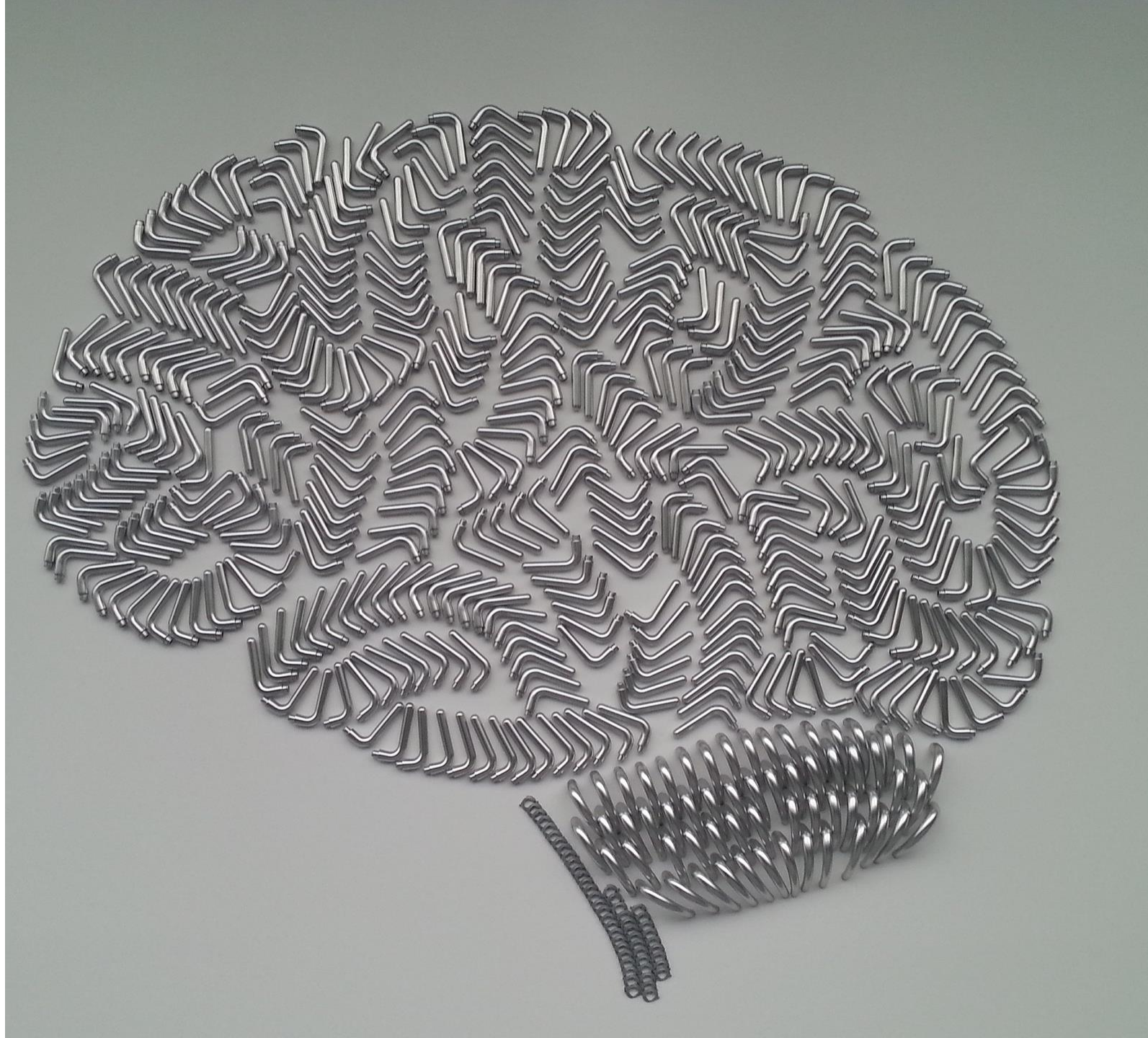
In derselben Art von Situation das richtige Verhalten aus einer Reihe von Möglichkeiten auswählen.

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**.

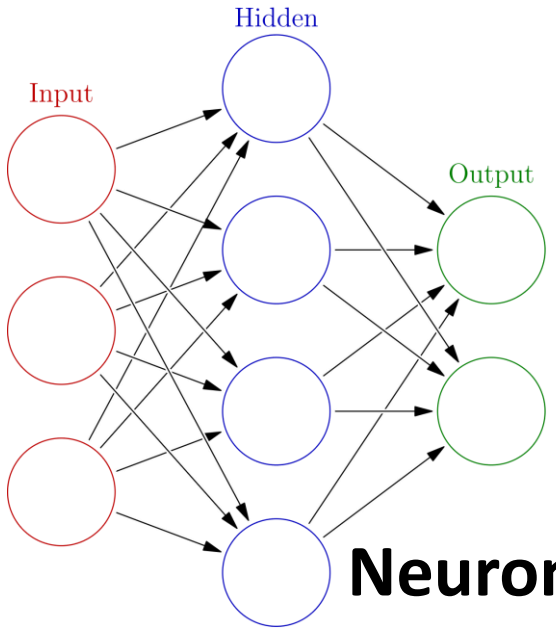


Computer lernen

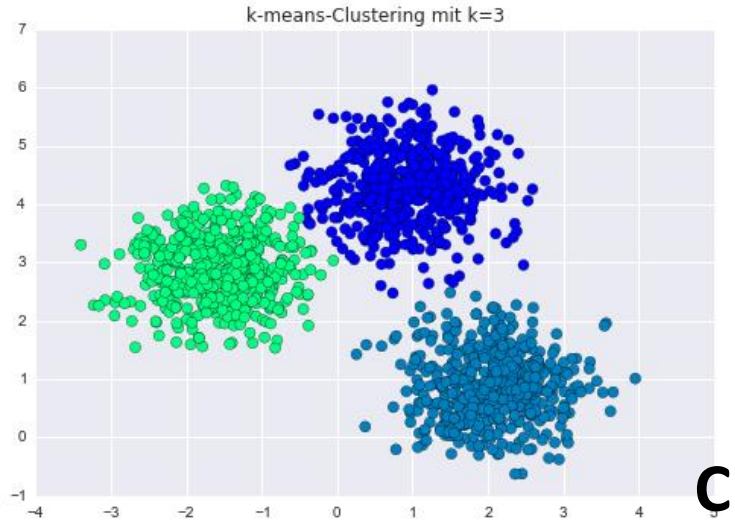
Damit ein Computer lernen kann, benötigt er ebenfalls eine **Struktur,** um Gelerntes abzuspeichern.

Optimal auch **Rückkopplung**.

Er lernt **generelle Regeln**.



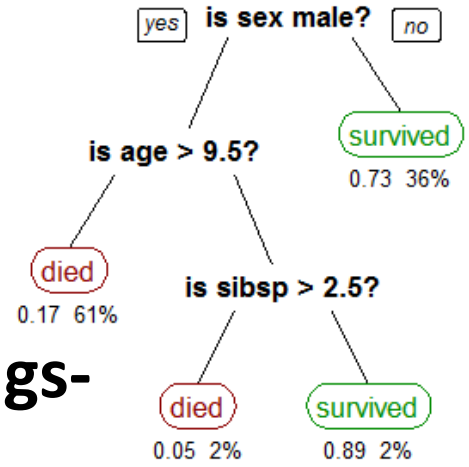
Neuronales Netz



Clustering

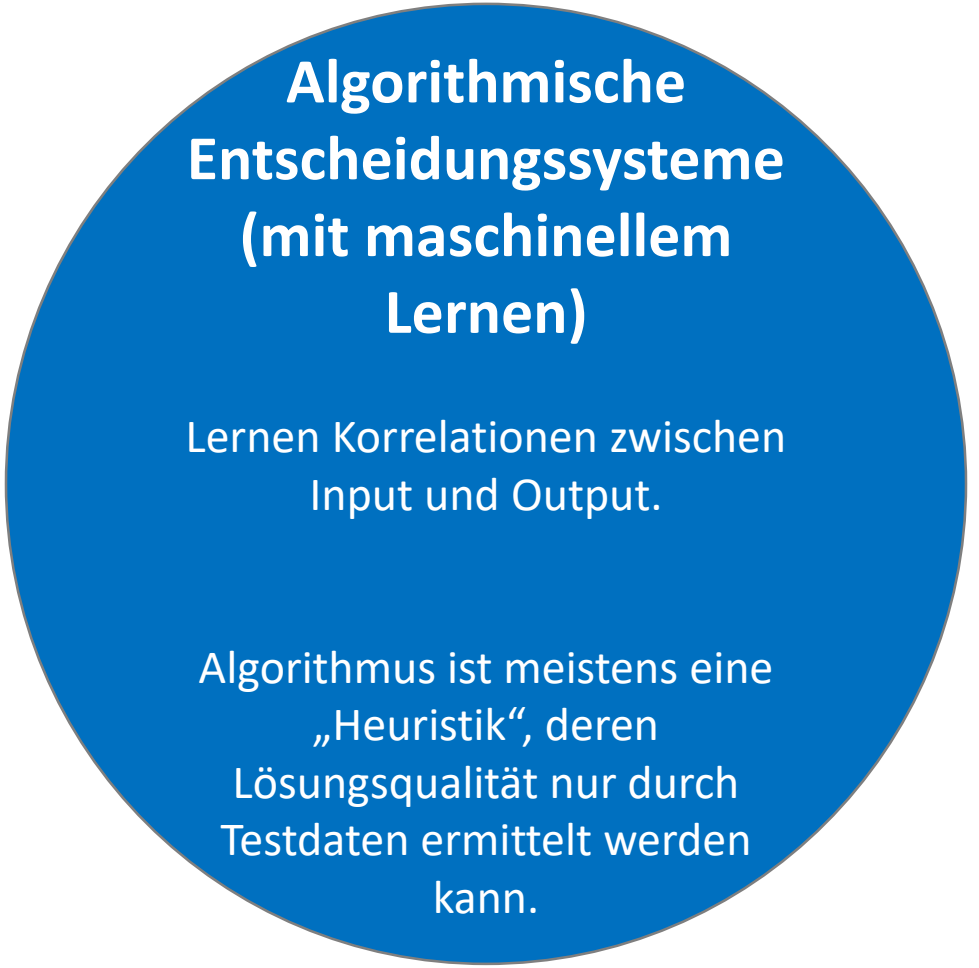
$$w_1 * \#Vh - w_2 * \#dayiVh + w_3 * I[g = male] + B[\ * 1.0 + \cdots$$

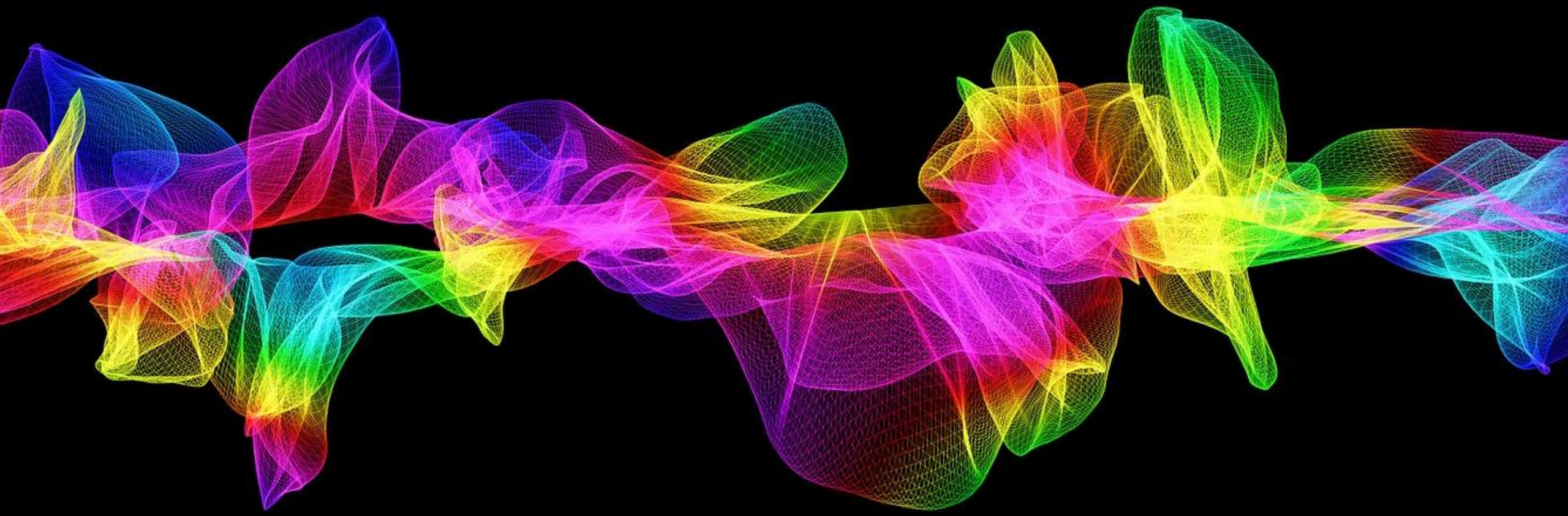
Formel



Entscheidungs-bäume

Algorithmen – eine Kategorisierung



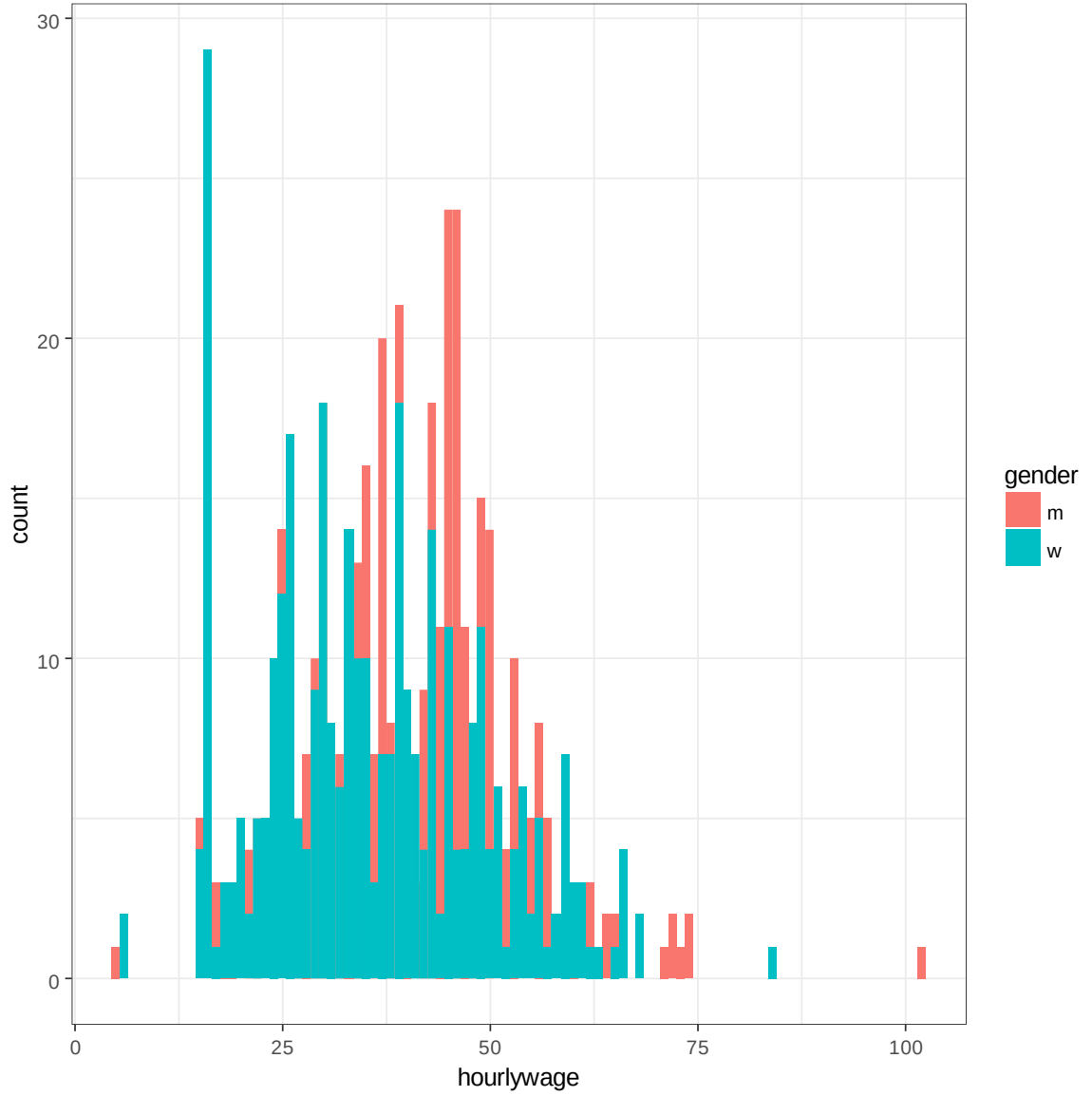


“Lernen” mit Korrelationen

Gehälter in Seattle

Sie bekommen Daten von einer Person – diese verdient weniger als $25 pro Stunde.

Basierend auf den Daten, ist die Person weiblich oder männlich?



@nettwerkerin
Prof. KA Zweig
TU Kaiserslautern

Lernen mit Formeln

Individuelle Risikobewertung der Rückfälligkeit von Kriminellen

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 zugeordnete) ethnische Zugehörigkeit.
- Wichtig: Beim Trainingsset ist bekannt, ob die Person rückfällig geworden ist oder nicht.

Regressionsansätze

- Algorithmendesigner entscheiden, welche der Daten vermutlich mit „Rückfallwahrscheinlichkeit“ korrelieren.
- Resultat sollte eine einzige Zahl sein.
- Je höher die Zahl, desto höher die Rückfallwahrscheinlichkeit.
- Beispiel Formel:

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

Allgemein

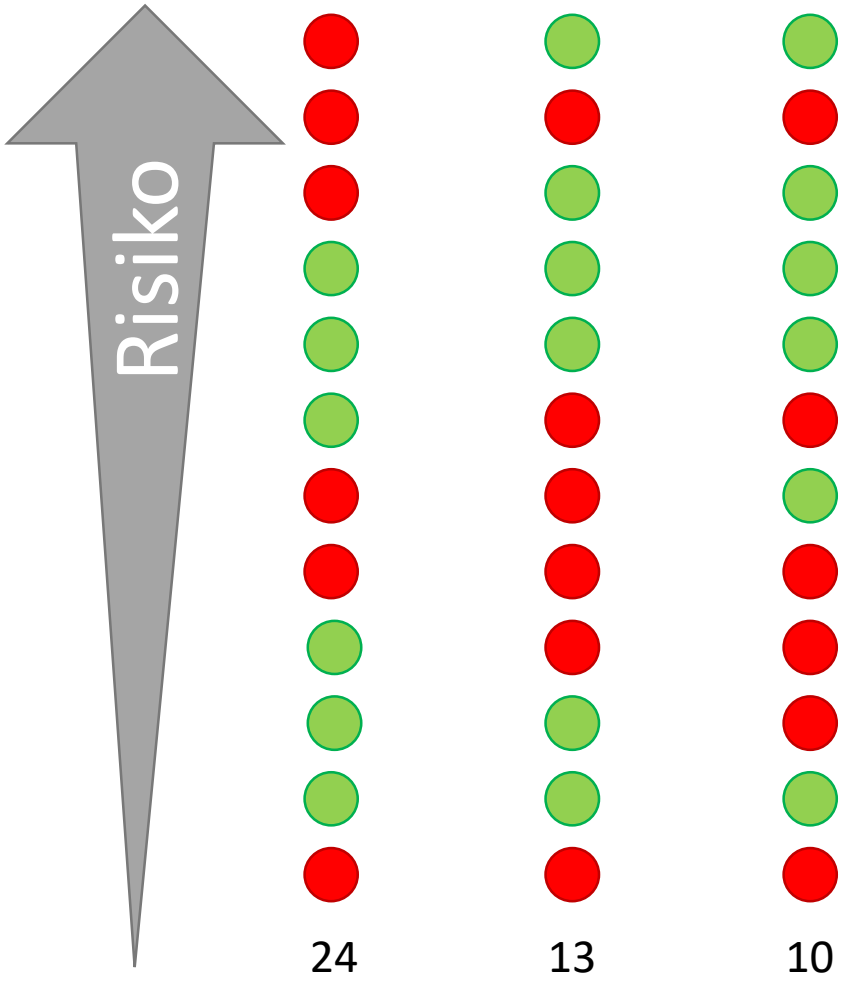
$$
\begin{aligned}
& w_1 * \text{ bisherige Verhaftungen} \\
- & w_2 * \text{ Anzahl Tage seit letzter Verhaftung} \\
+ & w_3 * \text{ (Wenn Mann, dann 1, sonst 0)} \\
+ & w_4 * \text{ (Wenn Raubüberfall, dann 1, sonst 0)} + ...
\end{aligned}
$$

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.

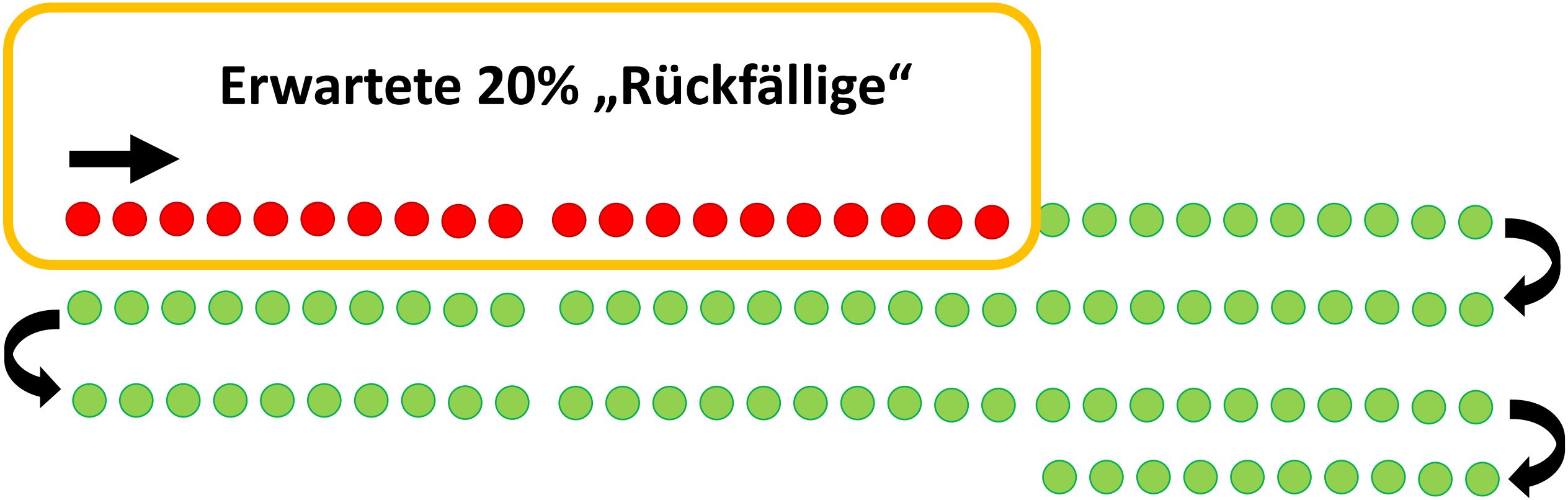
Optimale Sortierung: Alle roten oben, alle grünen darunter.

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

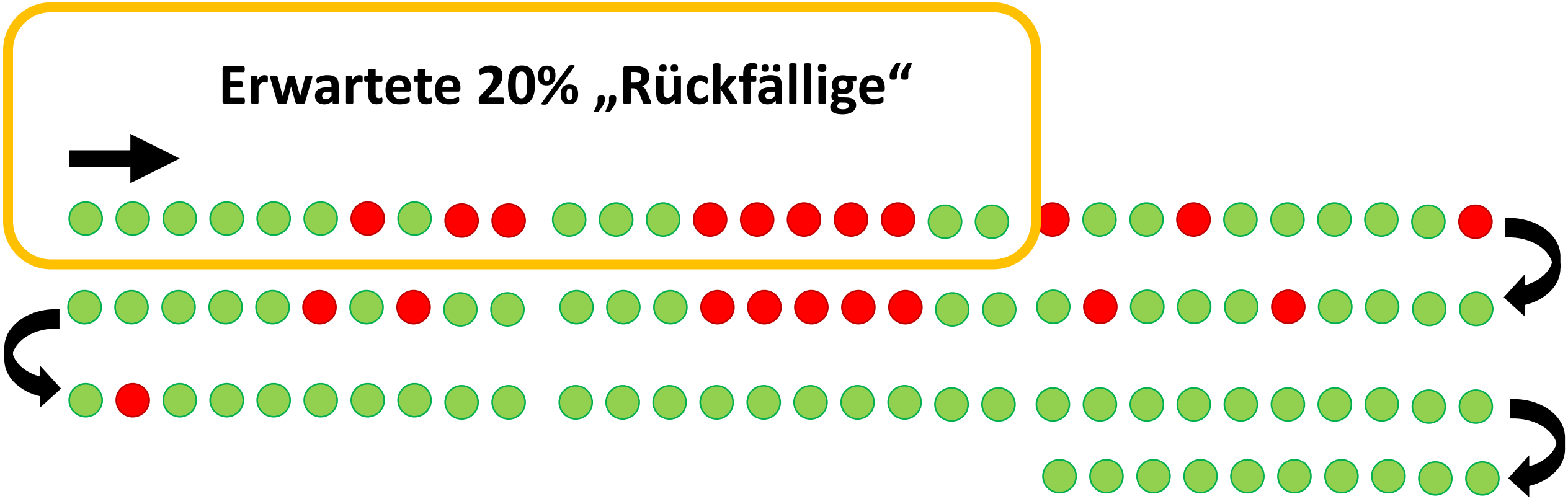
Oregon Recidivism Rate Algorithm

- 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)

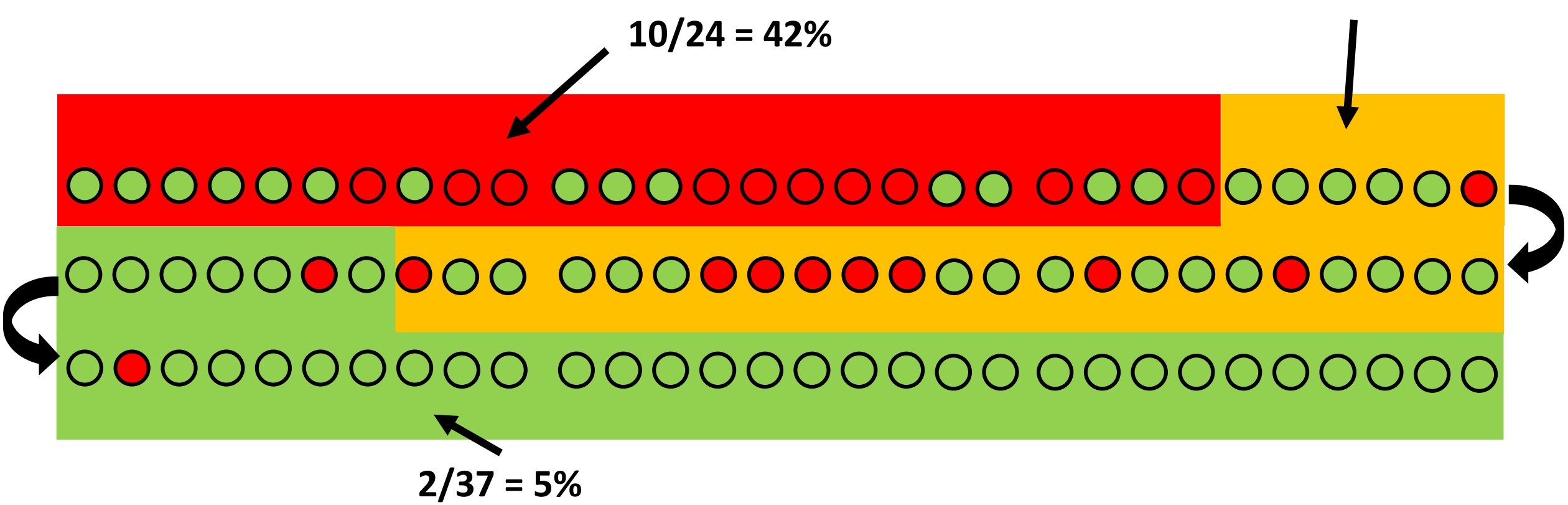


Einen Hund zum Jagen auszubilden, aber zum Schafehüten zu nutzen.

Das ist wie...

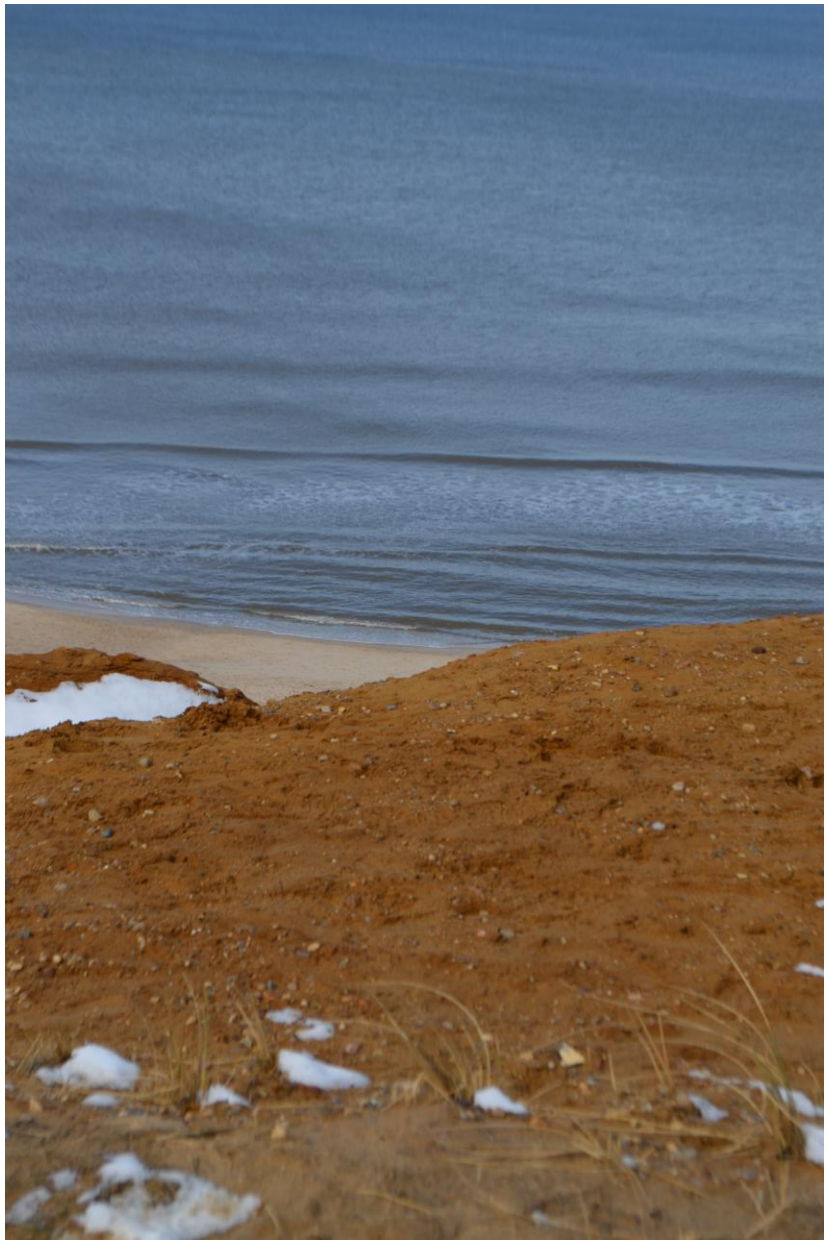
Vom Scoring zur Klassifikation

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



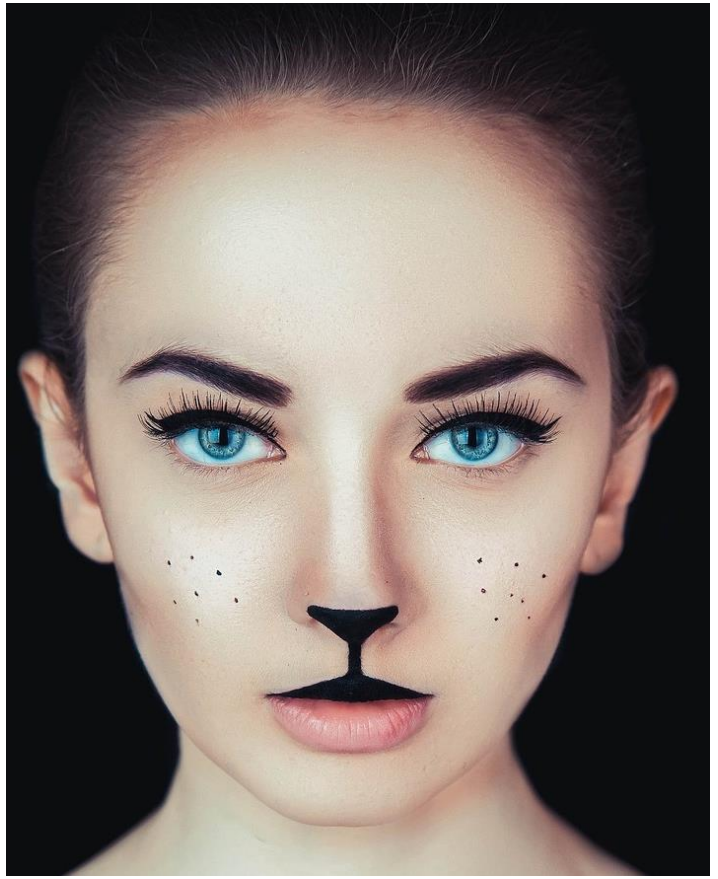
Statistische Vorhersagen über Menschen

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?

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“.

Algorithmen in einer demokratischen Gesellschaft

Generell

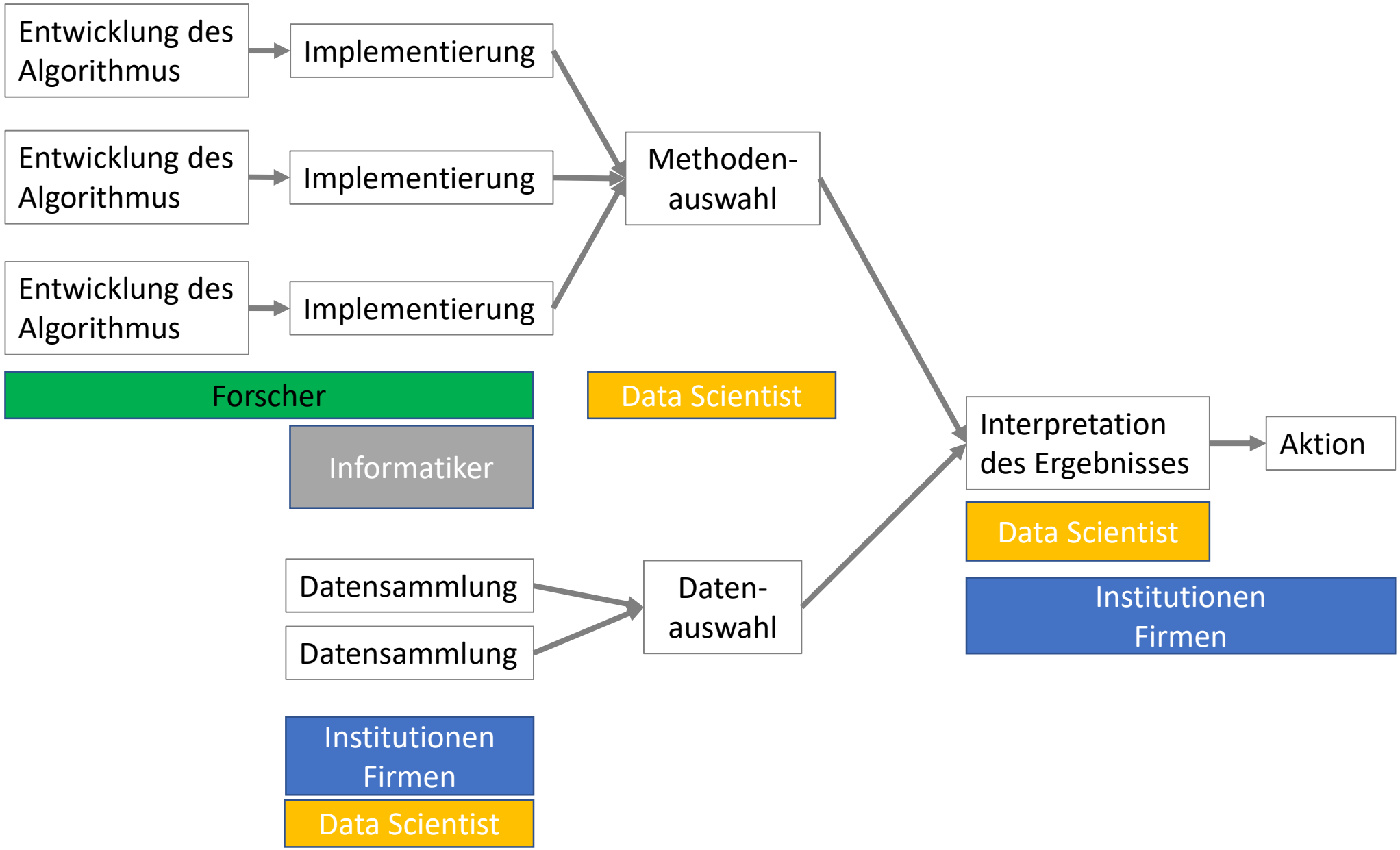
Prinzipiell können algorithmische Entscheidungssysteme für sehr viele, schwierige Fragestellungen in derselben Art gebaut werden:

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

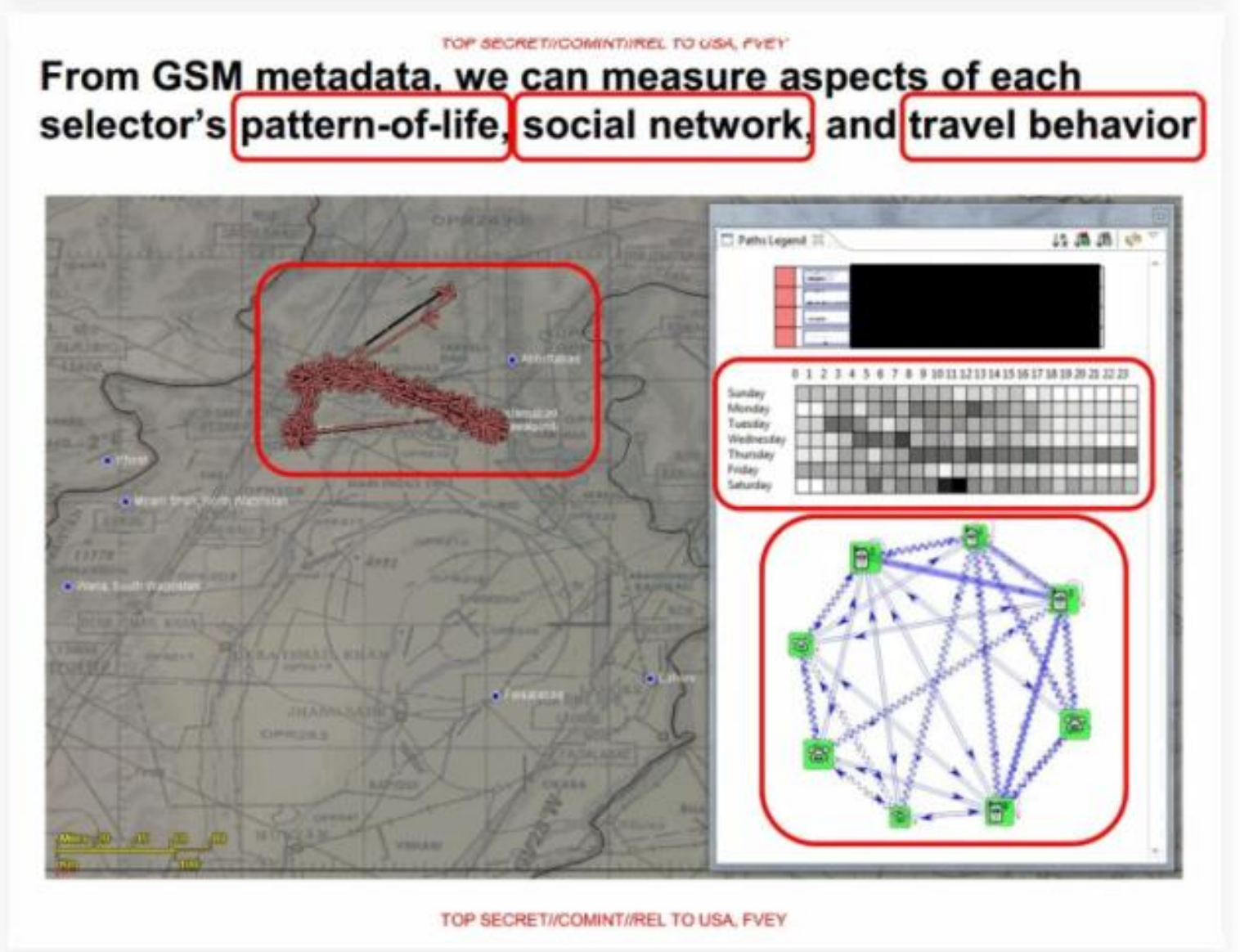
Ihre Aufgabe heute….

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

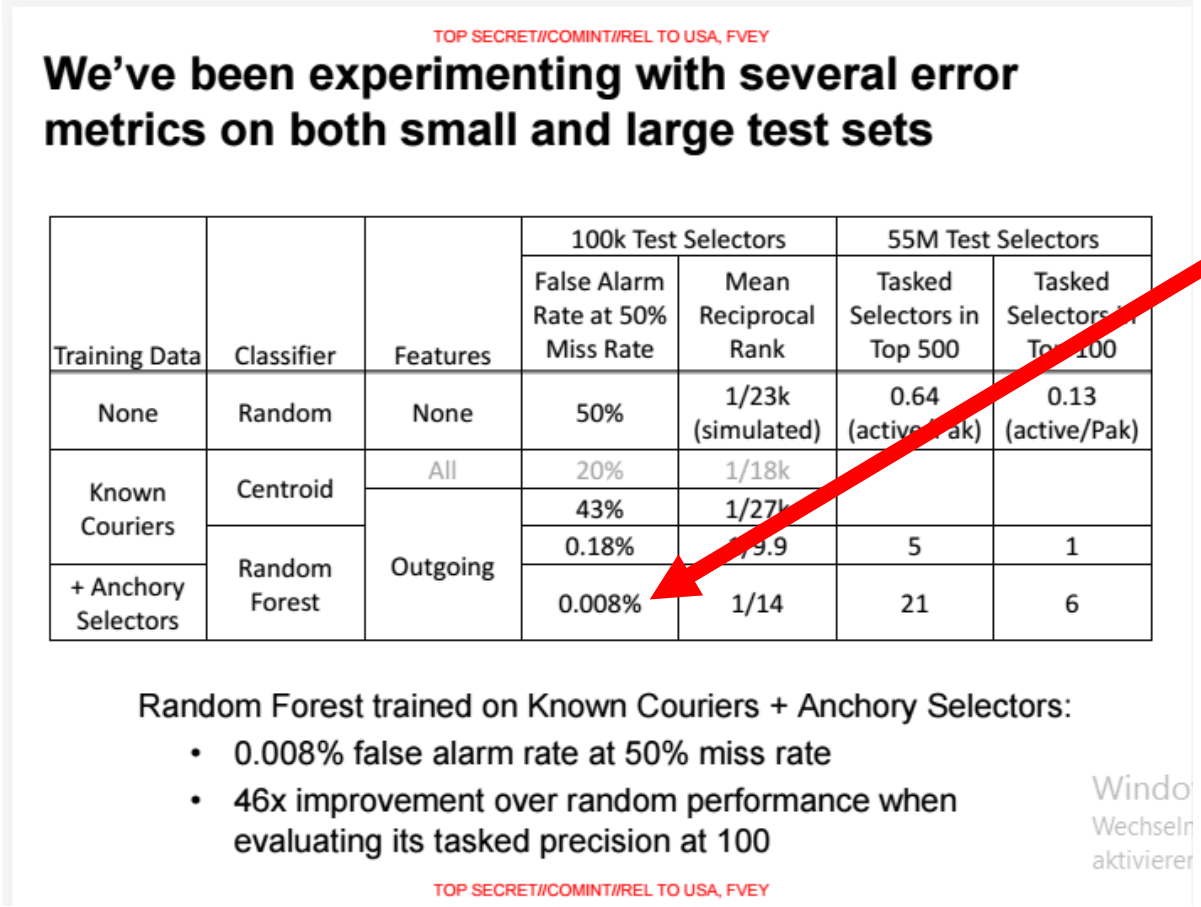
Designprozess



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:

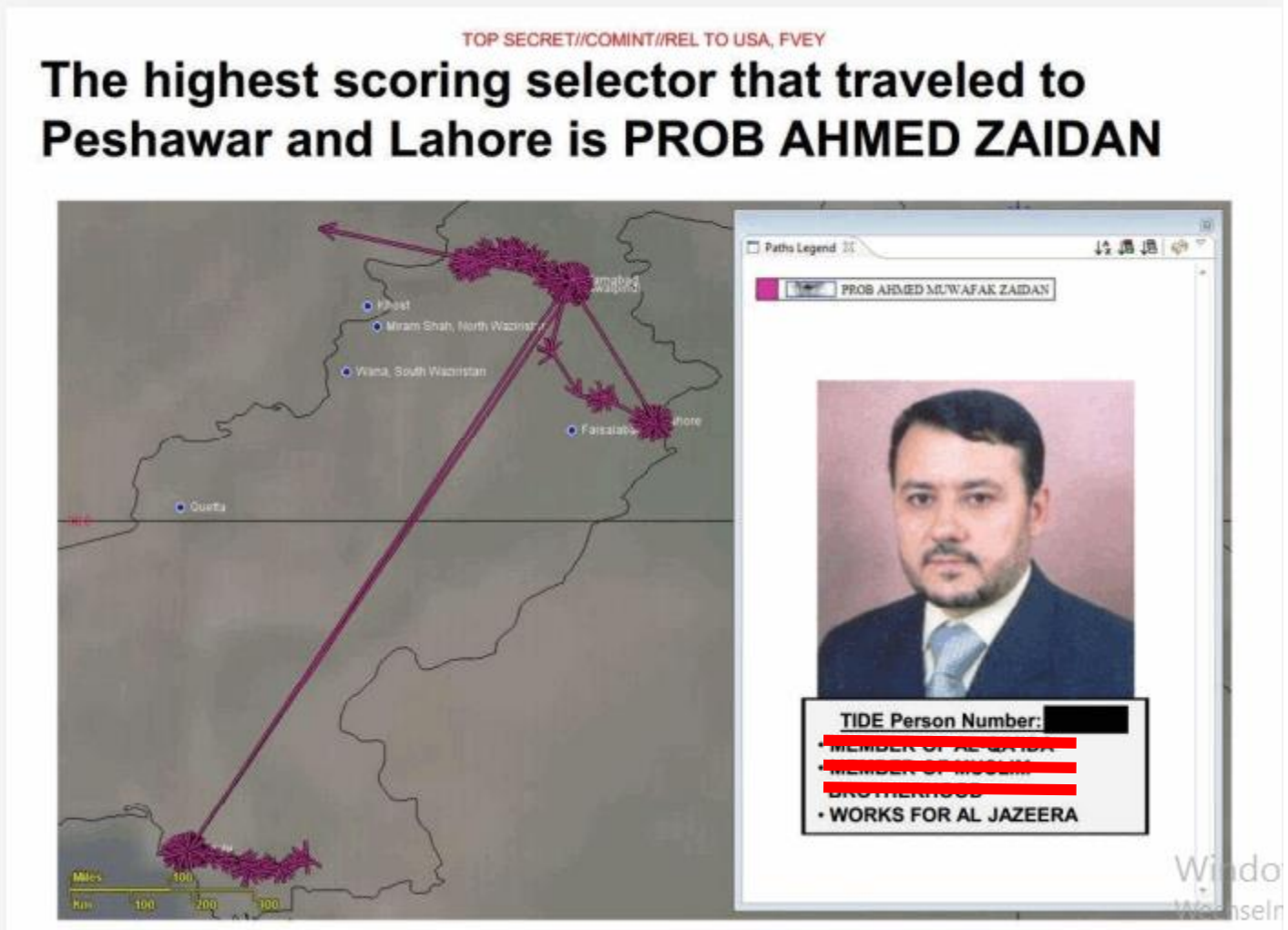
- 0.008% false alarm rate at 50% miss rate
- 46x improvement over random performance when evaluating its tasked precision at 100

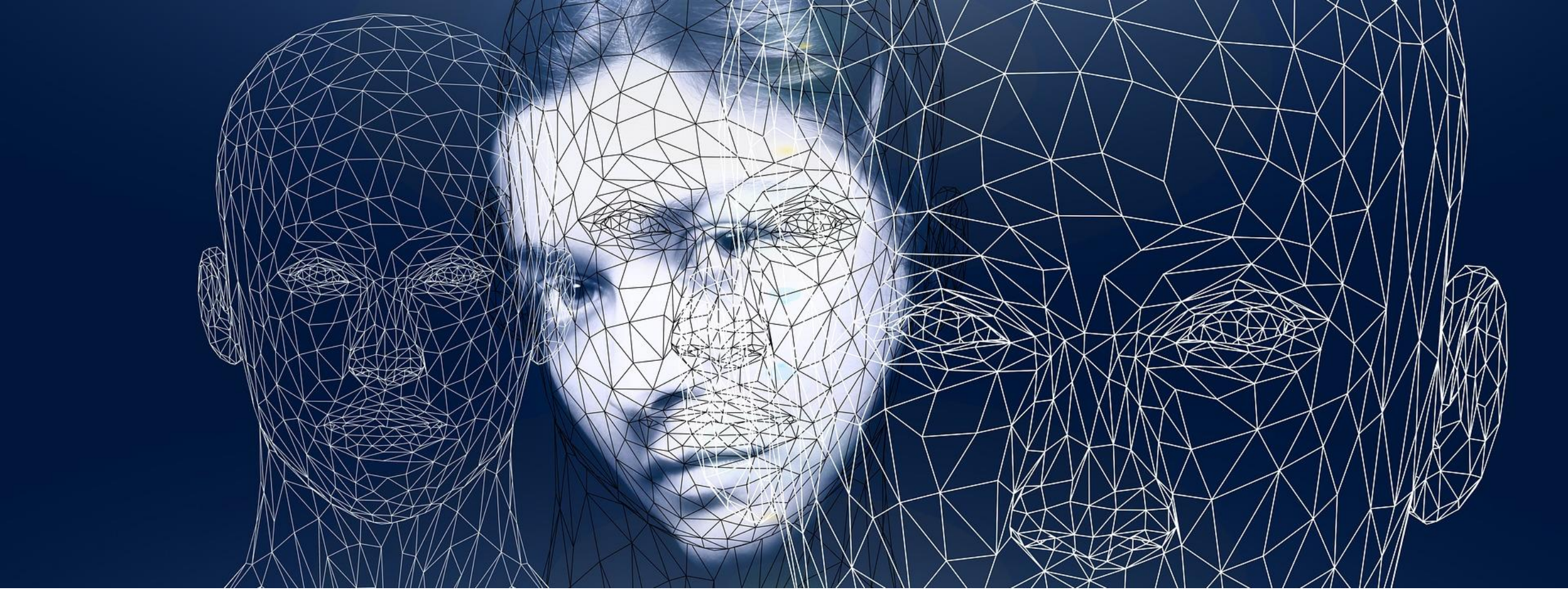
TOP SECRET//COMINT//REL TO USA, FVEY

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…





Sozio-informatische Gesamtbetrachtung

Probleme der Einbettung der ADM in den sozialen Prozess

- **Aufmerksamkeitsökonomie** von Entscheiderinnen und Entscheidern.
- **„Best practice“** erfordert Nutzung der Software.
- **Delegierung von Verantwortung!**
- Manchmal kann ein(e) falsch-negativ Beurteilte(r) **die Vorhersage prinzipiell nicht entkräften!**
  - Z.B. abgelehnte Bewerberin oder ins Gefängnis gesteckte Kriminelle

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.

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

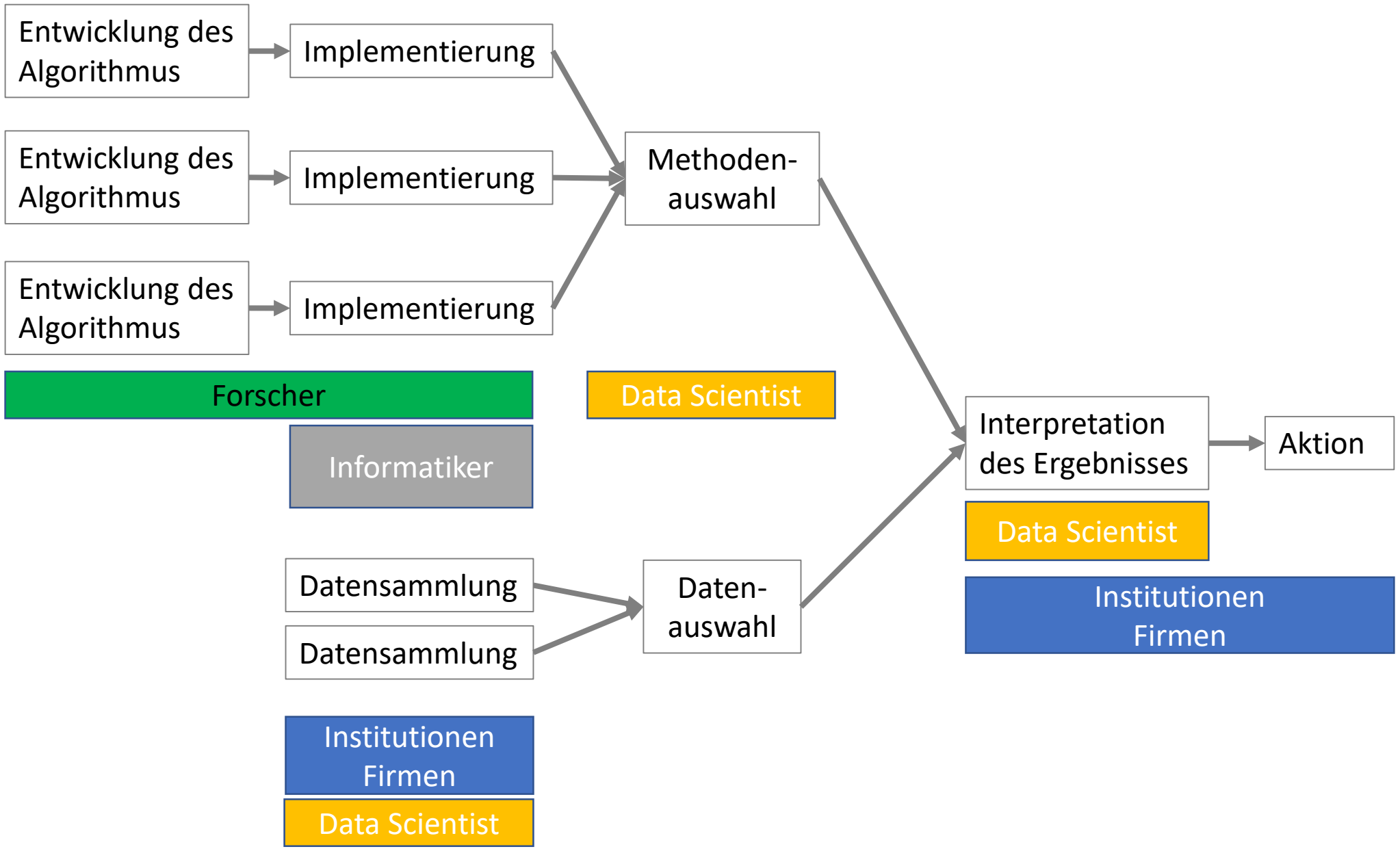
1. **Wer entscheidet, wann ein ADM System „gut“ ist?**
2. **ADM Systeme ergeben nur Wahrscheinlichkeiten, keine Wahrheiten.**
3. **ADM Systeme können diskriminieren.**
4. **ADM Systeme können soziale Prozesse verändern.**

@nettwerkerin
Prof. KA Zweig
TU Kaiserslautern

Quis custodiet ipsos algorithmos

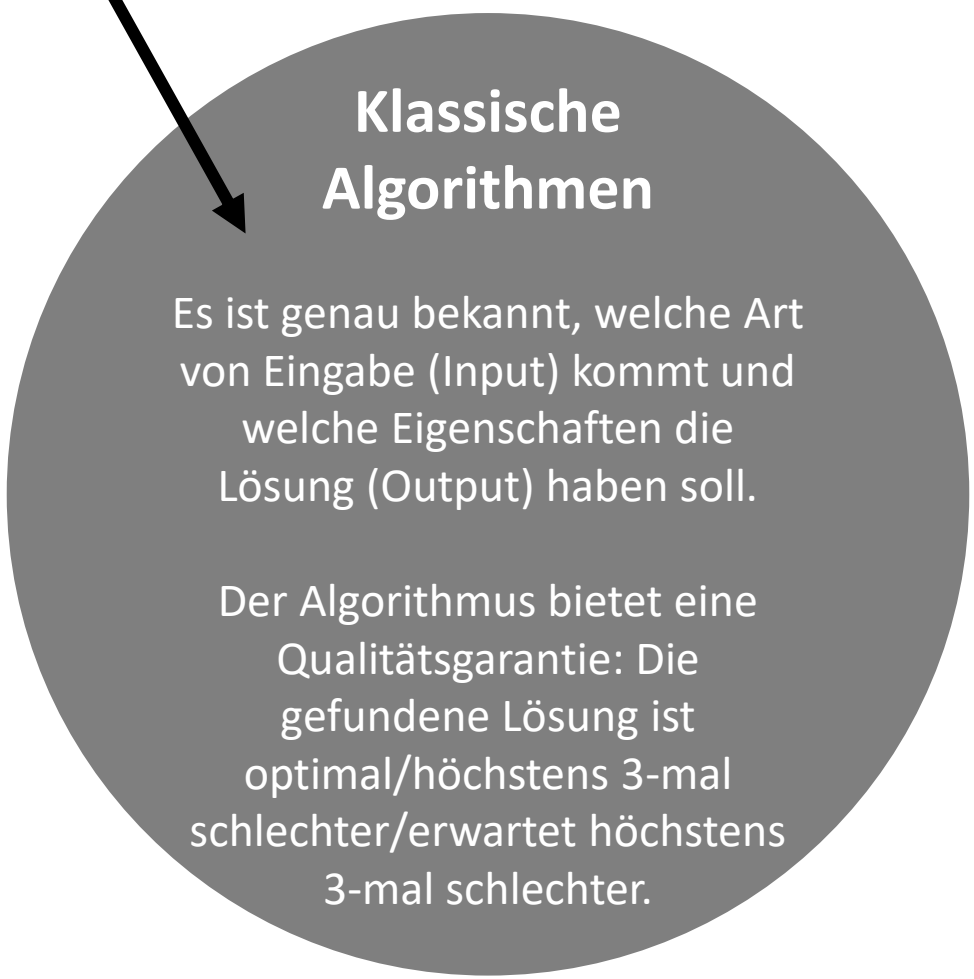
Der „Automated Decision Making“-TÜV vulgo: „Algorithmen TÜV“(Kenneth Cukier und Viktor Mayer-Schönberger: „Big Data“)

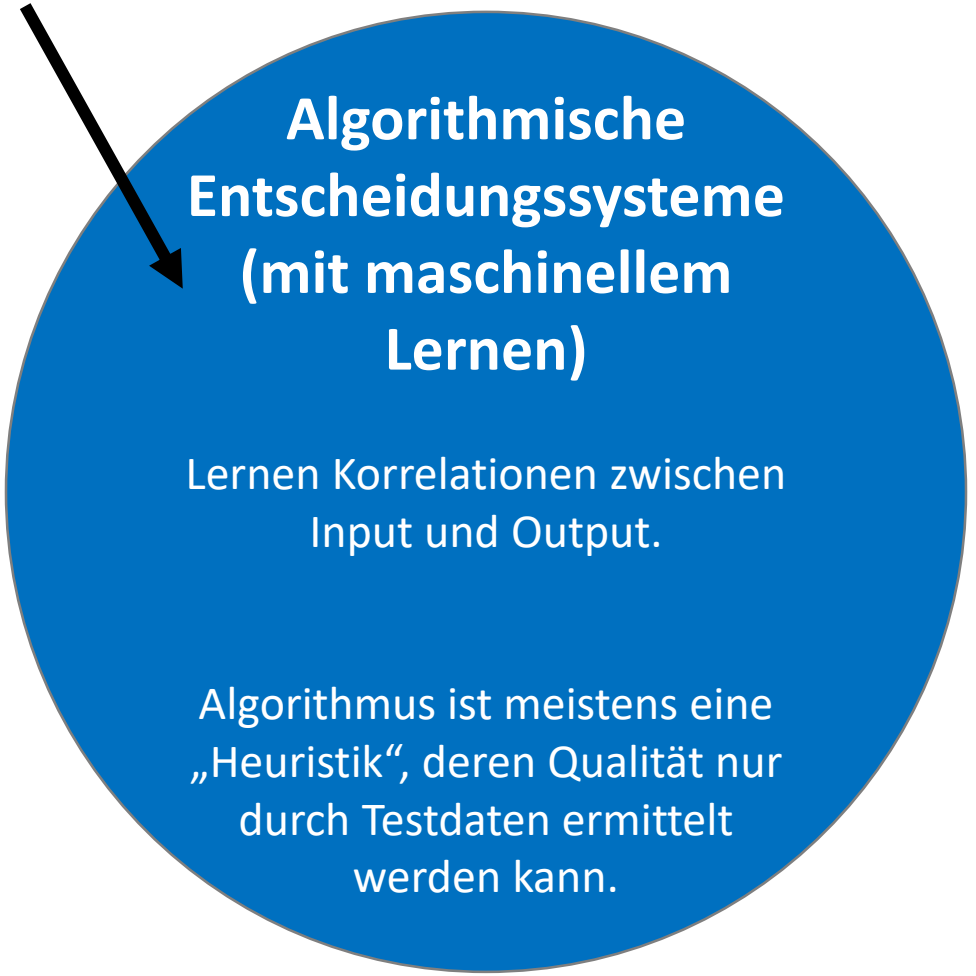
Verkettete Verantwortlichkeiten



Die hier haben wir einigermaßen im Griff mit bisherigen Verfahren und Institutionen

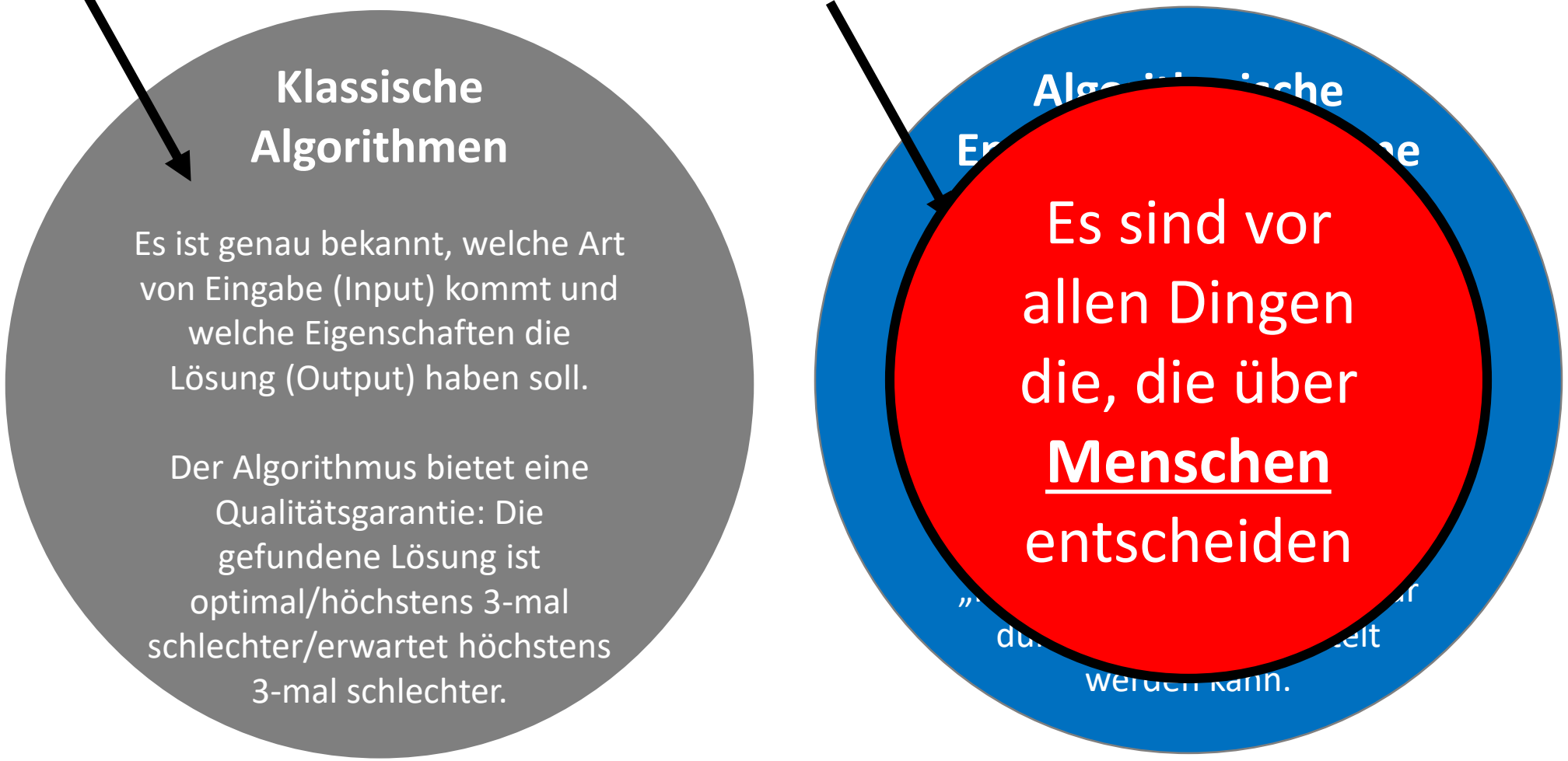
Auch hier sind viele unproblematisch: Die ohne direkten Bezug zum Menschen, z.B. Qualitätskontrolle, Bilderkennung i.A., Übersetzungen.





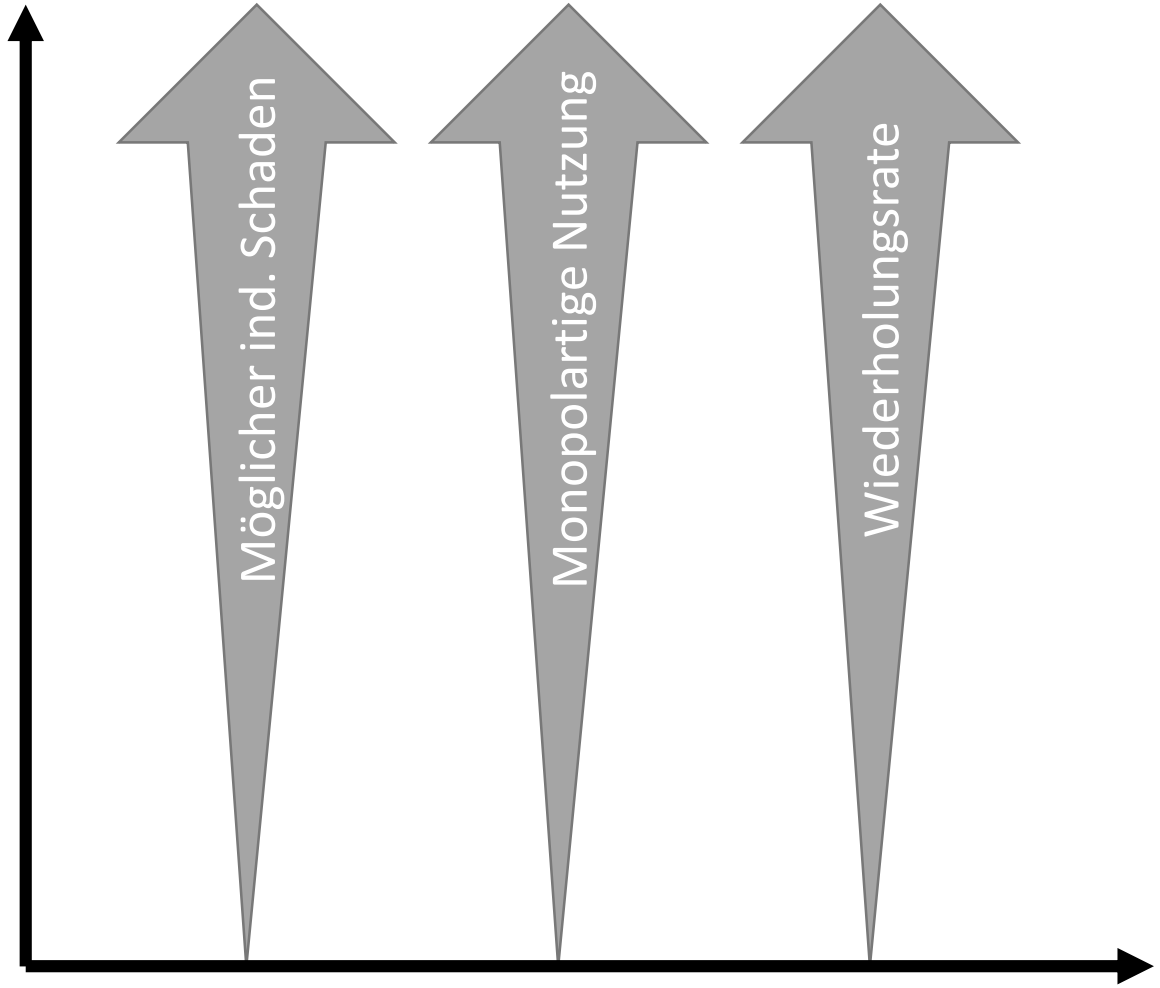
Die hier haben wir einigermaßen im Griff mit bisherigen Verfahren und Institutionen

Auch hier sind viele unproblematisch: Die ohne direkten Bezug zum Menschen, z.B. Qualitätskontrolle, Bilderkennung i.A., Übersetzungen.



Notwendigkeit von Technikfolgenabschätzung und Technikfolgenüberwachung

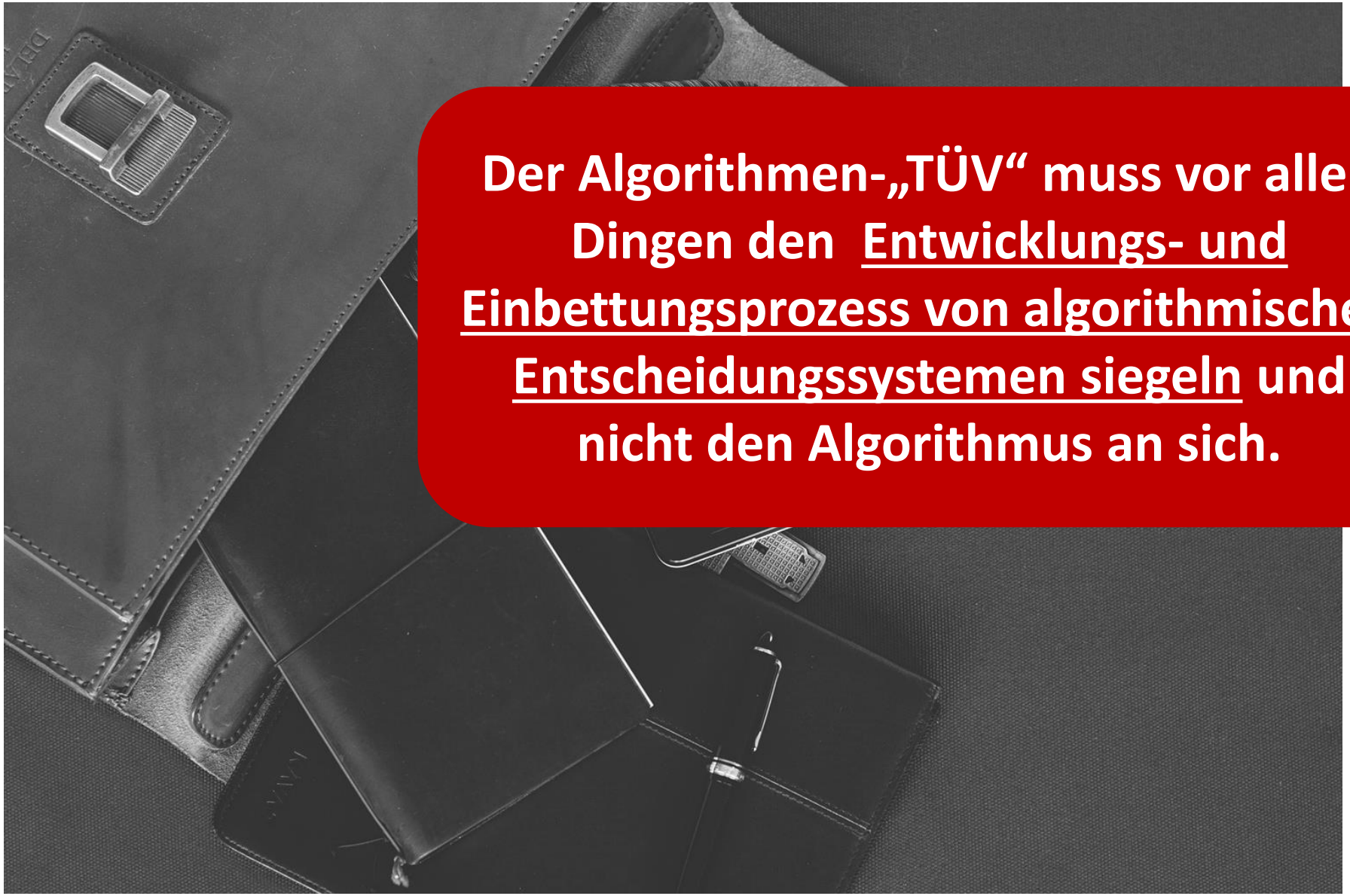
Technikfolgen-abschätzung und Technikfolgen-überwachung notwendig



Im Nachhinein, bei Verdachtsfall ausreichend?

Wie könnte ein „Algorithmen-TÜV“ aussehen?

- Unabhängig, mit Forschungsauftrag
- Identifikation der **kleinstmöglichen Menge** an zu überprüfenden Algorithmen
  - Die meisten Algorithmen sind harmlos;
  - Produkthaftung ermöglicht, dass andere, z.B. Versicherungen, Interesse an korrekten Algorithmen haben;
  - Wettbewerb ermöglicht, dass andere ‚neutralere‘ Algorithmen anbieten.
  - Ökosystem von verschiedenen Institutionen
  - Kein weiteres **Innovationshemmnis!**
- **Nicht der Algorithmus, sondern die Qualität der Entwicklungskriterien und die Evaluation der Einbettung von algorithmischen Entscheidungssystemen muss gewährleistet werden!**
- **Non-Profit**



Der Algorithmen-„TÜV“ muss vor allen Dingen den Entwicklungs- und Einbettungsprozess von algorithmischen Entscheidungssystemen siegeln und nicht den Algorithmus an sich.

Schlussformel

… zu Risiken und Nebenwirkungen der Digitalisierung befragen Sie bitte Ihren nächstgelegenen Data Scientist oder den deutschen „Algorithmen-TÜV“.

Weitere Informationen

1. Broschüre der Bayerischen Landesmedienanstalt
Kostenlos zu beziehen von der BLM
Googlen nach „BLM Dein Algorithmus - meine Meinung“

Prof. Dr. Katharina A. Zweig
zweig@cs.uni-kl.de
@nettwerkerin bei Twitter

2. Studie für die Bertelsmann-Stiftung (2018)