



HAL übernehmen Sie!

Einsatz von KI in der Leistungsbewertung
und im Personalwesen

Prof. Dr. Katharina Zweig

TU Kaiserslautern

Leiterin des Algorithm Accountability Labs

Mitgründerin und wissenschaftliche Beraterin von AlgorithmWatch

„Employment asse_ + software“

Let's take the emotion out of the process and replace it with a data-driven approach...

No more people problems

We deal with employee issues so you can focus on what you do best.

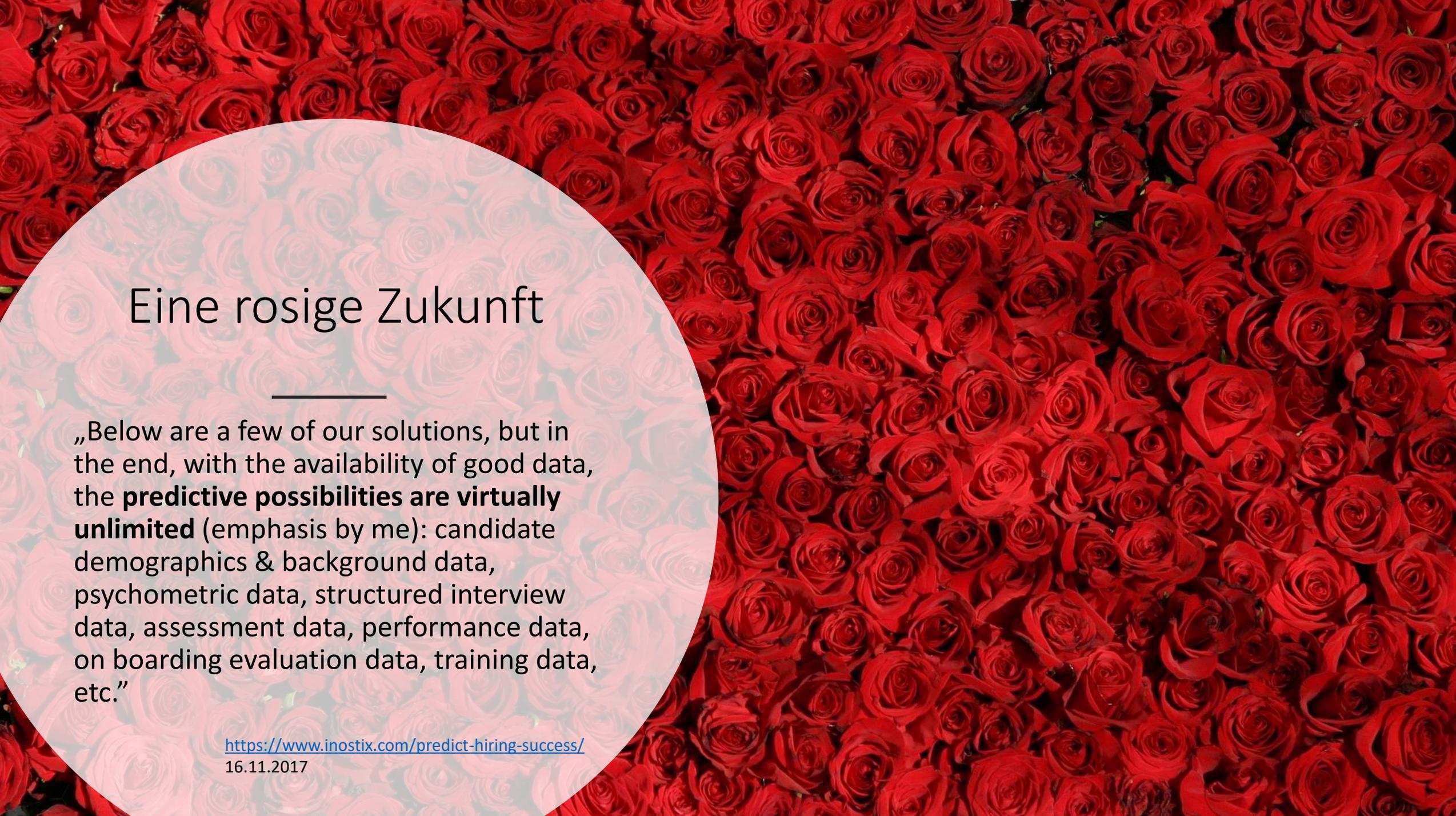


iNostix (by Deloitte),
16.11.2017


$$= a^2 \frac{(p(\theta) - Ci)^2}{(1 - C)^2}$$

Assessfirst.com,
16.11.2017

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

The background of the slide is a dense, repeating pattern of red roses. A large, semi-transparent white circle is overlaid on the left side of the image, containing the text.

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



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

Diskriminierung bei Bewerbungen

- Lebensläufe mit „deutschen“ Namen bekommen 14% Vorstellungsangebote als solche mit „türkischen“ Namen¹.
- US-amerik. Studie: Frauen mit Kopftuch erhalten weniger Jobangebote als solche ohne².



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

Menschen – so irrational!

- Richter müssen vorzeitige Haftentlassungsanträge begutachten.
- Studie: je weiter von der letzten Pause weg, desto weniger risikoreiche Entscheidungen¹.
- 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

Könnten Computer das besser?

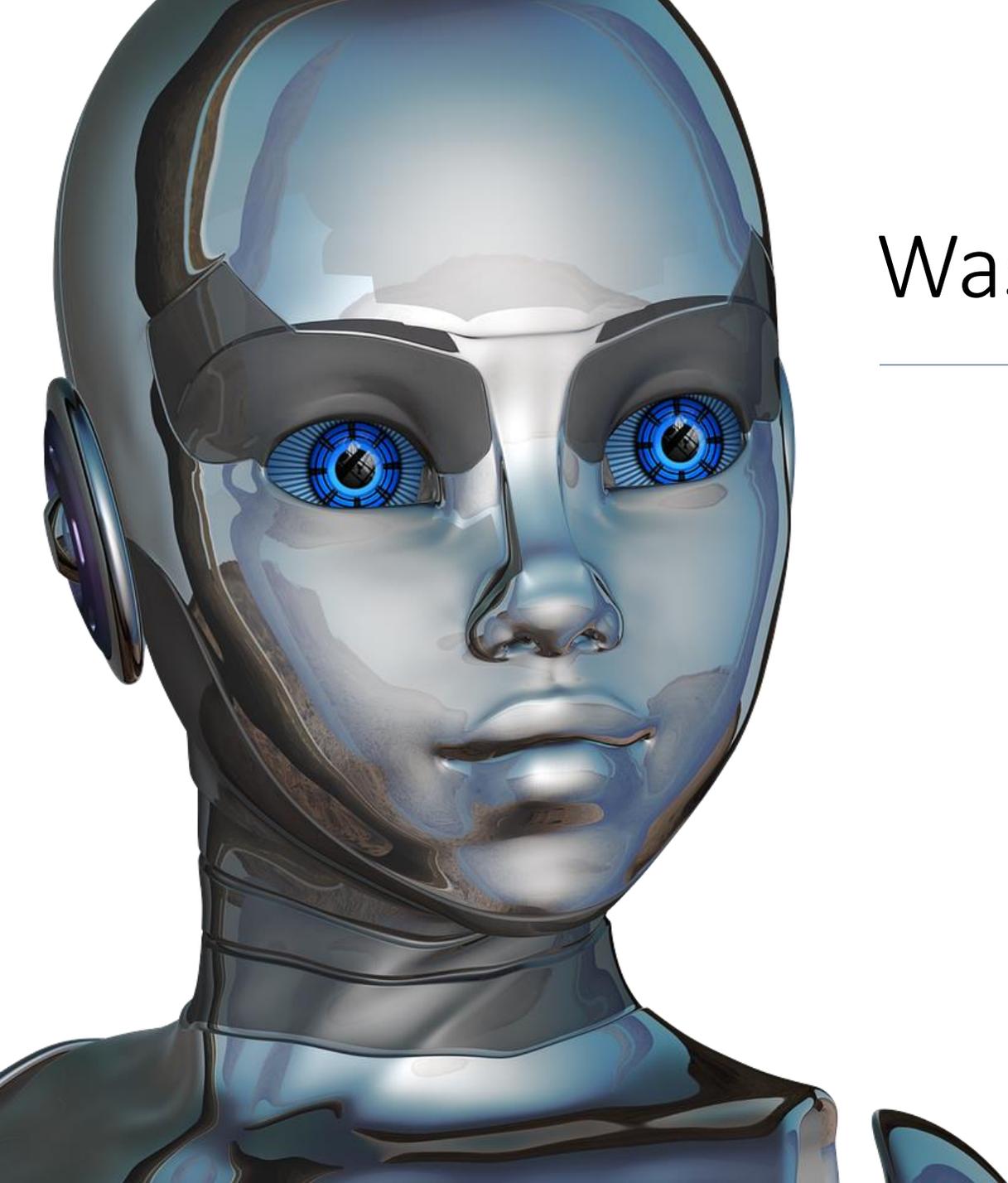
- Die ersten Firmen testen *algorithmische Entscheidungssysteme* (oder Entscheidungsunterstützungssysteme)¹.
- Eigenschaften, nach denen nicht diskriminiert werden darf, können vor ihnen besser verborgen werden.
- Sie sind objektiv und arbeiten nahezu fehlerfrei.
- (objektiv := „reproduzierbar dieselbe Entscheidung bei derselben Eingabe von Daten“)

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>





Können Computer lernen?



Was heißt Lernen?

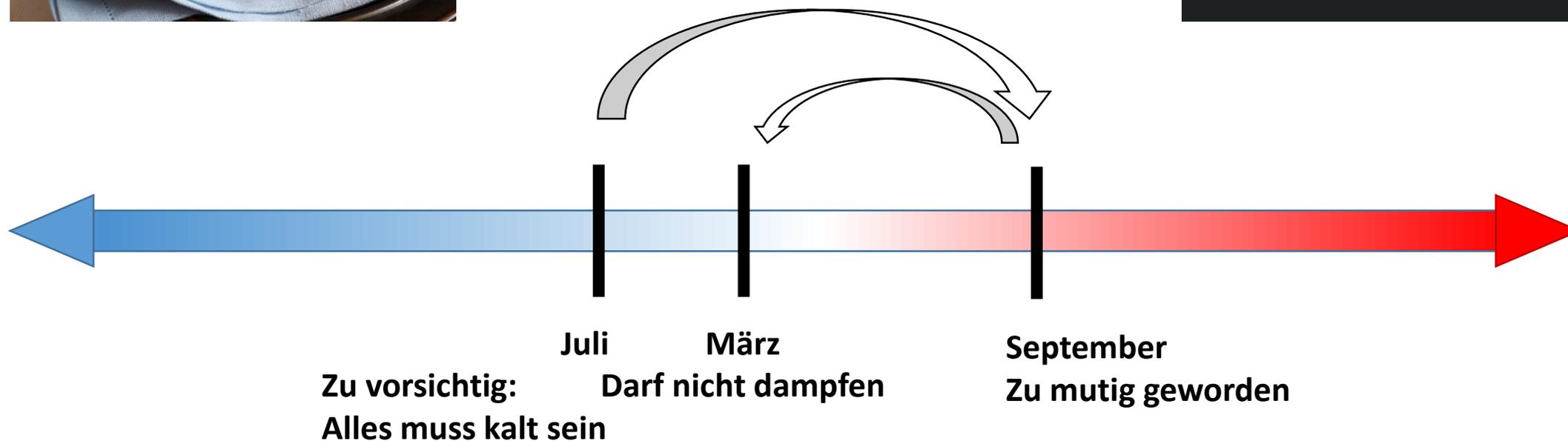
Einfach:

In derselben Situation ein vorher gezeigtes Verhalten wiederholen.

Generalisiert:

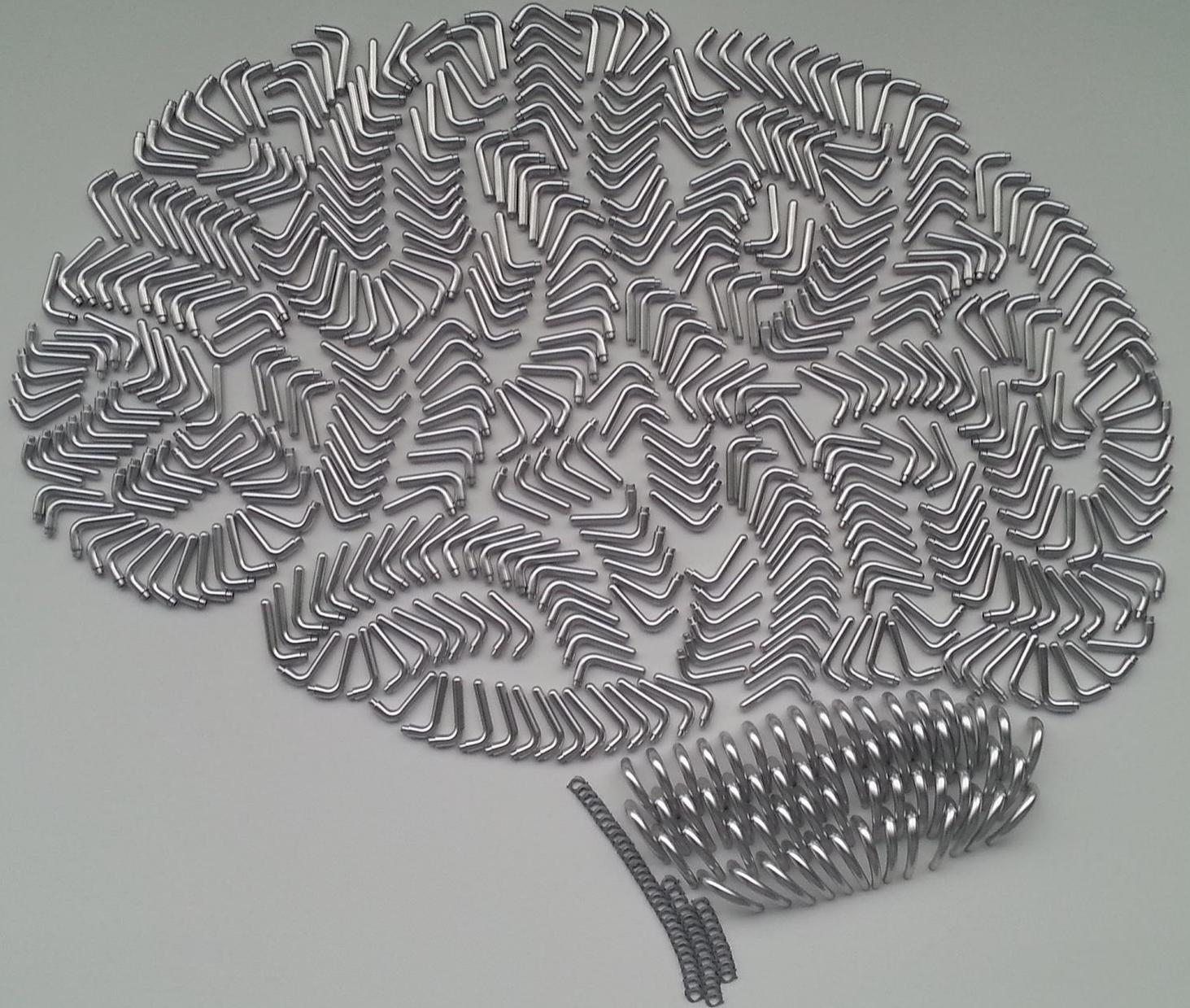
In derselben Art von Situation das richtige Verhalten aus einer Reihe von Möglichkeiten auswählen.

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**.

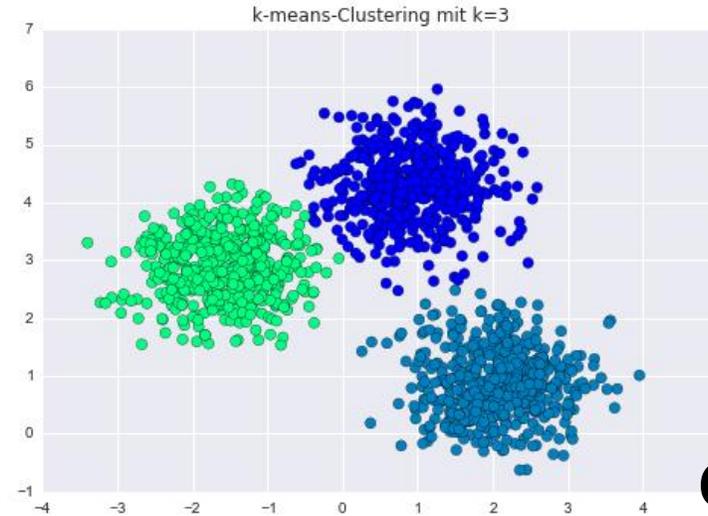
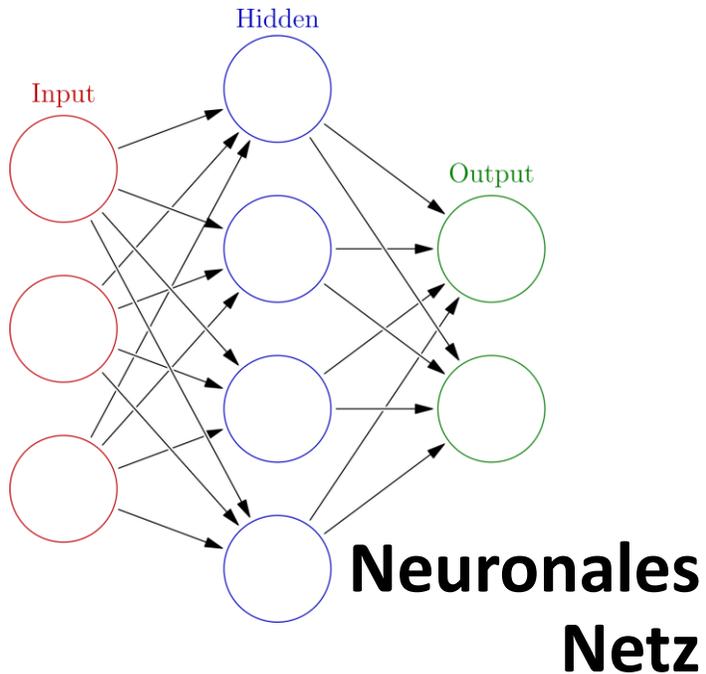


Computer lernen

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

Optimal auch **Rückkopplung**.

Er lernt **generelle Regeln**.

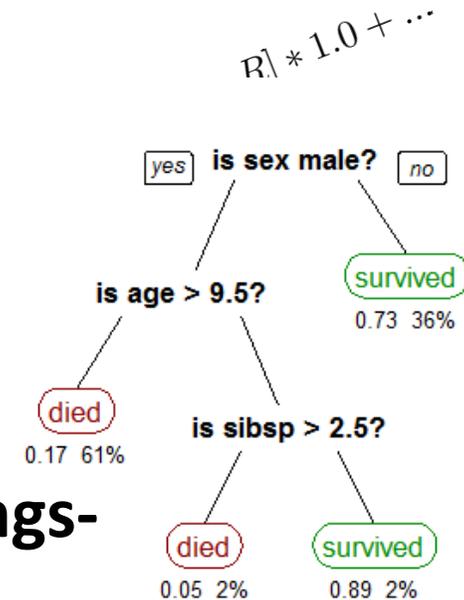


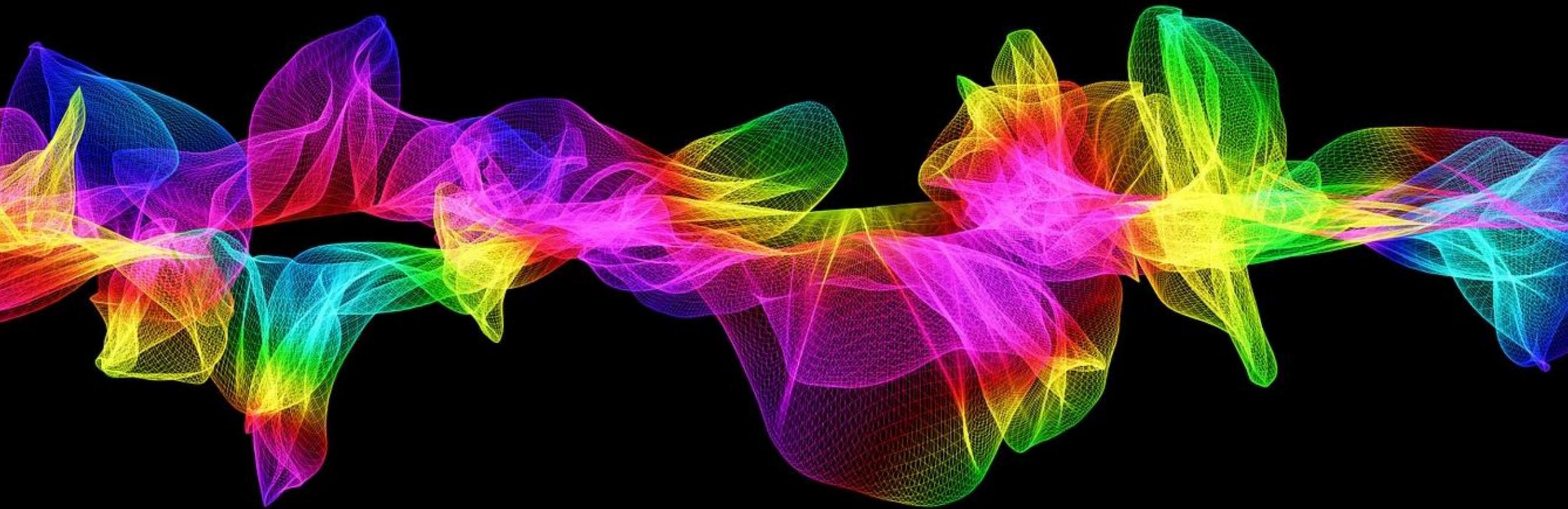
Clustering

Formel

$$w_1 * \#V_h - w_2 * \#day_i V_h + w_3 * I[g = male]$$

Entscheidungs-
bäume

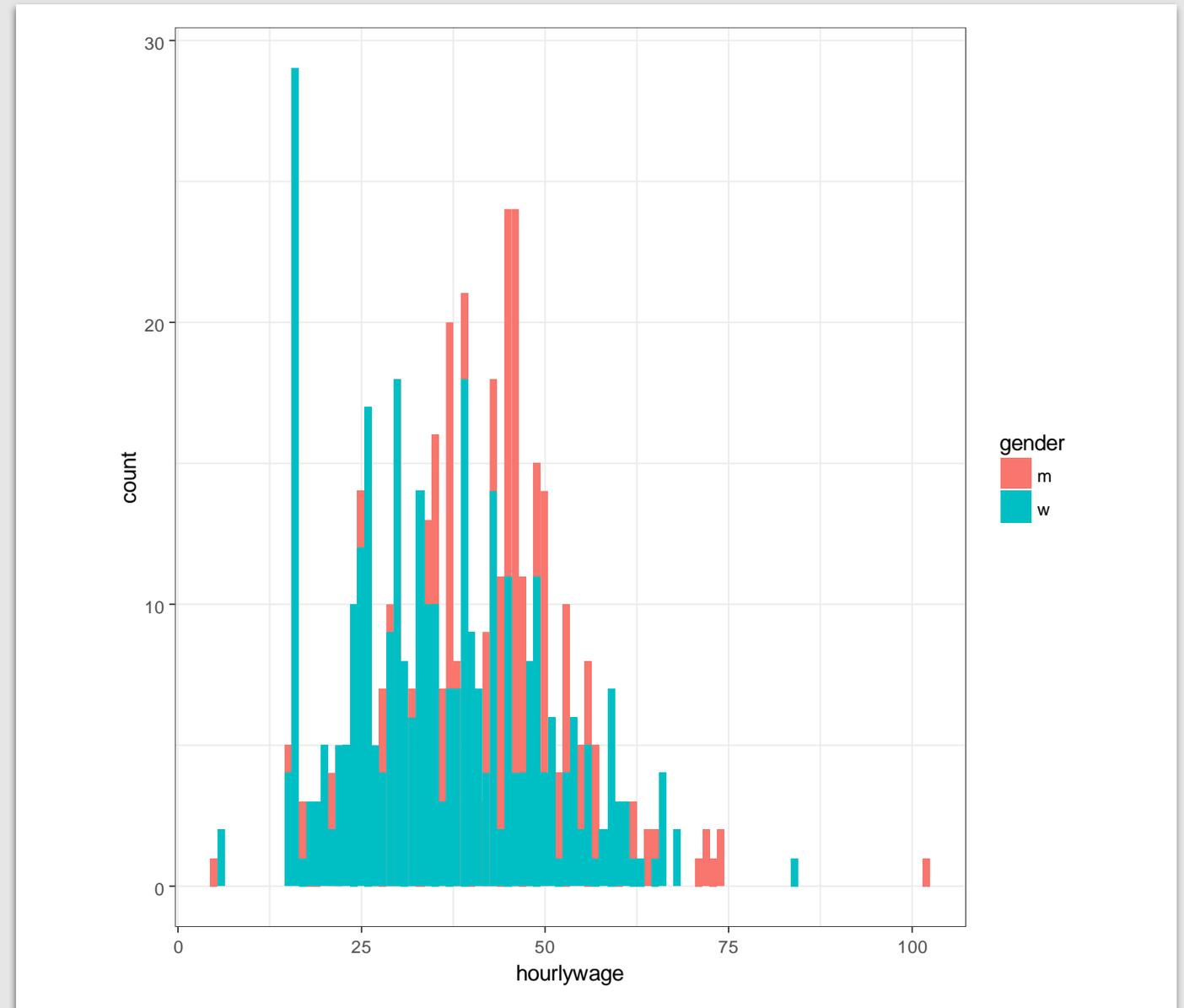




“Lernen” mit Korrelationen |

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?
- Was, wenn sie mehr als \$60/Stunde verdient?



$$X_{1/2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$



$$X^2 + px + q = 0$$



$$X_{1/2} = -\frac{p}{2} \pm \sqrt{\left(\frac{p}{2}\right)^2 - q}$$

Lernen mit Formeln

Am Beispiel der
Bewertung einer
Bewerbung

Wie sagt man die
Leistungsfähigkeit
einer Bewerberin
voraus?



Datengrundlagen

- Data Mining Methoden nutzen, z.B.:
 - Alter
 - Ausbildung
 - Leerzeiten
 - Arbeitgeberwechsel
 - Bewertung des Bewerbungsschreibens
 - Rechtschreibung
 - Ton
 - Wortvielfalt
 - Vielleicht auch Social Media Data
 -
- Wichtig: Beim Trainingsset ist bekannt, ob die Person nachher erfolgreich eingestellt worden ist oder nicht.



Regressionsansätze

- Die Algorithmen-designerinnen und -designer müssen nun entscheiden, welche der Daten vermutlich mit „erfolgreicher Einstellung“ korrelieren.
- Dies sollte am besten in einer einzigen Zahl münden, so dass man direkt sortieren kann.
- Je höher die Zahl, desto höher die Erfolgswahrscheinlichkeit.
- Beispiel Formel:

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

Allgemein

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

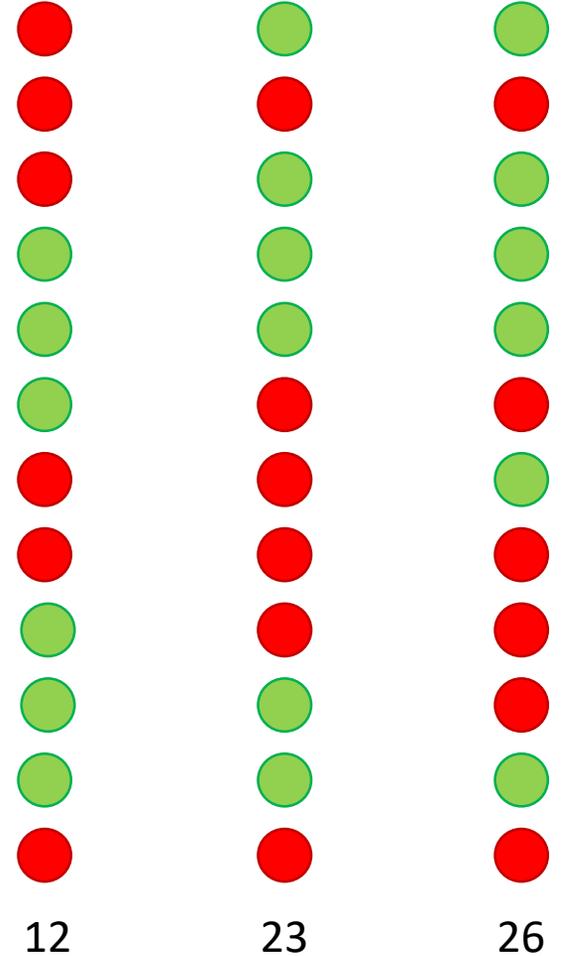
$$\begin{aligned} & w_1 * \text{Erfahrung im Job} \\ - & w_2 * \text{Anzahl Arbeitgeber} \\ + & w_3 * (\text{Wenn Auslandserfahrung, dann 1, sonst 0}) \\ + & w_4 * (\text{Wenn Fortbildung, dann 1, sonst 0}) + \dots \end{aligned}$$



Qualität eines Algorithmus |

„Lernen“ von Gewichten

- Algorithmus probiert Gewichte
- Bewertet jeweils, wie viele erwiesenermaßen erfolgreich Eingestellte möglichst weit oben stehen.
- Die Gewichtung, die das maximiert, wird für weitere Daten genommen.



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.

Ist das sinnvoll?

- ROC AUC vs. Positive Predictive Value
- Wenn die Stelle **sofort** besetzt werden muss, und nur 5 Bewerber da sind: **ja, ROC AUC**
- Wenn es langfristig um die Identifikation der besten Talente geht: **nein, PPV ist besser**



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**)



Können Algorithmen |
diskriminieren? |

Algorithmen...

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

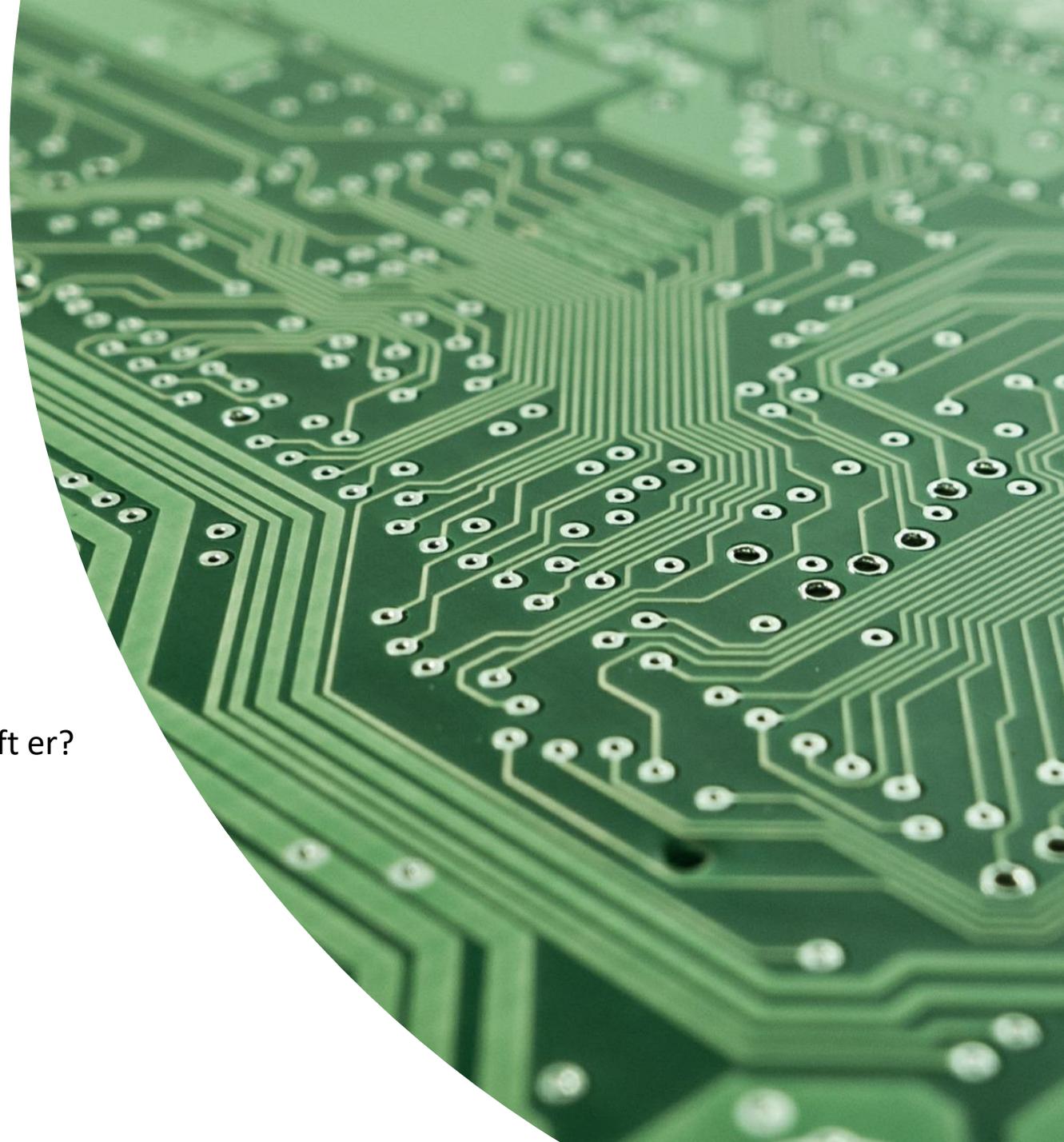
Diskriminierung

- Google zeigt weiblichen Surfern schlechtere Jobs an.
 - Wer ist dafür verantwortlich?
- Rückfälligkeitvorhersagealgorithmen 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“.



Probleme des technischen Systems

- Sind die **Daten** diskriminierungs- und fehlerfrei bzw. gut genug?
- Häufiges Problem: "Entity Recognition".
- Was heißt das:
 - „zu 70% leistungsfähig“?
 - Gehört zu den „30% Förderungswürdigen“?
- **Wie gut ist der HR-Algorithmus**, welche Annahmen trifft er?
- **Ist er vorurteilsfrei?** Garantiert?
- **Bleibt** er vorurteilsfrei? Garantiert?
- Wie geht er mit **Minderheiten** um?
- Wie werden **Ermessensspielräume** umgesetzt?





Sozio-informatische |
Gesamtbetrachtung

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!**
- **Grundlegende Modellierung** und **Datenqualität** kann schlecht sein.
- Manchmal kann ein(e) falsch-negativ Beurteilte(r) **die Vorhersage prinzipiell nicht entkräften!**
 - Z.B. abgelehnte Bewerberin

Kann Diversität gefördert werden?



- Hart vorgegebene Regeln wie „30% internationale Bewerber“: ja
- Automatisch: schwierig
- Algorithmen erkennen **häufige Muster**
- Erkennen nicht **positive** Abweichungen davon.

Einschätzung

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





Was können wir tun?

How to prevent HAL

- Was ist die Lage?
 - Datensammeln über den menschlichen Entscheidungsprozess
- Was soll verbessert werden, woran wird Qualität/Resultat gemessen?
- Welche backlash-Effekte könnte es geben?
- Welche Daten sollen genutzt werden?
- Sind die Entscheidungen menschlich nachvollziehbar?
- Gibt es Ermessensspielräume, die wegfallen?
- Algorithmen explorative nutzen.





Algorithmen in einer demokratischen Gesellschaft

Generell

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

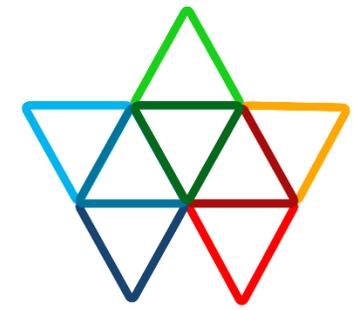
- Automatische Leistungsbewertung
- Kreditvergabe
- Schulische und universitäre Ausbildungen, die durch algorithmische Entscheidungssysteme unterstützt werden
- Gefährder-, Terroristenidentifikation
- ...



Quis custodiet ipsos algorithmos

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

Gründung von „Algorithm Watch“



ALGORITHM
WATCH



Lorena Jaume-Palasi, Mitarbeiterin im iRights.Lab



Lorenz Matzat, Datenjournalist der 1. Stunde, Gründer von lokaler.de, Grimme-Preis-Träger



Matthias Spielkamp, Gründer von iRights.info, ebenfalls Grimme-Preis-Träger, Vorstandsmitglied von Reporter ohne Grenzen.



Prof. Dr. K.A. Zweig, Junior Fellow der Gesellschaft für Informatik, Digitaler Kopf 2014, TU Kaiserslautern

Wie könnte ein „Algorithmen-TÜV“ aussehen?

- Unabhängige Prüfstelle mit Siegelvergabe
- Möglichst auch 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.
 - **Kein weiteres Innovationshemmnis!**
- **Non-Profit**

Beipackzettel für Algorithmen



Welches Problem „kuriert“ der Algorithmus?

Was ist das Einsatzgebiet des Algorithmus, was seine Modellannahmen?

Welche „Nebenwirkungen“ hat der Algorithmus durch seine Einbettung in einen sozialen Prozess?

Schlussformel

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

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/>