



# Risiko und Relevanz von Algorithmen

15.10.2016

Prof. Dr. Katharina A. Zweig

# Das kleine ABC der Informatik



Wann gefährden

**A**lgorithmen,

**B**ig Data und

**C**ünstliche Intelligenz

unsere Demokratie?



# A wie Algorithmus

Ein Algorithmus ist ein Problemlöser

# Problem



**INPUT**

**Der OUTPUT  
der uns sagt,  
wie Input  
mit Output  
zusammenhängt.**



**OUTPUT**

Input: By User:Bluemoose - Own work, [CC BY-SA 3.0](#)

Putput: By Yann (talk) - Own work, GFDL

Output: [CC BY-SA 3.0](#)

# Ein Algorithmus ist...



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



# Beispiele

# Problem: Maximum finden

- Wie finden wir die maximale Zahl im Raum?



# Problem: Sortieren



- Wie sortieren wir alle Ihnen ausgehändigten Zahlen?



# Sortieren 1: Aufsteigendes Sortieren



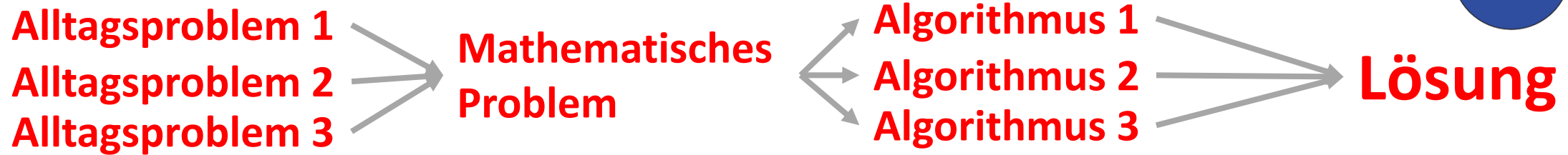
Solange noch Nachbarn nebeneinander stehen, deren Zahlen „falsch herum“ sortiert sind, tauschen sie miteinander.

# Sortieren 2: „Sortieren durch Einfügen“



- Fange mit einer Zahl an und lege sie auf einen (laaaaangen) Tisch
- Solange es noch weitere Zahlen gibt,
  - nimm die nächste,
  - geh am Tisch entlang und sortiere sie an der passenden Stelle ein.
- Alle Zahlen, die schon auf dem Tisch liegen, sind in der richtigen, relativen Reihenfolge.
- Daher: wenn alle auf dem Tisch liegen, sind sie vollständig sortiert.

# Problem-Algorithmus-Problem



- 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;

# Alle Sortierprobleme auf einen Schlag



- Gegeben eine Menge von Objekten oder Subjekten...
- und ein Sortierkriterium, das für je zwei von diesen besagt, welches nach links, welches nach rechts sortiert werden muss,...
- kann jeder beliebige Sortieralgorithmus die korrekte Lösung berechnen.
  
- Die oben genannten Sortieralgorithmen machen **nie** einen Fehler.
- Wir können diesen Algorithmen 100% vertrauen.
- Eine Interpretation der Ergebnisse (dies sind die relevantesten Nachrichten, die wichtigsten Freunde, die kaufenswertesten Produkte) liefert er **nicht**.



**Die Zuordnung einer Frage zu einem  
mathematischen Problem bezeichnet man als  
Modellierung.**



# Komplexe Algorithmen

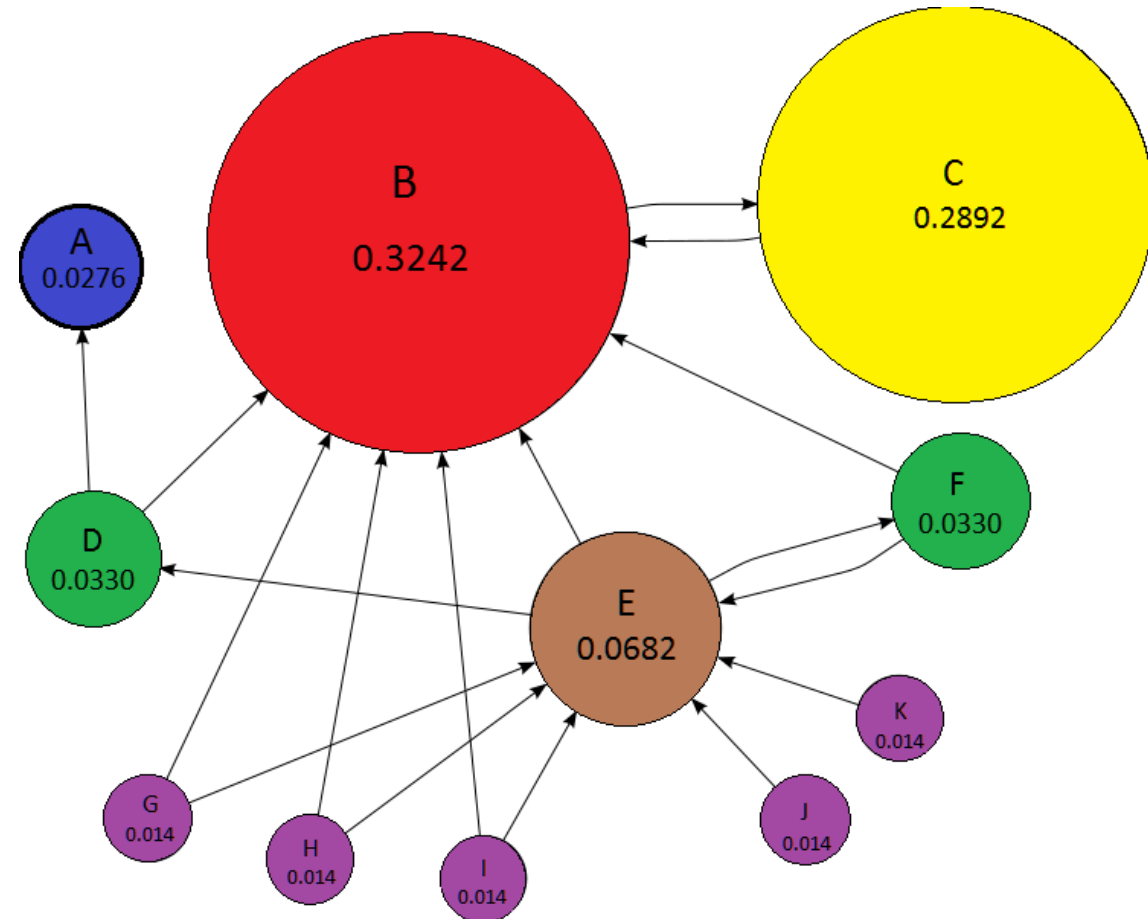
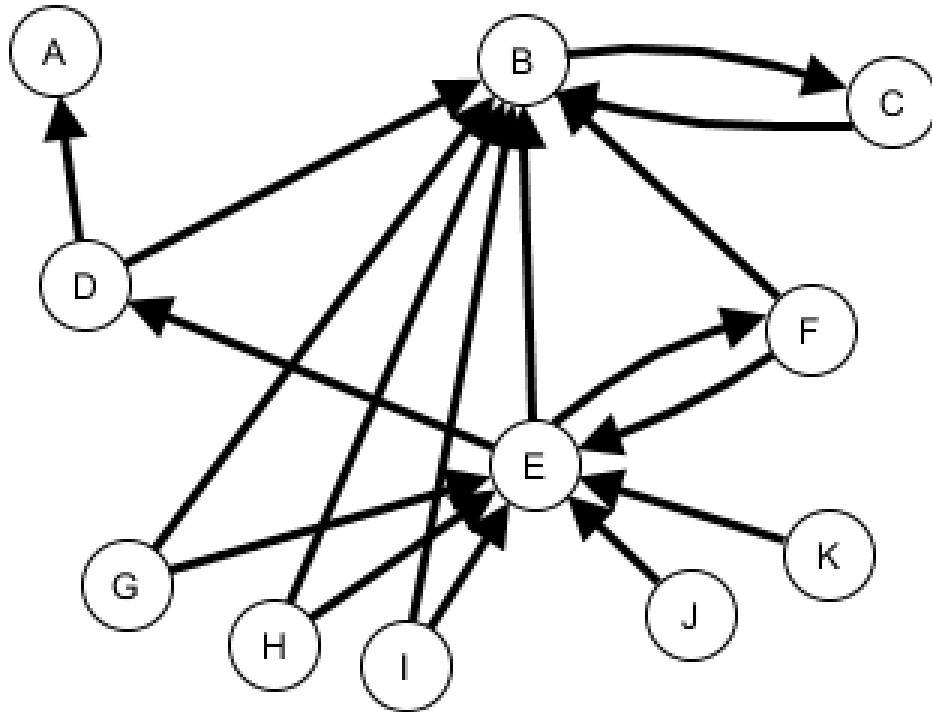
Beispiel: Suchmaschinenalgorithmen

# Suchmaschinen 101



1. Filtern aus allen ihnen bekannten Webseiten diejenigen, deren Text mit den angegebenen Suchbegriffen zusammenhängen.
2. Sortieren diese anhand:
  - Der Vernetzungsstruktur der Seiten untereinander
  - Dem Clickverhalten anderer Nutzer und Nutzerinnen bezüglich derselben Suche
  - Bei Personalisierung: auch nach dem eigenen, bisherigen Suchverhalten

# PageRank





# Idee hinter dem Algorithmus



Ein Modell menschlichen Verhaltens: der Random Surfer

- Ein Surfer klickt auf eine Webseite
- Folgt einem der Links auf der Webseite zufällig
- Von Zeit zu Zeit springt er auf eine völlig neue Webseite
  - Modelliert externes Wissen (z.B. Werbung, bekannte Seiten)



# Wann „stimmt“ dieser Algorithmus?

- Gibt nur dann relevante Ergebnisse, wenn Webseiten
  - Links auf ähnliche Seiten wie ihre eigene setzen,
  - Links auf relevante, meinungsangebende Seiten setzen, und
  - ihre Links **unabhängig** voneinander setzen.
- Unter dieser Bedingung ist der Algorithmus neutral und gibt das kollektive Wissen der Welt nutzbringend weiter.
- Die Veröffentlichung des Algorithmus führte prompt zu Manipulationen seitens der Webseitenbetreiber.
  - Zu große Offenheit der Algorithmen ist manchmal **schädlich**.



Können wir **allen**  
Algorithmen trauen?

# Predictive Policing



Wir haben schon  
auf Sie gewartet!



Vorhersagen,  
wann und wo  
Straftaten  
wahrscheinlich  
sind.

# Predictive Policing



Ein **Algorithmus**  
hat mir geflüstert,  
dass Du **fast** ein Krimineller bist.  
Dann komm mal mit!



Sozio-  
matik

Aber auch: Vorhersagen,  
ob ein Individuum  
straffällig werden könnte!

Beispiel USA:

- 1) Oregon
- 2) Andere Bundesstaaten

# Big Data



- Big Data 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
  - ....
  - Aber nicht: die (in den USA eindeutig zugeordnete) ‚race‘.

# Algorithmus



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

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

# Allgemein



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

- Wer bestimmt die Gewichte so, dass möglichst die einen hohen Wert bekommen, die rückfällig geworden sind?
- Dazu bedarf es Algorithmen der künstlichen Intelligenz.





C wie ... Künstliche Intelligenz

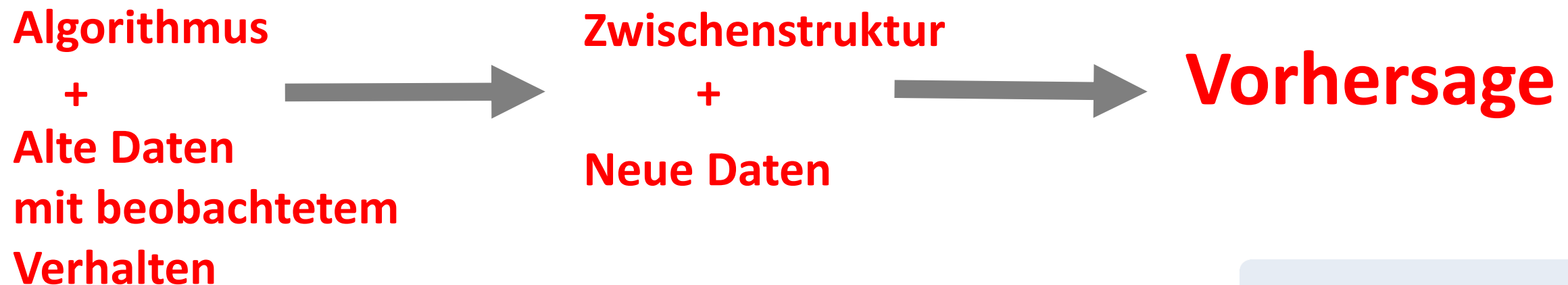
# Lernende Algorithmen





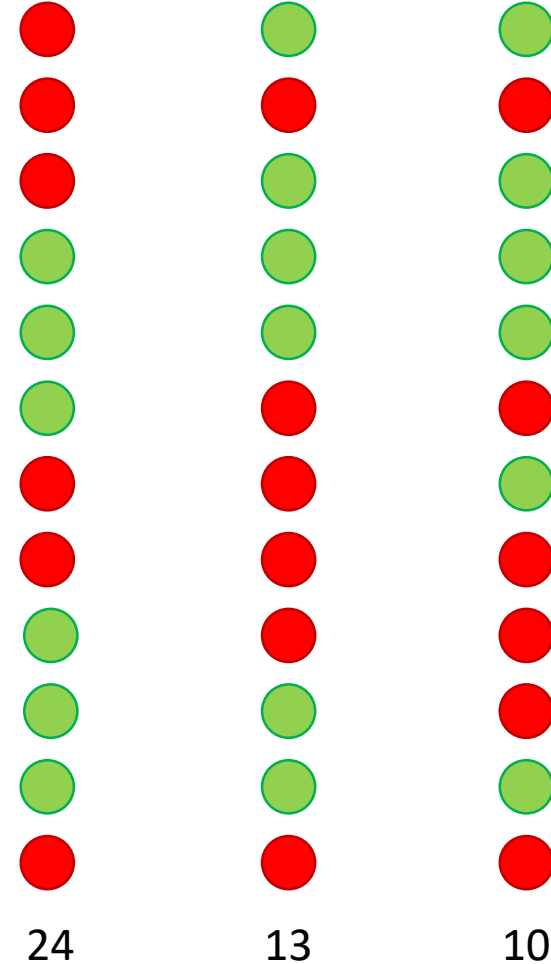
# Künstliche Intelligenz

- **Problem:** gegeben eine Menge von bekannten Daten, finde Muster, die auf neuen Daten vorhersagen, wie sich etwas oder jemand verhalten wird.
- Algorithmus baut – basierend auf bekannten Daten – eine Zwischenstruktur auf, die dann Vorhersagen für neue Daten generiert.
- Der Algorithmus wird „auf den Daten trainiert“.



# „Lernen“ von Gewichten

- Algorithmus probiert Gewichte
- Bewertet jeweils, wie viele bekannte Rückfällige möglichst weit oben stehen – für „alte“ Daten.
- Die Gewichtung, die das maximiert, wird für weitere Daten genommen.
- Kann im Wesentlichen für alles verwendet werden:
  - News Feed bei Facebook
  - Suchmaschinen
  - Produktempfehlung



# Oregon Recidivism Rate Algorithm

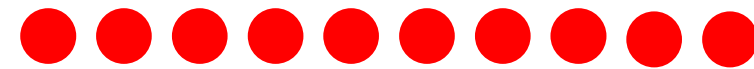
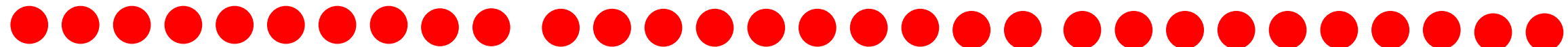
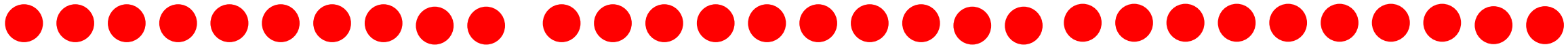
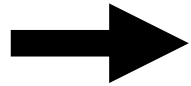


- Das oben genannte Qualitätsmaß dieses Algorithmus: 72 von 100 Paaren werden korrekt sortiert.
- Der in Oregon benutzte Algorithmus hat also, gegeben einen „Rückfall“ und einen „Nichtrückfall“, eine Chance von ca. 1:3 den Rückfall höher zu gewichten als den Nichtrückfall.
- Nur 28% aller so gemachten Prognosen sind falsch!
  - Das klingt doch ganz gut, oder?
- So werden aber keine Urteile gefällt!
- Problem: die Klassen sind ungleich verteilt!
  - 1000 Delinquenten
  - Ca. 200 werden rückfällig

# Optimale Sortierung



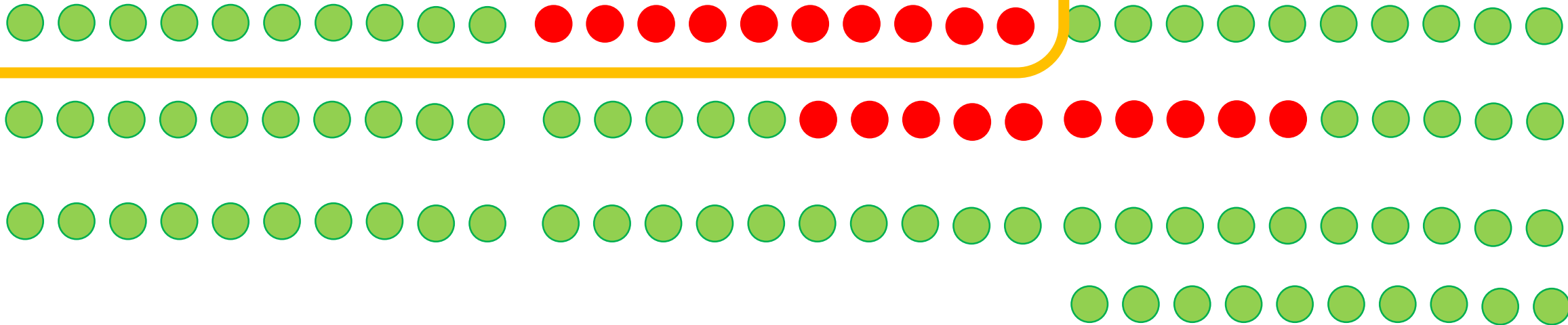
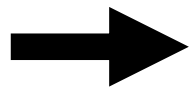
**Erwartete 20% „Rückfällige“**



# Mögliche Sortierung eines Algorithmus mit dieser „Güte“ (ca. 70/100 Paaren)



**Erwartete 20% „Rückfällige“**



# Problem: Unbalancierte Klassen



- Bei optimaler Sortierung: die ersten 200 rot – keine Fehlentscheidung.
- Jetzt: nur die Hälfte!
- Damit **50% Fehlentscheidungen**





# Rückfallvorhersagealgorithmus ist rassistisch (Propublica)



- In einer Studie von Propublica (anderer Algorithmus) war die Quote noch schlechter:
  - Nur 20% der (vorhergesagten) Gewalttäter begingen eine Straftat
  - Bei allen möglichen Straftaten war die Vorhersage etwas besser als ein Münzwurf.
  - Bei schwarzen Mitbürgern war die Vorhersage immer zu pessimistisch;
  - Bei weißen zu optimistisch.
- Northpoint Software ist eine Firma, der Algorithmus ist unbekannt.
- Rasse ist an sich keine Variable des Algorithmus...

# Zweig'sche 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**)

Wenn es **einfache Regeln zur Entscheidungsfindung gäbe, wären sie uns schon bekannt.**

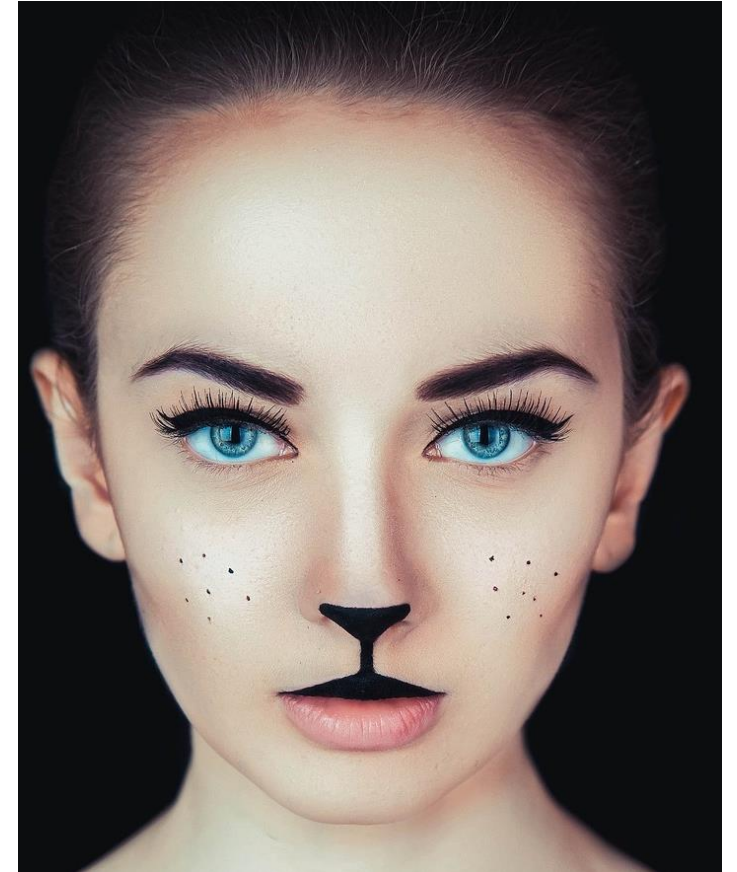


# Statistische Vorhersagen über Menschen

Was bedeutet das eigentlich?

# Zu 70% ein Krimineller....

- Wenn dieser Mensch eine Katze wäre und 7 Leben hätte, würde er in 5 davon wieder rückfällig werden...
- Nein!
- **Algorithmische Sippenhaftung**
  - Von 100 Personen, die „genau so sind wie dieser Mensch“, werden 70 wieder rückfällig;
  - Mitgefangen, mitgehungen;
  - In einer dem Delinquenten (der Delinquentin) völlig unbekanntem, algorithmisch bestimmten „Sippe“.



# Probleme



- Aufmerksamkeitsökonomie der Richter und Richterinnen.
- „Best practice“ erfordert Nutzung der Software.
- Eine Nichtbeachtung der Empfehlung und gleichzeitige Fehleinschätzung wirkt viel schwerer als eine Beachtung der Empfehlung.
- Grundlegende Modellierung und Datenqualität kann schlecht sein.
- Der ins Gefängnis geschickte Delinquent **kann die Vorhersage prinzipiell nicht entkräften!**
  - Dies gilt auch für: Kreditvergaben, Bildungsangebote, Jobs, Personen, die von Drohnen erschossen werden oder als Terrorist eingesperrt werden, ...

# 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		
+ Anchory Selectors	Random Forest		Outgoing	0.18%	1/9.9	5
		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

Windows  
Wechseln  
aktivieren

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

<https://theintercept.com/document/2015/05/08/skynet-courier/>

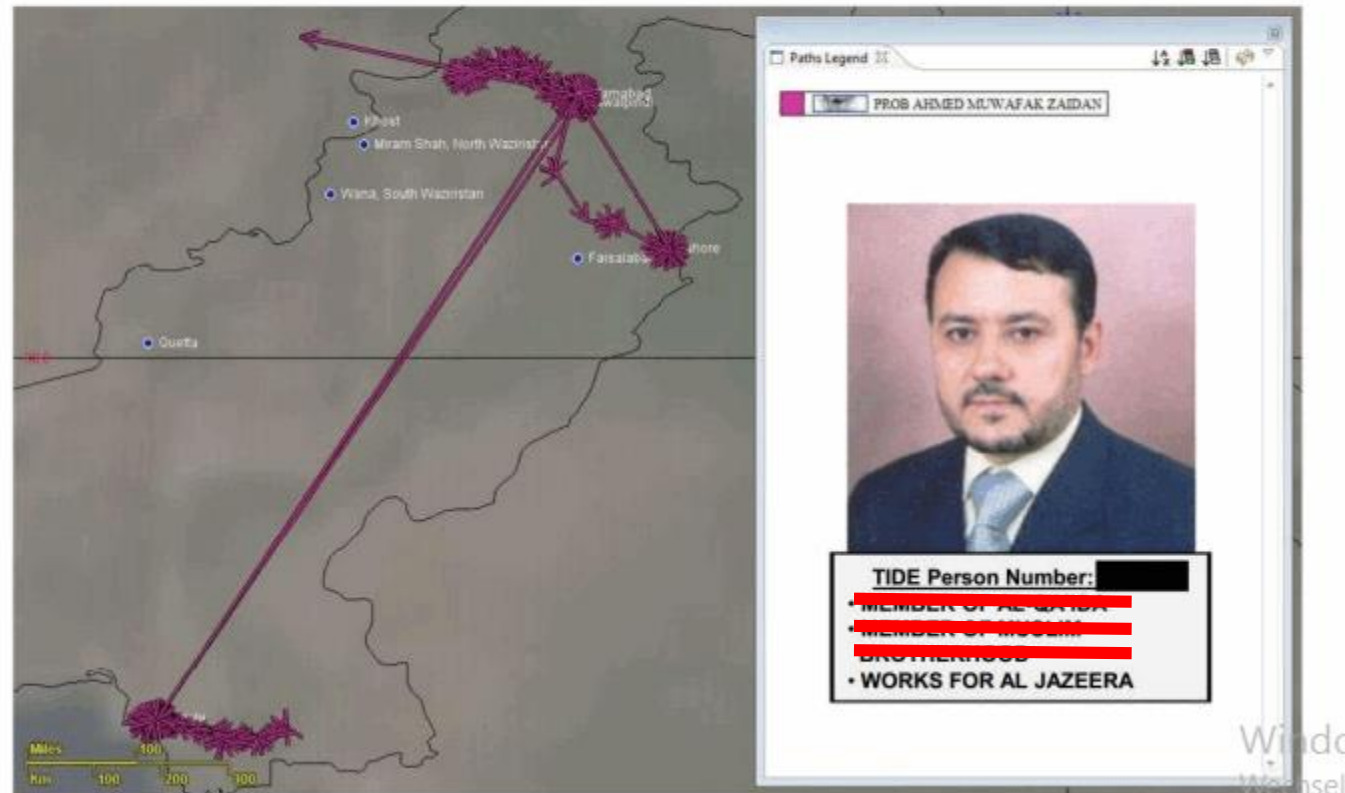
<https://theintercept.com/2015/05/08/u-s-government-designated-prominent-al-jazeera-journalist-al-qaeda-member-put-watch-list/>

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



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

## The highest scoring selector that traveled to Peshawar and Lahore is PROB AHMED Z Aidan



# Spielkampsche Regel



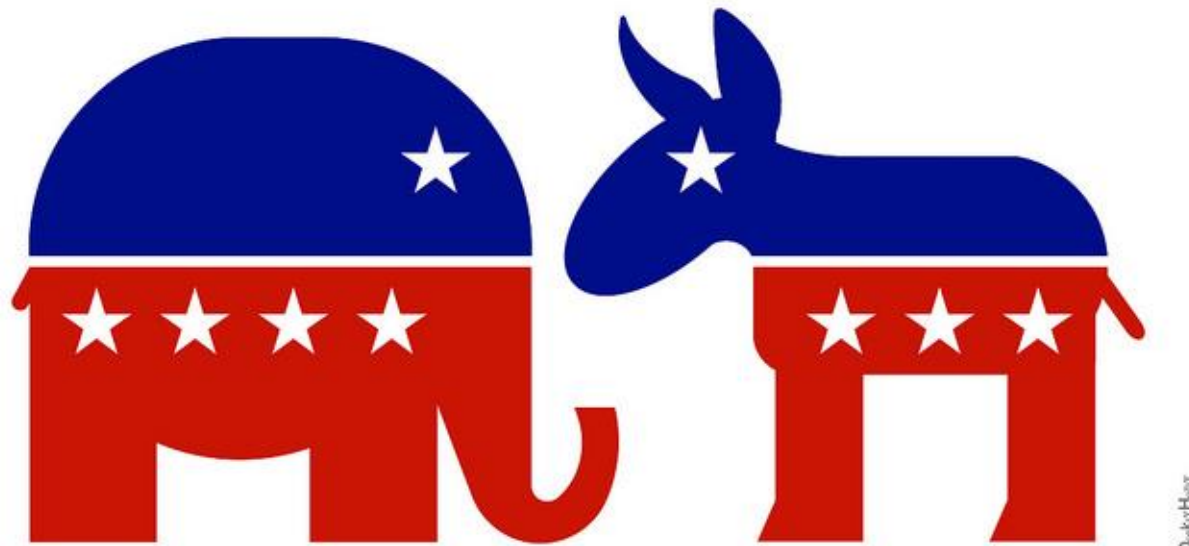
**Alle Algorithmen sind objektiv  
Bis auf die von Menschen gemachten!**





Können uns Algorithmen in  
unserer Meinung beeinflussen?

# Bevorzugt Google Demokraten?



DiederHunt

Studie von Trielli, Mussenden und Diakopolous<sup>1</sup>:

Unter 16 Präsidentschaftskandidaten (USA) gab es bei Demokraten unter den ersten 10 Suchergebnissen 7 positive Berichte, bei Republikanern nur 5,9.

1 <http://algorithmwatch.org/warum-die-google-suchergebnisse-in-den-usa-die-demokraten-bevorzugen/>

# Sind wir beeinflussbar über Algorithmen?



- Suchergebnisreihenfolgen:
  - Manipulierte Suchreihenfolgen werden vom Nutzer nicht bemerkt und können die Tendenz eines unentschlossenen Wähler beeinflussen (Epstein & Robertson, 2015)
- Facebooks „Vote“ bzw. „Ich habe gewählt“-Button
  - Studie von Bond et al. über den Effekt auf das Wahlverhalten.
  - Effekt war klein, aber hochgerechnet ca. 60.000 mehr Wahlstimmen.

Epstein, R. & Robertson, R. E.: "The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections", Proceedings of the National Academy of Science, 2015, E4512-E4521

Bond, R. M.; Fariss, C. J.; Jones, J. J.; Kramer, A. D. I.; Marlow, C.; Settle, J. E. & Fowler, J. H.: "A 61-million-person experiment in social influence and political mobilization", Nature, 2012, 489, 295-298



# „Redirect Method“ by Google Jigsaw

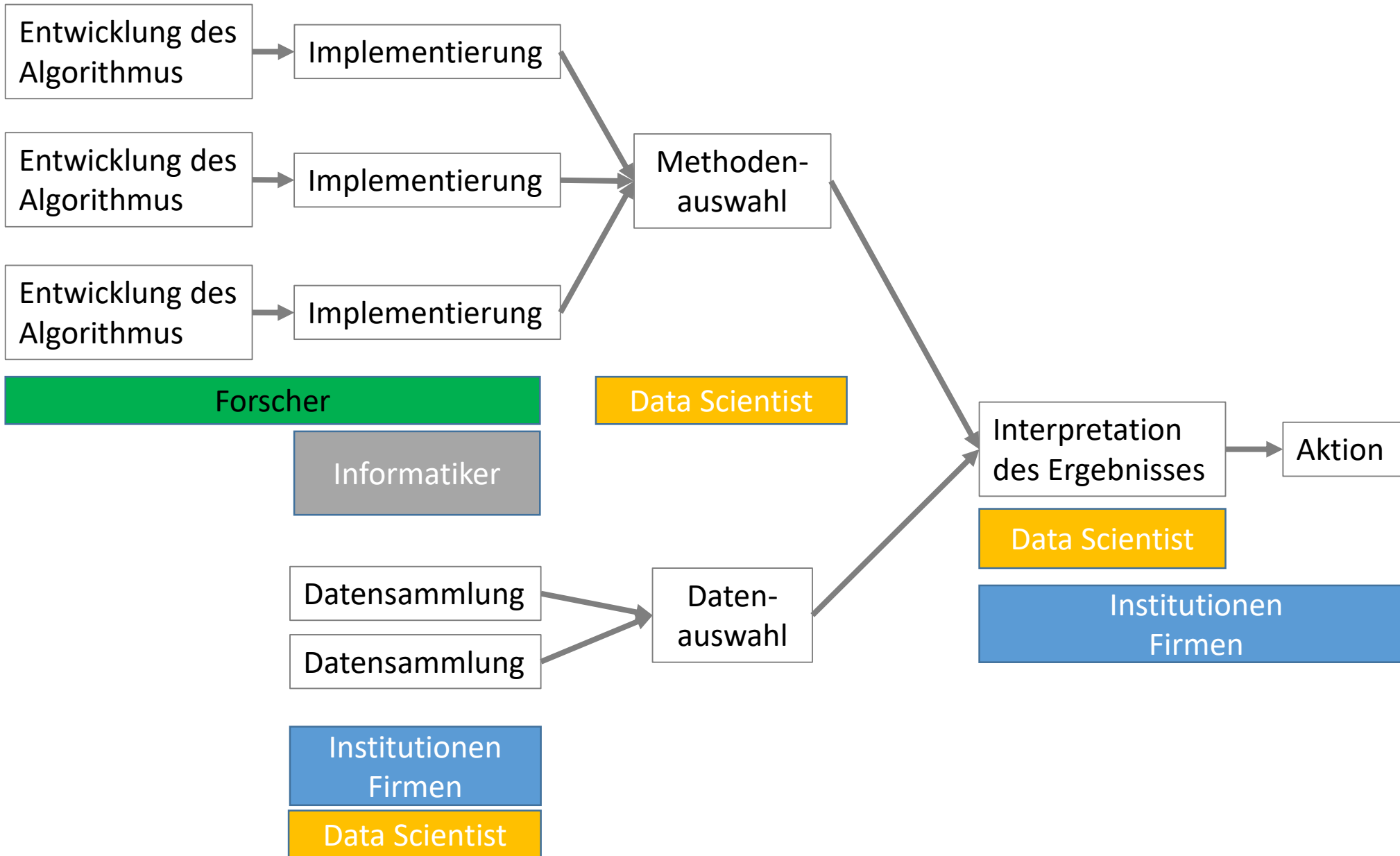
- Könnten wir Suchmaschinen nicht auch „umzudrehen“, die anti-demokratische Leute
- Jigsaw sammelte Anti-ISIS-YouTube-Videos, um zu sehen, welche Leute
- ...identifizierte Suchwörter, die von ISIS-Interessierten stammen,
- ...kreierte eine Werbekampagne für ihren YouTube-Kanal mit dem gesammelten Material
- ...und sah, wann an, wenn die oben genannten Stichworte kamen
- Sie erreichen mehr als 32.000 Interessierte, die sich insgesamt 500.000 Minuten Videomaterial ansahen.

**Ist es das, was wir wollen?**



# Algorithmen in einer demokratischen Gesellschaft

# Verkettete Verantwortlichkeiten



Wer überwacht die Auswirkungen auf die Gesellschaft?

Medien?  
Gesellschaft?  
Politik?  
Institutionen?  
Firmen?  
Recht?



# Quis custodiet ipsos algorithmos

Der „Automated Decision Making“-TÜV vulgo: „Algorithmen TÜV“



# Notwendige Eigenschaften

- 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?

# Schlussformel



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

# Gründung von „Algorithm Watch“



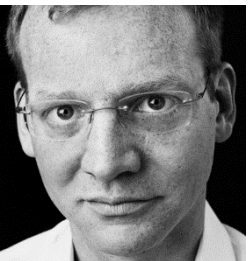
ALGORITHM  
WATCH



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

# Kontakt Daten

Prof. Dr. Katharina A. Zweig

TU Kaiserslautern

Gottlieb-Daimler-Str. 48

67663 Kaiserslautern

[zweig@cs.uni-kl.de](mailto:zweig@cs.uni-kl.de)

Algorithmwatch.org

