



# Medien und Meinungsbildung im Zeitalter der Algorithmen

Informationssitzung Medienrat/Verwaltungsrat der  
Bayerischen Landeszentrale für neue Medien (BLM)

**17.11.2016**

Prof. Dr. Katharina A. Zweig

# Haben Sie auch gezittert?



# Vor der Wahl: „Google manipuliert für Hillary“



- SourceFed behauptet, negative Suchvervollständigungen würden von Google bei Ms Clinton unterdrückt
- Epstein sieht massive Manipulation bei Google und betont, dass bis zu 20% der unentschlossenen Wähler durch Suchmaschinen manipuliert werden können.

# Nach der Wahl: Facebook war's!



- Ach nee.
- Doch nicht.
- Dann war's Facebook:
- Filter bubble und Echo chambers werden algorithmisch erzeugt, so dass wir in einer postfaktischen Welt leben.



# Das kleine ABC der Informatik

## Gefährden

**A**lgorithmen,

**B**ig Data und

**C**ünstliche Intelligenz

unsere Medienvielfalt und Meinungsfreiheit?



# 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.



# Beispiel

# Problem: Sortieren



# Sortieren 1: „Sortieren durch Einfügen“



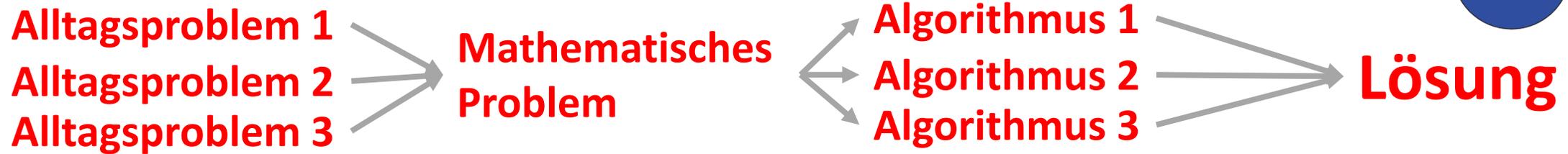
- Hier: Karten mit Zahlen darauf.
- Fange mit einer Karte an, lege sie auf den Tisch.
- Solange es noch unsortierte Karten gibt,
  - nimm die nächste,
  - geh am Tisch entlang und sortiere sie an der passenden Stelle ein.
- Alle Karten, die schon auf dem Tisch liegen, sind in der richtigen, relativen Reihenfolge.
- Daher: wenn alle auf dem Tisch liegen, sind sie vollständig sortiert.

# Sortieren 2: Aufsteigendes Sortieren



- Lege alle Karten in eine Reihe auf den Tisch.
- Gehe den Tisch entlang – wenn dabei zwei Karten in der falschen Reihenfolge nebeneinander liegen, vertausche sie. Tue dies bis zum Ende des Tisches und gehe wieder zum Anfang.
- Laufe solange immer wieder am Tisch entlang, bis im letzten Durchgang kein Tausch mehr nötig war.
- Wenn kein Tausch mehr nötig war, sind alle Karten 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;



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



# Komplexe Algorithmen

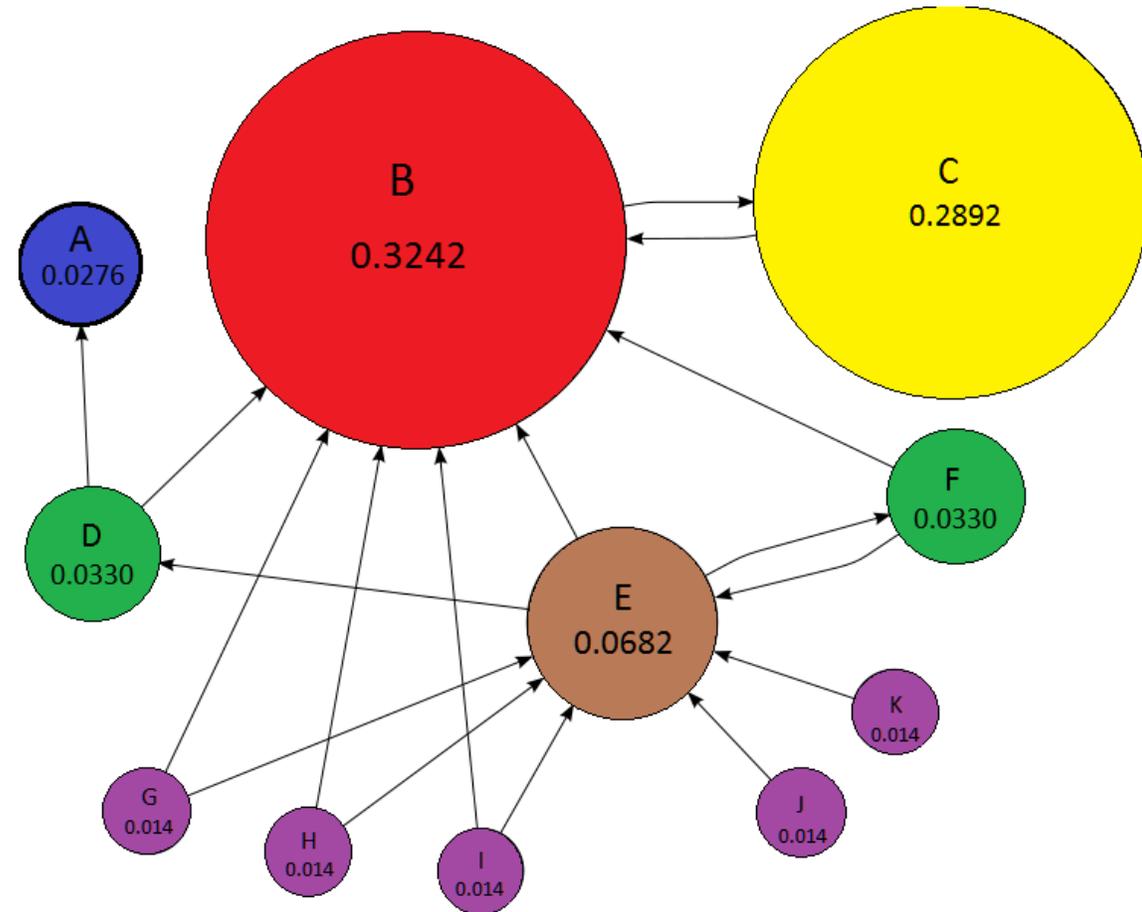
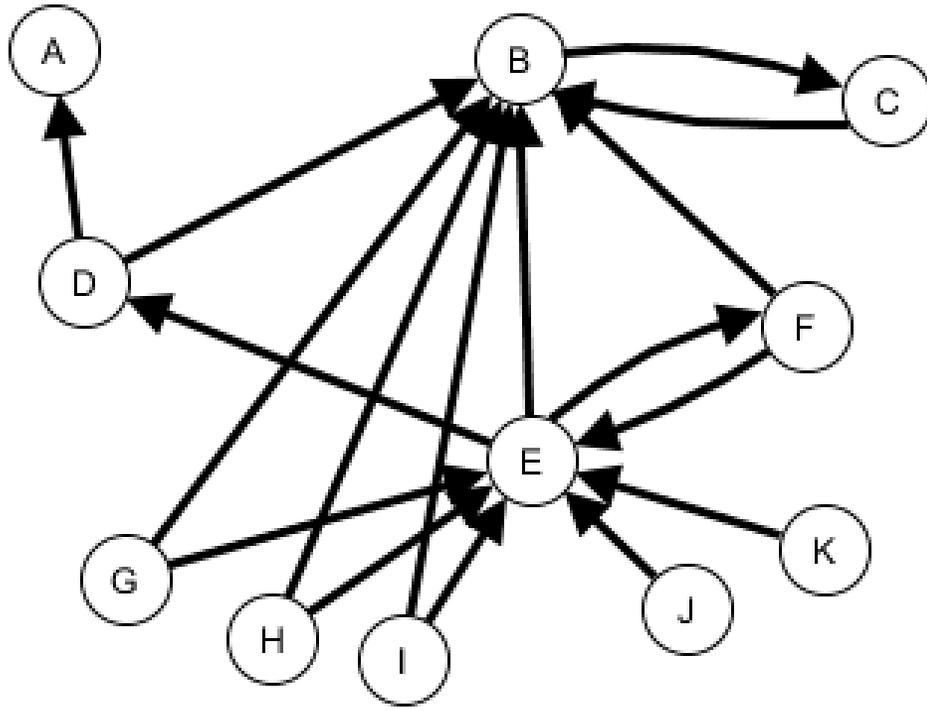
Beispiel: Suchmaschinenalgorithmen

# 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

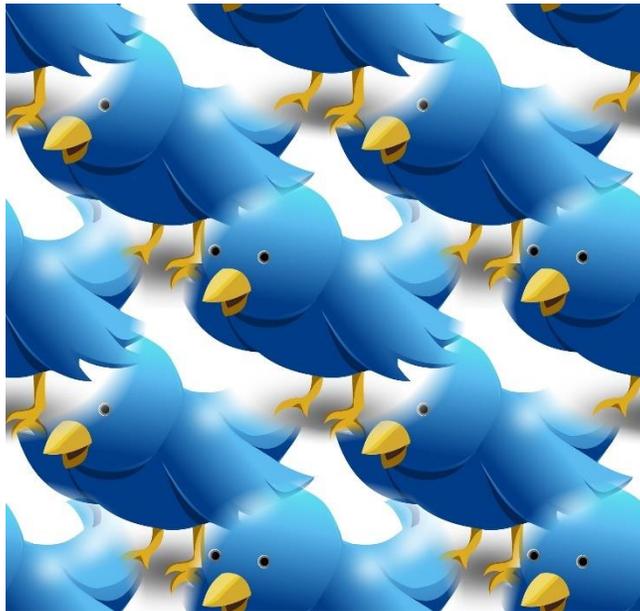


# Modellierungsannahme



- 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**.

# Twitter Netzwerke



- Wer folgt wem auf Twitter?
- Dieselbe Art von Informationen:
  - Wer verlinkt auf wen?
- Relevant:
  - Accounts mit hohen Follower-Zahlen, wie z.B. @RegSprecher Steffen Seibert
    - 525.000 Follower
  - Accounts mit „wichtigen“ Followern:
    - Seibert folgt selbst 100 Accounts, darunter
      - Zeit Online
      - Jan Josef Liefers
      - Barack Obama



# Zusammenfassung Algorithmen

# Algorithmen



- Eingefrorene Handlungsanweisungen, von Menschen erdacht, um mathematische Probleme durch Computer zu lösen.
- Ein mathematisches Problem kann durch verschiedene Algorithmen gelöst werden – per Definitionem kommt dabei immer exakt dasselbe Ergebnis heraus!
- Verschiedene Fragen können durch dasselbe mathematische Problem modelliert werden, z.B.:
  - Was ist die relevanteste Webseite?
  - Wer ist der einflussreichste Twitterer?
- Das Ergebnis **muss interpretiert** werden.





# B wie Big Data

Was heißt hier eigentlich „relevant“?

# Big Data



- Wie kann Relevanz modelliert und „quantifiziert“ werden?
- Big Data Methoden nutzen, z.B.:
  - Sprache der Anfrage, Niveau der Anfrage, Wörter, Wortkombination
  - Tageszeit und geographische Informationen, Gerätetyp
  - Ihre bisherigen Suchanfragen und Ihr persönliches Klickverhalten
  - Welche Seiten wurden angeklickt, wie lange betrachtet, kam die Nutzerin wieder zurück zu den Ergebnissen?
  - Metadaten der Nachrichten/Medien: wann erstellt, durch wen, wo publiziert, Verschlagwortung, Wahl der Wörter
  - Verhalten anderer Nutzer, „ liken “ auf sozialen Netzwerken, Interaktion mit Beiträgen

# Big Data



- Ganz allgemein:
  - Große Datenmengen
  - Außerhalb ihres spezifischen Zwecks genutzt
  - Daher im Einzelnen vermutlich fehlerbehaftet
  - Dank großer Masse und wenig individualisiertem Verhalten statistisch nutzbar



Frage + Big Data =  
mathematisches Problem?

Von der Schwierigkeit der Modellierung

# Relevanz – ein weites Feld



"A squirrel dying in front of your house may be **MORE RELEVANT TO YOUR INTERESTS** right now than people dying in Africa."

Mark Zuckerberg, CEO facebook,  
nach David Kirkpatrick: „the facebook EFFECT“,  
Simon & Schuster New York, New York, USA,  
2010, S. 181



# Big Data + Frage



- Die Algorithmen designerinnen und -designer müssen nun entscheiden, welche der Daten vermutlich mit „Relevanz“ korrelieren.
- Dies sollte am besten in einer einzigen Zahl pro Medium/Nachricht/Webseite münden, so dass man direkt sortieren kann.
- Beispiel Formel:
  - $3 * \text{bisherige Zugriffe}$
  - Anzahl Tage seit Publikation
  - + Beliebtheitsquotient des Publikationsortes
  - + Beliebtheitsquotient des Verfassers + ...

# Allgemein



$$\begin{aligned} & w_1 * \text{ bisherige Zugriffe} \\ + & w_2 * \text{ Anzahl Tage seit Publikation} \\ + & w_3 * \text{ Beliebtheitsquotient des Publikationsortes} \\ + & w_4 * \text{ Beliebtheitsquotient des Verfassers} + \dots \end{aligned}$$

- Wer bestimmt diese Gewichte, so dass insgesamt die „relevantesten“ (also die, die im Nachhinein am öftesten angeklickt werden) am weitesten nach oben sortiert werden?
- 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“.







# Relevanz von Algorithmen

# Neutrale Empfehlungsalgorithmen?

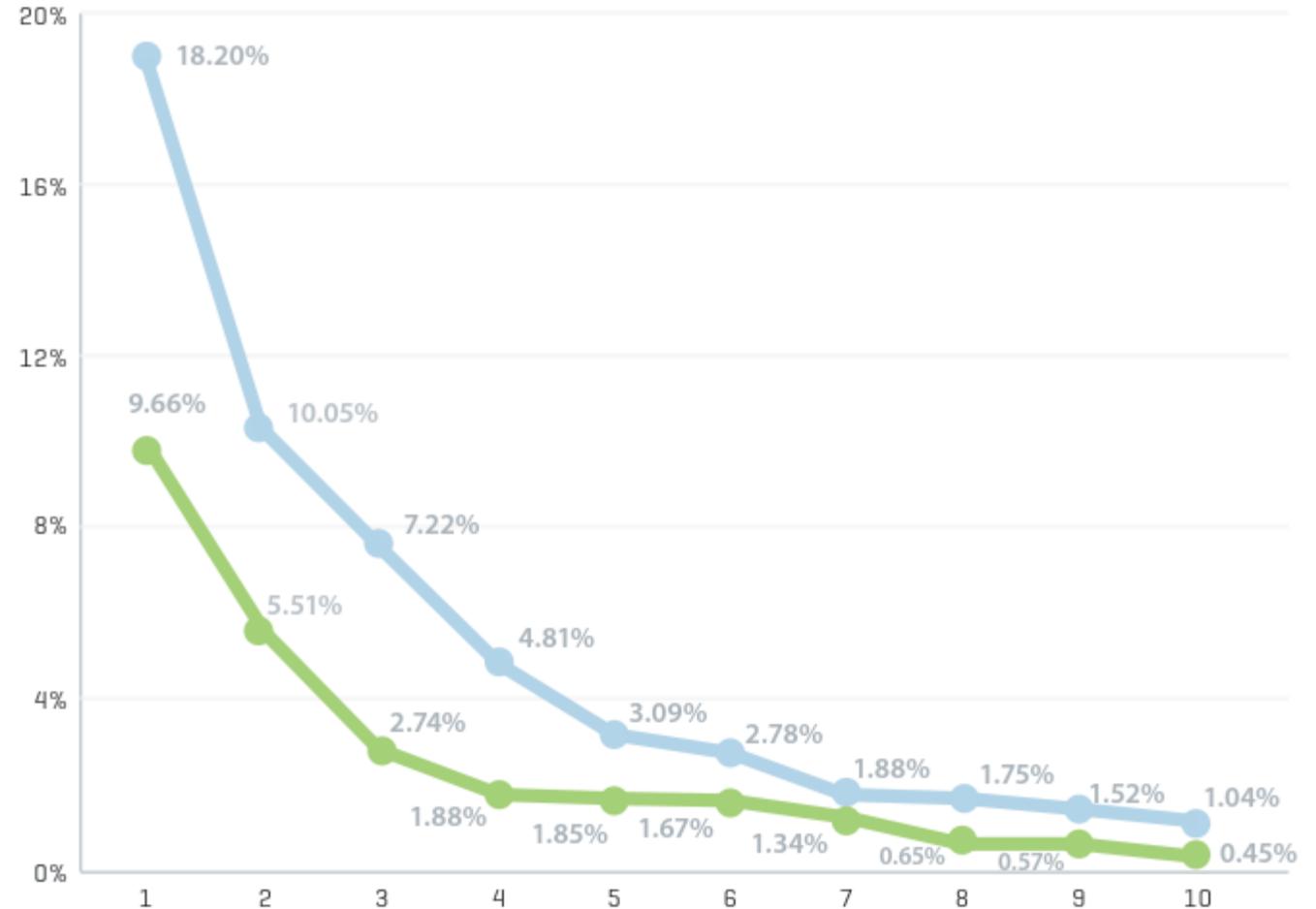


- Empfehlungsalgorithmen filtern, lernen und sortieren.
- Sie machen Modellierungsannahmen, wählen Variablen aus und lernen nur aus einer Teilmenge von Daten.
- All diese Schritte können mehr oder weniger gut gelingen, keiner davon ist neutral im Sinne von „objektiv“.
- Viele Betreiber behaupten, sie würden **nur** nach Relevanz sortieren.

# Vom Wert, der Erste zu sein

- Der erste bekommt ca. 18,2% (9,7%) aller User, der zweite nur noch 10,1% bzw. 5.51.
- 1. Platz fast doppelt so oft wie der 2., 2,5 mal so oft wie der dritte, fast fünfmal so oft wie der 4.

GOOGLE VS BING CLICK-THROUGH RATE



Paul Davison at Digital Relevance™: „A Tale of Two Studies: Establishing Google & Bing Click-Through Rates“, Study by Digital Relevance™ using client data from Jan-June 2011, available from [http://connect.relevance.com/a-tale-of-two-studies-establishing-google-bing-click\\_through-rates](http://connect.relevance.com/a-tale-of-two-studies-establishing-google-bing-click_through-rates) or [research@relevance.com](mailto:research@relevance.com); published 2013.

# Bevorzugt Google Demokraten?



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

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

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

# 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 68-million-person experiment in social influence and political mobilization”, Nature, 2012, 489, 295-298

# Zuccherosconi



Wollte Zuckerberg der jüngste  
Präsident der USA werden –

wer wäre rein technisch noch in  
der Lage, als Gegenkandidat über  
politische Themen zu punkten?

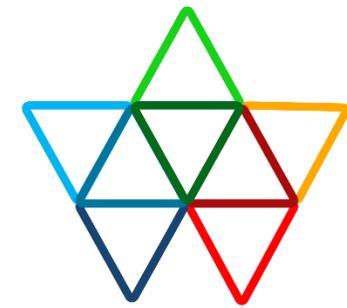




# Algorithmen in einer demokratischen Gesellschaft



# 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

# 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.