

## Einfluss der Datenbasis

Bennet Bleßmann

15. November 2018

1/14

## Einführung

Was ist die Datenbasis?

Die Datenbasis sind die Daten auf denen ein Algorithmus Entscheidungen trifft, sie umfasst vor allem die Daten auf denen ein Algorithmus trainiert, aber auch die Daten welche für die einzelnen Entscheidungen zur Verfügung stehen.

2/14

## Probleme

Zu viele Daten

- ▶ Alte Daten
- ▶ Keine Daten
- ▶ Verzerrte Daten
- ▶ Falsche Daten
- ▶ Nicht Repräsentative Daten
- ▶ Zu wenig Daten
- ▶ Zu viele Daten

3/14

## Lösung

Alte Daten

- ▶ Daten in einem Format Speichern welches nicht Veraltet, z.B. Geburtsjahr statt Alter
- ▶ Regelmäßig neue Daten sammeln und alte Daten löschen

4/14

## Lösung

Keine Daten/Zu wenig Daten

- ▶ Mehr Daten Sammeln
- ▶ Mehr Arten von Daten nutzen/sammeln

5/14

## Lösung

Verzerrte Daten

- ▶ Bei Credit-Scoring nicht nur Daten über negative Ereignisse sammeln sondern auch über positive Ereignisse

6/14

---

## Lösung

Falsche Daten

- ▶ Einsicht\Transparenz und Korrektur zulassen
- ▶ Mit anderen Quellen vergleichen
- ▶ Fehlerquellen minimieren

---

7/14

---

## Lösung

Nicht repräsentative Daten

- ▶ Statt allgemeine Daten zu nutzen, speziell für einen Zweck Daten Sammeln
- ▶ Maßnahmen gegen fehlende, zu wenige und verzerrte Daten

---

8/14

---

## Lösung

Zu viele Daten

---

9/14

---

## Ethik

---

10/14

---

## Diskussion

Was wäre wenn?

Angenommen die Zulassung und die Auswahl des Studienganges würde von einem Algorithmus abhängen würde.  
Wer von den Anwesenden könnte sich denken heute nicht hier zu sein (und warum)?

---

11/14

---

## Diskussion

Alltags Beispiele

Beispiel: Navi passt aufgrund einer *während der fahrt* mitgeteilten Baustelle die Strecke automatisch an und Informiert über die Anpassung.

---

12/14

---

## Diskussion

### Szenario

Einführung Im Rahmen der Vorrats Datenspeicherung sind sie ein vom BND eingesetzter Algorithmus zur Evaluierung der eingehenden Daten. Ihre Aufgabe ist es basierend auf den Vorhandenen Daten z.B. Einsätze zu initiieren, Nichts zu unternehmen oder an einen Menschen weiterzuleiten.

- ▶ Besuchen von vielen Seiten zum Thema **ISIS**
- ▶ Besuchen von vielen Seiten zu **Explosiven Gemischen**
- ▶ Kauf einer Angelweste
- ▶ Kauf von Rohren
- ▶ Kauf verschiedener Haushaltsmittel
- ▶ Last-Minute Kauf von Karten für einen großen Kongress

---

## References

- (2017). Special Issue Informatik Spektrum , 40(4) . Retrieved from <https://link.springer.com/journal/287/40/4/page/1>
- ACM US Public Policy Council. (12 January 2017). Statement on Algorithmic Transparency and Accountability. Retrieved from [https://www.acm.org/binaries/content/assets/public-policy/2017\\_usacm\\_statement\\_algorithms.pdf](https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf)
- Baeza-Yates, R. (June 2018). Bias on the Web. *Communications of the ACM*, 61(6), 54-61. doi:10.1145/3209581 Retrieved from <https://cacm.acm.org/magazines/2018/6/228035-bias-on-the-web/fulltext>
- Executive Office of the President. (May 2016). Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights. Retrieved from [https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016\\_0504\\_data\\_discrimination.pdf](https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf)
- Maynes, R., & Everdell, I. (16 May 2018). The evolution of Google search results pages & their effect on user behaviour. *Mediative* . Retrieved from <http://www.mediative.com/whitepaper-the-evolution-of-googles-search-results-pages-effects-on-user-behaviour/>
- Mercer, A., Deane, C., & McGeeney, K. (9 November 2016). Why 2016 election polls missed their mark. *Pew Research Center* . Retrieved from <http://www.pewresearch.org/fact-tank/2016/11/09/why-2016-election-polls-missed-their-mark/>
- Oltenu, A., Castillo, C., Diaz, F., & Kiciman, E. (20 December 2016). Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. Retrieved from <https://dx.doi.org/10.2139/ssrn.2886526>
- Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., ... Nosek, B. A. (25 August 2018). Many analysts, one dataset: Making transparent how variations in analytical choices affect results. *Pew Research Center* . doi:10.31234/osf.io/qlkvst Retrieved from <https://psyarxiv.com/qlkvst/>
- Vincent, J. (24 March 2018). Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day. *The Verge* . Retrieved from <https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist>