

Wenn Algorithmen irren lernen

pascoda

C&3W und TU Wien

privacyweek 2018

Motivation

Algorithmen übernehmen Entscheidungen in mehr & mehr Kontexten im täglichen Leben

Objektive Algorithmen?

Menschen glauben, Algorithmen wären objektiv.

Nope.

Algorithmus

... eindeutige Handlungsvorschrift zur Lösung eines Problems

... endlich viele, wohldefinierte Einzelschritte

Bsp.: Kochrezept, Sortieralgorithmus, Suchalgorithmus, Arbeitsabläufe

<https://de.wikipedia.org/wiki/Algorithmus>

Algorithmisches System

- ▶ Suchmaschinen
- ▶ Einstufungssysteme
- ▶ Machine Learning Systeme

Bühne frei: Bias!

- ▶ Problem-/Fragen-/Zieldefinition
- ▶ Daten
- ▶ Modell
- ▶ Visualisierung

Alles nicht Wert-frei!

Problem, Frage, Ziel

- ▶ welche Aufgabe soll gelöst werden?
- ▶ wessen Anwendungsfall ist es?
- ▶ wer steht im Zentrum?

Subjekte vs. Objekte; Entscheidende vs. Anwendende

Problem, Frage, Ziel

Bad News: Das Leben besteht aus lauter schlecht definierten Situationen, "wicked Problems" - und Menschen.

Problem, Frage, Ziel

Beispiele:

- ▶ Essens- und Küchenmanagement-System (Cathy O'Neil)
- ▶ AMS-System

Problem, Frage, Ziel

was für Entscheidungen, über wen, auf welcher Basis,
werden getroffen?

Problem, Frage, Ziel

- ▶ AMS-Daten nutzen, um Unterstützungsangebote zu planen
- vs.
- ▶ AMS-Daten nutzen, um Einzelpersonen zu beurteilen

und: was macht so eine AMS-Einstufung mit der eingestuften Person?

Bühne frei: Bias!

- ▶ Problem-/Fragen-/Zieldefinition
- ▶ **Daten**
- ▶ Modell
- ▶ Visualisierung

Alles nicht Wert-frei!

Daten

Neuerfassung vs. (historische) Bestandsdaten

- von wem, wo, was, wann, wie, ...?

Daten

Beispiele:

- ▶ unausgewogene Lernsets
- ▶ AMS-System

Bühne frei: Bias!

- ▶ Problem-/Fragen-/Zieldefinition
- ▶ Daten
- ▶ **Modell**
- ▶ Visualisierung

Alles nicht Wert-frei!

Modell, Aus- und Bewertung

Modelle reduzieren **immer**. Was bleibt, was "kann weg"?

- ▶ Modell vom Mensch an das System: AMS
- ▶ Modell vom System: bringt sich selbst Muster bei

Bühne frei: Bias!

- ▶ Problem-/Fragen-/Zieldefinition
- ▶ Daten
- ▶ Modell
- ▶ **Visualisierung**

Alles nicht Wert-frei!

Visualisierung

Suchergebnisse: über 50 % der Buchungen beim ersten Ergebnis, 92 % auf der ersten Seite (SABRE, 1982)

Berendt & Preibusch, 2017:
Entscheidungsprozesse/-parameter sichtbar machen.

What now?

sich bewusst sein: Algorithmen sind da.

- ▶ ausweichen, etwas anderes benutzen?
- ▶ benutzen & abwägen?
- ▶ "hacken"?
- ▶ low-tech Lösungen?
- ▶ irgendwie selbst verbessern?

Danke!

twitter: @pascoda

mastodon: @pascoda@chaos.social

mail: pascoda@posteo.net

slides: files.fairydust.space/pw18-algbias.pdf

c<3W und Technische Universität Wien

Leseempfehlungen

- ▶ Sara Wachter-Boettcher (2017). *Technically Wrong: Sexist Apps, Biased Algorithms, and Other Threats of Toxic Tech*
- ▶ Cathy O'Neil (2016). *Weapons of Math Destruction*
- ▶ Bettina Berendt, Sören Preibusch (2017). *Toward Accountable Discrimination-Aware Data Mining: The Importance of Keeping the Human in the Loop — and Under the Looking Glass*
- ▶ <https://www.sueddeutsche.de/digital/kuenstliche-intelligenz-computer-erklaer-mir-was-du-tu-3922868>