



# **Datengenerator für Daten mit Bias als Grundlage für Data Science Projekte**

**Studienarbeit**

für die Prüfung zum  
**Bachelor of Science**

des Studiengangs Informatik  
an der Dualen Hochschule Baden-Württemberg Stuttgart

von  
**Simon Jess, Timo Zaoral**

Juni 2022

**Bearbeitungszeitraum**  
**Matrikelnummer, Kurs**  
**Ausbildungsfirma**  
**Betreuer**

04.10.2021 - 10.06.2022  
8268544, 6146532, INF19C  
TRUMPF GmbH & Co KG, Ditzingen  
Prof. Dr. Monika Kochanowski

# Erklärung

Wir versichern hiermit, dass wir die vorliegende Studienarbeit mit dem Thema: *Datengenerator für Daten mit Bias als Grundlage für Data Science Projekte* selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt haben. Wir versichern zudem, dass die eingereichte elektronische Fassung mit der gedruckten Fassung übereinstimmt.

Stuttgart, Juni 2022

---

Vorname Nachname

---

Vorname Nachname

## **Abstract**

Fasst die Aufgabenstellung und Ergebnisse kompakt und übersichtlich in wenigen Zeilen zusammen (4-7 Zeilen).

# Inhaltsverzeichnis

Abkürzungsverzeichnis . . . . .	V
Abbildungsverzeichnis . . . . .	VI
Tabellenverzeichnis . . . . .	VII
<b>1 Einleitung</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Zielsetzung . . . . .	1
1.3 Aufbau der Arbeit . . . . .	2
<b>2 Stand der Technik</b>	<b>3</b>
2.1 Daten . . . . .	3
2.1.1 Datenqualität . . . . .	3
2.1.2 Bias . . . . .	3
2.2 Künstliche Intelligenz & Maschinelles Lernen . . . . .	3
2.2.1 Was ist Künstliche Intelligenz . . . . .	3
2.2.2 Teilgebiet Maschinelles Lernen . . . . .	3
2.2.3 Ethik in der Künstlichen Intelligenz . . . . .	3
2.3 Künstliche Intelligenz verbunden mit Bias . . . . .	3
2.3.1 Diskriminierung durch verzerrte Daten . . . . .	3
2.3.2 Gegenmaßnahmen . . . . .	3
<b>3 Praktischer Teil</b>	<b>4</b>
3.1 Szenarien . . . . .	4
3.1.1 Szenario 1 . . . . .	4
3.1.2 Szenario 2 . . . . .	4

3.2	Konzeption . . . . .	4
3.2.1	Grobkonzet . . . . .	4
3.2.2	Feinkonzept . . . . .	4
3.3	Umsetzung . . . . .	4
<b>4</b>	<b>Schluss</b>	<b>5</b>
4.1	Zusammenfassung . . . . .	5
4.2	Evaluierung . . . . .	5
4.3	Ausblick . . . . .	5

# Abkürzungsverzeichnis

**KI**      Künstliche Intelligenz

# Abbildungsverzeichnis

# Tabellenverzeichnis



# 1 | Einleitung

Daten sind die

Alter Text:

Das Arbeiten mit Daten wird immer relevanter und so werden auch in den Hochschulen und Universitäten immer mehr Vorlesungen im Bezug auf das Arbeiten mit Daten gehalten. Durch zunehmende Internetnutzung und neue Möglichkeiten Daten zu sammeln und zu speichern, wächst die Datenmenge stetig weiter an. Daten allein sind jedoch relativ wertlos, was sie wertvoll macht, ist die Arbeit mit den Daten. Bei dieser Arbeit handelt es sich in den meisten Fällen um die Entwicklung von Maschinellen Lernmodellen oder das Konzipieren und Trainieren einer Künstlichen Intelligenz (KI), die durch die gesammelten Daten trainiert werden kann. Durch die Verwendung der Daten als Lerndaten für eine KI, können Entscheidungen aber auch Vorhersagen auf Basis der in den Lerndaten enthaltenen Zusammenhänge getroffen. Da sich letztendlich alles um Zusammenhänge in Daten dreht, ist auch ein wichtiges Themenfeld im Bereich der KI die Ethik. Da Fragen gestellt werden wie "Inwieweit darf die KI Entscheidungen aufgrund von Datengrundlagen treffen". [1]

## 1.1 Motivation

- Anonymisierung/Pseudonymisierung bei besonders großen Datensätzen ist schwierig
- Nachvollziehbarkeit von Bias verzerrten Daten
- Veranschaulichung von Bias in Daten für die Allgemeinheit, um auf das Problem im Bereich ML aufmerksam zu machen

## 1.2 Zielsetzung

- Datengenerator für Bias verzerrte Daten
- Visualisierung von Bias in Lerndaten für ML
- Gesamt Produkt zur Erstellung von Daten und deren Bias Visualisierung für die Lehre
- 1 Satz, was sollen wir machen -> Stichwortliste mit Anforderungen
- Maschinelles Lernen hängt von den Trainingsdaten ab.
- Trainingsdaten können einen Bias Data enthalten.

## 1.3 Aufbau der Arbeit

## **2 | Stand der Technik**

### **2.1 Daten**

#### **2.1.1 Datenqualität**

#### **2.1.2 Bias**

- Begriffserklärung: Data Bias vs Bias Verzerrung (zu viel/zu wenig lernen im ml)
  - Arten von Bias:
    - Bias durch Abwesenheit - Wenn eine Info fehlt, kann das zu Diskriminierung führen.
    - Diskriminierung durch Menschen.
- Arten von Bias: Cognitive, Social, Perceptual und Motivational Bias [2]

### **2.2 Künstliche Intelligenz & Maschinelles Lernen**

#### **2.2.1 Was ist Künstliche Intelligenz**

#### **2.2.2 Teilgebiet Maschinelles Lernen**

- Supervised learning
- Unsupervised learning

#### **2.2.3 Ethik in der Künstlichen Intelligenz**

### **2.3 Künstliche Intelligenz verbunden mit Bias**

#### **2.3.1 Diskriminierung durch verzerrte Daten**

#### **2.3.2 Gegenmaßnahmen**

- Wenn der Parameter mit dem Bias entfernt wird, wird das Ergebnis erstmal schlechter.

## **3 | Praktischer Teil**

### **3.1 Szenarien**

#### **3.1.1 Szenario 1**

#### **3.1.2 Szenario 2**

### **3.2 Konzeption**

#### **3.2.1 Grobkonzept**

#### **3.2.2 Feinkonzept**

### **3.3 Umsetzung**

## 4 | Schluss

### 4.1 Zusammenfassung

- Fazit ziehen!!!

### 4.2 Evaluierung

### 4.3 Ausblick

## Literaturverzeichnis

- [1] T. Dullien, “Maschinelles lernen und künstliche intelligenz in der informationssicherheit,” *Datenschutz und Datensicherheit-DuD*, vol. 42, no. 10, pp. 618–622, 2018.
- [2] P. A., A. Jawaid, S. Dev, and V. M.S., “The patterns that dont exist : Study on the effects of psychological human biases in data analysis and decision making,” in *2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*, pp. 193–197, 2018.