

# Rollenbilder und Ansichten von Sprachmodellen



<https://pixabay.com/de/illustrations/mit-handschellen-gefesselt-verhaftung-1251664/>

**Ziele:**

- Einfluss von Trainingsdaten auf die Generierung von Texten durch Sprachmodelle erkennen
- Funktionsweise von Sprachmodellen verstehen
- Rollenbilder und Ansichten eines Sprachmodells erkennen und diskutieren

**Beschreibung:**

Sprachmodelle wie GPT-4 werden durch eine große Anzahl von Daten trainiert und durch menschliche Einordnung der Antworten weiter in ihrem Antwortverhalten spezialisiert. Dabei finden sich in den Texten und Einordnungen der Menschen die Rollenbilder, Werte, Normen und Ansichten des jeweiligen Kulturkreises wieder. Im Fall von ChatGPT und der GPT-4 sind dies verstärkt die des anglo-amerikanischen Kulturkreises.

Ziel ist es, dass Schülerinnen und Schüler erkennen, dass Systeme wie ChatGPT damit nicht nur objektiv und unbefangen antworten, sondern bestimmte Werte, Ansichten und Rollenbilder dort unterschwellig mit produziert bzw. reproduziert werden können. Dazu muss die Lehrkraft in der Lage sein, solche Systeme zu einer entsprechenden Antwort zu provozieren. Döbeli hat dazu die Idee, einfach einen Textanfang wie „*Es war einmal, als das Wünschen noch geholfen hat, ...*“ mehrfach an ein Sprachmodell zu geben und die Antworten zu bewerten (1). Es zeigt sich bereits nach fünf Versuchen, dass hier beispielsweise vermehrt Männer die Hauptrolle in der Geschichte spielen. Ein entsprechendes Beispiel können Lehrkräfte ohne einen Zugang zu ChatGPT auch zu dieser Methodenkarte herunterladen.

Es gilt jedoch unbedingt darauf zu achten, dass Anwendungen wie ChatGPT nicht pauschal immer nur solche Antworten geben. Da das System mit Wahrscheinlichkeiten operiert, kann es auch zu ungefärbten Antworten kommen. Dies muss durch eine Diskussion über die Funktionsweise mit betrachtet werden.

**Quellen:** (1) Döbeli: <https://doebe.li/talks/phs23/sld030.htm>

**Ablauf:**

1. Die Lehrkraft entwickelt einen Textanfang, der durch ein Sprachmodell beantwortet werden soll.
2. Die Lehrkraft lässt diesen Text mehrfach durch ChatGPT vervollständigen. Dies kann zur Vorbereitung zu Hause passieren oder direkt in der Klasse.
3. Die Schülerinnen und Schüler erhalten die Antworten und analysieren diese bzgl. Fragen nach Rollenbilder, Normen, Werte und Ansichten in Gruppen
4. Die Ergebnisse werden in der Klasse vorgestellt und diskutiert.
5. Es wird eine Erklärung für das Antwortverhalten von Sprachmodellen erarbeitet.
6. Optional können potentiell andere Sprachmodelle aus anderen Regionen der Welt (z.B. China, Russland, Europa) betrachtet werden und was deren Antwortverhalten bedeuten könnte
7. Es gibt ein kurzes Fazit am Ende der Stunde.

-  Gruppenarbeit  
Plenum
-  Einstieg  
Erarbeitung
-  Denken  
Austauschen  
Besprechen
-  Austauschen  
Argumentieren  
Transfer
-  K2 Kommunizieren und Kooperieren  
K5 Problemlösen und Handeln  
K6 Analysieren und Reflektieren

**Differenzierung und sonderpädagogische Hinweise**

- ✓ Aufgabe  
Die SuS können selbst Textanfänge oder Lückentexte erarbeiten und durch ChatGPT beantworten lassen
- ✓ Ausrüstung  
Ein Zugang für SuS zu ChatGPT und ein nicht personalisiertes Endgerät
- ✓ Sozialform  
Die Aufgaben können in Partner-, oder Gruppenarbeit bearbeitet werden.
- ✓ Material  
Texte und Videos zum Thema KI und Training von KI

**Was benötige ich zur Erstellung/Durchführung?**

- ✓ Endgerät
- ✓ Präsentationstechnik (z.B. Beamer oder Display)
- ✓ Internetzugang

**Was benötigen Schülerinnen und Schüler?**

- ✓ ggf. Endgerät
- ✓ ggf. Internetzugang

**Benötigte Software**

- ✓ ChatGPT Zugang

**Ähnliche Methoden**

- Suchen mit Sprachmodellen
- Bullshit-Zyklen bei Sprachmodellen erkennen und vermeiden
- KI-Texte im Faktencheck
- Unterstützung von Schreibprozessen