



UNIVERSITY  
OF COLOGNE

# Automatisierte Emotionsannotation

Annotation mit Sprachmodellen WS 23/24

# Gliederung

1. Datengrundlage: ChildTale-A Datensatz
2. Modellauswahl
3. Experimentdesign
4. Ergebnisse

# Datengrundlage: ChildTale-A Datensatz (Hermann & Lüdtké 2023)

- Umfassende Sammlung von Märchentexten aus den "Kinder- und Hausmärchen" der Gebrüder Grimm
- Speziell zusammengestellt für die Untersuchung und Analyse von Emotionen in Texten
- Enthält über **5.000 manuell annotierte Sätze**, die die emotionalen Inhalte der Märchen detailliert erfassen
- Annotationen umfassen emotionale Valenz (Skala -3 bis +3) und Arousal (Skala 1 bis 5), zur Bewertung des Ausmaßes und der Art der in jedem Satz ausgedrückten Emotionen

**Textsorte:** Märchen

**Sprache:** Deutsch.

**Quelle:** ChildTale-A Corpus, verfügbar auf [Zenodo](#).



# Datengrundlage: ChildTale-A Datensatz – manueller Annotationsprozess

- Durchgeführt von geschulten Master-Studierenden mit strukturierten Richtlinien
- Schwerpunkt auf den im Text ausgedrückten Emotionen
- Jeder Satz wurde hinsichtlich des Ausmaßes und der Art der Emotionen bewertet.

→ **Ziel:** Bereitstellung eines Goldstandards für die Analyse textlich kodierter Emotionen in literarischen Texten.

# Datengrundlage: ChildTale-A Datensatz – manueller Annotationsprozess

*“[...] 1. Bitte lesen Sie jeden einzelnen Satz und beurteilen sie den emotionalen Gehalt des Satzes. Dazu stehen Ihnen zwei Skalen zur Verfügung.*

*2. Auf der Valenzskala geben sie bitte an, ob der Satz eher etwas Negatives, etwas Neutrales oder etwas Positives beschreibt. Bitte nutzen sie die gesamte Skala von -3 bis +3, um möglichst genau anzugeben, ob der Satz z.B. nur leicht negativ (würde einem Wert von -1 entsprechen) oder sehr positiv ist (das würde z.B. einem Wert von +3 entsprechen).*

*3. Auf der zweiten Skala, der Arousal-Skala, geben sie bitte an, ob der Satz eher etwas Ruhiges bzw. Entspanntes oder eher etwas Erregtes bzw. Aufgeregtes beschreibt. Bitte nutzen sie hier die Abstufungen von 1-5 um ihre Einschätzung abzugeben. Niedrige Werte drücken aus, dass der Satz eher etwas Ruhiges/Entspanntes beschreibt, höhere Werte drücken aus, dass der Satz eher etwas Erregtes/Aufgeregtes beschreibt.”*

**– Hermann & Lüdtke 2023, Annotation Guidelines**

# Datengrundlage: ChildTale-A Datensatz – manueller Annotationsprozess

Satz	Valenz	Arousal
Nachdem es ihr der König versprochen hatte, tat sie die Augen zu und starb.	-3	5
Da ging sie hinauf, und glich einer Königstochter: und der König trat ihr entgegen und freute sich dass er sie wiedersah, und weil eben der Tanz anhub, so tanzten sie zusammen.	2,5	4

# Datengrundlage: ChildTale-A Datensatz – manueller Annotationsprozess

550 Sätze \* 8 Sekunden Annotation  $\approx$  **66 Minuten**

# Experimentdesign

Wie schlägt sich dagegen ein Sprachmodell?



# Experimentdesign

- Ursprünglich: Nutzung von GPT-4 für die Emotionsannotation im ChildTale-A Datensatz
- Wechsel zu GPT-3.5-Turbo (gpt-3.5-turbo-0613) aufgrund von Kostenüberlegungen und Probleme mit der API
  - > *context window* von 4096 Tokens
  - > *training data* bis September 2021
  - > *input*: \$0.0010 / 1K tokens ; *output*: \$0.0020 / 1K tokens

(GPT4-Turbo: *input*: \$0.01 / 1K tokens; *output*: \$0.03 / 1K tokens,  
GPT-4: *input*: \$0.03 / 1K tokens; *output*: \$0.06 / 1K tokens)

Source: OpenAI 2024

# Experimentdesign

1. Emotionsannotation mit GPT-3.5-Turbo *Zero Shot*
2. Emotionsannotation mit GPT-3.5-Turbo *Few Shot*

# Emotionsannotation mit GPT-3.5-Turbo Zero Shot

- Basierend auf Savelka et al. 2023: Minimale Anpassungen der ursprünglichen, für Menschen erstellten Annotationsrichtlinien.
  - > Entfernen von Richtlinienanteilen, die spezifisch für die menschliche Annotationsumgebung waren, aber für GPT-3.5 irrelevant sind
  - > Fokus auf die wesentlichen Aspekte der Emotionsannotation, um eine effiziente und relevante Nutzung von GPT-3.5 zu ermöglichen

# Emotionsannotation mit GPT-3.5-Turbo Zero Shot

1. "Pre-Processing":
  - a. Einlesen der Excel-Datei mit den einzelnen Sätzen mit polars und entfernen aller nicht benötigten Spalten
2. API-Anfrage (langchain + openai):
  - a. Angepasste Annotationsrichtlinie als System-Prompt einfügen
  - b. Satz als Input übergeben
3. Auswertung

# Emotionsannotation mit GPT-3.5-Turbo Zero Shot

""

1. Bitte lesen Sie jeden einzelnen Satz und beurteilen Sie den emotionalen Gehalt des Satzes.

2. Auf der Valenzskala geben Sie bitte an, ob der Satz eher etwas Negatives, etwas Neutrales oder etwas Positives beschreibt. Bitte nutzen Sie die gesamte Skala von -3 bis +3, um möglichst genau anzugeben, ob der Satz z.B. nur leicht negativ (würde einem Wert von -1 entsprechen) oder sehr positiv ist (das würde z.B. einem Wert von +3 entsprechen).

3. Auf der zweiten Skala, der Arousal-Skala, geben Sie bitte an, ob der Satz eher etwas Ruhiges bzw. Entspanntes oder eher etwas Erregtes bzw. Aufgeregtes beschreibt. Bitte nutzen sie hier die Abstufungen von 1-5 um ihre Einschätzung abzugeben. Niedrige Werte drücken aus, dass der Satz eher etwas Ruhiges/Entspanntes beschreibt, höhere Werte drücken aus, dass der Satz eher etwas Erregtes/Aufgeregtes beschreibt.

**Antworte in nur folgendem Format:**

**Valenz: <-3 oder -2,5 oder -2 oder -1,5 oder -1 oder -0,5 oder 0 oder 0,5 oder 1 oder 1,5 oder 2 oder 2,5 oder 3>**

**Arousal: <1 oder 1,5 oder 2 oder 2,5 oder 3 oder 3,5 oder 4 oder 4,5 oder 5>**

""

Kosten: \$3.68

Dauer 1h 33m 21s

Vorbild für fett markierten Teil: **Miranda 2023**

# Emotionsannotation mit GPT-3.5-Turbo Few Shot

""

Du bist ein Experte im Annotieren von Emotionen. Befolge die folgende Annotationsrichtlinien und annotiere den gegebenen Text.

---

1. Bitte lesen Sie jeden einzelnen Satz und beurteilen Sie den emotionalen Gehalt des Satzes.
2. Auf der Valenzskala geben Sie bitte an, ob der Satz eher etwas Negatives, etwas Neutrales oder etwas Positives beschreibt. Bitte nutzen Sie die gesamte Skala von -3 bis +3, um möglichst genau anzugeben, ob der Satz z.B. nur leicht negativ (würde einem Wert von -1 entsprechen) oder sehr positiv ist (das würde z.B. einem Wert von +3 entsprechen).
3. Auf der zweiten Skala, der Arousal-Skala, geben Sie bitte an, ob der Satz eher etwas Ruhiges bzw. Entspanntes oder eher etwas Erregtes bzw. Aufgeregtes beschreibt. Bitte nutzen sie hier die Abstufungen von 1-5 um ihre Einschätzung abzugeben. Niedrige Werte drücken aus, dass der Satz eher etwas Ruhiges/Entspanntes beschreibt, höhere Werte drücken aus, dass der Satz eher etwas Erregtes/Aufgeregtes beschreibt.

Antworte in nur folgendem Format:

Valenz: <-3 oder -2,5 oder -2 oder -1,5 oder -1 oder -0,5 oder 0 oder 0,5 oder 1 oder 1,5 oder 2 oder 2,5 oder 3>

Arousal: <1 oder 1,5 oder 2 oder 2,5 oder 3 oder 3,5 oder 4 oder 4,5 oder 5>

---

Es war einmal ein König, der hatte eine Frau mit goldenen Haaren, und sie war so schön, dass sich ihres Gleichen nicht mehr auf Erden fand.

Valenz: 2,5

Arousal: 1,5

Es geschah, dass sie krank lag, und als sie fühlte dass sie bald sterben würde, rief sie den König und sprach „wenn du nach meinem Tode dich wieder vermählen willst, so nimm keine, die nicht eben so schön ist, als ich bin, und die nicht solche goldene Haare hat, wie ich habe; das musst du mir versprechen.“

Valenz: -2,5

Arousal: 5

Nachdem es ihr der König versprochen hatte, tat sie die Augen zu und starb.

Valenz: -3

Arousal: 5

""

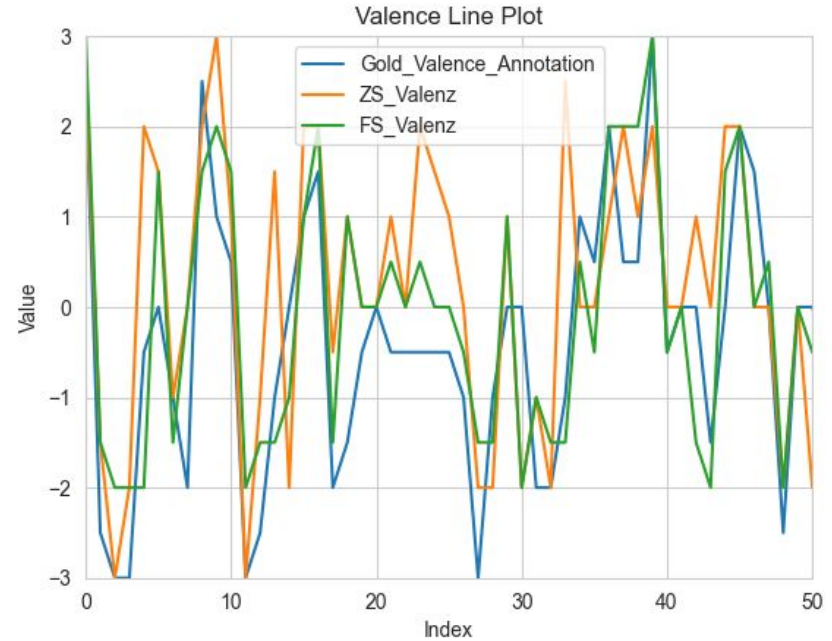
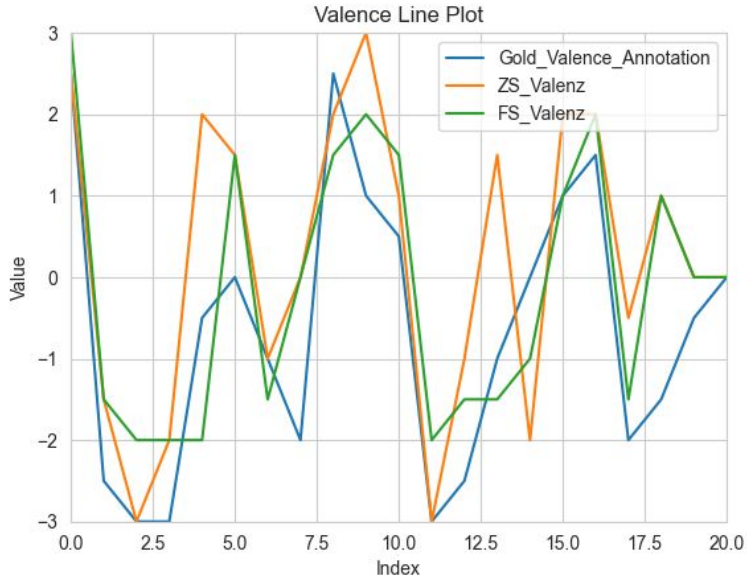
""

Kosten: \$5.74

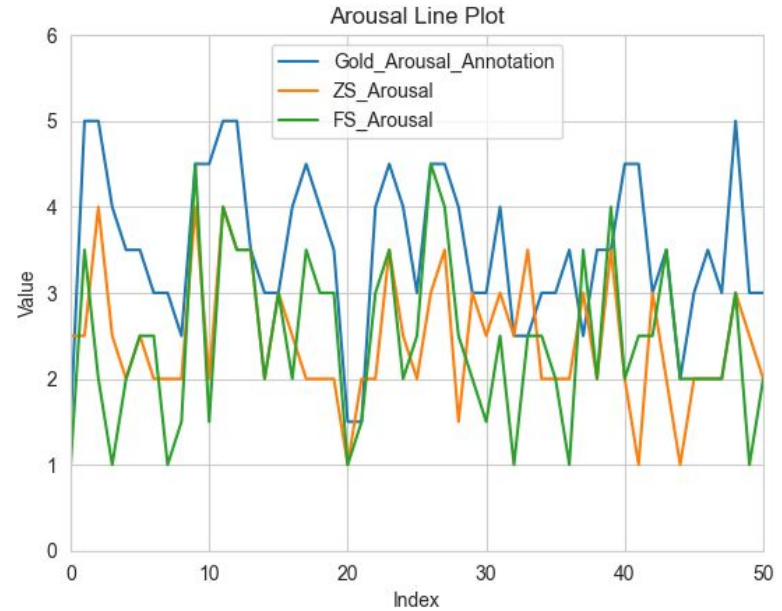
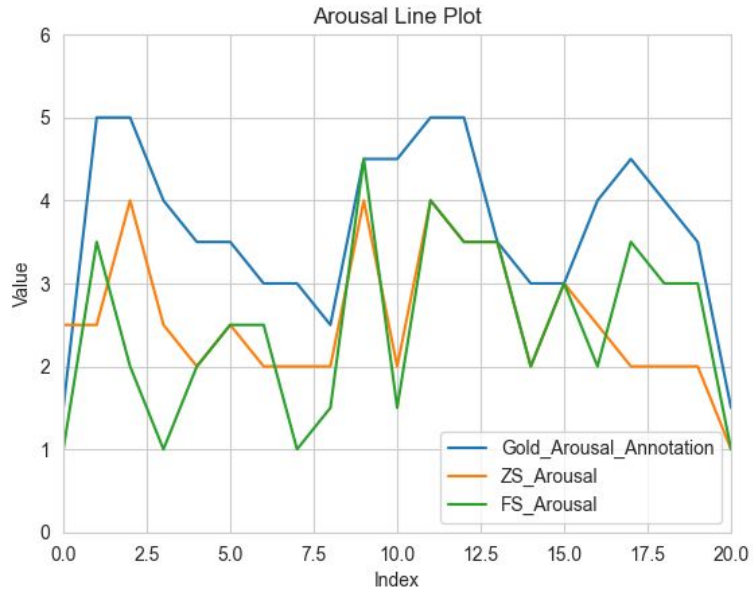
Dauer: 1h 30m 52s



# Ergebnisse

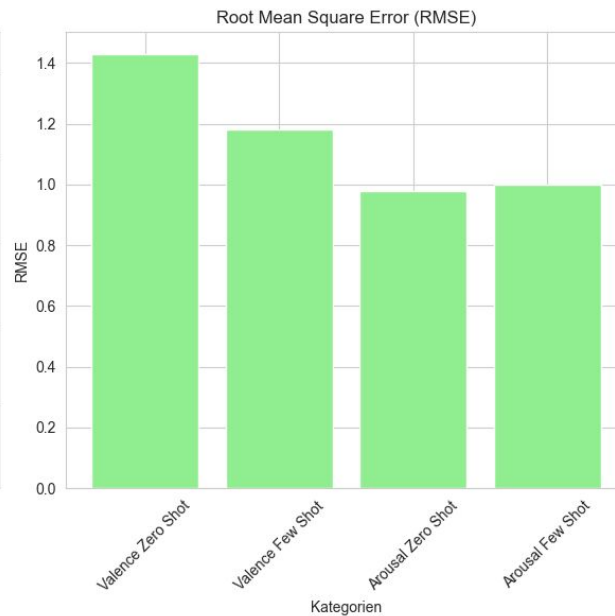
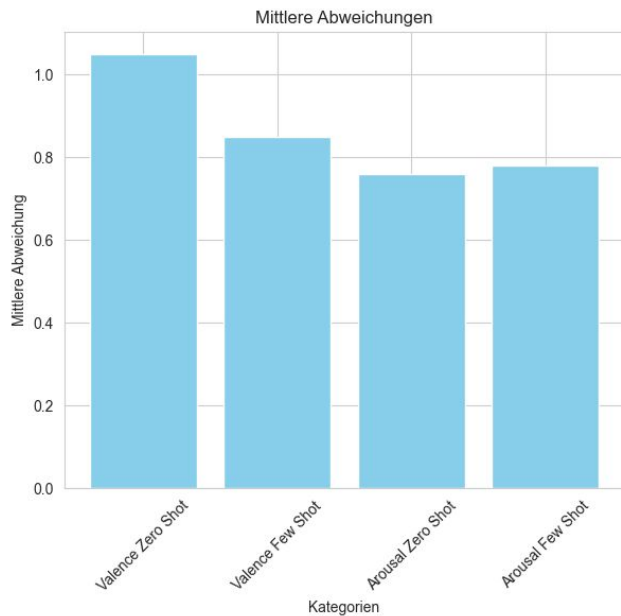


# Ergebnisse

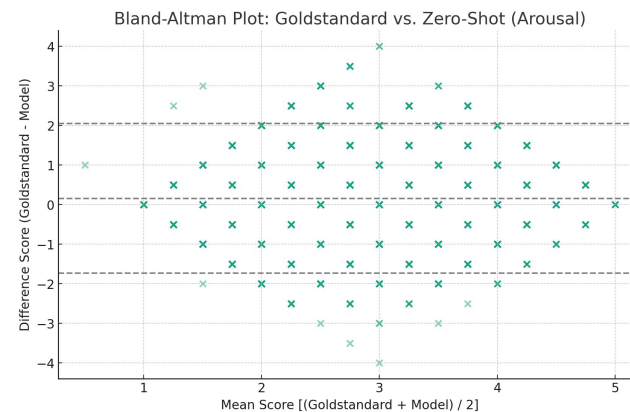
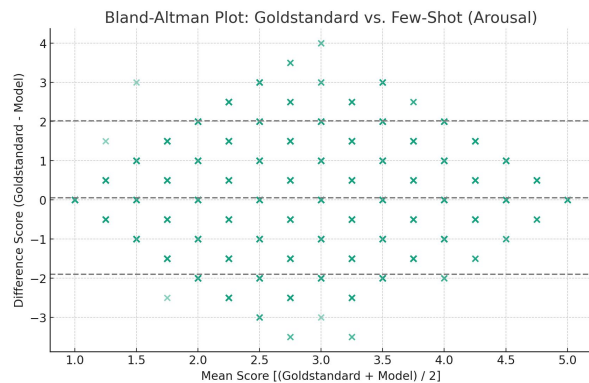
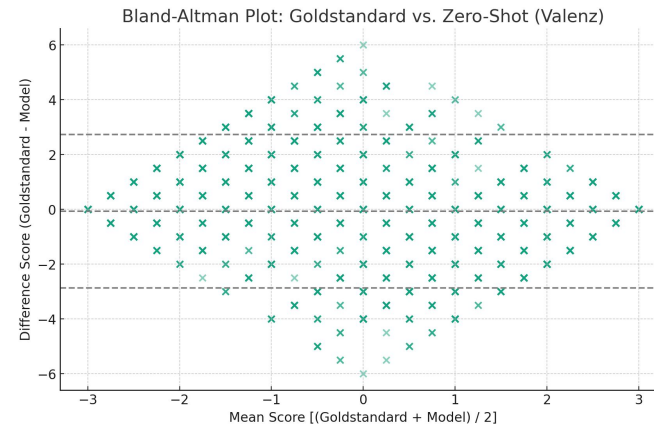
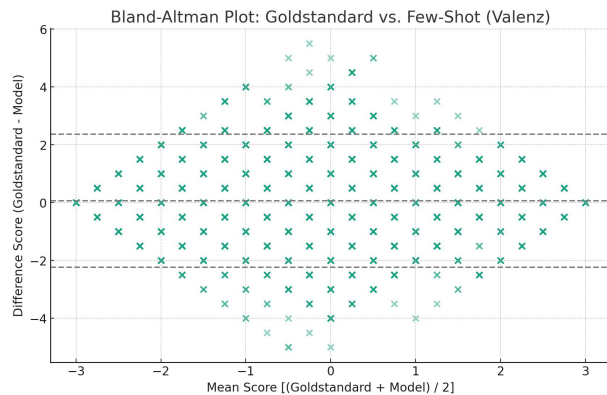




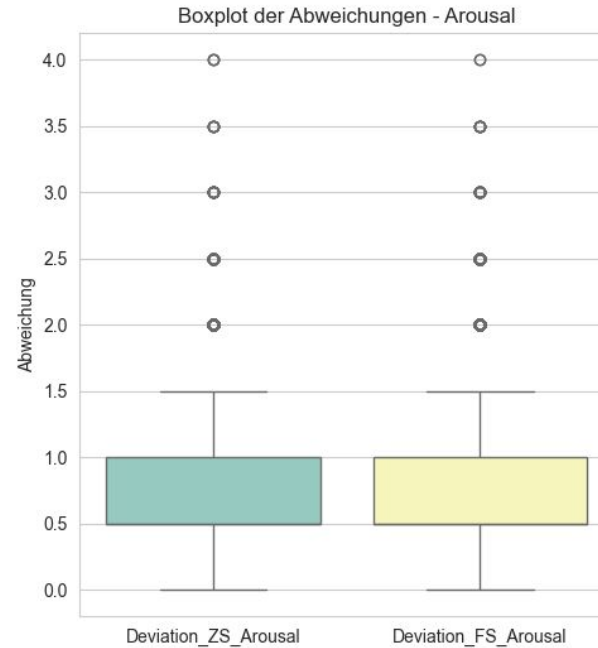
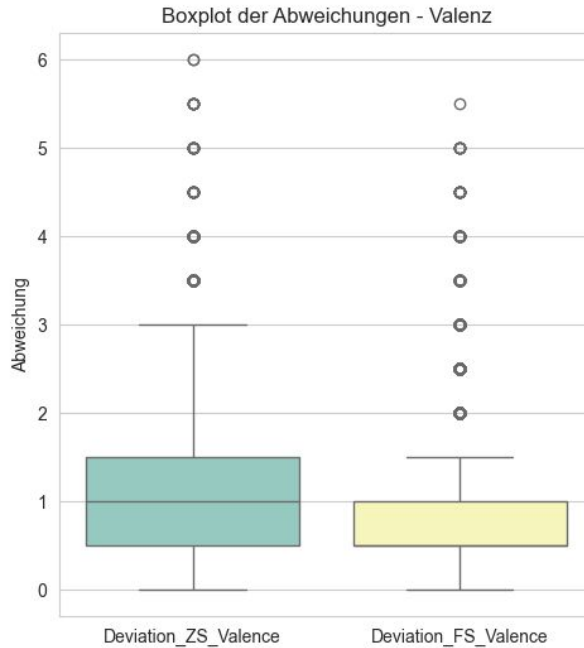
# Ergebnisse



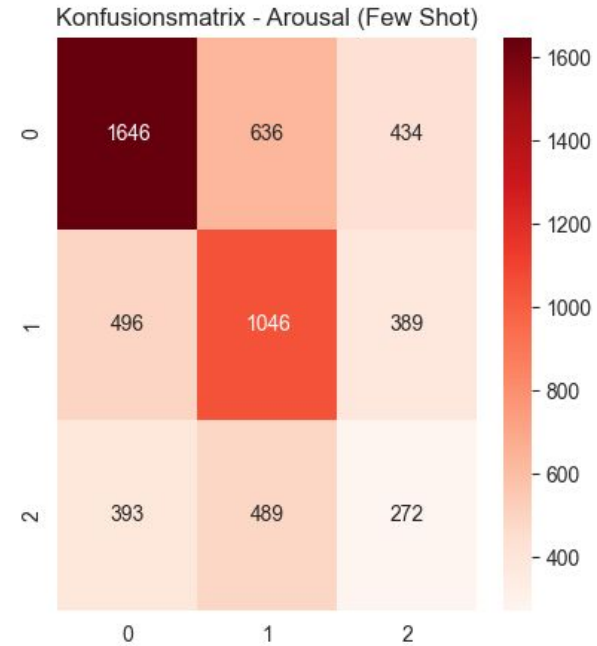
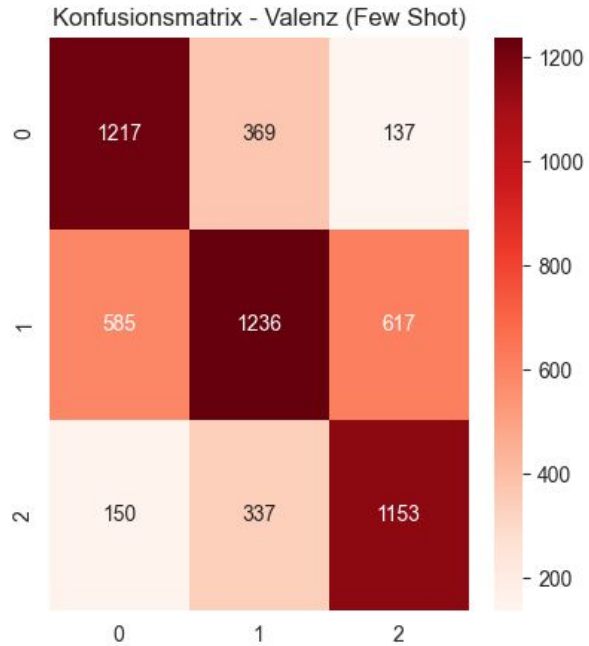
# Ergebnisse



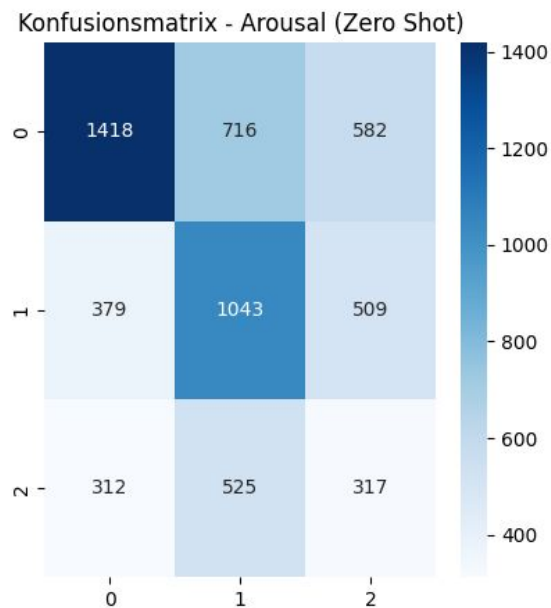
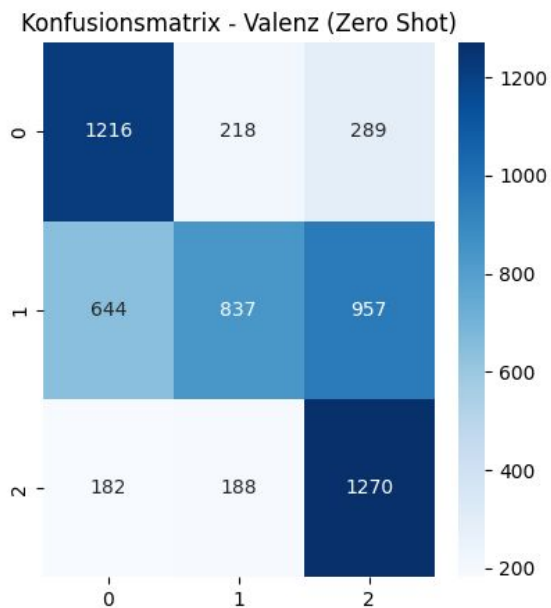
# Ergebnisse



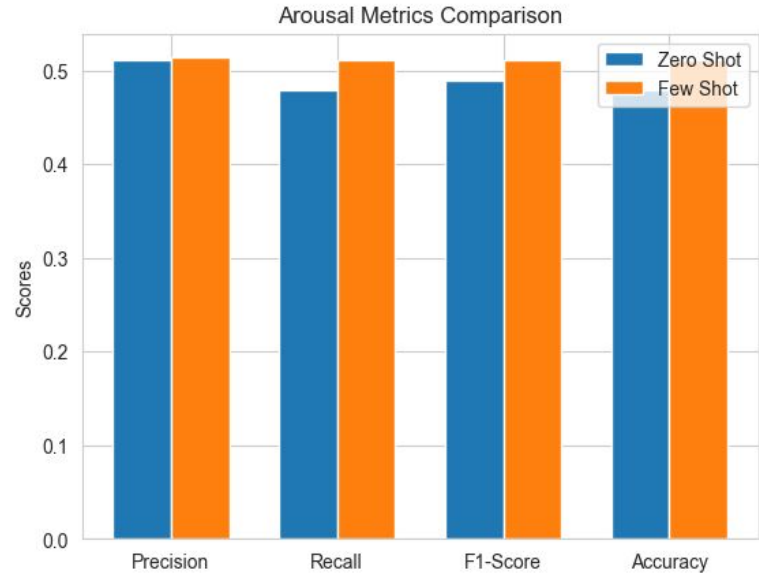
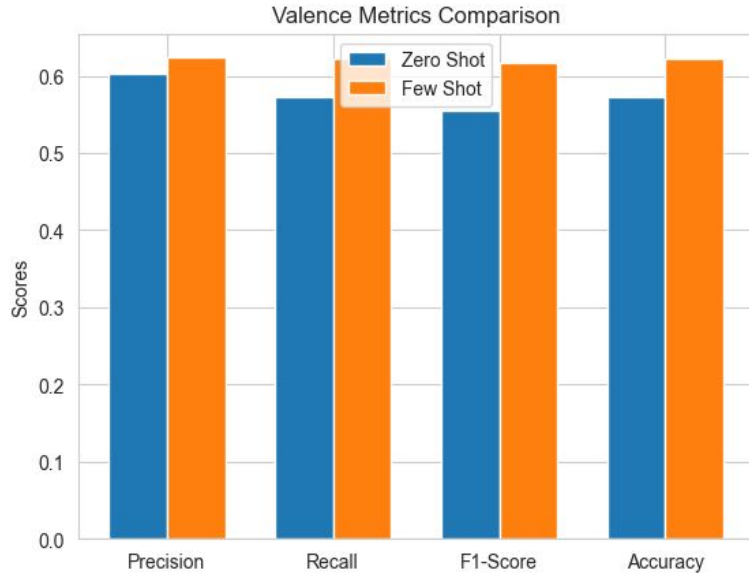
# Ergebnisse



# Ergebnisse



# Ergebnisse



# Quellen

- Herrmann, Berenike, und Jana Lüdtkke. 2023. „A Fairy Tale Gold Standard. Annotation and Analysis of Emotions in the Children’s and Household Tales by the Brothers Grimm“. HTML ,XML ,PDF. [https://doi.org/10.17175/2023\\_005](https://doi.org/10.17175/2023_005).
- Hermann, B., & Lüdtkke, J. (2023). Annotation Guidelines for the ChildTale-A Dataset. <https://zenodo.org/records/7737329> - zuletzt aufgerufen am 18.01.2024
- MIRANDA, LJ. 2023. „How Can Language Models Augment the Annotation Process?“ 24. März 2023. <https://lvmiranda921.github.io/notebook/2023/03/24/llm-annotation/>. - zuletzt aufgerufen am 18.01.2024
- OpenAI. 2024. „Pricing“. 18. Januar 2024. <https://openai.com/pricing> - zuletzt aufgerufen am 18.01.2024
- Savelka, Jaromir, Kevin D. Ashley, Morgan A. Gray, Hannes Westermann, und Huihui Xu. 2023. „Can GPT-4 Support Analysis of Textual Data in Tasks Requiring Highly Specialized Domain Expertise?“ In Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1, 117–23. <https://doi.org/10.1145/3587102.3588792>.
- 
- 
-



UNIVERSITY  
OF COLOGNE

Institut / Lehrstuhl / Dezernat

18.01.2024