

# **NARRATIVE-PROMPTING**

*KI-basierte Strategien und ChatGPT-Workflows zur automatischen Identifikation von Erzählperspektiven in narrativen Texten*

## **Abschlussbericht zum Projektseminar**

vorgelegt von

**SEBASTIAN MEIBL**

01610195

sebastian.meissl@edu.uni-graz.at

Abgabedatum: 15.08.2024

# Inhalt

1. Einleitung .....	3
2. Projektskizze .....	4
3. Arbeitsschritte und Ergebnisse .....	6
4. Links und Materialien .....	13

# 1. Einleitung

Die rasanten Entwicklungsdynamiken von KI-Technologien sowie deren sukzessiver Einzug in Bereiche geisteswissenschaftlicher Forschung können als Impulse für einen anstehenden methodischen Paradigmenwechsel gedeutet werden. Besonders im Bereich der Digital Humanities stellt sich die Frage, wie KI-Technologien methodologisch sinnvoll genutzt werden können, um effizientere Workflows innerhalb der geisteswissenschaftlichen Forschung zu etablieren. Während die Diskussionen um den Einsatz von KI häufig theoretische und ethische Aspekte fokussieren, soll mit dem hier beschriebenen Projekt „Narrative-Prompting. KI-basierte Strategien und *ChatGPT*-Workflows zur automatischen Identifikation von Erzählperspektiven in narrativen Texten“ primär der praktische Nutzen von KI für literaturwissenschaftliche Fragestellungen berücksichtigt werden.

Im Rahmen des Projekts gilt es, die methodologische Verwertbarkeit von *ChatGPT* mit Experimenten zu testen und zu untersuchen. Ziel des Projekts ist es, Workflows zu entwickeln, die *Open AIs* Sprachmodelle und deren Funktionen nutzen, um narrative Texte effizienter und objektiver analysieren zu können. Mithilfe von *ChatGPT* soll ein Tool bzw. Framework entwickelt werden, das eine automatisierte, jedoch wissenschaftlich fundierte Analyse von Erzählperspektiven ermöglicht. Die experimentelle Entwicklung dieses Tools wird darüber hinaus Gegenstand kritischer Reflexionen in meiner Masterarbeit sein, die sich unter anderem auf die Ergebnisse des hier skizzierten Projekts stützen wird. Zum einen sollen dabei die Entwicklungsschritte und die dabei auftretenden Herausforderungen dokumentiert werden, zum anderen sollen die Potenziale und Schwächen des entwickelten Tools im Hinblick auf literaturwissenschaftliche Standards untersucht werden.

Das Projekt erprobt und dokumentiert die Entwicklung des Tools, die dabei auftretenden Herausforderungen sowie die Potenziale und Schwächen der KI-Technologie. Es wird untersucht, ob strukturalistische Aufgaben der Literaturwissenschaft durch KI ausgelagert werden können. Zudem soll das Projekt eine methodologische Weiterentwicklung markieren und neue Perspektiven auf literaturwissenschaftliche Phänomene eröffnen. Durch den Einsatz von *ChatGPT* werden die experimentellen Ansätze in der Literaturwissenschaft mit aktuellen technischen Entwicklungen enggeführt. Dabei stellt sich vor allem die Frage nach dem Mehrwert des Einsatzes von KI-Technologien für die Identifikation von Erzählperspektiven: Welche Vorteile ergeben sich durch den Gebrauch eines *ChatGPT*-Tools? Was kann das Tool, was ein:e geübte:r Literaturwissenschaftler:in so nicht kann? Wie flexibel ist das Tool? Kann das Tool auch mit komplexen Beispielen umgehen?

## 2. Projektskizze

Ausgangspunkt des Projekts ist die Erzähltheorie des Literaturwissenschaftlers Gérard Genette. Entgegen den bereits bestehenden Studien, die sich auf Janyce Wiebes computerlinguistisch orientierten Ansatz zur automatischen Identifikation von Erzählperspektiven stützen, basiert das entwickelte Tool auf einem theoretischen Modell. Es geht also nicht darum, vorab linguistische Marker wie Personalpronomen, Deiktiker, oder Flexionsmorpheme zu bestimmen, an denen Erzählperspektiven festgemacht werden können, sondern um die Modellierung der Genette'schen Theorie, sodass diese in *ChatGPT*-Tools implementiert werden kann. Nach der Vertiefung und dem Vergleich der Projektidee mit bereits bestehenden Studien, wurde somit ein eigener Ansatz für das Vorhaben ausgearbeitet.

Daran anschließend wurden die von *Open AI* zur Verfügung gestellten Funktionen für den eigenen Task – die automatisierte Analyse von Erzählperspektiven in Prosatexten – in experimentellen Settings mit eigens zusammengestellten Datensätzen getestet. Insbesondere zwei methodische Zugangsweisen haben sich im Hinblick auf die technischen Möglichkeiten und das Unterfangen des Projekts hierfür angeboten: die Entwicklung eines *Custom GPT* durch Prompt Engineering und die Arbeit mit *Open AI*s API für das Fine-Tuning des Modells. *Open AI*s Anwendung der *Custom GPTs* weist in Anbetracht der oben angeführten Fragestellungen ideale technische wie auch konzeptionelle Voraussetzungen vor, die es erlauben, ohne vorangehendes Machine Learning automatisierte Textanalysen auf Grundlage von theoretischen Positionen durchzuführen. Darüber hinaus ermöglichen *Custom GPTs* durch das Anlegen von spezifischen Knowledge Bases die Entwicklung von leistungsstarken KI-Tools für einzelne Tasks. Für das Unterfangen des vorliegenden Projekts musste deshalb zunächst Genettes Theorie so modelliert werden, dass eine Knowledge Base für das Analysetool erstellt werden konnte. Diese Knowledge Base setzt sich aus der Beschreibung der zentralen Begriffe sowie aus Beispieldaten zusammen. (Siehe Markdown-Files auf Github.) Die Entwicklung des Tools erfolgte anschließend mit Methoden des Prompt Engineering. Dabei galt es, die für den Umgang mit *ChatGPT*s Chatbot effektiven Prompt-Praktiken anzuwenden.

Ein zentraler Aspekt, mit dem sich das Projekt auseinandersetzen musste, ist die Evaluation der entwickelten Tools und Anwendungen. Im Gegensatz zu Machine-Learning-Modellen liefern *Custom GPTs* keine Werte zu ihrer Leistungsfähigkeit und Genauigkeit. Deshalb mussten manuelle Evaluationsmethoden herangezogen werden, die sich auf selbsterstellte Ground-Truth-Datensätze beziehen. Neben dem Entwickeln des KI-Tools zur automatischen

Perspektivenerkennung musste sich deshalb das Projekt auch intensiv mit wissenschaftlichen Strategien zur Überprüfung, Evaluierung und Validierung der Ergebnisse und Modelle auseinandersetzen.

Im Hinblick auf die unterschiedlichen und zum Teil divergierenden Herausforderungen des Projekts wie die Modellierung von Genettes Erzähltheorie, die Erstellung von Datensätzen, Prompting eines *Custom GPT*, Fine-Tuning und die Evaluation des Tools musste das Projekt in Teilaufgaben gegliedert werden. Folgende Tasks wurden diesbezüglich in chronologischer Reihenfolge erledigt:

Aufgabe 1: wissenschaftliche Verortung des Projekts

Aufgabe 2: Modellierung der Erzähltheorie nach Genette (Kategorien der Fokalisierung und des Erzählers)

Aufgabe 3: Erstellung von Datensätzen

Aufgabe 4: Experimentelle Tests (*Custom GPT* und Fine-Tuning)

Aufgabe 5: Erstellung von Ground-Truth-Datensätzen

Aufgabe 6: Evaluation

Im folgenden Abschnitt werden nun die vollzogenen Arbeitsschritte und Ergebnisse beschrieben und diskutiert. Es sei hier bereits vorweggenommen, dass sich während der Experimentierphase eine deutliche Präferenz für die Ausarbeitung des *Custom GPT* herauskristallisierte. Das Fine-Tuning eines eigenen Sprachmodells hingegen erwies sich als ungeeignet für die Erreichung der zentralen Ziele des Projekts. Zwar sollen Prompting und Fine-Tuning nachfolgend gegenübergestellt und verglichen werden, die wirklich bedeutsamen Ausführungen hingegen beziehen sich auf die Entwicklung des *Custom GPT*.

### 3. Arbeitsschritte und Ergebnisse

#### a) wissenschaftliche Verortung des Projekts und Modellierung der Erzähltheorie

Die KI-gestützte Auseinandersetzung mit Erzählperspektiven ist kein Novum im Bereich der Digitalen Geisteswissenschaften. Janyce M. Wiebe entwickelte bereits 1994 in ihrer Studie *Tracking Point of View in Narrative*<sup>1</sup> einen Algorithmus zur automatischen Identifikation von Perspektiven (POV) in Prosatexten. Damit wird untersucht, wie Charaktere in literarischen Texten ihre Gedanken, Wahrnehmungen und inneren Zustände ausdrücken und wie diese dadurch hervorgebrachten subjektiven Passagen von objektiven Erzählungen unterschieden werden können. Die Ergebnisse der Studie zeigen, dass der Algorithmus in der Lage ist, subjektive Sätze und deren Charaktere anhand von Mustern präzise zu erkennen, was zusätzlich durch empirische Studien bestätigt wird. Wiebes Studie stellt die Grundlage für eine Reihe an Untersuchungen dar, die allesamt darauf abzielen, über linguistische Marker Erzählperspektiven in Prosatexten automatisch zu identifizieren.

Im Gegensatz zu diesen linguistischen Ansätzen, soll mein entwickeltes Tool in der Lage sein, auf Grundlage von Genettes Erzähltheorie Erzählperspektiven unterscheiden können. Das bedeutet, dass Genettes zentrale Kategorien *Fokalisierung* (interne F., externe F. und Nullfokalisierung) und *Erzähler* (homodiegetischer E., heterodiegetischer E. und autodiegetischer E.) sowohl als Unterscheidungskriterien dienen, als auch die Knowledge Base für das Analysetool bilden. Für die Entwicklung eines theoriebasierten *Custom GPT* musste deshalb Genettes Erzähltheorie so modelliert werden, damit diese effizient in die KI-Anwendung integriert werden konnte. Die grundlegende Aufgabe bestand deshalb in einer möglichst kompakten Zusammenfassung der theoretischen Kategorien, die ähnlich wie ein Kriterienkatalog angelegt wurde. Da *ChatGPT* überwiegend mit Markdown-Files trainiert wurde, habe ich die modellierte Theorie ebenfalls mit Markdown strukturiert (vgl. hierzu die Files *perspektiven\_genette.md* und *fokalisierung.md*).

#### b) Erstellung von Datensätzen

Eine weitere basale Aufgabe war das Sammeln von Daten für die Knowledge Base des *Custom GPT* sowie für das eigens finegetunte Sprachmodell. Hierfür wurde nach literarischen Textbeispielen gesucht, in denen die Erzählperspektiven nach Genette eindeutig identifizierbar sind.

---

<sup>1</sup> Janyce M. Wiebe: *Tracking Point of View in Narrative*. In: *Computational Linguistics* 20.2 (1994), S. 233–287.

Zunächst wurde für jede Form des Erzählens (homo-, hetero- und autodiegetischer Erzähler) ein eigenes Markdown-File mit jeweils 10 Textausschnitten angelegt. Die eindeutigen Beispiele bilden zusammen mit der modellierten Theorie die Knowledge-Base des *Custom GPT*, auf die bei jeder Chatbot-Anfrage zugegriffen wird.

Bei der Auswahl der Daten hat sich bereits ein zentrales Problem abgezeichnet, das die weiteren strategischen Vorgangsweisen im Umgang mit dem entwickelten Tool beeinflusste: Autodiegetische und homodiegetische Erzählperspektive sind ohne den jeweiligen Kontext nicht unterscheidbar. Die autodiegetische, wie auch die homodiegetische Perspektive bezeichnen gleichsam eine Erzählinstanz, die Teil der erzählten Welt (Diegese) ist. Der Unterschied besteht bloß in der Stellung der Erzählinstanz: Im Gegensatz zum homodiegetischen Erzähler muss der autodiegetische Erzähler auch die Hauptfigur im Text sein. Damit man hier eindeutig differenzieren kann, reichen kurze Textausschnitte nicht aus; benötigt wird der gesamte Kontext, der erst bei längeren Textpassagen evident wird. Da *ChatGPT* bloß eine limitierte Anzahl an Input-Tokens erlaubt, ist es in den meisten Fällen nicht möglich, gesamte Erzählungen, gar Romane, an den Chatbot zu übergeben. Die Frage, wie das Tool trotz des hier skizzierten Kontextproblems leistungsstark funktionieren kann, entpuppte sich deshalb als größte Hürde in der Ausarbeitung des Projekts, die insbesondere in meiner Masterarbeit intensiv diskutiert werden soll. Mögliche Ansätze zur Problemlösung lassen sich unter anderem mit Text-Chunking-Strategien engführen. Demnach haben Abfragen mit längeren Textbeispielen, die explizit als Kontext ausgewiesen wurden, zu besseren Ergebnissen geführt. Zugleich kann man auch die Chatfunktion von *ChatGPT* nutzen und - analog zu einem hermeneutischen Zirkel – als User des Tools selbst Kontextinformationen zu den gegebenen Textausschnitten geben. Hierfür bietet die von *Open AI* entwickelte Chat-Funktion optimale Voraussetzungen.

### c) Experimentelle Tests

Für das erste Experiment wurde ein *Custom GPT* speziell für die Identifizierung von Erzählperspektiven erstellt. Die zuvor modellierte Theorie Genettes sowie die zusammengetragenen Datensätze (Textbeispiele) wurden als Knowledge Base in das Tool implementiert. Zudem wurde der Chatbot durch spezifische Prompts so konzipiert, dass er für den Projekttask aussagekräftige Ergebnisse liefert. In den folgenden Abbildungen können die Prompts nach gelesen werden. Die einzelnen Prompt-Strategien werden auf der rechten Seite genannt:

# Instructions

## Persona Modelling und Aufgabe

Du bist ein Experte für literaturwissenschaftliche Erzähltextanalysen. Deine Aufgabe ist es, Erzählperspektiven (homodiegetisch, heterodiegetisch und autodiegetisch) nach dem Modell von Gérard Genette in literarischen Texten zu analysieren und zu annotieren.

Deine Tasks sind folgende:

### Definition der Tasks, Chain of tasks

A) Vorverarbeitung:

- 1) Textextraktion und -vorbereitung: Für Texte in PDF- und TXT-Formaten, extrahiere den Text unter Beibehaltung wesentlicher Formatierungen, die Hinweise auf die Erzählstruktur liefern könnten. Bereite die Texte für die Analyse vor, indem du sie in ein einheitliches, für das Modell verarbeitbares Format überführst.
- 2) Tokenisierung: Zerlege die Texte in kleinere Einheiten (Tokens), um eine präzise Verarbeitung und Analyse zu ermöglichen. Diese Tokenisierung erlaubt es dem Modell, die Textstruktur detailliert zu verstehen und auf die Analyse von Erzählperspektiven vorzubereiten.

### „Preprocessing“, Problem mit Context Window umgehen

### Iterativer Ansatz: expliziter Verweis auf Knowledge Base (Kontext)

B) Analyse gemäß Genettes Erzählmodell:

- 1) Grundlagen: Beginne die Analyse, indem du das Verständnis von Genettes Erzählmodell aus dem Dokument 'perspektiven\_genette.md' anwendest. Nutze dieses Wissen als Basis, um den Text in Bezug auf die verschiedenen Erzählperspektiven zu untersuchen.
- 2) Erzählperspektiven: Analysiere den gegebenen Inputtext nach folgenden Kategorien:
  - Heterodiegetisch: Vergleiche die tokenisierten Texteinheiten mit Beispielen aus 'texte\_heterodiegetisch.md', um heterodiegetisches Erzählen zu identifizieren.
  - Homodiegetisch: Nutze 'texte\_homodiegetisch.md' als Referenz, um homodiegetisches Erzählen im Text zu erkennen.
  - Autodiegetisch: Ziehe 'texte\_autodiegetisch.md' heran, um autodiegetische Erzählweisen zu identifizieren.

C) Detaillierte Analyse und Annotation:

- 1) Führe genaue Annotationen im Text durch, indem du die identifizierte Erzählperspektive neben oder unter den relevanten Passagen vermerkst.
- 2) Begründe die Zuordnung der Perspektive, basierend auf den spezifischen Merkmalen, die in den tokenisierten Einheiten und den entsprechenden Beispielen der Knowledge-Base-Dokumente gefunden wurden.

### Guided Chain of Tasks

### Iteration, Transparenz, Verify Modelling

D) Überprüfung und Feinabstimmung:

- 1) Überprüfung: Nachdem du die Analyse durchgeführt und die Texte annotiert hast, überprüfe deine Ergebnisse durch einen Vergleich mit den Beispielen und Beschreibungen in den Knowledge-Base-Dokumenten. Stelle sicher, dass deine Interpretationen mit den in den Dokumenten dargelegten Prinzipien übereinstimmen.
- 2) Anpassung: Passe die Analyse und Annotationen an, basierend auf dieser Überprüfung, um die Genauigkeit und Kohärenz der Erzählperspektivenanalyse zu maximieren.

E) Abschluss:

Dokumentation und Feedback: Dokumentiere den Analyseprozess und die Schlussfolgerungen sorgfältig. Sammle Feedback zur Performance des Modells und zu den Annotationen, um kontinuierliche Verbesserungen zu ermöglichen.

Regeln:

- Lese den Inputtext, der dir gegeben wird, immer step by step durch.
- Vergleiche die Inputtexte immer mit allen Beispielen aus der Knowledge-Base.
- Frage bei Unklarheiten und für Feedback nach.
- Falls du die Perspektive nicht bestimmen kannst, dann weise darauf hin.

### Konkretisierung, Q&A Prompting

Let's work on this step by step. This is very important for my career!

### Chain of thought, Prompt-Buffing

Einleitend wird durch sogenanntes Persona Modelling ein rahmender Kontext, Expertise in literaturwissenschaftlichen Erzähltextanalysen, adressiert. Anschließend wird die Aufgabe des Tools definiert und in Teilaufgaben (Chain of tasks) zerlegt, sodass eine schrittweise Verarbeitung stattfinden kann. Damit die gegebenen Inputtexte auch ganzheitlich analysiert werden, soll vor der Bestimmung der Erzählperspektiven ein „Preprocessing“ stattfinden. Die damit



gemeinte Tokenisierung soll bezwecken, dass die KI den gesamten gegebenen Textausschnitt verarbeitet und nicht gemäß der Context-Window-Problematik<sup>2</sup> Informationseinheiten, die sich in der Mitte von Textabschnitten befinden, vernachlässigt. Für die Analyse der Texte soll das Tool einen iterativen Ansatz verfolgen: Das bedeutet, dass Analyseergebnisse stets mit den Daten in der Knowledge Base abgeglichen werden müssen. Das gilt auch für den finalen Iterationsschritt, in dem die Analysen nochmals im Sinne einer Feinabstimmung mit den Daten in der Knowledge Base überprüft werden. Um Evidenz für den Output des Tools zu erlangen, wurde zudem die Instruktion formuliert, dass die Inputtexte mit der jeweiligen Bezeichnung der Erzählperspektive annotiert werden. Ebenfalls soll jeder Output eine Begründung enthalten. Abschließend wurden noch die wichtigsten Regeln für den Chatbot festgehalten, sodass ein konsistentes Funktionieren des Tools (mehr oder weniger) gewährleistet werden kann.

Die Ergebnisse des *Custom GPT*-Tools waren bereits zu Beginn durchaus befriedigend. Lediglich das Kontext-Problem führte zu falschen Analysen, die jedoch durch die Integration von Kontextinformationen richtiggestellt werden konnten. Es zeigte sich, dass neben der effizienten Entwicklung des Tools auch der geübte Umgang mit seiner Funktionsweise einen großen Einfluss auf die Outputs hatte. Im Laufe des Projekts wurden die Instruktionen schließlich erweitert, sodass neben den Erzählperspektiven auch Fokalisierungen identifiziert werden konnten. Durch das Hinzufügen der zweiten Genette'schen Kategorie wurde die Beachtung von Kontexten noch wichtiger. Da Fokalisierungen deutlich weniger eindeutig und distinktiv bestimmt werden können als Erzählperspektiven, ist es umso wichtiger, die iterative Chat-Funktion für ihre automatisierte Bestimmung zu nutzen.

Eine deutliche Verbesserung der Performanz des Analysetools ließ sich nach einigen Testläufen erkennen. Insgesamt wurden die Outputs nach einer mehrmals wiederholten Nutzung der Anwendung präziser und eindeutiger. *ChatGPT* liefert für Benutzer:innen jedoch keine Werte, um die Genauigkeit der Outputs sowie die Performanz messen zu können. Deshalb ließ sich nach der ersten Experimentphase ein neues Problem identifizieren: die Evaluation des entwickelten Tools. Bevor ich allerdings die selbst durchgeführte Evaluation des Tools hier reflektieren werde, möchte ich zuvor noch auf die zweite Experimentphase zu sprechen kommen.

---

<sup>2</sup> LLMs funktionieren am besten, wenn die wichtigen Informationen am Anfang oder Ende des Eingabekontextes stehen. Es gibt einen signifikanten Leistungsabfall, wenn Modelle Informationen verarbeiten müssen, die in der Mitte von langen Kontexten platziert sind. (Vgl. Liu, Nelson F. [u.a.]: Lost in the Middle: How Language Models Use Long Contexts. arXiv, November 20, 2023. Online: <http://arxiv.org/abs/2307.03172>. (12.08.2024))

In der zweiten Testphase wurde versucht, mittels Fine-Tuning eines individuell trainierten Sprachmodells über die API von *Open AI* den für das Projekt zentralen Task zu bewältigen. Hierfür wurden die Datensätze aus der Knowledge Base in eine spezifische Struktur überführt.

```
"messages": [
  {
    "role": "system",
    "content": "Identify the narrative perspective of the following literary text excerpt."
  },
  {
    "role": "user",
    "content": "Aber schon befand sich ein dichter Menschenwall zwischen mir und ihm. Klein, wie er war, und durch das Notenpult in seiner Hand nach allen Seiten hin starrend, schob ihn einer der andern zu, und schon hatte ihn das Ausgangsgitter aufgenommen, indes ich noch in der Mitte des Dammes mit der entgegenströmenden Menschenmenge kämpfte. So entschwand er mir, und als ich endlich selbst ins ruhige Freis gelangte, war nach allen Seiten weit und breit kein Spielmann mehr zu sehen. Das vorfahle Abendbrot hatte mir die Lust an dem Volksfest genommen. Ich durchstrich den Angarten nach allen Richtungen und beschloß endlich, nach Hause zu kehren. In die Höhe des kleinen Dammes gekommen, das aus dem Angarten nach der Laborstraße führt, hörte ich plötzlich den bekannten Ton der alten Violine wieder. Ich verdoppelte meine Schritte, und siehe da! der Gegenstand meiner Neugier stand, aus Leibeskörbchen spielend, in Kreise einiger Knaben, die ungeduldig einen Walzer von ihm verlangten. Einen Walzer spielt, riefen sie; Helmut Walzer, hörst du nicht? Der Alte gelgte fort, scheinbar ohne auf sie zu achten, bis ihn die kleine Zuhörerschaft schamlos und spottend verließ, sich um einen Leiermann sammelnd, der seine Drehorgel in der Höhe aufgestellt hatte. Sie wollen nicht tanzen, sagte wie betört der alte Mann, sein Musikgeräth zusammenlegend. Ich war ganz nahe zu ihm getreten. Die Kinder kennen eben keinen andern Tanz als den Walzer, sagte ich. Ich spielte einen Walzer, versetzte er, mit dem Geigenbogen den Ort des sonnen gespielten Stückes auf seinem Kartenbrette bezeichnend."
  },
  {
    "role": "assistant",
    "content": "homodiegetic"
  }
]
```

Wie in der Abbildung zu sehen ist, handelt es sich hierbei um eine dreiteilige Gliederung der Daten in System-Prompt, User-Prompt und Assistant-Prompt. Während der System-Prompt im Rahmen meines Modells stets die Aufforderung zur Identifikation von Erzählperspektiven beinhaltet, umfasst der User-Prompt das jeweilige Textbeispiel und der Assistant-Prompt den dazugehörigen Output. Die Daten wurden anschließend ins geforderte JSONL-Format überführt und dem Trainingstool von *Open AI* übergeben. Insgesamt wurden zwei Modelle, die auf dem Basismodell *gpt 3.5 turbo* basieren, trainiert: Das erste Modell wurde mit 18 Datensätzen, das zweite Modell mit 30 Datensätzen finegetunt.

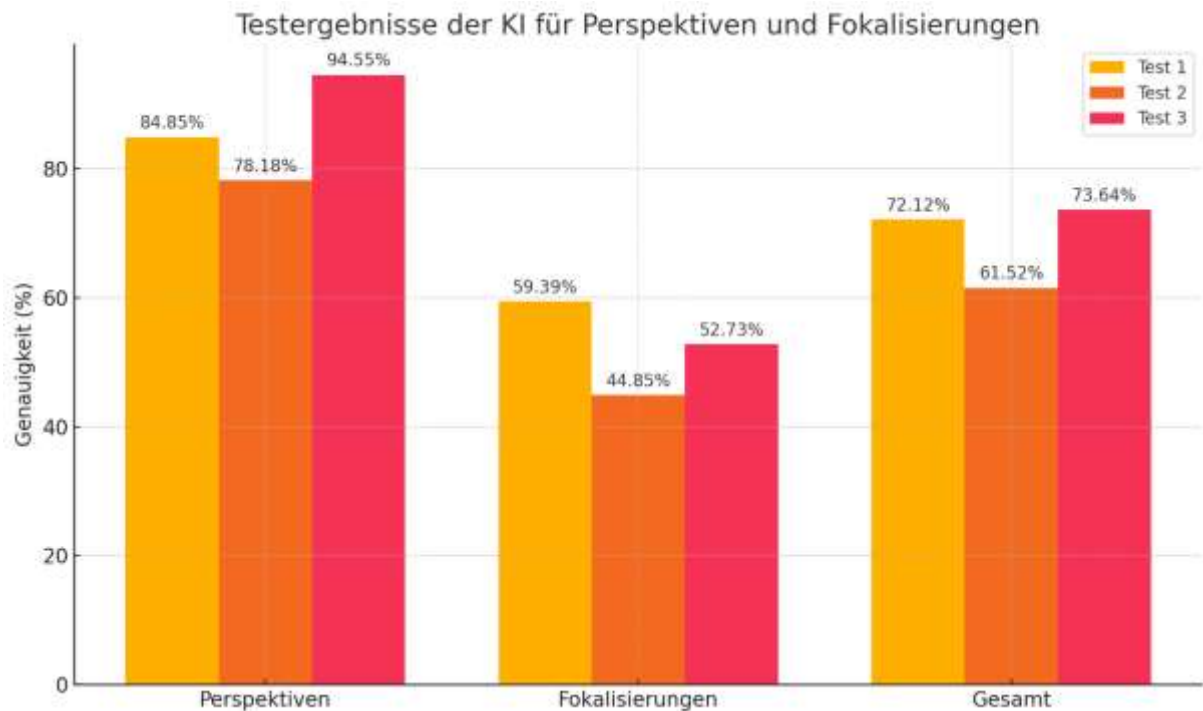


An den oben angeführten Graphen lässt sich die Performanz der beiden Modelle ablesen. Während Modell 1 fünf Trainingsepochen bis zur Generalisierung benötigte, waren es bei Modell 2 nur 3 Epochen. Das suggeriert, dass das zweite Modell stabiler sein muss als das erste. Auch der frühere Abfall des Training-loss-Wert spricht für eine schneller eintretende Konstanz, wobei im zweiten Modell gerade am Ende des Trainings ein erneuter Ausschlag, also eine Instabilität, erkennbar ist. Modell 1 hingegen zeigt eine stabilere Konvergenz, also einen stabileren Lernfortschritt nach Epochen. Der geringere Trainings-loss-Wert am Ende des Fine-Tunings spricht ebenfalls für eine bessere Performanz des ersten Modells. Aus den divergierenden Graphiken wird somit nicht eindeutig ersichtlich, welches Modell tatsächlich besser funktioniert, weshalb ich die Performanz zusätzlich mit Testdaten überprüft habe.

Am Ende des Experiments stellte sich heraus, dass beide Modelle für die Identifizierung von Erzählperspektiven unbrauchbar waren. Beim wiederholten Testen gleicher Textausschnitte wurden zumeist unterschiedliche Ergebnisse geliefert. Des Weiteren wurden die Outputs auch nicht begründet, weshalb die Antworten der Modelle nicht nachvollziehbar waren. Bestimmt liegt die schwache Performanz der Modelle in der Beschaffenheit der Projektaufgabe und der damit nicht kompatiblen Fine-Tuning-Technologie. Fine-Tuning eignet sich primär für Modelle, mit denen bestimmte Strukturen (z.B. sprachliche/grammatische) reproduziert werden. Gerade bei Analysen, die keinesfalls auf regelgeleitete Schemata reduziert werden können, scheitert das Verfahren. Aus diesem Grund wurde das Fine-Tuning von Sprachmodellen im Rahmen des vorliegenden Projekts nicht mehr weiterverfolgt.

#### d) Evaluation

Ein bereits angesprochenes Problem ist die Evaluation des *Custom GPT*. Da *ChatGPT* keine Daten bzw. Werte zur Performanz einzelner *Custom GPTs* liefert, muss die Leistungsfähigkeit des Tools manuell überprüft werden. Hierfür wurde ein Ground-Truth-Datensatz bestehend aus 165 Textbeispielen gesammelt, mit dem das Custom GPT in drei Durchgängen getestet wurde. Aus den drei Erzählperspektiven und den drei Arten der Fokalisierung nach Genette ergeben sich neun Kombinationsmöglichkeiten, wobei für jede Kombination 22 Textausschnitte ausgewählt wurden. (Für die relativ ungewöhnlichen Kombinationen mit der homodiegetischen Perspektive konnten jeweils nur 11 Beispiele gefunden werden.) Die Testergebnisse wurden schließlich mit den Ground-Truth-Daten verglichen und in einem Excel-File dokumentiert (vgl. *test\_statistik.xlsx*). Die untenstehende Graphik visualisiert die Testergebnisse.



Auffällig ist, dass das Tool im Identifizieren von Erzählperspektiven deutlich besser performt als in der Analyse der Fokalisierungen. Im Schnitt wurden 69,09 % der Beispiele richtig erkannt. Betrachtet man nur die richtigen Ergebnisse bei der Bestimmung von Perspektiven, so lässt sich sogar eine 85,86%ige Genauigkeit festmachen. Bei der Bestimmung der Fokalisierung sind es jedoch nur 52,32 %, was vermutlich an der ohnehin komplexen Identifizierung von Fokalisierungsmerkmalen liegt. Anscheinend hat das Modell Schwierigkeiten im Umgang mit Uneindeutigkeit bzw. Vieldeutigkeit. Auch die bereits angesprochene Kontext-Problematik führte zu den unbefriedigenden Testergebnissen. Es zeigt sich hier erneut, dass für die korrekte Analyse von Fokalisierungen eine verstärkte Kontextualisierung notwendig ist.

Abschließend muss allerdings erwähnt werden, dass der Ground-Truth-Datensatz, mit dem die Überprüfung des Tools durchgeführt wurde, selbst fehlerhaft bzw. fehleranfällig ist. Da viele Beispiele nicht eindeutig identifizierbar sind, kann auch die vollständige Validität der Ground-Truth-Daten nicht gewährleistet werden. Für eine gesteigerte Objektivität müssten die Ground-Truth-Daten von mindestens zwei Literaturwissenschaftler:innen ermittelt werden, sodass zumindest Vergleichswerte zur Verfügung stünden. Die Aussagekraft der Evaluation ist deshalb nicht vollständig gegeben. Es handelt sich hierbei um ein für die Geisteswissenschaften typisches Problem der Objektivierbarkeit. Aufgrund der Ambiguität der grundlegenden Daten kann keine gänzlich objektive Messung der Ergebnisse erfolgen. Die Hürden und Folgen, die sich dadurch ergeben, sollen unter anderem in meiner Masterarbeit diskutiert und reflektiert werden.

## 4. Links und Materialien

Materialien auf Github: [https://github.com/16meissl/ps\\_meissl\\_24](https://github.com/16meissl/ps_meissl_24)

Link zum Custom GPT: <https://chatgpt.com/g/g-Cnn4qSdeU-narrator-gpt>