

Do LLMs dream of antique hermeneutics? Critical remarks on automated text interpretation

Jens Dörpinghaus^{*†‡}, Michael Tiemann^{*†}

^{*} Federal Institute for Vocational Education and Training (BIBB), Bonn, Germany

[†] University of Koblenz, Germany

[‡] University of Pretoria,

Email: jens.doerpinghaus@bibb.de, tiemann@bibb.de

Abstract—This study investigates the potential of large language models (LLMs) to apply hermeneutical methods rooted in philosophy, theology, sociology and literary studies in a meaningful manner. Utilising a comparative experimental design, four LLMs were prompted to interpret a variety of texts, encompassing religious, philosophical, poetic, and conversational material. The findings indicate considerable variability, an absence of reproducibility, and a substantial reliance on prompt design, model type, and language. This suggests that LLMs do not employ coherent hermeneutical strategies. Despite the automation of formal features, limitations in context sensitivity, interpretive intentionality, and epistemic grounding render LLMs ill-suited to authentic hermeneutics. The study concludes that current LLMs are incapable of replicating the depth of human interpretive practice, and calls for further interdisciplinary research to define evaluation standards for machine-assisted exegesis.

I. INTRODUCTION

IT IS evident that all perception and self-perception are founded on acts of interpretation, extending beyond the mere reception of auditory and written statements. The term ‘hermeneutics’ generally encompasses theories that address the possibility, necessity, conditions, challenges, and objectives of processes involving understanding and interpretation. Thus, it can also be seen as a ‘theory of understanding’ or ‘theory of interpretation’, see [1]. Despite the numerous endeavours to define hermeneutics, a comprehensive consensus remains elusive. Maier has highlighted two predominant tendencies in the understanding and definition of hermeneutics, see [2]. Firstly, a rather mechanical tendency, and secondly, a definition of understanding and comprehension that is more common in German-speaking countries. In theology, this is an essential basis for exegesis, i.e. the interpretation of texts.

In antiquity and the Middle Ages, hermeneutics served as the science and art of interpreting fundamental texts, especially the Bible and laws. In modern times, its field of application has expanded, developing into a general doctrine of the prerequisites and methods of proper interpretation and a philosophy of understanding, see [3], [2]. Hermeneutics are also widely discussed and applied in areas like study of literature, psychology, social sciences, history and medical research.

It is important to note that hermeneutical concerns are ubiquitous. In the analysis of data (exegesis), a certain form of data hermeneutics is essential, and communication is pervasive.

Over the past several years, a new generation of generative AI tools has emerged. In Section II, we will demonstrate that researchers are concerned about which hermeneutics should be applied to AI-generated texts, or which tasks of hermeneutics can be carried out by LLMs. However, there is a paucity of research in this area, with only a small number of studies examining the hermeneutical capabilities of LLMs and generative AI in general.

The central research question guiding our study is as follows: Can contemporary LLMs apply hermeneutical approaches from Philosophy, Sociology, study of literature and Theology to textual data? Our research is particularly interested in methodological and technical reflections, while also raising novel questions that will serve as a foundation for a more robust theoretical framework in this field.

II. LITERATURE REVIEW

Some researchers argue to use hermeneutics on AI-generated texts [4], others propose hermeneutics as tool to optimize prompts [5]. However, there is a clear lack of hermeneutics for AI-generated texts as recognized in research [6], [7], [8]. While some researchers are generally positive on applying existing theory, e.g., by Heidegger or Lacan, to LLMs, see for example [9], there is no consent on this issue [10]. In general, researchers from non-technical domains seem to overestimate the features and capabilities of LLMs and AI in general, see for example [11]. Other researchers highlight the need for domain knowledge, for example in Bible teaching in confessional religious education using LLMs [12]. However, as our main research question is mainly concerned with whether LLMs can perform hermeneutical tasks, we will now turn to literature on this.

According to our best knowledge, there is only one preprint paper addressing this question directly [13] and being critical about hermeneutical and understanding capacities of LLMs. Other research focuses more on technical limitations or the correct place of LLMs, for example by seeing LLMs as “AI-interns” as [14] proposes. Another paper by [15] focuses on the theological and ethical bias of LLMs for interpreting texts from the Hebrew Bible and finds a bias which aligns with “contemporary socio-political and environmental concerns”. Some scholars work on hermeneutics within the medical domain. For example, Hans-Georg Gadamer [16] argues that

LLMs cannot take over patient interaction as it cannot take over moral practices. In general, we see the aspect of human interaction and experience as one of the key challenges in using LLMs for hermeneutical aspects. Although the aspect of machine relationships complementing human-human interactions was suggested [17] this does not solve any hermeneutical questions.

III. METHODOLOGICAL APPROACH

We will now discuss several methodological approaches from different domains. They all share a common foundation. For example, Körtner describes the reader as a central aspect of Biblical hermeneutics [18]. However, he also points at the connection between narrative approaches from literature and Biblical exegesis. Oeming describes some confusion about this approach and its naming. In general, we find this method as “New (Literary) Criticism, Literary Approach/Criticism, Holistic Approach, Rhetorical Criticism, Close Reading, Synchronical Approach, Narratology und Narrative Criticism” [19]. This approach tries to determine how biblical texts can be interpreted as literature, see [20]. In summary, “the literary critic looks at the text for what *it* says in itself by means of the patterning or shaping—the informing—of its content.” [21] Thus, we will – after a philosophical approach – focus on a selection of theological approaches. This selection must be somewhat arbitrary, but we tried to bridge between the reformators Zwingli and Calvin and a rather contemporary approach, liberation theology, as a contextual theology. However, as there is a clear link to narrative analysis, we will also discuss narrative approaches [1].

There is also some discussion between hermeneutics and sciences. Wyss, for example, discussed about the experience of time and self [22]. Other scholars discuss the world of living (‘Lebenswelt’) as paradigm for hermeneutics in the world of qualitative and quantitative research [23]. Thus we include a hermeneutics from the domain of sociology, namely objective hermeneutics.

Objective Hermeneutics, sometimes also labelled structural hermeneutics, derives from cultural anthropology. [24] discusses its connections to theories by Berger and Luckmann, Mead, and Chomsky. The idea is that how entities see the world they live in is reflected in the speech. On an individual level this would relate to an expression of one persons perception of reality, its ordering and the connections between entities. Within sociological theory there are also broader levels of groups of persons or organisations (meso level) up to the level of states (macro level). The structures that can be found by applying objective hermeneutics can be communicated intersubjectively and are answers to objective problems of action [24, p. 12]. For every structure there are numerous situations that pose problems, or crises. In fact, every single action is at first a situation of crises. The structure devises some solution and tests it. The way to re-construct structures within objective hermeneutics is thus interesting in two ways in view of its application to LLMs. The first interesting aspect is a “layer of normality” (“Normalitätsfolie”, [25, p. 136]).

This is one of the starting points for reconstructing a structure and refers to ideal types of normal behaviour. The second interesting aspect refers to the logic of cognition. In objective hermeneutics researchers explicitly use abductive reasoning. This can be seen as an inference to the best explanation while at the same time only having assumptions about the laws and rules from which to infer. This is seen as a creative process and might as such be particularly challenging for LLMs. For the inclusion of objective hermeneutics in this experiment this leads to some assumptions: If the LLMs are capable of applying an objective hermeneutics approach, they should all come to the same conclusion. Even though generally structures are reconstructed, these all refer to specific situations that find expressions in text. Thus, characters, places, times, etc. of these situations can also be reconstructed and should be comparable between different analysts.

The interest in narrative and ‘story’ began with Richard Niebuhr in 1941, see [26]. After that, it was introduced in the 1970s¹. While some scholars focused on ‘text-oriented approaches’, ‘reader-oriented approaches’ were also widely used. What they all have in common, however, is the usage of methods of criticism studies for the interpretation of narrative texts [29]. Different schools exist which are not only committed to different literary theories, but are also connected to different geographical areas, see [19] and in particular [30]. In narrative exegesis, the first step is to apply methods of text analysis: “The ‘what’ of a text (its content) and the ‘how’ of a text (its rhetoric and structure) are analyzed as a complete tapestry, an organic whole.” [20] For this purpose, the text is examined with respect to its content layer and effect layer to find out which structure, order and content the text wants to convey. In order to refine the exegesis, different elements of the text are examined: time, locations, characters, perspective, plot or plot line, narrator, listeners, and readers, key words, repetitions and other structural elements and stylistic devices.

Narrative approaches can be applied to any narrative texts – religious texts, for example Biblical texts, are widely considered to be suitable. Bailey et al. argue: “A Gospel is a narrative, fashioned out of selected traditions, that focuses on the activity and speech of Jesus as a way to reveal his character and develops a dramatic plot that culminates in the stories of his passion and resurrection.” [31] Other scholars emphasize the need to work with historical-critical methods, see for example [32] and [33]. Narrative approaches are also widely applied to biblical and theological texts [34], [35], [36], so the quintessential next step is to also consider hermeneutics in Theology, for example liberation theology [37] or the Swiss reformers Ulrich Zwingli and John Calvin [38], [39], [40], [41].

Despite having elected to examine only a limited number of approaches, it has become evident that, upon thorough discussion of the various hermeneutical approaches that have been posited, it can be concluded that all of them exhibit features that cannot be directly substituted with concepts that are quantifiable. However, our study reveals two fundamental

¹See also [27] and the overview by [28].

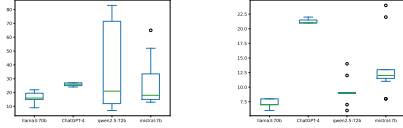


Fig. 1: Number of sections detected (direct) in Text 1 (left) and 2 (right)

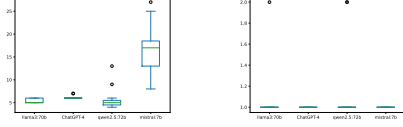


Fig. 2: Number of sections detected (direct) in Text 3 (left) and 4 (right)

aspects that all hermeneutics share: first, an understanding of human nature, and second, an understanding of language. While the question of whether AI can comprehend human nature remains a subject of ongoing discourse, the focus will be on the understanding of language. Danner summarizes: Hermeneutics is the study of interpretation, with a focus on comprehending meaning. The concept or phenomenon that is being understood is inherently present within the linguistic medium or can be articulated through it. The fundamental question, therefore, is: What is the origin of language? The fundamental question, therefore, is: what does precede language? The fundamental nature of language is predicated on its capacity to articulate elements of reality, that is to say, to “speak” of a reality that is already “out there” in the world. LLMs are unable to substitute the following aspects due to their inability to comprehend the context of reader and writer in a broader sense: (a) The reader and his perception or the listeners (unless we define LLMs as reader) (b) The context of a reader (unless we define LLMs as reader and define its context) (c) The author or narrator (d) With respect to both: Creator and created. Or the influence of the Holy Spirit (per Definition, as Luther and Zwingli described it), or Christ as hermeneutical principle (Calvin). In summary, it is evident that LLMs can only be utilized for the automation of fundamental principles of language understanding. However, as evidenced by our study of various hermeneutical approaches, different hermeneutics emphasize distinct aspects of language. Nevertheless, it is possible to identify several formalized structures that are, or may be, studied by hermeneutics. These include: (a) Plot as a broader concept, e.g., by selecting sections or units of meaning (b) Characters, e.g., the directly or indirectly acting persons (c) Space, while no single definition can be given. For example, according to Lefebvre and Soja [42] space can be defined as follows: “(Social) space is a (social) product.” [43]. They differ between the *firstspace*, which includes physical space, the *secondspace* as mental space, and the *thirdspace* as social space. This, in turn, redirects the focus back to the question of which social actor a LLM is. (d) The concept of perspectives or points of view is also crucial in this analysis.

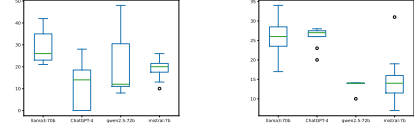


Fig. 3: Number of actors detected (direct) in Text 1 (left) and 2 (right)

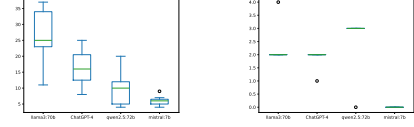


Fig. 4: Number of actors detected (direct) in Text 3 (left) and 4 (right)

However, it is crucial to acknowledge that an individual’s understanding of points of view is significantly influenced by the context of the reader. This reiterates the necessity for a comprehensive examination of the social actor in question.

Consequently, an initial approach to assess the capabilities of LLMs in applying hermeneutics entails a concentration on plot and characters, as this method facilitates the introduction of objectivity to the inquiry. It is imperative to establish the following assumptions: The assumption is made that a single LLM is capable of interpreting a text employing a specific hermeneutical approach. Conducting a repeated experiment should, under these conditions, yield a consistent result or at least a reasonably proximate outcome. Secondly, it is hypothesized that different models will yield analogous results. The number of actors in a text is a subject that merits discussion; however, especially in narrative texts, this number should be nearly fixed. Furthermore, if two models yield divergent units of meaning or sections, they are likely to employ different perspectives and, consequently, divergent hermeneutical methods.

Our approach is as follows: We will study four different texts from various backgrounds and repeatedly query a LLM ($n = 10$) to select sections or units of meaning in a text and to return the number of actors in a text. For a detailed exploration of different prompting strategies, we refer to Appendix A.

IV. EXPERIMENTAL RESULTS

To that end, the present study will examine four different texts. Initially, we examine a renowned German poem entitled “An die Freude” (“Ode to Joy”, 1785) by Friedrich Schiller. Subsequently, we examine a biblical text comprising three narratives (Acts 8:1-25), having both a narrative and theological dimensions. Finally, we turn to a section of the philosophical text by the German philosopher Immanuel Kant, titled “Beantwortung der Frage: Was ist Aufklärung?” (Answering the Question: What Is Enlightenment?, 1784). The final text is a concise sentence: “Guten Abend, meine Damen und Herren” (Good evening, ladies and gentlemen).

Reconstructing a structure from one sentence with objective hermeneutics is challenging, but entirely possible. The least

TABLE I: Details of detected sections with direct approach on Texts 1-4 with different languages (G=German, E=English)

L	Text	Model	Min	Q_1	Median	Q_3	Max
G	Text 1	llama3:70b	9.0	15.0	16.0	19.5	22.0
		ChatGPT-4	24.0	25.0	26.0	27.0	27.0
		qwen2.5:72b	7.0	12.0	21.0	71.5	83.0
		mistral:7b	13.0	15.0	18.0	33.0	65.0
G	Text 2	llama3:70b	6.0	7.0	7.0	8.0	8.0
		ChatGPT-4	21.0	21.0	21.0	21.5	22.0
		qwen2.5:72b	6.0	9.0	9.0	9.0	14.0
		mistral:7b	8.0	11.5	12.0	13.0	24.0
E	Text 2	llama3:70b	8.0	8.0	10.0	18.5	24.0
		ChatGPT-4	7.0	8.5	10.0	10.0	10.0
		qwen2.5:72b	5.0	5.0	6.0	6.0	6.0
		mistral:7b	18.0	18.0	18.0	18.5	19.0
G	Text 3	llama3:70b	5.0	5.0	5.0	6.0	6.0
		ChatGPT-4	6.0	6.0	6.0	6.0	7.0
		qwen2.5:72b	4.0	4.5	5.0	5.5	13.0
		mistral:7b	8.0	13.0	17.0	18.5	27.0
E	Text 3	llama3:70b	4.0	4.0	4.0	7.0	28.0
		ChatGPT-4	4.0	4.0	4.0	4.0	5.0
		qwen2.5:72b	4.0	4.0	4.0	5.0	5.0
		mistral:7b	17.0	17.0	17.0	20.0	20.0
G	Text 4	llama3:70b	1.0	1.0	1.0	1.0	2.0
		ChatGPT-4	1.0	1.0	1.0	1.0	1.0
		qwen2.5:72b	1.0	1.0	1.0	1.0	2.0
		mistral:7b	1.0	1.0	1.0	1.0	1.0

TABLE II: Details of detected sections with SE approach on Texts 2-3 with different languages (G=German, E=English)

L	Text	Model	Min	Q_1	Median	Q_3	Max
G	Text 2	llama3:70b	7.0	7.0	7.0	7.0	10.0
		ChatGPT-4	5.0	5.0	5.0	9.3	25.0
		qwen2.5:72b	4.0	4.3	6.0	10.0	23.0
		mistral:7b	5.0	11.0	12.0	13.0	19.0
E	Text 2	llama3:70b	5.0	5.0	6.0	6.0	6.0
		ChatGPT-4	5.0	5.0	5.0	5.0	6.0
		qwen2.5:72b	3.0	3.0	3.0	4.0	4.0
		mistral:7b	9.0	12.0	12.0	13.0	21.0
G	Text 3	llama3:70b	3.0	4.5	5.0	6.5	21.0
		ChatGPT-4	5.0	7.0	7.0	17.0	23.0
		qwen2.5:72b	3.0	4.0	4.0	6.0	14.0
		mistral:7b	6.0	9.0	12.0	15.5	23.0
E	Text 3	llama3:70b	3.0	7.0	8.0	9.0	52.0
		ChatGPT-4	3.0	4.0	4.0	4.0	5.0
		qwen2.5:72b	8.0	9.0	9.0	9.5	22.0
		mistral:7b	12.0	15.0	16.0	33.5	44.0

a model should arrive at is to reconstruct personen (one speaking, more than one being spoken to), possible places (a theater, a lecture room, a TV setting) and times (evening). There are at least two sequences within this sentence, on referring to people being addressed, the other to a time. There is a published example of such an analysis by Oevermann himself [44]. All texts can be found online and there are likely references to them being analysed with one or another of the included hermeneutics. Selecting LLMs to test is a another challenge. We followed [45] by choosing the most widely used ChatGPT-4 and llama3:70b which are widely compared and discussed, see e.g. [46], [47], [48]. In addition, we choose qwen2.5:72b as another large model and, to compare with a rather small model, mistral:7b.

A. Plot

As illustrated in Figures 1 and 2, the boxplots depict the range of detected sections within the text corpus. For a more detailed analysis, please refer to Table tab:sections. It is noteworthy that the ChatGPT 4 and llama3 models demonstrate the greatest consistency in their output. However, for text 2, which comprises three stories, the output is particularly

TABLE III: Details of found actors with direct approach on Texts 1-4 with different languages (G=German, E=English)

L	Text	Model	Min	Q_1	Median	Q_3	Max
G	Text 1	llama3:70b	21.0	23.0	26.0	35.0	42.0
		ChatGPT-4	0.0	0.0	14.0	18.5	28.0
		qwen2.5:72b	8.0	11.0	12.0	30.5	48.0
		mistral:7b	10.0	17.5	20.0	21.5	26.0
G	Text 2	llama3:70b	17.0	23.5	26.0	28.5	34.0
		ChatGPT-4	20.0	26.0	27.0	27.5	28.0
		qwen2.5:72b	10.0	14.0	14.0	14.0	14.0
		mistral:7b	7.0	11.5	14.0	16.0	31.0
E	Text 2	llama3:70b	14.0	15.0	17.0	23.0	26.0
		ChatGPT-4	9.0	10.5	13.0	18.0	21.0
		qwen2.5:72b	9.0	9.0	9.0	9.0	9.0
		mistral:7b	13.0	13.0	13.0	13.0	13.0
G	Text 3	llama3:70b	11.0	23.0	25.0	34.0	37.0
		ChatGPT-4	8.0	12.5	16.0	20.5	25.0
		qwen2.5:72b	4.0	5.0	10.0	12.0	20.0
		mistral:7b	4.0	5.0	6.0	6.5	9.0
E	Text 3	llama3:70b	2.0	3.5	5.0	5.5	7.0
		ChatGPT-4	7.0	9.0	9.0	10.5	12.0
		qwen2.5:72b	2.0	3.5	4.0	6.0	12.0
		mistral:7b	1.0	4.0	6.0	6.5	7.0
G	Text 4	llama3:70b	2.0	2.0	2.0	2.0	4.0
		ChatGPT-4	1.0	2.0	2.0	2.0	2.0
		qwen2.5:72b	0.0	3.0	3.0	3.0	3.0
		mistral:7b	0.0	0.0	0.0	0.0	0.0

problematic, though it should be noted that no model partitions every verse (25). This phenomenon is also evident in Text 4, which essentially comprises a single sentence, particularly in instances where two sections have been identified. In such instances, the LLM has demonstrated an inability to distinctly delineate the boundaries of the two sections. Consequently, the outcomes for the poem "An die Freude" and Kant's text appear somewhat arbitrary. The discrepancy in the analysis of different languages, as demonstrated in Text 2 and Text 3, is equally perplexing. These alterations, far from being mere nuances, introduce divergent interpretations.

The inconsistency of models in their own perspectives is evident, and the inability to compare different models is apparent. The application of a hermeneutical perspective appears to be both unclear and erratic. This discrepancy persists when employing an alternative prompting strategy, as evidenced in Table II. However, the number of suggested sections does undergo a substantial change. However, the value ranges do not offer any possible explanations within the text itself. In summary, three factors appear to influence the output: the model, the prompt, and the language.

B. Actors

As demonstrated in Figures 3 and 4, the boxplots illustrate the range of detected actors within the text corpus. For a more detailed analysis, please refer to Table [tab:personen]. The ensuing results proffer illuminating insights from a variety of perspectives. For instance, it is noteworthy that Texts 3 and 4 do not explicitly mention any actors, yet they do implicitly refer to actors through the use of language such as "Ladies and Gentlemen." However, the models detect a multitude of actors. The detection of more than 35 actors in a philosophical text is substantial, and the identification of four actors in a single sentence of greeting appears implausible. It is noteworthy that models can exhibit disagreement in their estimation of the number of actors mentioned in a narrative text or poem (texts

TABLE IV: Details of found actors with SE approach on Texts 2-3 with different languages (G=German, E=English)

L	Text	Model	Min	Q ₁	Median	Q ₃	Max
G	Text 2	llama3:70b	22.0	29.0	37.0	43.0	53.0
		ChatGPT-4	35.0	26.0	26.0	36.0	36.0
		qwen2.5:72b	14.0	18.5	19.0	19.0	34.0
		mistral:7b	8.0	16.0	18.0	25.0	34.0
E	Text 2	llama3:70b	24.0	27.5	31.0	34.5	47.0
		ChatGPT-4	23.0	27.5	32.0	35.0	36.0
		qwen2.5:72b	36.0	36.0	36.0	36.0	36.0
		mistral:7b	27.0	28.5	30.0	34.0	34.0
G	Text 3	llama3:70b	14.0	23.5	25.0	34.0	44.0
		ChatGPT-4	5.0	7.5	9.0	9.5	21.0
		qwen2.5:72b	3.0	8.0	20.0	27.5	35.0
		mistral:7b	6.0	15.0	18.0	19.0	23.0
E	Text 3	llama3:70b	1.0	1.0	1.0	2.0	26.0
		ChatGPT-4	8.0	9.5	10.0	11.5	12.0
		qwen2.5:72b	10.0	10.0	10.0	10.0	10.0
		mistral:7b	1.0	1.0	1.0	2.5	3.0

1 and 2). It is important to note that modifying the input language can substantially alter the output.

The lack of comparability and inconsistency of these models to other LLMs, as previously discussed in the context of plot, further complicates the analysis. It is evident that modifying the prompting strategy leads to alterations in the output, as illustrated in Table IV. It is important to note that none of these numbers are directly connected to the text itself. LLMs exhibit deficiencies in addressing fundamental hermeneutical elements that are essential for text exegesis.

V. CONCLUSIONS AND OUTLOOK

This study critically examined the extent to which large language models (LLMs) are capable of engaging in hermeneutical analysis, drawing from traditions in philosophy, theology, sociology, and literary theory. The present study evaluated the consistency, interpretive depth and methodological coherence of LLM outputs by testing multiple models on a range of textual genres, including poetic, theological, philosophical and conversational texts. The findings indicate that current LLMs exhibit significant variability and lack reproducibility in the application of hermeneutical principles. Across a range of models and prompting strategies, the interpretation of plot structures and identification of actors demonstrated inconsistency, both within and across iterations. This variability suggests that LLMs are not applying coherent hermeneutical frameworks, but rather relying on surface-level linguistic patterns that diverge significantly depending on language, prompt design, and model architecture.

Nevertheless, this is at odds with some of the recent literature that has been published, which has asserted that LLMs “can analyze nearly any textual statement.” [49] The experimental results obtained in this study appear to be in alignment with the arguments set forth by researchers who posit that LLMs are deficient in their capacity to perform even rudimentary logic-based tasks, such as counting and identifying general substructures in graphs [50], [51]. Therefore, it is debatable whether LLMs could offer any technical assistance with textual analysis, against, for example [52].

It has been demonstrated that there is a discernible discrepancy between certain components of LLM output and

hermeneutical procedures. These components include, but are not limited to, text segmentation and the identification of narrative elements. The models demonstrate an inability to comprehend text, see [53], and their apparent lack of understanding of context, intentionality, and reader-writer dynamics. These components are fundamental to traditional hermeneutics. In other words: “Should our interpretation capabilities be engaged? If yes, under what conditions? The rules of the language game should be spelled out; they should not be passed over in silence.” [54] Moreover, the validity of conventional hermeneutical principles is also open to question. For instance, the assumption that hermeneutics necessitates the interpretation of a human being with consciousness, engaging with interpretive dimensions involving principles, historical consciousness, praxis, or – in the context of theological hermeneutics – the influence of the Holy Spirit, is not necessarily indisputable. The assumption that an LLM could act as an exeget is not, in itself, sufficient to ensure that it will be able to identify all the important aspects of context necessary for the application of hermeneutics. It appears improbable that LLMs will ever be capable of this [55].

It is recommended that future work concentrate on the development of evaluation criteria for the purpose of assessing interpretive coherence in the context of machine-generated text analysis. This is a common debate also in the field of AI-generated text detection, see [56]. Furthermore, it is recommended that interdisciplinary frameworks be explored, with a view to more accurately reflecting the limitations of computational models within the domain of humanities or theological understanding. Moreover, the integration of insights from phenomenology and semiotics has the potential to provide supplementary conceptual foundations for the assessment of interpretive capacities of AI systems.

APPENDIX A PROMPTS

In the process of designing the prompts, adherence was observed to the general principle for success when engineering a prompt with clear and concise, step-by-step reasoning, as outlined in [45]. The direct approach posed a question that was explicit in its formulation. The ‘Single Explanation’ (SE) strategy is a rhetorical device that functions to provide an illustration of a concept or idea.

REFERENCES

- [1] Danner, Helmut, *Hermeneutik: Zugänge, Perspektiven, Positionen*, 1. Aufl ed. Freiburg: wbg Academic, 2021.
- [2] G. Maier, *Biblische Hermeneutik*. SCM, 2014.
- [3] M. Oeming, *Biblische Hermeneutik*. WBG, 2013.
- [4] R. Maciag, “The hermeneutics of artificial text,” in *Computer Sciences & Mathematics Forum*, vol. 8, no. 1. MDPI, 2024, p. 94.
- [5] L. Henrickson and A. Meroño-Peñuela, “Prompting meaning: a hermeneutic approach to optimising prompt engineering with chatgpt,” *AI & SOCIETY*, pp. 1–16, 2023.
- [6] L. Tiantian and M. Carl, “Toward practical hermeneutics of fourth paradigm ai for science,” *«Technology and language»*, vol. 5, no. 1, pp. 89–105, 2024.
- [7] K. Petković, “The truth and ai: Towards a new hermeneutic of the academic subject?” *Sociologija*, vol. 66, no. 4, pp. 581–611, 2024.

- [8] P. Pinell, "Does artificial intelligence speak our language?: A gadamerian assessment of generative language models," *Political Research Quarterly*, vol. 77, no. 3, pp. 713–728, 2024.
- [9] M. Heimann and A.-F. Hübener, "Circling the void: Using heidegger and lacan to think about large language models," *Cognitive Systems Research*, p. 101349, 2025.
- [10] A. Diržytė, "Large language models and the enhancement of human cognition: Some theoretical insights," *Filosofija. Sociologija*, vol. 36, no. 1, 2025.
- [11] M. Barnard and W. Otte, "Is the machine surpassing humans? large language models, structuralism, and liturgical ritual: A position paper," *International Journal of Practical Theology*, vol. 28, no. 2, pp. 289–306, 2024.
- [12] M. Chrostowski and A. J. Najda, "Chatgpt as a modern tool for bible teaching in confessional religious education: a german view," *Journal of Religious Education*, pp. 1–20, 2024.
- [13] R. Demicheli, "The hermeneutic turn of ai: Is the machine capable of interpreting?" *arXiv preprint arXiv:2411.12517*, 2024.
- [14] M. Gonzalez, "Beyond replacement: A hermeneutic phenomenological approach to human-ai collaboration," in *2024 IEEE International Symposium on Technology and Society (ISTAS)*. IEEE, 2024, pp. 1–4.
- [15] A. Elrod, "Uncovering theological and ethical biases in llms: An integrated hermeneutical approach employing texts from the hebrew bible," *HIPHIL Novum*, vol. 9, no. 1, pp. 2–45, 2024.
- [16] K. Karches, "Hermeneutics as impediment to ai in medicine," *Theoretical Medicine and Bioethics*, pp. 1–19, 2025.
- [17] E. Krapp, R. Neuhaus, M. Hassenzahl, and M. Laschke, "In a quasi-social relationship with chatgpt. an autoethnography on engaging with prompt-engineered llm personas," in *Proceedings of the 13th Nordic Conference on Human-Computer Interaction*, 2024, pp. 1–16.
- [18] Körtner, Ulrich H. J., *Der inspirierte Leser: zentrale Aspekte biblischer Hermeneutik*. Göttingen [München]: Vandenhoeck und Ruprecht [Bayerische Staatsbibliothek], 1994.
- [19] M. Oeming, *Biblische Hermeneutik: Eine Einführung*. Darmstadt: WBG, 2007.
- [20] J. Resseguie, *Narrative Criticism of the New Testament: An Introduction*. Grand Rapids: Baker Publishing Group, 2005.
- [21] N. R. Petersen, *Literary Criticism for New Testament Critics*. Minneapolis: Fortress Press, 1978.
- [22] Wyss, Dieter, *Zwischen Logos und Antilogos: Untersuchungen zur Vermittlung von Hermeneutik und Naturwissenschaft*. Göttingen [München]: Vandenhoeck & Ruprecht, 1980.
- [23] Schröer, Norbert and Bidlo, Oliver, "Die Entdeckung des Neuen: Qualitative Sozialforschung als Hermeneutische Wissenschaftssoziologie." Wiesbaden: VS Verlag für Sozialwissenschaften / Springer Fachmedien Wiesbaden GmbH, Wiesbaden, 2011, conference.
- [24] U. Oevermann, "Zur analyse der struktur von sozialen deutungsmustern," 1973. [Online]. Available: <https://publikationen.ub.uni-frankfurt.de/opus4/frontdoor/deliver/index/docId/4951/file/Struktur-von-Deutungsmuster-1973.pdf>
- [25] J. Reichertz, "Die objektive hermeneutik – darstellung und kritik," in *Qualitative Forschung*, E. König and P. Zedler, Eds. Beltz Verlag, 2002, pp. 123–156.
- [26] H. Niebuhr and D. Ottati, *The Meaning of Revelation*. Westminster: John Knox Press, 2006.
- [27] D. Ritschl and H. O. Jones, *Story als Rohmaterial der Theologie*. München: Kaiser, 1976.
- [28] B. Wacker, *Narrative Theologie?* München: Kösel, 1977.
- [29] G. Genette, *Figures III*. Paris: Éditions du Seuil, 2014.
- [30] A. Cornils, *Vom Geist Gottes erzählen: Analysen zur Apostelgeschichte*. Tübingen: Francke, 2006.
- [31] J. Bailey and L. Vander Broek, *Literary Forms in the New Testament: A Handbook*. Westminster: John Knox, 1992.
- [32] N. Perrin, "Historical criticism, literary criticism and hermeneutics: The interpretation of the parables of jesus and the gospel of mark today," in *Hermeneutical Inquiry*, D. E. Klemm, Ed. Atlanta, Georgia: Scholars Press, 1986, pp. 253–266.
- [33] J. A. Darr, "'Watch How you Listen' (Lk. 8:18): Jesus and the Rhetoric of Perception in Luke-Acts," in *The New Literary Criticism and the New Testament*, ser. JSNTS, E. S. Malbon and E. V. McKnight, Eds. Sheffield: Academic Press, 1994, vol. 109, pp. 87–107.
- [34] B. Dörpinghaus, "Die narratologische exegese alttestamentlicher erzähltexte: Neue perspektiven auf die jakobserzählung," *European journal of theology*, vol. 31, no. 2, pp. 186–214, 2022.
- [35] B. Dörpinghaus and H.-G. Wüch, "Relationships and forms in the social network of the jacob narrative: A narratological perspective," *Old Testament Essays*, vol. 36, no. 2, pp. 347–367, 2023.
- [36] B. Dörpinghaus, "Esau und jakob: Familien und figuren im wandel?" *European Journal of Theology*, vol. 32, no. 2, pp. 180–201, 2023.
- [37] B. Kern, *Theologie der Befreiung*. UTB, 2013.
- [38] G. W. Locher, *Zwingli und die schweizer Reformation*, ser. Die Kirche in ihrer Geschichte. Vandenhoeck & Ruprecht, 1982.
- [39] V. Leppin, "Zwingli, Ulrich," in *Theologische Realenzyklopädie*, H. Balz, G. Krause, and G. Müller, Eds. Berlin: de Gruyter, 2004, no. Bd. 36.
- [40] W. van't Spijker, *Calvin*, ser. Die Kirche in ihrer Geschichte. Vandenhoeck & Ruprecht, 2001.
- [41] W. Nijenhuis, "Calvin, Johannes," in *Theologische Realenzyklopädie*, H. Balz, G. Krause, and G. Müller, Eds. Berlin: de Gruyter, 1981, no. Bd. 07.
- [42] E. Soja, *Thirdspace: Journeys to Los Angeles and Other Real-and-Imagined Places*. Wiley, 1996.
- [43] H. Lefebvre and D. Nicholson-Smith, *The production of space*. Oxford: Blackwell, 1991, vol. 142.
- [44] U. Oevermann, "Zur sache. die bedeutung von adornos methodologischem selbstverständnis für die begründung einer materialen soziologischen strukturanalyse," in *Adorno-Konferenz*, L. von Friedeburg and J. Habermas, Eds., 1983, pp. 234–289.
- [45] R. Fechner and J. Dörpinghaus, "No train, no pain? assessing the ability of llms for text classification with no finetuning," in *Proceedings of the Position Papers of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS)*, Belgrade, Serbia, 2024, pp. 8–11.
- [46] K. Bhattarai, I. Y. Oh, J. M. Sierra, J. Tang, P. R. Payne, Z. Abrams, and A. M. Lai, "Leveraging gpt-4 for identifying cancer phenotypes in electronic health records: a performance comparison between gpt-4, gpt-3.5-turbo, flan-t5, llama-3-8b, and spacy's rule-based and machine learning-based methods," *JAMIA open*, vol. 7, no. 3, p. ooae060, 2024.
- [47] P. Ersoy and M. Erşahin, "Benchmarking llama 3 70b for code generation: A comprehensive evaluation," *Orclever Proceedings of Research and Development*, vol. 4, no. 1, pp. 52–58, 2024.
- [48] M. Merkel and J. Dörpinghaus, "A case study on the transformative potential of ai in software engineering on leetcode and chatgpt," *arXiv preprint arXiv:2501.03639*, 2025.
- [49] P. Törnberg, "How to use llms for text analysis," *arXiv preprint arXiv:2307.13106*, 2023.
- [50] T. Fu, R. Ferrando, J. Conde, C. Arriaga, and P. Reviriego, "Why do large language models (llms) struggle to count letters?" *arXiv preprint arXiv:2412.18626*, 2024.
- [51] L. Nguyen and Y. Yan, "Evaluating the structural awareness of large language models on graphs: Can they count substructures?" *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24)*, 2024.
- [52] R. H. Tai, L. R. Bentley, X. Xia, J. M. Sitt, S. C. Fankhauser, A. M. Chicas-Mosier, and B. G. Monteith, "An examination of the use of large language models to aid analysis of textual data," *International Journal of Qualitative Methods*, vol. 23, p. 16094069241231168, 2024.
- [53] W. S. Saba, "Llms' understanding of natural language revealed," *arXiv preprint arXiv:2407.19630*, 2024.
- [54] A. Videla, "Echoes of intelligence: Textual interpretation and large language models," *Queue*, vol. 21, no. 3, pp. 36–53, 2023.
- [55] P. Villalobos, A. Ho, J. Sevilla, T. Besiroglu, L. Heim, and M. Hobbhahn, "Position: Will we run out of data? limits of llm scaling based on human-generated data," in *Forty-first International Conference on Machine Learning*, 2024.
- [56] J. Wu, S. Yang, R. Zhan, Y. Yuan, L. S. Chao, and D. F. Wong, "A survey on llm-generated text detection: Necessity, methods, and future directions," *Computational Linguistics*, pp. 1–66, 2025.