অভিসরণ / ABHISARON
A Multilingual International Peer-Reviewed Research Journal
Volume 1, Issue II, July 2026, PP. 96-102
www.abhisaron.com
ISSN-3139-4027(Online)
DOI - https://dx.doi.org/10.67757/abhisaron.2026.v1i2.013

Statistical Hermeneutics: The Epistemic Status and Collaborative Potential of Large Language Models in Comparative Literary Studies

Author: Vaskar Chakraborty, PhD (Ongoing)
Assistant Professor, Department of Computer Science & Application, Siliguri Institute of Technology (S.I.T.)
Co-author: Dr. Kamanasish Bhattacharjee
Assistant Professor, Department of Computer Science & Engineering, NIT Agartala
Email: vaskar.scholar@gmail.com

Accepted: 06.07.2026
Published: 20.07.2026

Keywords: large language models; statistical hermeneutics; comparative literature; literary interpretation; machine reading; multilingual NLP; digital humanities; Gadamer; AI ethics

Abstract:

In this chapter, I explore who the reading of a large language model is and how it is connected with not hermeneutic, but computational interpretations of literature. The chapter adopts and builds on the concept of hermeneutics in the work of Hans-Georg Gadamer and Paul Ricoeur, and follows recent developments in literary and translation studies using large language models (LLM) to propose a new term, namely: statistical hermeneutics, as a description of the mode of epistemic action that LLMs perform—in an approximate manner—and without grounding. The chapter illustrates the ability of LLMs as cross-language readers, as well as the challenges encountered, using examples from Serbian poetry, Arabic fiction (Naguib Mahfouz) and Uighur translation, and Malay literature. The chapter ends with a plea for an "inter-disciplinary and collaborative" human scholarship that the LLM can facilitate, but cannot replace, in comparative literary study, as a "sovereign interpreter" of it.

Discussion:

1. Introduction

How do you define a large language model as a reader? It seems nave, but it is not. What can I use a large language model for, if anything, to help me to understand a text, to move me, and to place it in the cultural sands of understanding that preceded any individual reader and will outlast them? When GPT 4 offers fluent commentary on Gitanjali, or the ironies in Kafka's The Metamorphosis, it looks like interpretation on paper. Large language models have been used in a huge range of literary tasks; it has truly rewritten the scope of computational literary studies can take (Hatzel et al., 2023), though academics have largely steered behind the practical discussion. Many papers are now testing whether the model performed against a human task without asking, where does the concept of interpretation make this comparison? If a model outscored a student on a metaphor task, the score is real, but that doesn't tell us what type of understanding was responsible for the score. For example, the work of an LLM is not a position that neither Gadamerian hermeneutics nor Franco Moretti's distant reading can take. The term I propose here is statistical hermeneutics. I argue that it explains how an LLM can produce an interpretive look to us through probabilistic completion of a vast training corpus of samples without the historical situatedness, cultural memory, or personal stakes that philosophical hermeneutics recognize as important when discussing the role these tools are applied without mistaking their reach for depth.

2. From Distant Reading to Machine Hermeneutics: A Methodological Genealogy

The transformer wasn't the first to usher in the dream of reading at scale. The idea of "distant reading", proposed by Franco Moretti, was that aggregating and visualising huge collections of texts can reveal structural patterns which not even the most attentive reader could discover by following one text at a time, before the advent of large language models (Moretti, 2013). It's important to keep in mind the meager statement Moretti was making. He didn't say that the computer was reading anything. Computation provided the overall perspective; the interpreting remained in the scholar's mind. The previous generation of natural language processing techniques continued this legacy of topic-modelling, sentiment lexicons, stylometry, word embeddings. The tools worked well to identify regularities in a corpus but were not very good at generating a fluent literary commentary of a paragraph, could not truly engage with a metaphor, and were very weak at a text crossing a language boundary (Hatzel et al., 2023). If they had any connection to Hermeneutics, it was preparatory and to that extent, they cleared the ground for a human reader.

This is dramatically different with large scale, instruction-tuned transformer models. They can write longer, contextually appropriate critical commentary on a poem; they can translate a text from one language to another; they can identify figurative construction and construct a plausible-sounding narration of a speaker's attitude. The question of epistemic status does not change, nor is it masked by the fluency, as the question of where does the information come from, how much do you believe it when you read it, or how much do you trust the source? It's not the same thing to produce commentary that sounds like literary criticism, as opposed to criticism.

.

3. Statistical Hermeneutics: A Theoretical Framework

3.1 The Hermeneutic Inheritance

Hans-Georg Gadamer's philosophical hermeneutics has the most rigorous account of how to interpret a text in literature. According to Gadamer, understanding is not a technique that is used on the text from the outside but it is a dialogue that is encountered where the two horizons of the individual interpreter and the text meet (Gadamer, 2004). In readership, prejudices don't contaminate the texts, but are the condition that makes the difference of texts as difference. A true interpretation means something that Gadamer calls a "fusion of horizons" — where neither the reader nor the text is changed by the encounter. The point of the risk is that it has been interpreted in the way it was, it has shaken up the reader's assumptions if it had not, it wasn't much of an interpretation. Paul Ricoeur's account complements this by referring to the process of understanding – a process that moves from the initial naive reading of the text, to an explanatory distance in which the text becomes an object of analysis, and finally to an appropriation in which the reader has been transformed by the detour of analysis, entering the world of the text a different person (Ricoeur, 1976). What Gadamer and Ricoeur both stress, in their different ways, is that 'interpretation occurs over a period of time, for someone, and makes an impact'. To change nothing in the reader is difficult to tell from a non-reading.

3.2 What LLMs Do: The Case for Statistical Hermeneutics

All of this is not literally what a large language model does. There is no phenomenological horizon, no horizon in Gadamer's sense, to the LLM; no historically situated subject which it is on the other side of the exchange with. Instead, it's a statistical representation of the vocabulary used to interpret the vast number of real readers—the millions—who have used it, in all kinds of languages, over a very long period of time. The comments of a model on Gitanjali's devotional ambiguity are not "interpretation" of the poem but rather "re-combination" of the prior commentators' comments on it, which are influenced by the context in which the poem is being read, the architecture that reads the poem, and the language of the training data used by the model. The term statistical hermeneutics is supposed to name this third thing, rather than to be forced into one or the other of the categories that we already have. It is hermeneutic, because its product is a comment, paraphrase, theme, affective inference, etc., which is an interpretation. It's not statistical because it's not the output of a learned distribution (which is what statistical is) but rather it's the output of the situated understanding that hermeneutic theory is describing. Not in the sense of Gadamer, it's not interpretation, nor is it simple information retrieval. It is a distinct entity, and has its own set of successes and shortcomings.

Empirical evidence to support this treatment. The comparison of four LLMs that apply hermeneutical approaches to religious, philosophical, poetic and conversational text revealed significant variation – both in regard to the LLM used, the wording of the prompt and the language in which it was formulated – that the average reader of religious texts would hardly recognize as a coherent interpretive strategy (Dörpinghaus & Tiemann, 2025). The attention mechanism within a transformer is a loose computation analogue of the “relevance-weighting” approach to reading used by humans: it highlights some tokens relative to a query, but without the evaluative or cultural assumptions that make something relevant to a human reader, reader-first (Hatzel et al., 2023). The model is answering the shape of the prompt rather than the shape of the text (more or less literally).

.

4. LLMs Reading Across Traditions: Cross-Cultural Case Analysis

4.1 Metaphor and the Limits of Cross-Linguistic Reading

One of the more difficult tests you can set a reader is the sensitivity for connotation, cultural reference, and the exact pragmatic conditions in which a figurative reading differs from the literal reading; meter is one of the easier ones. GPT-4 has come up trumps on this one compared to many. It improved its performance over previous models on the Fig-QA test set and in a test set developed around metaphors taken from Serbian poetry, translated into English, blind human judges assessed its performance as superior to that of the group of college students (Ichien et al., 2024). That's no small feat. Indicates that the model has learned a bit about the nature of figurative language. The more difficult question to answer is what sort of sensitivity this is. The things that GPT-4 misreads are the regularities of the language that people used to talk about, translate and paraphrase in a vast amount of training text, including literary criticism, reader commentaries and previous translations, etc. That is a very strong weapon for the Diablo 2.That's an actual kick-ass tool for the Diablo 2. It's not, however, the same as understanding the significance of a certain figurative gesture in a particular poetic tradition. The differences and distances increase as soon as we cross a language frontier seriously. A model that is required to perform the Bengali originals (devotional register for bhakti, the cultural significance of a phrase such as amar sonar bangla in Bengali nationalism, the musicality inherent in the prosody) will work with cultural knowledge that is poorly and unevenly captured in the training corpus, which is dominated by texts in English. In some way the model reads. What it reads is the tradition of its own training distribution that has been filtered, and so on.

4.2 Translation as Interpretation: Arabic, Uighur, and Malay Traditions

Metaphor tests comprehension, but translation tests closer to judgement; a translator must first interpret a passage before translating it into another language. The situation in a number of non-European literary traditions is more or less uniform: gains over the older machine-translation baselines, and a constant deficit relative to the expert human translators, just at the point where cultural density is greatest.The results of the comparison between ChatGPT and Google Translate were obtained on the basis of a number of passages taken from the novel Zuqāq al-Midaqq by Naguib Mahfouz. The findings indicated that ChatGPT was successful in reproducing the translation in passages that carried cultural connotation, but failed in reproducing the translation of passages with the socio-linguistic register of Mahfouz' Cairo vernacular and cultural references that the human translator would recognize instinctively (Al-Kaabi et al., 2024). A comparative study of Uighur literary translation showed that ChatGPT, particularly when given a concise instruction prompt, outperforms Google Translate and Bing Translator in terms of semantic accuracy, fluency and the study's definition of cultural adequacy, while Bing Translator had a significantly higher error rate in terms of outright errors (Wang, 2025). Another twist is that training corpora are more often than not biased towards English and a few dominant languages, after which, AI systems tend to perform structural mapping instead of semantic, while human translators have greater interpretative flexibility, allowing for the oral literary forms and dialects to be underrepresented (Anuar & Sabri, 2025). All these findings together are the picture that statistical hermeneutics is supposed to describe. If fluency and semantic coverage are well represented in the multilingual training data, LLM's improve best, at the surface. They fall short, not only in the vocabulary, but at the level of cultural meaning, at the level of a relation that is positioned to a tradition, the horizon that Gadamer was attempting to define: Document-level prompting reduces this gap by allowing the model to keep more context in sight, but does not eliminate the gap (Karpinska & Iyyer, 2023).

.

5. The Epistemic and Ethical Stakes of Machine Reading

Combine pieces and a differentiated picture appears. These are actually effective models of pattern recognition between languages, and of making "surface" literary commentary. They are still confined, just as hermeneutic theory holds, where the real work is done: at the level of historical situatedness, cultural embeddedness, and an encounter with difference in a text that makes it different from one's own experience, Gadamer says, which is what constitutes real understanding. Some additional issues need to be mentioned. The above instability in the interpretations (sensitivity to the wording, to the model adopted, to the language used) suggests that no stable interpretive strategy (Dörpinghaus & Tiemann, 2025) can be detected. Have a human critic read the same passage twice and you'll find some differences in emphasis but one always driven by a predictable set of commitments (theoretical, cultural, personal). An LLM exhibits none of such stability; in fact, it isn't a judgement-led reader as is a human. There is also a version of the hallucination problem specifically to literary context, and not only accuracy is the issue. A model is a plausible but unsubstantiated interpretative claim—about what does or does not exist in the mind of an author, about what does or does not exist in any culture, attributing an intention to an author, asserting a cultural significance based on nothing—when it does that, it might be called a misreading. Creative misreading by the strong poet, in Harold Bloom, has come to be viewed as a productive gesture that can create new meaning out of an earlier text, one's wilful departure from it. Machine misreading is not a deliberated deviation from a familiar tradition, but a confabulation under distributional pressure, the fluency with which it is performed being able to fool readers who don't have the background to detect it. Lastly, who are these models, and who does well know their traditions? The asymmetry above was observed in both the Arabic and Uighur translation as well as Malay translation (Al-Kaabi et al., 2024; Anuar & Sabri, 2025; Wang, 2025). In the training of corpora, English and a few European languages are still the majority, and integrating such systems into scholarly practices can unwittingly reenforce the Eurocentric hierarchies that comparative literature has worked so hard since the 1950s to eradicate. The statistical hermeneutics is not neutral; it has imbalances brought in by its training data.

6. Towards a Collaborative Hermeneutics

It turns out that the machine is a reader of a special and peculiar type. It reads with breadth hardly anyone can exceed, from one language, let alone one so different and so varied, and with a fluency and a profound, although limited, sense of figurative and structural pattern. What this chapter's argument suggests it is not able to gain within its present structure, is the historical situatedness, the cultural embeddedness, and the ability to be transformed by the difference of a text: the very content of understanding that philosophical hermeneutics considers to be the content. The presentation of statistical hermeneutics as an alternative to the usual approach of collapsing into one of the extremes is presented. It embraces what these models can accomplish, but does not accept the use of statistical pattern-completion as an alternative to interpreting the models. Such tools would be very useful for comparative literature studies, both cross-linguistic and cross-cultural, in which the lack of access to texts and archives in one language has always been a constraint on a single scholar. However, this benefit will only be realized with responsibility within structures that ensure that the human factor, accountability, and an awareness of the cultural politics of interpretation remain intact.

.

The issue now is not whether or not LLMs should be part of literary scholarship, but were. The issue is no longer whether or not to have large language models in literary scholarship, but were. These are already available. But what sort of scholarship do we construct around them one in which machine reach expands human scope of interpretation or one in which an interpretive authority by itself slips off our radar, its epistemic underpinnings still not sufficiently explored? This chapter has advanced the latter thesis. It reads, it interprets, and, if done with care, it may be a comparative literary study of a scope which we have not previously attempted.

7. Conclusion

The machine turns out to be a reader of a particular and quite unusual kind. It reads with a breadth no individual scholar could match, across languages and traditions, with real fluency and a genuine, if partial, sensitivity to figurative and structural pattern. What it lacks and what the argument of this chapter suggests it cannot acquire within its current architecture is the historical situatedness, the cultural embeddedness, and the capacity to be changed by an encounter with a text's difference that philosophical hermeneutics treats as the actual substance of understanding.Statistical hermeneutics is offered as a way of holding both halves of that assessment at once, without collapsing into either extreme. It takes seriously what these models can already do, while refusing to treat statistical pattern-completion as a stand-in for interpretive understanding. Comparative literary scholarship stands to gain a great deal from these capabilities, particularly in multilingual and cross-cultural work, where limits of language access and archival scale have always constrained what an individual scholar could take on alone. But that gain will only be realised responsibly inside frameworks that keep human oversight, theoretical accountability, and an alertness to the cultural politics of interpretation firmly in place.

The question the field faces now is not whether large language models belong in literary scholarship. They are already there. The real question is what kind of scholarship we build around them: one where machine reach extends the scope of human interpretation, or one where interpretive authority quietly migrates to systems whose epistemic foundations, we have not yet properly examined. This chapter has argued for the former. The machine reads; the scholar interprets; and if that collaboration is conducted with care, the result may be a comparative literary study of a scope we have not previously had the means to attempt.

References:

1. Al-Kaabi, M. H., AlQbailat, N., Badah, A., Ismail, I. A., & Hicham, K. B. (2024). Examining the cultural connotations in human and machine translations: A corpus study of Naguib Mahfouz's Zuqāq al-Midaqq. Journal of Language Teaching and Research, 15(3), 707–718. https://doi.org/10.17507/jltr.1503.03
2. Anuar, A. F., & Sabri, M. A. Z. (2025). Artificial intelligence and Malaysian literature: A comparative study of AI and human translation. International Journal of Academic Research in Progressive Education and Development, 14(4), 977–988. https://doi.org/10.6007/IJARPED/v14-i4/26790
3. Bloom, H. (1973). The anxiety of influence: A theory of poetry. Oxford University Press.

4. Dörpinghaus, J., & Tiemann, M. (2025). Do LLMs dream of antique hermeneutics? Critical remarks on automated text interpretation. In M. Bolanowski, M. Ganzha, L. Maciaszek, M. Paprzycki, & D. Ślęzak (Eds.), Proceedings of the 20th Conference on Computer Science and Intelligence Systems (FedCSIS) (pp. 687–692). Polish Information Processing Society. https://doi.org/10.15439/2025F2839
5. Dunivin, Z. O. (2025). Scaling hermeneutics: A guide to qualitative coding with LLMs for reflexive content analysis. EPJ Data Science, 14(1), Article 28. https://doi.org/10.1140/epjds/s13688-025-00548-8
6. Gadamer, H.-G. (2004). Truth and method (J. Weinsheimer & D. G. Marshall, Trans.; 2nd rev. ed.). Continuum. (Original work published 1960)
7. Hatzel, H. O., Stiemer, H., Biemann, C., & Gius, E. (2023). Machine learning in computational literary studies. it – Information Technology, 65(4–5), 200–217. https://doi.org/10.1515/itit-2023-0041
8. Ichien, N., Stamenković, D., & Holyoak, K. J. (2024). Large language model displays emergent ability to interpret novel literary metaphors. Metaphor and Symbol. Advance online publication. https://doi.org/10.1080/10926488.2024.2380348
9. Karjus, A. (2023). Machine-assisted mixed methods: Augmenting humanities and social sciences with artificial intelligence (arXiv:2309.14379). arXiv. https://doi.org/10.48550/arXiv.2309.14379
10. Karpinska, M., & Iyyer, M. (2023). Large language models effectively leverage document-level context for literary translation, but critical errors persist. In P. Koehn, B. Haddow, T. Kocmi, & C. Monz (Eds.), Proceedings of the Eighth Conference on Machine Translation (pp. 419–451). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.wmt-1.41
11. Moretti, F. (2013). Distant reading. Verso.
12. Ricoeur, P. (1976). Interpretation theory: Discourse and the surplus of meaning. Texas Christian University Press.
13. Wang, Q. (2025). Evaluating Uighur literary translation: A comparative study of ChatGPT, Google Translate, and Bing Translator. PLOS ONE, 20(10), Article e0335261. https://doi.org/10.1371/journal.pone.0335261