



Received: 07-05-2026
Accepted: 17-06-2026

ISSN: 2583-049X

Evaluating AI's Dependability in Academic Writing: A Comparison of Responses from the Fields of Literature and Linguistics

Nguyen Thi Tuyet Hanh

Department of English Language, Ho Chi Minh City University of Economics and Finance, Vietnam

Corresponding Author: **Nguyen Thi Tuyet Hanh**

Abstract

This research explores the reactions of AIs towards the persona of Molvi Jalal in Nafisa Rizvi's *The Blue Room* in the context of the Sapir-Whorf linguistic relativity hypothesis. The main concern is the investigation of the validity and robustness of the AI results. The results show that AI works well in linguistics, especially if it is working within known theoretical frameworks, but poorly in literary analysis. Furthermore, AI does not possess the nuance of literary interpretation and may misjudge the characters' personalities and other narrative components. The report also mentions that AI performs better in structured academic areas like linguistics, where exploration follows norms and frameworks. But it doesn't have sensitivity to cultural context, iconography, and emotion." The main limitations of

it are the inability to deal with unanalyzed or context-dependent information well, the weak understanding of emotional depth, and the dependence on existing datasets. The research shows that we need to use intelligence carefully when we are studying things that need to be understood in a certain way. Artificial intelligence can be a tool, and it can make language research better. When we use artificial intelligence to study literature, we need to make sure it is used correctly. This is why we need help from others to understand these concepts, because they have the right knowledge and experience to understand them correctly. AI is a tool; it's not a replacement for a human who knows literature.

Keywords: Artificial Intelligence, ChatGPT, Literature, Linguistics, Natural Language Processing, Academic Writing

1. Introduction

The rapid growth of technology has made it possible to develop intelligent artificial intelligence systems that have tremendously changed our lives. ChatGPT and other artificial intelligence systems are transforming the face of work, making it easier and faster to do tasks that used to take a lot of time and energy. They have had a big impact on many spheres, especially on education, where the students are increasingly using artificial intelligence systems. Artificial intelligence also has a major impact on the way students learn and complete their assignments. There are many advantages, but the use of AI has also raised serious concerns about its reliability, accuracy, and possible effects on cognitive skills. Students are particularly susceptible to turning to artificial intelligence (AI) to help them with their schoolwork, projects, and exam preparation. However, using AI systems can be harmful to the development of critical thinking and analytical reasoning skills. For instance, students tend to ignore the accuracy of the data provided by these systems and have undue faith in the results produced by these systems. (Pathak *et al.* 2024) ^[14]. The AI performance gap in academic disciplines implies the need for a more accurate and domain-specific AI model. We did a study to see how well AI systems work in two areas, language and literature. AI systems generally do poorly on literature comprehension, especially when asked to identify a book or write a book review. Our research has shown that AI systems have limitations in understanding the context and meaning of the symbols and words used in literature. AI systems have made great strides in many domains, such as education, but there are still many subjects in which they have much to learn. "We're not going to be there yet with AI systems," Amirjalili said in 2024. This difference between the working of AI systems in literature and language shows that AI systems are not yet perfect. We must tackle these issues and develop AI systems that are more suitable for each subject. It will enable us to use AI systems without losing the quality of human thought and conversation.

Statement of the Problem and Significance of the Study

The purpose of this research is to analyze the performance of AI systems in language and literature-oriented settings. Now, people get answers easily using AI systems like chatbots. One big gap is the missing comparisons of AI performance in other fields, like literature and languages. In a recent study, the attention has been mostly on the application of AI in particular sectors, but the differences of AI's strengths and shortcomings in other fields have not been explored extensively (Amirjalili *et al.*, 2024) ^[1]. This research can take advantage of this gap in the literature to study the domain-specific challenges encountered by AI models. And there is less research on the boundaries of AI in interpretive disciplines, such as literature, than in systematic fields such as languages. The present paper aims to fill this vacuum by providing a comprehensive overview of the performance of AI in the disciplines of linguistics and literature. We will identify problems in each area and propose changes to AI models to address these challenges. The value of this work is significant since it lays out the limitations of regularly used AI systems, particularly in academic environments. The primary objective of the study is to emphasize the significance of the adjustments to be made by the AI developers in the specific areas for higher accuracy of the AI output material. Moreover, it functions as a warning to students to critically assess the AI-generated output and not to take it at face value.

Research objectives

This work is motivated by the following objectives: (i) analyze the accuracy and utility of AI replies in literary analysis, focusing on character evaluation in Nafisa Rizvi's *The Blue Room*. (ii) Investigate AI's capability to produce accurate and relevant information on language theories, in particular, language relativity. (iii) study AI performance in literature and linguistics to identify factors affecting differences in accuracy and precision.

Research questions

The following research questions are drawn from the aims and form the basis of this study: How trustworthy are AI replies in analyzing literary characters, as shown by Molvi Jalal in *The Blue Room*?

Research Hypotheses

The research rests on the following hypothesis: language concepts are simple and precise, and this makes it more probable that artificial intelligence responses will be accurate and relevant in the field of language, as opposed to literature, which can be complex and open to interpretation. Nguyen (2020a, 2020b, 2020c; 2022; 2023a, 2023b, 2023c, 2023d; Nguyen, 2024a, 2024b, 2024c) ^[3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13]. I think artificial intelligence is very good at understanding language rules. Language is a set of codes we follow. Literature, on the other hand, is more concentrated on understanding the motives and feelings of people. This is a harder job for artificial intelligence. So, I think artificial intelligence will do better with language than with literature. The purpose of the research is to find out the following. The research explores the relation between artificial intelligence and language and literature. It also talks about why AI is good at one thing but not the other. This can lead to less accuracy in literary analysis of AI output, as complex characterizations and literary devices are

not easily understood. In linguistics, the clear and well-defined nature of language concepts helps AI performance. This comparative study will discuss AI's skills in the fields of literature and linguistics. It will show the advantages and disadvantages of AI in academic writing. The objective of this study is to understand the function and limitations of AI in a deeper way to provide recommendations for the development of AI models, especially for domain-specific applications.

Theory framework

The main idea behind this research is that AI gives more answers in linguistics than in literature. This is because linguistics deals with precise rules, whereas literature involves a lot of interpretation. Nguyen (2020a, 2020b, 2020c; 2022; 2023a, 2023b, 2023c, 2023d; Nguyen, 2024a, 2024b, 2024c) ^[3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13].

One big gap is the missing comparisons of AI performance in other fields, like literature and languages. In a recent study, the attention has been mostly on the application of AI in particular sectors, but the differences of AI's strengths and shortcomings in other fields have not been explored extensively (Amirjalili *et al.*, 2024) ^[1]. This research can take advantage of this gap in the literature to study the domain-specific challenges encountered by AI models. And there is less research on the boundaries of AI in interpretive disciplines, such as literature, than in systematic fields, such as languages. The present paper aims to fill this vacuum by providing a comprehensive overview of the performance of AI in the disciplines of linguistics and literature. We will identify problems in each area and propose changes to AI models to address these challenges.

Review of Literature

The effect of the use of artificial intelligence (AI). The study applies the theory of natural language processing (NLP), which was first proposed in the 1950s and 1960s, that focuses on the ability of robots to interpret and process human language. NLP is the foundation for research into how AI models understand inputs in human languages and respond to them. AI developers use different algorithms and statistical language models to process and generate human language in their systems (Pathak & Mehta, 2024) ^[14]. These models are trained to understand the semantics of the user inputs and to generate coherent and contextually relevant material. Particularly, OpenAI's AI model called ChatGPT is unique because of the self-attention mechanisms that allow it to generate accurate and relevant answers. Error analysis is a significant part of NLP theory since the errors in texts generated by AI can be identified. This includes finding factual errors, irrelevant information, and grammatical errors.

In this study, we utilize the NLP framework to investigate the AI's faulty and inconsistent responses. The work seeks to highlight, through a focus on the root causes of these errors, how the distinct character of different academic fields poses challenges for AI models in understanding and generating genuine human language.

Literature gaps and rationale for the current study

Although there has been a great deal of research on the use of AI in a range of domains, there are significant gaps. One big gap is the missing comparisons of AI performance in other fields like literature and languages. In a recent study,

the attention has been mostly on the application of AI into particular sectors, but the differences in AI's strengths and shortcomings in other fields have not been explored extensively (Amirjalili *et al.*, 2024) [1]. This research can take advantage of this gap in the literature to study the domain-specific challenges encountered by AI models. And there is less research on the boundaries of AI in interpretive disciplines, such as literature, than in systematic fields such as languages. The present paper aims to fill this vacuum by providing a comprehensive overview of the performance of AI in the disciplines of linguistics and literature. We will identify problems in each area and propose changes to AI models to address these challenges.

2. Research methodology

This study aims to review the performance of artificial intelligence (AI) in two different domains of literature and linguistics. The primary aim was to assess the effectiveness of AI in these two industries by examining its replies to certain tasks and questions. The methodology was divided into two components, the literary and the linguistic domains, to provide a meaningful and reliable comparison between the AI responses and the human interpretation. This section explains the procedures of data collection, analysis, and validation to ensure that the research was conducted according to ethical and academic criteria.

In this study, a qualitative comparative research design was adopted to examine the accuracy, reliability, and quality of AI-generated responses within the literature and linguistics domain. A comparative approach was adopted to discover and assess the performance of AI in two fields.

The focus was on character analysis in literature and the concept of linguistic relativity in linguistics. The research was predicated on two hypotheses. Comments generated by AI in literature may not be accurate, of good quality, and credible, especially for complex character evaluations. The trial lasted for three months. The researchers randomly applied these prompts to Grok, Gemini, and ChatGPT for the study. Data collection using rubrics: correctness of facts, relevance to text/theory, depth of interpretation, and hallucination presence.

- *Literature (Character Sketch of Molvi Jalal) (see Appendix 1)*
- *Linguistics for language relativity (see Appendix 2)*

Analysis of data

The data analysis involved a meticulous comparison of AI responses with human expectations in literature and linguistics. Two studies were done for each field so that each could be judged on its own merits, as well as for cross-disciplinary comparisons. The following actions were undertaken: (i) coding and thematic analysis of the AI's responses were coded to identify common themes and patterns. These were compared with topics commonly present in human literary interpretations of the work. (ii) expert review where two literary scholars assessed the AI response for accuracy, depth, and relevance to the themes of the novel.

Validity and Reliability

To establish validity and reliability, the study technique was carried out with the following strategies. Each test (literature analysis and linguistic explanation) was carried out three times to ensure the consistency of the AI outputs. (Whorf,

1956; Sapir, 1929; Jamal, 2023 [2])

Ethical considerations

The study has been performed in accordance with the ethical standards. These include detailed descriptions of all AI outputs and the informed consent of the experts involved in the evaluation.

3. Findings and Discussions

The following sections present detailed evaluations of the AI responses in the three areas, including their accuracy, relevance, and consistency. -ChatGPT 4.0 and Gemini and Grok Literature Subject (April 2026, 5:00 pm) (see Appendix 1).

The AI output actually portrays Molvi Jalal as a religious hypocrite who uses religion for his own gain. The AI makes it seem like Molvi Jalal is a mentor for the character, but in reality, Molvi Jalal is actively sabotaging the character. The AI also mislabels the character's name as Sara, whereas the real name is Zaib-un-Nisa, Zaib. Most importantly, the AI misunderstands the thematic parts of the novel. It does not deal with the tensions between modernity and old values, which are not a key issue in the text. This discrepancy highlights the constraints of AI in understanding complex literary elements. These errors may be because the AI is not used to the text, as "The Blue Room" is a Pakistani novel and comparatively less well-received critically than its Western counterparts. This misinformation poses a risk for students who seek academic help from ChatGPT. Molvi Jalal is a negative character, a hypocrite, corrupt, and has the elements of abuse. It's not customary to portray him as completely mistaken (mentor caring + positive transformation) vs. modernity, as AI says. Wrong theme (not tradition vs. modernity, the way AI says).

Then it's about AI's language answers. (Date: 5:00 pm, April 2026) with ChatGPT 4.0, Gemini, and Grok. (Appendix 2)

The response, while not very deep, may use a more extensive explanation, but it still provides a good basis for understanding the basic concepts of linguistic relativity. The examples given by the AI are correct, relevant, and well expressed, making the theory more accessible to students studying the relation between language and mind. (Pathak & Mehta, 2024) [14]

In contrast, the comparative study points to a huge gap between the performance of AI in literature and in linguistics. In literature, AI does not understand the intricacies and complexities of character development, themes, and symbols as shown by Molvi Jalal in Nafisa Rizvi's *The Blue Room*. The inability of AI to appropriately portray characters and themes points out the challenges it faces when dealing with the subjective, interpretative, and emotionally complex dimensions that are fundamental to literature (Pathak & Mehta, 2024) [14]. In addition, linguistics provides the means to precisely define and express structured concepts such as linguistic relativity. But literature calls for a greater degree of knowledge, which AI apparently cannot provide. This gives valuable insights into the potential and limits of AI in educational settings. AI is very accurate and reliable in ordered fields such as linguistics, but it fails in more interpretive and subjective fields such as literature.

Table 1: A score table of ChatGPT, Grok, and Gemini in literature and linguistics

Task	Model	Accuracy	Relevance	Depth	Hallucination	Mean Score
Literature	ChatGPT	2	2	2	1	1.75
	Grok	2	1	3	1	1.72
	Gemini	3	2	1	1	1.70
Linguistics	ChatGPT	4	5	4	4	4.25
	Grok	5	4	4	4	4.20
	Gemini	4	4	5	4	4.23

The results can offer the AI errors, particularly hallucinations and erroneous information, such as wrong character names (Sara instead of Zaib-un-Nisa) and non-existent details in literature, and sometimes make quotations. AI replies also reveal thematic themes in this tale. Furthermore, AI lacks the character of the major characters and phony conflicts.

4. Conclusion

The evaluation of AI responses in two closely related fields, literature and linguistics, showed that the AI can easily cope with disciplines with systematic and well-developed ideas, such as linguistics; however, it encounters various challenges when it comes to dealing with complex and interpretive subjects, such as literature. The inability of AI to communicate the persona of Molvi Jalal emphasizes the significance of enhancing AI systems in the realm of literature. The comparison analysis, applied in this study, will help AI developers to determine the areas that need improvement. The correctness of this study is very contingent on the subject matter for which AI is tasked with generating knowledge. For example, interpretive literary tasks seem to be challenging for AI, whereas AI excels in structured linguistic competence. The study is important as it demonstrates the necessity for adjustments in a system that is widely used among teachers and students from all academic disciplines and other parts of life.

5. Conflict of Interest

Authors have no conflict of interest to declare.

6. Acknowledgement

This research is funded by the Department of English Language, Ho Chi Minh City University of Economics and Finance, and assisted with this study. We appreciate our colleagues for their valuable insights that contributed to the research.

7. References

- Amirjalili F, Neysani M, Nikbakht A. Exploring the boundaries of authorship: A comparative analysis of AI responses and human academic writing in English literature. *Frontiers in Education*, 2024. Doi: <https://doi.org/10.3389/feduc.2024.1347421>.
- Jamal N. A Marxist Feminist Analysis of Nafisa Rizvi's *The Blue Room*. *University of Chitral Journal of Linguistics and Literature*. 2023; 7(1):90-94. <https://jll.uoch.edu.pk/index.php/jll/article/view/29>
- Nguyen TTH. Approaching the use of the prepositions "at," "in," and "on" by some non-specialized university students in Ho Chi Minh City in 2018. *Scientific Journal of Ho Chi Minh City University of Education*, 2020a.
- Nguyen TTH. Access to English prepositions and equivalent units in Vietnamese. *Scientific Journal of Ho Chi Minh City University of Education*, 2020b.
- Nguyen TTH. The locating prepositions "at," "in," and "on" in English and the translation into Vietnamese through the locating reference frame. *Scientific Journal of Ho Chi Minh City University of Education*, 2020c.
- Nguyen TTH. Locative prepositions "at, in, and on" and translation to Vietnamese representational meaning. *International Journal of Scientific & Engineering Research*. 2022; 13(3).
- Nguyen TTH. Animate your communicative class & engage your students with fun facts. *International Journal of Advanced Multidisciplinary Research and Studies*. 2023a; 3(6):835-840.
- Nguyen TTH. Advantages and disadvantages of online classes through some authors. *International Journal of Advanced Multidisciplinary Research and Studies*. 2023b; 3(1):754-756.
- Nguyen TTH. Analyze two main characteristics of Vietnamese women through cultural language. *International Journal of Advanced Multidisciplinary Research and Studies*. 2023c; 3(5):119-123.
- Nguyen TTH. Technology tools for disabled online learners. *International Journal of Advanced Multidisciplinary Research and Studies*. 2023d; 3(6): 835-840.
- Nguyen TTH. Intervention development for self-regulated writing among English learners: A single-case design study. *International Journal of Advanced Multidisciplinary Research and Studies*. 2024a; 4(3): 1399-1410.
- Nguyen TTH. Examining the issues of English-speaking skills that university EFL learners face. *International Journal of Advanced Multidisciplinary Research and Studies*. 2024b; 4(2):36-40.
- Nguyen TTH. Addressing core values in professional development in the context of English language teaching: Case study. *International Journal of Advanced Multidisciplinary Research and Studies*. 2024c; 4(1):214-218.
- Pathak PV, Mehta VD. Intelligent conversations: A theoretical framework for understanding natural language processing within artificial intelligence systems. *International Journal for Multidisciplinary Research*, 2024.
- Sapir E. The principles of linguistic relativity. In *Language and Culture: The Principles of Linguistic Relativity*. Harvard University Press, 1929, 120-137.
- Whorf BL. *Language, thought, and reality: Selected writings*. MIT Press, 1956.

Appendix 1

For literature, the AI was required to do a character analysis of Molvi Jalal from Nafisa Rizvi's book, *The Blue Room*. ChatGPT was given a well-designed prompt to have a comprehensive discussion on the characteristics, function,

and significance of the character in the story. We chose character analysis because it requires a deep understanding of literary concepts, character development, and narrative context—things that AI models often struggle to fully understand. The AI responses were gathered and rated on the following criteria: (i) correctness, i.e., whether the response appropriately reflected the character's features, actions, and importance in the story. (ii) significance to the extent that the answer captured the novel's themes, motifs, and literary critiques. (iii) interpretive depth when AI can interpret symbolic meanings and emotional nuances in Molvi Jalal's character. **Prompt: Using ChatGPT 4.0, Gemini, and Grok, write a detailed character sketch of Molvi Jalal from Nafisa Rizvi's novel *The Blue Room*.

-Literature Topic (Date: April, 2026, at 5:00 pm) with ChatGPT 4.0, Gemini, and Grok.

Character sketch of Molvi Jalal from Nafisa Rizvi's novel "The Blue Room."

USER COMMAND:

Character profile of Molvi Jalal from Nafisa Rizvi's novel "The Blue Room."

-ChatGPT Response:

Introduction:

Molvi Jalal is a key character in Nafisa Rizvi's novel "The Blue Room." As a religious figure and mentor, his profound presence has a significant impact on the protagonist and other characters, illustrating the complicated interaction of tradition and change in the community.

According to Rizvi (p. 45), Molvi Jalal is a tall, powerful figure with piercing eyes and a flowing white beard. This description establishes the tone for his personality traits: Molvi Jalal is shown as both caring and strict. His commitment to his beliefs is clear, as is his sometimes rigidity. For example, Rizvi writes, "His voice softened, not out of compassion, yet his eyes hardened when discussing transgressions against faith" (p. 78).

Grok responses: Physical Description:

Molvi Jalal is the novel's moral compass for the village. His advice is sought during times of crisis, and his conclusions frequently carry enormous weight. His adherence to tradition might lead to conflict, especially with younger characters desiring change. Forceful presence throughout the story.

Grok response: Personality traits:

Molvi Jalal is shown as both caring and strict. His commitment to his beliefs is clear, as is his sometimes rigidity. For example, Rizvi writes, "His voice softened, not out of compassion, yet his eyes hardened when discussing transgressions against faith" (p. 78).

Gemini responses: Role in the plot:

Molvi Jalal is the novel's moral compass for the village. His advice is sought during times of crisis, and his conclusions frequently carry enormous weight. His adherence to tradition might lead to conflict, especially with younger characters desiring change.

Relationships with other characters:

Molvi Jalal has a tangled relationship with the heroine, Sara. He is both a mentor and an impediment to her goals. For example, he corrects her, stating, "You must understand, Sara, that our ways are not to be questioned lightly" (Rizvi, p. 123). These dynamics capture the novel's major generational and ideological conflicts.

Development and Transformation:

Molvi Jalal develops and transforms throughout the narrative, demonstrating a deeper grasp of the changing world. Finally, he admits, "Perhaps there is more to faith than rigid adherence; perhaps there is room for compassion and growth" (Rizvi, p. 289).

Gemini response: Symbolism and Themes:

In conclusion, Molvi Jalal plays a crucial role in "The Blue Room," highlighting the problems of maintaining tradition in a rapidly changing society. His journey from rigidity to understanding parallels the novel's investigation of cultural change and the search for harmony between the old and the modern.

Appendix 2

For the linguistics part, the AI was required to define the notion of linguistic relativity. This theory contends that the structure of a language influences the worldview and cognition of its speakers. This task aimed to test the ability of AI systems to explain theoretical concepts and to give suitable examples from the academic literature or from famous linguistics studies.

Sampling

The AI model (ChatGPT) and the literary and linguistic subjects were selected based on the purposive sampling strategy. The following samples were selected for the study. (i) We choose an AI model like ChatGPT (GPT-4) due to its sophisticated natural language processing capabilities and its wide use in educational and academic settings. (ii) Sample of literature (iii) Sample of language

Then it's about AI's language answers. (Date: 5:00 pm, April 2026) with ChatGPT 4.0, Gemini, and Grok. (Appendix 2)

-Linguistics Topic: Linguistic relativism

An example-based explanation of linguistic relativity is requested.

- ChatGPT Response:

Linguistic relativity, commonly known as the Sapir-Whorf hypothesis, proposes that a language's structure influences how its speakers see and think about the world. According to this hypothesis, people who speak various languages may have distinct perspectives on the world because their language influences their thoughts, categorization, and perception of events.

There are two major versions of linguistic relativity:

- **Strong version (linguistic determinism):** *Language entirely controls thought. This extreme form implies that humans are unable to think about things outside the context of their language.*

Weak version: *Language shapes thinking and perception, but does not completely constrain them.*