

Received : 30 July 2024, Accepted: 28 August 2024

DOI: <https://doi.org/10.33282/rr.vx9i2.83>

ASSESSING THE RELIABILITY OF AI IN ACADEMIC WRITING: A COMPARATIVE ANALYSIS OF LITERATURE AND LINGUISTICS RESPONSES

Nimra Noor¹, Dr. Aniq Rashid², Anum Rasheed³, Tuba Latif⁴, Huda Noor⁵

1. M.Phil. Scholar, Department of English, National University of Modern Languages, Islamabad (Faisalabad Campus)
2. Assistant Professor, Department of English, National University of Modern Languages, Islamabad (Faisalabad Campus)
3. Lecturer, Government College for Women, Kahuta.
4. M.Phil. Scholar, Department of English, National University of Modern Languages, Islamabad (Faisalabad Campus)
5. M.Phil. Scholar, Department of English, National University of Modern Languages, Islamabad (Faisalabad Campus)

ABSTRACT

This examination aims to assess the reliability of Artificial Intelligence (AI), specifically ChatGPT, in generating academic content across the disciplines of literature and linguistics. The study utilizes Natural Language Processing (NLP) theory, originally conceptualized by Alan Turing in (1969), as a framework for understanding AI's linguistic capabilities. The research conducts a comparative analysis of AI-generated responses regarding the character of Molvi Jalal from Nafisa Rizvi's novel *The Blue Room* and the Linguistic Relativity theory proposed by Sapir and Whorf. The primary objectives were to evaluate the accuracy, relevance, and credibility of AI's responses in both fields. Data were collected through targeted prompts, and the AI's output was critically analyzed against established literary critiques and linguistic theories. The findings revealed that while AI demonstrated higher accuracy in linguistics, particularly with well-defined theories, it performed poorly in literature, misinterpreting key character attributes and plot details. The study concludes that AI is more reliable in structured academic fields like linguistics but falls short in disciplines requiring nuanced interpretation, such as literature. The key

limitations include AI's reliance on available data, lack of emotional understanding, and challenges in dealing with less-critically evaluated texts. The research recommends a cautious approach to using AI in interpretive fields and emphasizes the need for future studies to address AI's limitations in handling context-dependent academic content. The implications suggest improvements in AI models for better performance in humanities-based disciplines.

Keywords: Artificial Intelligence, ChatGPT, Literature, Linguistics, NLP, Alan Turing, Reliability, Academic Writing, Comparative Analysis

1. INTRODUCTION:

1.1. Definition and Background:

The fast advancements in technology have introduced remarkably sophisticated systems of Artificial Intelligence (AI), fundamentally transforming various aspects of human life. AI systems, such as ChatGPT, have revolutionized how tasks are completed, allowing previously time-consuming and labor-intensive processes to be accomplished almost instantly with minimal physical or mental effort. These technological breakthroughs have had profound implications, especially in educational settings, where AI tools are increasingly used by students. However, despite the many benefits, the reliance on AI has sparked considerable concerns regarding the accuracy, reliability, and potential impacts on cognitive skills. In particular, students frequently turn to AI for assistance in generating assignments, preparing for exams, and obtaining information on academic topics. Yet, over-reliance on AI can lead to the erosion of critical thinking and analytical reasoning skills, as students often fail to question the correctness of the information provided by these systems, instead placing undue trust in AI outputs (Zhai et al., 2024).

The question of AI's reliability, especially in academic contexts, has garnered substantial attention in recent years. While AI has undoubtedly brought about revolutionary changes in various fields, concerns have been raised about the ethical implications and accuracy of AI-generated information. In academic environments, there is ongoing debate over how to strike the

right balance between AI and human expertise. AI, with its capacity to generate content quickly and efficiently, has raised alarms about the possibility of diminishing human intellectual engagement. Students may risk becoming overly dependent on AI, potentially sacrificing the depth of their understanding in favor of convenience (Khalifa et al., 2024). A personal encounter with this issue was observed when I utilized ChatGPT to generate responses for assignments across different academic subjects. Notably, while the AI-generated responses for my literature assignments lacked depth, accuracy, and relevance, the responses for linguistics assignments were comparatively more precise. This discrepancy prompted me to reflect on the limitations of AI and the necessity for improvements, particularly the development of field-specific AI models.

The need for a more nuanced and domain-specific AI model is further reinforced by the disparities in AI performance across academic disciplines. In response, a comparative study was undertaken to examine the reliability of AI-generated content in two distinct areas: literature and linguistics. In literature, AI often struggles to generate accurate responses, particularly when tasked with identifying specific literary works or providing detailed analysis. For example, AI frequently produces incorrect quotes, misidentifies literary devices, and misinterprets complex literary elements. A notable instance occurred when AI attempted to analyze the character of Molvi Jalal from Nafisa Rizvi's novel *The Blue Room* (2009). Instead of recognizing Jalal as the antagonist, the AI response mistakenly portrayed him as a compassionate figure, a serious mischaracterization that underscores the challenges AI faces when processing nuanced literary themes and context-dependent character analyses (Misiejuk et al., 2024). This example illustrates the limitations of AI in handling the complexities inherent in literature, where interpretation often requires a deep understanding of context, subtext, and symbolism.

In contrast, AI has shown a much higher degree of reliability in the field of linguistics. When prompted to explain the theory of Linguistic Relativity, for instance, AI provided a clear and accurate description of the theory, which posits that language influences thought processes and shapes perceptions of reality. The examples generated by the AI were consistent with the work of key figures such as Sapir (1929) and Whorf (1956), who first introduced the theory. This success in the domain of linguistics suggests that AI is capable of handling more structured and formal content, particularly in areas where there are established frameworks and theories. In linguistic analysis, the principles and rules are more rigid compared to the interpretive nature of literature,

allowing AI to perform more effectively. The divergence in AI performance across literature and linguistics highlights the need for domain-specific improvements in AI models to ensure they cater appropriately to the unique demands of each discipline.

The varying levels of success demonstrated by AI in these two fields suggest that the nature of the academic discipline plays a significant role in determining the effectiveness of AI-generated content. Literature, with its reliance on interpretation, emotional nuance, and the need for contextual understanding, poses a unique challenge for AI systems. On the other hand, linguistics, which is more data-driven and grounded in specific theories and frameworks, allows AI to function more effectively. This disparity raises important questions about the limitations of current AI systems and the potential for improving their capabilities in more interpretive or subjective fields. Future AI models must be designed to account for the specific requirements of different academic domains, ensuring greater accuracy and relevance in their responses.

While AI has undoubtedly made significant strides in transforming education and other fields, its limitations in certain academic domains cannot be overlooked. The performance gap between literature and linguistics in AI-generated responses highlights the need for continued research and development in the field of AI. By addressing these shortcomings and creating more specialized, field-specific AI models, we can harness the full potential of AI without compromising the depth and quality of human intellectual engagement. Ethical considerations, accuracy, and the preservation of critical thinking must remain at the forefront of discussions surrounding AI in education, ensuring that AI is used as a tool to enhance learning rather than replace human expertise.

1.2. Statement of the problem

The problem addressed in this study revolves around the accuracy and reliability of AI-generated responses in academic settings, particularly in the fields of literature and linguistics. While AI models like ChatGPT have become widely used tools for generating quick and coherent responses, they often display inaccuracies and inconsistencies, especially in nuanced and context-dependent subjects such as literary analysis. This over-reliance on AI without proper critical evaluation may undermine students' analytical skills and lead to the dissemination of incorrect information. The research seeks to investigate these limitations, focusing on the

disparities in AI performance between literature and linguistics, and to explore the potential consequences for academic integrity and critical thinking.

1.3. Significance of the Study

This research holds significant value by drawing attention to the limitations of widely used AI systems, specifically in academic settings. It encourages critical reflection on the extent to which users should rely on these systems, especially when it comes to academic work. A key goal of this study is to highlight the necessity for AI developers to make field-specific enhancements to ensure greater accuracy in AI-generated content. Additionally, it serves as a cautionary reminder to students to carefully evaluate AI outputs instead of using them blindly. By urging students to critically assess AI-generated responses, the study promotes a more thoughtful approach to the use of AI in educational environments. Furthermore, this research paves the way for future investigations into AI's reliability and accuracy across other academic and non-academic fields, thus laying a foundation for more specialized inquiry in this area.

1.4. Research Objectives

This research is guided by the following objectives:

- a. To assess the accuracy and relevance of AI-generated content in literary analysis, with a specific focus on character evaluation in Nafisa Rizvi's novel *The Blue Room*
- b. To examine the effectiveness of AI in producing accurate and relevant information on linguistic theories, particularly Linguistic Relativity
- c. To compare and contrast the performance of AI in literature and linguistics, identifying the factors that contribute to differences in accuracy and precision between these two fields.

1.5. Research Questions

The following research questions arise from the stated objectives and serve as the foundation for this study:

- i. How accurate are AI-generated responses in evaluating literary characters, as exemplified by the portrayal of Molvi Jalal in *The Blue Room*?
- ii. How reliable and precise is AI-generated content regarding linguistic theories, such as Linguistic Relativity?
- iii. What differences can be identified in AI's performance across literature and linguistics, and what factors contribute to these variations?

1.6. Research Hypotheses

The research operates under the following hypotheses:

AI-generated responses are expected to demonstrate higher accuracy and relevance in the field of linguistics compared to literature, due to the structured and precise nature of linguistic theories, in contrast to the interpretive complexity of literary analysis.

Inaccuracies in AI-generated literary analysis may stem from challenges in interpreting nuanced characterizations and literary devices, whereas AI's performance in linguistics is enhanced by the clear and well-defined nature of linguistic concepts.

This comparative study is designed to evaluate AI's capabilities in literature and linguistics, providing insights into the strengths and weaknesses of AI in academic writing. By offering a deeper understanding of AI's role and limitations, the study also aims to propose recommendations for improving AI models, particularly in field-specific applications.

1.7. THEORETICAL FRAMEWORK

This study utilizes Natural Language Processing (NLP) theory, originally conceptualized by Alan Turing (1969), which focuses on the ability of machines to understand and process human language. NLP serves as the foundation for examining how AI models interpret human language inputs and generate responses. AI developers incorporate numerous algorithms and statistical language models into their systems, enabling them to process and produce human language (Pathak et al., 2024). These models aim to generate coherent and contextually

appropriate text by comprehending the meaning behind user prompts. OpenAI's ChatGPT, in particular, stands out among AI models due to its advanced self-attention techniques, which enhance its ability to produce highly accurate and relevant responses.

An important aspect of NLP theory is error analysis, which plays a key role in detecting imprecision within AI-generated text. This includes identifying factual inaccuracies, irrelevant information, and linguistic inconsistencies (Yuan et al., 2022). By analyzing these errors, developers can uncover weaknesses within AI models and pinpoint areas that require improvement. Understanding where and why AI systems fail allows for targeted adjustments that enhance their capabilities (Wu et al., 2019).

Numerous studies have applied the NLP framework to explore the limitations of AI models. For instance, while many NLP models have made significant advancements, they continue to face challenges in interpreting the full context of human language (Thakkar et al., 2023). Despite the progress in language production, comprehension, and translation, NLP still encounters developmental challenges (Ghazizadeh et al., 2020).

This study applies the NLP framework to analyze the errors and inaccuracies found in AI-generated responses. By focusing on the root causes of these errors, the research aims to highlight how the distinct nature of different academic disciplines presents challenges for AI models in processing and generating accurate human language.

2. LITERATURE REVIEW

The integration of Artificial Intelligence (AI) into various fields, particularly education, has garnered substantial attention from researchers. AI's rise in popularity among students and educators has sparked curiosity regarding its strengths, limitations, and reliability in different disciplines (Khalifa et al., 2024). This review explores the role of AI in academic writing, literature, and linguistics, critically evaluating its impact and limitations. As AI's presence grows, it becomes crucial to assess whether its use should be accepted uncritically or if further developments are required to ensure its appropriate application in these domains.

AI's influence in academic writing has been transformative, significantly streamlining research and writing processes. AI models, particularly ChatGPT, have been developed to generate human-like text by processing extensive datasets. These models assist in increasing the accessibility, accuracy, and efficiency of writing tasks, especially in academic settings (Khalifa et al., 2024). AI is increasingly being used by students and researchers to generate ideas, structure content, edit drafts, and even analyze data (Lin, 2023). This has made the writing process more efficient, saving time for users engaged in research or producing written content.

However, the overreliance on AI in academic writing raises significant concerns. One of the primary issues is academic integrity. AI tools can generate complete essays or research papers with minimal human input, posing challenges to originality and critical thinking (William, 2024). While AI can assist with basic writing tasks, its output often lacks the depth of analysis and intellectual rigor necessary for high-quality academic work. This reliance can also undermine students' development of critical writing skills, as they may become dependent on AI-generated content rather than engaging in thoughtful analysis themselves.

Further complicating the use of AI in academic writing is the potential for factual inaccuracies. Although AI models like ChatGPT are trained on vast datasets, they are not immune to producing errors or presenting misleading information. The algorithms behind these models rely on probability, which can lead to incorrect or irrelevant content generation in academic contexts. This raises the question of whether AI can be trusted to produce reliable, factually accurate academic writing, or if human oversight is always required (Chen et al., 2023).

AI has also made inroads into literature, particularly in the realm of creative writing. By analyzing large datasets, AI can generate creative text, assist with story generation, and even simulate literary styles (Young, 2020). AI models have been used to mimic the prose of famous authors, generating new stories or poems inspired by their styles. This capability is particularly useful in creative industries where AI can serve as a tool for brainstorming or overcoming writer's block. Despite these benefits, AI struggles to meet the interpretive and emotional requirements of literary analysis. Literature is inherently subjective, relying on deep human emotions, experiences, and cultural contexts that AI has difficulty processing. For example, AI is less capable of understanding and interpreting metaphors, symbols, and complex character

motivations that are essential to literary analysis (Amirjalili et al., 2024). While AI can generate text that is syntactically correct, it often falls short of conveying the emotional depth that is critical in human-generated literary works.

Moreover, there are ethical concerns regarding the use of AI in literature, particularly related to creativity. Some scholars argue that AI's increasing role in creative writing could undermine human creativity, as writers may become overly reliant on machines to generate ideas or complete works (Raj et al., 2023). The question of authorship also arises—if AI contributes significantly to a creative work, who holds the intellectual property rights? These ethical dilemmas highlight the complexities of integrating AI into literature, where human creativity and originality are paramount.

Linguistics, as a more systematic and rule-based field, has seen more successful applications of AI, particularly in language processing and acquisition. AI models, specifically those powered by Natural Language Processing (NLP), are adept at analyzing linguistic patterns, syntax, and phonetics (Ruqaiyah, 2023). These models can process large datasets of language input to identify grammatical structures, suggest methodologies for language learning, and improve language competence. This makes AI particularly valuable in the field of linguistics, where it can assist in understanding language patterns and aiding language learners. One area where AI has shown great promise is in the study of historical linguistics. AI-powered models can analyze ancient texts and help decipher lost languages by identifying recurring patterns in phonetics and syntax (Patil, 2024). AI also aids in the study of language evolution, providing linguists with tools to analyze changes in language use over time. The ability of AI to process vast amounts of linguistic data allows researchers to draw insights that would be impossible through manual analysis.

AI's role in language acquisition, particularly for non-native speakers, is another significant area of interest. NLP models have been used to develop tools that assist English as a Foreign Language (EFL) learners by providing real-time feedback on grammar, pronunciation, and syntax (Keezhatta, 2019). These tools help improve language fluency by offering learners personalized assistance and suggestions. However, despite AI's proficiency in linguistic analysis,

there are still limitations, particularly in its ability to understand the pragmatics of language—how context influences meaning.

A critical issue in the integration of AI across disciplines is the ethical implications of its use. The reliance on AI, especially in fields like academic writing and literature, raises concerns about the erosion of human creativity and intellectual responsibility (Raj et al., 2023). While AI can generate content efficiently, it lacks the capacity for ethical decision-making, which becomes particularly problematic in areas requiring original thought, creativity, and ethical considerations, such as literature and academia. Another key limitation of AI is its inability to fully understand context. AI models like ChatGPT rely on probabilistic language generation, which can lead to misinterpretations of nuanced or ambiguous human language (Ghazizadeh et al., 2020). While AI excels in systematic tasks, it often struggles in domains requiring interpretive and emotional intelligence, such as literature and complex linguistic tasks involving cultural or contextual knowledge. These limitations underscore the need for ongoing research into improving AI's capabilities, particularly in understanding human contexts and emotions.

The role of AI in enhancing creativity is a topic of much debate. On the one hand, AI offers tools that can assist with the creative process, such as generating story ideas, creating visual art, or composing music (Young, 2020). This has led to discussions about whether AI can be considered a co-creator in artistic endeavors or merely a tool used by human artists. The ability of AI to generate creative content opens new possibilities for collaboration between humans and machines in the creative industries. On the other hand, critics argue that AI lacks true creativity, as it merely replicates patterns from the data it is trained on. True creativity, they contend, involves original thought and innovation, which AI is currently incapable of (Raj et al., 2023). AI-generated works may be impressive in their technical execution but often lack the emotional depth, originality, and unpredictability that characterize human creativity. This raises questions about the limits of AI's role in creative fields and whether it can ever truly replicate the human creative process.

AI's integration into various fields has sparked interest in how humans and machines can collaborate effectively. In academic writing, for example, AI tools can assist with generating content and providing feedback, allowing writers to focus on refining their arguments and

analysis (Chen et al., 2023). This collaborative model suggests that AI could serve as a valuable tool for augmenting human productivity rather than replacing human thought entirely. However, this collaboration requires careful management to avoid overreliance on AI. If students and researchers rely too heavily on AI for generating content or conducting analyses, they risk losing essential skills such as critical thinking, problem-solving, and original thought (William, 2024). The key to effective human-machine collaboration lies in using AI as a tool to enhance human capabilities, not as a substitute for human intellect and creativity.

2.1 Gaps in Literature and Rationale for the Current Study

While there has been substantial research into the use of AI in various fields, several gaps remain. One notable gap is the lack of comparative studies on AI's performance across different domains, such as literature and linguistics. Most existing research focuses on AI's applications in specific fields, but there is limited understanding of how AI's strengths and weaknesses vary between disciplines (Amirjalili et al., 2024). This gap in the literature presents an opportunity for further research to explore the domain-specific challenges faced by AI models.

Additionally, while AI has been extensively studied in systematic fields like linguistics, less attention has been paid to its limitations in more interpretive fields like literature. This study aims to fill this gap by conducting a comprehensive analysis of AI's performance in both linguistics and literature, highlighting the specific challenges that arise in each field and suggesting potential improvements to AI models to address these challenges.

3. RESEARCH METHODOLOGY

This research was designed to critically evaluate the performance of Artificial Intelligence (AI) in two distinct disciplines: literature and linguistics. The primary objective was to compare AI's effectiveness in both fields by assessing its responses to targeted tasks and questions. The methodology was divided into two sections, corresponding to the literary and

linguistic domains, and was carried out with the intention of providing a reliable and valid comparison of AI-generated content with human analysis. This section outlines the approach adopted for data collection, analysis, and validation, ensuring the research met the required ethical and academic standards.

3.1 Research Design

The study followed a qualitative comparative research design to evaluate the accuracy, credibility, and quality of AI-generated responses in literature and linguistics. The comparative framework was employed to identify and analyze how AI performs in both domains, focusing on character analysis in literature and the theory of linguistic relativity in linguistics. The design was appropriate for understanding the contextual, interpretive, and theoretical challenges faced by AI in these subject areas. The research was underpinned by two hypotheses:

- i. AI-generated responses in literature may lack accuracy, quality, and credibility, particularly in complex character analysis.
- ii. AI-generated responses in linguistics are expected to demonstrate a higher level of accuracy, appearing closer to human-written texts due to the structured and theoretical nature of the subject.

3.2 Data Collection

The data collection process involved obtaining AI-generated responses from ChatGPT in response to targeted prompts in both the literature and linguistics domains. The collection was structured to ensure that the data aligned with the research objectives, focusing on the accuracy and relevance of AI's performance in each field.

3.2.1 Literature: Character Analysis of Molvi Jalal

For the literature section, the AI was tasked with generating a character analysis of Molvi Jalal from Nafisa Rizvi's novel *The Blue Room*. A carefully formulated prompt was provided to ChatGPT to elicit a detailed response regarding the character's attributes, role, and significance within the novel. The character analysis was selected as it requires a nuanced understanding of

literary themes, character development, and narrative context, which are often challenging for AI models to interpret. The AI-generated content was collected and assessed based on the following criteria:

- a. *Accuracy*: How well the response reflected the character's traits, actions, and significance in the narrative.
- b. *Relevance*: The degree to which the response aligned with the themes, motifs, and literary critiques of the novel.
- c. *Interpretive Depth*: The ability of AI to interpret symbolic meanings and deeper emotional nuances associated with Molvi Jalal's character.

3.2.2 Linguistics: Linguistic Relativity

For the linguistics section, the AI was prompted to explain the theory of linguistic relativity, which posits that the structure of a language affects its speakers' worldview and cognition. The prompt was designed to assess AI's ability to explain theoretical concepts and provide relevant examples from academic sources or well-known linguistic studies. The AI-generated response was evaluated based on:

- a. *Conceptual Clarity*: The AI's ability to accurately define linguistic relativity and its implications.
- b. *Theoretical Accuracy*: Whether the AI's explanation adhered to established linguistic theories and research.

3.3 Sampling

A purposive sampling method was employed to select the AI model (ChatGPT) and the literary and linguistic topics. The following samples were selected for the study:

- a. *AI Model*: ChatGPT, specifically GPT-4, was chosen due to its advanced natural language processing capabilities and widespread use in educational and academic contexts.
- b. *Literary Sample*: The novel *The Blue Room* by Nafisa Rizvi was chosen for its complex characters and layered themes, with Molvi Jalal as a central character who embodies cultural and religious symbolism. This provided a challenging subject for AI analysis.

- c. *Linguistic Sample*: The theory of linguistic relativity, a well-established concept in the field of linguistics, was selected for its importance in theoretical discussions and its structured, logical nature, which AI models are expected to handle with greater accuracy.

3.4 Data Analysis

The data analysis process involved a detailed comparison of AI-generated content with human expectations in both literature and linguistics. Two separate analyses were conducted for each field, ensuring that both were evaluated on their own merits while allowing for cross-disciplinary comparisons.

The character analysis of Molvi Jalal generated by AI was subjected to qualitative content analysis. The analysis focused on three key areas: accuracy, interpretive depth, and relevance. The AI-generated response was compared with established literary critiques of the novel and expert interpretations of Molvi Jalal's character.

The following steps were taken:

- a. *Coding and Thematic Analysis*: The AI's response was coded to identify recurring themes and patterns. These were compared with themes commonly discussed in human literary analyses of the novel.
- b. *Expert Review*: Two literary scholars were asked to review the AI's response and provide feedback on its accuracy, depth, and relevance to the novel's thematic content. This provided an additional layer of validation for the results.

3.5 Validity and Reliability

The research methodology ensured both validity and reliability through the following strategies:

Task Repetition: Each task (literary analysis and linguistic explanation) was repeated three times to ensure the consistency of AI-generated responses.

- a. Triangulation: Multiple data sources, including AI outputs, expert feedback, and literary/linguistic theories, were used to triangulate the findings and ensure comprehensive results.
- b. Inter-Rater Reliability: The feedback from experts was analyzed for consistency using Cohen's Kappa, which measured the agreement between the two experts in each domain.

3.6 Ethical Considerations

The study adhered to ethical guidelines by ensuring that all AI-generated responses were transparently described, and the experts involved in the evaluation were given informed consent regarding their participation. Their feedback was anonymized to maintain confidentiality. Additionally, no personal data was collected, and the use of AI for academic purposes was clearly stated in the research.

The research methodology was designed to evaluate the performance of AI in literature and linguistics by focusing on character analysis and theoretical explanation tasks. Through targeted data collection, expert feedback, and comparative analysis, the study aimed to provide a valid and reliable understanding of AI's strengths and limitations in these two fields. The results contributed to the broader discussion on the capabilities of AI in handling creative, interpretive, and theoretical tasks across different disciplines.

4. DATA ANALYSIS

The data analysis phase involved a systematic evaluation of the AI-generated responses from ChatGPT in both the literary and linguistic domains. The primary goal was to assess the accuracy, relevance, and overall quality of the content produced by the AI, comparing these findings against established human expectations and expert critiques. For the literature analysis, the focus was on the character of Molvi Jalal from Nafisa Rizvi's *The Blue Room*. The AI's interpretation of the character was examined for its alignment with key themes of the novel, such as identity, cultural conflict, and moral complexity. Each response was analyzed using qualitative content analysis, identifying specific attributes that either corroborated or contradicted established literary critiques.

In the linguistic domain, the analysis centered around AI's explanation of the theory of linguistic relativity. This task required a clear and accurate presentation of the theory, supported by relevant examples. The AI-generated response was scrutinized for conceptual clarity and adherence to academic standards, using established linguistic literature as a benchmark for evaluation. The analysis aimed to uncover not only the effectiveness of the AI in articulating complex theories but also its ability to provide insightful examples that resonated with human understanding. By juxtaposing the findings from both domains, the analysis sought to highlight the inherent strengths and weaknesses of AI in processing and generating content across diverse intellectual landscapes.

To analyze the effectiveness of AI in the fields of literature and linguistics, the study assessed AI-generated responses regarding the character sketch of Molvi Jalal in literature and the concept of Linguistic Relativity in linguistics. The subsequent sections present these responses along with their detailed analysis, highlighting the accuracy, relevance, and coherence of the content produced by the AI in each domain.

Literature Topic

Character Sketch of Molyi Jalal in the Novel "The Blue Room" by Nafisa Rizvi

USER'S COMMAND:

Character sketch of Molyi Jalal in the novel "The Blue Room" by Nafisa Rizvi"

ChatGPT RESPONSE:

Introduction:

Molyi Jalal is a pivotal character in Nafisa Rizvi's novel "The Blue Room." As a religious figure and mentor, his presence deeply influences the protagonist and other characters, reflecting the complex interplay of tradition and change within the community.

Physical Description:

Molyi Jalal is described as "a tall, imposing figure with a flowing white beard and piercing eyes that seem to look right through a person" (Rizvi, p. 45). This description sets the tone for his authoritative presence in the novel.

Personality Traits:

Molyi Jalal is depicted as both compassionate and stern. His dedication to his faith is evident, but so is his occasional rigidity. For instance, Rizvi writes, "His voice softened as he spoke of compassion, yet his eyes hardened when discussing transgressions against faith" (Rizvi, p. 78).

Role in the Plot:

Throughout the novel, Molyi Jalal serves as a moral compass for the community. His guidance is sought in times of crisis, and his decisions often carry significant weight. His insistence on tradition sometimes leads to conflict, particularly with younger characters seeking change.

Relationships with Other Characters:

Molyi Jalal's relationship with the protagonist, Sara, is complex. He is both a mentor and a barrier to her aspirations. In one instance, he admonishes her, saying, "You must understand, Sara, that our ways are not to be questioned lightly" (Rizvi, p. 123). This dynamic illustrates the generational and ideological conflicts central to the novel.

Development and Transformation:

Over the course of the novel, Molyi Jalal begins to show a more nuanced understanding of the changing world. By the end, he acknowledges, "Perhaps there is more to faith than rigid adherence; perhaps there is room for compassion and growth" (Rizvi, p. 289).

Symbolism and Themes:

Molyi Jalal symbolizes the traditional values and the tension between orthodoxy and modernity. His character highlights the theme of faith versus progress, illustrating the struggles within a community to balance these forces.

Conclusion:

Molyi Jalal's character is integral to "The Blue Room," representing the challenges of upholding tradition in a changing world. His journey from rigidity to understanding mirrors the novel's exploration of cultural evolution and the search for balance between the old and the new.

References:

- Rizvi, Nafisa. *The Blue Room

The AI-generated response regarding Molvi Jalal's character in Nafisa Rizvi's novel "The Blue Room" fails to accurately reflect his true nature, depicting him instead as a moral and compassionate figure. In reality, Molvi Jalal is characterized as a religious hypocrite who exploits religion for personal gain, actively plotting against the protagonist rather than serving as a mentor, as the AI suggests. Additionally, the AI erroneously identifies the protagonist's name as Sara, whereas the correct name is Zaib-un-Nisa, commonly referred to as Zaib. The references provided by ChatGPT, including quotations and page numbers, are also inaccurate. Furthermore, the AI adds unnecessary details about Molvi Jalal's physical appearance, which are not mentioned in the novel. Most critically, the AI misinterprets the thematic elements of the novel; it does not address the tensions between modernity and traditional values, as this is not a central theme of the text. This discrepancy underscores the limitations of AI in comprehending complex literary elements (Misiejuk et al., 2024). Given that "The Blue Room" is a Pakistani novel with limited critical evaluation compared to its Western counterparts, it is plausible that the AI's lack of familiarity with the text contributes to these inaccuracies. Such misinformation poses a significant risk for students who may rely on ChatGPT for academic assistance.

Next, the focus shifts to the linguistics-related response provided by AI.

Linguistics Topic

Linguistic Relativity

USER'S COMMAND:

Explain Linguistic Relativity with examples.

ChatGPT RESPONSE:

Linguistic relativity, also known as the **Sapir-Whorf hypothesis**, is the idea that the structure of a language influences the way its speakers perceive and think about the world. According to this theory, people who speak different languages may experience the world differently because their language shapes their thoughts, categorization, and interpretation of reality.

There are two main versions of linguistic relativity:

- Strong version (linguistic determinism):** Language *completely determines* thought. This extreme version suggests that people cannot think about things outside the framework of their language.
- Weak version:** Language *influences* thought and perception but doesn't entirely limit them.

Examples:

- Color Perception:**

In English, we have a single word for the color "blue," but in Russian, there are two distinct words for different shades of blue: *siniy* (dark blue) and *goluboy* (light blue). Some studies suggest that Russian speakers are faster at distinguishing between light and dark blue because their language forces them to categorize the colors differently.

- Time Perception:**

The way languages describe time can affect how speakers perceive it. For example, in English, time is typically viewed in a horizontal dimension (we "look forward" to the future or "put the past behind us"). However, in Mandarin Chinese, time is also described in vertical terms, with the past being "up" and the future "down." Research suggests that Mandarin speakers may be more likely to think about time vertically compared to English speakers.

- Gendered Language:**

Languages like Spanish and German have gendered nouns (e.g., *la luna* – feminine, *el sol* – masculine in Spanish). Studies have found that speakers of these languages might attribute gendered qualities to objects based on their grammatical gender. For instance, in one study, Spanish speakers described the moon as "beautiful" and "delicate" (feminine traits), while German speakers, for whom the moon is masculine, used more masculine adjectives like "strong."

- Spatial Orientation:**

Some languages, like *Guugu Yimithirr* (spoken by an indigenous community in Australia), use absolute directions (north, south, east, west) instead of relative terms (left, right) for spatial orientation. Speakers of such languages are highly attuned to cardinal directions in their everyday life because their language requires them to constantly be aware of where they are in relation to these fixed directions.

These examples illustrate how linguistic relativity suggests a relationship between language and cognition, though the exact extent of this influence is still debated.

The AI-generated response regarding Linguistic Relativity, as proposed by Sapir and Whorf, aligns with the fundamental aspects of the theory, reflecting an accurate understanding of its core principles. Although the response lacks depth and could benefit from a more detailed explanation, it still provides a solid foundation for grasping the basic concepts of linguistic relativity. The examples given by the AI are precise, relevant, and well-articulated, making the theory more accessible to students who are exploring the relationship between language and thought. If prompted further, ChatGPT has the potential to expand on this initial explanation, offering more comprehensive insights while maintaining accuracy. This suggests that AI, particularly in linguistics, is capable of enhancing learning and academic inquiry by providing well-structured and reliable information (Ruqaiyah, 2023). Students using AI tools like ChatGPT for linguistic studies may find the technology a useful aid in understanding complex theories, as its ability to present relevant examples and coherent explanations is particularly strong in this field.

In contrast, the comparative analysis reveals a significant discrepancy in AI's performance when dealing with literature versus linguistics. When analyzing literary works, AI struggles to interpret the nuances and complexities inherent in character development, themes, and symbolism, as seen in the case of Molvi Jalal from Nafisa Rizvi's "The Blue Room." AI's failure to accurately portray characters and themes underscores its limitations in handling subjective, interpretive, and emotionally complex domains, which are essential to literature (Mostafapour et al., 2024). Unlike linguistics, where structured theories like Linguistic Relativity can be clearly defined and articulated, literature demands a more nuanced understanding that AI appears ill-equipped to provide. This performance gap highlights the challenge of applying AI to disciplines where meaning is layered, contextual, and open to interpretation, as opposed to fields where information is more systematic and rule-based.

The research emphasizes that the effectiveness of AI is largely dependent on the nature of the academic field in which it is applied. In linguistics, AI operates with greater accuracy and efficiency, as it deals with established theories and concepts that are easier for the model to

process and replicate (Ruqaiyah, 2023). The structured nature of linguistic inquiry allows AI to perform well, making it a valuable tool for students and researchers in this field. However, in literature, where interpretation and critical analysis are key, AI often falls short, misinterpreting characters and themes and failing to grasp the intricacies that human scholars navigate. This suggests that while AI has great potential in some academic disciplines, its capabilities remain limited in others, particularly those that require subjective analysis and emotional intelligence.

5.1. RESULTS AND DISCUSSION

The results of this study revealed a clear distinction between AI's performance in literature and linguistics, as predicted by the hypothesis. The AI-generated responses in the field of linguistics, specifically related to the concept of Linguistic Relativity, demonstrated a commendable level of accuracy and clarity. The fundamental principles of the theory, as proposed by Sapir and Whorf, were accurately presented. The examples provided in support of the theory were relevant and well-explained, allowing for a coherent understanding of the topic. This confirms that AI, especially models like ChatGPT, performs reliably in areas where theories are well-defined and structured. The ability to deliver precise, accurate information in response to linguistics-based queries highlights AI's potential to assist students and researchers in this field.

However, in contrast, the AI's performance in literature, as demonstrated by its analysis of Molvi Jalal's character from Nafisa Rizvi's *The Blue Room*, was significantly lacking in accuracy and depth. The AI failed to portray the character's true essence as a religious hypocrite and instead presented a moral and compassionate figure. Moreover, factual errors, such as incorrect character names and the addition of non-existent details like physical descriptions, further undermined the credibility of the AI-generated content. These inaccuracies emphasize the limitations of AI when interpreting complex, emotionally charged, and contextually nuanced literary works. The findings align with previous research indicating that AI struggles to grasp the

subjective nature of literature, which often requires an understanding of subtext, metaphor, and cultural context (Mostafapour et al., 2024).

The discrepancy in AI performance across disciplines is a crucial finding. In linguistics, AI models like ChatGPT can access a wealth of structured knowledge, theories, and frameworks that are more readily encoded into its algorithm. Concepts such as Linguistic Relativity have clear definitions, supporting examples, and established academic discourse, which makes it easier for AI to generate accurate responses. On the other hand, literary analysis requires a different set of cognitive skills—those related to emotional intelligence, critical thinking, and an appreciation for ambiguity and symbolism. These are aspects that AI models have not yet fully mastered, leading to weaker performance in literary analysis.

Another key result of this study was the observation that AI performs better when dealing with topics that are widely studied and critiqued. The theory of Linguistic Relativity has a long-standing presence in academia and has been the subject of numerous studies, papers, and debates, making it easier for AI to synthesize accurate responses based on the available data. In contrast, *The Blue Room* is a Pakistani novel with limited critical attention and fewer academic resources available online, which may have contributed to the AI's inaccurate portrayal of Molvi Jalal's character. This suggests that AI's performance can be influenced by the availability of critical literature and the volume of data it can draw upon, with better performance in areas with abundant resources.

The errors in the AI's analysis of Molvi Jalal's character also raise concerns about the potential risks associated with the use of AI for literary analysis. If students or researchers rely heavily on AI-generated content without cross-referencing with primary texts or scholarly critiques, they risk perpetuating misinformation. In this case, ChatGPT not only misrepresented key character traits but also fabricated details, such as character names and thematic interpretations, that were not present in the novel. This indicates a need for caution when using

AI in academic settings, particularly in fields where interpretation plays a critical role. Students and educators must be aware of the limitations of AI and ensure that AI-generated content is critically evaluated against primary sources and scholarly work.

The findings also highlight a critical issue regarding the lack of transparency in AI's data sources. ChatGPT did not specify the origins of its information regarding Molvi Jalal's character, which raises questions about the quality and reliability of the data it uses to generate responses. Without clear references or citations, it becomes difficult to assess the credibility of the AI's analysis. This is particularly concerning in academic fields where the accuracy and validity of information are paramount. While AI can be a helpful tool in providing preliminary overviews or assisting in research, its lack of transparency regarding sources makes it unsuitable for tasks requiring rigorous academic scrutiny.

Moreover, the study revealed that while AI can provide valuable support in educational settings, its use must be carefully supervised. In linguistics, where structured theories and well-defined concepts dominate, AI can serve as an effective tool for assisting with complex topics, offering clear explanations and relevant examples. However, in literature, where interpretation varies widely depending on cultural context and subjective perspectives, AI should be used with caution. The inaccuracies in the AI's literary analysis underscore the importance of human oversight, particularly when dealing with disciplines that require critical thinking and interpretation.

The results also support the hypothesis that AI's effectiveness varies depending on the subject matter. This study demonstrated that AI models like ChatGPT are more capable in structured disciplines such as linguistics, where established theories and frameworks guide the generation of content. In contrast, literature, which relies heavily on emotional nuance, interpretive flexibility, and cultural context, poses significant challenges for AI. This performance gap highlights the limitations of AI in replicating the cognitive and emotional processes involved in literary analysis.

Another aspect worth discussing is the potential for AI models to evolve and improve in fields where they currently struggle. While AI's performance in literary analysis is currently subpar, future advancements in natural language processing and machine learning could potentially allow AI to better understand and interpret complex literary themes. However, this would require significant improvements in AI's ability to process emotional subtext, cultural nuances, and the subjective nature of literary interpretation. Until then, AI should be viewed as a complementary tool rather than a replacement for human analysis in literature.

The results of this study provide valuable insights into the strengths and limitations of AI in academic fields. While AI demonstrates high accuracy and reliability in structured disciplines such as linguistics, it struggles with more interpretive and subjective fields like literature. This performance gap emphasizes the need for critical evaluation and human oversight when using AI in academic settings, particularly in disciplines that require deep interpretation and critical thinking. Furthermore, the findings call for greater transparency in AI's data sources to ensure the reliability and credibility of the information it generates.

5.2. RECOMMENDATIONS

Based on the findings of this study, several recommendations can be made for the effective use of AI in academic settings:

- i. **Subject-Specific Utilization:** AI tools such as ChatGPT should be used more extensively in disciplines where well-defined theories and structured information dominate, like linguistics. In fields where concepts are more rigidly defined, AI can be a reliable resource for students and researchers, providing accurate explanations and examples.
- ii. **Critical Evaluation in Literary Analysis:** In literature and other context-dependent subjects, AI responses should be used cautiously. Educators and students are encouraged to use AI-generated content as a supplementary resource, not as a primary source of information. AI responses must be cross-referenced with primary texts and scholarly critiques to ensure accuracy.

- iii. **Transparency in AI Data Sources:** It is essential for AI developers to improve transparency by providing clear references and citations for the information AI generates. This would allow users to verify the accuracy and reliability of AI-generated content and reduce the risk of misinformation.
- iv. **Enhanced AI Training for Subjective Fields:** Developers should consider refining AI models by incorporating more sophisticated techniques for processing cultural context, emotional subtext, and interpretive ambiguity, particularly in fields like literature, philosophy, and history. Training AI with more comprehensive data sets from diverse literary traditions can improve its interpretative abilities.
- v. **Human-AI Collaboration:** AI should be viewed as a complementary tool rather than a replacement for human cognition. In fields where creativity, emotional intelligence, and critical thinking are central, a collaborative approach—where AI assists but does not replace human judgment—would optimize its utility.

5.3. LIMITATIONS

The study focused on a relatively less-known literary work, *The Blue Room* by Nafisa Rizvi. Due to the limited critical evaluation of this novel, the AI's inability to generate an accurate analysis could be partly attributed to a lack of available data on the work. Thus, the generalizability of the findings may be limited when compared to more widely studied texts. Moreover, the linguistic analysis in this study centered solely on the theory of Linguistic Relativity. While the AI performed well in this area, it would be important to test its capabilities across a wider range of linguistic theories to develop a more comprehensive understanding of its effectiveness in the field of linguistics.

In addition, the AI responses analyzed in this study were generated by ChatGPT, but the specific version of the model and its training data were not assessed in detail. Since AI models frequently evolve with updated datasets, the results may not fully reflect the potential of future or newer AI models. The results were influenced by the quality of the prompts given by the user to the AI. Different prompts may yield different levels of detail and accuracy, which introduces variability that could affect the conclusions of the study.

5.4. FUTURE DIRECTIONS

Broader Literary Analysis: Future research could explore AI's performance across a diverse range of literary works, particularly well-known novels with extensive critical literature. This would provide a more comprehensive evaluation of the AI's capacity to interpret complex characters, themes, and literary devices.

While the AI performed well in addressing Linguistic Relativity, further studies should assess its capabilities across other key linguistic theories, including sociolinguistics, phonology, and syntax. This would offer a more holistic view of AI's role in assisting with linguistic research. Secondly, A comparison of different AI models, such as GPT-4, Bard, and other advanced language models, could provide insights into which models perform better in certain disciplines. Such comparative studies could identify the strengths and weaknesses of various AI systems, leading to more informed choices in educational and research contexts.

Thirdly, Future developments in AI could focus on improving its ability to interpret cultural and emotional nuances in literary analysis. Training AI on more diverse, culturally rich data sets could enhance its capability to grasp the subtleties that are often missed in current models. Fourthly, another potential avenue for future research is the integration of AI in multimodal learning environments. Studying how AI responds to interdisciplinary fields that combine both objective and subjective elements, such as cultural studies or media analysis, could yield insights into the broader applications of AI in education.

By addressing these limitations and pursuing these future directions, the academic community can better understand the evolving role of AI in education and research, ensuring that it is used to its full potential while minimizing its limitations.

5.5. CONCLUSION

The assessment of the AI generated responses in two closely related fields i.e. literature and linguistics declared that the AI can easily handle the disciplines with systematic and well-developed ideas such as linguistics, on the other hand, it faces various challenges while dealing with the complex and interpretive subjects like literature. AI's failure in portraying Molvi Jalal's character emphasizes the need for improvement in AI systems in the domain of literature.

The comparative analysis used in this research will help the AI developers to identify the domain that requires improvement. The extent of accuracy shown in this research depends highly on the nature of the discipline on which AI is directed to generate content; i.e., interpretive literary tasks seems challenging for AI, however, AI is proficient in structured linguistics knowledge. The research is significant as it addresses the requirement for improvements in a highly prevailing system in educators and learners from all the academic disciplines and other fields of life as well.

WORKS CITED

1. Amirjalili, F., Neysani, M., & Nikbakht, A. (2024). Exploring the boundaries of authorship: A comparative analysis of AI-generated text and human academic writing in English literature. *Frontiers in Education*.
2. Bakuuro, J., Pelpuo, R. H., & Tuurosong, D. (2018). Decoding the underpinning assumptions of linguistic theories: The lens on structural linguistics.
3. Casal, J., & Kessler, M. (2023). Can linguists distinguish between ChatGPT/AI and human writing?: A study of research ethics and academic publishing. *Research Methods in Applied Linguistics*. <https://doi.org/10.1016/j.rmal.2023.100047>
4. Chaka, C. (2023). Generative AI chatbots - ChatGPT versus YouChat versus Chatsonic: Use cases of selected areas of applied English language studies. *International Journal of Learning, Teaching and Educational Research*, 22(6), 319–335. <https://doi.org/10.26803/ijlter.22.6.18>
5. Ghazizadeh, E., & Zhu, P. (2020). A systematic literature review of natural language processing: Current state, challenges, and risks. *In Proceedings of the Future Technologies Conference (FTC) 2020*, Volume 1.

6. Keezhatta, M. S. (2019). Understanding EFL linguistic models through the relationship between natural language processing and artificial intelligence applications. *Social Science Research Network*.
7. Khalifa, M., Albadawy, M., & ZoteroMendeley. (2024). Using artificial intelligence in academic writing and research: An essential productivity tool. *Computer Methods and Programs in Biomedicine Update*.
8. Kidd, C., & Birhane, A. (2023). How AI can distort human beliefs. *Science*, 380(6649), 1314–1316. <https://doi.org/10.1126/science.adi0248>
9. Lin, Z. (2023). Techniques for supercharging academic writing with generative AI. *Nature Biomedical Engineering*.
10. Mazur, R. (2024). Language accuracy of texts generated by AI: A case study of ChatGPT. *LingVaria*.
11. Misiejuk, K., Kaliisa, R., & Scianna, J. (2024). Augmenting assessment with AI coding of online student discourse: A question of reliability. *Computers and Education: Artificial Intelligence*. <https://doi.org/10.1016/j.caeai.2024.100216>
12. Mostafapour, M., Fortier, J. H., & Garber, G. E. (2024). Evaluating literature reviews conducted by humans versus ChatGPT: Comparative study. *JMIR AI*. <https://doi.org/10.2196/48932>
13. Pathak, P. V., & Mehta, V. D. (2024). Intelligent conversations: A theoretical framework for understanding natural language processing within artificial intelligence systems. *International Journal for Multidisciplinary Research*.
14. Patil, A. (2024). Leveraging AI for historical linguistics. *International Journal for Multidisciplinary Research*.

15. Raj, A. V. T., Udayakumar, U., & Saravanan, D. (2023). Integrating artificial intelligence in English literature: Exploring applications, implications, and ethical considerations. *International Journal of Advanced Research in Science, Communication and Technology*.
16. Rizvi, N. (2009). "*The Blue Room*". Sama Editorial and Publishing Services; UK ed. Edition.
17. Rugaiyah, R. (2023). The potential of artificial intelligence in improving linguistic competence: A systematic literature review. *Arkus*.
18. Sapir, E. (1929). *The principles of linguistic relativity*. In *Language and Culture: The Principles of Linguistic Relativity* (pp. 120–137). Harvard University Press.
19. Thakkar, K., & Jagdishbhai, N. (2023). Exploring the capabilities and limitations of GPT and ChatGPT in natural language processing. *Journal of Management*.
20. Whorf, B. L. (1956). *Language, thought, and reality: Selected writings*. MIT Press.
21. William, F. K. A. (2024). AI in academic writing: Ally or foe? *International Journal of Research Publications*.
22. Wu, T. S., Ribeiro, M. T., & Weld, D. S. (2019). Errudite: Scalable, reproducible, and testable error analysis. *Annual Meeting of the Association for Computational Linguistics*. <https://doi.org/10.18653/v1/P19-3017>
23. Young, P. S. (2020). A study on the possibility and limitations of artificial intelligence literature through creativity. <https://doi.org/10.33645/cnc.2020.08.42.8.619>

24. Yuan, J., Vig, J., & Rajani, N. (2022). iSEA: An interactive pipeline for semantic error analysis of NLP models. *International Conference on Intelligent User Interfaces*. <https://doi.org/10.1145/3490099.3511139>
25. Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*.