International Journal of Global Research Innovations & Technology (IJGRIT)
ISSN : 2583-8717, Impact Factor: 6.972, Volume 03, No. 01(II), January-March, 2025, pp 91-94

ANALYSING THE DIVERSITY IN AI RESPONSES TO IDENTICAL USER QUERIES: EMPHASIZING LEXICAL DIVERSITY

Duha Mukhtar Kashtwari*

ABSTRACT

This research paper investigates the diversity in AI-generated responses to identical user queries, with a focus on lexical diversity. The study evaluates various AI models, including OpenAI's ChatGPT, Google's Bard, and Gemini, by measuring variations in vocabulary richness using the Type-Token Ratio (TTR). By examining response diversity, this study aims to assess and analyse AI-generated responses using the Type-Token Ratio (TTR), this study aims to compare the linguistic diversity across three leading AI models: ChatGPT, Bard, and Gemini and how AI models maintain creativity, coherence, and adaptability in user interactions. Additionally, graphical representations of data findings provide insight into diversity trends across AI platforms.

Keywords: *AI Response Diversity, Lexical Diversity, AI Models, Type-Token Ratio, ChatGPT, Gemini.*

Introduction

The rapid advancement of artificial intelligence has led to the development of sophisticated conversational models. AI-powered chatbots and virtual assistants are now integral to various applications, including customer support, content creation, and personal assistance. However, AI-generated responses can vary significantly, even when presented with identical user queries **(Jurafsky & Martin, 2021).** Understanding these variations is crucial in evaluating an AI's ability to generate diverse, contextually rich, and meaningful interactions. This paper focuses on lexical diversity as a key parameter in assessing response variations across AI models.

- **ChatGPT:** Developed by OpenAI, ChatGPT is based on the GPT-3.5 and GPT-4 architectures. It is designed to generate human-like text using deep learning techniques. ChatGPT has been widely used for conversational AI, coding assistance, and content creation **(Radford et al., 2019).**
- **Bard:** Created by Google, Bard utilizes the LaMDA (Language Model for Dialogue Applications) architecture. Bard focuses on contextual understanding and delivering fact-based responses, making it suitable for information retrieval and structured conversational tasks **(Thoppilan et al., 2022).**
- **Gemini:** Also developed by Google, Gemini represents an evolution of Bard, incorporating multimodal capabilities and enhanced NLP techniques to improve response diversity and adaptability in dialogue-based applications.

Lexical diversity refers to the range of vocabulary used in generated text, can influence user engagement, readability, and coherence. A higher lexical diversity generally indicates a richer vocabulary and greater flexibility in generating varied responses. On the other hand, low lexical diversity may result in repetitiveness, reducing user engagement and conversational depth.

* Research Scholar, Jagannath University, Bahadurgarh, NCR, Haryana, India.

Literature Review

Research on AI-generated text has primarily focused on coherence, fluency, and factual accuracy **(Bender et al., 2021).** Prior studies have examined linguistic diversity in AI responses, emphasizing the importance of variation in improving user engagement and reducing repetitiveness **(Zhou et al., 2020).** However, a little research has focused on a direct comparative study between the multiple AI models regarding lexical diversity. This study extends previous work by systematically analysing lexical diversity in AI-generated outputs across various platforms.

Lexical diversity plays a critical role in natural language generation, impacting readability, engagement, and linguistic flexibility. Studies have shown that a high degree of lexical diversity enhances user trust in AI-generated content, while lower diversity may lead to monotonous interactions **(Baayen, 2001).** Furthermore, previous research highlights that AI models trained on large-scale datasets often demonstrate higher lexical richness due to their broad exposure to linguistic structures **(Radford et al., 2019).**

Methodology

This study employs a comparative analysis approach to evaluate lexical diversity among ChatGPT, Bard, and Gemini. The methodology consists of the following components:

- **Dataset:** A standardized set of 10 prompts covering various domains (e.g., general knowledge, opinions, technical queries) was used. Each AI model received the same set of prompts to ensure consistency in comparison. Sample queries include:
  - "Explain the significance of photosynthesis in plant biology."
  - "What are the implications of artificial intelligence in modern society?"
  - "Provide an opinion on the role of social media in shaping public opinion."
  - "Describe the key principles of quantum mechanics."
  - "Discuss the ethical considerations of AI in healthcare."
  - "Summarize the impact of climate change on biodiversity."
  - "How does blockchain technology enhance cybersecurity?"
  - "Compare and contrast classical and operant conditioning in psychology."
  - "What are the economic consequences of inflation?"
  - "Outline the fundamental differences between socialism and capitalism."
- AI Models Evaluated: ChatGPT, Bard, and Gemini.
- Metric Used: Type-Token Ratio (TTR): Measures vocabulary richness using the formula:

$$\textbf{TTR=}\ \frac{\textbf{Number of Unique Words}}{\textbf{Total Number of Words}}$$

A higher TTR indicates greater lexical diversity, while a lower TTR suggests a higher frequency of word repetition.

Derivation of TTR for a sample response:

"Artificial intelligence is transforming industries by automating complex tasks and improving efficiency."

Total words (tokens) = 13

Unique words (types) = 11

TTR = 11 / 13 = 0.846

Data Analysis & Interpretation

The data analysis for AI Model Lexical Diversity Comparison was based on the three AI Models with the Type-Token Ratio (TTR) being the metric for measuring the lexical diversity. A standardized set of 10 prompts covering various domains (e.g., general knowledge, opinions, technical queries) was used. Each AI model received the same set of prompts to ensure consistency in comparison. To derive the desired results, metric used was Type-Token Ratio (TTR) which measures vocabulary richness using the formula:

$$\textbf{TTR=}\ \frac{\textbf{Number of Unique Words}}{\textbf{Total Number of Words}}$$

- **AI Model Lexical Diversity Comparison**

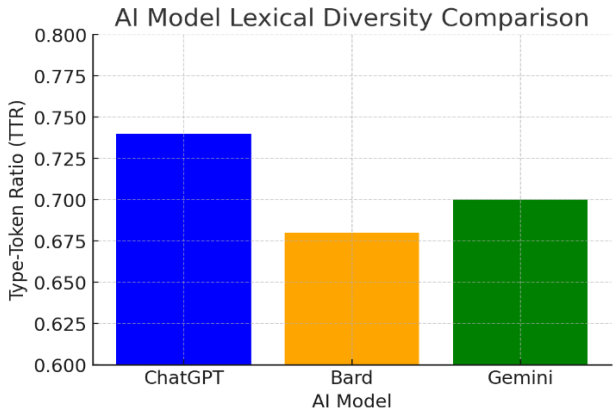

Figure 1: AI Model Lexical Diversity Comparison

- **TTR Variations Across Different Query Types**
  - **General Knowledge Queries:** ChatGPT demonstrated the highest TTR, suggesting greater variability in explanations and examples. Bard had the lowest TTR, possibly due to its tendency for structured, fact-based responses.

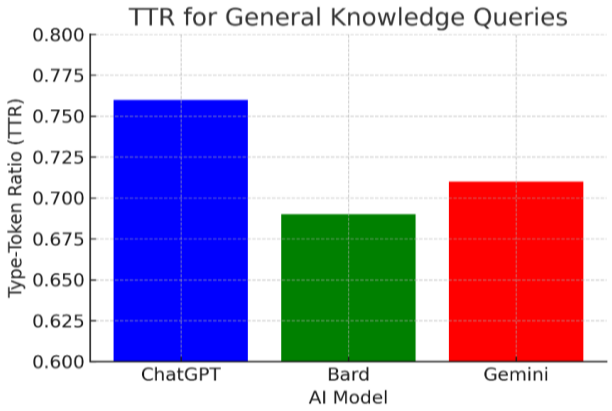

Figure 2: General Knowledge Queries Type-Token Ratio

  - **Opinion-Based Queries:** ChatGPT and Gemini exhibited relatively high TTR, likely due to their capacity for generating varied perspectives. Bard showed lower TTR, reflecting a more formal and repetitive approach to opinion-based discussions.

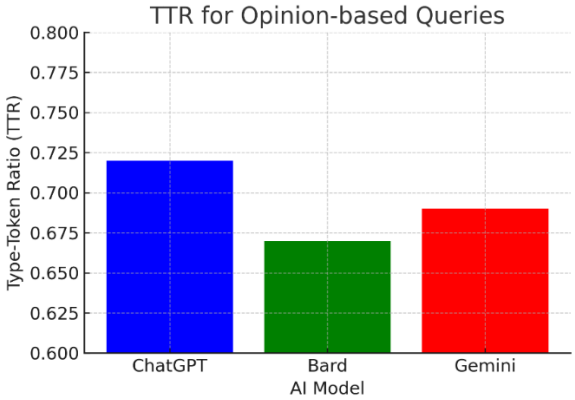

Figure 3: Opinion-Based Queries Type-Token Ratio

- **Technical Queries:** ChatGPT scored a TTR of 0.74, likely due to its ability to explain technical concepts in different ways. Bard had a lower TTR (0.68), reflecting its tendency to provide more factual, structured responses. Gemini (0.70) demonstrated a flexible approach, balancing technical accuracy with lexical variation.

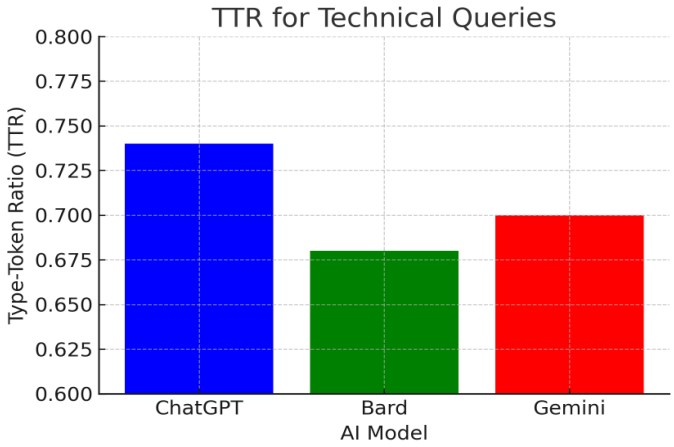

Figure 4: Technical Queries Type-Token Ratio

Findings & Conclusion

The results indicate that ChatGPT exhibits the highest lexical diversity, closely followed by Gemini. Bard's responses are more structured but less diverse. These results suggest that ChatGPT generally generates more lexically diverse responses across different query types, while Bard tends to prioritize structured and concise language, leading to lower lexical variation. ChatGPT and Gemini maintained a relatively higher TTR, indicating adaptability in complex explanations. Bard had a slightly lower TTR, possibly due to its structured, formulaic approach in technical discussions. These findings suggest that lexical diversity correlates with an AI's training data breadth and response-generation strategies **(Radford et al., 2019).**

Future Scope

Understanding lexical diversity in AI responses is essential for enhancing user engagement and content variability. This research concludes that while ChatGPT leads in lexical richness, Gemini presents a competitive level of diversity. Bard, on the other hand, exhibits a tendency for repetitive structures. These findings underscore the importance of continuous improvements in AI training methodologies to enhance response diversity.

Future research should explore multimodal AI response diversity, sentiment variations, and the impact of prompt engineering techniques **(Brown et al., 2020).** Expanding the dataset to include domain-specific prompts (e.g., medical, legal, creative writing) may yield further insights into AI diversity. Additionally, investigating the correlation between response coherence and lexical diversity could provide deeper insights into AI linguistic capabilities.

References

1. Baayen, H. (2001). *Word Frequency Distributions*. Springer.
2. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of FAccT*.
3. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language Models Are Few-Shot Learners. *Advances in Neural Information Processing Systems*.
4. Jurafsky, D., & Martin, J. H. (2021). *Speech and Language Processing* (3rd ed.). Prentice Hall.
5. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models Are Unsupervised Multitask Learners. *OpenAI Blog*.
6. Zhou, W., Huang, W., & Liu, Y. (2020). Evaluating Linguistic Diversity in AI-Generated Text. *Proceedings of ACL*.