

8-16-2024

## Exploring Diversity in Large Language Model Outputs: A Study on Product Reviews

Lars Thomsen

*Martin-Luther-University Halle-Wittenberg, lars.thomsen@wiwi.uni-halle.de*

Ralf Peters

*Institute for Information Systems and Operations Research, ralf.peters@wiwi.uni-halle.de*

Follow this and additional works at: [https://aisel.aisnet.org/treos\\_amcis2024](https://aisel.aisnet.org/treos_amcis2024)

---

### Recommended Citation

Thomsen, Lars and Peters, Ralf, "Exploring Diversity in Large Language Model Outputs: A Study on Product Reviews" (2024). *AMCIS 2024 TREOs*. 7.

[https://aisel.aisnet.org/treos\\_amcis2024/7](https://aisel.aisnet.org/treos_amcis2024/7)

This material is brought to you by the AIS TREO Papers at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2024 TREOs by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# Exploring Diversity in Large Language Model Outputs: A Study on Product Reviews

TREO Talk Paper

**Lars Thomsen**

Martin-Luther-University

lars.thomsen@wiwi.uni-halle.de

**René Theuerkauf**

Martin-Luther-University

rene.theuerkauf@wiwi.uni-halle.de

**Ralf Peters**

Martin-Luther-University

ralf.peters@wiwi.uni-halle.de

## Abstract

Past studies have demonstrated the impact of product reviews on consumer purchasing behavior. Consequently, there's a legitimate interest in strategically placing these reviews to strengthen one's position in a market. Generative Artificial Intelligence, particularly Large Language Models (LLMs), offers the ability to create texts of human-like quality, facilitating the generation of diverse reviews with manageable effort. A multitude of diverse reviews not only influences product visibility but also conversion rates. However, the control of diversity in review generation within the framework of LLMs has not been fully explored. Our work aims to investigate the diversity of LLM-generated texts using product reviews as a case study. We differentiate the diversity between lexical and semantic aspects. Lexical diversity refers to the ratio of unique words to the total number of words. It is assessed using the Self-BLEU score. Semantic diversity considers the meaning-based content of the output set. To quantize the diversity, the cosine similarities of the reviews embeddings transformed by a specific encoder are calculated. In our research approach, we investigate two core components in the use of LLMs to control diversity that have been identified as relevant in the literature. Using Google's state-of-the-art LLM Gemma, we apply the *prompt* in two variants to represent single- and multiple-output prompting and iterate over different values of the *temperature* to control the creativity of the output.

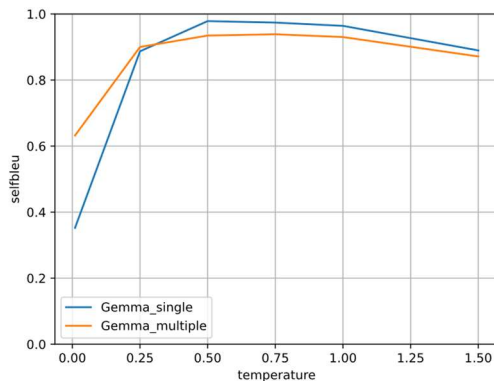


Figure 1. Self-BLEU Score\*

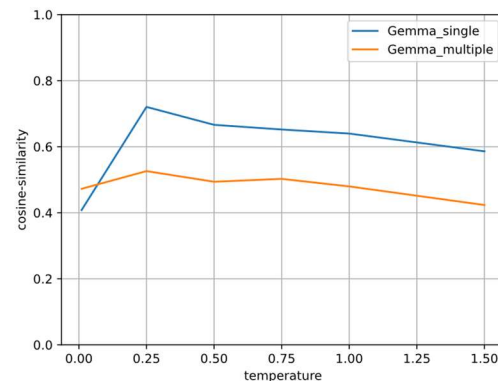


Figure 2. Cosine-Similarity Score

Regarding the lexical aspect, a high Self-BLEU score indicates lower diversity in the output set. Figure 1 illustrates a decrease in diversity up to a temperature of 0.75, followed by an increase with higher temperatures. Moreover, we show that multiple-output prompting leads to more diverse results. Considering the semantical aspect, a low cosine similarity score suggests greater diversity among the examined texts. Figure 2 describes the semantic diversity of the generated reviews.

Based on our preliminary findings, we plan to explore multiple-output prompting more profound by *increasing* the requested reviews per prompt. We will apply *few-shot prompting* to examine how example reviews influence the output. This entails using varying numbers of example reviews as input and assessing the generated output's diversity accordingly. We also intend to measure the diversity between the input reviews and the output reviews. Lastly, the consideration of *time complexity* becomes prominent due to the anticipated differences in token usage and increasing requirements.