Model Fine-Tuning, Evaluation, and Text Analysis

# 1. Introduction

This report summarizes the findings from a project focused on the fine-tuning and evaluation of a language model, alongside text analysis and visualization. The associated GitHub repository contains the source code, documentation, evaluation results, visualizations.

link

https://github.com/Nazeermastan/LLMtask

# 2. Approach and Methodology

## 2.1. Fine-Tuning the Model

The model was fine-tuned using a specific dataset, adjusting hyperparameters to optimize performance, and applying early stopping criteria to avoid overfitting.

## 2.2. Evaluation Metrics

Two primary metrics were used to evaluate the model:  
- Perplexity: A metric for measuring how well a model predicts the next word in a sequence. Lower values indicate better performance.  
- ROUGE Scores: Measures the overlap between generated and reference texts. Precision, recall, and F1 scores were computed for ROUGE-1 and ROUGE-L.

# 3. Results

## 3.1. Perplexity

Perplexity: 16.92

## 3.2. ROUGE Scores

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Reference | ROUGE-1 Precision | ROUGE-1 Recall | ROUGE-1 F1 | ROUGE-L Precision | ROUGE-L Recall | ROUGE-L F1 |
| Text 1 | 0.75 | 0.032 | 0.0619 | 0.75 | 0.032 | 0.0619 |
| Text 2 | 0.125 | 0.019 | 0.0333 | 0.125 | 0.019 | 0.0333 |
| Text 3 | 0.714 | 0.031 | 0.0599 | 0.714 | 0.031 | 0.0599 |
| Text 4 | 0.091 | 0.025 | 0.0392 | 0.091 | 0.025 | 0.0392 |
| Text 5 | 1.0 | 0.0097 | 0.0192 | 1.0 | 0.0097 | 0.0192 |

## 3.3. Visualization of Top 100 Word Pairs

The visualization depicting the top 100 word pairs highlights the most frequent co-occurring words in the generated text. The analysis offers insight into thematic patterns.

Visualization 1: Top 100 Word Pairs

![Visualization 1](visualization1.png)

Visualization 2: Another Textual Insight

![Visualization 2](visualization2.png)

# 4. Challenges

The following challenges were encountered:  
- Data Preprocessing: Handling special tokens and formatting the data for training took significant effort.  
- Computational Resources: Training the model on large datasets required optimization to stay within memory limits.  
- Overfitting: Careful regularization was needed to prevent overfitting, especially on smaller datasets.

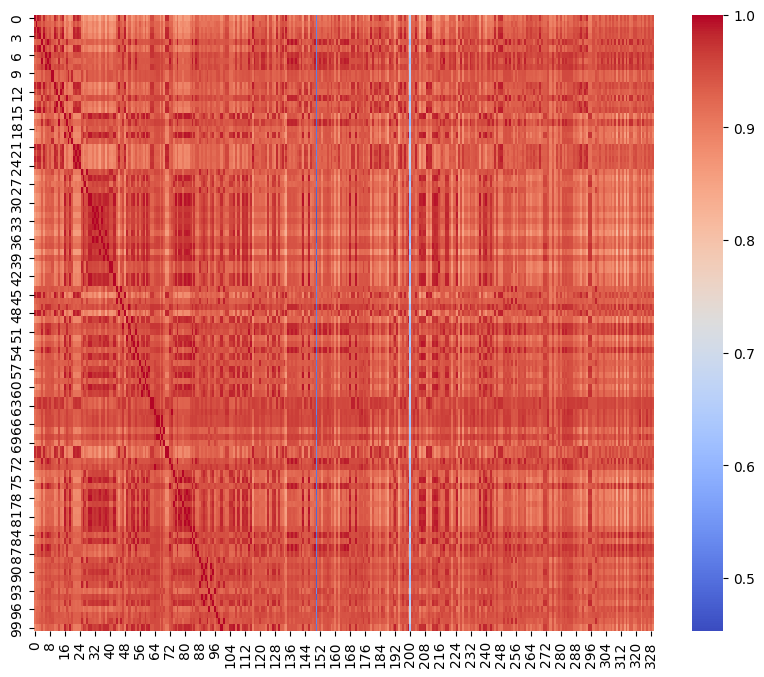
# 5. Areas for Improvement

Future improvements could include:  
- Parameter Tuning: Further optimization of hyperparameters may enhance model performance.  
- Data Augmentation: Expanding the dataset could improve generalization.  
- Evaluation Metrics: Incorporating additional metrics like BLEU or METEOR could provide deeper insights.  
- Interactive Visualization Enhancements: Improving the interactive components for deeper exploration of text relationships.

## 3.3. Visualization of Top 100 Word Pairs

The visualizations below provide insights into the relationships between word pairs in the generated texts.

Visualization 1: Top 100 Word Pairs



Visualization 2: Another Textual Insight

