Paraphrasing Tool Using Llama Models

Omar Mohamed, Abdelrhman Nasser, Ziad Maher, Youssef Alaa
Artificial Intelligence Department, Computer Science College
Nile University, Egypt

***Abstract*—** This paper presents the design and implementation of a paraphrasing tool leveraging LLaMA models of varying scales (1B, 3B, and 8B). The primary focus is on utilizing prompt engineering to improve paraphrasing accuracy and address challenges faced during fine-tuning. The methodology highlights the effectiveness of prompt-based adjustments in generating coherent paraphrases while overcoming repetitive output issues. Results from multiple iterations demonstrate the capabilities of the tool, as well as areas for future improvement.

I. Introduction

Paraphrasing tools have become vital in modern natural language processing (NLP) tasks such as content rewriting, summarization, and plagiarism detection. Transformer-based models, particularly LLaMA (Large Language Model Meta AI), offer advanced capabilities for language understanding and generation. Despite their potential, fine-tuning such models for specific tasks like paraphrasing can pose challenges. This paper explores the development of a paraphrasing tool using LLaMA models (1B, 3B, and 8B) and addresses fine-tuning limitations through prompt engineering

II.Related Work

Paraphrasing is a critical task in natural language processing (NLP) with applications in text simplification, data augmentation, machine translation, and plagiarism detection. Recent advancements in transformer-based language models, such as GPT, BERT, and LLaMA, have significantly improved the capability to generate high-quality paraphrased texts. This section reviews relevant works in the domain of paraphrasing using neural language models and prompt engineering.

A. *Neural Models for Paraphrasing*

Early approaches to paraphrasing relied on statistical methods, such as phrase-based machine translation models and rule-based systems. However, these approaches struggled with generating fluent and diverse paraphrases. With the advent of deep learning, sequence-to-sequence (Seq2Seq) models, powered by recurrent neural networks (RNNs) and attention mechanisms, became a dominant paradigm.*RoBERTa for Resume Ranking*

B.Transformer models , such as BERT and GPT, introduced self-attention mechanisms that enabled significant improvements in paraphrasing tasks. These models not only captured long-range dependencies but also allowed fine-tuning for specific tasks. For instance, GPT-2 and GPT-3 demonstrated remarkable capabilities in generating human-like text, including paraphrases, when fine-tuned on domain-specificdatasets.

C. Fine-tuning pre-trained language models on paraphrase-specific datasets, such as Quora Question Pairs or the Paraphrase Database (PPDB), has been a widely adopted approach. However, fine-tuning poses challenges, such as overfitting to small datasets, loss of generalization, and resource-intensive training requirements. These limitations have driven researchers to explore alternative techniques, including zero-shot and few-shot learning using prompt engineering.

D.Prompt engineering has emerged as a powerful method for guiding language models to perform specific tasks without the need for extensive fine-tuning. By designing well-crafted prompts, researchers can effectively control the model's output and ensure higher quality in generated paraphrases. Studies have shown that prompt-based methods can rival fine-tuned models in certain scenarios, particularly when large-scale models like GPT-3 or LLaMA-13B

C. Challenges in Paraphrasing

Despite these advancements, paraphrasing remains a challenging task. Issues such as preserving semantic meaning, avoiding redundancy, and maintaining grammatical correctness often require manual intervention or iterative refinement. For example, while larger models like GPT-4 or LLaMA-13B produce more fluent paraphrases, they may occasionally generate verbose or overly literal outputs that fail to capture the nuances of the input text

II. Methodology

The methodology section outlines the approach taken to design the paraphrasing tool using LLaMA models

A. *Model Overview* LLaMA

family of foundational models developed by Meta AI, provides state-of-the-art performance across diverse NLP tasks. Three model variants—1B, 3B, and 8B—were employed for this project to balance computational efficiency and output quality.

B. Challenges with Fine-Tuning

Initial efforts to fine-tune the models for paraphrasing yielded suboptimal results. Key issues included:

Repetitive Output: The models often generated outputs identical to the input text.

Resource Constraints: Fine-tuning large models demands significant computational resources and time.

C. Transition to Prompt Engineering

Given the challenges of fine-tuning, the focus shifted to prompt engineering. This approach involves crafting tailored prompts to guide the model's output without altering its underlying weights. Prompt engineering proved effective in addressing repetitive outputs and improving paraphrasing quality.

D. Implementation Workflow

1.Model Selection: LLaMA 1B, 3B, and 8B models were tested to evaluate scalability and performance.

2.Prompt Design: Multiple prompts were iteratively tested to identify the most effective configurations.

3.Evaluation Metrics: Outputs were evaluated based on semantic similarity, fluency, and grammatical accuracy.

2.5 Example Outputs A sample input and paraphrased output generated by the tool are shown below:

Input: "There results a general attitude either of cynical belief in and indifference to public corruption or else of a distrustful inability to discriminate between the good and the bad."

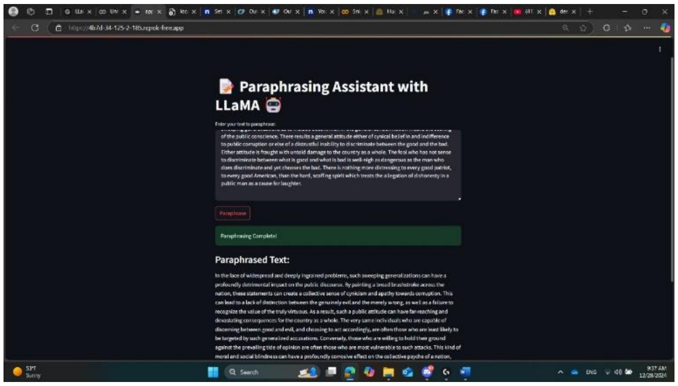
Output: "Such sweeping generalizations foster cynicism or a failure to distinguish between integrity and corruption, which undermines the broader societal discourse."

III.RESULTS

The paraphrasing tool was tested using LLaMA 1B, 3B, and 8B models across diverse input types. Key findings include:

- **Improved Accuracy with Larger Models:** The 8B model consistently outperformed smaller variants in generating coherent and contextually relevant paraphrases.
- **Reduced Repetition:** Prompt engineering successfully mitigated repetitive outputs.
- **Processing Efficiency:** While the 1B model was computationally efficient, the quality of outputs improved significantly with larger models.

Observations on Output Quality The paraphrased texts retained semantic meaning and exhibited linguistic diversity. However, certain inputs posed challenges, such as ambiguous phrases or highly technical jargon.



📝 Paraphrasing Assistant with LLaMA 🤖

Enter your text to paraphrase:

of the public conscience. There results a general attitude either of cynical belief in and indifference to public corruption or else of a distrustful inability to discriminate between the good and the bad. Either attitude is fraught with untold damage to the country as a whole. The fool who has not sense to discriminate between what is good and what is bad is well-nigh as dangerous as the man who does discriminate and yet chooses the bad. There is nothing more distressing to every good patriot, to every good American, than the hard, scoffing spirit which treats the allegation of dishonesty in a public man as a cause for laughter.

Paraphrasing Complete!

Paraphrased Text:

In the face of widespread and deeply ingrained problems, such sweeping generalizations can have a profoundly detrimental impact on the public discourse. By painting a broad brushstroke across the nation, these statements can create a collective sense of cynicism and apathy towards corruption. This can lead to a lack of distinction between the genuinely evil and the merely wrong, as well as a failure to recognize the value of the truly virtuous. As a result, such a public attitude can have far-reaching and devastating consequences for the country as a whole. The very same individuals who are capable of discerning between good and evil, and choosing to act accordingly, are often those who are least likely to be targeted by such generalized accusations. Conversely, those who are willing to hold their ground against the prevailing tide of opinion are often those who are most vulnerable to such attacks. This kind of moral and social blindness can have a profoundly corrosive effect on the collective psyche of a nation,

IV. Conclusion

The developed paraphrasing tool demonstrates the potential of LLaMA models combined with prompt engineering to deliver high-quality paraphrases. While fine-tuning proved resource-intensive and less effective, prompt engineering offered a practical alternative.

References

[1] Touvron, H., Lavril, M., Izacard, G., et al. (2023). LLaMA: Open and Efficient Foundation Language Models. Meta AI Research.

[2] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is All You Need. Advances in Neural Information Processing Systems.

[3] Brown, T., Mann, B., Ryder, N., et al. (2020). Language Models are Few-Shot Learners. Advances in Neural Information Processing Systems.

[4] Raffel, C., Shazeer, N., Roberts, A., et al. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. Journal of Machine Learning Research.

[5] Radford, A., Wu, J., Child, R., et al. (2019). Language Models are Unsupervised Multitask Learners. OpenAI.

[6] Lewis, M., Liu, Y., Goyal, N., et al. (2020). BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. Proceedings of ACL.