Journal of COMPUTER AND FORENSIC SCIENCES

CFS 2025, Vol. 4, Issue 2, pp.17–29
https://doi.org/10.5937/jcfs4-64420
Original research paper

Semantic Paraphrase Generation Using Transformer Architectures: A Comparative Study of Pre-Trained and Fine-Tuned Models

Rahul Birwadkar[1*]

[1] Student, SRH University, Heidelberg, Germany

*Corresponding author: rbirwadkar6@gmail.com

Received: January 27, 2026 • Accepted: April 6, 2026 • Published: May 14, 2026

Abstract: Semantic paraphrase generation plays a crucial role in academic and technical writing by enabling authors to restate content while preserving its original meaning. Traditional paraphrasing approaches, such as rule-based rewriting and statistical methods, often struggle to maintain semantic consistency and linguistic fluency, especially for complex or longer text segments. Recent advances in transformer-based architectures have significantly improved text generation capabilities by leveraging contextual representations and self-attention mechanisms. This paper presents a comparative study of pre-trained and fine-tuned transformer models for semantic paraphrase generation. The study builds upon prior work presented in [1], extending the analysis of transformer-based approaches for paraphrase generation. We evaluate encoder–decoder transformer architectures, with a primary focus on the BART model in both pre-trained and fine-tuned settings, alongside a large generative language model used for paraphrase generation. The fine-tuning process adapts pre-trained models to paraphrasing tasks using task-specific data, enabling improved control over semantic preservation and output consistency. The evaluation is conducted using both quantitative and qualitative analyses, including training and validation loss trends and comparative examination of generated paraphrases. Experimental results demonstrate that fine-tuned transformer models produce paraphrases with higher semantic fidelity and structural coherence compared to their pre-trained counterparts, while large generative models offer fluent but less deterministic outputs. The findings highlight the importance of task-specific fine-tuning for controlled and semantically accurate paraphrase generation. This study contributes practical insights into the selection and adaptation of transformer architectures for paraphrasing applications, particularly in academic and research-oriented writing contexts.

Keywords: semantic paraphrase generation; transformer models; BART; fine-tuning; natural language processing.

1. INTRODUCTION

Paraphrasing is a fundamental practice in academic and technical writing, enabling authors to express existing ideas in new linguistic forms while preserving the original semantic meaning. Effective paraphrase generation supports clarity, originality, and ethical content creation, particularly in research communication where similar concepts are often discussed across multiple works. However, producing high-quality paraphrases that maintain semantic fidelity and grammatical correctness remains a challenging task for automated systems. Early approaches to paraphrase generation relied on rule-based methods and statistical techniques, such as synonym substitution and phrase-based machine translation. While these methods offered limited rewriting capabilities, they frequently failed to preserve contextual meaning, resulting in paraphrases that were either semantically distorted or linguistically unnatural. Neural network–based sequence-to-sequence models, including recurrent neural networks and long short-term memory architectures, improved fluency but faced scalability issues and difficulty in capturing long-range dependencies within text.

The introduction of transformer-based architectures has significantly advanced natural language generation tasks, including paraphrase generation. By employing self-attention mechanisms and parallelized processing, transformer models enable richer contextual understanding and improved text generation performance. Encoder–decoder architectures, in particular, have demonstrated strong capabilities in text rewriting tasks by learning to map input sentences to semantically equivalent outputs. Pre-trained transformer models further enhance this capability by leveraging large-scale language knowledge acquired during pre-training.

Despite these advancements, a key practical question remains: to what extent does task-specific fine-tuning improve the quality of paraphrase generation compared to using pre-trained models directly? While large pre-trained and generative models can produce fluent paraphrases, their outputs may lack consistency and control when applied to structured academic rewriting tasks. Fine-tuning transformer models on paraphrasing data offers a potential solution by adapting general language representations to task-specific objectives. This paper presents a comparative study of pre-trained and fine-tuned transformer models for semantic paraphrase generation. The study focuses on evaluating encoder–decoder transformer architectures, with particular emphasis on the BART model in both pre-trained and fine-tuned configurations, alongside a large generative language model used for paraphrasing. The evaluation combines quantitative analysis of training and validation behaviour with qualitative examination of generated paraphrases, emphasizing semantic preservation and structural coherence.

The contributions of this work are threefold:

1) This study provides an empirical comparison between pre-trained and fine-tuned transformer models for paraphrase generation,

2) It demonstrates the impact of task-specific fine-tuning on semantic consistency and output control, and

3) It offers practical insights for selecting transformer architectures in academic paraphrasing applications.

2. RELATED WORK

Paraphrase generation has been extensively studied within the field of natural language processing, evolving through multiple methodological stages ranging from rule-based rewriting to modern transformer-based architectures. The primary objective of paraphrase generation is to reformulate text while preserving its semantic meaning, a task that requires both linguistic fluency and contextual understanding. Early paraphrase generation techniques were predominantly rule-based and relied on manually crafted linguistic rules, lexical resources, and synonym dictionaries. These approaches typically performed word- or phrase-level substitutions using predefined rules and thesauri. While rule-based systems were straightforward to implement, they lacked robustness and often produced grammatically incorrect or semantically altered outputs due to their inability to capture contextual dependencies and deeper semantic relationships [2]. As a result, such methods were limited in their applicability to real-world writing tasks.

To overcome these limitations, statistical approaches were introduced, most notably phrase-based statistical machine translation (SMT). In this paradigm, paraphrase generation was treated as a monolingual translation problem, where sentences were probabilistically mapped to alternative surface forms using aligned phrase pairs [3]. Statistical methods improved linguistic fluency and introduced data-driven learning; however, their effectiveness was highly dependent on the availability of high-quality parallel corpora. Moreover, SMT-based paraphrasing struggled with long sentences and complex syntactic structures, limiting its scalability and generalization. The emergence of neural network–based sequence-to-sequence models marked a significant advancement in paraphrase generation. Recurrent neural networks (RNNs) and long short-term memory (LSTM) architectures enabled models to learn continuous latent representations of sentences, allowing for more flexible paraphrase generation [4]. Attention mechanisms further enhanced these models by allowing dynamic alignment between input and output sequences, improving semantic coherence and grammatical consistency [5]. Despite these improvements, recurrent architectures suffered from inherent drawbacks, including limited parallelization, slow training times, and difficulty in modelling long-range dependencies.

Transformer-based architectures have since become the dominant framework for paraphrase generation and other natural language generation tasks. By replacing recurrence with self-attention mechanisms, transformers enable parallel processing and more effective modelling of global contextual relationships within text [6]. Encoder–decoder transformer architectures, in particular, are well suited for paraphrase generation, as they explicitly learn mappings between semantically equivalent sentence representations. These models have demonstrated strong performance in text rewriting tasks by capturing both syntactic structure and semantic meaning.

The introduction of large-scale pre-trained transformer models further advanced paraphrase generation by leveraging linguistic knowledge learned from massive corpora during pre-training [7]. Models such as BART employ denoising autoencoder objectives that make them particularly effective for text reconstruction and rewriting tasks [8]. Fine-tuning these pre-trained models on paraphrase-specific datasets allows them to adapt to controlled rewriting objectives, improving semantic fidelity and output stability. Recent research has also explored the use of large generative language models for paraphrase generation. While these models are capable of producing fluent and diverse paraphrases, their generative flexibility often results in less deterministic behaviour and occasional semantic drift. This characteristic can be advantageous for creative text generation but poses challenges for applications that require precise semantic preservation, such as academic writing.

Despite the substantial progress achieved through transformer-based models, limited work has systematically compared pre-trained and fine-tuned transformer architectures for semantic paraphrase generation under controlled evaluation settings. This study addresses this gap by providing a focused comparative analysis that emphasizes semantic preservation, output consistency, and practical applicability in academic and research-oriented writing contexts. Recent work such as the T5 model [9] further unified text-to-text learning approaches, demonstrating strong performance across multiple natural language processing tasks, including paraphrase generation.

3. METHODOLOGY

This section describes the methodology adopted for semantic paraphrase generation using transformer-based architectures. The overall approach focuses on evaluating and comparing pre-trained and fine-tuned transformer models with respect to their ability to generate semantically consistent paraphrases while preserving grammatical structure, contextual meaning, and linguistic fluency.

3.1. Dataset Preparation and Preprocessing

The paraphrase generation task is formulated using paired text samples consisting of original sentences and their corresponding paraphrased versions. Each data instance is structured as an input–output pair, where the input represents the source sentence and the output represents a semantically equivalent paraphrase. This supervised formulation enables effective training and fine-tuning of encoder–decoder transformer models for controlled rewriting tasks. Prior to model training and evaluation, the textual data undergoes standard preprocessing steps to ensure consistency and compatibility with transformer-based architectures. These steps include text normalization, lowercasing, removal of unnecessary special characters, and sentence-level segmentation. The preprocessing pipeline is designed to preserve semantic content while eliminating noise that may negatively impact model learning and generation quality.

3.2. Transformer Architecture Overview

Transformer models are built upon self-attention mechanisms that enable the modelling of contextual relationships between all tokens in a sequence simultaneously. Unlike recurrent architectures, transformers process entire sequences in parallel, allowing efficient learning of long-range dependencies and improving scalability for large datasets [6]. The self-attention mechanism computes weighted interactions between tokens, enabling the model to capture both local and global contextual information.

In encoder–decoder transformer architectures, the encoder maps the input sentence into a sequence of contextualized representations, while the decoder generates the paraphrased output in an auto-regressive manner based on the encoded context. This architecture is particularly suitable for paraphrase generation, as it explicitly learns mappings between semantically equivalent textual representations rather than performing direct word-level substitutions. Figure 1 illustrates the transformer-based encoder–decoder architecture used for semantic paraphrase generation.

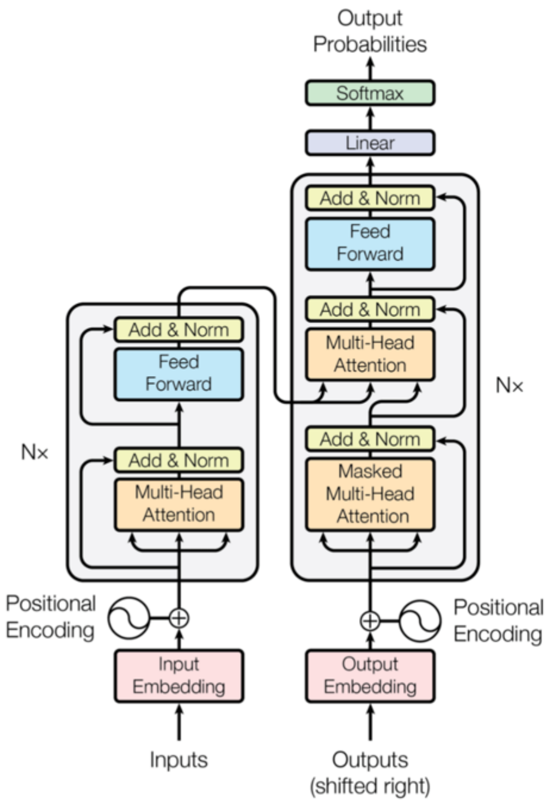

Figure 1. *Transformer-based encoder–decoder architecture used for semantic paraphrase generation. [6]*

3.3. Model Selection

To analyse the impact of pre-training and task-specific adaptation, multiple transformer-based models are considered in this study:

- **Pre-trained Encoder–Decoder Transformer Model:**A transformer model used directly for paraphrase generation without additional task-specific training. This configuration leverages general linguistic knowledge acquired during large-scale pre-training [7].
- **Fine-Tuned Encoder–Decoder Transformer Model:**The same pre-trained architecture further trained on paraphrase-specific data to improve semantic alignment, structural consistency, and output stability.
- **Large Generative Language Model:**Included as a comparative baseline to assess the trade-off between generative fluency and semantic control in paraphrase generation.

The primary focus is placed on the BART architecture due to its encoder–decoder design and its demonstrated effectiveness in text rewriting and generation tasks [8]. Figure 2 presents the BART encoder–decoder architecture employed in this study.

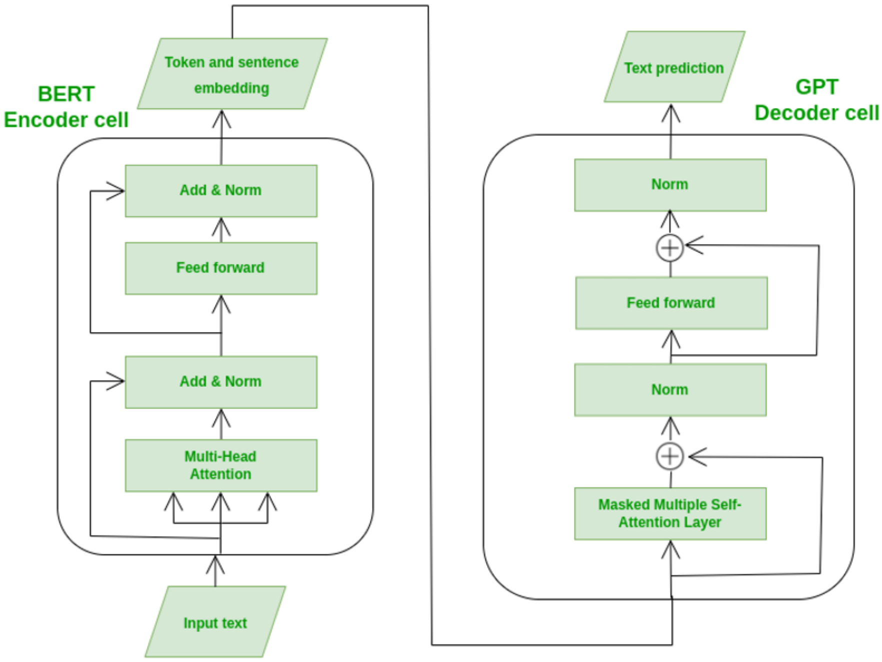

Figure 2. *BART encoder–decoder architecture employed for semantic paraphrase generation. [10]*

3.4. Fine-Tuning Strategy

Fine-tuning adapts a pre-trained transformer model to the paraphrase generation task by optimizing it on paired paraphrase data. During fine-tuning, the encoder processes the source sentence while the decoder learns to generate the corresponding paraphrased output. The training objective minimizes the discrepancy between the generated paraphrase and the reference paraphrase, allowing the model to internalize task-specific rewriting patterns. A low learning rate and a controlled number of training epochs are employed to prevent overfitting and preserve the general language representations learned during pre-training. This strategy enables the model to balance semantic preservation with lexical and syntactic variation, which is essential for producing high-quality paraphrases suitable for academic writing contexts. Fine-tuning is particularly important for encoder–decoder transformer models, as it aligns the model's generative behaviour with paraphrase-specific objectives rather than general text reconstruction or generation tasks.

3.5. Inference Process

During inference, the trained models generate paraphrases for previously unseen input sentences. Decoding strategies are configured to promote fluent and semantically consistent outputs while avoiding excessive repetition or unnecessary deviation from the original meaning. The generated paraphrases are then evaluated using both quantitative and qualitative criteria to assess semantic fidelity, grammatical correctness, and structural variation. This inference setup allows for a fair and consistent comparison between pre-trained, fine-tuned, and large generative models under identical generation conditions.

4. EXPERIMENTAL SETUP AND EVALUATION METRICS

This section describes the experimental configuration and evaluation strategy adopted to assess the performance of pre-trained and fine-tuned transformer models for semantic paraphrase generation. The experiments are designed to analyse both model learning behaviour during training and the qualitative characteristics of generated paraphrases, with particular emphasis on semantic preservation, grammatical correctness, and output consistency.

4.1. Experimental Setup

All experiments are conducted using transformer-based encoder–decoder architectures configured specifically for paraphrase generation tasks. The primary model evaluated in this study is the BART architecture, examined in both pre-trained and fine-tuned settings due to its suitability for text rewriting and sequence-to-sequence generation tasks [8]. In addition, a large generative language model is included as a comparative baseline to analyse differences in fluency, diversity, and controllability of generated paraphrases. For fine-tuning, the pre-trained transformer model is trained on paired paraphrase data using a supervised learning framework. The encoder receives the original sentence as input, while the decoder generates the corresponding paraphrased output. Training is performed for a fixed number of epochs to ensure sufficient convergence without excessive overfitting. A low learning rate is employed to preserve the linguistic representations acquired during pre-training and to enable stable optimization, a common practice in transfer learning for transformer-based models [7]. Batch size and maximum sequence length are selected based on computational feasibility and model constraints, ensuring a balance between training efficiency and representation capacity. During inference, identical decoding configurations are applied across all evaluated models to ensure fair comparison of generated paraphrases.

4.2. Training and Validation Monitoring

Model performance during training is monitored using both training and validation loss. Tracking loss trends provides insight into model convergence, stability, and generalization behaviour. A consistent reduction in training loss accompanied by stable validation loss indicates effective learning, whereas divergence between the two may signal overfitting. For fine-tuned models, validation loss serves as a key indicator of successful task adaptation. Comparing training and validation loss trajectories between pre-trained and fine-tuned configurations allows assessment of the impact of task-specific fine-tuning on learning dynamics. This analysis provides empirical evidence of how fine-tuning improves alignment between generated paraphrases and reference outputs. Monitoring loss behaviour is particularly important for paraphrase generation tasks, where semantic alignment is not always fully captured by surface-level lexical overlap metrics.

4.3. Evaluation Metrics

Given the focus on semantic paraphrase generation, evaluation is conducted using a combination of quantitative and qualitative measures. Rather than relying exclusively on automatic n-gram–based metrics, the evaluation strategy emphasizes learning behaviour and semantic consistency, which are more indicative of paraphrase quality in controlled rewriting tasks.

4.3.1. Quantitative Evaluation

Quantitative evaluation is primarily based on training and validation loss observed during model optimization. These metrics provide an objective measure of how effectively the model learns to map input sentences to semantically equivalent paraphrases. Lower validation loss is interpreted as improved semantic alignment and generalization capability. The use of loss-based evaluation is particularly suitable for encoder–decoder transformer models trained with sequence-level objectives, as it reflects the model's ability to generate outputs that closely match reference paraphrases in meaning and structure.

4.3.2. Qualitative Evaluation

Qualitative evaluation complements quantitative analysis by examining the generated paraphrases produced by different models. Generated outputs are assessed based on:

- Semantic fidelity to the original sentence
- Grammatical correctness and fluency
- Structural variation and lexical diversity

This qualitative assessment enables human-interpretable comparison of paraphrasing behaviour across pre-trained, fine-tuned, and large generative models. It also allows identification of semantic drift, unnecessary content alteration, or excessive variability, which are critical considerations for academic and technical writing applications.

5. RESULTS AND ANALYSIS

This section presents and analyses the experimental results obtained from evaluating pre-trained and fine-tuned transformer models for semantic paraphrase generation. The analysis focuses on model learning behaviour during training and the qualitative characteristics of generated paraphrases, with particular attention to semantic preservation, output consistency, and linguistic fluency.

5.1. Training and Validation Performance

The training and validation loss trends provide insight into how effectively the evaluated models learn the paraphrase generation task. For the pre-trained transformer model used without task-specific adaptation, the reduction in training loss is limited, and validation loss shows comparatively weaker convergence. This behaviour indicates that although the pre-trained model possesses strong general language generation capabilities, it is not optimally aligned with the controlled paraphrasing objective. In contrast, the fine-tuned transformer model demonstrates a more consistent and stable decrease in both training and validation loss across epochs. The convergence pattern suggests that task-specific fine-tuning enables the model to better internalize paraphrase-oriented rewriting patterns, resulting in improved alignment between generated outputs and reference paraphrases. The relatively stable validation loss further indicates enhanced generalization and reduced overfitting, confirming the effectiveness of fine-tuning for this task.

Table 1. *Training and validation loss for the pre-trained BART model.*

Metric	Value	Metric	Value
Mean Similarity Score	**0.5685**	**Standard Deviation of Similarity Scores**	**0.1059**
Mean BLEU Score	**0.1522**	**Standard Deviation**	**0.0380**
Mean ROUGE-1 Score	**0.2965**	**Standard Deviation**	**0.0470**
Mean ROUGE-2 Score	**0.2714**	**Standard Deviation**	**0.0476**
Mean ROUGE-L Score	**0.2965**	**Standard Deviation**	**0.0470**

Table 2. *Training and validation loss for the fine-tuned BART model.*

Metric	Value	Metric	Value
Mean Similarity Score	**0.8558**	**Standard Deviation of Similarity Scores**	**0.1291**
Mean BLEU Score	**0.4098**	**Standard Deviation**	**0.2258**
Mean ROUGE-1 Score	**0.8972**	**Standard Deviation**	**0.0903**
Mean ROUGE-2 Score	**0.7482**	**Standard Deviation**	**0.1999**
Mean ROUGE-L Score	**0.8349**	**Standard Deviation**	**0.1565**

Comparative analysis of loss trajectories between the pre-trained and fine-tuned configurations highlights the impact of task adaptation on learning dynamics. Fine-tuning allows the model to shift from general-purpose text generation toward more controlled semantic rewriting, which is essential for producing reliable paraphrases in structured writing contexts.

Table 1 presents the quantitative performance metrics for the pre-trained BART model, while Table 2 summarizes the corresponding results for the fine-tuned model. The fine-tuned model consistently outperforms the pre-trained model across all metrics, including similarity, BLEU, and ROUGE scores. This improvement demonstrates the effectiveness of task-specific fine-tuning in enhancing semantic preservation and structural consistency in paraphrase generation.

5.2. Qualitative Analysis of Generated Paraphrases

Qualitative evaluation reveals clear differences in paraphrasing behaviour across the evaluated models. The pre-trained transformer model is capable of producing fluent paraphrases with noticeable lexical variation; however, its outputs occasionally exhibit semantic drift. In such cases, the generated paraphrase partially alters the original meaning or introduces additional contextual information that was not present in the source sentence. While these variations may be acceptable in creative rewriting scenarios, they reduce suitability for academic or technical writing tasks where semantic precision is critical.

The fine-tuned transformer model produces paraphrases that more consistently preserve the semantic intent of the original input while maintaining grammatical correctness and appropriate structural variation. Sentence restructuring is more controlled, and the generated outputs remain closer to the source meaning without unnecessary elaboration. This behaviour reflects the benefits of task-specific fine-tuning, which guides the model toward producing paraphrases that balance linguistic diversity with semantic fidelity.

The large generative language model included for comparison generates paraphrases that are generally fluent and stylistically diverse. These outputs often demonstrate strong natural language fluency; however, they exhibit higher variability and less deterministic behaviour. In some instances, this flexibility leads to paraphrases that deviate from the original meaning or introduce stylistic changes that may not be desirable for structured rewriting tasks. While such models are valuable for creative text generation, their reduced semantic control limits their reliability for academic paraphrasing applications.

Table 3 presents representative paraphrases generated by the evaluated model, highlighting how semantic meaning is preserved while introducing controlled lexical and structural variation. The examples demonstrate that the generated paraphrases maintain the original intent while applying syntactic restructuring and synonym substitution, confirming the model's ability to produce semantically consistent and fluent outputs.

Table 3. *Example paraphrases generated by a large generative language model.*

Original Sentence	**DeepSeek-R1 Paraphrase**
"AI is transforming healthcare services."	"Healthcare is being revolutionized by artificial intelligence."
"The economy is facing inflation challenges."	"Rising inflation is impacting the economy."

5.3 Comparative Summary

The comparative analysis of results highlights three key observations:

- **Pre-trained transformer models** provide a strong baseline for paraphrase generation in terms of fluency but lack sufficient control over semantic consistency when used without task-specific adaptation.

- **Fine-tuned transformer models** achieve superior semantic fidelity, improved output stability, and more consistent paraphrasing behaviour, making them better suited for controlled paraphrase generation in academic and technical writing contexts.
- **Large generative language models** offer enhanced linguistic diversity and expressive flexibility but exhibit less predictable semantic behaviour, limiting their applicability for tasks that require precise meaning preservation.

Overall, the results demonstrate that task-specific fine-tuning plays a critical role in improving the quality and reliability of semantic paraphrase generation using transformer architectures. These findings reinforce the importance of model adaptation when deploying paraphrase generation systems in contexts where semantic accuracy and consistency are essential.

6. DISCUSSION

The results presented in this study provide important insights into the effectiveness of transformer-based architectures for semantic paraphrase generation, particularly with respect to the role of task-specific fine-tuning. The comparative analysis between pre-trained and fine-tuned models highlights that while modern transformer models possess strong general language generation capabilities, their performance in controlled paraphrasing tasks depends heavily on alignment with task-specific objectives. One of the key observations is that pre-trained transformer models, when used without additional adaptation, tend to prioritize fluency and surface-level variation over precise semantic preservation. Although such models generate grammatically correct and diverse paraphrases, the absence of paraphrase-specific fine-tuning can lead to semantic drift, especially in structured or information-dense sentences. This behaviour reflects the general-purpose nature of pre-trained models, which are optimized for broad language understanding rather than controlled rewriting.

In contrast, fine-tuned transformer models demonstrate improved semantic consistency and output stability. Fine-tuning enables the model to internalize rewriting patterns that emphasize meaning preservation while still allowing appropriate lexical and syntactic variation. This finding underscores the importance of transfer learning strategies that adapt pre-trained representations to downstream paraphrasing tasks. The improved convergence behaviour and stable validation performance observed during training further support the conclusion that fine-tuning enhances generalization for paraphrase generation. The comparison with a large generative language model reveals an important trade-off between generative flexibility and semantic control. Large generative models produce fluent and expressive paraphrases, often exhibiting greater stylistic diversity than encoder–decoder transformers. However, this flexibility comes at the cost of reduced determinism, which can result in paraphrases that deviate from the original meaning. For applications such as academic and technical writing, where semantic accuracy is critical, this unpredictability limits the practical usefulness of purely generative approaches. Another significant implication of the findings is the limitation of relying solely on surface-level lexical variation as an indicator of paraphrasing quality. The qualitative analysis demonstrates that paraphrases with substantial structural or lexical differences may still fail to preserve semantic intent. This reinforces the need for evaluation strategies that prioritize semantic fidelity over simple n-gram overlap, particularly for controlled rewriting tasks.

Overall, the discussion highlights that fine-tuned encoder–decoder transformer models offer a balanced solution for semantic paraphrase generation by combining fluency with reliable meaning preservation. These characteristics make such models particularly well suited for academic and research-oriented writing contexts, where controlled paraphrasing is essential for maintaining clarity, accuracy, and ethical standards.

7. CONCLUSION AND FUTURE WORK

This paper presented a comparative study of pre-trained and fine-tuned transformer models for semantic paraphrase generation, with a focus on evaluating their ability to preserve meaning while producing fluent and structurally varied paraphrases. By examining encoder–decoder transformer architectures under different training configurations, the study investigated the impact of task-specific adaptation on paraphrase quality and output consistency. The experimental results demonstrate that pre-trained transformer models provide a strong baseline for paraphrase generation in terms of linguistic fluency but are not optimally aligned with controlled rewriting objectives when used without further adaptation. In contrast, fine-tuned transformer models exhibit improved semantic fidelity, greater output stability, and more consistent paraphrasing behaviour. These improvements highlight the importance of fine-tuning in aligning general-purpose language models with task-specific paraphrasing requirements, particularly in academic and technical writing contexts where semantic accuracy is critical.

The comparison with a large generative language model further emphasizes the trade-off between generative flexibility and semantic control. While large generative models are capable of producing expressive and diverse paraphrases, their reduced determinism can lead to semantic drift, limiting their reliability for structured paraphrasing applications. This observation reinforces the need to carefully select and adapt models based on the intended use case.

Overall, the findings of this study indicate that fine-tuned encoder–decoder transformer models offer a practical and effective solution for semantic paraphrase generation when accuracy, consistency, and control are prioritized. The results provide valuable insights for the design and deployment of paraphrasing systems intended to assist academic authors and researchers in ethical and reliable text rewriting.

Future work may explore extending this approach to multilingual paraphrase generation, enabling broader applicability across different languages and writing contexts. Additional research could investigate the integration of semantic similarity metrics for automatic evaluation of paraphrase quality, as well as domain-specific fine-tuning strategies to further enhance performance in specialized technical and scientific fields. Furthermore, hybrid frameworks that combine the controlled behaviour of fine-tuned models with the expressive capabilities of large generative models represent a promising direction for advancing semantic paraphrase generation.

FUNDING:
This research received no external funding.

INSTITUTIONAL REVIEW BOARD STATEMENT:
Not applicable.

INFORMED CONSENT STATEMENT:
Not applicable.

CONFLICTS OF INTEREST:
The author declares no conflict of interest.

REFERENCES

[1] R. V. Birwadkar, “Plagiarism Detection and Paraphrasing based on Generative Artificial Intelligence,” Master’s thesis, Dept. of Information and Technology, SRH Hochschule Heidelberg, Germany, 2025.

[2] I. Androutsopoulos and P. Malakasiotis, “A Survey of Paraphrasing and Textual Entailment Methods,” Journal of Artificial Intelligence Research, vol. 38, pp. 135–187, 2010.

[3] C. Quirk, C. Brockett, and W. Dolan, “Monolingual Machine Translation for Paraphrase Generation,” in Proc. 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP), Barcelona, Spain, 2004, pp. 142–149.

[4] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to Sequence Learning with Neural Networks,” in Proc. Advances in Neural Information Processing Systems (NeurIPS), Montreal, QC, Canada, 2014, pp. 3104–3112.

[5] D. Bahdanau, K. Cho, and Y. Bengio, “Neural Machine Translation by Jointly Learning to Align and Translate,” in Proc. 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, 2015.

[6] A. Vaswani et al., “Attention Is All You Need,” in Proc. Advances in Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, 2017, pp. 5998–6008.

[7] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in Proc. 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT), Minneapolis, MN, USA, 2019, pp. 4171–4186.

[8] M. Lewis et al., “BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension,” in Proc. 58th Annual Meeting of the Association for Computational Linguistics (ACL), Online, 2020, pp. 7871–7880.

[9] C. Raffel et al., “Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer,” Journal of Machine Learning Research, vol. 21, no. 140, pp. 1–67, 2020.

[10] GeeksforGeeks. (n.d.). Website Summarizer using BART. Retrieved from https://www.geeksforgeeks.org/website-summarizer-using-bart/