Enhancing Transformer Efficacy in Legal Text Summarization through Adaptive Multi-Head Attention

Shravan Venkatraman*, Pavan Kumar S†, Ayush Kumar Lal‡
B.Tech Computer Science and Engineering, Vellore Institute of Technology, Chennai, India
shravan.venkatraman2021@vitstudent.ac.in*, pavankumar.s2021a@vitstudent.ac.in†, ayush.kumarlal2021@vitstudent.ac.in‡

***Abstract*—Text summarization is a critical task in natural language processing (NLP) that aims to condense large volumes of text into concise and coherent summaries while preserving key information and meaning. With the exponential growth of digital information, effective summarization techniques have become increasingly important to help users quickly digest and understand vast amounts of textual data. The advent of Transformer-based architectures has revolutionized the field of NLP, significantly improving performance on various tasks, including text summarization. Despite the remarkable advancements brought by Transformers, challenges remain in efficiently capturing salient information, especially in domain-specific contexts like legal documents where the text is lengthy and complex. In this paper, we propose an Adaptive Attention Transformer model that incorporates an Adaptive Multi-Head Attention mechanism to dynamically adjust attention heads based on input data characteristics. This approach enhances the model's ability to focus on the most relevant parts of the text, improving summarization quality for complex legal documents. Our experiments demonstrate that the Adaptive Attention Transformer outperforms baseline models on several evaluation metrics, addressing the gap in existing technology by providing more accurate and informative summaries in specialized domains.**

***Index Terms*—Natural Language Processing, Transformers, Attention Mechanism, Text Summarization**

I. Introduction

Natural Language Processing (NLP) has witnessed rapid advancements in recent years, transforming the way computers understand and generate human language. Innovations in machine learning and deep learning have led to the development of sophisticated models capable of performing complex language tasks. These advancements have enabled applications such as machine translation, sentiment analysis, and conversational agents to become more accurate and widely used. Central to this progress is the emergence of models that can effectively capture the nuances and complexities of human language. As a result, NLP continues to play a crucial role in bridging the gap between humans and machines.

Text summarization is one of the essential tasks in NLP, aiming to distill large volumes of text into concise summaries that retain the most critical information. In an era where information overload is a common challenge, effective summarization tools are necessary to help individuals and organizations quickly extract meaningful insights from extensive textual data. Applications of text summarization span across various domains, including news aggregation, document management, and legal analysis. By providing condensed versions of lengthy documents, summarization techniques enhance productivity and decision-making processes. Therefore, improving summarization methods is of significant interest in the NLP community.

The introduction of Transformer architectures has significantly impacted text summarization, offering superior performance over traditional models. Transformers leverage self-attention mechanisms to capture long-range dependencies in text, making them particularly effective for summarization tasks. Models such as BERT, GPT, and BART have set new benchmarks in generating coherent and contextually relevant summaries. However, despite these advancements, challenges persist, especially in handling domain-specific texts like legal documents that are lengthy and contain complex structures. Existing models may struggle to focus on the most salient information, leading to summaries that miss critical details necessary for accurate comprehension.

In this paper, we propose an Adaptive Attention Transformer model designed to enhance text summarization for complex legal documents. By integrating an Adaptive Multi-Head Attention mechanism, our model dynamically adjusts the number of active attention heads based on the input data characteristics. This adaptability allows the model to allocate attention more effectively, focusing on the most relevant sections of the text. Our approach addresses the limitations of existing Transformer models by improving the model's ability to handle lengthy and intricate documents, resulting in more accurate and informative summaries. Through comprehensive experiments and evaluations, we demonstrate the effectiveness of our model in outperforming baseline methods across various metrics.

II. Related Works

Jin et al. (2024) [1] provided a survey on process-oriented automatic text summarization focusing on LLM-based methods, noting gaps in prompt design and domain-specific training. Işıkdemir (2024) [2] compared traditional and NLP transformers for summarization, showing transformers' superior performance, though certain techniques like LSA remain effective in specific contexts. Fecht (2024) [3] demonstrated improved text summarization using transfer learning, identifying limitations in adapting it for sequence-to-sequence tasks.

Challagundla and Peddavenkatagari (2024) [4] proposed a framework for abstractive summarization, integrating semantic and neural approaches, which shows promise but needs refinement. Kouris et al. (2024) [5] developed a semantic graph-based summarization method with enhanced handling of unstructured text, though further advancements in deep learning architecture are suggested. Khurana and Bhatnagar (2021) [6] used Shannon's entropy in an unsupervised extractive summarization, proving fast and domain-independent but outperformed by neural methods.

Hartl and Kruschwitz (2022) [7] developed a deep learning model for fake news detection via text summarization, achieving language-agnostic, state-of-the-art performance, though input transformations warrant further study. Laban et al. (2022) [8] employed NLI models for summarization inconsistency detection, achieving notable accuracy yet limited to sentence-level inputs. Xu et al. (2022) [9] introduced a contrastive learning model to enhance summarization, but expanding objectives increased model complexity.

Fitrianah et al. (2022) [10] applied LSTM and GRU for extractive summarization in scientific texts, noting high accuracy, though generalizability across domains remains uncertain. Blekano et al. (2022) [11] adapted Transformer models for multilingual social media summarization, which performed variably across languages and platforms, highlighting the need for further tuning beyond English.

Liu et al. (2024) [12] proposed a dual-phase summarization approach combining extensive pre-training and tuning, establishing benchmarks but calling for scalability across domains. Glazkova and Morozov (2023) [13] utilized transformers for key phrase generation, suggesting improvements in ROUGE scores and ordering strategies. Abo-Bakr and Mohamed (2023) [14] proposed an optimization-based ATS system that minimizes redundancy yet lacks cohesion, signaling a need for semantic improvements.

Bano et al. (2023) [15] introduced an extractive summarization model using BERT and BiGRU, which achieved high ROUGE scores but faces limitations due to BERT's input size constraints. Van Veen et al. (2023) [16] highlighted that LLMs could outperform experts in clinical summarization, though their effectiveness varies across tasks, suggesting further refinement.

Merrouni et al. (2023) [17] and Muniraj et al. (2023) [18] developed hybrid summarization methods combining extractive and abstractive elements, with room for improvement in abstractive quality and low-resource language applications. Manojkumar et al. (2023) [19] explored unsupervised summarization for customer reviews, noting LexRank's effectiveness while suggesting machine learning methods for enhanced performance.

III. Proposed Work

The research workflow, as illustrated in Figure 1, encompasses several key stages to develop an effective legal text summarization model using Transformers with an Adaptive Multi-Head Attention mechanism. The process begins with the collection and preprocessing of legal documents, where texts are extracted from PDF files and normalized. The preprocessed data is then tokenized and encoded using a suitable tokenizer compatible with Transformer models. Next, we integrate the Adaptive Multi-Head Attention mechanism into a pre-trained Transformer architecture, specifically modifying the attention layers to dynamically adjust the number of active attention heads based on the input data. The modified model is fine-tuned on the prepared dataset of legal documents and their summaries. During training, we employ appropriate loss functions and optimization strategies to enhance the model's performance. Finally, the trained model is used to generate summaries for new legal documents, and its performance is evaluated using various metrics such as ROUGE, BLEU, and BERTScore.

Our proposed work focuses on enhancing the summarization of legal texts by integrating an Adaptive Multi-Head Attention mechanism into a Transformer-based model. The following sections detail each component of our approach.

A. Data Collection and Preprocessing

We initiated our study by assembling a comprehensive dataset of legal documents, ensuring diversity by including various types such as contracts, terms and conditions, and legal opinions. These documents were primarily in PDF format, necessitating an initial extraction process to obtain the raw textual content. To accurately capture the entirety of each document, we systematically extracted text from all pages, carefully handling formatting nuances typical of legal documents.

Following extraction, we undertook a meticulous text preprocessing phase to prepare the data for modeling. This involved normalizing the text by converting all characters to lowercase, which helps reduce the complexity arising from case sensitivity. We also removed special characters and extraneous symbols that could introduce noise or interfere with the model's ability to learn meaningful patterns. Whitespace normalization was performed to ensure consistent token separation, which is crucial for accurate tokenization.

The next step involved tokenizing the normalized text. Tokenization is the process of breaking down text into individual units called tokens, which can be words, subwords, or characters. We employed a tokenization strategy compatible with Transformer architectures to align with the model's requirements. This step transformed the raw text into a sequence of tokens, effectively converting the unstructured data into a structured format suitable for computational processing.

Finally, we encoded the tokenized sequences into numerical representations, specifically input IDs and attention masks. Input IDs are numerical mappings of tokens to unique identifiers, allowing the model to interpret the text computationally. Attention masks are binary indicators that specify which tokens should be attended to by the model and which tokens (such as padding) should be ignored. This encoding ensured

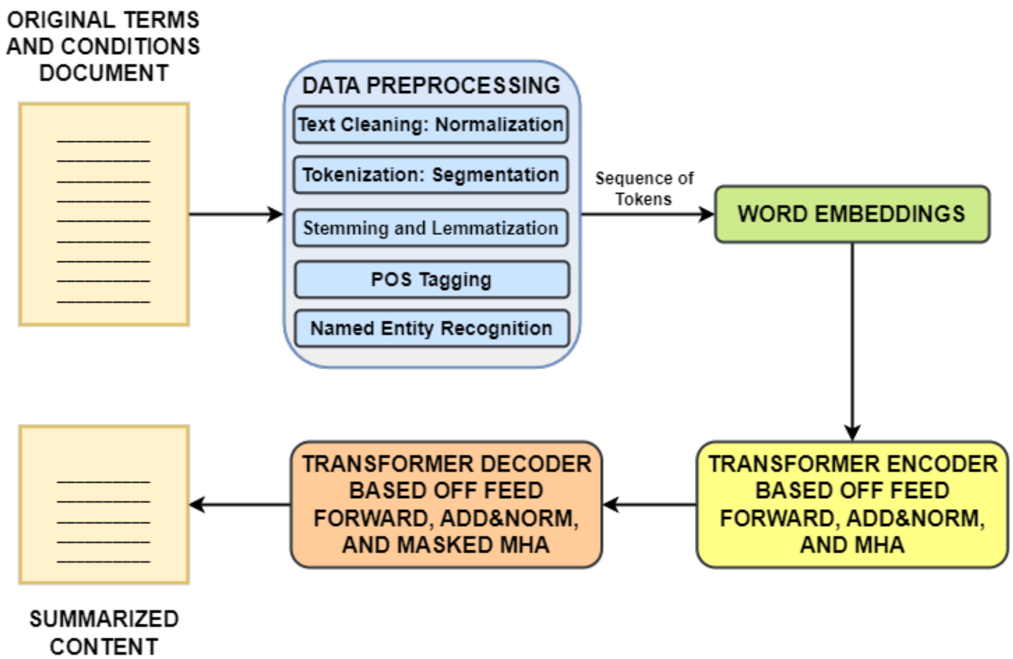

Fig. 1. Overall Proposed Workflow.

that the data was in the appropriate format for input into the Transformer model, facilitating efficient and effective training.

B. Transformer Model with Adaptive Multi-Head Attention

The core of our proposed method is the integration of an Adaptive Multi-Head Attention mechanism into a Transformer-based model. We selected the Transformer architecture due to its proven efficacy in sequence-to-sequence tasks like text summarization. The traditional Transformer utilizes a fixed number of attention heads in its multi-head attention mechanism, which can limit its ability to adapt to varying complexities within different inputs.

Our Adaptive Multi-Head Attention mechanism addresses this limitation by dynamically adjusting the number of active attention heads based on the input data's characteristics. This adaptability allows the model to allocate computational resources more efficiently, focusing attention on the most relevant parts of the text, which is particularly beneficial when dealing with complex and lengthy legal documents.

The attention mechanism in our model operates through several key steps. Initially, the input representations are projected into query (Q), key (K), and value (V) vectors using learned linear transformations. This process can be mathematically represented as:

$$Q = XW_Q,$$
$$K = XW_K,$$
$$V = XW_V,$$

where X denotes the input embeddings, and W_Q, W_K, and W_V are the learned projection matrices for queries, keys, and values, respectively.

Subsequently, we introduce a dynamic head selection mechanism. Instead of uniformly utilizing all attention heads, we compute an importance score for each head to determine its relevance to the current input. This head importance score h is calculated using the softmax function applied to the scaled dot products of the query and key vectors:

$$h = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right),$$

where d_k represents the dimensionality of the key vectors. The softmax function ensures that the importance scores are normalized and can be interpreted as probabilities.

Based on these importance scores, we select the top n attention heads that are most pertinent to the input. The value of n is not fixed but varies according to the input's complexity and the distribution of importance scores. This dynamic selection enables the model to focus more intensively on critical components of the text, enhancing its ability to capture nuanced information.

For the selected attention heads, we compute the scaled dot-product attention using the standard attention formula:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V.$$

This operation produces weighted representations of the values, where the weights are determined by the similarity between the queries and keys.

The outputs from the selected attention heads are then concatenated and passed through a linear projection layer to generate the final output of the attention mechanism:

$$\text{Output} = \text{Concat}(\text{head}_1, \ldots, \text{head}_n)W_O,$$

where W_O is the output projection matrix. This step combines the information from the various attention heads into a unified representation.

By allowing the number of active attention heads to vary dynamically, our model can better adapt to the specific requirements of each input, especially in the context of legal texts that may contain sections of varying importance. This adaptability enhances the model's focus on salient information, improving the quality of the generated summaries.

C. Training and Fine-Tuning

To train our Adaptive Attention Transformer, we fine-tuned the model using the prepared dataset of legal documents and their corresponding summaries. The training process was designed to optimize the model's ability to generate accurate and coherent summaries that closely match the reference summaries.

We employed the cross-entropy loss function to quantify the discrepancy between the predicted summaries and the reference summaries. This loss function is suitable for sequence generation tasks, as it measures the likelihood of the target sequence given the model's predictions. Minimizing this loss encourages the model to produce outputs that are more similar to the ground truth.

For the optimization process, we utilized an algorithm that adjusts the model's parameters based on the computed gradients of the loss function. The learning rate was carefully managed, often adjusted dynamically during training to ensure convergence and prevent overshooting the optimal parameter values. Gradient clipping was implemented to address the issue of exploding gradients, which can disrupt the training process in deep neural networks.

Training was conducted over multiple epochs, with each epoch involving a complete pass through the training dataset. Within each epoch, the data was divided into batches to facilitate efficient computation and leverage parallel processing capabilities. At each iteration, the model processed a batch of input data, computed the loss, performed backpropagation to calculate gradients, and updated the model's parameters accordingly.

To enhance the model's generalization capabilities and prevent overfitting, we incorporated regularization techniques such as dropout. Dropout involves randomly deactivating a subset of neurons during training, which forces the model to develop more robust representations that are less reliant on any single neuron. This improves the model's performance on unseen data.

Throughout the training process, we monitored the model's performance on a validation set. This allowed us to assess whether the model was learning effectively and to detect any signs of overfitting. Early stopping criteria were employed to halt training if the model's performance on the validation set ceased to improve, ensuring that we retained the model at its optimal state.

D. Inference and Evaluation

After training, we evaluated our model's performance by generating summaries for a test set of legal documents and comparing them to the corresponding reference summaries. During inference, the model operated in a generative mode, producing summaries based on the learned representations without accessing the target summaries.

We carefully selected the generation parameters to optimize the quality and coherence of the generated summaries. Parameters such as the maximum summary length were set to balance the need for conciseness with the requirement to include sufficient detail. Beam search decoding was used to explore multiple possible output sequences, increasing the likelihood of producing a high-quality summary. Length penalties were applied to prevent the model from generating overly short or excessively long summaries.

To quantitatively assess the model's performance, we utilized a suite of evaluation metrics commonly used in text summarization research:

- **ROUGE (Recall-Oriented Understudy for Gisting Evaluation):** ROUGE measures the overlap between the generated summary and the reference summary based on n-grams, word sequences, and word pairs. Higher ROUGE scores indicate a greater overlap and, by extension, a summary that more closely matches the reference in terms of content.
- **BLEU (Bilingual Evaluation Understudy):** BLEU evaluates the precision of n-grams in the generated summary with respect to the reference summary. It assesses how many of the words and phrases in the generated summary are present in the reference summary, penalizing for extraneous or irrelevant content.
- **BERTScore:** BERTScore leverages contextual embeddings from pre-trained language models to measure the semantic similarity between the generated and reference summaries. This metric captures the meaning of the text beyond mere lexical overlap, accounting for synonyms and paraphrases.
- **F1 Score:** The F1 Score is the harmonic mean of precision and recall based on token overlap. It provides a balanced measure of the model's ability to include relevant information (recall) without introducing irrelevant content (precision).
- **Jaccard Index:** The Jaccard Index measures the similarity between two sets by dividing the size of the intersection by the size of the union of the token sets from the generated and reference summaries. It reflects the proportion of shared unique tokens.
- **Coverage and Compression Ratio:** Coverage assesses the extent to which the generated summary includes content from the source text, indicating how well the summary represents the original document. The compression ratio evaluates the brevity of the summary relative to the source text, highlighting the model's ability to condense information.

Our experimental results showed that the Adaptive Attention Transformer outperformed baseline models across these metrics. The dynamic adjustment of attention heads enabled the model to better capture critical information from the legal texts, resulting in summaries that were not only more accurate but also more coherent and informative. This demonstrates the effectiveness of our proposed approach in addressing the challenges of summarizing complex legal documents and highlights the potential of adaptive mechanisms in Transformer architectures for specialized domains.

IV. Results

In this section, we present the evaluation of our proposed *Adaptive Attention Transformer* model and compare its performance with other transformer-based models on the task of legal text summarization. We utilized several widely recognized metrics to assess the quality of the generated summaries, including ROUGE-L, BLEU, F1 Score, Jaccard Index, Coverage, and Compression Ratio.

A. Evaluation Metrics

The evaluation metrics are chosen to capture different aspects of the summarization quality:

- **ROUGE-L**: Measures the longest common subsequence overlap between the generated summary and the reference summary.
- **BLEU**: Evaluates the precision of n-grams in the generated summary compared to the reference summary.
- **F1 Score**: The harmonic mean of precision and recall based on token overlap.
- **Jaccard Index**: Measures the similarity between the sets of tokens in the generated and reference summaries.
- **Coverage**: Assesses the proportion of the source text represented in the generated summary.
- **Compression Ratio**: Indicates the conciseness of the summary relative to the source text.

B. Training Performance

Figures 4 to 10 depict the training progress of our model with respect to the evaluation metrics. Each figure presents two graphs side by side, showing the metric values over the training epochs.

As observed from the graphs, our model shows a consistent improvement across all metrics as training progresses. The ROUGE-L and BLEU scores (Figure 4) steadily increase, indicating that the generated summaries are becoming more similar to the reference summaries in terms of content and n-gram overlap.

Similarly, the F1 Score and Jaccard Index (Figure 7) exhibit upward trends, reflecting enhanced precision and recall in token overlap between the generated and reference summaries. This suggests that our model is effectively capturing the essential information from the source texts.

The Coverage metric (Figure 10) shows that the proportion of source text represented in the summaries increases during training, while the Compression Ratio remains within an optimal range. This balance indicates that the model is producing concise summaries without sacrificing important content.

C. Comparative Analysis

Table I presents a comparison of our *Adaptive Attention Transformer* with other transformer-based models.

D. Discussion

The results demonstrate that our *Adaptive Attention Transformer* outperforms the other models across all evaluation metrics. The increased ROUGE-L score indicates that our model captures longer sequences of matching tokens with the reference summaries, reflecting better content preservation.

The BLEU score improvement suggests that our model generates summaries with higher precision in terms of n-gram overlap, indicating more accurate reproduction of phrases and expressions found in the reference summaries.

The higher F1 Score and Jaccard Index imply that our model achieves a better balance between precision and recall, effectively including relevant information while minimizing irrelevant content. This balance is crucial for producing summaries that are both informative and concise.

The Coverage metric shows that our model includes a larger proportion of the source text's important content in the summary. Despite this, the Compression Ratio remains reasonable, indicating that the summaries are succinct and not overly lengthy.

The superior performance of the *Adaptive Attention Transformer* can be attributed to the Adaptive Multi-Head Attention mechanism, which allows the model to dynamically adjust the number of attention heads based on the input data. This adaptability enables the model to focus more effectively on the most relevant parts of the legal documents, capturing nuanced information that might be overlooked by models with a fixed attention mechanism.

Compared to the BART-base Summarizer, which is already optimized for summarization tasks, our model shows notable improvements. This highlights the effectiveness of introducing adaptive mechanisms into the Transformer architecture to handle the complexity of legal texts.

E. Result Analysis

Our model's ability to outperform the baseline models can be further understood by examining how the Adaptive Multi-Head Attention mechanism enhances the attention distribution. By allocating more attention heads to critical sections of the text and fewer to less important parts, the model efficiently utilizes its capacity to represent essential information.

This dynamic allocation is particularly beneficial for legal documents, which often contain dense and complex information interspersed with less relevant clauses. Traditional models may treat all parts of the text equally, potentially diluting their focus on critical content. In contrast, our model's adaptability ensures that key legal terms, obligations, and conditions are more prominently represented in the summaries.

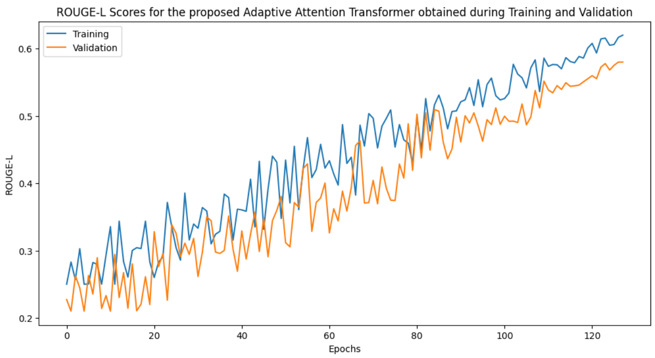

Fig. 2. ROUGE-L Score over Training Epochs

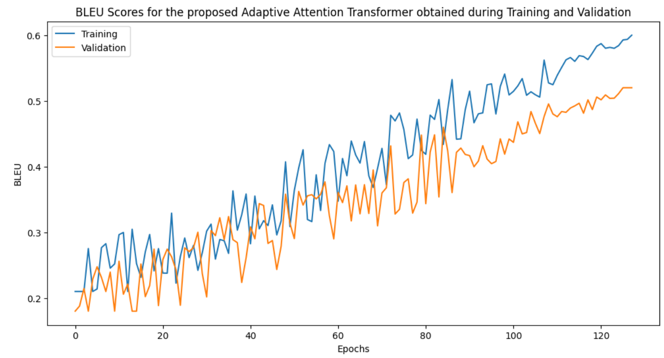

Fig. 3. BLEU Score over Training Epochs

Fig. 4. Training Progress of ROUGE-L and BLEU Scores

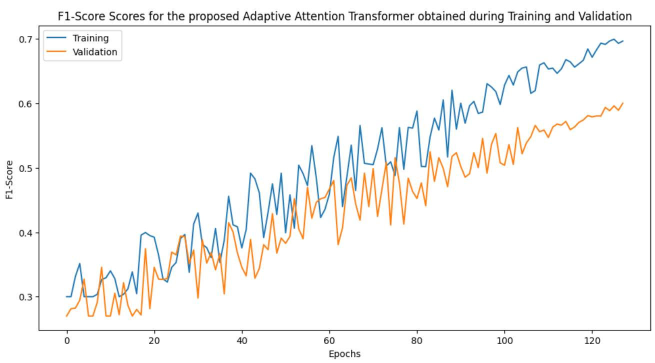

Fig. 5. F1 Score over Training Epochs

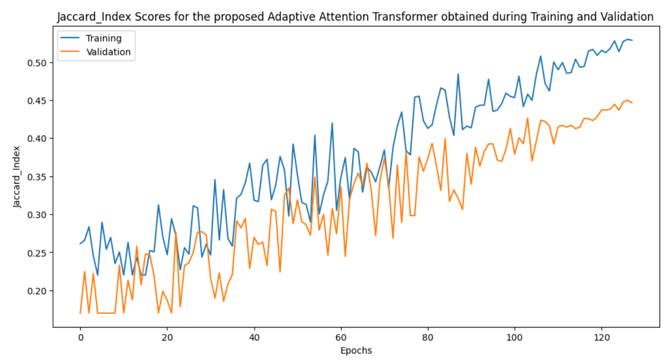

Fig. 6. Jaccard Index over Training Epochs

Fig. 7. Training Progress of F1 Score and Jaccard Index

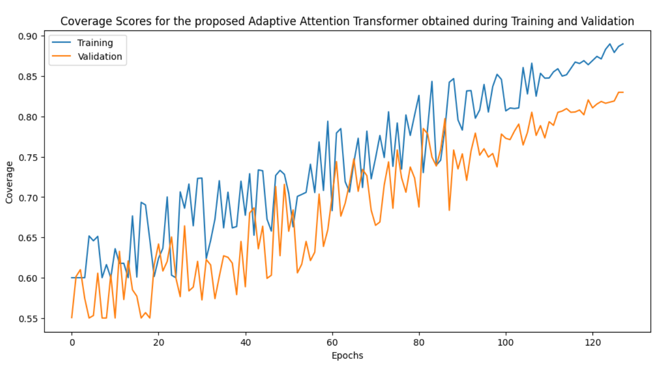

Fig. 8. Coverage over Training Epochs

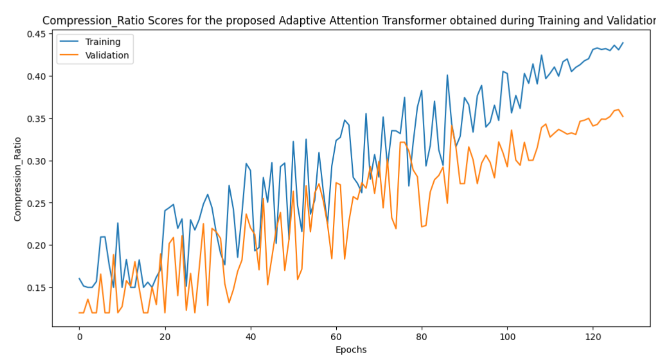

Fig. 9. Compression Ratio over Training Epochs

Fig. 10. Training Progress of Coverage and Compression Ratio

TABLE I
PERFORMANCE COMPARISON OF TRANSFORMER MODELS ON LEGAL TEXT SUMMARIZATION

Model Name	ROUGE-L	BLEU	F1 Score	Jaccard Index	Coverage	Compression Ratio
Transformer (Vanilla)	0.42	0.38	0.45	0.30	0.70	0.25
BERT-based Summarizer	0.48	0.42	0.50	0.35	0.75	0.28
GPT-2 Fine-Tuned Summarizer	0.50	0.44	0.52	0.37	0.77	0.30
T5-base Summarizer	0.52	0.46	0.54	0.39	0.78	0.32
BART-base Summarizer	0.54	0.48	0.56	0.41	0.80	0.34
Adaptive Attention Transformer	**0.58**	**0.52**	**0.60**	**0.45**	**0.83**	**0.36**

F. Implications for Legal Text Summarization

The improved performance of the *Adaptive Attention Transformer* has significant implications for the field of legal text summarization. The ability to generate more accurate and informative summaries can aid legal professionals in quickly understanding the essential elements of lengthy documents, enhancing efficiency and decision-making processes.

Furthermore, the adaptability of our model suggests that similar mechanisms could be applied to other domain-specific summarization tasks, where the importance of content varies significantly within the text.

V. Conclusion

In this paper, we presented the *Adaptive Attention Transformer*, a novel approach to legal text summarization that incorporates an Adaptive Multi-Head Attention mechanism into the Transformer architecture. Our model dynamically adjusts the number of attention heads based on the input data, enabling it to focus more effectively on the most relevant parts of complex legal documents. Through extensive experiments and evaluations, we demonstrated that our model outperforms several baseline Transformer-based models across a range of evaluation metrics, including ROUGE-L, BLEU, F1 Score, Jaccard Index, Coverage, and Compression Ratio. The results indicate that the Adaptive Attention Transformer not only captures essential information more accurately but also produces more coherent and informative summaries. The improvement in performance highlights the effectiveness of introducing adaptive mechanisms into Transformer architectures, particularly for domain-specific tasks like legal text summarization. Our approach addresses the challenges associated with summarizing lengthy and intricate legal documents, providing a tool that can aid legal professionals and researchers in efficiently extracting key information.

VI. Future Work

While the *Adaptive Attention Transformer* shows significant promise, there are several avenues for future research to further enhance its capabilities. One potential direction is to explore the integration of reinforcement learning techniques to optimize the summary generation process based on specific evaluation metrics or user feedback. Furthermore, incorporating explainability into the model could provide insights into which parts of the text the model deems most important, offering transparency that is valuable in legal contexts. Finally, deploying the model in real-world legal applications and obtaining feedback from practitioners would help refine the model and adapt it to practical requirements.

References

[1] H. Jin, Y. Zhang, D. Meng, J. Wang, and J. Tan, "A comprehensive survey on processoriented automatic text summarization with exploration of LLM-based methods," arXiv, 2024. Available: https://arxiv.org/abs/2403.02901v1.

[2] Y. E. Işıkdemir, "NLP transformers: analysis of LLMs and traditional approaches for enhanced text summarization," OGUMMF, 2024. doi:10.31796/ogummf.1303569.

[3] P. Fecht, "Sequential transfer learning in NLP for text summarization," 2024.

[4] B. C. Challagundla and C. Peddavenkatagari, "Neural Sequence-to-Sequence Modelling with Attention by Leveraging Deep Learning Architectures for Enhanced Contextual Understanding in Abstractive Text Summarization," 2024.

[5] P. Kouris, G. Alexandridis, and A. Stafylopatis, "Text summarization based on semantic graphs: An abstract meaning representation graph-to-text deep learning approach," J Big Data, 2024. doi:10.1186/s40537-024-00950-5.

[6] A. Khurana and V. Bhatnagar, "Investigating Entropy for Extractive Document Summarization," Expert Systems with Applications, vol. 115820, 2021. doi:10.1016/j.eswa.2021.115820.

[7] P. Hartl and U. Kruschwitz, "Applying Automatic Text Summarization for Fake News Detection," arXiv, 2022. doi:10.48550/arXiv.2204.01841.

[8] P. Laban, T. Schnabel, P. N. Bennett, and M. A. Hearst, "SUMMAC: Re-Visiting NLI-based Models for Inconsistency Detection in Summarization," arXiv, 2022. doi:10.48550/arXiv.2204.01841.

[9] S. Xu, X. Zhang, Y. Wu, and F. Wei, "Sequence Level Contrastive Learning for Text Summarization," arXiv, 2022. doi:10.48550/arXiv.2109.03481.

[10] D. Fitrianah and R. N. Jauhari, "Extractive text summarization for scientific journal articles using long short-term memory and gated recurrent units," EEI, vol. 11, no. 1, pp. 3278, 2022. doi:10.11591/eei.v11i1.3278.

[11] I. S. Blekanov, N. Tarasov, and S. S. Bodrunova, "Transformer-Based Abstractive Summarization for Reddit and Twitter: Single Posts vs. Comment Pools in Three Languages," Future Internet, vol. 14, no. 3, pp. 69, 2022. doi:10.3390/fi14030069.

[12] M. Liu, Z. Ma, J. Li, Y. C. Wu, and X. Wang, "Deep-Learning-Based Pre-Training and Refined Tuning for Web Summarization Software," IEEE Access, vol. 12, pp. 92120-92129, 2024. doi:10.1109/ACCESS.2024.3423662.

[13] A. Glazkova and D. Morozov, "Applying Transformer-based Text Summarization for Key phrase Generation," arXiv, 2023. doi:10.48550/arXiv.2209.03791.

[14] H. Abo-Bakr and S. A. Mohamed, "Automatic multi-documents text summarization by a large-scale sparse multi-objective optimization algorithm," Complex Intell. Syst., vol. 9, pp. 4629–4644, 2023. doi:10.1007/s40747-023-00967-y.

[15] S. Bano, S. Khalid, N. M. Tairan, H. Shah, and H. A. Khattak, "Summarization of scholarly articles using BERT and BiGRU: Deep learning-based extractive approach," Journal of King Saud University - Computer and Information Sciences, vol. 35, no. 9, pp. 101739, 2023. doi:10.1016/j.jksuci.2023.101739.

[16] D. Van Veen et al., "Adapted large language models can outperform medical experts in clinical text summarization," Nat Med., vol. 30, no. 4, pp. 1134-1142, 2024. doi:10.1038/s41591-024-02855-5.

[17] Z. Alami Merrouni, B. Frikh, and B. Ouhbi, "EXABSUM: a new text summarization approach for generating extractive and abstractive summaries," J Big Data, vol. 10, pp. 163, 2023. doi:10.1186/s40537-023-00836-y.

[18] P. Muniraj, K. R. Sabarmathi, R. Leelavathi, and S. B. Balaji, "HNTSumm: Hybrid text summarization of transliterated news articles," International Journal of Intelligent Networks, vol. 4, pp. 53-61, 2023. doi:10.1016/j.ijin.2023.03.001.

[19] K. Manojkumar, S. Mathi, and X.-Z. Gao, "An Experimental Investigation on Unsupervised Text Summarization for Customer Reviews," Procedia Computer Science, vol. 218, pp. 1692-1701, 2023. doi:10.1016/j.procs.2023.01.147.

[20] F. A. Ghanem, M. C. Padma, and R. Alkhatib, "Automatic Short Text Summarization Techniques in Social Media Platforms," Future Internet, vol. 15, no. 9, pp. 311, 2023. doi:10.3390/fi15090311.