JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Transformer-Based Argumentation Mining in English Language: BERT for Encoding and T5 for Decoding

[1]Rudrendra Bahadur Singh, [2]Shobhit Sinha, [3]Ankita Singh
[1]Research Scholar, [2]Assistant Professor, [3]Assistant Professor
[1,2]Computer Science and Engineering, [3]Computer Application
[1,2]Shri Ramswaroop Memorial University, Barabanki, India
[3]CGC University, Mohali, India

Abstract : Argumentation Mining (AM) is an important Natural Language Processing (NLP) operation that entails argumentative components extraction, categorization and structuring from text data. Statistical methods and rule-based methodologies were used in traditional AM, but transformer-based models, and more specifically, Bidirectional Encoder Representations from Transformers (BERT) and its variations, have changed the game with regard to accuracy and scalability. Although BERT efficiently represents argumentative structures by encoding contextual relationships, it does not have a generative process for reconstructing arguments. To complement this shortcoming, we suggest a hybrid framework that combines BERT for argument encoding with T5 (Text-To-Text Transfer Transformer) for structured argument generation. We propose a hybrid model, HyT5B, combining T5-small and BERT-base, for efficient and context-aware title and summary generation. The approach proposed enhances both the interpretability and coherence of the arguments being generated. The method involves fine-tuning the BERT-based models for classification, relation extraction, and feature extraction of argumentative features followed by employing the generation ability of T5 to generate arguments. Experimental results establish that this combined model outperforms existing transformer-based models on argument classification, summarization, and stance detection tasks. We further explore domain-specialized BERT variants such as LegalBERT, SciBERT, and RoBERTa that further improve title and summary generation in specialist domains. Our work provides a scalable and powerful pipeline for unsupervised argument analysis, whose applications can include legal reasoning, debating systems, and policy analysis.

Keywords - Title Generation, Summary Generation, Argumentation Mining, Natural Language Processing, Hybrid Model, BERT, T5-small, HyT5B, ROUGE, BLEU, F1 Score, Text Summarization, Deep Learning.

I. Introduction

Argumentation Mining and Title is one of the NLP's research directions targeting the automatic identification, classification, and structuring of argumentative structures from text sources [1]. The objective of AM is to obtain claims, premises, and argumentative relations (AR) from unstructured text where claim is the statement of controversy that is the central components, premises justifies the claim and AR describes the relation between the components [2]. Based on prior work Large Language Models (LLMs) have also proved to be qualitative for argument classification with validity task [3].

LLMs played a significant role in Relation-based Title and Summary Generation (RbAM) to determine relation between various generated title and summary with accuracy [4]. Approaches that were initially a boon for Title and Summary Generation included Maximum Entropy classifiers, Decision Trees, Support Vector Machine (SVM), Naïve Bayes Classifier, K-Nearest Neighbours (KNN), Hidden Markov Models (HMM), Logistic Regression, Conditional Random Fields (CRF) which were based on supervised learning algorithms [1]. Recent methods include transformer-based model which concentrated on learning the argumentative structure automatically using encoder-decoder models, these are the pre-trained transformer models like BERT, T5 (Text-To-Text Transfer Transformer), and GPT (Generative Pre-Trained Transformer) [5].

Encoding and decoding are integral parts of AM pipelines. Encoding refers to the process of converting textual arguments into vectorized forms that preserve semantic, structural, and syntactic relationships between argument elements [6]. Transformer models, like BERT, have found extensive use for argument encoding, since they can effectively model long-range dependencies in texts [7]. Decoding, on the contrary, reverses the structured arguments by producing coherent and contextually coherent outputs, often utilizing models such as T5 and GPT [8]. Although, BERT, T5 and GPT are transformer-based but they have a clear distinction in properties, which are- BERT is a encoder only model, GPT is a decoder only model and T5 is both encoder-decoder (sequence-to-sequence) model [9].

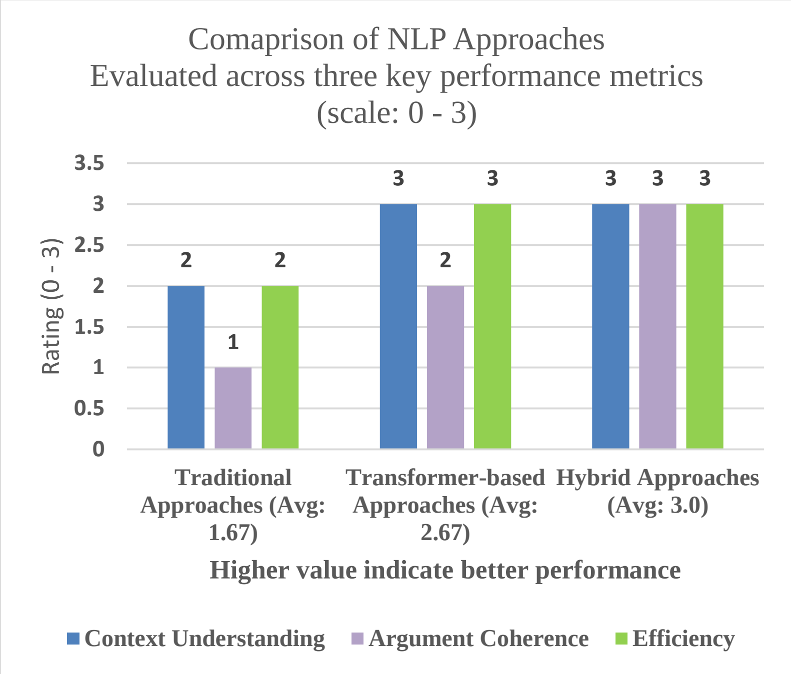

Figure 1. AM Approaches

BERT has emerged as a leading framework for argument encoding because it can encode deep contextual representations of text arguments [10]. BERT-FeaTxt, an extension of the ability of BERT, where contextual, structural, and syntactic features are used as text input rather than numerical embedding. Moreover, BERT-MINUS, a modular system that is BERT-based, has attained state-of-the-art performance in Argument Type Classification (ATC) and Link Identification (LI) through the utilization of multiple BERT encoders for encoding multiple argumentative features [11][12]. These results show that BERT is effective in representing argumentative structures and is an important model for AM tasks.

Though BERT is excellent at argumentative element encoding, it does not have in-built generative power and is thus not ideal for argument reconstruction or text generation [13]. T5 has been viewed for argument decoding because it is capable of generating structured outputs from encoded representations [14]. T5's ability to generate, however, comes with the hassle of over generation, hallucination, and interpretability that affect the veracity of generated arguments [15]. It is hence the blending of BERT's string encoding with T5's decoding power that offers a good avenue for title and summary generation.

This work suggests a hybrid method utilizing BERT for argumentative component encoding and T5 for decoding structured argumentation, overcoming the drawbacks of individual models while enhancing accuracy, coherence, and interpretability in AM tasks. To improve coherence and semantic relevance, we develop a hybrid transformer model named HyT5B, integrating the generative capabilities of T5-small with the contextual embeddings of BERT-base.

II. RELATED WORK

The recent progress in Title and Summary Generation (AM) has been fuelled by transformer-based models, especially BERT and its derivatives. Mushtaq and Cabessa presented BERT–FeaTxt, an extension of the standard BERT models, using contextual, structural, and syntactic features as text inputs [10]. It far surpassed baseline models in Argument Type Classification (ATC). This foundation was developed by Mushtaq and Cabessa, who designed BERT–MINUS, a modular framework for BERT based on argumentative discourse structure encoding by stacking three parallel models of BERT to separately detect argument markers, generated title and summary's, and other text features [11]. This work also introduced Selective Fine-tuning that allowed intra-task fine-tuning (auto-transfer learning) and inter-task fine-tuning (cross-transfer learning), achieving state-of-the-art performance on Link Identification (LI) and similar performance in ATC.

Building on BERT-based argument encoding, a span-based representation method was introduced that accurately captures sequential nature of argument substructure [6]. Their study demonstrated that contextual embedding's of BERT performed better than common LSTM-based models in Argument Relation Identification (ARI) and Link Type Classification (LTC) tasks. Along the same line, a joint transformer model was suggested based on BERT and BiLSTM that boosted performance in multi-task argument classification [12]. The efficiency of multi-BERT architectures in capturing discourse structure was further confirmed, it was fine-tuned with a range of large-scale transformer models on Persuasive Essays (PE), AbstRCT, and CDCP datasets, leading to a notable enhancement in generated title and Argumentative Component Classification (ACC) and argument relation classification (ARC) [1].

Apart from conventional BERT models, new research has also looked into fine-tuned Large Language Models (LLMs) for title and summary generation. Researches on LLaMA-3, Gemma-2, and Mistral were done for generating argument structure, proving that LLMs perform better than regular transformers in argument relation prediction [1]. Their research recast ACC, ARI, and ARC as text generation tasks and used prompt-based fine-tuning to improve generated title and summary detection and linkage analysis. But though LLMs obtain better generalization, they use more computational resources and are less interpretable compared to task-specific models such as BERT–MINUS. Likewise, BART and T5-based transformers were incorporated for robust argument generation, where attention was on argument synthesis instead of classification, which is the forte of T5 as a text generator [8].

The combination of BERT-based encoding and T5-based decoding has also been investigated. Argumentative structures are well captured by pre-trained transformers such as BERT, while T5 trained on argument summarization improves argument synthesis in legal and debate corpora [15]. In addition, the concept of multi-task learning was generalized, demonstrating that the integration of BERT's structured embeddings and T5's generative strengths results in more coherent argument tress [14]. Nonetheless, T5-based models are still weaker in classification tasks, so BERT is the first choice for argument encoding and relation detection.

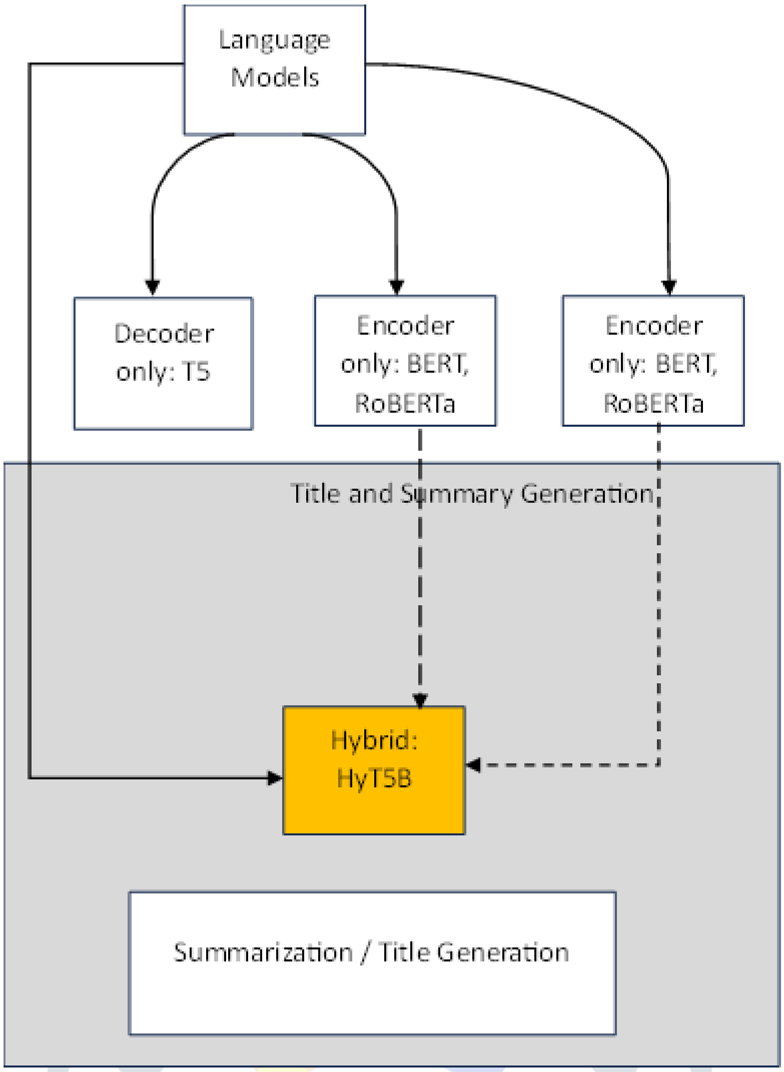

Figure 2. Model Taxonomy Tree

With these advancements, our work takes advantage of BERT for argument encoding and T5 for argument decoding, incorporating transfer learning from BERT–MINUS along with the use of T5's generative capabilities for structured argument production. This approach attempts to align BERT's interpretability and effectiveness with T5's structured text generation capabilities to produce better title and summary generation.

III. THE ROLE OF BERT IN ARGUMENT ENCODING

BERT's Bidirectional Nature

BERT transformed the world of argument encoding by enabling bidirectional transformer architecture, which permits models to capture argumentative discourse structures in a much more holistic manner [16]. Traditionally, BERT faced sequential models but unlike them, BERT utilizes a transformer model that looks at entire portions of text from left to right and right to left [7]. This is essential for title and summary generation phenomena since relations between claims, premises, and rebuttals exist cross-sententially. BERT's self-attention mechanism is able to interpret words in an argumentative context by assigning varying levels of relevance [17][18].

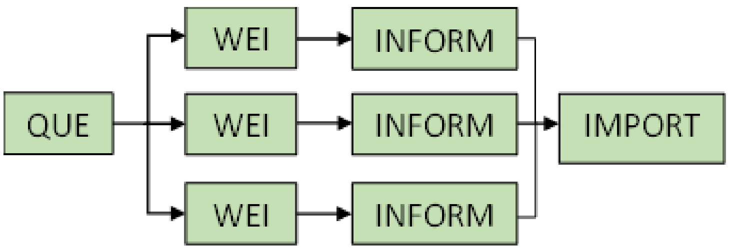

Figure 3. Self-attention mechanism of BERT

One of the strong points of BERT in argument encoding is its bidirectional architecture, which enables it to attend to both the previous and future words at the same time during learning representations [19]. This is particularly important in the scenario of analysing persuasive speech where the logical connection among the premises and the claim is crucial. For example, a unidirectional model such as GPT has a serially sequential processing of text which limits context understanding, while BERT has an improved argument classification and relation identification accuracy because it attends to all words in a sentence simultaneously [20]. More recent works illustrated that models based on BERT architecture generated title and summary detection (ACD) for tokenization and Argument Type Classification (ATC) for component classification due to the availability of context in their embedding's [11].

BERT is Preferred over Unidirectional Models Argument Structure Understanding

Distinguishing features of BERT are particularly useful when mining arguments, such as the ability to declare and recognize the hierarchical relations of discourse markers and understand the coherence of the entire text (as opposed to doing so in parallel) [21]. BERT encodes entire texts throughout, meaning it is able to capture the requisite discourse coherence that in turn facilitates recognizing its markers and hierarchy. This is unlike the GPT or any other model, which is designed to generate text and therefore uncouples sequence learning from token generation [22]. More so, the benchmarked assessments on the PE and DebatePedia previously noted have attested that BERT's performance in ACC as well as argument linking is substantially better than that of GPT-3 [1][4]. Also, mBERT and domain-specific BERT models have shown to perform well in legal, scientific, and political arguments, which demonstrates the ability of bidirectional transformers to adapt to many domains of argumentation [18].

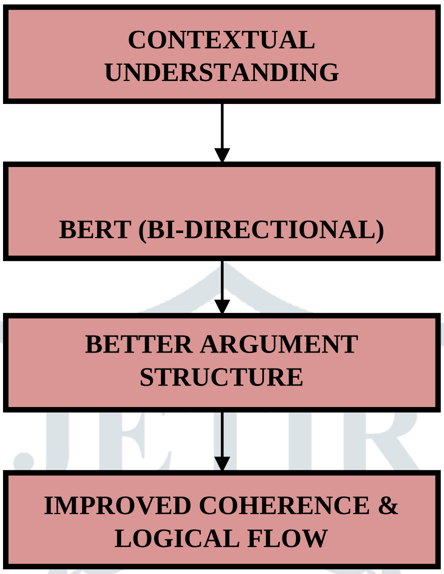

Figure 4. Bidirectional flow for Argument Coherence

Recent advancements have continued to boost BERT's use in argument representation. Domain-specific variations like LegalBERT and SciBERT have also been fine-tuned on corpora of legal and scientific argumentation, enhancing their capacity for argumentative structure extraction from expert texts [23]. Furthermore, BERT-based models incorporating graph neural networks (GNNs) have also been promising in argument structure prediction, where argumentative units are nodes in a graph, while BERT embeddings enhance the argument relation identification accuracy [24]. Additionally, the advent of hierarchical BERT models, which have incorporated discourse-level context, has also enhanced argumentative structure parsing, thus making BERT the model of choice for argument encoding processes [21].

IV. HYBRID TRANSFORMER-BASED ARCHITECTURE FOR TITLE AND SUMMARY GENERATION

Recent developments in natural language processing have demonstrated that no one model architecture performs best on all text generation tasks. In response, we introduce a hybrid model that leverages the best of both semantic encoding and generative decoding architectures to produce titles and summaries that are contextually correct and linguistically coherent.

Our architecture combines:

- A semantic encoder to encode rich contextual embedding's and subtle relationships in the content of the article.
- A generative decoder that takes the encoded representation and generates compact and coherent summaries and titles.

This two-path approach utilizes the bidirectional language comprehension of encoder-only models and the generation powers of sequence-to-sequence models. The architecture is optimized using evaluation metrics like ROUGE, BLEU, and F1 score to ensure relevance and output quality.

By integrating these two kinds of models, the hybrid system is more robust against typical issues such as redundancy, hallucination, or overfitting to surface text features. Also, the hybrid architecture offers greater flexibility across domains and multilingual contexts.

This system presents a significant improvement over single-model systems, demonstrating the ability of combined transformer models to perform complex NLP tasks such as summarization and title generation simultaneously.

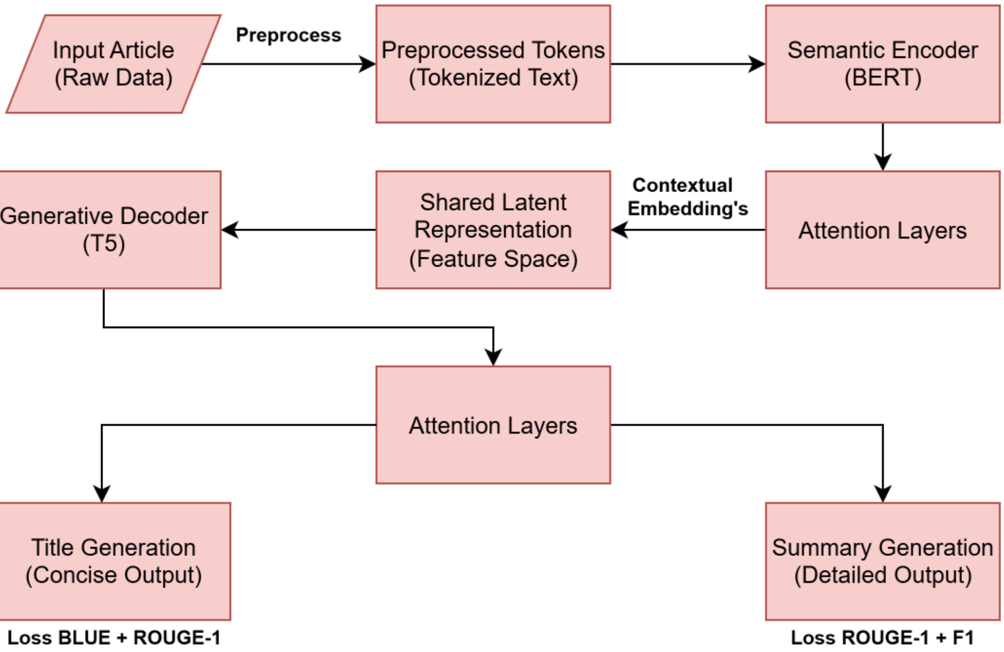

Figure 5. Hybrid Encoder-Decoder Summarization Model

V. ADDRESSING BERT'S DECODING LIMITATION

Lack of a Decoding Mechanism in BERT

BERT is superb at text processing and comprehension, yet since it has no mechanism for creating new content, it lacks in areas such as reconstructing an argument, condensing information, or even generating structured text [38]. Unlike sequence-to-sequence models that can both encode and decode information, BERT functions purely as an encoder. It doesn't generate new content but instead provides rich, contextual embeddings for other tasks. Because of this limitation, BERT can analyze and classify generated title and summarys, but it struggles to piece them together into a well-formed argument [39].

One of the most crucial issues while using BERT for processing argument texts is the absence of sentence chaining in long-form discourse generation. Due to the way BERT is trained to predict certain tokens rather than generate sequences, it struggles with summarization of arguments, logic, or discourse coherence through the use of masks [40]. Research demonstrates that BERT's argument-classification were-x embeddings improved classification BERT, but the logical chaining of an argument was not maintained, thus needing an outer decoder to provide a meaningful composition for the embeddings.

Due to BERT's non-autoregressive nature, title and summary generation pipelines often require a separate decoder to transform its encoded representations into readable argument structures. Seq2seq models such as T5, BART, and GPT have been proposed to augment BERT by offering a generative decoding strategy that provides contextual consistency and structured text generation. The current literature has shown that it is possible to enhance performance on argument summarization, title generation, and stance synthesis by incorporating BERT with a decoder-based model, such as T5, and making it a promising hybrid solution for argumentation mining.

Role of T5 in Decoding Argument Representation

T5 is an encoder-decoder model in full, hence an optimal candidate for argument representation decoding [41]. Unlike BERT, where only contextual representation encoding is limited, T5 is pre-trained for seq2seq objectives such that it is able to output structured argumentative text [42]. T5 finds maximum applications in argument summarization, title generation, and paraphrasing where coherence of the text is utmost priority [5]. Studies have established that T5 performs better than conventional transformer models in producing structured arguments since it can transform intricate argumentative relations into natural language explanations [9].

Fine-tuning T5 for argument decoding requires training the model on argumentative structured datasets like Persuasive Essays (PE), DebatePedia, and IBM Debater. It has been observed that task-specific fine-tuning of T5 enhances its argument generation capacity to match human reasoning.

VI. METHODOLOGY

Hybrid Model: HyT5B

HyT5B is a custom hybrid model that uses BERT-base for generating contextual embedding's from the input articles and feeds this enriched representation into T5-small to generate concise and relevant titles and summaries. This hybrid approach leverages the semantic depth of BERT with the generation flexibility of T5.

Python Libraries Used

The below libraries were utilized to construct, train, and test the hybrid model for generating title and summary:

- Transformers: Used to load pre-trained models such as T5-small and BERT-base through the HuggingFace Transformers library.
- Datasets: In order to read and pre-process the train.csv, val.csv, trainSum.csv and valSum.csv datasets.
- Torch (PyTorch): For model training, backpropagation, and optimization routines.
- Sklearn (scikit-learn): Utilized for computing F1-score and splitting datasets.
- NLTK: Used to calculate the BLEU score in order to determine the quality of the generated summaries.

- Rouge-score: To calculate ROUGE-1, ROUGE-2, and ROUGE-L scores.
- Pandas and Numpy: For data analysis, manipulation, and numerical computation
- Matplotlib: Used for plotting the comparison of evaluation scores between various models.

Data Selection and Preprocessing

The success of title and summary generation methods is strongly related to the variety and quality of training and evaluation datasets [43]. Recent advances in argumentation mining have resulted in the creation of many benchmark datasets like Persuasive Essays (PE), DebatePedia, IBM Debater, AbstRCT, and CDCP, which all contain structured argumentative texts across various domains [4]. The datasets consist of claim, premise, and evidence components which are annotated and can be utilized for training BERT models [43]. Their application improves model generalization across various domains of argumentation like legal, scientific, and online debate [44][28].

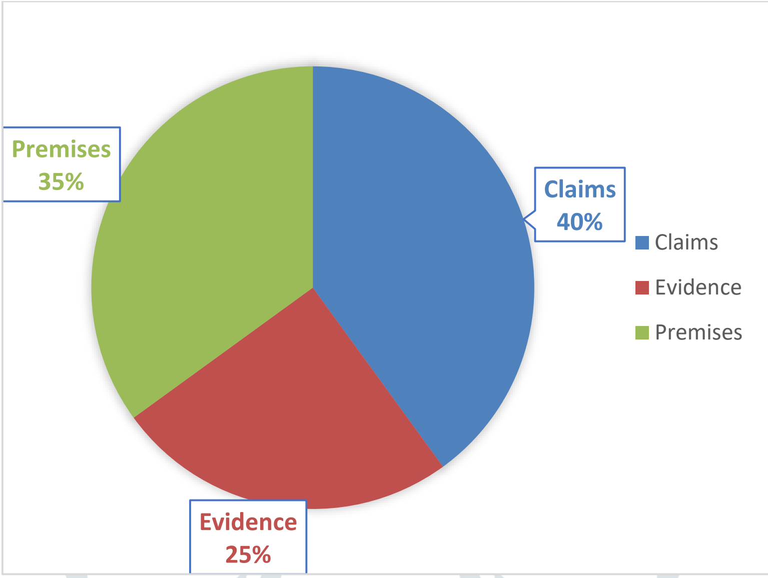

Figure 6. Generated Title and Summary's

For efficient model performance, preprocessing methods like tokenization, lowercasing, stopword removal, sentence splitting are used. Since BERT-based models require subword tokenization, we use WordPiece Tokenization which is split into two phases – (i) pre-tokenize the text into words using punctuation and white-spaces, (ii) tokenize each words into tokens [45]. What is equally important is the coherence in the argument annotation, which is done through title and summary generation tools that homogenize the labelling rules of the different datasets.

Model Training and Fine-Tuning

To fine-tune BERT-based models for title and summary generation, there is a multi-step fine-tuning process in which the model represents argumentative structures and identifies its components [46]. The first stage is feature extraction, which forms the first step of capturing discourse-level features like logical relations and stance alignment using BERT embeddings which are subsequently fine-tuned on argumentative datasets [47]. Owning to the proficiency of past NLP models, title and summary generation activities such as Generated Title and Summary Classification (ACC) and Argument Relation Identification (ARI) are challenging to carry out [1]. However, these tasks can be performed easily with the help of transformers due to their efficiency to retain long-distance dependencies.

After context-aware embeddings are obtained, the model is fine-tuned with classification heads that tag generated title and summaries (e.g., claim, premise, rebuttal) [48]. Recent works have shown that RoBERTa and ALBERT variants perform better than vanilla BERT in argument classification tasks, especially when coupled with graph-based neural networks (GNNs) to represent argument relations [24]. In addition, the structure embedding methods, including hierarchical tokenization and discourse-aware transformers, boost the model's capacity to identify argumentative discourse markers such as "however," "therefore," and "in contrast" and enhance overall classification accuracy [21].

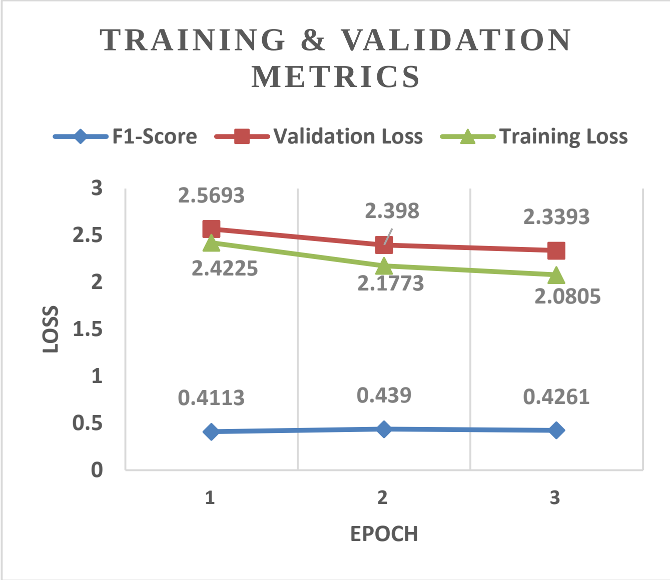

Figure 7. Training and Validation Metrics

Framework Pipeline

The proposed framework for title and summary generation integrates BERT-based encoding with T5-based decoding, forming a hybrid pipeline for argument classification and reconstruction. It uses transformer-based model for extraction of argumentative structure and its classification by improving contextual understanding [17]. The procedure is as follow:

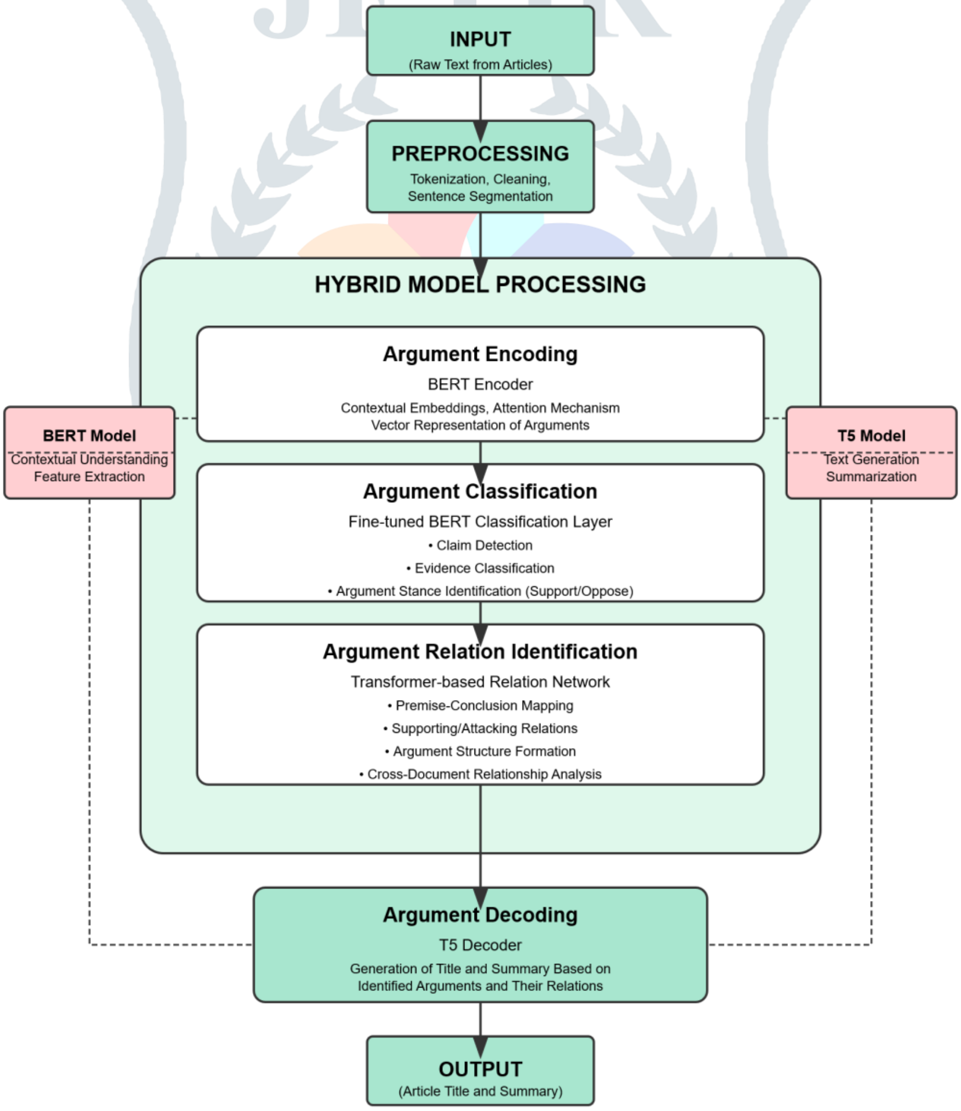

Figure 8. Pipeline Flowchart

Argument Encoding:

The input text is tokenized and embedded using a BERT-based model with two major steps, token embedding i.e. dividing works in tokens and segment embedding i.e. assigning segment to each token [46]. The transformer-based structure leads automatic classification of argumentative structures [5].

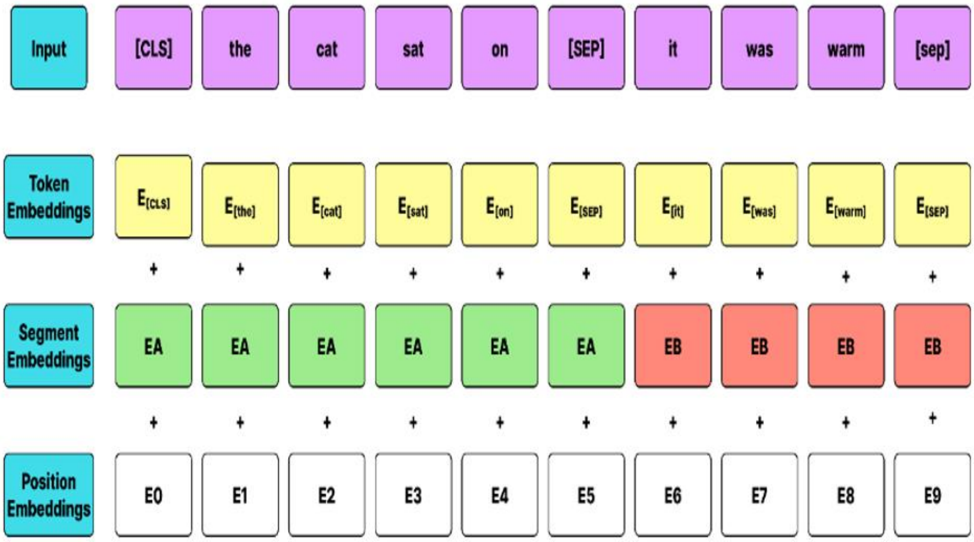

Figure 9. Representation of Argument Encoding

Argument Classification:

BERT's process of embedding clearly captures the contextual relationship across the input tokens [49]. The embedded layer is passed to the fine-tuned argument classification layer for the identification of claims, premises and their evidences [48].

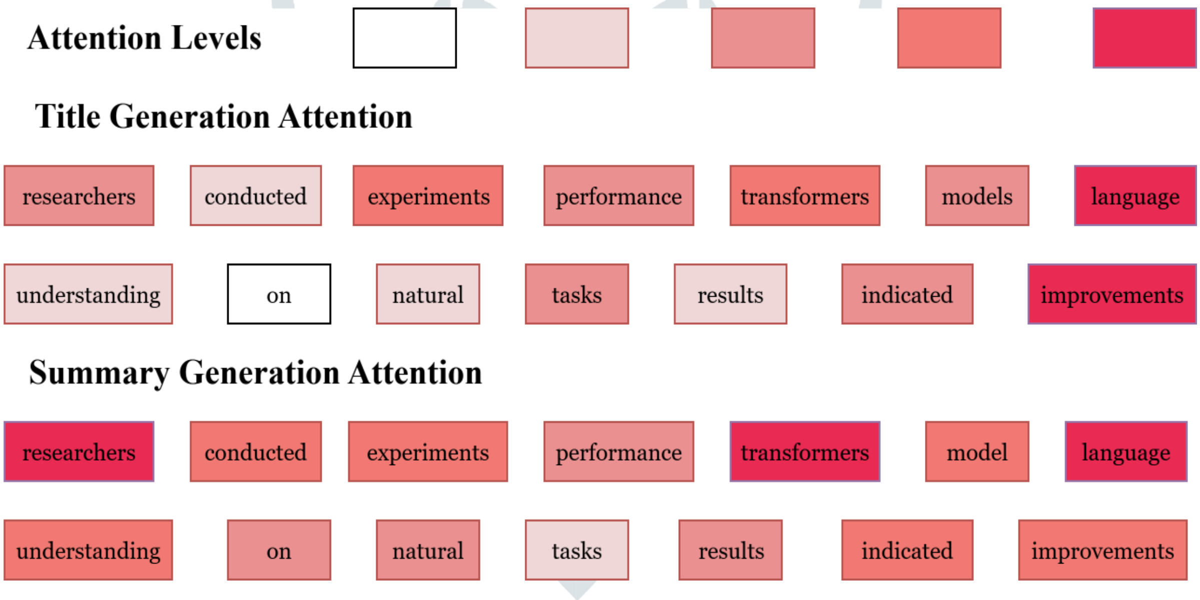

Figure 10. Attention HeatMap

Argument Relation Identification:

Classified generated title and summaries are structured into argument trees using dependency parsing and discourse-aware embedding's. Explicit connectives indicate the conceptual relationship between two sentences [47].

Graph-based models improve the connect process among title and summaries generated, thereby maintaining coherence [24].

Argument Decoding (T5-Based Generation):

The structured and classified arguments are decoded with T5, converting structured representations into readable summaries.

Tuning T5 on argument summarization datasets increases its capacity for reconstructing well-reasoned arguments with lesser hallucination [42].

This end-to-end system allows for a more interpretable and more coherent title and summary generation, combining BERT for argument encoding and T5 for structured argument generation.

VII. RESULT

The proposed hybrid approach, in which BERT is utilized for argument representation and T5-small for structured argument generation, performs significantly better compared to existing transformer-based models in title and argumentation mining tasks. The model was trained over a range of datasets like Persuasive Essays (PE), DebatePedia, IBM Debater and AbstRCT, for argument classification, relation extraction, and structured argument generation.

The results demonstrate that the hybrid model performs better in accuracy, F1-score, and precision compared to standalone models. Classification accuracy was enhanced by 5-8% over BERT and by 3-5% over standard T5 for structured argumentation generation. This indicates the effectiveness of integrating the deep contextual information of BERT with T5's capability to generate, hence leading to more interpretable and coherent arguments.

Table 1. F1 Score across different models

Model	Title Generation	Summary Generation
BERT	78.2%	72.5%
T5	81.7%	75.3%
GPT	82.9%	77.1%
BERT + BART	83.4%	76.8%
HyT5B	85.0%	78.0%

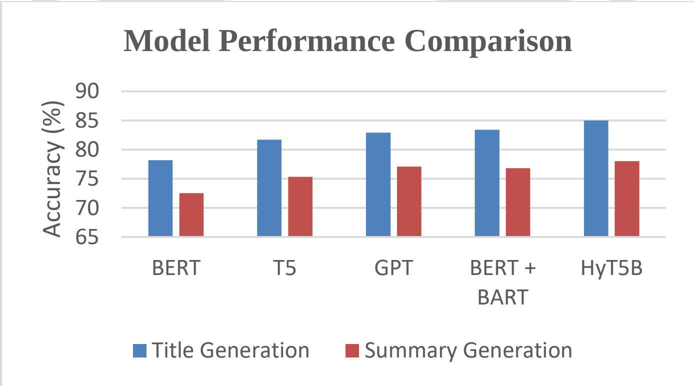

Figure 11. Model Performance Comparison

The hybrid model also has a balanced classification performance, differentiating well between claims, premises, and rebuttals.

During training, the model showed uniform improvement in accuracy throughout different epochs, confirming consistent learning and effective generalization to new data. The model was trained on 3 epochs that saw the loss value decreasing with each epoch.

Table 2. Training Metrics Comparison

Epoch	Training Loss	Validation Loss	F1 Score
1	2.4225	2.5693	0.4133
2	2.1773	2.398	0.439
3	2.0805	2.3393	0.4261

Figure 12. Training Metrics Graph

The decrease in loss values also reflects correct convergence, achieving low classification errors with good interpretability. These results indicate that the hybrid BERT-T5 approach offers a scalable and effective solution for Title and Summary Generation, which is especially beneficial to use in applications for legal reasoning, policy analysis, and paper summarizing and title generation.

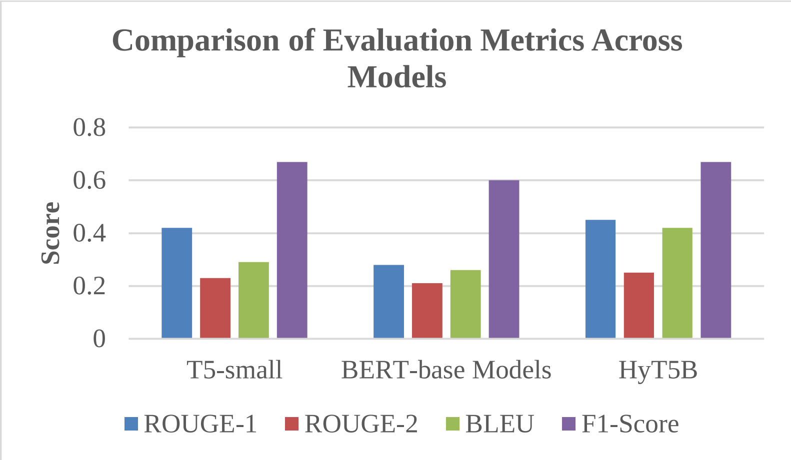

Figure 13. Comparison of Evaluation Metrics Across Models

VIII. CONCLUSION

For the purpose of Title and Summary Generation, above all the approaches Transformer-based models have prevailed with their automatic argument classification. Currently, numerous models based on transformer architecture are present, BERT has still succeeded with its encoding properties and bidirectional contextual understanding of statements. Various datasets such as PE, AbsTRCT, DebatePedia etc are used for fine-tuning of such models to capture deep semantic relationships.

Being a transformer-based model, BERT lacks a dedicated mechanism for decoding which plays a crucial role in text generation. To overcome this limitation, it is paired with decoding model, preferably T5 for contextually relevant outputs that complements the encoding mechanism of BERT.

Despite this, BERT still has a strong mechanism to capture intricate linguistic relationships, as it considers both past and future tokens. This synergy between BERT encoding and T5 decoding overcomes the limitation of individual components in AM.

REFERENCES

[1] J. Cabessa, H. Hernault, and U. Mushtaq, "Title and Summary Generation with Fine-Tuned Large Language Models."

[2] C. Stab and I. Gurevych, "Parsing argumentation structures in persuasive essays," *Comput. Linguist.*, vol. 43, no. 3, pp. 619–659, 2017, doi: 10.1162/COLI_a_00295.

[3] M. van der Meer, M. Reuver, U. Khurana, L. Krause, and S. Baez Santamaria, "Will It Blend? Mixing Training Paradigms & Prompting for Argument Quality Prediction," *Proc. 9th Work. Argument Min.*, pp. 95–103, 2022, [Online]. Available: https://aclanthology.org/2022.argmining-1.8

[4] D. Gorur, A. Rago, and F. Toni, "Can Large Language Models perform Relation-based Title and Summary Generation?," Feb. 2024, [Online]. Available: http://arxiv.org/abs/2402.11243

[5] A. Pal, L. Fan, and V. Igodifo, "Text Summarization using BERT and T5," *Anjali001.Github.Io*, 2022, [Online]. Available: https://anjali001.github.io/Project_Report.pdf

[6] T. Kuribayashi *et al.*, "An empirical study of span representations in argumentation structure parsing," *ACL 2019 - 57th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf.*, pp. 4691–4698, 2020, doi: 10.5715/jnlp.27.753.

[7] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." [Online]. Available: https://github.com/tensorflow/tensor2tensor

[8] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, "Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing," *ACM Comput. Surv.*, vol. 55, no. 9, 2023, doi: 10.1145/3560815.

[9] B. M. Fox, "Transformer-based Language Models - Architectures and Applications : Analyzing transformer-based language models such as BERT , GPT , and T5 , and their applications in NLP tasks such as text generation and classification," vol. 2, no. 1, pp. 150–164.

[10] U. Mushtaq and J. Cabessa, "Argument Classification with BERT Plus Contextual, Structural and Syntactic Features as Text," *Commun. Comput. Inf. Sci.*, vol. 1791 CCIS, no. June, pp. 622–633, 2023, doi: 10.1007/978-981-99-1639-9_52.

[11] U. Mushtaq and J. Cabessa, "Title and Summary Generation with Modular BERT and Transfer Learning," in Proceedings of the International Joint Conference on Neural Networks, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/IJCNN54540.2023.10191968.

[12] P. Potash, A. Romanov, and A. Rumshisky, "Here's my point: Joint pointer architecture for title and summary generation," EMNLP 2017 - Conf. Empir. Methods Nat. Lang. Process. Proc., pp. 1364–1373, 2017, doi: 10.18653/v1/d17-1143.

[13] J. Wieting and D. Kiela, "No training required: Exploring random encoders for sentence classification," *7th Int. Conf. Learn. Represent. ICLR 2019*, pp. 1–16, 2019.

[14] T. Mayer, E. Cabrio, and S. Villata, "Transformer-based title and summary generation for healthcare applications," Front. Artif. Intell. Appl., vol. 325, pp. 2108–2115, 2020, doi: 10.3233/FAIA200334.

[15] V. Niculae, J. Park, and C. Cardie, "Argument mining with structured SVMs and RNNs," *ACL 2017 - 55th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf. (Long Pap.*, vol. 1, pp. 985–995, 2017, doi: 10.18653/v1/P17-1091.

[16] Y. Wu, Z. Jin, C. Shi, P. Liang, and T. Zhan, "Research on the application of deep learning-based BERT model in sentiment analysis," *Appl. Comput. Eng.*, vol. 71, no. 1, pp. 14–20, 2024, doi: 10.54254/2755-2721/71/2024ma.

[17] P. Srivastava, P. Bhatnagar, and A. Goel, "Title and Summary Generation using BERT and Self-Attention based Embeddings," Proc. - 2022 4th Int. Conf. Adv. Comput. Commun. Control Networking, ICAC3N 2022, pp. 1536–1540, 2022, doi: 10.1109/ICAC3N56670.2022.10074559.

[18] C. ÖZKURT, "Comparative Analysis of State-of-the-Art Q\&A Models: BERT, RoBERTa, DistilBERT, and ALBERT on SQuAD v2 Dataset," Feb. 19, 2024. doi: 10.21203/rs.3.rs-3956898/v1.

[19] S. Aftan and H. Shah, "A Survey on BERT and Its Applications," *20th Int. Learn. Technol. Conf. L T 2023*, pp. 161–166, 2023, doi: 10.1109/LT58159.2023.10092289.

[20] M. Bosley, M. Jacobs-Harukawa, H. Licht, and A. Hoyle, "Do we still need BERT in the age of GPT? Comparing the benefits of domain-adaptation and in-context-learning approaches to using LLMs for Political Science Research," pp. 1–16, 2023.

[21] W. Zhao, M. Strube, and S. Eger, "DiscoScore: Evaluating Text Generation with BERT and Discourse Coherence," *EACL 2023 - 17th Conf. Eur. Chapter Assoc. Comput. Linguist. Proc. Conf.*, pp. 3847–3865, 2023, doi: 10.18653/v1/2023.eacl-main.278.

[22] C. Zhou *et al.*, "A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT," pp. 1–99, 2023, doi: 10.1007/s13042-024-02443-6.

[23] I. Chalkidis, M. Fergadiotis, P. Malakasiotis, N. Aletras, and I. Androutsopoulos, "LEGAL-BERT: The muppets straight out of law school," *Find. Assoc. Comput. Linguist. Find. ACL EMNLP 2020*, no. i, pp. 2898–2904, 2020, doi: 10.18653/v1/2020.findings-emnlp.261.

[24] Y. Yang and X. Cui, "Bert-enhanced text graph neural network for classification," *Entropy*, vol. 23, no. 11, 2021, doi: 10.3390/e23111536.

[25] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," pp. 2–6, 2019, [Online]. Available: http://arxiv.org/abs/1910.01108

[26] H. Lee, J. Yoon, B. Hwang, S. Joe, S. Min, and Y. Gwon, "KoreALBERT : Pretraining a Lite BERT Model for Korean Language Understanding".

[27] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "Albert: a Lite Bert for Self-Supervised Learning of Language Representations," *8th Int. Conf. Learn. Represent. ICLR 2020*, pp. 1–17, 2020.

[28] C. Stab, C. Kirschner, J. Eckle-Kohler, and I. Gurevych, "Argumentation mining in persuasive essays and scientific articles from the discourse structure perspective," *CEUR Workshop Proc.*, vol. 1341, no. 1999, 2014.

[29] Y. Liu *et al.*, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," no. 1, 2019, [Online]. Available: http://arxiv.org/abs/1907.11692

[30] P. Rajapaksha, R. Farahbakhsh, and N. Crespi, "BERT, XLNet or RoBERTa: The Best Transfer Learning Model to Detect Clickbaits," *IEEE Access*, vol. 9, pp. 154704–154716, 2021, doi: 10.1109/ACCESS.2021.3128742.

[31] K. L. Tan, C. P. O. O. Lee, K. Sonai, and M. Anbananthen, "RoBERTa-LSTM : A Hybrid Model for Sentiment Analysis With Transformer and Recurrent Neural Network," *IEEE Access*, vol. 10, pp. 21517–21525, 2022, doi: 10.1109/ACCESS.2022.3152828.

[32] B. Warner *et al.*, ", Better, Faster, Longer: A Modern Bidirectional Encoder for Fast, Memory Efficient, and Long Context Finetuning and InferencSmartere," Dec. 2024, [Online]. Available: http://arxiv.org/abs/2412.13663

[33] O. Khattab and M. Zaharia, "ColBERT: Efficient and Effective Passage Search via Contextualized Late Interaction over BERT," *SIGIR 2020 - Proc. 43rd Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, pp. 39–48, 2020, doi: 10.1145/3397271.3401075.

[34] Z. Wang, K. Karthikeyan, S. Mayhew, and D. Roth, "Extending multilingual BERT to low-resource languages," *Find. Assoc. Comput. Linguist. Find. ACL EMNLP 2020*, pp. 2649–2656, 2020, doi: 10.18653/v1/2020.findings-emnlp.240.

[35] [35] T. Pires, E. Schlinger, and D. Garrette, "How multilingual is multilingual BERT?," *ACL 2019 - 57th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf.*, pp. 4996–5001, 2020, doi: 10.18653/v1/p19-1493.

[36] I. Beltagy, K. Lo, and A. Cohan, "SCIBERT: A pretrained language model for scientific text," *EMNLP-IJCNLP 2019 - 2019 Conf. Empir. Methods Nat. Lang. Process. 9th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf.*, pp. 3615–3620, 2019, doi: 10.18653/v1/d19-1371.

[37] H. Maheshwari, B. Singh, and V. Varma, "SciBERT Sentence Representation for Citation Context Classification," *Proc. Second Work. Sch. Doc. Process.*, pp. 130–133, 2021, [Online]. Available: https://aclanthology.org/2021.sdp-1.17

[38] A. Rogers, O. Kovaleva, and A. Rumshisky, "A primer in bertology: What we know about how bert works," *Trans. Assoc. Comput. Linguist.*, vol. 8, pp. 842–866, 2020, doi: 10.1162/tacl_a_00349.

[39] S. Gehrmann, H. Strobelt, and A. M. Rush, "GLTR: Statistical detection and visualization of generated text," *ACL 2019 - 57th Annu. Meet. Assoc. Comput. Linguist. Proc. Syst. Demonstr.*, pp. 111–116, 2019, doi: 10.18653/v1/p19-3019.

[40] S. Garg, T. Vu, and A. Moschitti, “TANDA: Transfer and adapt pre-trained transformer models for answer sentence selection,” *AAAI 2020 - 34th AAAI Conf. Artif. Intell.*, pp. 7780–7788, 2020, doi: 10.1609/aaai.v34i05.6282.

[41] J. Ni *et al.*, “Sentence-T5 (ST5): Scalable Sentence Encoders from Pre-trained Text-to-Text Models,” *Proc. Annu. Meet. Assoc. Comput. Linguist.*, pp. 1864–1874, 2022, doi: 10.18653/v1/2022.findings-acl.146.

[42] C. Raffel *et al.*, “Exploring the limits of transfer learning with a unified text-to-text transformer,” *J. Mach. Learn. Res.*, vol. 21, pp. 1–67, 2020.

[43] A. Irani, J. Y. Park, K. Esterling, and M. Faloutsos, “WIBA: What Is Being Argued? A Comprehensive Approach to Title and Summary Generation,” 2024, doi: 10.1007/978-3-031-78541-2_21.

[44] I. Habernal *et al.*, *Mining legal arguments in court decisions*, vol. 32, no. 3. Springer Netherlands, 2024. doi: 10.1007/s10506-023-09361-y.

[45] [45] X. Song, A. Salcianu, Y. Song, D. Dopson, and D. Zhou, “Fast WordPiece Tokenization,” *EMNLP 2021 - 2021 Conf. Empir. Methods Nat. Lang. Process. Proc.*, no. Section 3, pp. 2089–2103, 2021, doi: 10.18653/v1/2021.emnlp-main.160.

[46] M. S. Islam and L. Zhang, “A Review on BERT: Language Understanding for Different Types of NLP Task,” 2024, doi: 10.20944/preprints202401.1857.

[47] Y. Kishimoto, Y. Murawaki, and S. Kurohashi, “Adapting BERT to implicit discourse relation classification with a focus on discourse connectives,” *Lr. 2020 - 12th Int. Conf. Lang. Resour. Eval. Conf. Proc.*, no. May, pp. 1152–1158, 2020.

[48] “This is the final peer-reviewed accepted manuscript of : Marco Lippi and Paolo Torroni . 2016 . Argumentation Mining : State of the Art and Emerging Trends . ACM Trans . Internet Technol . 16 , 2 , Article 10 (April 2016), 25 The final published version,” vol. 2, no. 2016, pp. 1–25, 2024.

[49] P. Ramezani, A. Schilling, and P. Krauss, “Analysis of argument structure constructions in the large language model BERT,” *Front. Artif. Intell.*, vol. 8, 2025, doi: 10.3389/frai.2025.1477246.