

Transformer-Based Semantic Analysis of English Literary and Educational Texts

Bakytbek Sadykov^{1*} and Ermek Turgunov²

^{1*}Department of English Language, Osh State University, Osh, Kyrgyzstan.

²Department of English Language, Osh State University, Osh, Kyrgyzstan.

Abstract: Gaining a deep insight into English literary and educational texts is difficult owing to their contextual, symbolic, emotional, and conceptual expressions. Conventional methods for text interpretation have been reliant on manual analysis and linguistic features at the surface level and have proved inadequate in analyzing sets of texts. In this study, a semantic analysis approach based on transformers has been suggested to assist in the interpretation of English literary and educational texts through contextual, thematic, emotional, and conceptual expressions. The suggested approach has been tested with a set of 10,000 English texts, including 5,000 literary texts such as novels, stories, poems, and drama, and 5,000 educational texts such as textbooks and educational materials. The methodology was related to text preprocessing, context analysis using transformers, and interpretation of the semantics of obtained patterns on the basis of English language and literature approaches. The results prove the effectiveness of the suggested methodology in terms of improved understanding of the meaning of the texts through identification of the most significant literary and educational patterns. As for literary analysis, the main semantic patterns were represented by the thematic aspect (32%), emotional aspect (26%), relation between characters and events (22%), and symbolic-contextual meaning (20%). Moreover, there is an additional potential of transformer-based semantic analysis in terms of educational content organization, improving reading comprehension, and conducting language learning activities.

Keywords: transformer-based semantic analysis, english literary texts, educational text analysis, natural language processing, contextual representation, digital humanities, artificial intelligence in education

(Received: 30 December 2025; Revised: 11 February 2026; Accepted: 06 March 2026; Published: 30 March 2026)

Introduction

Finding the meaning is one of the major goals of English language studies, especially for literary analysis and education, where the meaning of a linguistic text is constructed using linguistic structures, context relations, cultural expressions, and conceptual representations. The traditional way of analyzing texts relies mostly on human interpretation, linguistic theories, and statistical tools. But these approaches struggle to analyze large text collections because of ambiguity, complex expressions, and different contexts. The new approach of natural language processing has provided the use of transformer architecture that improves the semantic comprehension of texts by learning the context relationships of words, sentences, and discourse (Junaid et al., 2022; Kamil & Çakır, 2025).

Literary texts of English involve intricate components such as metaphors, symbolisms, emotional storylines, interactions between characters, and themes, which make them difficult to interpret. The semantic analysis through transformer technology offers chances to uncover hidden patterns, themes, and emotional meanings in literary texts that would aid in developing research in digital humanities. It has been proven from prior studies how effective transformer models have been in analyzing figurative language,

sentiment patterns, and semantic relations within text data (Monika et al., 2025; Tzimiris et al., 2025). The same is true for educational texts, where semantic analysis is important in order to determine knowledge structure, etc. (Shiha et al., 2025).

Despite the great success that has been achieved by transformers in NLP tasks, the comprehensive usage of this tool for the analysis of English literary and educational texts is still underexplored. Most existing solutions use transformer models to perform sentiment analysis, determine semantic similarity, or deal with domain-specific text analysis, but not with literary interpretation alongside educational interpretation (Abdalgader et al., 2024; Yang et al., 2020). The key issue discussed in this research is related to the complexity of the analysis of the deeper context and conceptual links within English texts by traditional methods of analysis. Thus, this research proposes a transformer-based approach to the semantic analysis of English literature and education.

Research Objective

1. To create a method to extract contextual semantic representation from English language texts by using transformers.
2. To study thematic, emotional, and conceptual patterns of literary English texts.
3. To explore semantic relations and knowledge structures of educational English texts.
4. To assess the efficiency of transformer-based semantic analysis by means of linguistic and computational evaluations.

Research Questions

RQ1: To what extent is the transformer-based semantic analysis able to capture contextual meanings, thematic patterns, and semantic relations within literary English texts?

RQ2: In what way does the transformer-based semantic analysis help interpret and organize English educational texts?

RQ3: What is the advantage of transformer-based semantic analysis over conventional text analysis techniques in regard to the semantic analysis of English texts?

The rest of the paper is arranged as follows: Section I starts with the introduction of the paper, where the research background, motivations, objectives, and research questions will be elaborated. Section II involves the review of the relevant literature associated with transformer-based semantic analysis and its application in natural language processing. Section III contains the details about the dataset used and the proposed transformer-based semantic analysis model. Section IV deals with the semantic analysis and the influence of the results in the field of education. Section V consists of the research results, limitations, and future research directions. Finally, Section VI concludes the paper.

Literature Review

Transformer-based semantic analysis has turned out to be a promising technique for studying contextual connections, theme identification, and other aspects related to meanings in text data. In recent research, researchers have investigated the use of transformers for extracting semantic information from different fields. A transformer-based method for thematic analysis of university reading lists has been proposed, showing the power of modern language models in discovering semantic themes in academic texts (Shiha et al., 2025). Moreover, transformer-based models have performed better in figurative language classification, which requires contextual meaning interpretation (Junaid et al., 2022).

Transformer-based semantic search research has focused on their ability to provide context-aware representations and semantic matching (Kamil & Çakır, 2025).

The development of transformers has greatly affected the field of natural language processing by eliminating the constraints posed by the conventional sequential models through contextual representation and attention modeling. Semantic analysis methods using transformers have been seen to improve relation extraction performance as they allow for more effective relation identification among textual items (Evans et al., 2024). The use of transformer-based deep learning models has helped improve the performance of sentiment analysis by capturing the implicit semantics and variations in context from text data (Kokab et al., 2022). Transformer-based semantic relation classification approaches have shown strong potential in identifying relationships among textual components through advanced language representations (Lee et al., 2022).

The findings in previous works concerning semantic analysis have revealed the effectiveness of transformer models for interpreting various types of textual information. New research has also used transformer models for performing sentiment analysis in review-based corpora, proving their effectiveness in interpreting users' sentiments and context-specific expressions (Saad et al., 2024). Models based on sentence transformers have been utilized for applying semantic analysis to discover text reusability and calculating phrase similarities (Mehak et al., 2023). Sentiment analysis models based on transformers have also been studied for working with multilingual texts, proving their efficiency in dealing with different languages and linguistic structures (Ozturk & Ozcan, 2022). Approaches for automatic measurement of sentiments combining transformer models have proved to be reliable for assessing large corpora of texts (Zhao & Wong, 2024). Novel transformer architectures have also been explored for working with linguistic phenomena like sarcasm detection and sentiment classification, proving the robustness of the models in difficult tasks of semantic analysis (Shah et al., 2023). All these results prove the effectiveness of transformer models for analyzing English literary and educational texts by interpreting context meanings and semantic relations.

In the reviewed papers, there is a confirmation of the higher potential of transformer-based models in relation to semantic analysis due to context relations, patterns, and deeper meanings. There are numerous examples of the successful use of transformer-based models for literary text analysis, sentiment analysis, semantic similarity analysis, and text understanding in education.

Methodology

Dataset Description

A double dataset will be employed in conducting this research, which is made up of English literary texts and English educational texts. The English literary texts will be made up of English novels, English short stories, English poetry books, and English plays. There are diverse linguistic characteristics of these texts, including narratives, feelings, symbolism, and themes. English educational texts include English textbooks, academic learning material, educational passages, and articles on education. These texts vary in terms of knowledge representation and linguistic aspects as well. The following is the dataset distribution as shown in table 1.

Table 1: Distribution of Literary and Educational Text Datasets Used for Transformer-Based Semantic Analysis

Text Category	Source Type	Number of Documents
Literary Texts	Novels, Short Stories, Poetry, Drama	5,000
Educational Texts	Textbooks, Academic Materials, Learning Resources	5,000
Total	Combined Dataset	10,000

Before semantic analysis, the collected texts undergo preprocessing operations including removal of unnecessary symbols, sentence segmentation, tokenization, normalization, and linguistic formatting.

Semantic Interpretation and Analysis Procedure

The process of semantic interpretation considers how the interpretation of meanings, themes, emotions, and conceptual relationships through the use of transformers is performed. In order to perform the analysis, the processed texts are examined based on their context. When considering the literary texts, the analysis was performed in terms of themes, emotions, narrative relationships, and symbolism found in novels, short stories, poetry, and drama. In the case of educational texts, the analysis allowed identifying the key concepts, knowledge relationships, and instructional meanings to help understand the text better. The identified semantic patterns were then evaluated from the perspective of English language and literature.

Proposed Transformer-Based Semantic Analysis Framework

The suggested framework has been designed to consist of four stages of transformer-based semantic analysis of English literary and educational texts. First, in the first stage, literary and educational texts are collected from different sources and preprocessed to get them structured through such processes as cleaning, tokenization, normalization, and sentence segmentation without losing any important linguistic information. Second, in the second stage, a pretrained transformer is used to obtain semantic representations based on relationships between words, sentences, and paragraphs. Third, in the third stage, semantic feature interpretation is performed to detect literary themes, emotional patterns, conceptual relationships, contextual meanings, and educational knowledge structures. Finally, in the fourth stage, the extracted semantic information is analyzed linguistically and computationally to evaluate the effectiveness of the transformer-based approach in improving contextual understanding and interpretation of English literary and educational texts.

Figure 1 represents the structure of the framework for semantic analysis using a transformer model for English text. This is the process from the acquisition of literary and educational text through preprocessing, contextual representation using a transformer model, and the semantic analysis process where concepts such as themes, emotions, relations, and education are discovered.

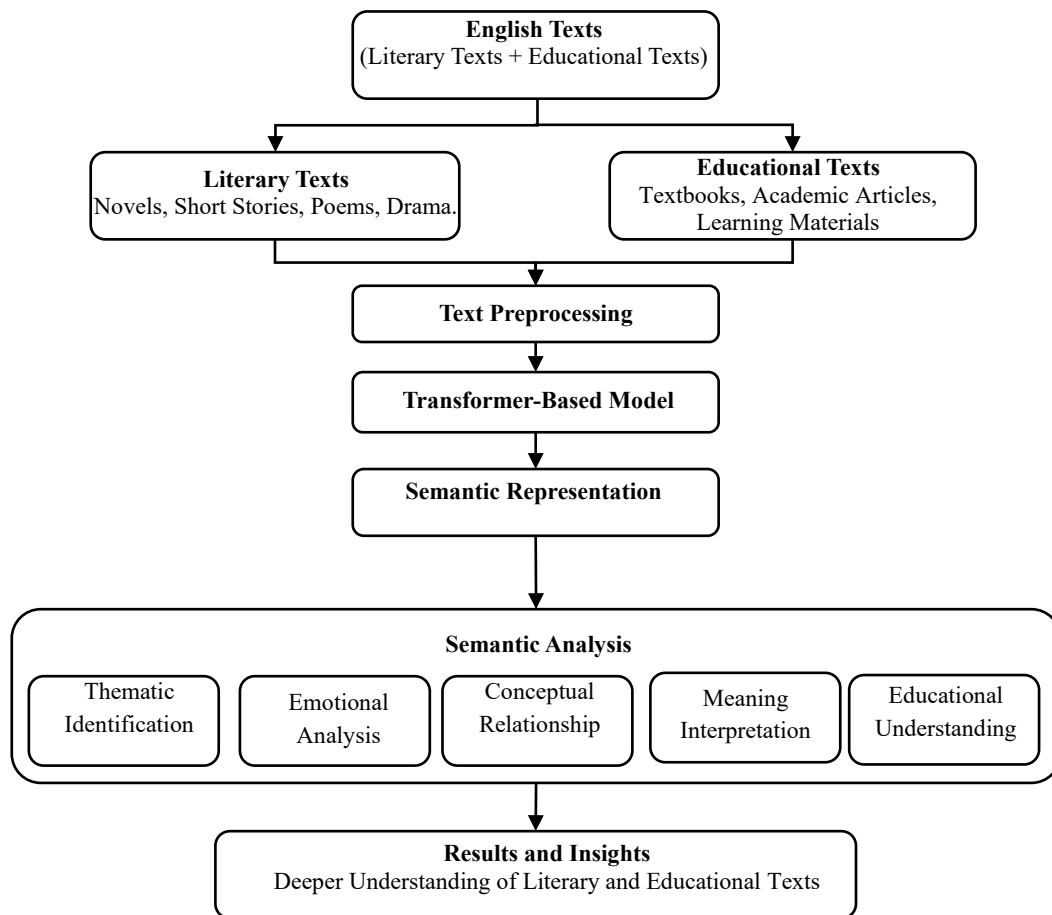


Figure 1: Conceptual Framework of Transformer-Based Semantic Analysis for English Literary and Educational Texts

Results

Enhanced Understanding of Literary and Educational Texts Through AI-Assisted Semantic Analysis

The performance of the suggested transformer-based semantic analysis framework has been analyzed by evaluating conventional text analysis methodologies against the proposed methodology in analyzing English literary and educational texts. Conventional text analysis techniques primarily rely upon keyword detection, statistics, and surface linguistic structure, which generally pose challenges in terms of detecting implicit meaning, emotion, thematic structure, and concept connections. Literary text contains structures such as metaphors, symbols, story structures, and meaningful interpretations that require analysis, whereas educational text requires proper representation of knowledge.

The transformer-based semantic analysis method improves text analysis in view of the fact that both the word/sentence relations and context are taken into consideration in the text analysis. The transformer-based semantic analysis helps understand the themes of literature, emotions, concepts, and meanings of education by analyzing the contextual links contained in English texts. Comparison between traditional semantic analysis and transformer-based semantic analysis is presented in table 2.

Table 2: Comparative Analysis of Semantic Understanding Before and After Transformer-Based Analysis

Analysis Aspect	Conventional Semantic Analysis	Transformer-Based Semantic Analysis
Contextual Meaning Understanding	Limited interpretation based on individual words and direct sentence patterns	Improved understanding through contextual relationships within complete texts
Literary Theme Identification	Requires manual examination and keyword-based extraction of themes	Enhanced identification of thematic structures and narrative patterns
Emotional Interpretation	Limited capability in recognizing implicit emotions and contextual expressions	Improved identification of emotional patterns and deeper meanings
Conceptual Relationship Understanding	Difficulty in connecting related concepts across different text segments	Effective identification of relationships among concepts and ideas
Educational Content Analysis	Requires manual organization of learning concepts and instructional information	Improved representation and organization of educational knowledge structures

Semantic Interpretation of English Literary Texts

In addition to the semantic analysis carried out using the transformer model, its capability to detect literary features in English literary texts was explored further. It was capable of detecting thematic features, emotions, and meaning from novels, short stories, poems, and drama texts. Semantic analysis showed that literary texts have interconnections among theme, emotion, and narrative. The transformer-based analysis successfully revealed important literary patterns, including thematic structures, emotional expressions, character relationships, and symbolic meanings embedded within literary texts.

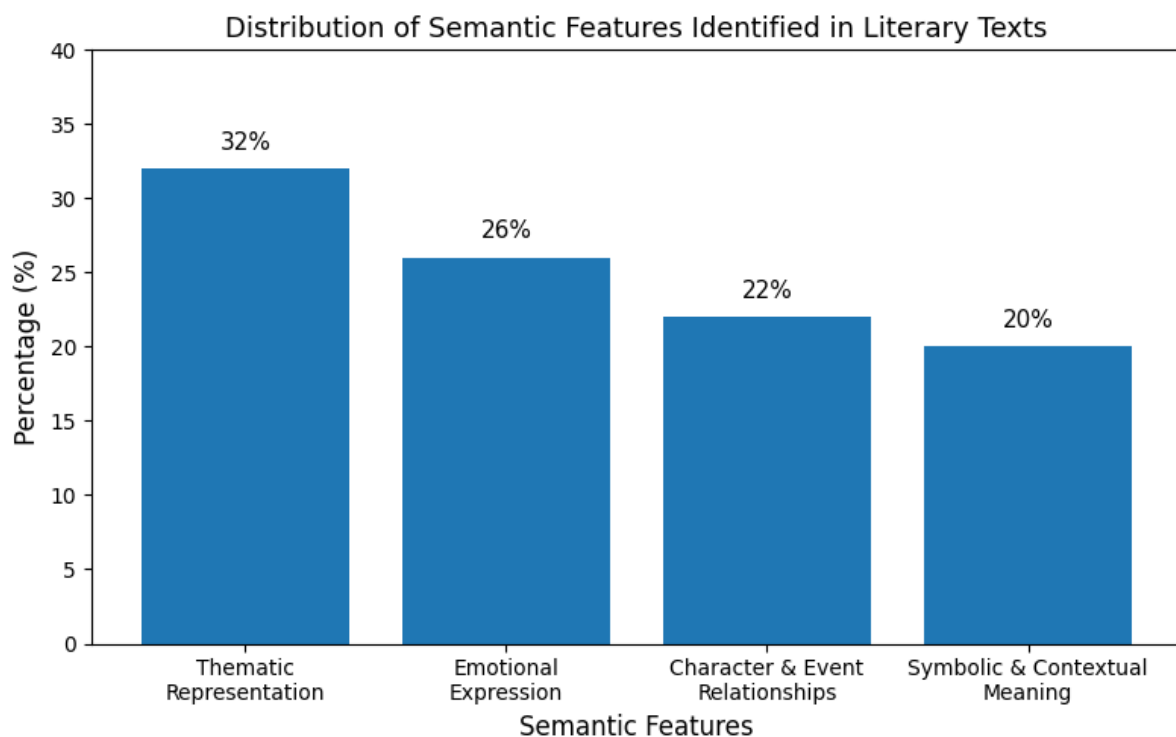


Figure 2: Distribution of Semantic Features Identified in Literary Texts

Figure 2 shows the distribution of the semantic features extracted from the English literary texts based on the semantic analysis of literature using the presented transformer-based approach. The analysis revealed that thematic representation is the main feature, while emotional expression, connections between characters and events, as well as symbolic-contextual meanings are other important semantic features of literature.

Educational Implications of Semantic Analysis Findings

The results show the possibility of applying transformer-based semantic analysis to improve English education by facilitating the deeper understanding of contextual and conceptual relations in educational materials. This proposed approach could be utilized in several educational applications such as automatic classification of educational materials, developing learners' reading comprehension abilities, and custom language learning. It is possible to use the ability of the model to identify important notions and text complexity to choose suitable educational material depending on learners' requirements. In addition, it will allow educators to evaluate the vast number of educational materials and adapt their educational approach accordingly. Using transformer-based semantic analysis connects the spheres of artificial intelligence, English education, and digital humanities together. It could be stated that the results of the study reveal that transformer-based semantic analysis could provide new perspectives on analyzing challenging texts and identifying important notions for English language learners.

Discussion

The findings of the present study reveal that transformer-based semantic analysis is a rather effective method for comprehending literary and educational texts in English. The ability of the transformer models to analyze contextual connections between words, sentences, and even bigger textual units allows for a deeper analysis of literary meanings, themes, emotions, and education-related concepts. It is evident from the results that semantic analysis assisted by artificial intelligence can create a connection between computational linguistics and English by helping scholars investigate patterns and semantic relations in texts. Nonetheless, some limitations have to be taken into consideration. Literary analysis cannot be conducted without cultural, historical, philosophical, and other aspects of life that can be partially included in computational models. Also, the transformer-based approach presupposes the availability of computational power, big training sets, and proper optimization of the model. The effectiveness of semantic analysis also depends on the quality and diversity of the chosen literary and educational texts. From the standpoint of English education, these results suggest that the application of semantic analysis based on transformers is capable of acting as a helpful analytical tool for the teacher and learner in the exploration of difficult meanings and literary aspects.

In addition to that, future research will allow enhancing transformer-based semantic analysis through the creation of explainable AI models providing clear explanations of literary meanings and the decision-making processes of such models. The use of multilingual transformers for comparative literature research is one of the potential ways of doing it. Moreover, implementation of multimodal learning models using text, images, and audio as inputs can also lead to gaining more insights into educational and literary sources. Development of adaptive AI systems aimed at personalization of English language studies through the analysis of learners' needs and recommendations of corresponding educational content is also an important future direction of research.

Conclusion

In this research work, the use of transformer-based semantic analysis technique in helping to interpret English literary and educational texts is assessed. As the results show, the application of such an approach can be helpful in terms of identifying contexts, themes, emotions, and concepts that can be hardly found by traditional means of text analysis. The suggested method has been analyzed on the basis of the database containing 10,000 English texts, including 5,000 literary texts and 5,000 educational texts. When analyzing literary texts, it has been found that thematic representation constitutes the biggest part of semantic features, being equal to 32%, emotional expression takes 26%, relationships between characters and events – 22%, and symbolic-contextual meanings – 20%.

Moreover, the findings emphasize the importance of transformer-based semantic analysis in English language studies, as it helps to structure educational material, define key concepts, and enhance the process of reading comprehension. This method creates a link between artificial intelligence, digital humanities and language education by adding extra help for researchers, teachers and students while analyzing texts. Nevertheless, the research acknowledges that the process of interpretation in literature also involves cultural, historical, and philosophical aspects that cannot be fully interpreted by means of computational tools. Future studies may concentrate on creating explainable and multimodal AI techniques that would use the information from text, image, and context for more extensive interpretation of literary and educational resources. In general, the research proves that transformer-based semantic analysis can become a good supplement for modern English language studies and learning technologies.

References

- Abdalgader, K., Matroud, A. A., & Hossin, K. (2024). Experimental study on short-text clustering using transformer-based semantic similarity measure. *PeerJ Computer Science*, 10, e2078.
<https://doi.org/10.7717/peerj-cs.2078>
- Evans, M. T. C., Latifi, M., Ahsan, M., & Haider, J. (2024). Leveraging semantic text analysis to improve the performance of transformer-based relation extraction. *Information*, 15(2), 91.
<https://doi.org/10.3390/info15020091>
- Junaid, T., Sumathi, D., Sasikumar, A. N., Suthir, S., Manikandan, J., Khilar, R., ... & Raju, M. J. (2022). A comparative analysis of transformer based models for figurative language classification. *Computers and Electrical Engineering*, 101, 108051.
<https://doi.org/10.1016/j.compeleceng.2022.108051>
- Kamil, M., & Çakır, D. (2025). Advances in transformer-based semantic search: Techniques, benchmarks, and future directions. *Turkish Journal of Mathematics and Computer Science*, 17(1), 145-166.
<https://doi.org/10.47000/tjmcs.1633092>
- Kokab, S. T., Asghar, S., & Naz, S. (2022). Transformer-based deep learning models for the sentiment analysis of social media data. *Array*, 14, 100157. <https://doi.org/10.1016/j.array.2022.100157>
- Lee, Y., Son, J., & Song, M. (2022). BertSRC: transformer-based semantic relation classification. *BMC Medical Informatics and Decision Making*, 22(1), 234.
<https://doi.org/10.1186/s12911-022-01977-5>
- Mehak, G., Muneer, I., & Nawab, R. M. A. (2023). Urdu text reuse detection at phrasal level using sentence transformer-based approach. *Expert Systems with Applications*, 234, 121063.
<https://doi.org/10.1016/j.eswa.2023.121063>

- Monika, R. P., Hemalatha, S., Sayed, M. Y., Mohammad, R., Preshiya, D. J., & Bala, B. K. (2025, October). Leveraging NLP with Sentiment-Aware Transformers to Analyze Literary Texts for Emotion and Stylistic Patterns. In *2025 International Conference on Communication, Computer, and Information Technology (IC3IT)* (pp. 1-7). IEEE.
<https://doi.org/10.1109/IC3IT66137.2025.11341044>
- Ozturk, O., & Ozcan, A. (2022, May). Sentiment analysis in turkish using transformer-based deep learning models. In *The International Conference on Artificial Intelligence and Applied Mathematics in Engineering* (pp. 1-15). Cham: Springer International Publishing.
https://doi.org/10.1007/978-3-031-31956-3_1
- Saad, T. B., Ahmed, M., Ahmed, B., & Sazan, S. A. (2024, May). A novel transformer based deep learning approach of sentiment analysis for movie reviews. In *2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT)* (pp. 1228-1233). IEEE. <https://doi.org/10.1109/ICEEICT62016.2024.10534588>
- Shah, S. M. A. H., Shah, S. F. H., Ullah, A., Rizwan, A., Atteia, G., & Alabdulhafith, M. (2023). Arabic sentiment analysis and sarcasm detection using probabilistic projections-based variational switch transformer. *IEEE Access*, *11*, 67865-67881. <https://doi.org/10.1109/ACCESS.2023.3289715>
- Shiha, R. B., Atwell, E., & Abbas, N. (2025, November). A Transformer-Based Framework for Thematic Analysis of University Reading Lists. In *International Conference on Innovative Techniques and Applications of Artificial Intelligence* (pp. 405-410). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-11442-6_28
- Tzimiris, S., Nikiforos, S., Nikiforos, M. N., Mouratidis, D., & Kermanidis, K. L. (2025). A comparative evaluation of transformer-based language models for topic-based sentiment analysis. *Electronics*, *14*(15), 2957. <https://doi.org/10.3390/electronics14152957>
- Yang, X., He, X., Zhang, H., Ma, Y., Bian, J., & Wu, Y. (2020). Measurement of semantic textual similarity in clinical texts: comparison of transformer-based models. *JMIR medical informatics*, *8*(11), e19735. <https://doi.org/10.2196/19735>
- Zhao, X., & Wong, C. W. (2024). Automated measures of sentiment via transformer-and lexicon-based sentiment analysis (TLISA). *Journal of Computational Social Science*, *7*(1), 145-170.
<https://doi.org/10.1007/s42001-023-00233-8>