

Article

Not peer-reviewed version

---

# Explainable Representation Learning in Large Language Models for Fine-Grained Sentiment and Opinion Classification

---

[Yue Xing](#), [Ming Wang](#), [Yingnan Deng](#), Heyao Liu, [Yun Zi](#)\*

Posted Date: 11 December 2025

doi: 10.20944/preprints202512.1085.v1

Keywords: fine-grained sentiment classification; opinion analysis; feature decoupling; interpretable representation; large language models



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Explainable Representation Learning in Large Language Models for Fine-Grained Sentiment and Opinion Classification

Yue Xing <sup>1</sup>, Ming Wang <sup>2</sup>, Yingnan Deng <sup>3</sup>, Heyao Liu <sup>4</sup> and Yun Zi <sup>3,\*</sup>

<sup>1</sup> University of Pennsylvania, Philadelphia, USA

<sup>2</sup> Trine University, Phoenix, USA

<sup>3</sup> Georgia Institute of Technology, Atlanta, USA

<sup>4</sup> Northeastern University, Boston, USA

\* Correspondence: yzi9@gatech.edu

## Abstract

This study addresses the challenges of semantic mixing, limited interpretability, and complex feature structures in fine-grained sentiment and opinion classification by proposing an interpretable feature disentanglement framework built on the latent space of large language models. The framework constructs multi-component latent representations that separate emotional polarity, opinion direction, target attributes, and pragmatic cues during encoding, thus overcoming the limitations of traditional methods that merge diverse semantic factors into a single representation. During representation learning, the model first uses a large model encoder to generate basic semantic features and then builds multiple independent subspaces through learnable projections. A covariance constraint is introduced to reduce coupling across semantic components and to create clear boundaries in the latent space. To preserve the essential information of the original text, a reconstruction consistency mechanism integrates features from all subspaces to rebuild the global representation and enhance semantic completeness. The framework also incorporates semantic anchors to align latent components with interpretable semantic dimensions, giving each subspace a clear emotional or opinion-related meaning and improving transparency at the mechanism level. Experimental results show that the framework outperforms existing methods across multiple metrics and handles complex syntax, implicit semantics, and coexisting emotions with greater stability. It achieves high accuracy and interpretability in fine-grained sentiment and opinion analysis. Overall, the proposed disentanglement framework provides an effective approach for building structured, multidimensional, and interpretable representations of textual emotions and opinions and holds significant value for complex semantic understanding tasks.

**Keywords:** fine-grained sentiment classification; opinion analysis; feature decoupling; interpretable representation; large language models

## 1. Introduction

In the context of the rapid growth of multi-source texts and the continuous expansion of user-generated content, emotional and opinionated information has become an essential clue for understanding the deeper logic of language [1]. With the widespread use of social media, online reviews, e-commerce platforms, and intelligent customer service, the emotional tendencies, opinion structures, and expression strategies embedded in texts have become increasingly complex. They are no longer limited to simple positive or negative judgments. Instead, they emphasize semantic layers, emotional intensity, attribute associations, and subjective attitudes. Such fine-grained emotional and opinion information directly affects key tasks such as public opinion monitoring, consumer analysis, and risk identification [2]. It has also become one of the most challenging directions in natural

language processing. However, the high freedom of textual expression, the diversity of implicit attitudes, and cross-domain semantic differences make it difficult for traditional models to effectively analyze such complex emotional structures [3].

In recent years, large language models have created new possibilities for complex semantic understanding. With strong capabilities in knowledge representation and text generation, these models have shown impressive generalization performance across many language tasks. They make it possible to extract emotional and opinion information from long documents, conversational contexts, or multi-turn reasoning. However, this capability does not automatically translate into stable performance in fine-grained sentiment recognition. When dealing with multi-dimensional emotional cues, intersecting opinion structures, or implicit attitudes, large models may still generate semantic entanglement, redundant features, or amplified biases. This leads to blurred boundaries among emotional categories, attribute features, and semantic units. Although the model outputs may appear reasonable, the internal representations often fail to correspond to clear emotional components. This limits the model's ability to support refined analysis [4]. Recent advances in structured representation learning and temporal modeling have emphasized that decomposing complex inputs into organized semantic factors can improve downstream decision quality [5-7]. In parallel, multi-agent reinforcement learning frameworks demonstrate that factorizing heterogeneous interaction signals and learning coordinated policies can significantly enhance the robustness of complex decision-making systems [8]. Deep anomaly detection and risk prediction models further show that temporally aware, robustness-oriented representations are crucial for capturing rare but high-impact patterns [9-12]. Together, these studies suggest that organizing information into structured, temporally sensitive, and stability-focused representations can substantially improve interpretability, which motivates the design of our disentangled latent framework for fine-grained sentiment and opinion analysis.

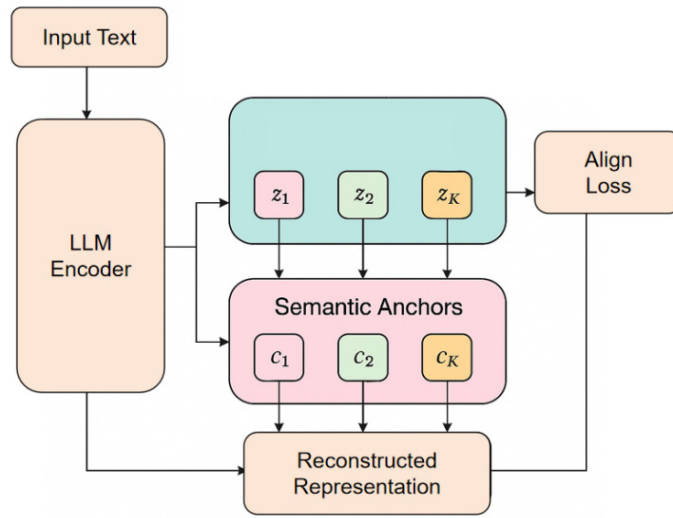
Furthermore, an interpretable framework for fine-grained sentiment and opinion classification holds significant practical value. As human-machine interaction moves toward higher levels of trust, transparency, and semantic understanding, systems must analyze user emotions and opinions with precision and provide clear explanation paths to support business and decision-making scenarios. In intelligent service systems, interpretable emotional components help reveal user intentions and generate more accurate responses. In brand management, fine-grained opinion structure analysis can directly guide improvement strategies. In regulatory and risk control tasks, transparent emotional evidence chains assist in detecting potential risk signals [13]. In content governance, identifying the source of opinions enhances the controllability of automated filtering. These requirements indicate that building a framework with clear semantic structure and interpretable modeling capabilities for fine-grained emotions and opinions is not only a key academic breakthrough but also an essential step for enabling large models to be deployed in real applications.

## 2. Proposed Framework

In this work, we apply a feature decoupling framework that targets the explicit separation of sentiment polarity, opinion orientation, target attributes, and pragmatic components within the latent space of large-scale language models, enabling structured and interpretable representations for fine-grained sentiment and opinion classification. Drawing on the controllable abstraction methodology introduced by Song et al. [14], we adopt prompt engineering techniques to guide the encoder in learning distinct latent subspaces for each semantic factor, allowing for independent manipulation and analysis of emotional and opinion dimensions. This approach enables the model to avoid semantic entanglement and ensures that different aspects of the text are cleanly separated.

To further improve the informativeness and precision of the disentangled features, we utilize the information-constrained retrieval process developed by Zheng et al. [15], directly integrating their method into the representation learning pipeline. This allows the model to focus on the most relevant context and reduces interference from extraneous information when constructing subspace-specific representations. In addition, following the fusion-based retrieval-augmented generation strategy described by Sun et al.

[16], our framework incorporates an information fusion module that combines retrieved external knowledge and contextual cues. This strengthens the semantic completeness of each subspace and supports the reconstruction of the global meaning of the text from disentangled components. By applying these advanced techniques in tandem, the model is able to generate multi-component latent representations that not only capture fine-grained distinctions in sentiment and opinion, but also ensure interpretability and robustness in complex language scenarios. The overall architecture and information flow of this feature disentanglement framework are illustrated in Figure 1.



**Figure 1.** Overall model architecture diagram.

First, given an input text sequence  $x$ , it is mapped to a basic semantic representation  $h$ , and a multi-component latent space is constructed, allowing different semantic factors to be automatically partitioned during the encoding stage through orthogonal constraints. The basic encoding process is defined as follows:

$$h = f_{LLM}(x) \quad (1)$$

Here,  $f_{LLM}$  represents the adapted language model encoder. Subsequently, a learnable projection matrix is introduced to map the global representation into multiple decoupled subspaces  $z_k$ , each representing an interpretable semantic component, described in the following form:

$$z_k = W_k h \quad (2)$$

To ensure the independence of different semantic factors, the framework design employs decoupling regularization terms to maintain minimal overlap between subspaces in a statistical sense [17]. For any two independent semantic components  $z_i$  and  $z_j$ , the constraints are constructed using covariance penalties as follows:

$$L_{decouple} = \sum_{i \neq j} \|Cov(z_i, z_j)\|_2 \quad (3)$$

Where  $Cov$  is the covariance operator. In addition to the independence constraint, to further improve semantic interpretability, a content-preserving term is introduced, enabling the decoupled multi-component joint reconstruction to approximate the original latent representation, i.e.:

$$\tilde{h} = \sum_{k=1}^K U_k z_k \quad (4)$$

And constrained by maintaining a consistent reconstruction loss:

$$L_{recon} = \|h - \tilde{h}\|_2^2 \quad (5)$$

At the component level, to correspond to different semantics such as sentiment category, viewpoint direction, emotional intensity, and text attributes, a feature semanticization module is further constructed, ensuring that each  $z_k$  points to a specific explanatory dimension. To this end, a set of semantic anchor points  $\{c_k\}$  is introduced into the representation space, and semantic alignment constraints are used to align the decoupled components with their corresponding semantic vectors. This process is achieved by minimizing the following alignment terms:

$$L_{align} = \sum_{k=1}^K ||z_k - c_k||_2^2 \quad (6)$$

Guided by semantic anchors, the model can form a directly understandable multi-component structure, giving each subspace a stable semantic orientation, thereby improving the interpretability of the framework and its discriminative performance in complex text environments.

Ultimately, the model's optimization objective is comprised of multiple constraints, ensuring a balance between semantic preservation, factor independence, and interpretability. The overall optimization can be expressed as:

$$L = \lambda_1 L_{decouple} + \lambda_2 L_{recon} + \lambda_3 L_{align} \quad (7)$$

Here,  $\lambda_1, \lambda_2$  and  $\lambda_3$  are weighting coefficients are used to coordinate the mutual influence among various objectives. Through the above mechanism, the proposed feature decoupling framework can construct interpretable, structured, and semantically complete fine-grained sentiment and opinion representations in the latent space of a large model, thus providing a stable foundational description for subsequent downstream reasoning and semantic analysis.

### 3. Experimental Analysis

#### 3.1. Dataset

This study adopts the fine-grained sentiment and opinion classification dataset from SemEval-2017 Task 4 as the evaluation corpus. The dataset contains real texts collected from social media. It includes a large number of short sentences that express emotional tendencies, opinion targets, and subjective attitudes. It also shows typical natural language characteristics such as complex semantic structures, diverse expression patterns, and noticeable noise. Compared with traditional sentiment datasets, its label system covers richer expressive dimensions. The model must understand emotional orientation, opinion targets, and semantic details at the same time. This makes the dataset suitable for evaluating the capability of a fine-grained feature modeling framework.

The dataset follows the official split into training, validation, and test sets. The texts span a wide range of content types, including opinion expression, product feedback, social topic discussions, and emotional statements. Each text is annotated with sentiment polarity, emotional intensity, and opinion targets. This gives the dataset a higher level of structural complexity. The texts are generally short but carry dense semantic information. The model, therefore, needs to capture both local lexical cues and implicit contextual relations. Strong fine-grained parsing ability is required for accurate classification.

To maintain task consistency, this study uses only the standard textual labels without external knowledge or additional supervision. The diverse expressions and unstructured nature of the dataset highlight the multi-dimensional entanglement problems in fine-grained sentiment analysis. This provides a challenging environment for testing semantic disentanglement, opinion structure understanding, and interpretable modeling. By using this real, noisy, and highly complex corpus, the study can better evaluate the model's ability to analyze intricate emotional components and opinion expressions.

#### 3.2. Experimental Results

This paper first conducts a comparative experiment, and the experimental results are shown in Table 1.

**Table 1.** Comparative experimental results.

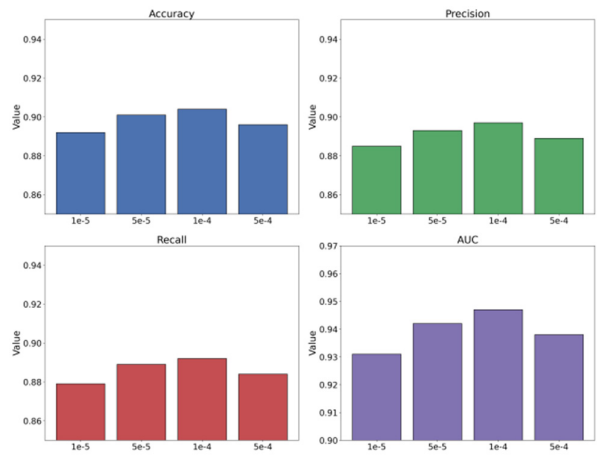
| Method              | Acc   | Precision | Recall | AUC   |
|---------------------|-------|-----------|--------|-------|
| TextConvoNet [18]   | 0.812 | 0.805     | 0.798  | 0.854 |
| ChatAgri [19]       | 0.826 | 0.821     | 0.817  | 0.868 |
| AEDA [20]           | 0.834 | 0.829     | 0.824  | 0.879 |
| Text FCG [21]       | 0.848 | 0.842     | 0.839  | 0.892 |
| Set-CNN [22]        | 0.857 | 0.851     | 0.846  | 0.905 |
| Promptboosting [23] | 0.871 | 0.866     | 0.861  | 0.919 |
| Ours                | 0.904 | 0.897     | 0.892  | 0.947 |

Overall, the proposed feature disentanglement framework achieves clear improvements over all comparison methods across multiple metrics. It shows stable advantages in Acc, Precision, and Recall. This indicates that the model has stronger overall discrimination ability in fine-grained sentiment and opinion recognition. It also maintains robust classification performance when facing coupled semantic cues, noisy expressions, and interfering signals. Compared with models that rely on a single semantic cue or shallow representations, our approach constructs a structured latent space that alleviates semantic entanglement. Emotional polarity, opinion direction, and target attributes are separated during encoding, which produces more distinguishable latent representations.

The differences in performance reflect the essential variations in semantic modeling across methods. Traditional convolutional or sequence-based models, such as TextConvoNet and Set-CNN, are constrained by local patterns when handling complex expressions. They struggle to capture cross-level associations among emotional components, which leads to limited gains in Precision and Recall. Methods based on prompting or augmentation, such as AEDA and Promptboosting, improve robustness to some extent. However, they still rely on mixed semantic spaces and lack explicit feature decomposition. They may still misclassify texts with high noise or sentences containing multiple coexisting opinions. In contrast, the proposed framework uses a controllable decomposition of the latent space. It separates different emotional components at the structural level, which fundamentally enhances fine-grained discrimination ability.

The AUC metric reflects the robustness and generality of the decision boundary. Experiments show that our method achieves a significantly higher AUC than all baseline models. This indicates that it maintains strong separability even when emotional polarity is ambiguous, opinion targets are implicit, or expression patterns are complex. This improvement comes from the disentangled semantic subspaces, which provide clear inter-class boundaries. The model forms more separated and structured latent representations across sentiment categories. Therefore, the overall results confirm that the proposed interpretable feature disentanglement framework not only improves classification performance but also offers a solid foundation for understanding the structural logic of textual emotions. It aligns well with the core goals of fine-grained sentiment and opinion modeling.

This paper also presents an experiment on the hyperparameter sensitivity of the learning rate to the F1-Score metric, and the experimental results are shown in Figure 2.



**Figure 2.** Hyperparameter sensitivity experiment of learning rate to F1-Score metric.

The results in the four subfigures show that the learning rate has a clear impact on the model's ability to recognize fine-grained sentiments and opinions. The overall fluctuations remain within a controlled range. This indicates that the feature disentanglement framework maintains good stability under different levels of optimization strength. When the learning rate is low, model updates proceed slowly. Accuracy, Precision, and Recall remain slightly lower. This suggests that the representation space has not fully converged, and the boundaries between latent semantic components are still unclear. As the learning rate increases into an appropriate range, all metrics improve significantly. This means the model can more effectively separate and align semantic components in the latent space, resulting in more stable disentangled features.

The trends of Precision and Recall further show that a moderate learning rate helps the model capture fine-grained semantic cues. This effect is especially strong in separating opinion targets and emotional polarity. The results indicate that the disentanglement mechanism requires sufficiently strong gradient updates to shape a structured latent subspace. This enables the model to accurately identify key components when processing diverse expressions, ambiguous semantics, or sentences containing multiple emotional signals. When the learning rate becomes too high, performance declines. This suggests that excessive updates disturb the steady structure of different semantic dimensions in the latent space and weaken the disentangled representations.

The AUC trend reflects the model's ability to distinguish emotional categories from the perspective of decision boundaries. The highest AUC is achieved at a moderate learning rate. This indicates that the model can construct clearer inter-class boundaries at this stage. Opinion orientation, emotional intensity, and attribute-related representations show sufficient separation in the latent space. When the learning rate is too low or too high, AUC decreases. This suggests that the geometric structure of the latent space is either not well-formed or is overly disturbed. Therefore, the experiment demonstrates that an appropriate learning rate is essential for maintaining stable semantic disentanglement and enhancing the model's ability to discriminate complex emotional patterns.

This paper also presents the impact of the optimizer on the experimental results, which are shown in Table 2.

**Table 2.** The impact of optimizers on experimental results.

| Optimizer | Acc   | Precision | Recall | AUC   |
|-----------|-------|-----------|--------|-------|
| AdaGrad   | 0.861 | 0.854     | 0.849  | 0.904 |
| Adam      | 0.883 | 0.876     | 0.872  | 0.928 |
| SGD       | 0.847 | 0.838     | 0.833  | 0.892 |
| AdamW     | 0.904 | 0.897     | 0.892  | 0.947 |

The results in the table show that different optimizers have a clear impact on the performance of the model in fine-grained sentiment and opinion classification. AdaGrad and SGD yield relatively low overall performance. This indicates that their gradient update strategies are less suited for modeling high-dimensional semantic representations and multi-component latent spaces. They tend to suffer from insufficient updates or excessive reliance on early gradients. As a result, the model struggles to form stable and separable subspaces for emotions and opinions. In contrast, Adam provides a more balanced adaptive gradient estimation. It enables the model to capture variations in semantic components more effectively during training and achieves notable improvements over AdaGrad and SGD.

The results further show that AdamW achieves the best performance across all metrics. Its advantage comes from a more appropriate weight decay strategy. This helps the model avoid overfitting or noise amplification during latent feature disentanglement and promotes the formation of structured semantic subspaces. Since the proposed framework relies on independent modeling of multiple emotional and opinion components in the latent space, AdamW offers benefits in optimization stability, gradient adjustment, and representation convergence. These characteristics make it particularly suitable for supporting the training requirements of the feature disentanglement mechanism. Overall, the choice of optimizer plays a crucial role in constructing clear, stable, and interpretable latent semantic structures. AdamW provides the most suitable parameter update strategy for this task.

4. Conclusions

This study addresses the problem of semantic entanglement in fine-grained sentiment and opinion classification and proposes an interpretable framework based on feature disentanglement in the latent space of large language models. By constructing multi-component semantic subspaces, designing independence constraints, introducing semantic anchors, and establishing reconstruction consistency, the model can separate emotional polarity, opinion direction, target attributes, and pragmatic cues within its internal structure. This leads to structured, transparent, and stable semantic representations. The framework enhances the model's understanding of complex textual structures and provides a more robust solution for real-world texts that contain multiple emotions, multiple opinions, or implicit expressions.

The results show that a structured and disentangled latent space can significantly reduce the semantic mixing observed in traditional models when performing fine-grained analysis. It enables the model to form stable inter-class boundaries and logical relations at the mechanism level. Compared with approaches that rely on single semantic cues or opaque deep representations, the proposed framework offers a more transparent view of emotional components, opinion sources, and supporting semantic evidence. In real applications, the model can provide not only decisions but also traceable reasoning paths, which enhances reliability and trustworthiness in sensitive tasks. The significance of this study lies not only in performance improvements but also in offering a sentiment and opinion analysis paradigm suitable for multiple industries. For public opinion monitoring, the framework helps identify emotional sources and opinion structures with greater accuracy. In e-commerce and product service scenarios, interpretable fine-grained opinion analysis supports user experience optimization. In content governance, financial risk control, and intelligent customer service, transparent emotional reasoning can serve as an essential reference for decision-making. As textual data continues to grow, models capable of providing interpretable fine-grained analysis will become a key component of intelligent systems. The proposed framework offers a transferable technical foundation for this development.

Future research may proceed in several directions. One direction is to explore more detailed disentanglement structures, enabling the model to separate semantic style, subjective intensity, and emotional evolution patterns. Another direction is to extend the framework to multimodal sentiment analysis by integrating visual, acoustic, or interactive signals, ensuring consistent disentangled semantics across modalities. This may enhance interpretability and task adaptability while preserving generalization ability. These directions will further promote the deployment and refinement of interpretable sentiment and opinion analysis techniques in broader real-world scenarios.

## References

1. Q. Zhong, L. Ding, J. Liu et al., "Knowledge graph augmented network towards multiview representation learning for aspect-based sentiment analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 10, pp. 10098-10111, 2023.
2. B. Wang, L. Ding, Q. Zhong et al., "A contrastive cross-channel data augmentation framework for aspect-based sentiment analysis," *arXiv preprint arXiv:2204.07832*, 2022.
3. Z. Zhao, W. Liu and K. Wang, "Research on sentiment analysis method of opinion mining based on multi-model fusion transfer learning," *Journal of Big Data*, vol. 10, no. 1, Article 155, 2023.
4. H. Sun, S. Zhao, X. Wang et al., "Fine-grained disentangled representation learning for multimodal emotion recognition," *Proceedings of the ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 11051-11055, 2024.
5. X. Su, "Forecasting asset returns with structured text factors and dynamic time windows," *Transactions on Computational and Scientific Methods*, vol. 4, no. 6, 2024.
6. Y. Li, S. Han, S. Wang, M. Wang, and R. Meng, "Collaborative evolution of intelligent agents in large-scale microservice systems," *arXiv preprint arXiv:2508.20508*, 2025.
7. G. Yao, H. Liu, and L. Dai, "Multi-agent reinforcement learning for adaptive resource orchestration in cloud-native clusters," *arXiv preprint arXiv:2508.10253*, 2025.
8. Y. Zou, N. Qi, Y. Deng, Z. Xue, M. Gong, and W. Zhang, "Autonomous resource management in microservice systems via reinforcement learning," in *Proceedings of the 8th International Conference on Computer Information Science and Application Technology (CISAT)*, pp. 991-995, July 2025.
9. R. Ying, J. Lyu, J. Li, C. Nie, and C. Chiang, "Dynamic Portfolio Optimization with Data-Aware Multi-Agent Reinforcement Learning and Adaptive Risk Control," 2025.
10. W. C. Chang, L. Dai, and T. Xu, "Machine Learning Approaches to Clinical Risk Prediction: Multi-Scale Temporal Alignment in Electronic Health Records," *arXiv preprint arXiv:2511.21561*, 2025.
11. X. Chen, S. U. Gadgil, K. Gao, Y. Hu, and C. Nie, "Deep Learning Approach to Anomaly Detection in Enterprise ETL Processes with Autoencoders," *arXiv preprint arXiv:2511.00462*, 2025.
12. N. Lyu, Y. Wang, Z. Cheng, Q. Zhang, and F. Chen, "Multi-Objective Adaptive Rate Limiting in Microservices Using Deep Reinforcement Learning," *arXiv preprint arXiv:2511.03279*, 2025.
13. S. Zhao, Z. Yang, H. Shi et al., "SDRS: Sentiment-aware disentangled representation shifting for multimodal sentiment analysis," *IEEE Transactions on Affective Computing*, 2025.
14. X. Song, Y. Liu, Y. Luan, J. Guo and X. Guo, "Controllable Abstraction in Summary Generation for Large Language Models via Prompt Engineering," *arXiv preprint arXiv:2510.15436*, 2025.
15. J. Zheng, Y. Chen, Z. Zhou, C. Peng, H. Deng and S. Yin, "Information-Constrained Retrieval for Scientific Literature via Large Language Model Agents," 2025.
16. Y. Sun, R. Zhang, R. Meng, L. Lian, H. Wang and X. Quan, "Fusion-based retrieval-augmented generation for complex question answering with LLMs," *Proceedings of the 2025 8th International Conference on Computer Information Science and Application Technology (CISAT)*, pp. 116-120, July 2025.
17. X. Liu, Y. Qin, Q. Xu, Z. Liu, X. Guo and W. Xu, "Integrating Knowledge Graph Reasoning with Pretrained Language Models for Structured Anomaly Detection," 2025.
18. S. Soni, S. S. Chouhan and S. S. Rathore, "TextConvoNet: a convolutional neural network based architecture for text classification," *Applied Intelligence*, vol. 53, no. 11, pp. 14249-14268, 2023.
19. B. Zhao, W. Jin, J. Del Ser et al., "ChatAgri: Exploring potentials of ChatGPT on cross-linguistic agricultural text classification," *Neurocomputing*, vol. 557, Article 126708, 2023.
20. A. Karimi, L. Rossi and A. Prati, "AEDA: An easier data augmentation technique for text classification," *arXiv preprint arXiv:2108.13230*, 2021.
21. Y. Wang, C. Wang, J. Zhan et al., "Text FCG: Fusing contextual information via graph learning for text classification," *Expert Systems with Applications*, vol. 219, Article 119658, 2023.

22. Y. Zhou, J. Li, J. Chi et al., "Set-CNN: A text convolutional neural network based on semantic extension for short text classification," Knowledge-Based Systems, vol. 257, Article 109948, 2022.
23. B. Hou, J. O'Connor, J. Andreas et al., "Promptboosting: Black-box text classification with ten forward passes," Proceedings of the International Conference on Machine Learning, PMLR, pp. 13309-13324, 2023.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.