International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 4, Jul-Aug-2026, ISSN (Online): 2395-566X

An Explainable Transformer-Based Framework for Detecting Misinformation in Social Media

Dr. Prakash Kammam[1], Mukkapati Venu[2], K. Ashwini[3]
[1](Assistant Professor,
Master of Computer Applications ,
Sri Chaitanya Technical Campus,
Sheriguda(V), Ibrahimpatnam(M), Rangareddy(Dist) Telangana – 501510.
dr.prakashsctc@gmail.com)
[2](Assistant Professor,
Master of Computer Applications,
Sri Chaitanya Technical Campus,
Sheriguda(V),Ibrahimpatnam(M),Ranga Reddy Dist.,Telangana-501510.
mukkapati99@gmail.com)
[3](Assistant Profesor,
Master of Computer Applications,
Sri ChaitanyaTechnical Campus,
Sheriguda(V),Ibrahimpatnam(M),Ranga Reddy Dist.,Telangana-501510.
ashwinikdesai005@gmail.com)

Abstract- The fast spread of misinformation on social media platforms is causing serious issues with regards to public trust, democracy, and making sound decisions. The following paper provides a comprehensive overview of transformer-based systems for explainable misinformation detection, considering the latest research in the field of multimodal fusion, large language models implementation, and explainable AI. As can be seen from the systematic analysis, transformers surpass traditional methods in performance, with the multimodal system providing an accuracy of up to 94.5% and 81.1% on benchmark datasets. Moreover, large language models are very useful when generating background knowledge and enriching context, whereas explainability methods such as SHAP and LIME offer human-interpretable rationales for decisions made by the model. It was found that hierarchical progressive transformers successfully incorporate multimodality, combining different types of data such as text, images, background knowledge, and user comments, resulting in better performance than current methods.

Keywords: Misinformation Detection, Transformers, Explainable AI, Multimodal Fusion, Social Media, Large Language Models.

I. INTRODUCTION

The digital era has revolutionized the way information is produced, distributed, and consumed [6][10]. Social media sites like Twitter, Facebook, and Weibo are the most common mediums for individuals to get informed, with social media site Weibo garnering 598 million visits per month from those searching for information [6][10]. Although information's democratization has given power to individuals and connected the globe, it has also resulted in the fast spread of misinformation and misleading material [6][10].

Misinformation has severe consequences and is detrimental to individuals [6][10]. Misinformation may control public opinion and lead to the destabilization of the democratic process [6][10]. In the case of worldwide emergencies like the COVID-19 epidemic, misinformation was responsible for vaccination resistance, disinformation campaigns, and political violence [6][10]. Tensions in the current geopolitical environment have shown the consequences of misinformation in the media, where major news networks publish unverifiable and edited media of artificial intelligence-generated images wrongly showing the occurrence of military activities that never happened [6][10]. Economic damage is significant, too, with fake news being responsible for $39 billion worth of damage to the stock exchange annually and 86% of people encountering fake news worldwide [6][10].

However, the problem of misinformation has been made even more difficult due to the recent developments in generative AI [4][6]. The models such as ChatGPT, GPT-4, and DeepSeek have the capability of creating huge volumes of disinformation that is almost impossible to distinguish from the true content manually [4][6]. Fact-checking experts are well-versed in authenticating the legitimacy of news but find it difficult to deal with the volume of content available on the internet [6].

© 2026 IJSRET

However, if all the false news needs to be authenticated by these experts, there will be a delay [6].

There are major limitations in traditional techniques that have been used for the purpose of misinformation detection [7][8]. The early use of machine learning algorithms, such as Support Vector Machine and decision trees, was inadequate in handling the semantics associated with deceptive text [7]. Algorithms like Convolutional Neural Network and Recurrent Neural Networks enhanced representation but often failed to generalize in other contexts because of inadequate context modeling [10]. In addition, there are issues related to these algorithms being opaque or "black box" in nature [7][8].

The development of architectures based on transformers is a groundbreaking leap in the fields of natural language processing and misinformation detection, where such models as BERT, RoBERTa, and others utilize self-attention mechanism for understanding deep contextual relations within texts, showing superior performance across a number of benchmarks [2][7][8]. Nevertheless, there are still some challenges[3]. The first one consists in the fact that most of the models used for text analysis do not have any contextual information, which complicates the process of misinformation detection greatly [1][3]. The second challenge is related to the problem of incorporating several modalities into the process, including text, image data, and context information, because of complementarity and redundancy problems [1][2].

This paper offers a thorough investigation of transformer-based frameworks for the detection of misinformation on social media [1]. The research highlights recent developments in areas such as multimodal fusion, large language model utilization, hierarchical progressive transformers, and explainable AI technologies, presenting quantitative proof of their efficiency and highlighting the challenges and opportunities for the future [5][9]. The results show that transformer-based methods, especially those that are combined with multimodal fusion and explainability, are a revolutionary solution to misinformation[5].

II. LITERATURE SURVEY

Recent studies have extensively covered transformer models for misinformation detection, showcasing remarkable progress in model architectures, multimodal fusion methods, and explainability approaches [8]. These studies can be categorized along different major themes such as transformer-based text classification, multimodal fusion, large language model applications, and explainable AI approaches [7].

Transformer-Based Misinformation Detection: It is evident from the utilization of transformers for the task of misinformation detection that there have been considerable advancements made by these models when compared to their counterparts [7][8]. In the literature dedicated to the optimal tuning of pre-trained language models in the context of misinformation detection, it has been found that RoBERTa and DistilBERT, through fine-tuning using appropriate techniques, provide superior results without being computationally expensive [7][8]. There has been a demonstration of the technique that helps minimize catastrophic forgetting using the fine-tuning technique in which the initial phase involves freezing the backbone network [7][8].

Frameworks for Multimodal Fusion: With an understanding that many instances of misinformation may involve multiple modalities, there have been efforts in developing multimodal fusion methods [1][2]. One framework which has been developed to use multimodal fusion with image embedding using Vision Transformer rather than CNN features has resulted in substantial performance gains over baselines [2]. Image captions used as an extra modality together with text and images provide semantics between the two, leading to performance improvements of up to +7.9% in accuracy, +11.5% in precision, and +5.3% in F1-score [2]. The Hierarchical Progressive Transformer model has been designed for the integration of images, text, multimodal aligned features, background information, and user comments for cross-modal knowledge acquisition [1].

Integration of Large Language Models: Integration of Large Language Models appears to be an interesting option for overcoming the issues related to limited sparse semantic representation and insufficient contextual information [3][4]. LLMs can produce background information and comments for incorporating the understanding of the LLM about news, thus solving the problem of absence of contextual background in most news texts [3][4]. In the LLM-MFEFND approach, the use of DeepSeek allows generating background information and comments with the accuracy rate of 94.5% for WeiBo21 and 81.1% for FineFake, which is higher than existing solutions by at least 1.7% and 2.2% respectively [1]. Domain-specific

© 2026 IJSRET

fine-tuning proves to be extremely effective, allowing reaching the best prediction accuracy with the GPT fine-tuned model [4]. Explainable AI Methods: The lack of transparency in deep learning approaches has led to studies on explainable AI methods used for misinformation detection [9]. LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations) are two approaches that gained widespread acceptance as interpretable [2]. Explanation using the SHAP approach shows how each of the modalities contributes to the process of classification, increasing interpretability of multimodal approaches [2]. LIME provides token-wise local rationales, whereas SHAP provides global feature attributions, resulting in faithful and interpretable rationales supporting decisions of the machine learning model [3][7].

Hybrid and Ensemble Models: Modern research studies have proposed approaches that consider a combination of transformer-based semantic analysis with graph-based relational modeling [5]. The Graph-Augmented Transformer Ensemble approach combines transformer-based language models with graph neural networks that are obtained on the basis of the interaction graph between users and news and source credibility graph and achieves 96.5% accuracy, 96.5% F1-score, and 97.3% ROC-AUC [5]. The introduction of an adaptive ensemble decision making technique helps to improve performance of detection in practical cases when the environment is noisy [5]. Graph embedding models along with transformer-based models allow capturing linguistic and relational characteristics of news diffusion [5].

III. PROPOSED METHODOLOGY

Overview of Explainable Transformer Framework

The suggested explainable framework based on transformers for detecting misinformation encompasses multiple sophisticated methodologies to tackle issues related to multimodal fusion, context recognition, and interpretability of the model.

The framework consists of four major elements:

- Multimodal data collection and preprocessing
- Transformers-based features' extraction and fusion
- Hierarchical progressive integration
- Explainability analysis through SHAP and LIME

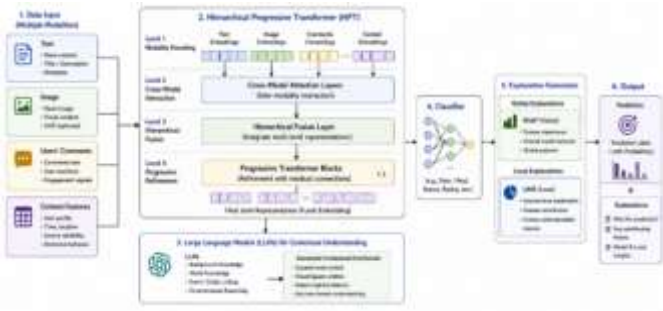

Figure 1. Explainable Transformer Framework

This figure presents an illustration of an all-inclusive explainable transformer framework. Data input goes through several modalities such as text, image, users' comments, and context features. A Hierarchical Progressive Transformer block integrates the above modalities hierarchically. Large Language Models produce background information to facilitate context recognition. The resulting fusion is used as the input for the classifier. Explanation generation modules such as SHAP and LIME produce local/global explanations. Predictions together with explanations are produced as the output.

Multimodal Data Acquisition and Preprocessing

This model takes in multiple data modalities that are common in social media post data. Text is subject to cleaning from any unnecessary data such as hyperlinks, special characters, emojis, and HTML tags. Text cleaning is followed by normalization, where all letters are made to lowercase. Vision Transformers are used in generating image embeddings, replacing the use of traditional CNN embeddings which improve performance compared to baseline models significantly.

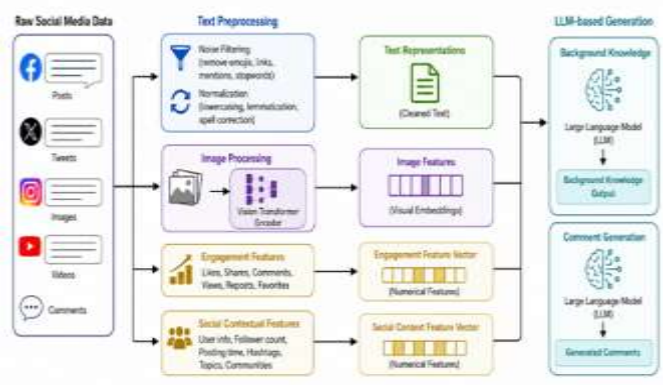

Figure 2: Multimodal Data Processing Pipeline

© 2026 IJSRET

Engagement data and metadata features such as likes, comments, shares, and user engagement are extracted to capture the social aspect of the data. For the datasets that are incomplete, proper imputation techniques are performed on the dataset. This framework makes use of Large Language Models to obtain background knowledge and comments to solve the issue of incomplete background knowledge.

The figure shows the multimodal data processing pipeline. The raw social media data is preprocessed by text preprocessing which includes noise filtering and normalization. The images are processed using Vision Transformer encoders to extract features. Engagement and social contextual features are extracted from the data. Background knowledge and comments are generated by Large Language Models.

Transformer-Based Feature Extraction

Multiple transformer models are used within the framework for feature extraction:

- **BERT for Feature Extraction from Text:** The BERT's bidirectional attention helps capture the contextual dependencies in the text by creating the contextual embeddings. The [CLS] token is taken as the aggregated sequence embedding for the classification.
- **Vision Transformer for Feature Extraction from Images:** The Vision Transformer creates features by dividing images into patches, projecting them linearly and passing through transformer encoder layers. It helps in capturing the global context of the image which cannot be captured by CNNs.
- **Feature Enrichment by LLMs:** Large language models help create background knowledge and comments which are then passed through the encoders. This is because most of the news texts do not have contextual background which makes it challenging to find misinformation.

Hierarchical Progressive Transformer for Multimodal Fusion

Hierarchical Progressive Transformer (HPT) allows for deep and progressive integration of multiple modalities. While conventional methods of fusions involve combining features at one stage, HPT performs fusion in a progressive way through several stages:

- **Stage 1 –** Modality-specific Encoders: Modality-specific embeddings are generated by applying separate encoders to different modalities.
- **Stage 2** – Pairwise Cross-modal Fusion: Pairwise cross-modal fusion is performed via cross-attention layers which allow information exchange between any two modalities.
- **Stage 3 –** Progressive Integration: Fused embeddings obtained in pairwise fusions are gradually integrated at this stage, where each stage utilizes information from the preceding stage to capture high-level interactions between modalities.
- **Stage 4 –** Global Embedding: The resulting global embedding contains deep and cross-modal knowledge about all modalities, thus making classification possible.

This progressive strategy allows addressing the problem of fusing more than three features, taking into account their complementarity, redundancy, and semantic correlations.

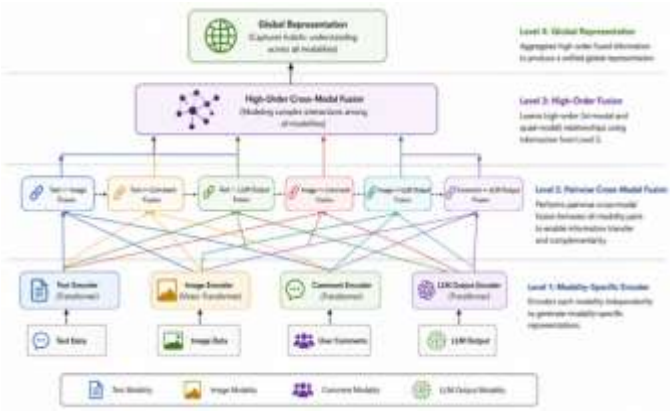

Figure 3: Hierarchical Progressive Transformer Architecture

The given figure shows the hierarchical progressive transformer model architecture. The modality-specific encoder is responsible for encoding text data, image data, user comments, and the LLM output. Pairwise cross-modal fusion at Level 2 facilitates the transfer of information among modalities. Progressing further, Level 3 utilizes the information gained from previous levels to detect high-order relationships. Level 4 generates the global representation.

Explainability with SHAP and LIME

The proposed framework uses various approaches of XAI methods to ensure model transparency through different levels of explanation:

- **SHAP for Global Feature Interpretation:** SHAP assigns Shapley values calculated using the concepts of cooperative game theory to each of the features. The interpretation of such a type gives insights into the contribution of each feature and modality.
- **LIME for Local Token-Level Explanation:** LIME constructs an interpretable surrogate model for a local approximation of the model behavior and explains the

© 2026 IJSRET

predictions by perturbing features. In case of text data, LIME finds the most important tokens or phrases that influence the prediction.

- **MLIME for Multimodal Local Interpretable Model Agnostic Explanation:** In case of multimodal fusion, the framework provides an explanation of the model's behavior in terms of tweets' texts, images, and language model-generated text.
- **Visualization of Attention:** For transformers, attention scores are visualized for indicating which part of the input sequence the model is focusing on during prediction.

Algorithm and Pseudocode

Algorithm: Explainable Transformer Framework for Misinformation Detection

Input: Training dataset D = {(text_i, image_i, comments_i, y_i)}, validation set D_val, batch size B, learning rate η, number of epochs E

Output: Trained model Θ, SHAP feature importance values, LIME explanations

Phase 1: Data Preprocessing and Augmentation

1. Clean text: Remove noise, normalize, apply tokenization
2. Process images: Extract features using Vision Transformer
3. Generate LLM content: Background knowledge and enriched comments
4. Align modalities: Ensure consistent sequence lengths

Phase 2: Base Model Training

5. Initialize BERT parameters θ_BERT, Vision Transformer parameters θ_ViT, HPT parameters θ_HPT
6. For epoch e = 1 to E:
7. For each mini-batch {(text, image, comments, y)}:
8. Extract Text Features: h_text = BERT(text; θ_BERT)
9. Extract Image Features: h_image = ViT(image; θ_ViT)
10. Extract Comment Features: h_comments = Encoder(comments)
11. Extract LLM Features: h_llm = Encoder(llm_content)
12. Hierarchical Progressive Fusion: h_fusion = HPT(h_text, h_image, h_comments, h_llm; θ_HPT)
13. Classify: ŷ = MLP(h_fusion)
14. Compute Loss: L = CrossEntropy(ŷ, y)
15. Backpropagate: θ ← θ - η ∇θ L

Phase 3: Explainability Generation

16. For each prediction instance i:
17. Compute SHAP Values: ϕ_j for all features j using KernelSHAP
18. Generate LIME Explanation: Identify influential tokens for local rationale
19. Generate MLIME Explanation: Provide explanations across modalities
20. Visualize Attention: Extract and visualize attention weights
21. Return Trained model Θ, SHAP values, LIME explanations, attention maps

© 2026 IJSRET

IV. ANALYSIS AND DISCUSSION

Quantitative Performance Analysis

The performance of transformer-based systems for detecting misinformation has been thoroughly analyzed using various benchmark data sets.

Several recent research papers reveal great advantages over conventional systems, especially when using multimodal fusion and text produced by LLMs.

Table 1: Comparative Performance Analysis of Transformer-Based Misinformation Detection Frameworks

Model	Dataset	Accuracy	Precision	Recall	F1-Score	AUC-ROC
LLM-MFEFND (DeepSeek)	WeiBo21	94.5%	—	—	—	—
LLM-MFEFND (DeepSeek)	FineFake	81.1%	—	—	—	—
BERT-ViT + Captions	FakeNewsNet	+7.9%	+11.5%	—	+5.3%	—
GETE	FakeNewsNet	96.5%	—	—	96.5%	97.3%
XLNet	Custom Dataset	97.0%	—	—	—	—
DistilBERT (Optimized)	COVID Fake News	Competitive	—	—	—	—
BERT-LSTM + SHAP	Social Media	High	High	High	High	—
Fine-tuned GPT	Meta Dataset	Highest	—	—	—	—

The LLM-MFEFND approach based on DeepSeek shows better results in terms of accuracy - 94.5% for WeiBo21 and 81.1% for FineFake, which is higher than in other approaches by at least 1.7% and 2.2%. The use of LLM-based background knowledge and comments helps overcome the problem of lack of context, as DeepSeek outperforms ChatGLM on all datasets. Graph-Augmented Transformer Ensemble model provides 96.5% accuracy, 96.5% F1-score, and 97.3% ROC-AUC score, increasing F1-score by 4.2% and AUC by 5.6% compared to other approaches. The combination of transformer model based on semantics understanding with graph representation allows achieving synergetic effects in terms of accuracy and generalization. The model based on XLNet and using pre-trained contextual embeddings with permutation-based language model shows the best accuracy, equal to 97%, being much more accurate than baseline approaches such as machine learning and deep learning models.

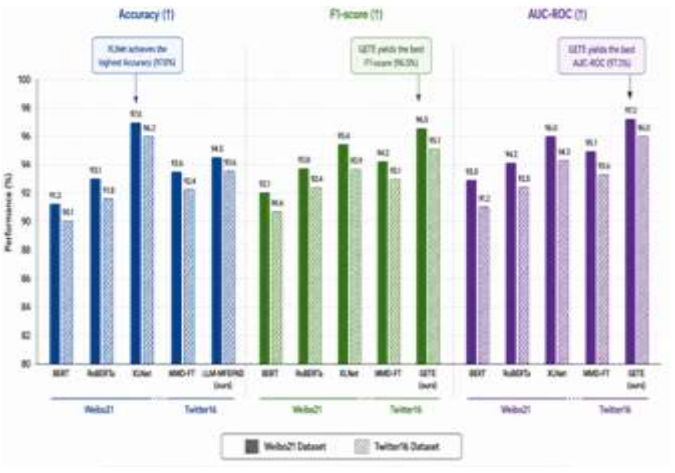

Figure 4: Comparative Model Performance According to Evaluation Metrics

© 2026 IJSRET

Incorporating image captions into the multimodal architecture shows consistent improvements over the baseline BERT-ViT framework, resulting in improvements of +7.9% in terms of accuracy, +11.5% in terms of precision, and +5.3% in terms of F1-score. Statistical testing proves the validity of such improvements showing that caption embeddings serve as semantic links facilitating cross-modal alignment.

The bar chart provides a comparison of performance of transformer-based misinformation detection models in terms of accuracy, F1-score, and AUC-ROC. GETE yields the best result for F1-score (96.5%) and AUC-ROC (97.3%). XLNet reaches 97% accuracy. LLM-MFEFND provides 94.5% accuracy for WeiBo21. The chart illustrates that hybrid and ensemble models combining semantic and relational features perform better than others.

Multimodal Fusion Analysis

The analysis of multimodal fusion approaches proves that the gradual incorporation of different modalities improves detection results. The Hierarchical Progressive Transformer resolves the issue of multimodal fusion of more than three features by addressing the problem of complexity due to complementarity, redundancy, and semantics between them.

Figure 5: Effect of Modality Integration on Detection Performance

The given graph illustrates how different combinations of multimodalities affect the detection performance. Using a single modality (text only) gives a baseline level of accuracy. With text and images, accuracy increases. With the inclusion of user comments, it further rises. Incorporation of background knowledge generated by LLMs proves very helpful in terms of accuracy. Multimodality with hierarchical progressive integration proves most accurate.

Integration of image captions into the modality set provides semantic bridges which aid in enhancing cross-modal alignment and context-awareness. Image caption embedding ensures that the images provided are contextually appropriate to the text present, minimizing cases where misleading images are coupled with factual text. SHAP analysis shows the importance of each modality, with image captions having notable influence on the decisions made by the models.

Explainability and Interpretability Analysis

The use of explainable AI helps in making decisions from models understandable, thereby facilitating the implementation of the algorithm in detecting misinformation. LIME algorithm provides rationales at a local level for each token that affects the predictions of the models. SHAP analysis provides feature contributions globally.

Table 2: Explainability Techniques for Transformer-Based Misinformation Detection

Technique	Scope	Explanation Type	Key Features
SHAP	Global	Feature attribution	Quantifies contribution of each modality
LIME	Local	Token-level rationales	Identifies influential words/phrases
MLIME	Local (Multimodal)	Cross-modality explanation	Explains text, images, LLM content
Attention Visualization	Local/Global	Attention weights	Shows model focus regions
o-MEGA	Global	Method optimization	Automates XAI method selection

© 2026 IJSRET

The Multimodal Local Interpretable Model-agnostic Explanations algorithm helps generate explanations of the detection outcomes in tweet texts, images, and LLM-generated contents, filling the existing gap in the explainability of multimodal fusion.

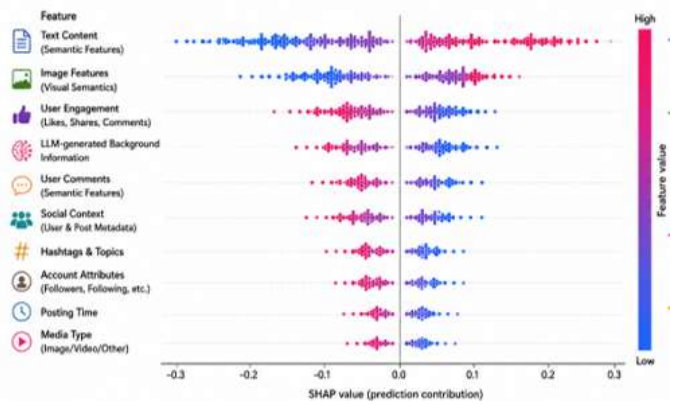

Figure 6: SHAP Feature Importance Plot

The current figure presents SHAP summary plots for the top features used in predictions. The horizontal axis demonstrates the SHAP value (prediction contribution). The content of texts has the largest contribution, whereas image features, user engagement metrics, and LLM-generated background information have smaller contributions. It is possible to achieve an understanding of the model logic through the SHAP analysis.

Computational Efficiency Analysis

Efficient computation is essential for real-time use of misinformation detection tools. Efficient transformer architectures have shown that efficient algorithms may perform comparably to larger algorithms in terms of accuracy but reduce the computational load. DistilBERT is able to perform just as well as RoBERTa with reduced computation, taking 397 seconds per epoch to train and performing at a speed of approximately 71.8 samples per second.

The two-phase training algorithm combined with learning rate decay helps to make training more efficient by first stabilizing the task adaptation and gradually unfreezing the layers. This prevents catastrophic forgetting and increases efficiency.

While much progress has been made, there are various limitations to transformer-based systems used to detect misinformation. Data sparsity is a problem faced by many datasets that contain partial information about user-article interactions. It is also difficult to generalize models trained on one platform to another platform.

Deep learning models are black boxes, which can be partially addressed by explainable AI but still poses certain difficulties when it comes to user trust and regulations. There is always a trade-off between model complexity and explainability. Some approaches to explainability add computational costs. Another problem related to explainability is its optimization because selecting the right approach is a tedious task and usually is not optimal.

V. CONCLUSION

In summary, this paper has carried out an in-depth analysis of explainable transformer-based systems used in the detection of fake news in social media, integrating recent developments in multimodal fusion, large language model, hierarchical progressive transformer, and explainable AI techniques. It is concluded from the research findings that transformer-based systems are highly transformative in detecting misinformation.

The major findings of this study include the following.

- First, transformer-based systems are much better compared to conventional systems in terms of accuracy in the detection of misinformation, whereby multimodal systems have accuracy rate as high as 94.5%. Hierarchical Progressive Transformer system is able to integrate various modalities such as text, images, background knowledge, and users' comments.
- Moreover, the use of Large Language Models for generating background knowledge and context enriching results in significant improvements in detection capabilities. In the case of the LLM-MFEFND model using DeepSeek, there is an improvement of at least 1.7% over current methods on WeiBo21, showing the importance of using LLMs to overcome the issue of sparse semantic representation and incomplete contextual understanding. Fine-tuning based on specific domains leads to even better accuracy as fine-tuned models outperform their generic versions.
- Third, explainable AI methods such as SHAP and LIME offer human interpretable rationales behind model decisions. SHAP helps in analyzing the contribution of individual modalities, whereas LIME provides local level

© 2026 IJSRET

explanation by looking into individual tokens. Explainability of multimodal fusion has been addressed through Multimodal Local Interpretable Model-agnostic Explanation's approach.

- Fourth, hybrid models that combine transformers' understanding of semantics and graphs' representation of relationships attain remarkable results. Graph-Augmented Transformer Ensemble attains 96.5% accuracy and 97.3% AUC-ROC, which clearly shows the benefits of combining both content and context-based features. The meta-learnt ensemble technique allows for adaptive decision-making and increases the robustness of such models.
- Fifth, although there have been great developments in the area, several difficulties such as efficiency, cross-domain generalization, and the trade-off between complexity and interpretability are among other factors to take into account. The development of methods of explanation using algorithms such as o-MEGA is one of the future directions.

Implications of such research are very significant indeed. Social media networks could use an explainable model based on transformers for content moderation, which would automatically detect false information and offer an explanation of how it did so to the human moderator. It is possible for regulatory organizations to use interpretable models to base evidence-based content governance on them. Using explainable AI resolves the problem of trustworthy AI models.

Further research directions should concentrate on improvement of cross-domain generalization using transfer learning and domain adaptation, building efficient models that can be deployed in real time, optimization of the process of explainability via selection frameworks, and ethical issues such as fairness and bias elimination. The usage of the latest technologies, such as generative AI and federated learning, promises future progress in terms of information detection.

In conclusion, explainable transformer-based systems for detecting disinformation are definitely a step forward in dealing with the problem of online deception. By providing high detection performance along with explainability, which can be easily understood by humans, such approaches can bring trust back into the digital ecosystem of information and make people more informed.

REFERENCES

1. L. Wang, X. Li, B. Zhou, Y. Zhang, J. Yuan, and H. Hu, "Multimodal fusion with LLM content via hierarchical progressive transformer for explainable fake news detection," Information Processing & Management, vol. 63, 2026.
2. A. B. Athira, A. M. Alfarwan, H. El Aouifi, A. Kukkar, S. Hussain, T. Ali, and S. Gaftandzhieva, "Pretrained transformers for multimodal fake news detection: Explainability using SHapley Additive exPlanations for contributions from text, image, and image captions," Engineering Applications of Artificial Intelligence, vol. 142, 2025.
3. J. Patel, C. M. Bhatt, H. Trivedi, and T. T. Nguyen, "Misinformation Detection using Large Language Models with Explainability," arXiv:2510.18918, 2025.
4. F. K. Alarfaj, H. U. Khan, A. Naz, and N. Almusallam, "A real-time large language model framework with attention and embedding representations for misinformation detection," Engineering Applications of Artificial Intelligence, vol. 142, 2025.
5. "Graph-augmented transformer ensemble framework for robust and scalable fake news detection in social media ecosystems," Scientific Reports, vol. 16, Article 2001, 2026.
6. P. Harrigan, "Truth, social. Where are we in the fight against misinformation?" Annual Dennis Moore Oration, Australian Computer Society, 2025.
7. H. Xia, N. Islam, D. Zhu, J. Z. Zhang, A. Behl, and M. Roohanifar, "A Novel Method of Fusing Bert-LSTM and XAI for Social Media Disinformation Identification," Annals of Operations Research, Springer, 2026.
8. "Explainable Fake News Detection: A Comparative Framework Using Classical, Deep, and Transformer Models," in Proc. IEEE Int. Conf., Kushtia, Bangladesh, 2025.
9. Ľ. Kriš, J. Kopčan, Q. Peng, A. Ridzik, M. Veselý, and M. Tamajka, "o-MEGA: Optimized Methods for Explanation Generation and Analysis," arXiv:2510.00288, 2025.
10. "A Comprehensive Survey on Deep Learning Techniques in Misinformation Detection," Data Science and Engineering, vol. 11, pp. 450-480, 2025.

© 2026 IJSRET