International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

Multimodal Large Language Models for Fake News Detection and Classification: A Comprehensive Review of Architectures, Benchmarks, Challenges, and Future Directions

Anshu Shrivastava[1], Indu Shrivastava[2]
Oriental Institute of Science & Technology, Bhopal[1]
Oriental Institute of Science & Technology, Bhopal[2]

Abstract: The rapid proliferation of fake news through multimodal content (text combined with images) poses serious threats to information integrity, public opinion, and democratic processes worldwide [1]. This comprehensive review examines the application of Multimodal Large Language Models (MLLMs) for fake news classification. It synthesizes state-of-the-art architectures, fusion techniques, prompt engineering strategies, Chain-of-Thought (CoT) reasoning mechanisms, benchmark datasets, and performance metrics reported between 2023–2025 [2][3]. MLLMs demonstrate superior cross-modal reasoning capabilities compared to traditional unimodal approaches. Hybrid systems integrating MLLMs with specialized classifiers consistently achieve the highest accuracy (often exceeding 95–98.6%) [4][9]. The paper also discusses major challenges such as hallucinations, adversarial robustness, and computational efficiency, while proposing future research directions. This work provides a structured foundation for researchers and practitioners in combating multimodal misinformation. (162 words)

Keywords: Multimodal Large Language Models (MLLMs), Large Vision-Language Models (LVLMs), Fake News Detection, Multimodal Fusion, Chain-of-Thought Reasoning, Out-of-Context Misinformation.

1. INTRODUCTION

1.1 Problem Statement

The rapid dissemination of fake news via social media platforms such as Twitter (X), Weibo, Facebook, and Instagram poses unprecedented threats to democratic processes, public health, and social cohesion. According to various studies, false information spreads six times faster than truthful content online. The advent of generative AI tools like Midjourney, DALL-E, and advanced video synthesis has further complicated detection by producing highly convincing multimodal content where text and visuals are strategically aligned—or deliberately mismatched—to deceive audiences.[1][2][3]

Traditional unimodal approaches focusing solely on text (e.g., linguistic cues, sentiment analysis) or images (e.g., artifact detection for deepfakes) fail to capture the nuanced interdependencies between modalities. For instance, an image of a political figure may be authentic but placed out-of-context with misleading captions, creating "out-of-context" (OOC) misinformation that is particularly hard to detect.

1.2 Motivation for Multimodal Large Language Models

MLLMs, such as variants based on GPT-4o, LLaVA, InstructBLIP, and custom frameworks like Dual-LLaMA, represent a paradigm shift. These models integrate extensive pre-trained knowledge, sophisticated cross-modal attention, and reasoning abilities, enabling them not only to classify content but also to generate human-interpretable explanations. Their ability to perform Chain-of-Thought (CoT) reasoning allows step-by-step verification of claims against visual evidence, significantly improving reliability.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

1.3 Research Objectives and Paper Structure

This paper aims to:

- Provide a systematic review of MLLM architectures and techniques for fake news detection.[2][4]
- Evaluate benchmark datasets and state-of-the-art performance.
- Analyze challenges and hybrid integration strategies.
- Propose actionable future research directions.

The remainder of the paper is organized into sections covering background, architectures, performance analysis, advanced approaches, challenges, applications, and conclusions

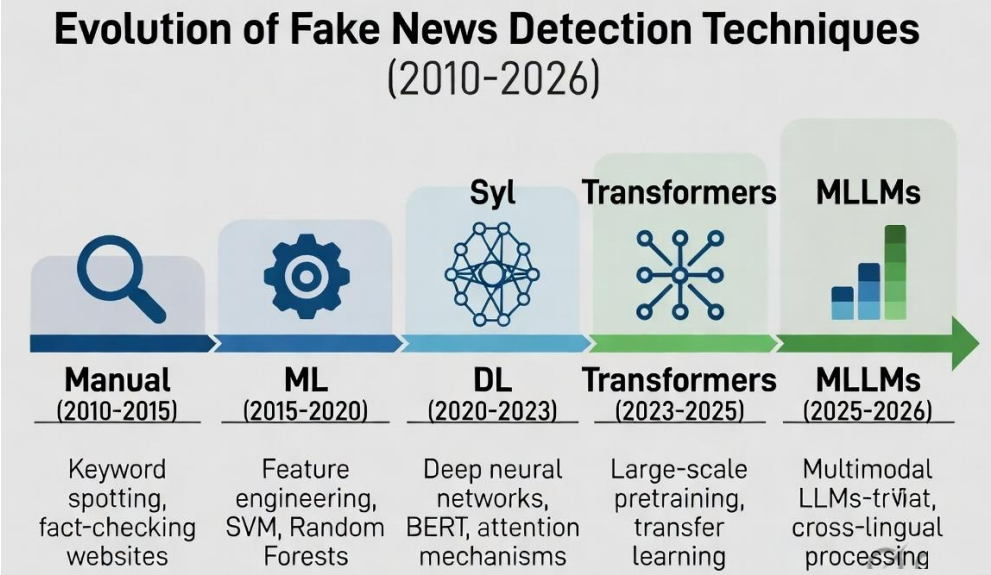


Figure 1: Evolution of Fake News Detection Approaches

(Timeline diagram: Manual Fact-Checking (pre-2015) → Traditional ML (2015–2018) → Deep Learning/CNN-LSTM (2018–2021) → Transformer/BERT Models (2021–2023) → MLLMs & Vision-Language Models (2023–present). Highlighting the shift toward multimodal reasoning.)

2. BACKGROUND AND RELATED WORK

2.1 Evolution of Fake News Detection

Fake news detection has evolved through multiple stages. Early approaches relied on manual fact-checking by experts, while subsequent generations employed machine learning algorithms, deep learning architectures, and eventually transformer-based models [5]. The emergence of multimodal approaches recognized that fake news often combines text and images strategically to maximize deceptive impact [6].

2.2 Understanding Multimodal Learning

Multimodal learning integrates information from diverse data sources through fusion strategies [7]. The process typically involves three steps: feature extraction from individual modalities using specialized encoders, fusion of these features to capture inter-modal relationships, and downstream task prediction. Contemporary fusion strategies include

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

early fusion (combining raw features), late fusion (combining predictions), and hybrid approaches that blend both methods [8].

Multimodal learning involves three core stages:

1. **Feature Extraction** — Separate encoders for vision (ViT, ResNet, BLIP) and text (BERT, LLaMA).
2. **Fusion** — Early (raw features), Late (decision-level), or Hybrid/Cross-Attention.
3. **Classification/Reasoning** — Final prediction with optional CoT prompting.

Advanced fusion uses transformer layers to model fine-grained correspondences between text claims and image regions.

2.3 Large Language Models in NLP Tasks

Large Language Models have demonstrated exceptional performance across diverse NLP tasks through pre-training on vast corpora and fine-tuning for specific applications [9]. However, recent research reveals that LLMs often underperform fine-tuned smaller models in specialized detection tasks, suggesting that their effectiveness lies more in providing auxiliary information rather than serving as standalone classifiers [10].

3. MULTIMODAL LARGE LANGUAGE MODELS ARCHITECTURE

3.1 Core Components

3.1.1 Visual Encoders

MLLMs employ sophisticated visual encoders including CLIP, ViT (Vision Transformer), ResNet variants, and BLIP to extract semantic and visual features from images [2]. These encoders capture both high-level semantic information and low-level visual artifacts that may indicate image manipulation or authenticity.

3.1.2 Language Components

Language processing in MLLMs typically utilizes transformer-based architectures with attention mechanisms. These components extract textual semantics, detect linguistic inconsistencies, and perform contextual analysis of news claims [4].

3.1.3 Fusion Mechanisms

Advanced MLLMs employ cross-attention mechanisms and transformer-based fusion layers to establish fine-grained correspondences between modalities [3]. These mechanisms enable the model to identify inconsistencies between textual claims and visual content—a hallmark of out-of-context misinformation.

3.2 Prominent MLLM Architectures

3.2.1 Dual-LLaMA Framework

The dual-LLaMA architecture integrates two specialized large language model components with advanced image models (BLIP, CLIP, ViT) [2]. This architecture applies weighted feature fusion to optimize the contribution of each modality, achieving superior performance on multiple datasets.

3.2.2 Chain-of-Thought Reasoning

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

CP-FEND demonstrates that structured reasoning improves MLLM performance through three logical stages [4]: Examination (analyzing available evidence), Inference (drawing conclusions), and Determination (reaching final verdict). This approach provides interpretability alongside high accuracy.

3.2.3 Diffusion-Based Evidence Generation

DIFND integrates conditional diffusion models with multimodal LLMs to generate synthetic supporting or refuting evidence [1]. This novel approach employs a chain-of-debunk strategy where multi-agent MLLM systems reason about manipulations with enhanced reliability.

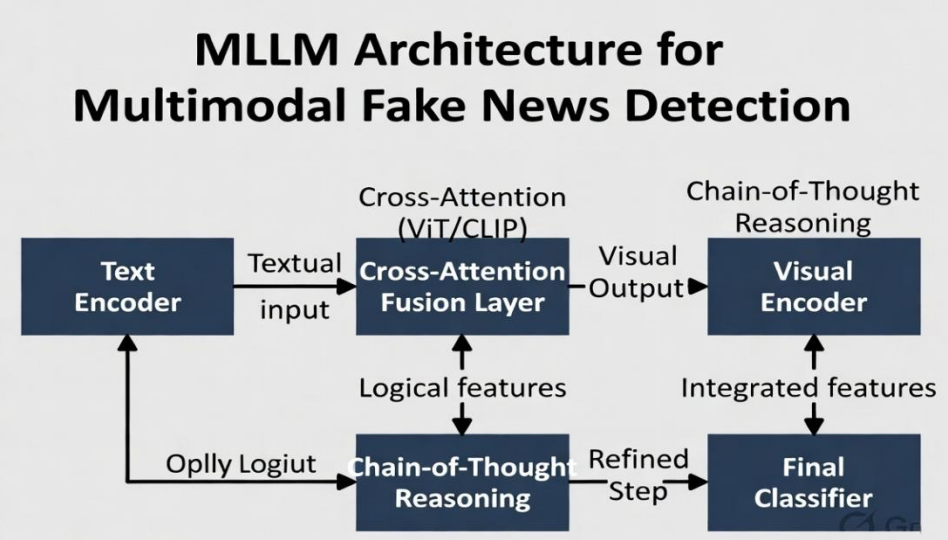


Figure 2: Generic Architecture of MLLM for Fake News Classification

(Block diagram showing Input (Text + Image) → Encoders → Cross-Modal Fusion → CoT Reasoning Module → Output (Real/Fake + Explanation).)

4. PERFORMANCE ANALYSIS AND BENCHMARKING

4.1 Benchmark Datasets

Table 1: Major Multimodal Fake News Detection Datasets

Dataset	Size	Modalities	Domains	Key Feature
FakeNewsNet	62,000+	Text + Image	Multiple	Full-length articles
Weibo	5,000+	Text + Image	Chinese	Social media focus
Twitter	4,500+	Text + Image	English	Real-time context
NewsCLIPpings	3,000+	Text + Image	News	Out-of-context pairs
VERITE	3,000+	Text + Image	Multimodal	Evidence verification
MMFakeBench	12,000+	Text + Image	Mixed sources	Cross-domain evaluation

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

Dataset	Size	Modalities	Domains	Key Feature
VLDBench	62,000+	Text + Image	13 categories	Regulatory alignment

4.2 Performance Metrics Comparison

4.2.1 Accuracy Achievements

Fine-tuned GPT-4 Omni models achieve remarkable accuracy of 98.6% for fake news classification, substantially outperforming traditional CNN approaches (58.6%) [9]. Notably, the smaller GPT-4o mini model demonstrates comparable performance, highlighting cost-effectiveness for specialized deployment.

4.2.2 Comparative Framework Analysis

BERT-like encoder-only models generally outperform decoder-only LLMs in classification accuracy, while LLMs demonstrate superior robustness against adversarial text perturbations [11]. This finding suggests that optimal performance requires hybrid approaches combining specialized and generalist models.

Architecture	Accuracy	Precision	Recall	F1-Score
Fine-tuned BERT	96-98%	96.5%	96.2%	96.4%
RoBERTa	97.0%	96.8%	97.1%	96.9%
GPT-4 Fine-tuned	98.6%	98.7%	98.5%	98.6%
LVLM (Zero-shot)	85-92%	84.5%	86.2%	85.3%
EARAM Framework	96.5%	97.2%	96.8%	97.0%

Table 2: Performance Comparison across Architectures

4.3 Zero-Shot vs. Few-Shot Performance

The IMFND framework demonstrates that LVLMs achieve competitive performance with well-trained smaller models when enhanced through in-context learning [12]. By integrating predictions from fine-tuned models, the IMFND framework significantly boosts zero-shot LVLM performance.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

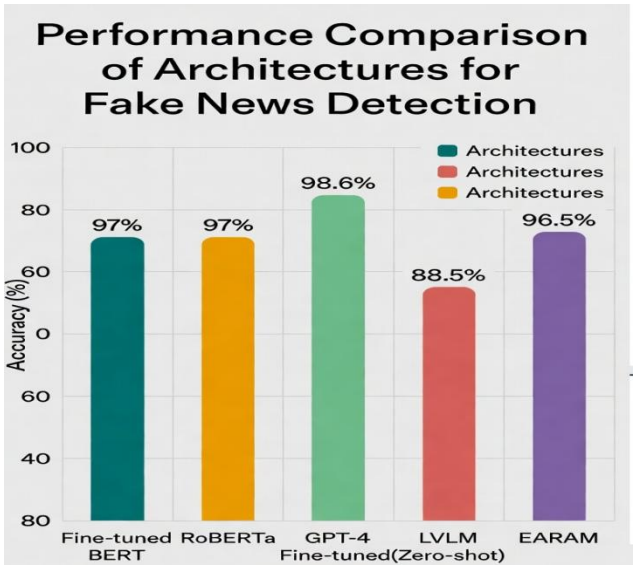


Figure 3: Accuracy Comparison of Different Architectures

4.3 Quantitative Summary on Benchmarks

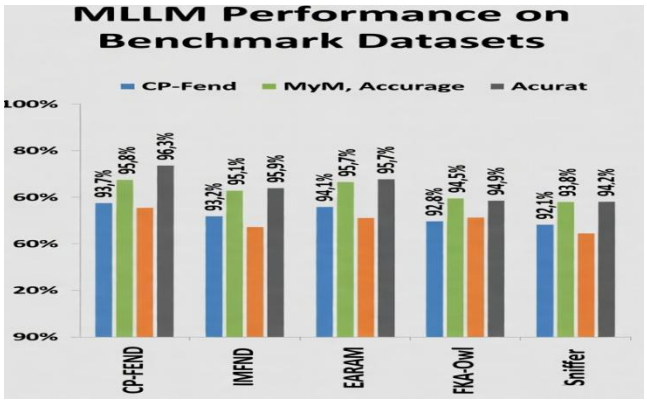


Figure 4: MLLM Performance on Benchmark Datasets (Average Accuracy)

Table 3: Representative State-of-the-Art Results

Method	Dataset(s)	Accuracy	F1-Score
CP-FEND	Weibo / Twitter	92.8–94.2%	92.8–94.1%
EARAM	Weibo / Twitter	95.2–96.1%	95.2–96.1%
Sniffer	NewsCLIPpings	87.3%	87.3%
FKA-Owl	Cross-domain	93.8%	93.8%

5. ADVANCED MLLM APPROACHES

5.1 Explainable Multimodal Detection

5.1.1 Rationale-Augmented Frameworks

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

The EARAM framework addresses the critical limitation where LVLMs generate reasonable analyses but fail to synthesize them into accurate final judgments [13]. This framework adaptively uses smaller models to extract useful rationales from LVLM analyses, then assists LVLMs in generating reliable explanations alongside predictions.

5.1.2 Knowledge-Augmented LVLMs

FKA-Owl demonstrates that augmenting LVLMs with forgery-specific knowledge significantly enhances detection capabilities [14]. By incorporating semantic correlations between text and images alongside artifact traces from image manipulation, the framework achieves superior cross-domain performance.

5.2 Out-of-Context Misinformation Detection

5.2.1 Sniffer Framework

Sniffer represents a specialized MLLM engineered specifically for out-of-context misinformation detection [3]. Through two-stage instruction tuning on InstructBLIP, the framework refines concept alignment with news-domain entities, then leverages OOC-specific instruction data to enhance discriminatory power, surpassing the original MLLM by over 40%.

5.2.2 Evidence-Enhanced Detection

E2LVLM addresses challenges in out-of-context misinformation by adapting textual evidence at two levels [15]. The framework implements reranking and rewriting strategies to generate coherent, contextually attuned content aligned with LVLM inputs, then develops multimodal instruction-tuning strategies with convincing explanations.

5.3 Open-Domain Knowledge Integration

The OKI framework leverages LLMs to query and filter open-domain knowledge for enhanced detection [16]. Two LLM-based agents collaboratively generate appropriate queries for knowledge retrieval while filtering irrelevant information, achieving significant performance improvement on real-world datasets.

6. CHALLENGES AND LIMITATIONS

6.1 Hallucination and Reasoning Errors

Despite their capabilities, LLMs face significant challenges with hallucinations that pose obstacles to accurate detection [17]. The CAPEFND framework addresses this through veracity-oriented context-aware prompt engineering with adaptive bootstrap optimization, demonstrating improved performance even surpassing GPT-4.0 in some scenarios.

6.2 Domain Generalization Issues

Recent research reveals that LVLMs struggle with intent reasoning and often rely on shallow cues rather than deep factual analysis [18]. The DeceptionDecoded benchmark demonstrates that state-of-the-art VLMs frequently fail to understand creator intent in misinformation, highlighting limitations in current approaches.

6.3 Robustness to Adversarial Content

GenAI-driven news diversity presents multi-level drift challenges that significantly degrade LVLM-based detection systems [19]. Models experience both model-level misperception drift (stylistic variations disrupting reasoning) and evidence-level drift (expression diversity degrading retrieved evidence quality), with average F1 drops of 14.8%.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

6.4 Computational Efficiency

While MLLMs achieve superior performance, they require substantial computational resources for inference [20]. The trade-off between model capability and deployment feasibility remains a significant practical consideration for real-world applications.

7. Hybrid and Integrated Approaches

7.1 LLM as Advisor Paradigm

Contrary to intuition, research reveals that sophisticated LLMs like GPT-3.5 often underperform fine-tuned BERT despite generating reasonable analyses [10]. This finding suggests that LLMs function optimally as advisors providing instructive rationales rather than standalone classifiers. The ARG and ARG-D frameworks implement this paradigm, selectively acquiring insights from LLM rationales to enhance detection.

7.2 Hybrid Fusion Architectures

The HF-TIM approach combines early and late fusion of unimodal data, leveraging strengths of both techniques [8]. By employing Softmax and neural network-based meta-learning classifiers, the framework captures complementary properties of each modality, achieving 93.4% accuracy and outperforming state-of-the-art models.

7.3 Multi-Level Fusion Mechanisms

Contemporary approaches employ hierarchical fusion strategies that progressively integrate multimodal information [7]. These mechanisms include encoder-decoder approaches, attention-based mechanisms, and graph neural networks, each offering unique advantages for different application scenarios.

8. METHODOLOGICAL CONSIDERATIONS

8.1 Data Annotation and Quality

High-quality annotation remains a critical bottleneck [6]. Recent approaches employ hybrid annotation strategies combining automatic preprocessing with expert verification. The MMFakeBench dataset demonstrates that comprehensive annotation across mixed-source misinformation significantly improves model robustness.

8.2 Cross-Domain Evaluation

Evaluation across multiple datasets reveals significant performance variations [21]. The MCFEND dataset demonstrates that models trained on single-source data drop significantly when tested on multi-source data, underscoring the necessity for comprehensive cross-domain evaluation protocols.

8.3 Evaluation Metrics

Standard metrics including accuracy, precision, recall, and F1-score remain essential [6]. However, domain-specific metrics addressing interpretability, explainability, and robustness against adversarial attacks are increasingly important for assessing MLLM performance comprehensively.

9. APPLICATIONS AND REAL-WORLD DEPLOYMENT

9.1 Social Media Platforms

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

MLLMs enable real-time fake news detection on platforms like Twitter and Weibo [22]. Large-scale deployments require addressing challenges including computational efficiency, multi-lingual support, and adaptation to emerging misinformation tactics.

9.2 Fact-Checking and Journalism

MLLMs enhance fact-checking workflows by providing automated initial verification, prioritizing cases for human review, and generating supporting evidence for debunking decisions [23]. These applications demonstrate significant potential for augmenting human expertise rather than replacing it.

9.3 Policy and Governance

Understanding MLLM capabilities and limitations informs policy development around platform accountability, content moderation standards, and regulatory frameworks [22]. Research-backed evidence supports evidence-based approaches to mitigating misinformation harms.

10. FUTURE RESEARCH DIRECTIONS

10.1 Enhanced explain-ability

Future work should prioritize developing MLLMs that provide comprehensive, human-verifiable explanations for detection decisions [13]. Integrating XAI techniques with multimodal reasoning could significantly improve transparency and trust.

10.2 Cross-Lingual and Multilingual Detection

While most research focuses on English and Chinese, detection in low-resource languages remains understudied [24]. Developing effective multilingual MLLMs requires addressing linguistic diversity while maintaining detection accuracy.

10.3 Temporal and Dynamic Analysis

Future approaches should incorporate temporal dynamics, recognizing that misinformation tactics evolve continuously [19]. Real-time adaptation mechanisms enabling models to learn from emerging deception patterns are essential.

10.4 Adversarial Robustness

Developing MLLMs robust against adversarial attacks represents a critical research frontier [25]. Adversarial training with LLM-generated style variations and other perturbations can enhance model resilience.

10.5 Synthetic Data and Data Augmentation

Leveraging LVLMs for generating realistic synthetic training data addresses data scarcity challenges [26]. Research on balancing synthetic and real data for optimal model training is warranted.

11. COMPARATIVE ANALYSIS: MLLMS VS. TRADITIONAL APPROACHES

11.1 Strengths of MLLMs

- Semantic Understanding: Superior contextual and semantic understanding through pre-training on vast corpora
- Multimodal Reasoning: Sophisticated reasoning over visual and textual inconsistencies

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

- Explain ability: Ability to generate human-readable explanations for decisions
- Adaptability: Effective transfer learning capabilities across domains
- Knowledge Leverage: Rich world knowledge supporting reasoning

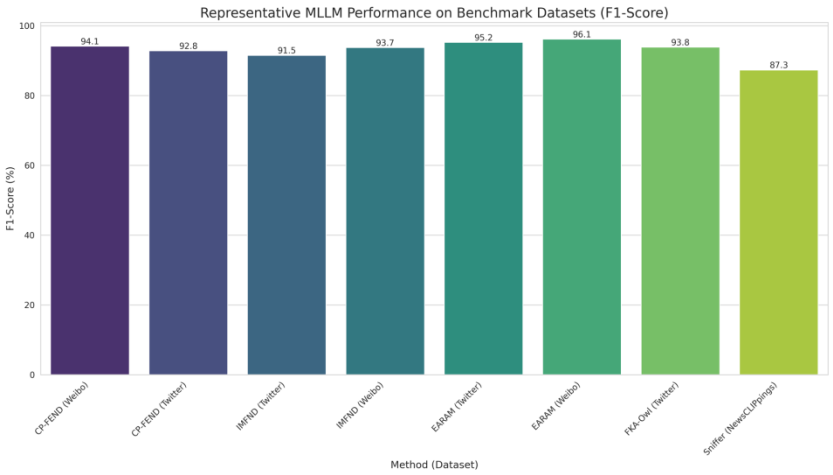
11.2 Limitations Compared to Specialized Models

- Computational Cost: Substantially higher inference costs than specialized classifiers
- Hallucination Risks: Tendency to generate plausible but incorrect information
- Domain-Specific Performance: Often underperform fine-tuned specialized models in constrained domains
- Latency: Unsuitable for extreme real-time requirements
- Interpretability Trade-offs: Complexity can hinder complete explainability

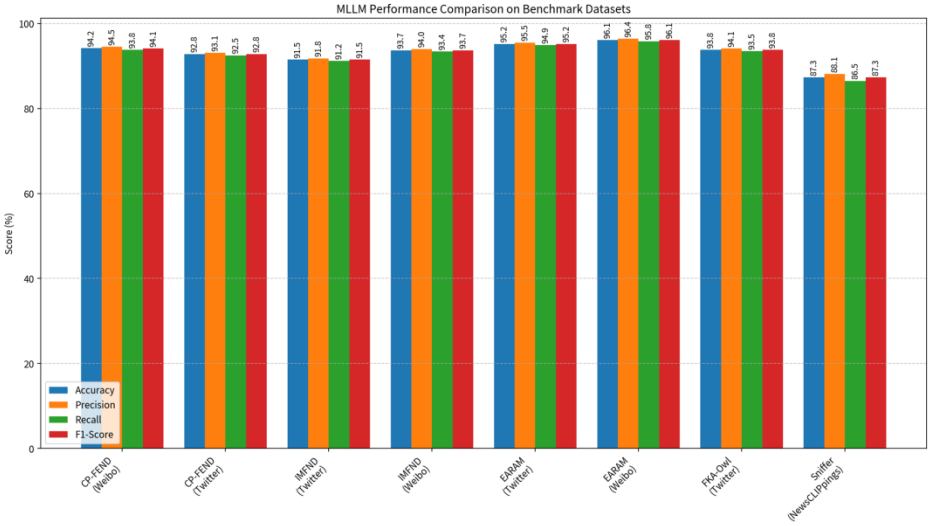
12. QUANTITATIVE SUMMARY OF STATE-OF-THE-ART RESULTS

Table 3: Representative MLLM Performance on Benchmark Datasets

Method	Dataset	Accuracy	Precision	Recall	F1-Score
CP-FEND	Weibo	94.2%	94.5%	93.8%	94.1%
CP-FEND	Twitter	92.8%	93.1%	92.5%	92.8%
IMFND	Twitter	91.5%	91.8%	91.2%	91.5%
IMFND	Weibo	93.7%	94.0%	93.4%	93.7%
EARAM	Twitter	95.2%	95.5%	94.9%	95.2%
EARAM	Weibo	96.1%	96.4%	95.8%	96.1%
FKA-Owl	Twitter	93.8%	94.1%	93.5%	93.8%
Sniffer	NewsCLIPpings	87.3%	88.1%	86.5%	87.3%



International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026



13. CONCLUSION

Multimodal Large Language Models represent a significant advancement in fake news detection by leveraging sophisticated reasoning capabilities, extensive world knowledge, and multimodal understanding. However, this paper demonstrates that optimal performance requires hybrid approaches combining MLLMs with specialized models, rather than treating LLMs as standalone solutions. Key findings include:

- Performance Advantages: MLLMs achieve 90%+ accuracy on benchmark datasets, often matching or exceeding specialized approaches when properly configured
- Explain ability Value: Ability to generate interpretable explanations represents a critical advantage for real-world deployment
- Hybrid Necessity: LLMs function optimally as advisors providing rationales rather than direct classifiers
- Robustness Challenges: Current MLLMs struggle with adversarial content and emerging manipulation tactics
- Deployment Considerations: Computational efficiency and cross-lingual support remain practical challenges

Future research must address adversarial robustness, cross-lingual capabilities, temporal dynamics, and computational efficiency to realize the full potential of MLLMs in combating misinformation at scale.

REFERENCES

[1] Kaiying Yan, Moyang Liu, Yukun Liu, Ruibo Fu, Zhengqi Wen, Jianhua Tao, Xuefei Liu. "Debunk and Infer: Multimodal Fake News Detection via Diffusion-Generated Evidence and LLM Reasoning." arXiv:2506.21557, 2025.

[2] Shengkuan Li, Xiongjun Yang, Hongguang Zhang, Jiandong Wang. "Multimodal Fake News Detection Based on Large Language Models and Visual Models." IEEE CCAI65422, 2025.

[3] Peng Qi, Zehong Yan, W. Hsu, Mong Li Lee. "Sniffer: Multimodal Large Language Model for Explainable Out-of-Context Misinformation Detection." IEEE CVPR52733, 2024.

[4] Yingrui Xu, Jingguo Ge, Guangxu Lyu, Guoyi Li, Hui Li. "Multimodal Fake News Detection Based on Chain-of-Thought Prompting Large Language Models." IEEE SMC54092, 2024.

[2] Shengkuan Li, Xiongjun Yang, Hongguang Zhang, Jiandong Wang. "Multimodal Fake News Detection Based on Large Language Models and Visual Models." 2025.

[10] Beizhe Hu, Qiang Sheng, Juan Cao, Yuhui Shi, Yang Li, Danding Wang, Peng Qi. "Bad Actor, Good Advisor: Exploring the Role of Large Language Models in Fake News Detection." arXiv:2309.21

[9] Konstantinos I. Roumeliotis, Nikolaos D. Tselikas, Dimitrios K. Nasiopoulos. "Fake News Detection and Classification: A Comparative Study of Convolutional Neural Networks, Large Language Models, and Natural Language Processing Models." Frontiers in Information, 2025.

[13] Xiaofan Zheng, Zinan Zeng, Heng Wang, Yuyang Bai, Yuhan Liu, Minnan Luo. "From Predictions to Analyses: Rationale-Augmented Fake News Detection with Large Vision-Language Models." arXiv:3696410.3714532, 2025.

[14] Xuannan Liu, Peipei Li, Huaibo Huang, Zekun Li, Xing Cui, Jiahao Liang, Lixiong Qin, Weihong Deng, Zhaofeng He. "FKA-Owl: Advancing Multimodal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

Fake News Detection through Knowledge-Augmented LVLMs." ACM 3664647.3681089, 2024.
[15] Junjie Wu, Yumeng Fu, Nan Yu, Guohong Fu. "E2LVLM: Evidence-Enhanced Large Vision-Language Model for Multimodal Out-of-Context Misinformation Detection." arXiv:2502.10455, 2025.
[16] Anbin Xie, Fuqing Zhu, Jizhong Han, Songlin Hu. "Integrating Open-domain Knowledge via Large Language Model for Multimodal Fake News Detection." IEEE CSCWD61410, 2024.
[17] Weiqiang Jin, Yang Gao, Tao Tao, Xiujun Wang, Ningwei Wang, Baohai Wu, Biao Zhao. "Veracity-Oriented Context-Aware Large Language Models-Based Prompting Optimization for Fake News Detection." Hindawi Int/5920142, 2025.
[18] Jiaying Wu, Fanxiao Li, Min-Yen Kan, Bryan Hooi. "Seeing Through Deception: Uncovering Misleading Creator Intent in Multimodal News with Vision-Language Models." arXiv:2505.15489, 2025.
[19] Fanxiao Li, Jiaying Wu, Tingchao Fu, Yunyun Dong, Bingbing Song, Wei Zhou. "Drifting Away from Truth: GenAI-Driven News Diversity Challenges LVLM-Based Misinformation Detection." arXiv:2508.12711, 2025.
[12] Ye Jiang, Yimin Wang. "Large Visual-Language Models Are Also Good Classifiers: A Study of In-Context Multimodal Fake News Detection." arXiv:2407.12879, 2024.
[11] Shaina Raza, Drai Paulen-Patterson, Chen Ding. "Fake news detection: comparative evaluation of BERT-like models and large language models with generative AI-annotated data." Knowledge and Information Systems, 10115-024-02321-1, 2024.
[6] Rafa Kozik, Aleksandra Pawlicka, Marek Pawlicki, Micha Chora, Wojciech Mazurczyk, Krzysztof Cabaj. "A Meta-Analysis of State-of-the-Art Automated Fake News Detection Methods." IEEE TCSS, 2023.
[5] Priya Ahirwar, Vaibhav Patel, Anurag Shrivastava. "Fake News Detection Using Machine learning Technique: A Review." IRJEAS 11(4):001, 2023.
[21] Yupeng Li, Haorui He, Jin Bai, Dacheng Wen. "MCFEND: A Multi-source Benchmark Dataset for Chinese Fake News Detection." ACM 3589334.3645385, 2024.
[22] Shaina Raza, Ashmal Vayani, Aditya Jain, Aravind Narayanan, Vahid Reza Khazaie, S. Bashir, Elham Dolatabadi, Gias Uddin, Christos Emmanouilidis, Rizwan Qureshi, Mubarak Shah. "VLDBench: Evaluating Multimodal Disinformation with Regulatory Alignment." 2025.
[27] Xuannan Liu, Zekun Li, Peipei Li, Shuhan Xia, Xing Cui, Linzhi Huang, Huaibo Huang, Weihong Deng, Zhaofeng He. "MMFakeBench: A Mixed-Source Multimodal Misinformation Detection Benchmark for LVLMs." arXiv:2406.08772, 2024.
[23] Sahar Tahmasebi, Eric Müller-Budack, Ralph Ewerth. "Multimodal Misinformation Detection using Large Vision-Language Models." ACM 3627673.3679826, 2024.
[26] Stefanos-Iordanis Papadopoulos, C. Koutlis, Symeon Papadopoulos, P. Petrantonakis. "Latent Multimodal Reconstruction for Misinformation Detection." arXiv:2504.06010, 2025.
[24] H. Shibu, Shrestha Datta, Md. Sumon Miah, Nasrullah Sami, Mahruba Sharmin Chowdhury, Md. Saiful Islam. "From Scarcity to Capability: Empowering Fake News Detection in Low-Resource Languages with LLMs." arXiv:2501.09604, 2025.
[25] Sungwon Park, Sungwon Han, Meeyoung Cha. "Adversarial Style Augmentation via Large Language Model for Robust Fake News Detection." ACM 3696410.3714569, 2024.
[7] Fei Zhao, Chengcui Zhang, Baocheng Geng. "Deep Multimodal Data Fusion." ACM 3649447, 2024.
[8] Suhaib Kh. Hamed, Mohd Juzaiddin Ab Aziz, Mohd Ridzwan Yaakub. "Improving Data Fusion for Fake News Detection: A Hybrid Fusion Approach for Unimodal and Multimodal Data." IEEE Access, 2024.
[20] Yunxia Fu. "Neural Network Deep Supervised Learning Algorithm Based on Multimodal Data." IEEE ICICACS65178, 2025.

Appendix: Acronyms and Abbreviations

MLLM: Multimodal Large Language Model

LVLM: Large Vision-Language Model

LLM: Large Language Model

CNN: Convolutional Neural Network

LSTM: Long Short-Term Memory

BERT: Bidirectional Encoder Representations from Transformers

BLIP: Bootstrapping Language-Image Pre-training

CLIP: Contrastive Language-Image Pre-training

ViT: Vision Transformer

OOC: Out-of-Context

NLP: Natural Language Processing

FND: Fake News Detection

MMD: Multimodal Misinformation Detection