

# Synergizing LLMs with Global Label Propagation for Multimodal Fake News Detection

Shuguo Hu<sup>1</sup>, Jun Hu<sup>2\*</sup>, Huaiwen Zhang<sup>1</sup>

<sup>1</sup>Inner Mongolia University, <sup>2</sup>National University of Singapore  
shuguo.hu@mail.imu.edu.cn, jun.hu@nus.edu.sg, huaiwen.zhang@imu.edu.cn

## Abstract

Large Language Models (LLMs) can assist multimodal fake news detection by predicting pseudo labels. However, LLM-generated pseudo labels alone demonstrate poor performance compared to traditional detection methods, making their effective integration non-trivial. In this paper, we propose Global Label Propagation Network with LLM-based Pseudo Labeling (GLPN-LLM) for multimodal fake news detection, which integrates LLM capabilities via label propagation techniques. The global label propagation can utilize LLM-generated pseudo labels, enhancing prediction accuracy by propagating label information among all samples. For label propagation, a mask-based mechanism is designed to prevent label leakage during training by ensuring that training nodes do not propagate their own labels back to themselves. Experimental results on benchmark datasets show that by synergizing LLMs with label propagation, our model achieves superior performance over state-of-the-art baselines. Our code is available online <sup>1</sup>.

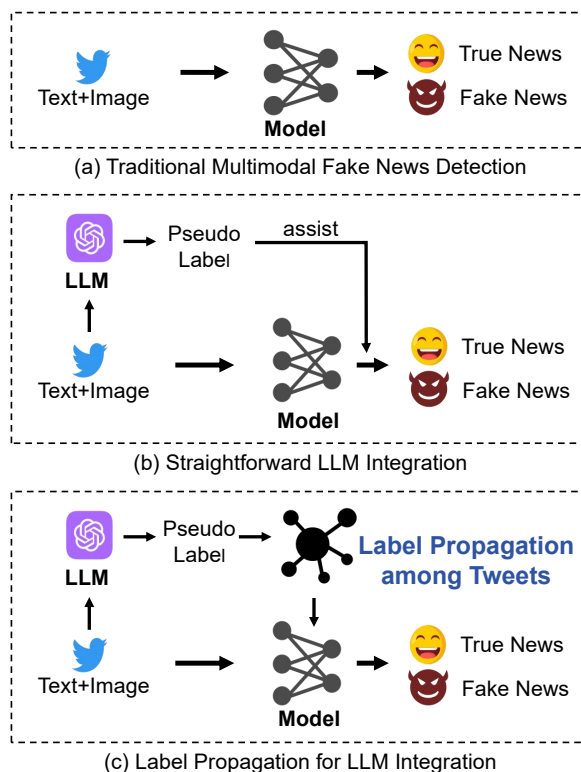


Figure 1: Illustrations of different methods.

## 1 Introduction

Detecting and mitigating the spread of multimodal fake news is a critical task for safeguarding the authenticity of information in the digital age (Zhang and Ghorbani, 2020), as shown in Figure 1(a). In recent years, the rapid growth of social media platforms has significantly accelerated the spread of misinformation, underscoring the urgent need for effective detection techniques (Shu et al., 2017; Zhou and Zafarani, 2020; Shu et al., 2019; Pérez-Rosas et al., 2017; Zhou et al., 2019). Large Language Models such as GPT-4 (Achiam et al., 2023; Brown et al., 2020) have demonstrated strong capabilities in language understanding and reasoning tasks (Xiong et al., 2024; Xu et al., 2024; Chu

et al., 2024; Bai et al., 2024), making them promising tools for enhancing fake news detection systems (Wu et al., 2024; Sun et al., 2024; Hu et al., 2024a; Su et al., 2023).

A straightforward approach to leveraging LLMs for multimodal fake news detection involves directly combining predictions from existing models with LLM outputs, as illustrated in Figure 1(b). However, LLM-generated pseudo labels may significantly underperform compared to existing multimodal fake news detection models (see Table 1), indicating that this direct combination approach requires further refinement. Therefore, it is essential to explore more effective methods for integrating LLM capabilities into fake news detection tasks.

To address these limitations, we propose a

\*Corresponding author: Jun Hu.

<sup>1</sup><https://github.com/TSCenter/GLPN-LLM>

novel framework that integrates LLM-generated pseudo labels via Label Propagation (LP) (Zhu and Ghahramani, 2002) techniques, as shown in Figure 1(c). LP enhances classification performance by propagating labels or pseudo labels between samples (Zhu and Ghahramani, 2002; Iscen et al., 2019; Zhao et al., 2023). Importantly, LP can remain effective even when pseudo label accuracy is moderate (Sun et al., 2025), making it well-suited for incorporating LLM-generated pseudo labels in fake news detection, where individual LLM predictions may be imperfect.

Our framework, Global Label Propagation Network with LLM-based Pseudo Labeling (GLPN-LLM), introduces a mask-based global label propagation module that works alongside an LLM-based pseudo label generation module. The global label propagation module can utilize LLM-generated pseudo labels, enhancing prediction accuracy by propagating label information among all samples. For label propagation, a mask-based mechanism is designed to prevent label leakage during training by ensuring that training nodes do not propagate their own labels back to themselves. Experimental results on benchmark datasets show that by synergizing LLMs with label propagation, our model achieves superior performance over state-of-the-art baselines, demonstrating its effectiveness for fake news detection.

In summary, our contributions are threefold:

- We propose Global Label Propagation Network with LLM-based Pseudo Labeling (GLPN-LLM), a novel multimodal fake news detection framework that integrates LLM capabilities via label propagation techniques.
- We introduce a mask-based global label propagation mechanism that prevents label leakage during training while effectively propagating label information across all samples.
- We conduct experiments on three benchmark datasets, demonstrating that our framework achieves superior performance compared to state-of-the-art baselines with significant improvements in accuracy and F1 scores.

## 2 Related work

In this section, we review related work in multimodal fake news detection and label propagation techniques.

### 2.1 Multimodal Fake News Detection

Early fake news detection methods primarily focused on text-based classification (Shu et al., 2017; Wang et al., 2018). Recent work in fake news detection has also benefited from multimodal learning (Jin et al., 2017), which has achieved success in many applications (Gao et al., 2025; Hu et al., 2025; Zhang et al., 2023; Yang et al., 2024; Fang et al., 2023). Recent studies show growing attention on multimodal representations (Yang et al., 2021; Singhal et al., 2019) and various multimodal methods have been proposed, such as adversarial training for modality-invariant feature learning (Wang et al., 2018), multimodal attention mechanisms (Qian et al., 2021), and multimodal graph-based approaches (Wang et al., 2020; Zhao et al., 2023).

### 2.2 Label Propagation

Label Propagation (LP) (Zhu and Ghahramani, 2002) spreads label information across a graph to predict unlabeled nodes, assuming connected nodes may share labels (Zhu et al., 2003). It has been extended to improve performance, such as kernelized LP for non-linear relationships (Zhou et al., 2003) and post-processing approaches that correct predictions through error correlation (Huang et al., 2021). Some recent work (Zhang et al., 2022; Yang et al., 2023) combines LP with Graph Neural Networks (GNNs) (Kipf and Welling, 2016; Velickovic et al., 2018; Mao et al., 2023; Hu et al., 2024b)—which show promising performance for social media applications (Liang et al., 2024; Hu et al., 2024c, 2021; Zhang et al., 2024; Sun et al., 2023; Sang et al., 2025b,a; Qiao et al., 2024; Zhang and He, 2025)—achieving encouraging results.

## 3 Method

In this section, we introduce our Global Label Propagation Network with LLM-based Pseudo Labeling (GLPN-LLM) framework.

### 3.1 Global Label Propagation Network with LLM-based Pseudo Labeling

LLM-generated pseudo labels underperform compared to existing multimodal fake news detection models. This underperformance makes their effective integration into detection systems a significant challenge. Therefore, as simpler direct combination methods require further refinement, we explore more advanced strategies to fully leverage the po-

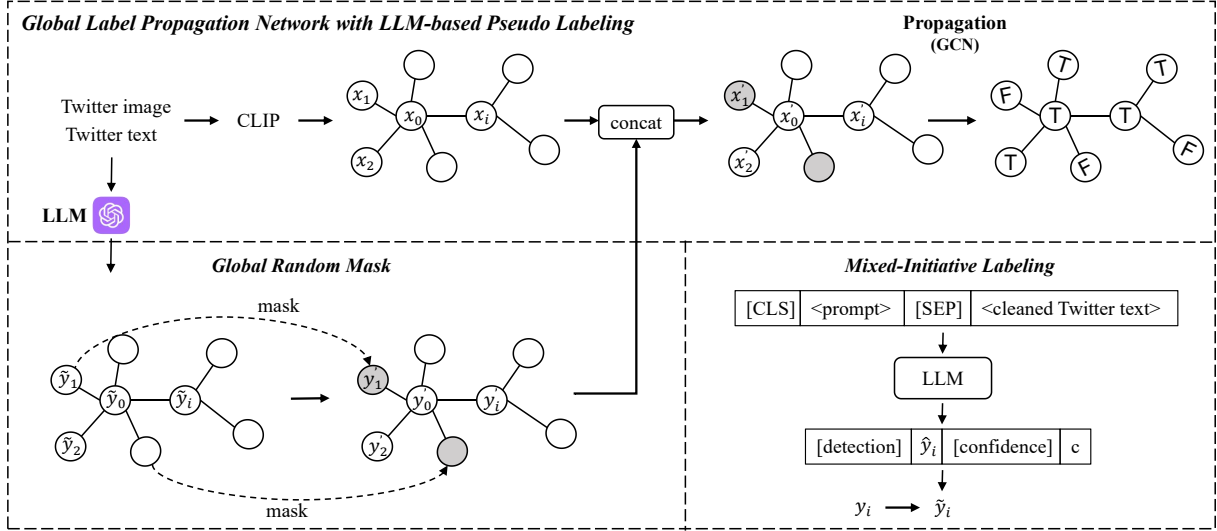


Figure 2: Overview of the GLPN-LLM framework for fake news detection. The framework synergizes LLM-generated pseudo labels with a global label propagation mechanism, leveraging multimodal features.

tential of LLMs. To address these limitations, we propose the Global Label Propagation Network with LLM-based Pseudo Labeling (GLPN-LLM) framework. This novel framework enables comprehensive label propagation across the entire graph and leverages Large Language Models to generate pseudo labels for the test set. By integrating these components, GLPN-LLM ensures full data utilization and improves label alignment between text-image representations and their corresponding labels, thereby significantly enhancing the effectiveness of fake news detection.

### 3.2 Multimodal Feature Extraction

To effectively capture the multimodal characteristics of news items, we employ CLIP (Radford et al., 2021) for feature extraction. CLIP is a powerful model that jointly learns visual and textual representations by aligning them in a shared embedding space. Given a news item comprising an image and its corresponding text, we utilize CLIP’s dual encoders to generate high-dimensional feature vectors for both modalities.

Specifically, the image encoder produces the visual feature vector  $\mathbf{v}_i \in \mathbb{R}^{d_v}$ , while the text encoder yields the textual feature vector  $\mathbf{t}_i \in \mathbb{R}^{d_t}$ . These feature vectors are then concatenated to form a unified representation  $\mathbf{x}_i \in \mathbb{R}^{d_t+d_v}$ , where:

$$\mathbf{x}_i = \mathbf{t}_i \oplus \mathbf{v}_i \quad (1)$$

Here,  $\oplus$  denotes the concatenation operation. By leveraging CLIP’s robust feature extraction capabilities, our framework generates unified feature

vectors that effectively integrate textual (semantic) and visual information from each news item. Ensuring these modalities are well-represented and aligned is crucial for enhancing the overall performance of fake news detection.

### 3.3 Cross-Modal Graph Construction

We construct a cross-modal graph following the graph construction method proposed in FCN-LP (Zhao et al., 2023). Each node in the graph represents a distinct news item, characterized by a unified feature vector. Edges between nodes are established based on multiple similarity metrics to encapsulate both intra- and inter-modal relationships. These similarity measures include: 1) *concatenated feature similarity*: This integrates both textual and visual embeddings by calculating the cosine similarity between the concatenated feature vectors of two news items. 2) *image-to-text similarity*: This measures the semantic similarity between the image feature of one news item and the text feature of another. 3) *text-to-image similarity*: This assesses the similarity from text to image across different news items. 4) *image-to-image similarity*: This captures the similarity within the same modality by comparing image features of different news items. 5) *text-to-text similarity*: This evaluates the similarity between text features of different news items. An edge is created between two nodes  $i$  and  $j$  if any of the aforementioned similarity scores exceed a predefined threshold ( $\theta = 0.95$ ). This threshold ensures that only strongly related news items are connected.

By constructing the graph using these comprehensive similarity measures, we ensure that label information can be effectively propagated across multimodally related news items. This cross-modal graph structure leverages the full spectrum of available data, enabling robust and accurate alignment between text-image representations and their corresponding labels. Consequently, the Global Label Propagation Network can maximize data utilization and enhance the overall performance.

### 3.4 Label Integration into Node Features

To facilitate effective label propagation across the entire dataset, it is essential to integrate label information directly into the node features within the cross-modal graph. This involves incorporating both the ground truth labels available for training data and the pseudo labels generated for unlabeled or test data. We introduce the Label Integration Module to achieve this, which seamlessly embeds these varied label types into node feature representations, thereby enhancing the semantic alignment between labels and multimodal data.

The core of this module is the construction of an augmented feature vector for each node  $i$ . Specifically, the label-based feature  $y'_i$ , whose composition is detailed in Section 3.5, is concatenated with the original node features  $x_i$ :

$$x'_i = x_i \oplus y'_i \quad (2)$$

This integration ensures that both feature and label information are jointly represented within each node, enabling the model to leverage label semantics during the propagation process.

**Mixed-Initiative Labeling** To further enhance the label propagation process and ensure the inclusion of high-confidence pseudo labels, we introduce the Mixed-Initiative Labeling approach. This method leverages a pre-trained LLM to generate reliable pseudo labels for unlabeled data, thereby augmenting the label information available for propagation within the graph.

The Mixed-Initiative Labeling process begins with constructing a structured prompt that incorporates both the context and the specific task requirements. The prompt is formulated as

$$X = [\text{cls}] \langle \text{prompt} \rangle [\text{SEP}] \langle \text{cleaned Twitter text} \rangle \quad (3)$$

$$Y = [\text{detection}], \hat{y}, [\text{confidence}], c \quad (4)$$

The  $\langle \text{prompt} \rangle$  provides the necessary context for the LLM to perform the detection task, while the

$\langle \text{cleaned Twitter text} \rangle$  represents the preprocessed content of the news item. Detailed examples of the prompts used are provided in Table 3.

Upon receiving the structured input, the LLM processes the prompt and generates two key outputs: 1) *Detection Label* ( $\hat{y}$ ) indicates the authenticity of the news item, categorizing it as either true or fake. 2) *Confidence Score* ( $c$ ) reflects the LLM’s confidence in its prediction, quantifying the probability associated with the generated label.

To ensure the reliability of the propagated labels, we employ a confidence-based filtering mechanism. Specifically, pseudo labels are selected for integration into the graph based on their associated confidence scores ( $c$ ). Once selected, the pseudo labels ( $\hat{y}$ ) are converted into one-hot encoded vectors and integrated into the graph’s node features. This integration is performed as follows:

$$\tilde{y}_i = \begin{cases} y_i & \text{if node } i \text{ is truly labeled,} \\ \hat{y}_i & \text{if unlabeled node } i \text{ is high-confidence,} \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

where nodes without high-confidence pseudo labels retain zero vectors as label embeddings, ensuring only reliable label information is propagated.

### 3.5 Global Random Mask for Optimization

During inference, we set the label-based features  $y'_i = \tilde{y}_i$ . However, during training,  $y'_i$  is obtained via a Global Random Mask (GRM) to prevent label leakage. Without a mechanism like the Global Random Mask, if a node’s label is included in its input features (due to label integration), the model might learn to predict this label trivially during training, without sufficiently leveraging the graph structure or original content features. This is a form of label leakage. Using these propagated labels as part of the node input features can inadvertently leak label information into the training process, resulting in biased training outcomes. GRM addresses this issue by randomly masking the label information of a subset of nodes and computing the loss only for these nodes during each training iteration, where label masking prevents label leakage.

The GRM operates by selectively masking the label embeddings of a randomly chosen subset of nodes in the unified graph during each training epoch. Specifically, given a mask ratio  $\rho$  (e.g.,  $\rho = 0.3$ ), a proportion of  $\rho \times N$  nodes are randomly selected, where  $N$  is the total number of nodes

in the graph. For each selected node  $i$ , its label embedding  $\tilde{y}_i$  is replaced with a zero vector:

$$y'_i = \tilde{y}_i \cdot m_i \quad (6)$$

where  $m_i$  is a binary mask scalar defined as:

$$m_i = \begin{cases} 0 & \text{if node } i \text{ is selected for masking,} \\ 1 & \text{otherwise.} \end{cases} \quad (7)$$

Note that this masking operation only affects the label embeddings, leaving the original feature vectors  $\mathbf{x}_i$  unchanged. This masking prevents label leakage during training by ensuring that training nodes do not propagate their own labels back to themselves, thus preventing the model from relying on this information for classification.

Given  $y'_i$ , we obtain  $x'_i$  using Equation 2 as each node’s feature vector, then apply a GCN (Kipf and Welling, 2016) on the graph to predict labels for each node. Since  $x'_i$  contains both multimodal content features and label information, both types of information are propagated through the graph by the GCN for fake news detection. The GCN output is optimized using cross-entropy loss with the Adam optimizer (Kingma, 2014).

## 4 Experiment

In this section, we conduct experiments on public benchmark datasets to verify the effectiveness of GLPN-LLM and perform detailed analysis to assess the contribution of each proposed component.

### 4.1 Datasets

We evaluate our method on three widely-used benchmark datasets: Twitter (Boididou et al., 2015), PHEME (Zubiaga et al., 2017), and Weibo (Jin et al., 2017). The Twitter dataset contains 17,000 tweets (15,000 for training, 2,000 for testing). PHEME consists of 1,414 training tweets and 608 testing tweets related to five major news events. The Weibo dataset includes 4,141 training samples and 1,125 test samples from the Sina Weibo platform. Details of these datasets are provided in Appendix A.

### 4.2 Baselines

We compare our GLPN-LLM framework against several state-of-the-art methods, including LLM (GPT-4o) (Hurst et al., 2024), EANN (Wang et al., 2018), SpotFake (Singhal et al., 2019), and MVAE (Khattar et al., 2019), which leverage multimodal features for fake news detection. We also include SAFE (Zhou et al., 2020), MCAN (Wu et al.,

2021), HMCAN (Qian et al., 2021), FCN (Zhao et al., 2023), FCN-LP (HMCAN), FCN-LP (CLIP), FCN-LP (HMCAN) + LLM, and FCN-LP (CLIP) + LLM (Zhao et al., 2023), where HMCAN and CLIP in parentheses denote the multimodal feature encoders. FCN-LP + LLM is a naive solution, where FCN-LP directly uses LLM-generated pseudo labels. For our method, we report the results of GLPN-LLM (HMCAN) and GLPN-LLM (CLIP), which use different multimodal encoders HMCAN and CLIP, respectively. For experiments in ablation studies and detailed analysis, we use CLIP by default and omit the parentheses.

### 4.3 Evaluation Metrics

We use Accuracy, Precision, Recall, and F1 Score to evaluate detection performance. Following standard practice, F1 metrics are reported as macro averages, where higher values indicate better performance.

### 4.4 Overall Evaluation

The performance results on the three datasets are shown in Table 1. Based on the results, we have the following observations:

- Among the baselines that do not use label propagation techniques, HMCAN and FCN consistently outperform the others. Note that LLM also shows poorer performance compared to these methods.
- By incorporating label propagation techniques, FCN-LP achieves a noticeable improvement in performance across all datasets over baselines that do not use LP, which validates the efficacy of label propagation on this task.
- FCN-LP + LLM leverages LLM-generated pseudo labels in a naive way, and only achieves marginal performance improvement over FCN-LP, showing that it is non-trivial to synergize LLMs with label propagation for fake news detection.
- Our framework, GLPN-LLM, introduces a mask-based global label propagation module that works alongside an LLM-based pseudo label generation module, outperforming previous methods with substantial performance improvement. Specifically, GLPN-LLM (HMCAN) demonstrates substantial improvements

Methods	Twitter				PHEME				Weibo			
	Accuracy ↑	Precision ↑	Recall ↑	F1 ↑	Accuracy ↑	Precision ↑	Recall ↑	F1 ↑	Accuracy ↑	Precision ↑	Recall ↑	F1 ↑
LLM	75.39±3.32	75.66±3.44	80.92±2.83	78.20±5.66	74.38±6.68	78.66±5.31	75.16±4.32	76.87±5.65	80.86±3.11	82.16±2.86	81.33±3.66	81.75±2.95
EANN	71.53±0.91	71.38±1.23	63.82±2.11	68.91±1.58	70.17±0.79	71.28±1.32	67.36±2.17	69.10±1.83	79.18±0.76	80.31±1.23	78.52±0.32	79.44±2.13
SpotFake	77.16±1.57	75.32±1.14	87.83±0.63	85.14±0.07	81.37±2.38	79.53±2.27	81.22±2.43	79.43±0.75	86.39±2.51	86.12±0.53	87.17±2.63	83.22±1.41
MVAE	74.56±1.58	80.15±2.69	76.34±0.83	81.57±1.98	77.83±1.27	73.82±2.05	73.45±2.62	72.21±0.54	71.86±0.25	70.32±0.69	70.32±2.84	70.53±1.60
SAFE	76.66±3.00	76.32±1.94	75.41±2.12	76.37±2.85	81.25±1.34	79.22±2.76	79.11±1.45	79.69±2.67	84.91±2.12	83.81±1.58	82.19±1.16	83.01±1.70
MCAN	80.91±2.33	82.68±2.48	76.67±0.94	82.26±1.32	80.74±1.89	79.21±2.23	79.64±1.53	80.15±0.86	86.50±3.00	88.10±2.10	84.60±1.80	86.15±1.60
HMCAN	83.91±1.49	81.68±2.08	84.67±1.21	82.57±1.62	86.36±1.83	83.18±1.41	83.81±2.51	83.49±1.07	86.75±2.95	88.40±3.00	84.65±1.80	87.20±1.20
FCN	82.86±1.27	78.64±1.68	87.39±0.85	82.78±0.47	80.36±1.93	84.43±1.27	89.12±0.12	86.71±1.88	82.92±0.54	83.17±1.00	88.45±2.13	86.74±0.41
FCN-LP (HMCAN)	84.57±1.62	83.58±1.66	85.22±2.06	84.04±0.82	87.25±1.18	84.48±1.91	84.78±1.95	84.50±0.85	87.15±1.32	88.82±2.56	86.55±1.98	88.11±1.66
FCN-LP (CLIP)	85.32±2.56	81.52±2.82	89.32±0.99	85.24±1.93	84.68±0.81	86.32±1.55	89.85±1.22	87.97±0.88	84.47±1.66	88.41±0.26	91.18±0.69	89.78±0.84
FCN-LP (HMCAN) + LLM	85.30±1.88	84.70±2.25	86.30±2.53	85.10±0.98	87.35±1.37	84.55±2.76	85.85±2.19	84.60±1.28	87.55±1.94	89.30±2.85	88.60±2.39	88.75±1.97
FCN-LP (CLIP) + LLM	85.93±1.83	81.92±2.64	90.44±2.17	85.97±2.72	84.89±2.81	87.80±2.36	90.55±2.28	89.21±2.59	84.88±2.78	88.55±2.43	91.68±1.95	89.85±2.67
GLPN-LLM (HMCAN)	87.60±1.23	<b>86.52±0.92</b>	88.88±1.88	86.86±0.85	<b>88.29±0.38</b>	86.92±0.88	88.14±1.11	86.87±1.32	<b>90.66±1.32</b>	89.06±0.13	92.20±0.17	91.46±0.16
GLPN-LLM (CLIP)	<b>88.83±2.23</b>	84.02±2.43	<b>92.68±2.81</b>	<b>89.03±2.18</b>	86.47±1.97	<b>89.24±2.74</b>	<b>92.13±2.56</b>	<b>90.66±2.32</b>	86.74±1.32	<b>89.83±0.88</b>	<b>93.27±0.65</b>	<b>91.52±0.66</b>

Table 1: Performance comparison of different methods on the Twitter, PHEME, and Weibo datasets. The highest value in each column is marked in bold.

over FCN-LP (HMCAN) + LLM, and GLPN-LLM (CLIP) also shows considerable gains compared to FCN-LP (CLIP) + LLM. These consistent improvements are observed across all three datasets. This shows that GLPN-LLM can effectively integrate LLM capabilities via label propagation techniques for multimodal fake news detection.

## 4.5 Ablation Study

To assess the contribution of each component in our GLPN-LLM framework, we conduct an ablation study, summarized in Table 2. By introducing our mask-based global label-propagation module, GLPN surpasses FCN-LP, which utilizes basic label propagation techniques, demonstrating the superiority of our label propagation strategy. GLPN-LLM extends GLPN by coupling our mask-based global label-propagation module with an LLM-based pseudo-label generation module, effectively integrating LLMs to achieve superior performance in multimodal fake news detection.

## 4.6 Detailed Analysis

### 4.6.1 Effect of Mask Rate

We examine the impact of the mask rate on the performance of GLPN-LLM, as shown in Figure 3. The mask rate determines the proportion of label information that is masked during each training iteration, influencing how much label data is available for propagation. As illustrated in Figure 3, we vary the mask rate from 0.1 to 0.9 and observe the corresponding changes in accuracy, precision, and

recall. Our results indicate that a mask rate of 0.5 yields the best performance, with the model achieving an accuracy of 86.80%, precision of 86.3%, and recall of 91.8%. This optimal mask rate suggests that balancing the availability and masking of label information is crucial for effective label propagation, allowing the model to generalize well without overfitting to specific label patterns.

### 4.6.2 Impact of the Quantity of LLM-Generated Pseudo Labels

We analyze how the number of LLM-generated pseudo labels affects detection performance, as shown in Figure 3. We experiment with varying the percentage of pseudo labels integrated into the graph, ranging from the top 1% to the top 90% based on confidence scores. The results, depicted in Figure 3, demonstrate that incorporating pseudo labels up to the top 5% of confidence scores yields the highest performance improvements. Increasing the percentage beyond 5% does not lead to further gains and may even degrade performance due to the inclusion of lower-confidence labels, which can introduce noise into the label propagation process. Therefore, selecting a top 5% threshold ensures that only high-confidence pseudo labels are utilized, enhancing the reliability and effectiveness of label propagation within the graph.

### 4.6.3 Analysis of Prompt Designs

We analyze the effect of prompt specificity by varying the level of contextual information provided, as shown in Table 3 and Table 4 in the appendix. In general, more detailed prompts tend to yield better

Method	Twitter				PHEME				Weibo			
	Accuracy $\uparrow$	Precision $\uparrow$	Recall $\uparrow$	F1 $\uparrow$	Accuracy $\uparrow$	Precision $\uparrow$	Recall $\uparrow$	F1 $\uparrow$	Accuracy $\uparrow$	Precision $\uparrow$	Recall $\uparrow$	F1 $\uparrow$
FCN-LP	85.32	81.52	89.32	85.24	84.68	86.32	89.85	87.97	84.47	88.41	91.18	89.78
GLPN	85.47	82.27	90.57	86.30	85.58	87.58	89.99	86.96	85.07	88.95	91.61	90.76
GLPN-LLM	<b>88.83</b>	<b>84.02</b>	<b>92.68</b>	<b>89.03</b>	<b>86.47</b>	<b>89.24</b>	<b>92.13</b>	<b>90.66</b>	<b>86.74</b>	<b>89.83</b>	<b>93.27</b>	<b>91.52</b>

Table 2: Ablation study of different components of GLPN-LLM for multimodal fake news detection. The highest value in each column is marked in bold.

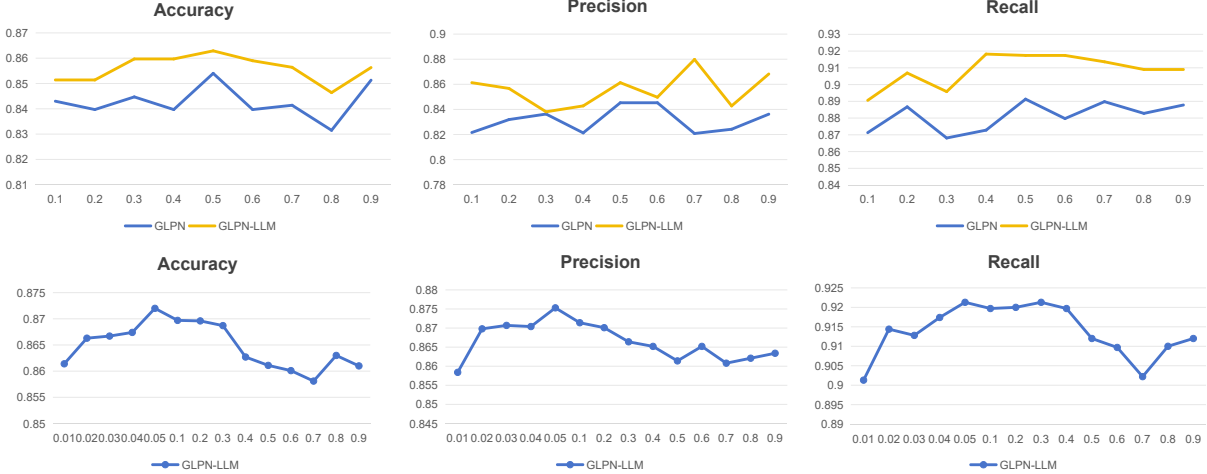


Figure 3: The effect of the mask rate and the impact of the quantity of LLM-generated pseudo labels. The first row shows how the mask rate parameter affects Accuracy, Precision, and Recall. The second row investigates how incorporating different proportion of LLM-generated pseudo labels from the test set into the label propagation process affects performance. The presented results are averaged across the three benchmark datasets.

performance as they provide clearer and more comprehensive guidance for the model. Prompts with sufficient detail ensure higher confidence scores and more reliable pseudo-label generation, resulting in improved recall rates and F1 scores, as shown in Figure 4. In contrast, simple prompts, which lack necessary contextual information, often lead to poorer performance, lower confidence, and reduced label accuracy. However, it is important to strike a balance—while detailed prompts are beneficial, excessive complexity or overly intricate phrasing may introduce noise, potentially confusing the model and diminishing the effectiveness of label generation. Clear, concise, and well-structured prompts remain optimal for achieving consistent and reliable results.

#### 4.7 Case Study

We present a case study demonstrating the classification process of our GLPN-LLM framework for real and fake news. As shown in Figure 5, the model relies on textual content, image features extracted using CLIP, and LLM-generated pseudo labels. Real news is labeled ‘R’, Fake news ‘F’.

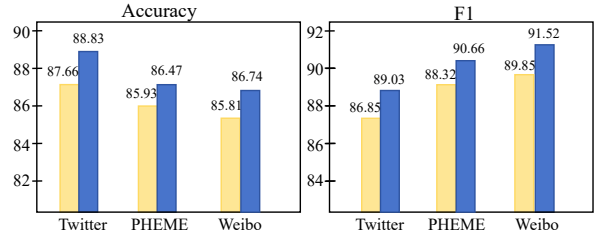


Figure 4: Effect of prompt detail on performance. Detailed prompts achieve higher accuracy and F1 scores than simple prompts. Yellow bars represent simple prompts, while blue bars represent detailed prompts.

We report the model’s predicted probability for the challenging instances.

In the first example, the original model misclassifies real news as fake. However, after applying the GLPN-LLM framework, the global label propagation module successfully corrects the classification, identifying the news as real. The model relied on the consistency of features across the dataset, ultimately leading to the correct classification despite initially low credibility scores for individual features. In the second example, the model initially misclassifies fake news as real. With the

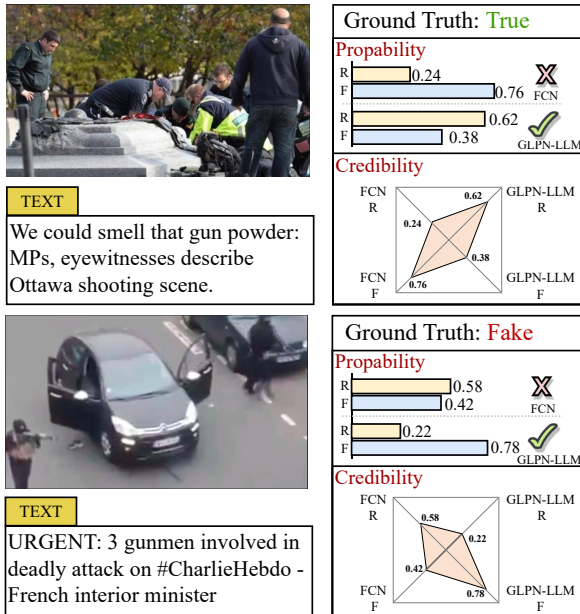


Figure 5: Case study - examples of challenging instances with their corresponding images and text.

GLPN-LLM framework, the integration of text, image features, and LLM-generated pseudo labels helps the model accurately identify the news as fake. The label propagation mechanism ensures that the fake news label is correctly spread across related data points, rectifying the original misclassification. The model’s decision is reinforced by the alignment between multiple features, resulting in the correct classification outcome.

#### 4.8 Visualization

Figure 6 presents t-SNE visualizations of the feature embeddings on the test set for three model configurations. The first column shows the embeddings produced by FCN; real and fake-news clusters largely overlap, indicating that the model struggles to distinguish the two classes. This overlap suggests that relying solely on FCN does not capture the nuanced differences between real and fake news. The second column depicts the embeddings from GLPN. With our mask-based global label propagation module, the separation between real and fake news becomes more pronounced—particularly on the Weibo dataset—demonstrating that GLPN yields embeddings in which the two classes are more distinct. The most substantial improvement appears when we integrate LLM-generated pseudo-labels (GLPN-LLM, third column). In this configuration the clusters are clearly separated, with minimal overlap, as illustrated on the Twitter dataset. This

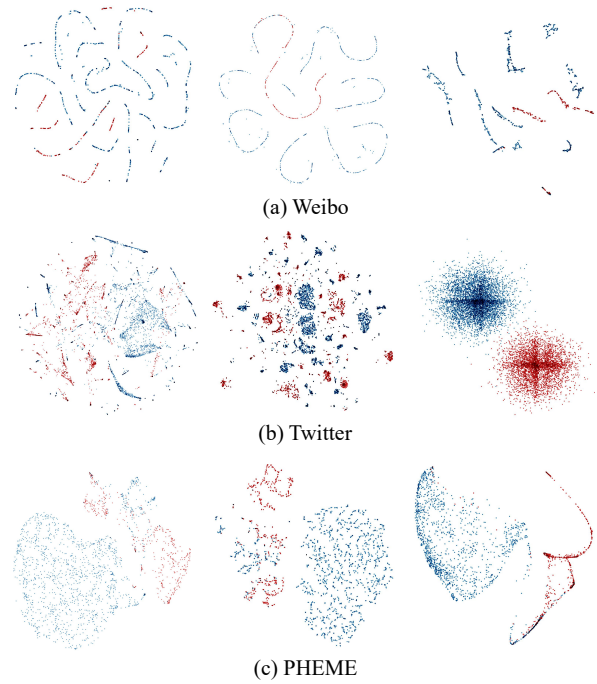


Figure 6: t-SNE visualizations of feature embeddings on the test set. The first column shows t-SNE embeddings from FCN, the second column shows embeddings from GLPN, and the third column shows embeddings from GLPN-LLM. Each point is color-coded according to its label.

observation underscores the synergistic effect of combining label propagation with the rich semantic signals provided by LLM-generated labels.

#### 5 Conclusion

In this paper, we present a framework, Global Label Propagation Network with LLM-based Pseudo Labeling (GLPN-LLM), for multimodal fake news detection. While LLM-generated pseudo labels alone demonstrate poor performance compared to traditional detection methods, our approach effectively integrates LLM capabilities via label propagation techniques. Experiments on three benchmark datasets demonstrate that GLPN-LLM consistently outperforms state-of-the-art baselines with significant improvements, highlighting the effectiveness of synergizing LLMs with label propagation for fake news detection. In the future, our work will focus on exploring approaches to improve GLPN-LLM’s scalability to larger and more complex datasets, while examining its adaptability across diverse social media platforms and content modalities to enhance practical applicability in real-world deployment scenarios.



## Acknowledgement

This work was supported in part by the National Natural Science Foundation of China under Grants 62206137, 62206200, 62276257, 62036012, in part by the Program for Young Talents of Science and Technology in Universities of Inner Mongolia Autonomous Region under Grant NJYT23105, and in part by the National Natural Science Foundation of Inner Mongolia under Grant 2025JQ012.

## Limitation

**Dependency on Backbone Models** The effectiveness of our GLPN-LLM framework is closely tied to the performance of the underlying backbone models, namely FCN. While these models provide strong feature representations, any limitations in their ability to capture comprehensive semantic relationships can directly impact the label propagation process. Consequently, the overall detection accuracy is highly dependent on the quality and robustness of these backbone models. Additionally, the reliance on specific backbones may limit the adaptability of our framework to other feature extraction architectures that might offer different advantages.

**Reliance on High-Confidence Pseudo Labels** Our approach relies on the generation of high-confidence pseudo labels by the LLM to enhance label propagation. However, the accuracy of these pseudo labels is contingent upon the LLM's ability to produce reliable predictions. Inaccurate or biased pseudo labels can introduce noise into the label propagation process, potentially degrading the model's performance. Ensuring the reliability of pseudo labels is crucial, and future work may explore more robust methods for pseudo label verification and refinement to mitigate this limitation.

## References

- Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.
- Jiaxin Bai, Yicheng Wang, Tianshi Zheng, Yue Guo, Xin Liu, and Yangqiu Song. 2024. Advancing abductive reasoning in knowledge graphs through complex logical hypothesis generation. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, ACL 2024, Bangkok, Thailand, August 11-16, 2024, pages 1312–1329. Association for Computational Linguistics.
- Christina Boididou, Katerina Andreadou, Symeon Papadopoulos, Duc Tien Dang Nguyen, Giulia Boato, Michael Riegler, Yiannis Kompatsiaris, et al. 2015. Verifying multimedia use at mediaeval 2015. In *MediaEval 2015*, volume 1436. CEUR-WS.
- Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.
- Yuqi Chu, Lizi Liao, Zhiyuan Zhou, Chong-Wah Ngo, and Richang Hong. 2024. Towards multimodal emotional support conversation systems. *arXiv preprint arXiv:2408.03650*.
- Quan Fang, Xiaowei Zhang, Jun Hu, Xian Wu, and Changsheng Xu. 2023. Contrastive multi-modal knowledge graph representation learning. *IEEE Trans. Knowl. Data Eng.*, 35(9):8983–8996.
- Junyu Gao, Mengyuan Chen, and Changsheng Xu. 2025. Learning probabilistic presence-absence evidence for weakly-supervised audio-visual event perception. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Beizhe Hu, Qiang Sheng, Juan Cao, Yuhui Shi, Yang Li, Danding Wang, and Peng Qi. 2024a. Bad actor, good advisor: Exploring the role of large language models in fake news detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 22105–22113.
- Jun Hu, Bryan Hooi, and Bingsheng He. 2024b. Efficient heterogeneous graph learning via random projection. *IEEE Trans. Knowl. Data Eng.*, 36(12):8093–8107.
- Jun Hu, Bryan Hooi, Bingsheng He, and Yinwei Wei. 2025. Modality-independent graph neural networks with global transformers for multimodal recommendation. In *AAAI-25, Sponsored by the Association for the Advancement of Artificial Intelligence, February 25 - March 4, 2025, Philadelphia, PA, USA*, pages 11790–11798. AAAI Press.
- Jun Hu, Bryan Hooi, Shengsheng Qian, Quan Fang, and Changsheng Xu. 2024c. MGDCF: distance learning via markov graph diffusion for neural collaborative filtering. *IEEE Trans. Knowl. Data Eng.*, 36(7):3281–3296.
- Jun Hu, Shengsheng Qian, Quan Fang, Youze Wang, Quan Zhao, Huaiwen Zhang, and Changsheng Xu. 2021. Efficient graph deep learning in tensorflow with tf\_geometric. In *MM '21: ACM Multimedia Conference, Virtual Event, China, October 20 - 24, 2021*, pages 3775–3778. ACM.

- Qian Huang, Horace He, Abhay Singh, Ser-Nam Lim, and Austin R. Benson. 2021. Combining label propagation and simple models out-performs graph neural networks. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net.
- Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, et al. 2024. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*.
- Ahmet Iscen, Giorgos Tolias, Yannis Avrithis, and Ondrej Chum. 2019. Label propagation for deep semi-supervised learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5070–5079.
- Zhiwei Jin, Juan Cao, Han Guo, Yongdong Zhang, and Jiebo Luo. 2017. Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 795–816.
- Dhruv Khattar, Jaipal Singh Goud, Manish Gupta, and Vasudeva Varma. 2019. Mvae: Multimodal variational autoencoder for fake news detection. In *The world wide web conference*, pages 2915–2921.
- Diederik P Kingma. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.
- Yuxuan Liang, Wentao Zhang, Zeang Sheng, Ling Yang, Jiawei Jiang, Yunhai Tong, and Bin Cui. 2024. HGAMLP: heterogeneous graph attention MLP with de-redundancy mechanism. In *40th IEEE International Conference on Data Engineering, ICDE 2024, Utrecht, The Netherlands, May 13-16, 2024*, pages 2779–2791. IEEE.
- Qiheng Mao, Zemin Liu, Chenghao Liu, and Jianling Sun. 2023. Hinormer: Representation learning on heterogeneous information networks with graph transformer. In *Proceedings of the ACM Web Conference 2023, WWW 2023, Austin, TX, USA, 30 April 2023 - 4 May 2023*, pages 599–610. ACM.
- Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. 2017. Automatic detection of fake news. *arXiv preprint arXiv:1708.07104*.
- Shengsheng Qian, Jinguang Wang, Jun Hu, Quan Fang, and Changsheng Xu. 2021. Hierarchical multi-modal contextual attention network for fake news detection. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, pages 153–162.
- Hezhe Qiao, Hanghang Tong, Bo An, Irwin King, Charu Aggarwal, and Guansong Pang. 2024. Deep graph anomaly detection: A survey and new perspectives. *arXiv preprint arXiv:2409.09957*.
- Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR.
- Lei Sang, Honghao Li, Yiwen Zhang, Yi Zhang, and Yun Yang. 2025a. Adagin: Adaptive graph interaction network for click-through rate prediction. *ACM Trans. Inf. Syst.*, 43(1):3:1–3:31.
- Lei Sang, Yu Wang, Yi Zhang, Yiwen Zhang, and Xindong Wu. 2025b. Intent-guided heterogeneous graph contrastive learning for recommendation. *IEEE Trans. Knowl. Data Eng.*, 37(4):1915–1929.
- Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 19(1):22–36.
- Kai Shu, Suhang Wang, and Huan Liu. 2019. Beyond news contents: The role of social context for fake news detection. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pages 312–320. ACM.
- Shivangi Singhal, Rajiv Ratn Shah, Tanmoy Chakraborty, Ponnurangam Kumaraguru, and Shin’ichi Satoh. 2019. Spofake: A multi-modal framework for fake news detection. In *2019 IEEE fifth international conference on multimedia big data (BigMM)*, pages 39–47. IEEE.
- Jinyan Su, Claire Cardie, and Preslav Nakov. 2023. Adapting fake news detection to the era of large language models. *arXiv preprint arXiv:2311.04917*.
- Chuxiong Sun, Jie Hu, Hongming Gu, Jinpeng Chen, Wei Liang, and Mingchuan Yang. 2025. Scalable and adaptive graph neural networks with self-label-enhanced training. *Pattern Recognit.*, 160:111210.
- Peijie Sun, Le Wu, Kun Zhang, Xiangzhi Chen, and Meng Wang. 2023. Neighborhood-enhanced supervised contrastive learning for collaborative filtering. *IEEE Transactions on Knowledge and Data Engineering*, 36(5):2069–2081.
- Yanshen Sun, Jianfeng He, Limeng Cui, Shuo Lei, and Chang-Tien Lu. 2024. Exploring the deceptive power of llm-generated fake news: A study of real-world detection challenges. *arXiv preprint arXiv:2403.18249*.
- Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph attention networks. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net.
- Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, and Jing Gao. 2018. Eann: Event adversarial neural networks for multi-modal fake news detection. In *Proceedings of*

- the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pages 849–857. ACM.
- Youze Wang, Shengsheng Qian, Jun Hu, Quan Fang, and Changsheng Xu. 2020. Fake news detection via knowledge-driven multimodal graph convolutional networks. In *Proceedings of the 2020 on International Conference on Multimedia Retrieval, ICMR 2020, Dublin, Ireland, June 8-11, 2020*, pages 540–547. ACM.
- Jiaying Wu, Jiafeng Guo, and Bryan Hooi. 2024. Fake news in sheep’s clothing: Robust fake news detection against llm-empowered style attacks. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3367–3378.
- Yang Wu, Pengwei Zhan, Yunjian Zhang, Liming Wang, and Zhen Xu. 2021. Multimodal fusion with co-attention networks for fake news detection. In *Findings of the association for computational linguistics: ACL-IJCNLP 2021*, pages 2560–2569.
- Siheng Xiong, Ali Payani, Ramana Kompella, and Faramarz Fekri. 2024. Large language models can learn temporal reasoning. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2024, Bangkok, Thailand, August 11-16, 2024*, pages 10452–10470. Association for Computational Linguistics.
- Yifan Xu, Xiaoshan Yang, Yaguang Song, and Changsheng Xu. 2024. Libra: Building decoupled vision system on large language models. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*. OpenReview.net.
- Xiaocheng Yang, Mingyu Yan, Shirui Pan, Xiaochun Ye, and Dongrui Fan. 2023. Simple and efficient heterogeneous graph neural network. In *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023*, pages 10816–10824. AAAI Press.
- Xiaoshan Yang, Baochen Xiong, Yi Huang, and Changsheng Xu. 2024. Cross-modal federated human activity recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 46(8):5345–5361.
- Xiaoyu Yang, Yuefei Lyu, Tian Tian, Yifei Liu, Yudong Liu, and Xi Zhang. 2021. Rumor detection on social media with graph structured adversarial learning. In *Proceedings of the twenty-ninth international conference on international joint conferences on artificial intelligence*, pages 1417–1423.
- Huaiwen Zhang, Zihang Guo, Yang Yang, Xin Liu, and De Hu. 2023. C<sup>2</sup>st: Cross-modal contextualized sequence transduction for continuous sign language recognition. In *IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris, France, October 1-6, 2023*, pages 20996–21005. IEEE.
- Huaiwen Zhang, Xinxin Liu, Qing Yang, Yang Yang, Fan Qi, Shengsheng Qian, and Changsheng Xu. 2024. T<sup>3</sup>rd: Test-time training for rumor detection on social media. In *Proceedings of the ACM on Web Conference 2024, WWW 2024, Singapore, May 13-17, 2024*, pages 2407–2416. ACM.
- Wentao Zhang, Ziqi Yin, Zeang Sheng, Yang Li, Wen Ouyang, Xiaosen Li, Yangyu Tao, Zhi Yang, and Bin Cui. 2022. Graph attention multi-layer perceptron. In *KDD ’22: The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, August 14 - 18, 2022*, pages 4560–4570. ACM.
- Xichen Zhang and Ali A Ghorbani. 2020. An overview of online fake news: Characterization, detection, and discussion. *Information Processing & Management*, 57(2):102025.
- Zhen Zhang and Bingsheng He. 2025. Aggregate to adapt: Node-centric aggregation for multi-source-free graph domain adaptation. In *Proceedings of the ACM on Web Conference 2025, WWW 2025, Sydney, NSW, Australia, 28 April 2025- 2 May 2025*, pages 4420–4431. ACM.
- Wanqing Zhao, Yuta Nakashima, Haiyuan Chen, and Noboru Babaguchi. 2023. Enhancing fake news detection in social media via label propagation on cross-modal tweet graph. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 2400–2408.
- Dengyong Zhou, Olivier Bousquet, Thomas Lal, Jason Weston, and Bernhard Schölkopf. 2003. Learning with local and global consistency. *Advances in neural information processing systems*, 16.
- Xinyi Zhou, Jindi Wu, and Reza Zafarani. 2020. : Similarity-aware multi-modal fake news detection. In *Pacific-Asia Conference on knowledge discovery and data mining*, pages 354–367. Springer.
- Xinyi Zhou and Reza Zafarani. 2020. A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, 53(5):1–40.
- Xinyi Zhou, Reza Zafarani, Kai Shu, and Huan Liu. 2019. Fake news: Fundamental theories, detection strategies and challenges. In *Proceedings of the twelfth ACM international conference on web search and data mining*, pages 836–837.
- Xiaojin Zhu and Zoubin Ghahramani. 2002. Learning from labeled and unlabeled data with label propagation. *ProQuest number: information to all users*.
- Xiaojin Zhu, Zoubin Ghahramani, and John D Lafferty. 2003. Semi-supervised learning using gaussian fields and harmonic functions. In *Proceedings of the*

Arkaitz Zubiaga, Maria Liakata, and Rob Procter. 2017. Exploiting context for rumour detection in social media. In *Social Informatics: 9th International Conference, SocInfo 2017, Oxford, UK, September 13-15, 2017, Proceedings, Part I 9*, pages 109–123. Springer.

## A Datasets

We validate the proposed methods on three real social media datasets: Twitter (Boididou et al., 2015), PHEME (Zubiaga et al., 2017), and Weibo (Jin et al., 2017):

**Twitter** (Boididou et al., 2015) is used to detect fake content on social media by analyzing approximately 17,000 unique tweets. The tweets are originally from the widely-used fake news dataset MediaEval Verifying Multimedia Use benchmark (Boididou et al., 2015). Each tweet includes textual content and a related image, with labels for real/fake. Following the benchmark (Zhao et al., 2023), we split the dataset into train and test sets with 15,000 and 2,000 tweets, respectively.

**PHEME** (Zubiaga et al., 2017) includes tweets sourced from Twitter, specifically targeting five significant breaking news events. Each event includes a substantial collection of tweets, along with their textual content, associated images, and real/fake labels. Following the setup in the benchmark (Zhao et al., 2023), we adopt 1,414 and 608 tweets as the training and test sets, respectively.

**Weibo** (Jin et al., 2017) originates from Sina Weibo, a widely used microblogging platform in China. Following the setup in the benchmark (Zhao et al., 2023), we adopt 4,141 and 1125 tweets as the training and test sets, respectively.

## B Baselines

We benchmark our proposed GLPN-LLM framework against several state-of-the-art methods that utilize multimodal features and label propagation techniques. The combined effectiveness of the backbone and label propagation is significantly influenced by the backbone’s ability to extract and integrate features, with weaker backbones limiting overall performance. Building upon recent advancements in label propagation (Zhao et al., 2023), we include only methods that employ state-of-the-art backbone architectures.

We provide a comprehensive comparison with

leading multimodal and label propagation approaches.

- EANN (Wang et al., 2018) employs attention mechanisms to integrate textual and visual features for fake news detection, focusing on attention-based fusion of modalities to enhance prediction accuracy.
- SpotFake (Singhal et al., 2019) utilizes multimodal information to identify fake news by analyzing both textual content and accompanying images, optimizing feature alignment across text and images for effective detection.
- MVAE (Khattar et al., 2019) is a multimodal variational autoencoder that captures the joint distribution of text and image data, improving fake news classification through the combination of textual and visual information.
- MCAN (Wu et al., 2021) is a multi-modal contextual attention network that fuses inter-modality and intra-modality relationships, enhancing fake news detection by modeling the dependencies between text and image modalities.
- SAFE (Zhou et al., 2020) combines multimodal feature extraction with cross-modal similarity measures to learn tweet representations, directly measuring the similarity across modalities to achieve effective alignment for detecting fake news.
- HMCAN (Qian et al., 2021) utilizes a hierarchical multi-modal contextual attention network to capture rich hierarchical semantics, enhancing fake news detection by modeling both high-level and fine-grained relationships in the data.
- FCN (Zhao et al., 2023) utilizes CLIP for feature extraction. It constructs a cross-modal tweet graph to unify text and image features and then employs a Graph Convolutional Network (GCN) for classification.
- FCN-LP (Zhao et al., 2023) uses CLIP for feature extraction, followed by fixed label propagation. It builds a cross-modal tweet graph to unify text and image features, utilizing iterative label propagation to refine predictions.

To ensure fairness in our comparisons, we follow the benchmark setup of FCN-LP (Zhao et al., 2023) and use the same similarity threshold  $\theta$  across all datasets to construct the cross-modal tweet graph.

## C Implementation Details

For our GLPN-LLM framework, the core graph neural network is a Graph Convolutional Network (GCN) (Kipf and Welling, 2016). We adopt Adam (Kingma, 2014) as the optimizer with a  $1e-3$  learning rate. The hyperparameters for label propagation mask rate  $\lambda_{\text{mask}}$  and LLM pseudo rate  $\lambda_{\text{pseudo}}$  are set to 0.50 and 0.05, respectively. The multimodal common space dimension learned by GCN is set to 512. To enable a consistent comparison with the baseline, we follow the settings of previous work FCN-LP (Zhao et al., 2023) and set the similarity threshold to  $\theta = 0.95$  for the Twitter, PHEME and Weibo datasets when building the cross-modal tweet graph. We train the model 5 times and report the average and standard deviation for accuracy, precision, recall, and F1 score.

## D Efficiency

The LLM module operates independently of the Label Propagation module and utilizes an API to significantly reduce computational overhead. Its computational complexity is  $O(N_{\text{test}} \cdot T_{\text{avg}})$ , where  $N_{\text{test}}$  is the size of the test set and  $T_{\text{avg}}$  is the average number of tokens per sample (input + output). Taking the Twitter dataset as an example, the LLM exclusively processes tweets from the test set. Based on GPT-4o API rates, the processing cost per tweet is approximately \$0.00074, which renders the module highly efficient for deployment. This high efficiency ensures that the module is both cost-effective and scalable for practical deployment.

The LP module introduces minimal additional cost. This is achieved by injecting pseudo labels (converted to one-hot vectors) into node features and then propagating them across the graph. Its complexity for  $K$  iterations on a graph with  $M$  edges is  $O(K \cdot M)$ . Consequently, the label propagation step increases the overall computation time by only about 2%, while substantially boosting performance.

Overall, the framework’s computational costs, both in time and monetary terms, remain low and are more than justified by the resulting accuracy improvements, making the approach practical for real-world scenarios.

Table 3: Prompts for Mixed-Initiative Labeling

Role	Content
system	<p>You are a professional misinformation evaluation expert with extensive experience in detecting and evaluating fake news. Your primary task is to assess the authenticity of the provided news content.</p> <p>You must adhere to the following strict evaluation guidelines:</p> <ul style="list-style-type: none"> <li>• If the news is clearly true, label it as 1 (true).</li> <li>• If the news contains ambiguity, unverifiable information, or suspicious claims, you must classify it as 0 (false).</li> <li>• Alongside your classification, provide a confidence score (between 0% and 100%) that reflects your certainty in the decision.</li> <li>• Your confidence score should be lower (e.g., 50%-70%) when the news includes unclear or mixed signals, and higher (e.g., 80%-100%) when you are certain.</li> </ul> <p>Output format: Result: R, Confidence: C%, where R is 1 (true) or 0 (false), and C is the confidence score. Be precise, concise, avoid unnecessary explanations, and give me the reason.</p>
user	<p><b>BREAKING:</b> SkyBusiness reports another five hostages seen escaping #sydneysiege. Additional context: The event is unfolding in Sydney’s central business district. Initial unverified reports mention hostages being rescued, but official statements have not yet been released.</p>
system	<p>Result: 1, Confidence: 85%</p> <p>Reason: Based on credible news reports and consistent information across major media outlets, the claim of hostages escaping appears highly plausible. Minor uncertainty remains due to the absence of official verification.</p>
user	<p><b>CONFIRMED:</b> NASA discovers alien life on Mars.</p>
system	<p>Result: 0, Confidence: 30%</p> <p>Reason: This claim lacks supporting evidence from verified scientific sources, and NASA has not released any official confirmation regarding such a discovery. The headline seems sensationalized or misleading.</p>
user	<p><b>ALERT:</b> Severe storms expected to hit California tomorrow, warns National Weather Service.</p>
system	<p>Result: 1, Confidence: 95%</p> <p>Reason: The information originates from the National Weather Service, a highly reliable and authoritative source. Severe weather forecasts for tomorrow are consistent across official meteorological channels.</p>

Table 4: Simple Prompt for Mixed-Initiative Labeling

<b>Role</b>	<b>Content</b>
system	You are tasked with determining whether the provided news content is true or false. Output format: Result: R, Confidence :c, where R is 1 (true) or 0 (false).
user	BREAKING: SkyBusiness reports another five hostages seen escaping #sydneyseige.
system	Result: 1, Confidence: 49%
user	CONFIRMED: NASA discovers alien life on Mars.
system	Result: 0, Confidence: 20%
user	ALERT: Severe storms expected to hit California tomorrow.
system	Result: 1, Confidence: 63%