

Unified Evidence Enhancement Inference Framework for Fake News Detection

Lianwei Wu^{1,2}, Linyong Wang^{1,2}, Yongqiang Zhao³

¹ASGO, School of Computer Science, Northwestern Polytechnical University, Xi'an, China

²Research & Development Institute of Northwestern Polytechnical University in Shenzhen

³School of Computer Science, Peking University, Beijing, China

wlw@nwpu.edu.cn, linyongwang@mail.nwpu.edu.cn, yongqiangzhao@stu.pku.edu.cn

Abstract

The current approaches for fake news detection are mainly devoted to extracting candidate evidence from comments (or external articles) and establishing interactive reasoning with the news itself to verify the falsehood of the news. However, they still have several drawbacks: 1) The interaction object is coarse-grained, which mainly drives the entire news to participate in interaction, but ignores the learning of potential suspicious segments in news; 2) The reasoning ways are relatively single, making it difficult to explore the various possible correlations between news and candidate evidence. To this end, we propose Unified Evidence Enhancement Inference framework (UEEI) to discover and infer high-quality evidence for detection. Specifically, UEEI first promotes the interaction fusion between comments and news from the perspectives of semantics and emotion, thereby learning potential suspicious fragments in news. Then, the model constructs entity-level and relationship-level retrievals to screen sufficient candidate evidence from external sources. Finally, we measure coherence between suspicious fragments and candidate evidence by multi-view reasoning, and further infer explainable evidence. Experiments on three public datasets confirm the effectiveness and interpretability of our UEEI.

1 Introduction

In recent years, the growth and dissemination of fake news on social media have caused serious negative impacts. Especially in major events, such as the US elections [Grinberg *et al.*, 2019], the Russo-Ukrainian War [Kreft *et al.*, 2023], and the Israeli-Palestinian conflict [Kyriakidou *et al.*, 2023], the deliberate and deceptive dissemination of fake news poses a significant threat to social stability and national security. In addition, the endless emergence of fake news makes it difficult for individuals to distinguish between truth and falsehood at short notice, posing a huge challenge to curb fake news [Van der Linden, 2023]. In light of these impacts and challenges, how to automatically detect fake news has become an imperative problem.

The current studies for fake news detection have made significant progress, which has gone through three stages. In addition

to **feature engineering stage** of manually constructing features, with the popularization of deep learning, **automatic fake news detection** has greatly developed. Most methods design reasonable neural networks to extract features from the perspective of news content, which tends to learn semantic [Qian *et al.*, 2021], emotional [Zhang *et al.*, 2021], stance-based [Xie *et al.*, 2021], and intentional [Zhou *et al.*, 2022] features for detection. Besides, several methods pursue the patterns of news propagation, which establish graph neural networks [Xu *et al.*, 2022] to explore the differences in propagation structures. Considering that the first two stages only give detection results but cannot supply interpretability, there is no evidence to prove news authenticity [Chien *et al.*, 2022]. **Interpretable detection** receives extensive concern, which relies on acquiring knowledge from external sources and interacting with the news to capture corresponding evidence to reveal the errors of fake news [Guo *et al.*, 2023].

Nevertheless, although these methods have improved the model performance, they still have several disadvantages: **1) They overlook the targeted exploration of potential suspicious fragments in news.** The existing methods directly utilize the entire news content and interact with external evidence to uncover conflicting features between both. We know that the false parts in a news may only be hidden in a certain paragraph or sentence. The current methods are coarse-grained in terms of the entire news without specific screening, which makes it difficult to focus on suspicious fragments of the news; **2) The reasoning ways are relatively single.** Existing methods generally devise simple attention interaction or semantic (or entity) alignment as inference strategies to discover consistency features between news content and potential evidence. These strategies mainly focus on global-level semantic similarity, without considering more possible relationships, which are easy to affect the accuracy of model inference.

To this end, we propose a Unified Evidence Enhancement Inference framework (henceforth, UEEI) to explore high-quality evidence from external sources and accurately locate the errors of the news for explainable detection. Specifically, in UEEI, 1) To explore potential suspicious fragments in news content, we develop hierarchical conflict discovery layer (HCD), which first learns key semantics from news and main viewpoints from comments, and then promotes interaction fusion from perspectives of semantic and emotions between key semantics and viewpoints, thereby fine-grained learning suspicious fragments in news; 2) To retrieve candidate evidence sentences

from Wikipedia, external evidence enhancement layer (Ex3) is constructed, which designs different retrieval strategies from both entity-level and relationship-level, ensuring retrieval recall and accuracy, respectively; 3) Finally, we design multi-view coherence inference layer (MCI) that measures the consistency between suspicious fragments and candidate evidence from three aspects of causality, global, and local, which fully considers the possible relationships between the two, and further infers explainable evidence to confirm the false parts of the news. The experiments on three datasets demonstrate the superiority of UEEI. Our main contributions could be summarized as follows:

- A novel and unified evidence inference framework for explainable fake news detection is explored, which uncovers suspicious fragments within the news, obtains candidate evidence from external sources, and further explores valuable evidence to reveal the falsehood of fake news.
- HCD layer explores the interactive fusion between news content and comments from both semantic and emotional views, enabling fine-grained learning of suspicious fragments. Unlike traditional simple reasoning methods, MCI layer comprehensively considers multiple consistency relationships and infers useful explainable evidence.
- Experimental results on three competitive datasets confirm that our UEEI achieves state-of-the-art performance and provides valuable and user-understandable evidence.

2 Related Work

Automatic Fake News Detection The existing methods mainly focus on extracting credibility-indicative features around news content and social context by constructing different neural networks. **Content-based methods** are inclined to learn features around semantics [Qian *et al.*, 2021], emotions [Wan *et al.*, 2023], writing styles [Zhou *et al.*, 2023], and stances [Yang *et al.*, 2022] from news content, relying on BiLSTM [Chen *et al.*, 2023], attention mechanisms [Qian *et al.*, 2021], and large language models [Hu *et al.*, 2023]. **Social context-based methods** rely on graph neural networks to learn structural features of news propagation, thereby learning differences in propagation patterns among different news [Wei *et al.*, 2022]. Dou *et al.* [2021] explored various signals from rich social contexts (e.g., users’ behavior history and social engagements) by joint content and graph modeling to demonstrate the effectiveness of propagation structures. In summary, the methods for automatic detection can deeply learn high-level representations in news content and social contexts to improve detection.

Explainable Fake News Detection The current methods [Wu *et al.*, 2023a; Wang and Shu, 2023] mainly involve retrieving potential evidence from external sources and establishing different inference interaction mechanisms with the unverified news, so as to reveal the erroneous parts of the news. Different inference mechanisms mainly exploit similarity comparison [Yao *et al.*, 2023], semantic matching [Wu *et al.*, 2023a], entity alignment [Krishna *et al.*, 2022], and consistency modeling [Wu *et al.*, 2023c] strategies to respectively learn similarity semantics, common semantic fragments, associated entity information, and global consistency semantics between news and potential evidence. These methods have achieved a certain degree of

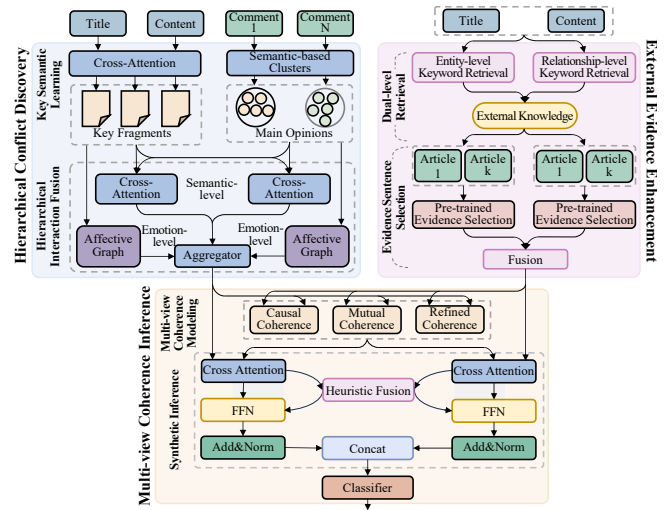


Figure 1: The architecture of UEEI. It explores suspicious semantics between news and comments, extracts candidate evidence from external articles, and establishes coherence inference to learn explainable evidence.

interpretability, but they are still confronted with problems of ignoring the exploration of suspicious segments within the news and relatively single reasoning ways. Thus, we propose a novel evidence inference model that explores suspicious segments in news at a fine-grained level, and considers multiple consistency relationships between suspicious segments and candidate evidence to comprehensively learn effective evidence for detection.

3 The Proposed Model

3.1 Input Encoding Layer

The inputs of UEEI include three types: the title and content of the news, and all comments on the news. For any sequence with n words, it is represented as $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$, where x_t is a d -dimensional vector obtained by pre-training BERT model aiming at the t -th word. For the encoding of each sequence \mathbf{X}_i , we adopt self-attention networks to capture contextual dependencies between words within a sequence, thereby learning the output \mathbf{E}_i . Here, the encoding sequences of title, content, and comments are denoted as \mathbf{E}_T , \mathbf{E}_C , and \mathbf{E}_R , respectively.

3.2 Hierarchical Conflict Discovery Layer (HCD)

With the continuous exposure of news, its comments are easily rich clues questioning the news credibility. To capture them, we construct HCD layer, which first relies on key semantics learning to obtain key semantics of news and main opinions in comments, and then establishes interaction between from both semantic and emotion levels to explore questionable fragments.

Key Semantic Learning

Considering that some news is too long to capture valuable semantics, we build cross-attention networks to facilitate interaction between news title and content for capturing key semantics.

We utilize self-attention mechanism as cross-attention networks to capture the dependencies between any two tokens so

as to learn their contextual information, which is formalized as:

$$\mathbf{H} = \text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{Q}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (1)$$

where \mathbf{Q} , \mathbf{K} , and \mathbf{V} are query, key, and value matrices, respectively. In our settings, we set $\mathbf{Q} = \mathbf{E}_T$ and $\mathbf{K} = \mathbf{V} = \mathbf{E}_C$. d_k is the column in comment encoding matrix. To enhance parallelism of the networks, we employ multi-head attention to linearly project queries, keys, and values:

$$\text{hd}_i = \text{Attention}(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V) \quad (2)$$

$$\mathbf{H}_F = \text{MultiH}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = [\text{hd}_1; \text{hd}_2; \dots; \text{hd}_m]\mathbf{W}^o \quad (3)$$

where \mathbf{W}_i^Q , \mathbf{W}_i^K , \mathbf{W}_i^V , and \mathbf{W}^o are all trainable parameters and ; is concatenation. \mathbf{H}_F is key fragments of the news.

We know that comments generally break down into several opinions. We design semantic-based cluster (i.e., single-pass incremental cluster) to learn mainstream viewpoints in all comments, which does not have any initial setting of clustering values, but determines whether to classify new comments into a new type by a similarity threshold θ_{simi} . In this way, we obtain K clusters with different viewpoints $\mathbf{H}_R = \{\mathbf{H}_{R1}, \mathbf{H}_{R2}, \dots, \mathbf{H}_{RK}\}$.

Hierarchical Interaction Fusion

When encountering differences of opinions, people are prone to emotional conflicts. To identify more questionable fragments from comments, we establish hierarchical interaction between news and comments from semantic- and emotional-level.

Semantic-level Interaction To discover semantic contradictions, we employ cross attention to enhance interaction between key fragments of news and comment mainstream viewpoints:

$$\mathbf{H}'_F = \text{Attention}(\mathbf{H}_F, \mathbf{H}_F, \mathbf{H}_R) \quad (4)$$

$$\mathbf{H}'_R = \text{Attention}(\mathbf{H}_R, \mathbf{H}_R, \mathbf{H}_F), \quad \mathbf{H}_{FR} = [\mathbf{H}'_F; \mathbf{H}'_R] \quad (5)$$

where \mathbf{H}_{FR} is semantic-level questionable fragments.

Emotion-level Interaction The study [Zhang *et al.*, 2022b] finds that news comments contain rich emotional words to deliver opinions. Thus, we devise emotion-level interaction to capture emotional questionable features between comments and news.

Affective Graph Construction. We build affective graph around news and comments separately. Specifically, given a news or comment sequence $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$, we leverage sentiment dictionary SenticNet [Cambria *et al.*, 2020] to evaluate emotional scores between any two words and obtain an adjacency matrix $\mathbf{E}' \in \mathbb{R}^{n \times n}$, where each element $e_{i,j}$ in \mathbf{E}' is:

$$e_{i,j} = |u(x_i) - u(x_j)|, u(x_i) \in [-1, 1] \quad (6)$$

where $u(x_i)$ denotes the emotional score of the i -th word. $|\cdot|$ is an absolute value operation. Especially, the greater the opposition between two emotional words, the higher their corresponding edge weights. In this way, words with conflicting emotions in the sequence will receive sufficient attention.

Additionally, given that emotional words with different contexts may convey different emotions (like the context may contain different degree adverbs, negations, etc.), we exploit syntactic dependency trees (graph forms) to resolve the entire sequence to learn structural features between emotional words.

Each element in dependent adjacency matrix \mathbf{D} is:

$$d_{i,j} = \begin{cases} 1, & v(i,j) = 1 \\ 0, & v(i,j) = 0 \end{cases} \quad (7)$$

where $v(i,j) = 1$ indicates that there is an edge between words x_i and x_j in syntactic dependency tree. $v(i,j) = 0$ means there are no dependency relationships between two words.

To fuse sequence-structure and emotional features, we adopt coordination mechanism to control overall proportion between two graphs, thereby forming emotional enhancement graph.

$$\mathbf{E}^{(l)} = \alpha \mathbf{E}'^{(l)} + (1 - \alpha)\mathbf{D}^{(l)} \quad (8)$$

where α is the hyperparameter. $\mathbf{E}^{(l)}$ is the emotional enhancement graph obtained after l iterations.

Aggregator Fusion. Naturally, emotional information in a sequence is more discrete than semantic features, which may lead to sparse issue in the construction of affective graph. To alleviate it and enhance deep fusion between emotional and semantic-level questionable information, we design two aggregator fusion modes:

1) **Attention Aggregator Fusion.** We use semantic-level questionable information to attend to aggregation of affective graph for exploring common questionable features \mathbf{O}_a .

$$o_i = \mathbf{W}_o(\text{ReLU}(\mathbf{W}'_o(\mathbf{H}_{FR} \parallel \mathbf{E}_i^{(l)}))) \quad (9)$$

$$\beta_i = \text{softmax}(o_i) = \frac{\exp(o_i)}{\sum_{k=1}^N \exp(o_k)} \quad (10)$$

$$\mathbf{O}_a = \sum_{k=1}^N \exp(\beta_i \mathbf{E}_k^{(l)}) \quad (11)$$

where \mathbf{W}_o and \mathbf{W}'_o are trainable parameters. N is the number of the nodes of affective graph.

2) **Max Aggregator Fusion.** Questionable information is more likely to appear in extreme emotional expressions. Therefore, we also explore element-wise maximization to aggregate significant emotion information, which adopts dynamic weight to balance proportion between semantic- and emotion-level features, and then maximizes the mining of major emotionally questionable features.

$$\mu_m = \sigma(\mathbf{H}_{FR} \mathbf{W}_{\mu 1} + \mathbf{E}^{(l)} \mathbf{W}_{\mu 2} + \mathbf{W}_{\mu}) \quad (12)$$

$$\mathbf{O}_m = \text{Tanh}(\mu_m \text{Max}(\mathbf{E}_1^{(l)}, \mathbf{E}_2^{(l)}, \dots, \mathbf{E}_N^{(l)}) \mathbf{W}_m + (1 - \mu_m) \mathbf{E}_k^{(l)} \mathbf{W}_n + \mathbf{b}_m) \quad (13)$$

where all \mathbf{W} and \mathbf{b} are trainable parameters. Finally, we integrate the two fusion modes to maximize questionable information between news and comments, i.e., $\mathbf{O} = [\mathbf{O}_a; \mathbf{O}_m]$.

3.3 External Evidence Enhancement Layer (Ex3)

We first design dual-level keyword retrieval to respectively improve retrieval accuracy and recall, to comprehensively capture relevant articles. Then, evidence selection screens fine-grained sentences from articles as external evidence.

Dual-level Keyword Retrieval

Entity-level Keyword Retrieval Entity-level retrieval endeavors to collect and optimize retrieval keyword seeds, which could retrieve wide-coverage results and ensure retrieval recall.

Step 1: Keyword Matching: 1) Title-guided matching: We employ news title as keyword seeds for retrieval from Wikipedia. 2) Co-occurrence words extracting: Due to title and content as

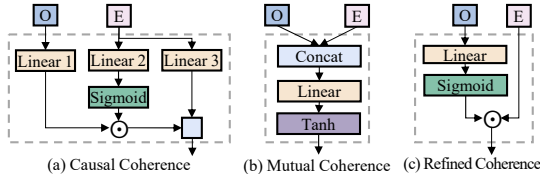


Figure 2: Three coherence modeling in the MCI layer of UEEI.

a whole, we extract their co-occurrence words and sort them by quantity, and choose the top-K words as seeds for retrieval.

Step 2: Keyword Seed Optimization. If there are few feedback results in step 1, keyword seed optimization will be carried out by removing stop words and filtering keywords with low importance in sequence. Filtering way relies on TF-IDF to filter out the least important words for keyword optimization.

Relationship-level Keyword Retrieval It endeavors to boost retrieval accuracy by extracting correlation words in news (like subject-predicate), which are more effective in depicting news opinions and making it easier to retrieve highly relevant articles, so they are more inclined to improve retrieval accuracy.

Inspired by the work [Zhong *et al.*, 2020], we use semantic role labeling (SRL) [Carreras and Màrquez, 2005] to parse news titles and content, establish connections between arguments to construct graphs, and then exploit graph-based reasoning methods to obtain keyword seeds in news, which contain rich association relationships between key semantics. These seeds can be used to search for more accurate related articles. Finally, we integrate the articles retrieved from the two levels to form relevant article set \mathbf{A} .

Evidence Selection

This block aims to extract evidence sentences \mathbf{E}_c from \mathbf{A} .

$$G(\mathbf{X} \cup \mathbf{P}_i) = \mathbf{E}_{c_i} \quad (14)$$

where \mathbf{E}_{c_i} is the i -th sentence of \mathbf{E}_c . Evidence selection could be regarded as a semantic matching task, where each sentence in an article is compared with a given news to determine the probability of that sentence becoming evidence. Here, we utilize BART [Lewis *et al.*, 2020] to improve sentence selection. It inputs news \mathbf{X} and sentences \mathbf{S} of a retrieved article into model in the following way: [CLS] \mathbf{X} [SEP] \mathbf{S} [EOS]. For outputs, following the work [Liao *et al.*, 2023], we employ the BIO form to classify irrelevant tokens as O, where the first token in evidence sentence is set as B evidence, and the rest tokens of a sentence as I evidence. We perform the above operations on all articles, and ultimately obtain the most likely evidence set with p sentences as $\mathbf{E} = \{e_1, e_2, \dots, e_p\}$.

3.4 Multi-view Coherence Inference Layer (MCI)

We propose MCI layer that includes multiple coherence modeling and synthetic inference. The former learns coherence relationships of internal conflicts between news and external evidence to discover different incoherence semantics, and the latter is devoted to fine-grained inferring explainable evidence.

Multiple Coherence Modeling

To reveal the inconsistency between conflicting semantics of news and external evidence, multiple coherence modeling

promotes deep interaction between both. It mainly includes causal coherence, mutual coherence, and refined coherence, which respectively attempt to explore incoherence features of different perspectives: causal, global, and local hotspots.

Three coherence blocks are all extended self-attention mechanism (Eqs. 15 and 16), which focuses on adjusting nonlinear transformation $f(\cdot)$ to adapt to different coherence strategies.

$$\text{att}_i = \text{softmax}(v_i^T f(h_i)) \quad (15)$$

$$r = \sum_{i=1}^n \text{att}_i h_i \quad (16)$$

Causal Coherence. It primarily maps the conflicting information of news and external potential evidence into the same dimensional space, and then conducts interactive comparisons to explore the causal relationship between them.

$$\alpha'(\mathbf{O}) = \mathbf{W}_\alpha \mathbf{O} + \mathbf{b}_\alpha, \beta(\mathbf{E}) = \mathbf{W}_\beta \mathbf{E} + \mathbf{b}_\beta, \gamma(\mathbf{E}) = \mathbf{W}_\gamma \mathbf{E} + \mathbf{b}_\gamma \quad (17)$$

$$\mathbf{G}_c = f_c(\mathbf{O}, \mathbf{E}) = \sigma(\beta(\mathbf{E})) \odot \alpha'(\mathbf{O}) + \gamma(\mathbf{E}) \quad (18)$$

where all \mathbf{W} and \mathbf{b} are parameters and \odot is element-wise multiplication.

Mutual Coherence. To discover the incoherence features between conflicting information and potential evidence from a global perspective, we design mutual coherence that maps and transforms them together to learn their inconsistent features:

$$\mathbf{G}_m = f_m(\mathbf{O}, \mathbf{E}) = \tanh(\mathbf{W}_m[\mathbf{O}||\mathbf{E}] + \mathbf{b}_m) \quad (19)$$

Refined Coherence. For highly focused conflicting information, we explore refined coherence, which first highly condenses the conflicting information in the news, and then further interacts with external evidence to refine their inconsistent features:

$$\mathbf{G}_r = f_r(\mathbf{O}, \mathbf{E}) = \sigma(\mathbf{W}_r \mathbf{E} + \mathbf{b}_r) \odot \mathbf{O} \quad (20)$$

where \mathbf{W}_r and \mathbf{b}_r are trainable parameters. Thus, the above is integrated as multi-view consistent semantics \mathbf{G} between conflicting information and external evidence: $\mathbf{G} = \mathbf{G}_c + \mathbf{G}_m + \mathbf{G}_r$.

Synthetic Inference

To further synthesize explainable evidence, we propose synthetic inference to drive contextual association between multi-view consistent semantics and external evidence.

Cross-Attention Networks. It explores common false parts of news \mathbf{H}_{go} and highly-correlated evidence \mathbf{H}_{ge} respectively by deep interaction association of different semantics.

$$\mathbf{H}_{go} = \text{Attention}(\mathbf{G}, \mathbf{G}, \mathbf{O}), \mathbf{H}_{ge} = \text{Attention}(\mathbf{G}, \mathbf{G}, \mathbf{E}) \quad (21)$$

Heuristic Fusion. It establishes full integration between common false parts and highly-correlation evidence, thereby compounding accurately interpretable evidence.

$$\mathbf{H}_{goge} = [\mathbf{H}_{go}; |\mathbf{H}_{go} - \mathbf{H}_{ge}|; \mathbf{H}_{go} \odot \mathbf{H}_{ge}; \mathbf{H}_{ge}] \quad (22)$$

Next, we apply feed forward network to the fusion results for adding non-linear features while scale-invariant features, which involves a single hidden layer with a ReLU.

$$\mathbf{I}_{go} = \text{FFN}(\mathbf{H}_{go}, \mathbf{H}_{goge}) = \max(0, [\mathbf{H}_{go}; \mathbf{H}_{goge}] \mathbf{W}_1 + \mathbf{b}_1) \mathbf{W}_2 + \mathbf{b}_2 \quad (23)$$

$$\mathbf{I}_{ge} = \text{FFN}(\mathbf{H}_{ge}, \mathbf{H}_{goge}) = \max(0, [\mathbf{H}_{ge}; \mathbf{H}_{goge}] \mathbf{W}_3 + \mathbf{b}_3) \mathbf{W}_4 + \mathbf{b}_4 \quad (24)$$

$$\mathbf{I}_{goe} = [\mathbf{I}_{go}; \mathbf{I}_{ge}] \quad (25)$$

Finally, we adopt softmax function to emit probability distribution for training, where a loss drives the model to minimize

| | PolitiFact | GossipCop | Weibo |
|------------|------------|-----------|-------|
| #True News | 399 | 4,219 | 436 |
| #Fake News | 345 | 3,393 | 311 |
| Total | 744 | 7,612 | 747 |

Table 1: Statistics of three publicly available datasets.

cross-entropy error for a training instance with golden label y :

$$\text{Loss} = - \sum y \log p \tag{26}$$

$$p = \text{softmax}(\mathbf{W}_p \mathbf{I}_{goe} + \mathbf{b}_p) \tag{27}$$

4 Experiments

In this section, we endeavor to answer the following questions:

- **Q1:** Could UEEI achieve more excellent performance?
- **Q2:** Does each layer contribute to improving detection?
- **Q3:** How much does the exploration of potentially suspicious fragments of news boost model performance?
- **Q4:** What are the advantages of our multi-view coherence inference compared to existing reasoning ways?
- **Q5:** Is the obtained evidence reasonable and interpretable?

4.1 Datasets and Evaluation Metrics

As shown in Table 1, PolitiFact and GossipCop are two English datasets [Shu *et al.*, 2020] and Weibo is a Chinese dataset [Liu *et al.*, 2018]. **PolitiFact** is a factual verification website that deals with political topics and rating credibility of voter officials’ claims. This dataset mainly includes two types of labels: real and fake. The labels are evaluated by professional journalists. **GossipCop** is collected from the Gossip Cop website for factual checks of celebrity reports. The labels are the same as PolitiFact dataset. **Weibo** is collected from Sina Weibo platform, which mainly involves some hot news and is marked as true news and rumors. **Evaluation Metrics.** We leverage F1-macro (F1-Ma), and F1-micro (F1-Mi), and F1 to measure our model. In dataset partitioning, we hold out 75% of the news as training set and the remaining 25% as test set.

4.2 Settings

For preprocessing, we tokenize sequences and convert tokens to lowercase, filtering out nonalphanumeric characters. For model configuration, we adopt BERT-base as embeddings. The embedding size d is 768. K clusters of viewpoints are from [3, 4, 5], and α is 0.65. K in the top- K words is 10; In self-attention networks, attention heads and blocks are set to 6 and 4, respectively, and the dropout of multi-head attention is 0.6. We adopt Adam optimizer as the model optimizer. The learning rate is uniformly set to 10^{-4} . We utilize L2-regularizers with the fully-connected layers and the mini-batch size is 32.

4.3 Performance Comparison (Q1)

Comparative Baselines

We compare UEEI with the following baselines: **TextRNN** [Yang *et al.*, 2016] employs LSTM to learn contextual representations for detection. **TCNN-URG** [Qian *et al.*, 2018] integrates

user responses into two-level CNN to learn contextual semantics to detect fake news. **BERT** [Kenton and Toutanova, 2019] focuses on gaining long-term dependencies within news for detection. **DeClarE** [Popat *et al.*, 2018] explores potential evidence between claims and articles for debunking claims. **HAN** [Ma *et al.*, 2019] is hierarchical attention networks, which learns coherence between news and related articles to obtain interpretability. **EHAN** [Wu *et al.*, 2021] builds inference networks to identify semantic conflicts from related articles as evidence. **MAC** [Vo and Lee, 2021] leverages dual-view attention to explore word-level and document-level evidence for detection. **GET** [Xu *et al.*, 2022] develops graph networks to obtain semantic structures and model evidence to detect fake news. **MUSER** [Liao *et al.*, 2023] builds multi-step retrieval and learns dependencies between pieces of evidence for detection.

Overall Performance

As shown in Table 2, we observe that:

- In the methods of automatic feature learning (i.e., the first three baselines), TCNN-URG achieves the optimal performance, indicating that relying on comments as external features can markedly improve model performance.
- Compared to automatic methods, the methods of exploring evidence (like DeClarE) achieve better performance, showing the improvements from 0.8% to 4.9% in F1-Mi on the three datasets. This reflects the fact that capturing conflicting semantics between news and external articles as credibility indicators is conducive to facilitating detection.
- Our UEEI consistently outperforms all baselines, showing up to 3.4% improvements in all metrics on three datasets than the latest baseline (MUSER). This not only highlights the superiority of our model, but also demonstrates that the route of exploring fine-grained error parts of the unverified news and capturing high-quality evidence is effective.

4.4 Discussion

4.5 Ablation Study (Q2)

To assess the necessity of each module in our model, we ablate UEEI into: **-HCD**, **-Ex3**, and **-MCI** are the removal of HCD, Ex3, and MCI layers. **-Ex3(entity)**, **-Ex3(relation)**, and **-Ex3(select)** are denoted as entity-level retrieval, relationship-level retrieval, and evidence selection blocks ablated from UEEI, respectively. As shown in Table 3, we could observe that:

- The separation of MCI results in a significant change, with up to 5.2% degradation on three datasets, which sufficiently demonstrates the value of establishing consistency reasoning from the perspectives of causality, global, and local.
- In Ex3 layer, the removal of entity- and relationship-level retrieval blocks results in varying degrees of model degradation, which confirms respectively the effectiveness of both boosting retrieval recall and accuracy through entity and relationship aspects. **-Ex3(select)** is weaker than our model, which reveals the rationality of extracting key evidence sentences from documents.
- Overall, the ablation of each module weakens the overall performance of the model, showing a decrease from 1.4% to 5.2% in F1-Ma on three datasets, which reveals the effectiveness of various modules that constitute our model.

| Methods | PolitiFact | | | | GossipCop | | | | Weibo | | | |
|-------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | F1-Ma | F1-Mi | F1-T | F1-F | F1-Ma | F1-Mi | F1-T | F1-F | F1-Ma | F1-Mi | F1-T | F1-F |
| TextRNN | 61.0 | 60.9 | 61.6 | 60.3 | 62.9 | 62.8 | 63.6 | 62.0 | 74.1 | 73.7 | 77.1 | 70.1 |
| TCNN-URG | 62.1 | 61.9 | 63.7 | 60.1 | 64.4 | 64.3 | 65.0 | 63.6 | 70.9 | 70.4 | 74.1 | 66.7 |
| BERT | 59.7 | 59.8 | 60.8 | 58.6 | 61.7 | 61.3 | 63.5 | 57.8 | 69.9 | 69.8 | 71.9 | 67.8 |
| DeClarE | 65.4 | 65.1 | 65.6 | 65.1 | 66.0 | 65.7 | 68.6 | 62.9 | 74.6 | 74.5 | 76.5 | 72.4 |
| HAN | 66.1 | 66.0 | 67.9 | 64.3 | 70.2 | 70.0 | 72.2 | 67.8 | 68.9 | 68.7 | 71.1 | 66.2 |
| EHIAN | 66.4 | 66.3 | 67.4 | 65.0 | 70.5 | 70.2 | 73.1 | 67.3 | 75.3 | 75.2 | 77.0 | 73.4 |
| MAC | 67.8 | 67.5 | 70.0 | 65.3 | 72.9 | 72.7 | 72.5 | 70.5 | 73.4 | 73.2 | 70.9 | 75.5 |
| GET | 69.4 | 69.2 | 72.5 | 66.9 | 73.3 | 73.1 | 75.1 | 71.2 | 73.3 | 73.1 | 75.1 | 71.2 |
| MUSER | 73.2 | 72.9 | 75.7 | 70.2 | 77.6 | 77.5 | 78.4 | 76.8 | 75.6 | 75.4 | 77.6 | 73.0 |
| UEEI(Ours) | 76.4 | 75.6 | 78.9 | 73.6 | 80.4 | 80.1 | 81.0 | 79.2 | 82.7 | 82.7 | 85.0 | 81.7 |

Table 2: Performance comparison of UEEI against the baselines on the three datasets.

| Methods | PolitiFact | | | | GossipCop | | | | Weibo | | | |
|----------------|------------|-------|-------|-------|-----------|-------|-------|-------|-------|-------|-------|-------|
| | F1-Ma | F1-Mi | F1-T | F1-F | F1-Ma | F1-Mi | F1-T | F1-F | F1-Ma | F1-Mi | F1-T | F1-F |
| -HCD | 0.729 | 0.715 | 0.755 | 0.706 | 0.771 | 0.768 | 0.776 | 0.772 | 0.798 | 0.801 | 0.814 | 0.786 |
| -Ex3 | 0.713 | 0.708 | 0.741 | 0.690 | 0.756 | 0.754 | 0.762 | 0.756 | 0.783 | 0.784 | 0.799 | 0.771 |
| -MCI | 0.718 | 0.704 | 0.746 | 0.695 | 0.760 | 0.759 | 0.766 | 0.762 | 0.787 | 0.790 | 0.804 | 0.775 |
| -Ex3(entity) | 0.745 | 0.740 | 0.775 | 0.722 | 0.791 | 0.787 | 0.795 | 0.779 | 0.813 | 0.814 | 0.832 | 0.803 |
| -Ex3(relation) | 0.736 | 0.734 | 0.757 | 0.695 | 0.780 | 0.778 | 0.785 | 0.769 | 0.801 | 0.802 | 0.824 | 0.793 |
| -Ex3(select) | 0.741 | 0.737 | 0.771 | 0.717 | 0.785 | 0.781 | 0.790 | 0.773 | 0.805 | 0.808 | 0.828 | 0.798 |
| UEEI | 0.764 | 0.756 | 0.789 | 0.736 | 0.804 | 0.801 | 0.809 | 0.795 | 0.827 | 0.827 | 0.850 | 0.817 |

Table 3: Results of ablation analysis of our UEEI on the three datasets.

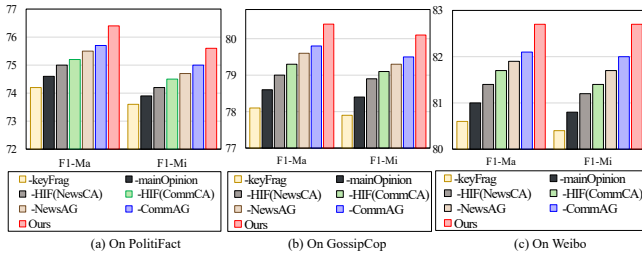


Figure 3: Performance evaluation of internal structures of HCD layer.

Evaluation of Hierarchical Conflict Discovery (Q3)

To analyze the effectiveness of the model’s fine-grained learning of suspicious fragments in news, we explore the contribution of each block in HCD layer. We divide contributions into two categories: semantic-based and emotion-based. Semantic-based contribution: **-keyFrag**, **-mainOpinion**, **-HIF(NewsCA)**, **-HIF(CommCA)** represent respectively that the model replaces news key fragments, main opinions in comments, as well as cross-attention aiming to news content and comments in hierarchical interaction fusion as concatenation operations. Emotion-based contribution: **-NewsAG** and **-CommAG** are the replacement of affective graphs for news and comments by BiLSTM, respectively. As shown in Figure 3, we observe that:

- The detachment of modules related to suspicious fragments induces varying degrees of degradation. Especially, the removal of clustering algorithm aiming at comments has been greatly affected. These reflect the superiority of capturing

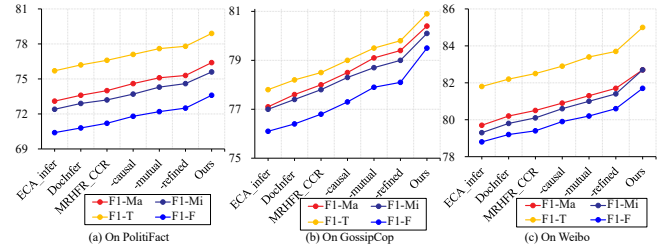


Figure 4: Comparison of our MCI layer with other inference strategies.

suspicious fragments and utilizing comments.

- The alteration of affective graph contributes to a decrease of at least 0.6% in F1-Mi, indicating that using graph neural networks (GNN) is more effective in learning emotions than sequence models (BiLSTM). The reason is that emotional words in news or comments are discrete, and they are not just simple context relationships, and we can extract complex long-range correlations between emotion words using GNN.
- The changes in both types of modules cause a reduction in performance, which illustrates the effectiveness of each module and mutual promotion of semantic and emotional layers in extracting suspicious fragments in the news.

Superiority of Multi-view Coherence Inference Layer (Q4)

To further validate the superiority of MCI layer, we replace it with advanced inference modules: **ECA_infer** [Zhang *et al.*, 2022a] is that two extended attentions with feed-forward networks perform feature- and relation-level inference. **DocInfer**

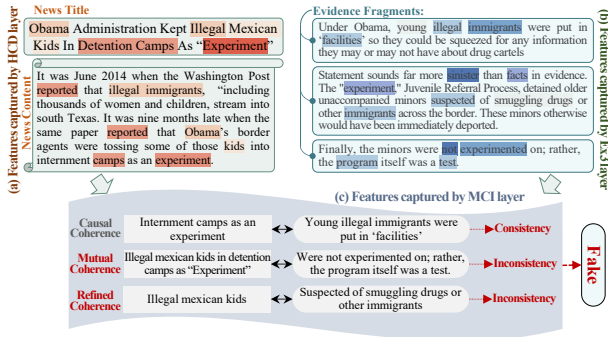


Figure 5: Visualization of explainable features learned by our model.

[Mathur *et al.*, 2022] is a textual inference module based on document graph, which adopts reinforcement learning to provide evidence. **MRHFR_CCR** [Wu *et al.*, 2023b] is coherence constraint reasoning for inferring coherence between two sequences. Additionally, we also decompose MCI layer to analyze the contribution of its internal structures. Specifically, **-causal**, **-mutual**, and **-refined** respectively denote that causal coherence, mutual coherence, and refined coherence are removed from multi-view coherence inference layer. As shown in Figure 4, we observe that: 1) By comparing with the first three coherence inference models, our model achieves more unparalleled performance, showing at least 2.4% improvement in F1-Ma, which demonstrates the superiority of our MCI layer. 2) The removal of each module in MCI layer results in a substantial performance decline, which reflects their effectiveness and organic synergy. we analyze the reasons as: Our multi-view coherence mechanisms could capture multiple potential relationships, and our synthetic inference layer could retain their significant parts to improve detection ability.

4.6 Case Study (Q5)

We visualize outputs of UEEI to perceive interpretability and introduce user participation to verify comprehensibility of the evidence.

Visualization of Explainable Features Learned by UEEI

To display the interpretability of our model, we visualize the features captured by its different layers, and further compare the final outputs with traditional reasoning strategies. The learned features are shown in Figure 5, and we observe that:

- HCD not only captures key semantics of the news (like "illegal Mexican kids"), but also gains conflicts ("experiment"), which confirms that our model could finely capture suspicious segments. Ex3 layer explores multiple potential evidence, such as "young illegal immigrants were put in 'facilities'" and "the minors were not experimented", which lively presents the effectiveness of our model in learning evidence.
- The outputs of MCI give greater weight to the most-likely erroneous fragments ("kids in detention camps as experiment") and strengthen the evidence, such as "not experimented on; the program itself was a test". However, traditional reasoning only focuses on shared features ('internment camps as an experiment'), without providing sufficient evidence. These fully confirm that our model extends higher-quality interpretability.

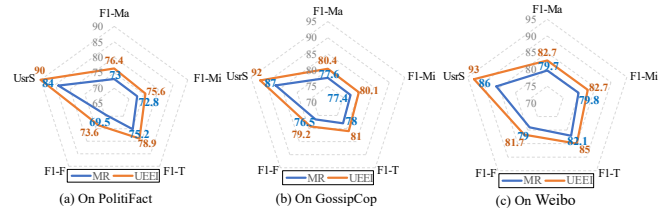


Figure 6: Evaluation of UEEI and MR. UsrS is user satisfactory [0, 100].

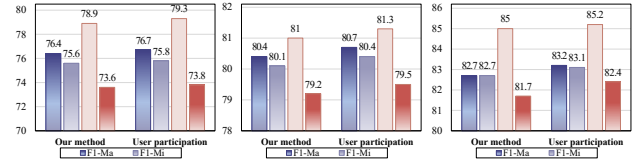


Figure 7: Performance comparison between our UEEI and the method of user participation in evidence retrieval (i.e., human in the loop).

Model Validation with User Participation

To validate the comprehensibility of the evidence, we study from two perspectives: 1) Users rely on candidate evidence captured by UEEI and the latest retrieval way (MR) [Liao *et al.*, 2023] to rate the unverified news for verifying the two types of evidence; 2) Adopting user participation to manually select more favorable evidence from candidate evidence to downstream for measuring the effectiveness of man-in-the-loop. In detail, we randomly choose 120 news from three datasets and retrieve Top-5 evidence articles, with 5 real-world users as participants. As shown in Figures 6 and 7, and we observe that: From Figure 6, compared to MR, users gain better performance in classifying news based on our captured evidence, showing at least a 2.7% boost in F1-Mi, and our user satisfaction degree is significantly higher than that of MR, which demonstrates that our retrieved evidence is easier to understand and interpretable. From Figure 7, evidence manually screened by users outperforms evidence automatically retrieved by UEEI, with improvements of up to 0.7% on the three datasets, which confirms the effectiveness of user participation in detecting fake news. Furthermore, human-in-the-loop strategy is time-consuming and labor-intensive, and we need to carefully consider the combination of automatic detection and user participation.

5 Conclusion

In this paper, we propose unified evidence enhancement inference model (UEEI) for fake news detection based on the progressive process of suspicious fragment learning, evidence enhancement, and coherence inference. We first promote hierarchical interaction between news and comments to explore internal suspicious fragments, then enhance external evidence retrieval around two levels, and finally drive multi-view coherence learning between suspicious semantics and external evidence, thereby inferring false parts of news. Experiments on three datasets demonstrate the effectiveness and interpretability of our UEEI. In the future, considering current prevalence of multimodal news, we seek to design inference mechanisms to explore inconsistency between multimodal news for detection.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants U22B2036 and 62202381, in part by Shenzhen Science and Technology Program and Guangdong Basic and Applied Basic Research Foundation (2021A1515110717, 2024A1515010087), General Program of Chongqing Natural Science Foundation (No. CSTB2022NSCQ-MSX1284), Sponsored by CAAI-Huawei MindSpore Open Fund, the National Postdoctoral Innovative Talents Support Program for L. Wu. We would like to thank the anonymous reviewers for their constructive comments.

References

- [Cambria *et al.*, 2020] Erik Cambria, Yang Li, Frank Z Xing, Soujanya Poria, and Kenneth Kwok. Senticnet 6: Ensemble application of symbolic and subsymbolic ai for sentiment analysis. In *Proceedings of the 29th ACM international conference on information & knowledge management*, pages 105–114, 2020.
- [Carreras and Màrquez, 2005] Xavier Carreras and Lluís Màrquez. Introduction to the conll-2005 shared task: Semantic role labeling. In *Proceedings of the ninth conference on computational natural language learning (CoNLL-2005)*, pages 152–164, 2005.
- [Chen *et al.*, 2023] Mu-Yen Chen, Yi-Wei Lai, and Jiunn-Woei Lian. Using deep learning models to detect fake news about covid-19. *ACM Transactions on Internet Technology*, 23(2):1–23, 2023.
- [Chien *et al.*, 2022] Shih-Yi Chien, Cheng-Jun Yang, and Fang Yu. Xflag: Explainable fake news detection model on social media. *International Journal of Human-Computer Interaction*, 38(18-20):1808–1827, 2022.
- [Dou *et al.*, 2021] Yingdong Dou, Kai Shu, Congying Xia, Philip S Yu, and Lichao Sun. User preference-aware fake news detection. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2051–2055, 2021.
- [Grinberg *et al.*, 2019] Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. Fake news on twitter during the 2016 us presidential election. *Science*, 363(6425):374–378, 2019.
- [Guo *et al.*, 2023] Hao Guo, Weixin Zeng, Jiuyang Tang, and Xiang Zhao. Interpretable fake news detection with graph evidence. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pages 659–668, 2023.
- [Hu *et al.*, 2023] Beizhe Hu, Qiang Sheng, Juan Cao, Yuhui Shi, Yang Li, Danding Wang, and Peng Qi. Bad actor, good advisor: Exploring the role of large language models in fake news detection. *arXiv preprint arXiv:2309.12247*, 2023.
- [Kenton and Toutanova, 2019] Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, pages 4171–4186, 2019.
- [Kreft *et al.*, 2023] Jan Kreft, Monika Boguszewicz-Kreft, and Daria Hliebova. Under the fire of disinformation. attitudes towards fake news in the ukrainian frozen war. *Journalism Practice*, pages 1–21, 2023.
- [Krishna *et al.*, 2022] Amrith Krishna, Sebastian Riedel, and Andreas Vlachos. Proofver: Natural logic theorem proving for fact verification. *Transactions of the Association for Computational Linguistics*, 10:1013–1030, 2022.
- [Kyriakidou *et al.*, 2023] Maria Kyriakidou, Marina Morani, Stephen Cushion, and Ceri Hughes. Audience understandings of disinformation: navigating news media through a prism of pragmatic scepticism. *Journalism*, 24(11):2379–2396, 2023.
- [Lewis *et al.*, 2020] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2020.
- [Liao *et al.*, 2023] Hao Liao, Jiahao Peng, Zhanyi Huang, Wei Zhang, Guanghua Li, Kai Shu, and Xing Xie. Muser: A multi-step evidence retrieval enhancement framework for fake news detection. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 4461–4472, 2023.
- [Liu *et al.*, 2018] Qiang Liu, Feng Yu, Shu Wu, and Liang Wang. Mining significant microblogs for misinformation identification: an attention-based approach. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 9(5):1–20, 2018.
- [Ma *et al.*, 2019] Jing Ma, Wei Gao, Shafiq Joty, and Kam-Fai Wong. Sentence-level evidence embedding for claim verification with hierarchical attention networks. Association for Computational Linguistics, 2019.
- [Mathur *et al.*, 2022] Puneet Mathur, Gautam Kunapuli, Riyaz Bhat, Manish Shrivastava, Dinesh Manocha, and Maneesh Singh. Docinfer: Document-level natural language inference using optimal evidence selection. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 809–824, 2022.
- [Popat *et al.*, 2018] Kashyap Popat, Subhabrata Mukherjee, Andrew Yates, and Gerhard Weikum. Declare: Debunking fake news and false claims using evidence-aware deep learning. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2018.
- [Qian *et al.*, 2018] Feng Qian, Chengyue Gong, Karishma Sharma, and Yan Liu. Neural user response generator: fake news detection with collective user intelligence. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 3834–3840, 2018.
- [Qian *et al.*, 2021] Shengsheng Qian, Jinguang Wang, Jun Hu, Quan Fang, and Changsheng Xu. 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, 2021.
- [Shu *et al.*, 2020] Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big data*, 8(3):171–188, 2020.
- [Van der Linden, 2023] Sander Van der Linden. *Foolproof: Why misinformation infects our minds and how to build immunity*. WW Norton & Company, 2023.
- [Vo and Lee, 2021] Nguyen Vo and Kyumin Lee. Hierarchical multi-head attentive network for evidence-aware fake news detection. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 965–975, 2021.
- [Wan *et al.*, 2023] Mingyu Wan, Yin Zhong, Xuefeng Gao, Sophia Yat Mei Lee, and Chu-Ren Huang. Fake news, real emotions: Emotion analysis of covid-19 infodemic in weibo. *IEEE Transactions on Affective Computing*, 2023.
- [Wang and Shu, 2023] Haoran Wang and Kai Shu. Explainable claim verification via knowledge-grounded reasoning with large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 6288–6304, 2023.
- [Wei *et al.*, 2022] Lingwei Wei, Dou Hu, Wei Zhou, and Songlin Hu. Uncertainty-aware propagation structure reconstruction for fake news detection. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 2759–2768, 2022.
- [Wu *et al.*, 2021] Lianwei Wu, Yuan Rao, Ling Sun, and Wangbo He. Evidence inference networks for interpretable claim verification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 14058–14066, 2021.
- [Wu *et al.*, 2023a] Junfei Wu, Weizhi Xu, Qiang Liu, Shu Wu, and Liang Wang. Adversarial contrastive learning for evidence-aware fake news detection with graph neural networks. *IEEE Transactions on Knowledge and Data Engineering*, 2023.
- [Wu *et al.*, 2023b] Lianwei Wu, Pusheng Liu, and Yanning Zhang. See how you read? multi-reading habits fusion reasoning for multi-modal fake news detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 13736–13744, 2023.
- [Wu *et al.*, 2023c] Lianwei Wu, Pusheng Liu, Yongqiang Zhao, Peng Wang, and Yangning Zhang. Human cognition-based consistency inference networks for multi-modal fake news detection. *IEEE Transactions on Knowledge and Data Engineering*, 2023.
- [Xie *et al.*, 2021] Jianhui Xie, Song Liu, Ruixin Liu, Yinghong Zhang, and Yuesheng Zhu. Sern: Stance extraction and reasoning network for fake news detection. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2520–2524. IEEE, 2021.
- [Xu *et al.*, 2022] Weizhi Xu, Junfei Wu, Qiang Liu, Shu Wu, and Liang Wang. Evidence-aware fake news detection with graph neural networks. In *Proceedings of the ACM Web Conference 2022*, pages 2501–2510, 2022.
- [Yang *et al.*, 2016] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, pages 1480–1489, 2016.
- [Yang *et al.*, 2022] Ruichao Yang, Jing Ma, Hongzhan Lin, and Wei Gao. A weakly supervised propagation model for rumor verification and stance detection with multiple instance learning. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1761–1772, 2022.
- [Yao *et al.*, 2023] Barry Menglong Yao, Aditya Shah, Lichao Sun, Jin-Hee Cho, and Lifu Huang. End-to-end multimodal fact-checking and explanation generation: A challenging dataset and models. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2733–2743, 2023.
- [Zhang *et al.*, 2021] Xueyao Zhang, Juan Cao, Xirong Li, Qiang Sheng, Lei Zhong, and Kai Shu. Mining dual emotion for fake news detection. In *Proceedings of the web conference 2021*, pages 3465–3476, 2021.
- [Zhang *et al.*, 2022a] Liang Zhang, Jinsong Su, Yidong Chen, Zhongjian Miao, Min Zijun, Qingguo Hu, and Xiaodong Shi. Towards better document-level relation extraction via iterative inference. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 8306–8317, 2022.
- [Zhang *et al.*, 2022b] Wei Zhang, Wanling Zhu, Jia Nie, Frank Andrasik, and Xara Naomi Blom. The effect of emotion background on pathological internet users’ comments on online news: Evidence from online text analysis. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 16(4), 2022.
- [Zhong *et al.*, 2020] Wanjun Zhong, Jingjing Xu, Duyu Tang, Zenan Xu, Nan Duan, Ming Zhou, Jiahai Wang, and Jian Yin. Reasoning over semantic-level graph for fact checking. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6170–6180, 2020.
- [Zhou *et al.*, 2022] Xinyi Zhou, Kai Shu, Vir V Phoha, Huan Liu, and Reza Zafarani. “this is fake! shared it by mistake”: Assessing the intent of fake news spreaders. In *Proceedings of the ACM Web Conference 2022*, pages 3685–3694, 2022.
- [Zhou *et al.*, 2023] Xinyi Zhou, Jiayu Li, Qinzhou Li, and Reza Zafarani. Linguistic-style-aware neural networks for fake news detection. *arXiv preprint arXiv:2301.02792*, 2023.