RoE-FND: A Case-Based Reasoning Approach with Dual Verification for Fake News Detection via LLMs

Anonymous ACL submission

Abstract

The proliferation of deceptive content online necessitates robust Fake News Detection (FND) systems. While evidence-based approaches leverage external knowledge to verify claims, existing methods face critical limitations: noisy evidence selection, generalization bottlenecks, and unclear decision-making processes. Recent efforts to harness Large Language Models (LLMs) for FND introduce new challenges, including hallucinated rationales and conclusion bias. To address these issues, we propose **RoE-FND** (**R**eason **o**n **E**xperiences FND), a framework that reframes evidence-based FND as a logical deduction task by synergizing LLMs with experiential learning. RoE-FND encompasses two stages: (1) *self-reflective knowledge building*, where a knowledge base is curated by analyzing past reasoning errors, namely the exploration stage, and (2) *dynamic criterion retrieval*, which synthesizes task-specific reasoning guidelines from historical cases as experiences during deployment. It further cross-checks rationales against internal experience through a devised dual-channel procedure. Key contributions include: a case-based reasoning framework for FND that addresses multiple existing challenges, a training-free approach enabling adaptation to evolving situations, and empirical validation of the framework's superior generalization and effectiveness over state-of-the-art methods across three datasets.

1 Introduction

Contemporary media platforms, such as news feeds and microblogs, are witnessing a growing prevalence of deceptive and manipulative material. This includes dubious assertions, "alternative facts", or even entirely fabricated news stories (Shu et al., 2017; Fisher et al., 2016). The proliferation of such content erodes public trust and exacerbates societal polarization, making automated Fake News Detection (FND) systems a critical line of defense (Shu et al., 2020). Early methods relied on shallow textual features like lexical statistics (Castillo et al., 2011) or syntactic patterns (Feng et al., 2012). Recent works that based on deep-learning techniques prepare adequate news samples from the real world, including authentic news and those suspicious or fabricated (Shu et al., 2020; Popat et al., 2017; Jin et al., 2017). Researchers can thus develop methods to capture semantic patterns of deception (Zhang et al., 2021; Kaliyar et al., 2021; Zhu et al., 2022). However, these approaches lack support from related factual information, limiting their adaptability to evolving manipulation tactics.

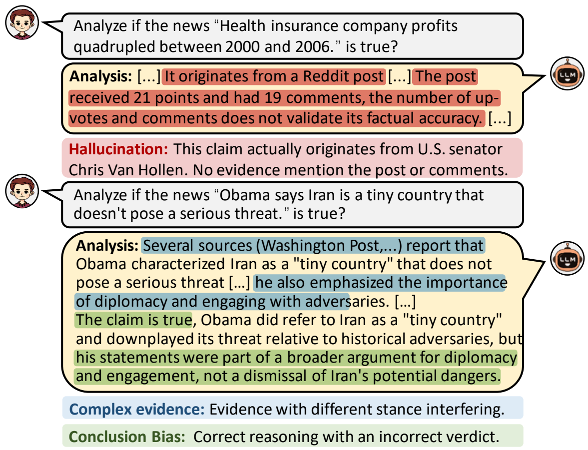

Figure 1: Examples of common mistakes made by LLMs when analyzing the news's authenticity. We omit the less important content using [...].

This challenge has spurred interest in evidence-based FND, where models verify claims against external knowledge sources (Li et al., 2016; Wu et al., 2021a). Current approaches mainly employ two strategies: 1) Evidence filtering via lexical similarity (Rashkin et al., 2017) or neural retrievers (Yang et al., 2022), and 2) Learning joint representations of news-evidence pairs (Ma et al., 2019; Wu et al., 2021a). While these methods achieve domain-specific success, our analysis reveals critical limitations. Firstly, current evidence selection mechanisms lack explicit reasoning traces, making them vulnerable to noisy or adversarial evidence. Secondly, models are trained on platform-biased or domain-specific datasets, this brings a generalization bottleneck that causes performance drops when tested on news from different domains or platforms (Zhu et al., 2022). Thirdly, the interpretability gap brought by end-to-end neural architectures obscures the decision logic, hindering practical deployment in high-stakes scenarios.

Large Language Models (LLMs) offer promising potential for addressing these challenges with their emergent textual reasoning capabilities (Wei et al., 2022a) and zero-shot generalization ability that requires no model parameters tuning (Kojima et al., 2022). Recent attempts directly employ LLMs as fact-checkers (Pan et al., 2023b) or claim verifiers (Caramancion, 2023), yet introduce new challenges. For instance, LLMs may generate plausible but factually incorrect rationales, a phenomenon known as hallucinated reasoning (Huang et al., 2023). Besides, harder than fact-checking tasks that only require determining the evidence's stance (Thorne et al., 2018), evidence-based FND via LLMs often struggle with complex evidence synthesis that may contain supporting, refuting, and misguiding materials. Thirdly, recent research reveals the conclusion bias problem where correct reasoning inexplicably leads to incorrect verdicts (Hu et al., 2024). We present two examples in Figure 1 illustrating these challenges. These limitations highlight the need for a principled framework that can harness the reasoning power of LLMs while mitigating their vulnerabilities.

In this paper, we reframe evidence-based FND as a logical deduction task and present a framework, namely RoE-FND (**R**eason on **E**xperiences FND), that employs LLMs as logical reasoning units and synergize their power with case-based experiential learning. Unlike prior LLM-based approaches that perform one-off verification (Pan et al., 2023b; Caramancion, 2023), RoE-FND comprises two stages and introduces several key innovations: 1) *Self-reflective experience building*: during the exploration stage, the model constructs a knowledge base through self-questioning the wrong answer. 2) *Dynamic criterion retrieval*: at the deployment stage, task-specific advice is dynamically synthesized from relevant historical cases, assisting in finding better rationales. 3) *Dual-channel verification mechanism*: RoE-FND cross-checks generated rationales against both external evidence and internal experience patterns.

We conduct extensive experiments on three challenging datasets, i.e. CHEF (Hu et al., 2022), Snopes (Popat et al., 2018), and PolitiFact (Shu et al., 2020) to validate the effectiveness of RoE-FND. The results demonstrate improvements in various metrics over state-of-the-art methods. We also propose a fine-tuning strategy for RoE-FND, which brings significant improvements. Quantitative results and analysis of generated content underscore the advantages of our design, particularly in terms of robustness, interpretability, and generalization. In summary, our work makes several key contributions:

- We propose RoE-FND, formalizing experiential reasoning for FND through a case-based strategy. It can produce accurate explanations for predictions of the news's authenticity.
- We devise a training-free strategy that incorporates self-reflective experience curation and dynamic criterion adaptation, offering a novel approach to leveraging LLMs for FND.
- Extensive experiments on multiple settings validate the advantages of our framework. Detailed studies offer insights into the mechanisms driving the framework's success.

2 Related Works

2.1 Evidende-based Fake News Detection

Evidence-based fake news detection conducts knowledge comparison between news and relevant evidence materials (Zhou and Zafarani, 2020). Retrieved materials are usually without further relevance checking. Famous datasets of the field, Snopes (Popat et al., 2017) and PolitiFact (Rashkin et al., 2017) comply with this retrieval pipeline using Microsoft Bing API to retrieve evidence. Many customized detection approaches have been suggested, building upon these datasets. DeClarE (Popat et al., 2018) proposes the earliest evidence-based FND method, it jointly learns representations of news content with evidence materials. Many works leverage the merits of hierarchical attention for evidence-news interaction, e.g., HAN (Ma et al., 2019), EHIAN (Wu et al., 2021b), MAC (Vo and Lee, 2021). GET proposes a graph neural network to model distant semantic correlation among news and evidence materials (Xu et al.,

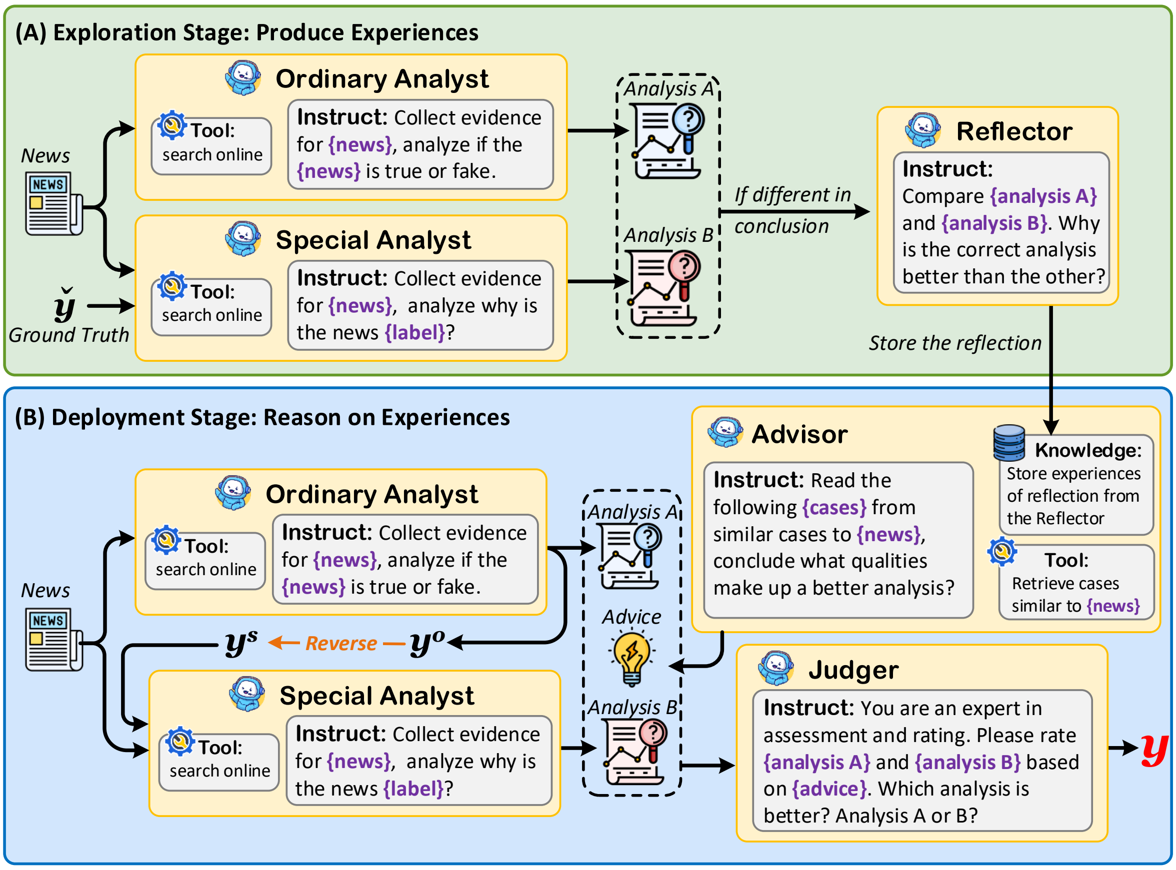

Figure 2: Framework design of **RoE-FND**. **Exploration Stage**: construct a knowledge base through self-reflective experience building. **Deployment Stage**: dynamically synthesizes advice from historical cases to enhance detection.

2022). MUSER retrieves the key evidence information for news verification in a multi-round retrieval process (Liao et al., 2023). SEE builds a dynamic network to adaptively exploit the coarse evidence materials (Yang et al., 2024).

2.2 Large Language Model Inference

Large language models (LLMs) are trained on extensive corpora and are designed to align with human preferences (OpenAI; Touvron et al.; Anthropic, 2024). LLMs rely on devised prompts to solve tasks. Advanced prompting methods enhance the reasoning abilities of LLMs. Chain-of-Thought (CoT) achieved considerable improvements by segmenting tasks (Wei et al., 2022b). CoT-SC searches for the best solution within multiple traces of CoT (Wang et al., 2022). ReAct prompts the LLMs to work in a paradigm of “observe, thought, then acting”, it achieves better results in complex tasks (Yao et al., 2022). ReWOO improves ReAct’s solution by asking the LLMs to plan thoroughly before acting (Xu et al.). Several studies offer fresh perspectives by equipping LLMs with experiences of similar cases during inference (Yang et al., 2023; Sourati et al., 2023; Guo et al., 2023). Therefore, we propose to coordinate advanced inference techniques for our framework. Especially, utilizing experiences of similar cases to address the aforementioned challenges of employing LLMs for FND.

3 Proposed Approach

In this section, we detail the methodology of the proposed framework RoE-FND, which leverages LLM as reasoning units to solve evidence-based fake news detection tasks. As shown in Figure 2, RoE-FND operates in two stages: the exploration stage and the deployment stage. Different components are implemented via calling LLM with different task instructions and tool usage.

3.1 Exploration: Produce Experiences

In the exploration stage, we structure the workflow of the RoE-FND to produce experiences by exploring samples in the training set. Directly instructing LLMs to complete the exploration job in one response is unrealistic due to the job’s complexity.

Therefore, we design sub-tasks to achieve the goal.

3.1.1 Step 1: Dual-channel Analyzing

In the exploration stage, we assign `Analyzer` to analyze the news's authenticity. The `Analyzer` is equipped with a search tool, which enables it to retrieve factual information by querying the news in search engines during its rationale. It concludes the news's authenticity by comprehending its analysis, obtaining a predicted label $y \in \{true, false\}$.

Although the `Analyzer` can produce step-by-step rationale to compose its analysis, it still risks hallucinated reasoning. Thereby, we harness the wrong analysis or hallucinated reasoning to produce experiences. We reveal the label of the news to a special `Analyzer` ahead, thus it ensures a correct conclusion and possibly has a correct analysis. By imposing objectives to the LLM, it improves the generation's faithfulness (Dhuliawala et al.).

Thus, the news is fed into a dual-channel procedure, with an ordinary and a special `Analyzer` each. We represent the procedure as:

$$\begin{aligned} \ddot{y}, \ddot{\mathcal{A}} &= \texttt{OriginalAnalyst}\left(\mathcal{N}\right), \\ \tilde{y}, \tilde{\mathcal{A}} &= \texttt{SpecialAnalyst}\left(\mathcal{N}, \breve{y}\right), \end{aligned} \tag{1}$$

where $\breve{y} \in \{true, false\}$ denotes the ground truth label of the news $\mathcal{N}$, ideally $\tilde{y} = \breve{y}$. The `Analyzer` produces a predicted label and analysis as output.

3.1.2 Step 2: Comparative Reflection

Previous studies have demonstrated the self-reflection abilities of LLMs (Luo et al., 2023; Pan et al., 2023a), which can notice the hallucination or logical inconsistency within LLM generations. However, they normally directly ask LLM to examine the content that needs reflection (Asai et al.; Jeong et al., 2024). Different from them, in RoE-FND, we innovate to boost the reflection by giving the `Reflector` agent with vanilla analysis $\mathcal{A}$ and crafted analysis $\tilde{\mathcal{A}}$ for comparisons.

Specifically, we locate samples that are erroneously analyzed in $\ddot{\mathcal{A}}$ but correct in $\tilde{\mathcal{A}}$, i.e., satisfy $\ddot{y} \neq \tilde{y} = \breve{y}$ in the dual-channel procedure. It is reasonable to assume that, for these samples, the LLM makes logic errors or hallucinations in $\ddot{\mathcal{A}}$, thereby we delegate `Reflector` to identify the mistakes within $\ddot{\mathcal{A}}$ while providing it with $\tilde{\mathcal{A}}$ as comparison and reference. We have:

$$\mathcal{R} = \texttt{Reflector}\left(\mathcal{A}, \hat{\mathcal{A}}\right), \tag{2}$$

where all experiences $\mathcal{R}$ are stored for future reference in the deployment stage.

3.2 Deployment: Reason on Experiences

Our goal during the deployment stage is to enhance LLM's detection fidelity by leveraging reflection experiences in the exploration stage. The methodology is inspired by the paradigm of utilizing old experiences to understand and solve new problems (Kolodner, 1992).

3.2.1 Step 1: Variant Dual-channel Analyzing

In the deployment stage, RoE-FND firstly employs almost the same dual-channel analyzing procedure as Equation 1. However, due to the unavailability of the news's ground truth label, in this dual-channel analyzing procedure, we first generate ordinary analysis and prediction y^o. Then, we reverse the ordinary prediction to have an opposite predicted label y^s for the special `Analyst`. The procedure by step can be formulated as below:

$$\begin{aligned} y^o, \mathcal{A}^o &= \texttt{OridinaryAnalyst}(\mathcal{N}) \\ y^s &= \begin{cases} true, & if\ y^o = false, \\ false, & if\ y^o = true. \end{cases} \\ \tilde{y}^s, \mathcal{A}^s &= \texttt{SpecialAnalyst}(N, y^s). \end{aligned} \tag{3}$$

3.2.2 Step 2: Advice Generation

In this step, we devise an agent `Advisor` with a knowledge base and a tool to retrieve information from it. `Advisor`'s job is to read the experiences from similar cases within the knowledge base, then generalize advice for its colleague to determine the analyses from the dual-channel procedure.

Specifically, given news article $\mathcal{N}$ for detection, `Advisor` retrieves similar news cases in the knowledge base constructed by storing experiences of the `Reflector`'s work and retrieves these cases' reflection. Let the k-th entry within the knowledge base be $(\mathcal{N}_k, \mathcal{R}_k)$, `Advisor` utilize a retrieval tool that based on semantical similarity calculation to retrieve n cases. `Advisor` then instructed to comprehend these cases and provide advice. Assume it retrieve n cases from the knowledge base, the generation of advice $\mathcal{C}$ can be presented as:

$$\begin{aligned} &\mathcal{C} = \texttt{Advisor}\left(\mathcal{N}, [\mathcal{R}_1, \ldots, \mathcal{R}_n]\right), \\ &\quad s.t.\ \mathcal{R}_k \in S = \{(\mathcal{N}_1, \mathcal{R}_1), \ldots, (\mathcal{N}_n, \mathcal{R}_n)\} \\ &\qquad S = \underset{S \subseteq \mathcal{D}, |S| = n}{\arg\max} \sum_{(\mathcal{N}_x, \mathcal{R}_x) \in S} sim(\mathcal{N}_x, \mathcal{N}), \end{aligned} \tag{4}$$

where $\mathcal{D}$ denotes the knowledge base, $sim(\cdot)$ denotes the similarity score of two news articles calculated by the retrieval tool.

Dataset	# Train	# Val	# Test
CHEF	5,754	666	666
PolitiFact	1,919	631	650
Snopes	2,604	869	868

Table 1: Statistics of the divided FND datasets, where the symbol “#" denotes "the number of".

3.2.3 Step 3: Determine the Better Analysis

`Advisor` generates advice by comprehending experiences of similar cases that are harvested by cross-comparing erroneous and correct analyses with reflection, the advice thereby can guide the LLM to identify better analysis from the dual-channel procedure.

In this step, we assign `Judger`, who is expertise in rating and assessment, to select the correct analysis while retaining advice from `Advisor`. `Judger` critically compares two analyses with opposing conclusions and finally determines which analysis is better. The chosen analysis's conclusion is regarded as the prediction result. `Judger`'s procedure to output the final prediction is represented as:

$$\mathcal{J} = \texttt{Judger}\left(\mathcal{A}^o, \mathcal{A}^s, \mathcal{C}\right),$$
$$y = \begin{cases} y^o, & \textit{if } \mathcal{J} \textit{ indicates } \mathcal{A}^o \geq \mathcal{A}^s, \\ \tilde{y}^s, & \textit{otherwise}, \end{cases} \tag{5}$$

where we utilize the symbol $\geq$ to represent that `Judger` determines the original analysis is better than the specially crafted analysis.

4 Experiment

4.1 Experimental Setups

Dataset Preparation. We employ three famous datasets. CHEF (Hu et al., 2022) is a Chinese dataset collected from the real world with multiple domains of news. PolitiFact (Rashkin et al., 2017), and Snopes (Popat et al., 2018) for our experiments are collected from fact-checking websites. We follow previous works' settings to split the dataset, detailed statistics are presented in Table 1.

Tools modules. RoE-FND involves a search tool for collecting evidence and a tool to retrieve cases within the knowledge base. For fair comparisons with baselines, we replace the searched results with evidence within datasets. The retrieval tool utilizes all-MiniLM-L6-v2 [1] embedding. We retrieve one case from the knowledge base.

Implementation details. We utilize the OpenAI platform [2] and DeepSeek platform [3] to run RoE-FND with their models. The LLaMa-series and Qwen-series are run locally using 8 NVIDIA RTX4090 GPUs with Ollama [4].

4.2 Performance Comparisons

4.2.1 Comparisons of Different LLM Options

In this section, we evaluate the performance of RoE-FND across various LLM configurations. Figure 3 illustrates RoE-FND's detection accuracy using different LLMs. The shaded portion of each bar represents the baseline performance achieved by directly prompting the LLMs without RoE-FND, while the unshaded portion highlights the improvements attributable to RoE-FND. Although the extent of improvement varies across datasets, RoE-FND consistently enhances detection accuracy across all LLMs and datasets. Figure 4 (a) and (b) further explore the impact of model size (under 72B parameters) within two prominent model families. In the Qwen2.5 series, the 1.5B model underperforms significantly, likely due to limitations in instruction-following capabilities. The 72B model achieves the best performance, despite with only a slight improvement over the 7B model. The lines in both figures indicate that increasing model size below 72B does not consistently yield performance gains. Overall, both models exhibit stable detection accuracy for sizes larger than 7B.

4.2.2 Comparison with Baseline Approaches

We compare RoE-FND with multiple methods, including DeClarE (Popat et al., 2018), EHIAN (Wu et al., 2021b), GET (Xu et al., 2022), ReRead (Hu et al., 2023), MUSER (Liao et al., 2023), ProgramFC (Pan et al., 2023b), and SEE (Yang et al., 2024). Details of them are presented in Section 2.

In Table 2, we report accuracy (ACC), F1-macro (F1), precision (PR), and recall (RC) following previous methods. The supervised-trained FND methods exhibit varying levels of performance across the datasets. RoE-FND outperforms both supervised-trained evidence-based FND methods and standalone LLMs when integrated with powerful LLMs like DeepSeekv3 and GPT-4o-mini. This highlights the potential of combining RoE-FND with advanced LLMs to achieve state-of-the-art performance in FND tasks. RoE-FND achieves the

[1] https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2

[2] https://platform.openai.com/

[3] https://platform.deepseek.com/

[4] https://ollama.com/

Method	CHEF				Snopes				PolitiFact			
	ACC	F1	PR	RC	ACC	F1	PR	RC	ACC	F1	PR	RC
Trained evidence-based FND methods												
DeClarE	0.589	0.581	0.583	0.568	0.786	0.725	0.610	0.852	0.652	0.653	0.667	0.637
EHIAN	0.571	0.600	0.583	0.516	0.828	0.784	0.617	0.882	0.679	0.676	0.686	0.675
GET	0.588	0.602	0.585	0.582	0.814	0.771	0.721	0.854	0.694	0.691	0.687	0.708
ReRead	0.789	0.776	0.826	0.745	0.816	0.714	0.652	0.789	0.693	0.681	0.711	0.718
MUSER	0.608	0.612	0.603	0.631	0.841	0.745	0.699	0.798	0.729	0.732	0.735	0.728
SEE	0.763	0.776	0.751	0.802	0.824	0.786	**0.773**	0.845	0.706	0.705	0.688	0.724
Training-free LLM methods												
4o-mini	0.740	0.773	0.687	0.883	0.745	0.452	0.604	0.361	0.575	0.395	0.698	0.275
DeepSeekv3	0.780	0.784	0.790	0.778	0.736	0.253	0.712	0.154	0.470	0.631	0.485	0.901
ProgramFC	0.694	0.708	0.723	0.697	0.741	0.619	0.542	0.723	0.678	0.684	0.725	0.741
RoE-FND with different LLMs												
GPT-3.5	0.643	0.676	0.618	0.745	0.711	0.585	0.503	0.698	0.665	0.603	0.758	0.500
LLaMa3	0.722	0.688	0.806	0.612	0.750	0.675	0.554	0.865	0.726	0.720	0.741	0.701
Qwen2.5	0.828	0.827	0.846	0.784	0.824	0.747	0.644	0.888	**0.733**	0.509	**0.875**	0.359
DeepSeekv3	0.890	0.892	**0.876**	0.908	**0.863**	**0.791**	0.708	**0.896**	0.711	0.658	0.767	0.576
4o-mini	**0.891**	**0.893**	**0.876**	**0.911**	0.860	0.774	0.732	0.822	0.712	**0.735**	0.662	**0.827**

Table 2: Performance comparison between our method with baseline methods. We report accuracy (ACC), F1-Macro (F1), precision (PR), and recall (RC)

.

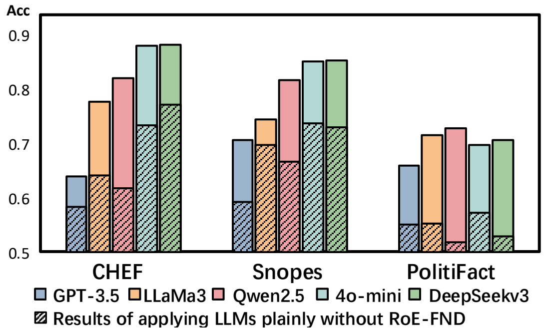

Figure 3: Performance comparisons of RoE-FND utilizing different LLMs on three datasets.

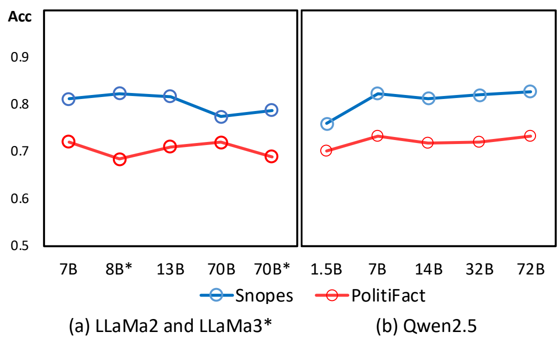

Figure 4: Performance comparison of model size's impact on RoE-FND by two famous models. (a) Testing of LLaMa2 and LLaMa3 (marked by * symbol). (b) Testing of Qwen2.5 models.

best results on nearly all metrics and datasets, with significant performance improvement on CHEF.

Training Dataset	Method	Snopes		PolitiFact	
		ACC	↓(%)	ACC	↓(%)
Snopes	GET	0.814	-	0.603	42.2
	SEE	0.824	-	0.584	59.2
	MUSER	0.841	-	0.667	27.1
	RoE-FND	**0.860**	-	0.702	**4.7**
PolitiFact	GET	0.692	38.8	0.694	-
	SEE	0.665	49.1	0.706	-
	MUSER	0.733	31.6	**0.729**	-
	RoE-FND	0.843	**4.7**	0.712	-

Table 3: Results of cross-datasets testing. ↓(%) indicates the relative decrease compared to training and testing on the same dataset in percentage.

On Snopes and PolitiFact, RoE-FND addresses the detection bias of directly prompting LLMs, while achieving the best detection accuracy. In conclusion, our experiments demonstrate that RoE-FND can effectively leverage the capabilities of LLMs to improve FND detection across multiple datasets.

4.2.3 Cross-datasets Testing Performance

In applications, a trained FND framework is likely to encounter news samples that differ significantly from those in the training dataset. Therefore, the generalization capability of a method is crucial for its practical effectiveness. We conducted cross-dataset testing to evaluate the generalization ability of RoE-FND. The results, presented in Table 3, demonstrate that RoE-FND exhibits superior gener-

Ablation Setting	CHEF		Snopes		PolitiFact	
	ACC	↓(%)	ACC	↓(%)	ACC	↓(%)
w/o `Reflector`	0.610	26.3	0.633	23.2	0.549	23.8
w/o `Summarizer`	0.748	9.6	0.721	12.5	0.555	23.0
w/o Dual-channel	0.725	12.4	0.736	10.7	0.645	10.5
w/o Case Retrieval	0.694	16.2	0.793	3.8	0.673	6.6
Baseline setting	**0.828**	-	**0.824**	-	**0.721**	-

Table 4: Ablation studies of the proposed RoE-FND. ↓(%) indicates the relative decrease compared to the baseline setting in percentage.

Method	CHEF		Snopes		PolitiFact	
	ACC	F1	ACC	F1	ACC	F1
OpenAI o1	0.892	0.888	0.852	**0.800**	0.833	**0.833**
DeepSeekR1	0.882	0.880	0.820	0.791	0.800	0.792
RoE-FND	0.891	0.893	0.863	0.791	0.702	0.735
+ *fine-tuned*	**0.904**	**0.908**	**0.876**	0.788	**0.891**	0.765

Table 5: Results of advanced reasoning LLMs and RoE-FND enhanced by fine-tuning.

alization performance compared to baseline methods. This highlights its robustness and adaptability when facing unfamiliar data distributions.

4.3 Framework Design Exploration

4.3.1 Ablation Study

In Table 4, we investigate the impact of various ablations. For *w/o* `Reflector`, we store analyses without reflection, which are then presented to the `Judger` as high-quality exemplars, simulating a few-shot learning approach. This ablation results in a significant performance drop. For *w/o* `Advisor`, we feed the reflections to the `Judger` directly. The performance drop underscores the importance of `Advisor` in maintaining efficiency. For *w/o Dual-channel*, we replace the dual-channel procedure with a single ordinary `Analyst`, and without comparative reading during reflection. It diminishes the accuracy, demonstrating the value of the dual-channel approach. For *w/o Case Specified Experience*, we retrieve random cases from the knowledge base. The result emphasizes the necessity of targeted case retrieval for optimal results.

4.3.2 Enhancement via Fine-tuning

We devise a solution to enhance the performance of RoE-FND further. The `Analyst` is the most critical component of RoE-FND, thereby we propose a fine-tuning strategy that leverages larger LLMs to improve the capabilities of smaller LLMs. Specifically, we employ a larger LLM, e.g. DeepSeekv3 (671B), to generate high-quality analyses on the training dataset by performing the role of the special `Analyst`. These analyses are then used to fine-tune a smaller model, Qwen2.5 (7B), using LoRA (Low-Rank Adaptation). The fine-tuned model serves as the `Analyst` in RoE-FND, significantly improving its reasoning and analytical capabilities.

We compare the results of enhanced RoE-FND with two state-of-the-art reasoning LLMs: OpenAI o1 (Jaech et al., 2024) and DeepSeek-R1 (Guo et al., 2025). The results are presented in Table 5. While both reasoning models demonstrate significant improvements over baseline methods, the ordinary RoE-FND still performs competitively, achieving results very close to them. Notably, while they take around one minute to process a sample, RoE-FND only takes 22.8s on average. The RoE-FND with fine-tuning enhancement outperforms both reasoning models on three datasets in detection accuracy. The result indicates that enhanced RoE-FND achieves more effective reasoning in FND tasks, leveraging its unique architecture to deliver fast and accurate results.

4.4 Analysis of Generated Content

4.4.1 Case Analysis

We present a challenging case with corresponding outputs in Figure 5 to illustrate these findings. In this example, GPT-4 and DeepSeek-v3 demonstrate a strong ability to dissect the news and analyze each component with supporting evidence. In contrast, Qwen2.5 follows a similar analytical process but struggles to draw meaningful conclusions due to its overemphasis on descriptive alignment. The special `Analyst`, however, achieves more accurate analysis by imposing the conclusion ahead. `Advisor` also emphasizes the importance of logical arguments from similar cases, enabling the `Judger` to recognize the distinctive traits of the analysis and make a correct final prediction. Through the devised procedure of the RoE-FND framework, a smaller LLM like Qwen2.5 can correctly handle challenging detections as larger LLMs.

4.4.2 Gains from the Exploration Stage

`Advisor` generates suggestions based on insights drawn from similar historical cases We present the most frequently mentioned keywords from these suggestions in Figure 6. While there is some overlap in the most common advice keywords across the three datasets, the distribution and emphasis of these keywords vary, likely due to the distinct

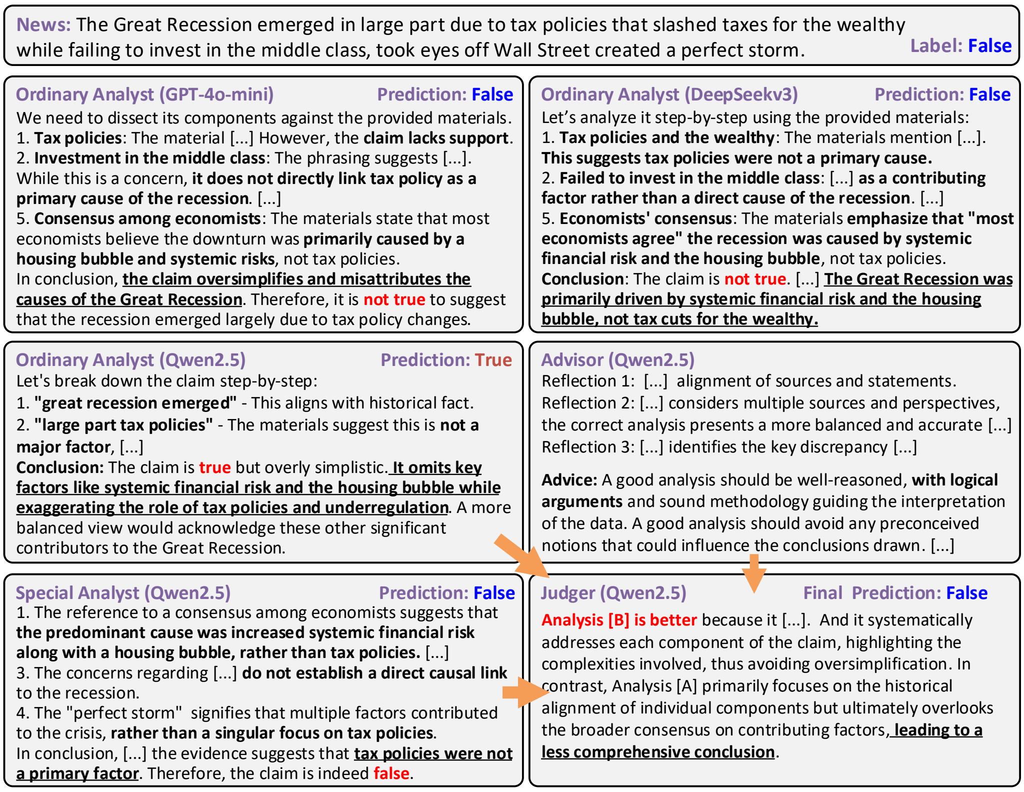

Figure 5: A challenging case from PolitiFact and generation from multiple LLMs. We omit less important content by [..] and highlight key points by colors and boldface.

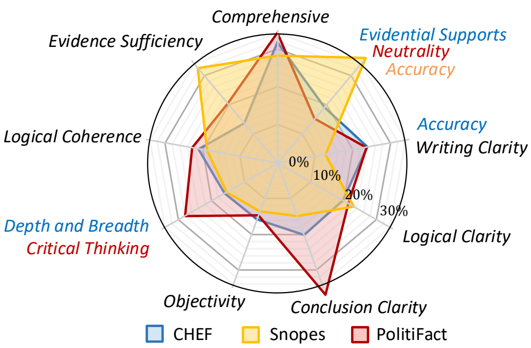

Figure 6: Statistics of the most common keywords from advice by the `Advisor`. (Black words are shared, colored words are dataset-specified.)

characteristics of each dataset. For example, PolitiFact primarily focuses on news related to policies and political parties, which demands analyses that are neutral, critically constructed, and free from ambiguous conclusions. In contrast, societal news in Snopes requires rigorous evidential support to verify news. Despite these differences, stringent standards such as "comprehensiveness" remain a consistent requirement. These findings demonstrate that, even without a formal training process, RoE-FND effectively builds specialized knowledge through its exploration stage.

5 Conclusion and Future Works

We introduce RoE-FND, a novel framework that synergies LLMs with case-based experiential learning. By reframing evidence-based FND as a logical deduction task, RoE-FND leverages self-reflective error analysis to construct a knowledge base, synthesizes advice from historical cases, and ensures the reliability of conclusions through dual-channel verification. Experimental results on CHEF, Snopes, and PolitiFact highlight RoE-FND's superior performance in both effectiveness and interpretability compared to existing methods.

Several future directions can further enhance the framework. First, incorporating real-time case updates will improve its adaptability in practical applications. Second, the framework can be expanded to multi-modal content, e.g. images, to address a broader range of fake news scenarios. Third, potential biases in LLM-generated rationales can be mitigated through training or human-AI collaboration. In summary, our work demonstrates the value of integrating experiential learning with logical reasoning, providing a unique and effective approach to tackling the challenges of fake news detection.

6 Limitations

While RoE-FND demonstrates promising results in evidence-based FND, several limitations warrant consideration for future improvements.

Reliable evidence availability. In RoE-FND, we assign the task of filtering evidence to `Analyzer` in the deployment stage by inserting an extra step into its action chain before analyzing. Thereby addressing the problem of noise or biased evidence. However, the framework may be vulnerable to massive fabricated adversarial evidence materials, which will misguide the LLMs' rationales.

LLM reasoning fidelity. The framework assumes that LLMs can generate logically consistent rationales when guided by retrieved criteria. However, LLMs are known to suffer from hallucination and contextual bias, which may introduce errors even with dual-channel checks. For example, in our experiments, performance variations were observed across different LLM families (e.g., GPT-4 vs. LLaMa2), suggesting that results are sensitive to the base model's reasoning capabilities.

Scalability problem. The employment of LLMs brings difficulties in scalability. RoE-FND relies on multiple LLM generations to process each sample, resulting in higher computational latency compared to traditional trained methods. Although the framework can be deployed by pipeline, it might be less suitable for applications that require swift decision-making. Besides, running RoE-FND locally without leveraging LLM APIs may bring a heavy equipment burden.

Potential risks of RoE-FND. First, RoE-FND may struggle to detect AI-generated news, as the subtle manipulation traces in such content (e.g., stylistic inconsistencies) are often better identified by trained models with feature extraction. Logical reasoning alone may fail to capture these nuances, limiting the framework's effectiveness against increasingly sophisticated AI-driven misinformation. Second, although we assume the logical deduction is homogeneous for either language, the cultural generalization in LLMs may be different. For instance, culturally specific idioms may be neglected during reasoning and cause detection failures. Moreover, RoE-FND may inherit the bias or ethical shortness of LLMs that caused by training.

Addressing the above limitations is critical for enhancing the robustness, scalability, and practical applicability of the framework in diverse and dynamic misinformation scenarios.

References

Anthropic. 2024. Introducing the next generation of claude. Technical report, Anthropic.

Akari Asai, Zeqiu Wu, Yizhong Wang, Avirup Sil, and Hannaneh Hajishirzi. Self-rag: Learning to retrieve, generate, and critique through self-reflection. In *The Twelfth International Conference on Learning Representations*.

Kevin Matthe Caramancion. 2023. News verifiers showdown: a comparative performance evaluation of chatgpt 3.5, chatgpt 4.0, bing ai, and bard in news fact-checking. In *2023 IEEE Future Networks World Forum (FNWF)*, pages 1–6. IEEE.

Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web*, pages 675–684.

Shehzaad Dhuliawala, Mojtaba Komeili, Jing Xu, Roberta Raileanu, Xian Li, Asli Celikyilmaz, and Jason E Weston. Chain-of-verification reduces hallucination in large language models. In *ICLR 2024 Workshop on Reliable and Responsible Foundation Models*.

Song Feng, Ritwik Banerjee, and Yejin Choi. 2012. Syntactic stylometry for deception detection. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 171–175.

Marc Fisher, John Woodrow Cox, and Peter Hermann. 2016. Pizzagate: From rumor, to hashtag, to gunfire in dc. *Washington Post*, 6:8410–8415.

Chunxi Guo, Zhiliang Tian, Jintao Tang, Pancheng Wang, Zhihua Wen, Kang Yang, and Ting Wang. 2023. Prompting gpt-3.5 for text-to-sql with de-semanticization and skeleton retrieval. In *Pacific Rim International Conference on Artificial Intelligence*, pages 262–274.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

Beizhe Hu, Qiang Sheng, Juan Cao, Yuhui Shi, Yang Li, Danding Wang, and Peng Qi. 2024. 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.

Xuming Hu, Zhijiang Guo, GuanYu Wu, Aiwei Liu, Lijie Wen, and S Yu Philip. 2022. Chef: A pilot chinese dataset for evidence-based fact-checking. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3362–3376.

Xuming Hu, Zhaochen Hong, Zhijiang Guo, Lijie Wen, and Philip Yu. 2023. Read it twice: Towards faithfully interpretable fact verification by revisiting evidence. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2319–2323.

Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, et al. 2023. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *arXiv preprint arXiv:2311.05232*.

Aaron Jaech, Adam Kalai, Adam Lerer, Adam Richardson, Ahmed El-Kishky, Aiden Low, Alec Helyar, Aleksander Madry, Alex Beutel, Alex Carney, et al. 2024. Openai o1 system card. *arXiv preprint arXiv:2412.16720*.

Soyeong Jeong, Jinheon Baek, Sukmin Cho, Sung Ju Hwang, and Jong C Park. 2024. Adaptive-rag: Learning to adapt retrieval-augmented large language models through question complexity. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 7029–7043.

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.

Rohit Kumar Kaliyar, Anurag Goswami, and Pratik Narang. 2021. Fakebert: Fake news detection in social media with a bert-based deep learning approach. *Multimedia tools and applications*, 80(8):11765–11788.

Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2022. Large language models are zero-shot reasoners. *Advances in neural information processing systems*, 35:22199–22213.

Janet L Kolodner. 1992. An introduction to case-based reasoning. *Artificial intelligence review*, 6(1):3–34.

Yaliang Li, Jing Gao, Chuishi Meng, Qi Li, Lu Su, Bo Zhao, Wei Fan, and Jiawei Han. 2016. A survey on truth discovery. *ACM Sigkdd Explorations Newsletter*, 17(2):1–16.

Hao Liao, Jiahao Peng, Zhanyi Huang, Wei Zhang, Guanghua Li, Kai Shu, and Xing Xie. 2023. 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.

Junyu Luo, Cao Xiao, and Fenglong Ma. 2023. Zero-resource hallucination prevention for large language models. *arXiv preprint arXiv:2309.02654*.

Jing Ma, Wei Gao, Shafiq Joty, and Kam-Fai Wong. 2019. Sentence-level evidence embedding for claim verification with hierarchical attention networks. Association for Computational Linguistics.

OpenAI. Chatgpt. Technical report, OpenAI.

Liangming Pan, Michael Saxon, Wenda Xu, Deepak Nathani, Xinyi Wang, and William Yang Wang. 2023a. Automatically correcting large language models: Surveying the landscape of diverse self-correction strategies. *arXiv preprint arXiv:2308.03188*.

Liangming Pan, Xiaobao Wu, Xinyuan Lu, Anh Tuan Luu, William Yang Wang, Min-Yen Kan, and Preslav Nakov. 2023b. Fact-checking complex claims with program-guided reasoning. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6981–7004.

Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, and Gerhard Weikum. 2017. Where the truth lies: Explaining the credibility of emerging claims on the web and social media. In *Proceedings of the 26th international conference on world wide web companion*, pages 1003–1012.

Kashyap Popat, Subhabrata Mukherjee, Andrew Yates, and Gerhard Weikum. 2018. Declare: Debunking fake news and false claims using evidence-aware deep learning. *arXiv preprint arXiv:1809.06416*.

Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pages 2931–2937.

Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2020. 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.

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.

Zhivar Sourati, Filip Ilievski, Hng-n Sandlin, and Alain Mermoud. 2023. Case-based reasoning with language models for classification of logical fallacies. In *The 61st Annual Meeting Of The Association For Computational Linguistics*.

James Thorne, Andreas Vlachos, Christos Christodoulopoulos, and Arpit Mittal. 2018. Fever: a large-scale dataset for fact extraction and verification. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 809–819.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothee Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models.

Nguyen Vo and Kyumin Lee. 2021. 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.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2022. Self-consistency improves chain of thought reasoning in language models. *arXiv preprint arXiv:2203.11171*.

Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, et al. 2022a. Emergent abilities of large language models. *Transactions on Machine Learning Research*.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022b. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

Lianwei Wu, Yuan Rao, Yuqian Lan, Ling Sun, and Zhaoyin Qi. 2021a. Unified dual-view cognitive model for interpretable claim verification. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 59–68.

Lianwei Wu, Yuan Rao, Xiong Yang, Wanzhen Wang, and Ambreen Nazir. 2021b. Evidence-aware hierarchical interactive attention networks for explainable claim verification. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pages 1388–1394.

Binfeng Xu, Zhiyuan Peng, Bowen Lei, Subhabrata Mukherjee, Yuchen Liu, and Dongkuan Xu. Rewoo: Decoupling reasoning from observations for efficient augmented language models, may 2023. *arXiv preprint arXiv:2305.18323*.

Weizhi Xu, Junfei Wu, Qiang Liu, Shu Wu, and Liang Wang. 2022. Evidence-aware fake news detection with graph neural networks. In *Proceedings of the ACM web conference 2022*, pages 2501–2510.

Yuzhou Yang, Yangming Zhou, Qichao Ying, Zhenxing Qian, and Xinpeng Zhang. 2024. Search, examine and early-termination: Fake news detection with annotation-free evidences. In *ECAI 2024*, pages 1463–1470. IOS Press.

Zhiwei Yang, Jing Ma, Hechang Chen, Hongzhan Lin, Ziyang Luo, and Yi Chang. 2022. A coarse-to-fine cascaded evidence-distillation neural network for explainable fake news detection. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 2608–2621.

Zonglin Yang, Xinya Du, Erik Cambria, and Claire Cardie. 2023. End-to-end case-based reasoning for commonsense knowledge base completion. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 3509–3522.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. 2022. React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*.

Xueyao Zhang, Juan Cao, Xirong Li, Qiang Sheng, Lei Zhong, and Kai Shu. 2021. Mining dual emotion for fake news detection. In *Proceedings of the web conference 2021*, pages 3465–3476.

Xinyi Zhou and Reza Zafarani. 2020. A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Comput. Surv.*, 53(5).

Yongchun Zhu, Qiang Sheng, Juan Cao, Qiong Nan, Kai Shu, Minghui Wu, Jindong Wang, and Fuzhen Zhuang. 2022. Memory-guided multi-view multi-domain fake news detection. *IEEE Transactions on Knowledge and Data Engineering*, 35(7):7178–7191.

A Implementation Details

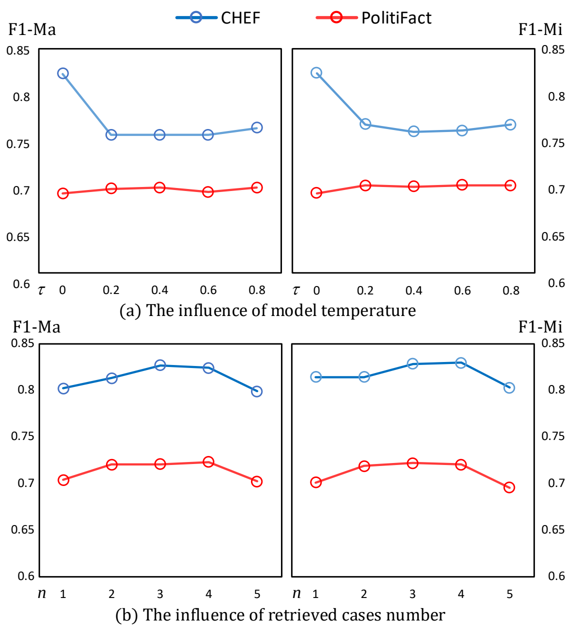

Figure 7: Results of different LLM temperatures and retrieved cases number.

Parameter settings. The impact of LLM temperature value and hyperparameter of RoE-FND is depicted in Figure 7. For all LLMs employed, we

Analyst

Step One:

```
You are an expert at retrieving useful information. Do not fabricate information source if it is
not mentioned. Your response should be a numbered list of retrievals on a new line. For the
claim: \n {news} \n Inspect following materials: \n {evidence} \n
The relevant retrievals with useful information are:
```

Step Two:

```
You are an expert in credibility analysis and factual reasoning. I want you to verify a claim.
Think step-by-step to produce an answer. Return your reasoning process and conclusion.
Materials: \n {output from the last step} \n
Based on the materials above, is the below claim true?
Claim: {news} \n
```

Step Three:

```
Your respond should be a single word, with no premable or explanation.
Analysis: {output from the last step}\n
Based on the above analysis, is the claim true? (yes or no )
```

Special Analyst

Step Two :

```
You are an expert in credibility analysis and factual reasoning. I want you to verify a claim.
Think step-by-step to produce an answer. Return your reasoning process and conclusion.
Materials: \n {output from the last step} \n
Based on the materials above, is the below claim true?
Claim: {news} \n
You should answer the question with a specific conclusion: The claim/news is {conclusion}
```

Step One and Three are identical to the Analyst

Figure 8: Prompts of ordinary and special `Analyst`.

apply the default temperature and top_p value, the length of generation is set under 2,048 tokens. Information on model sizes is listed in Table 6. The Ollama platform uses Q4 quantization. We utilize DeepSeek and OpenAI models via the officially provided API calls.

Dataset settings. We use the dataset setting of the previous methods. Specifically, for CHEF, we delete the label of "NEI" (Not Enough Information) from the dataset to make it a binary classification task following previous work. All datasets are split into training, validation, and test sets in the ratio of 6:2:2. For RoE-FND we develop prompts on the training set and experiment on the testing set. All datasets are downloaded from their official published sources: CHEF https://github.com/THU-BPM/CHEF, Snopes https://www.mpi-inf.mpg.de/dl-cred-analysis/, PolitiFact https://www.mpi-inf.mpg.de/dl-cred-analysis/.

Prompt design. In RoE-FND, we design the prompts in four parts: 1) role assignment, 2) output formation, 3) task instruction, and 4) input wrapper. The prompts of ordinary and special `Analyst` are presented in Figure 8. The prompts of `Reflector`,

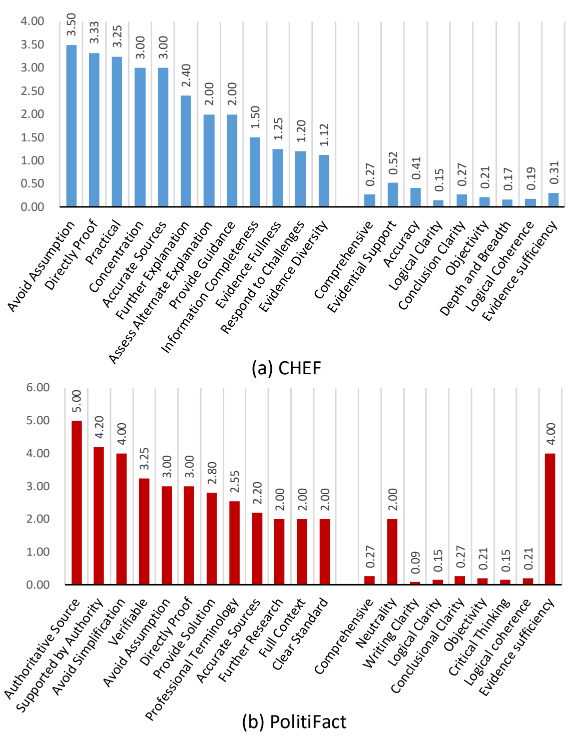

Figure 9: statistics of the most advantageous criteria keywords for the `Judger` to select the correct analysis.

Model	Parameters	Platform	License
LLaMa2	6.74B, 13B, 69B	Ollama	LLAMA2 COMMUNITY LICENSE
LLaMa3	8.03B, 70.6B	Ollama	META LLAMA3 COMMUNITY LICENSE
Qwen2.5	1.54B, 7.62B, 14.8B, 32.8B	Ollama	Apache License Version 2.
Qwen2.5	72.7B	Ollama	Qwen RESEARCH LICENSE
DeepSeekv3	671B	DeepSeek	MIT LICENSE
DeepSeekR1	671B	DeepSeek	MIT LICENSE
gpt-3.5-turbo	-	OpenAI	https://openai.com/policies/terms-of-use/
gpt-4o	-	OpenAI	https://openai.com/policies/terms-of-use/
gpt-4o-mini	-	OpenAI	https://openai.com/policies/terms-of-use/
o1-mini	-	OpenAI	https://openai.com/policies/terms-of-use/

Table 6: Information of the utilized models.

Reflector

```
You are an expert in critical thinking and analysis.
Given a claim, related materials, and two fact-checking analyses that verify the claim's
authenticity,
Find out on why the correct analysis is better. Respond as concisely as possible.

Claim: {news}
Related Materials: {evidence}
Original Analysis: {analysis1}
Correct Analysis: {analysis2}
Why is the correct analysis better?
```

Judger

Step One:

```
You are an expert in assessment and rating. Your answer should be within 100 words.
A better fake news detection analysis often have certain qualities, for example: \n
{advice}\n
Read the below two fake news detection analysis (Analysis [A] and Analysis [B]) that conclude the
authenticity of a news claim differently.
Analysis [A]: {analysis1}\n
Analysis [B]: {analysis2}\n
Which analysis is better and why?
Answer in the following format:
Analysis [X] is better because ...
```

Step Two:

```
Respond in one word to answer question
{output from the last step}\n
According to the above text, which analysis is better? [A] or [B]?
```

Advisor

```
Read the following fact-checking experiences from different cases, conclude what qualities make up
a better analysis? Respond as concise as possible.

Experiences:
{retrieved cases}

Generally speaking, what makes an analysis good? Compose your answer within 200 words.
```

Figure 10: Prompt of `Reflector`, `Advisor`. and `Judger`

`Advisor`, and `Judger` are presented in Figure 10.

B Supplement Results

Generation of `Analyst`. The `Analyst` generates analysis about the news' authenticity. Although the conclusions may vary, authenticity analysis adheres

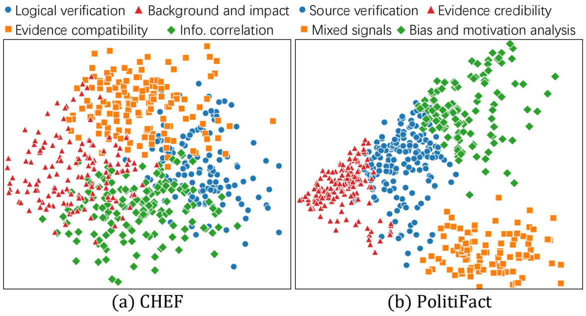

Figure 11: Visualizations of the `Analyzer`'s methodologies with keywords extraction and K-means clustering on Chinese and English datasets.

to certain methodologies. To uncover the underlying patterns of the methodologies, we engage an LLM to extract the keywords that anchor the methodologies. Subsequently, we employ K-means clustering on the embeddings of the extracted keywords to identify several centroids. This process categorizes the methodologies into four distinct genres, as illustrated in Figure 11. The clustering outcomes demonstrate that the generated analyses are underpinned by robust logic across various dimensions. This comprehensive foundation ensures the reliability and interpretability of the proposed framework.

Case Specific Criteria is Prefered than Common Criteria. The summarized criteria contain a list of keywords and their explanation, which are provided to the `Judger` as standard for it to select the better analysis. We ask the `Judger` to score analyses by each criteria keyword. Then, it considers the score difference and selects the winning analysis. Therefore, the criteria keywords that distinguish the two analyses benefit the detection most. We define the advantage of a criteria keyword as the difference between the score of the correct analysis and that of the incorrect one. This advantage is then normalized by dividing it by the keyword's occurrence count. Consequently, theoretically, criteria with greater advantages will have higher scores, capped at a maximum of ten. In Fig 9, we visualize head advantageous criteria keywords and common keywords from Figure 6. Notably, common criteria keywords offer limited advantages during the `Judger`'s selection process. The results underscore the importance of retrieving experiences from similar cases, as analyses may not be distinguishable based on common criteria alone.

C Licenses and Tterm for Use.

All-MiniLM-L6-v2, Ollama uses MIT license. Table 6 lists the utilized models' information. The above artifacts are utilized consistent with their intended use.