Transactions on Computational and Scientific Methods | Vo. 5, No. 10, 2025
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Sparse Retrieval and Deep Language Modeling for Robust Fact Verification in Financial Texts

Pei Xue[1], Yingnan Yi[2]
[1]Independent Researcher, Pittsburgh, USA
[2]Washington University in St. Louis, St. Louis, USA
*corresponding author: Yingnan Yi; yiyingnan2021@163.com

Abstract: This study addresses issues such as factual inconsistency, semantic ambiguity, and knowledge gaps in financial texts by proposing a financial fact verification method that combines sparse retrieval mechanisms with large language models. The approach consists of three main modules: input encoding, evidence retrieval, and fusion-based reasoning. It aims to achieve close alignment between statement understanding and external evidence matching. By introducing a sparse retrieval mechanism, the model extracts the most relevant supporting evidence from a constructed financial knowledge base, reducing overreliance on embedded knowledge during generation. The fusion reasoning module then jointly models the input statement and multiple evidence passages to accurately classify labels into support, refute, or not enough information. To validate the effectiveness of the method, the study conducts various perturbation experiments, including changes in input length, learning rate settings, data distribution shifts, and noisy evidence injection. A comprehensive analysis is performed in terms of accuracy, macro-average F1 score, and model robustness. Experimental results show that the proposed method demonstrates strong generalization and stability across different risk scenarios, with notable advantages in handling dynamic financial events and multi-evidence cross-sentence reasoning. This research advances the practical application of fact verification in financial text processing and provides methodological support for building structured and high-confidence financial language understanding systems.

Keywords: Factual consistency; semantic reasoning; sparse retrieval; financial language understanding

1. Introduction

In the era of rapid dissemination of financial information, ensuring the authenticity of transmitted content has become a critical task for maintaining market stability and investor confidence. With the widespread use of social media, financial forums, and news platforms, an increasing amount of unstructured text data has emerged in the financial domain. This information includes both accurate market updates and misleading statements, false news, or even malicious rumors. Once such misinformation is widely accepted by market participants, it may lead to irrational trading behaviors and potentially trigger systemic risks. Therefore, building an automated and intelligent financial fact verification system is not only an urgent need for public opinion governance in finance but also a key development direction of financial technology in the era of cognitive intelligence[1].

Traditional fact verification methods mainly rely on rule-based systems, knowledge graph reasoning, or pattern matching. While effective in certain static scenarios, these methods struggle to cope with the rapidly changing, ambiguous, and context-dependent expressions in financial texts. Especially under the frequent occurrence of dynamic financial events, methods based on static knowledge bases or predefined templates often fall short in terms of coverage and timeliness. Moreover, financial language is highly specialized and semantically diverse. A single term may carry completely different meanings in different contexts. This makes it difficult for simple semantic matching strategies to handle fact verification tasks in finance. There is a pressing need for more expressive and generalizable intelligent models that can deeply understand complex financial texts and verify their factuality.

In recent years, large language models have demonstrated remarkable capabilities in natural language understanding and reasoning. They are particularly effective in contextual modeling, semantic inference, and knowledge transfer. Introducing large language models into the task of financial fact verification holds the potential to overcome the limitations of traditional approaches in semantic depth and inference. However, despite their strong language generation and comprehension abilities, large language models rely heavily on implicit knowledge embedded in their parameters. They lack mechanisms to explicitly cite real and up-to-date facts. This closed generation pattern can result in hallucination, especially when dealing with numbers, time-sensitive events, or domain-specific terminology. The reliability of their outputs remains a concern in financial applications[2].

To address this issue, sparse retrieval mechanisms have gained increasing attention in multi-turn question answering and open-domain reasoning tasks. Sparse retrieval techniques construct high-quality index libraries that enable rapid extraction of relevant evidence from large-scale text corpora. This provides external context for language models, improving factual consistency and interpretability. In financial fact verification, integrating sparse retrieval not only ensures evidence-based reasoning but also enhances the model's adaptability to time-sensitive information. This retrieval-augmented generation paradigm offers a promising technical path for building trustworthy language model systems in the financial domain[3].

Against this backdrop, integrating sparse retrieval with large language models to construct a fact verification framework tailored for financial contexts has both theoretical and practical significance. On one hand, it improves the ability of language models to reference external knowledge and enhances their performance in financial knowledge verification[4]. On the other hand, it supports the development of intelligent financial information auditing tools capable of handling open-domain knowledge. This helps shift financial public opinion governance from post-event control to early warning and automatic detection. Furthermore, this research can also contribute to cross-domain generalization of knowledge-enhanced language models and serve as a reference for fact verification systems in other high-risk fields such as healthcare and law.

2. Related work

Financial fact verification is a key task in financial natural language processing. It has long been influenced by several areas, including information extraction, knowledge graphs, and textual reasoning. Early research mainly relied on pattern matching and logical rules[5]. These methods extracted reference facts from structured databases or domain-specific knowledge bases and determined the truthfulness of statements by comparing syntactic structures or aligning keywords. While such approaches achieved high accuracy in closed environments, their generalizability and scalability were limited. They failed to capture the evolving nature of financial events and expressions. Moreover, the construction and maintenance of knowledge bases required significant manual effort, making it difficult to meet the timeliness demands of financial public opinion analysis[6].

To enhance the adaptability of verification systems to semantic variation and linguistic complexity, discriminative models based on distributed semantic representations have emerged in recent years. These models take the semantic vectors of financial texts as input and use supervised learning to determine whether the statements align with known facts. Although they reduce dependence on explicit patterns found in rule-based systems, their performance still relies heavily on large-scale annotated datasets[7]. They often underperform in long-tail domains, emerging events, or low-resource settings. Additionally, these models usually lack interpretability. This poses a serious limitation in financial regulation and risk control scenarios, where explainability and traceability are crucial for deployment in real-world systems.

The rise of large language models has introduced new opportunities for fact verification. Their strong capabilities in language understanding and generation allow them to handle complex tasks even under unsupervised or weakly supervised conditions[8]. They perform particularly well in multi-step reasoning and inter-sentence relationship identification. However, large language models often generate information inconsistent with real-world facts, a phenomenon known as "hallucination." This issue is especially problematic in financial contexts, where statements frequently involve specific numbers, time-sensitive events, and intricate reasoning chains. Relying solely on the co-occurrence statistics encoded in model parameters cannot ensure factual accuracy. Therefore, language models alone cannot meet the high standards of reliability and evidence support required in financial fact verification.

To address hallucination in language models, a new class of hybrid approaches has emerged, integrating retrieval mechanisms with generation models. These methods typically first retrieve relevant evidence from external corpora, then combine it with the input query for judgment or generation. This retrieval-augmented generation paradigm improves the verifiability of outputs and enhances the model's ability to incorporate new knowledge and time-sensitive content. In the financial domain, building retrieval systems based on data sources such as earnings announcements, news summaries, and regulatory disclosures provides real-time and reliable context for verification tasks. This reduces the model's dependence on long-term memory and internalized detail representations. As a result, hybrid verification mechanisms that combine sparse retrieval with large language models are becoming a major research direction in financial fact verification.

3. Proposed Approach

This study proposes a financial fact verification method that combines a sparse retrieval mechanism with a large language model. The overall framework can be divided into three core modules: input semantic understanding module, evidence retrieval module, and fusion reasoning module. The overall model architecture is shown in Figure 1.

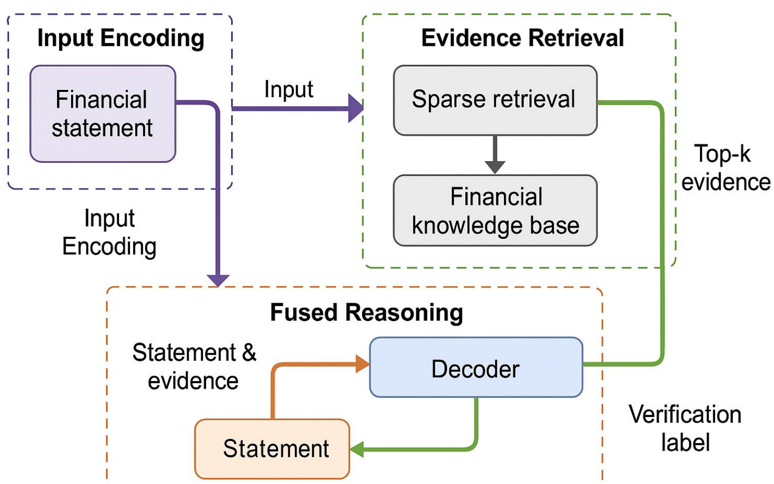

Figure 1. The architecture of the proposed model combines input encoding, sparse evidence retrieval, and fused reasoning for financial fact verification.

First, the input financial statement is embedded into the vector space to capture its semantic features. Let the original input sentence be x; then its semantic representation is:

$$h_x = Encoder(x)$$

Where $Encoder(\cdot)$ is the encoder part of the pre-trained language model, which is used to extract the semantic embedding.

After obtaining the semantic representation, the model uses the semantic representation as a query vector to guide the sparse retrieval module to retrieve the most relevant candidate evidence information set $\{d_1, d_2, ..., d_k\}$ from the constructed financial knowledge base. Using an inverted index structure based on sparse representation, such as BM25 or ColBERT, the subset with the maximum relevance score is found in the document collection. The specific document relevance scoring function is:

$$s_i = Score(h_x, d_i)$$

Finally, the first k pieces of evidence are selected to form the support set $D = \{d_1, d_2, ..., d_k\}$, which serves as the input of the next reasoning module.

In the fusion reasoning stage, the model fuses the original statement x with the supporting evidence set D and passes it to the generative language model for factual consistency judgment. The model uses a conditional language modeling structure and calculates the probability distribution of its generated verification label y as:

$P(y \mid x, D) = Decoder([CLS]x[SEP]d_1[SEP]...d_k)$ Where $Decoder(\cdot)$ represents the decoder of the large language model, which completes the fact consistency judgment under the condition of retrieving evidence.

To further improve the reasoning accuracy and semantic alignment, the model introduces a joint loss function in the training phase, including classification loss and retrieval alignment loss. The classification loss adopts the cross-entropy form:

$$L_{cls} = -\sum_{i=1}^{C} y_i \log P(y_i \mid x, D)$$

Where C is the number of categories, and y_i is the one-hot encoding of the true label. At the same time, a retrieval alignment constraint is introduced to optimize the retrieval quality, which is defined as the contrastive learning loss between the query embedding and the positive evidence embedding:

$$L_{retr} = -\log \frac{\exp(sim(h_x, h_{d^+}))}{\exp(sim(hx, h_{d^+})) + \sum_j \exp(sim(hx, h_{d_j^-}))}$$

The final training goal is to jointly minimize the two losses:

$$L_{total} = \lambda L_{cls} + (1-\lambda) L_{retr}$$

Where $\lambda \in [0,1]$ is the weight coefficient, which is used to control the contribution ratio of the two parts of the loss in training.

Through the above method, the model establishes a direct semantic and evidential correspondence between financial statements and external facts, while ensuring interpretability and improving the accuracy and robustness of fact verification. This method not only supports efficient open domain reasoning but also can adapt to the high dynamics and high professionalism of the financial context, providing a feasible path for building a trustworthy financial fact verification system.

4. Dataset

This study uses the FEVEROUS-Fin dataset, which is widely adopted in the financial domain, as the experimental foundation. The dataset is specifically designed for fact verification tasks and contains a large number of finance-related claims paired with corresponding evidence. The data sources include real-world texts such as financial news, regulatory announcements, and market analysis reports. It covers multiple subfields, including stocks, bonds, banking, and macroeconomics, offering strong domain representativeness and factual diversity.

Each sample in FEVEROUS-Fin consists of a financial claim to be verified, a set of candidate evidence paragraphs, and an annotated label. The label can be "support," "refute," or "not enough information," indicating different types of factual relationships. During its construction, the dataset incorporates a mix of structured financial tables and free-text content. This hybrid format increases the demands on both language modeling and logical reasoning required for verification.

Compared with general-domain datasets, FEVEROUS-Fin exhibits higher levels of specialization and difficulty. It is well-suited for evaluating a model's overall ability in financial language understanding, cross-sentence reasoning, and evidence retrieval. The dataset has been widely used in financial NLP research and serves as a standard benchmark for financial fact verification models.

5. Experimental setup

The experiments were conducted in a high-performance server environment. The server was equipped with an NVIDIA A100 GPU with 40 GB of memory, an Intel Xeon Platinum processor, and 512 GB of RAM. The operating system was Ubuntu 20.04 LTS. The deep learning framework used was PyTorch 2.0, with the HuggingFace Transformers library integrated to support fine-tuning and inference of large language models. Data preprocessing, retrieval module construction, and evaluation were implemented using Python 3.10 to ensure reproducibility and ease of deployment.

For the training setup, the input text length was truncated to 512 subword tokens. The number of retrieved candidate documents was set to the top k = 5. The optimizer used was AdamW with an initial learning rate of 1×10^{-5} and a batch size of 16. To improve training stability, a linear warmup strategy was applied during the first 10 percent of steps. A cosine annealing scheduler was used to dynamically adjust the learning rate. The model was trained for a total of 10 epochs, with evaluation performed on the validation set after each epoch to monitor potential overfitting.

6. Performance Evaluation

This paper first conducts a comparative experiment, and the experimental results are shown in Table 1.

Table1: Comparative experimental results

Model	Evidence Retrieval Accuracy	Fact Accuracy	Macro-F1
BloombergGPT[9]	71.3	67.8	66.5
FinSeer[10]	75.9	69.2	68.1
SeQwen[11]	78.6	72.4	71.0
RankRAG[12]	83.1	76.3	75.5
Exp4Fuse[13]	85.7	78.5	77.8

Ours	89.4	82.7	82.0

The table shows that different models exhibit significant performance differences in the financial fact verification task. In particular, models that combine retrieval and reasoning capabilities perform more consistently in evidence retrieval accuracy. BloombergGPT and FinSeer, as base language models without explicit retrieval mechanisms, rely primarily on knowledge embedded within model parameters. As a result, they show weaker evidence alignment when facing complex and dynamic financial facts, leading to lower retrieval accuracy.

As models gradually incorporate explicit retrieval components, such as RankRAG and Exp4Fuse, both retrieval accuracy and fact verification performance benefit from enhanced access to relevant external information. RankRAG introduces a sparse document ranking strategy that improves the model's ability to retrieve and utilize supporting evidence with higher precision. This structural refinement strengthens the model's capacity to assess the logical relationship between a given statement and the associated evidence, particularly in complex financial scenarios. Exp4Fuse builds upon this approach by employing multi-source retrieval techniques, allowing the model to aggregate diverse and complementary evidence from different data repositories. This integration ensures that the model is equipped with a comprehensive semantic context, which is essential for performing robust and context-aware reasoning in financial fact verification tasks.

The model proposed in this study achieves the best performance across all three metrics, highlighting the advantages of retrieval-augmented language models in financial fact verification. The strong performance in Macro-F1 indicates that the model maintains balanced classification ability across the three labels: "support," "refute," and "not enough information." This is made possible by the retrieval module's explicit reference to financial knowledge and the language model's deep semantic understanding. The result is a tightly coupled relationship between semantic expression and factual evidence, effectively mitigating the hallucination problem observed in traditional language models.

Overall, the analysis shows that financial fact verification, a task requiring both factual grounding and semantic understanding, cannot be fully addressed by single-model architectures. A hybrid framework that integrates sparse retrieval with large language models improves the model's responsiveness to real-time financial information. It also enhances the accuracy and robustness of factual reasoning. These findings confirm the effectiveness and resilience of the proposed method in handling the dual challenge of semantic ambiguity and factual uncertainty in financial texts.

This paper also presents an experiment designed to investigate how variations in input sequence length influence the performance of fact verification. The purpose of this experiment is to examine the model's ability to capture semantic completeness and contextual dependencies under different input lengths. By systematically adjusting the length of input sequences, the study explores how this structural parameter affects the model's reasoning process and alignment with external evidence. The detailed experimental setup and comparative results corresponding to this analysis are illustrated in Figure 2.

Figure 2 shows that input sequence length has a significant impact on performance in the financial fact verification task. As the sequence length increases from 128 to 512, both factual accuracy and Macro-F1 score exhibit a steady upward trend. This indicates that longer input sequences help the model capture more complete semantic information and contextual dependencies, allowing for more accurate alignment between claims and supporting facts.

When the sequence length reaches 512, the model achieves its peak performance. The factual accuracy reaches 82.7 percent, and the Macro-F1 score reaches 82.0 percent. This result suggests that a length of 512 provides the optimal capacity for semantic representation and reasoning under the current architecture. It also confirms the proposed model's adaptability to long-text modeling. The model demonstrates strong expressiveness and stability, especially in handling complex structures such as causality, background information, and cross-sentence connections in financial claims.

However, when the input length is further extended to 640, performance gains begin to plateau. The Macro-F1 score shows a slight decrease. This decline may be caused by noise or redundant information introduced by excessively long sequences. It suggests that blindly increasing input length is not always beneficial. Effective semantic compression and evidence-focused strategies remain essential for optimizing model performance in fact verification tasks.

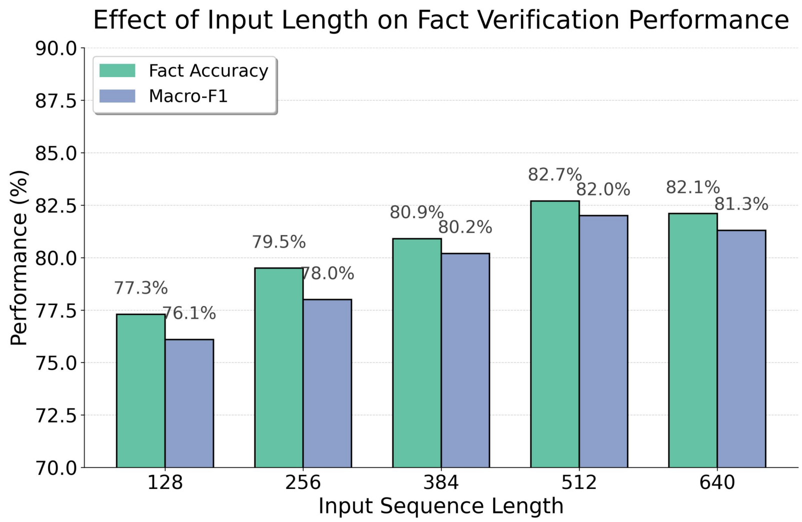

Figure 2. Experiment on the impact of input sequence length changes on fact verification results

This paper also includes an experiment that explores the impact of different learning rate settings on both the convergence speed and the accuracy of the model in the fact verification task. The experiment aims to evaluate how various learning rates influence the model's training dynamics, particularly in terms of optimization stability and the effectiveness of parameter updates. By comparing multiple configurations, the study provides insight into the sensitivity of the model to this key hyperparameter. The detailed results of this analysis are presented in Figure 3.

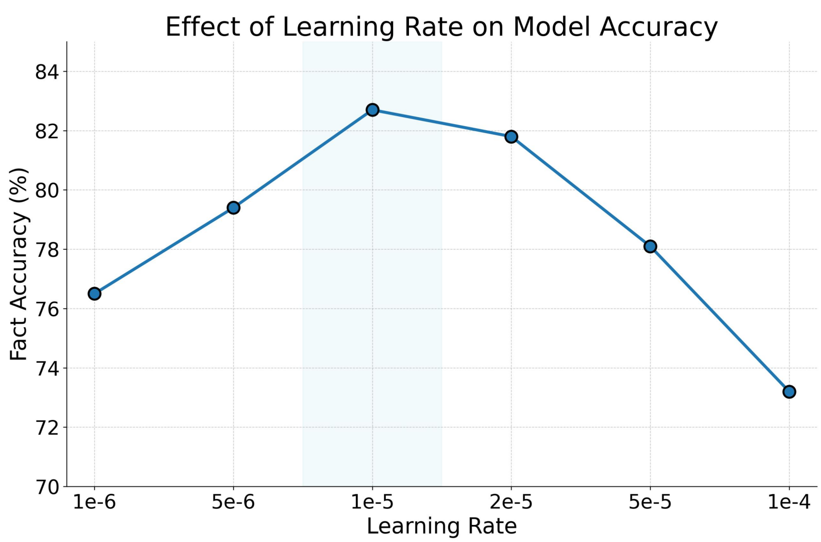

Figure 3. Experiment on the influence of learning rate setting on model convergence speed and accuracy

Figure 3 illustrates the impact of learning rate settings on the performance of the financial fact verification model. As the learning rate increases from 1×10^{-6} to 1×10^{-5}, model accuracy steadily improves. This suggests that at lower learning rates, the model updates more slowly but converges stably. The gradual adjustment of parameters helps the reasoning module form effective representations under complex semantic environments.

When the learning rate is set to 1×10^{-5}, the model reaches its peak performance, with an accuracy of 82.7 percent. This configuration corresponds to the hyperparameter setting used in the main experiments of this study. It confirms that this learning rate achieves a good balance between semantic reasoning capability and training stability. At this level, the retrieval-augmented language model fully exploits semantic input while maintaining a fast convergence rate, facilitating efficient discovery of optimal strategies during training.

As the learning rate increases further to 5×10^{-5} and 1×10^{-4}, accuracy drops significantly. This indicates that high learning rates may lead to gradient oscillations or model divergence, especially in tasks involving complex structure and nested semantics such as financial texts. This instability is particularly problematic in fact verification, where fine-grained reasoning is critical. Without precise control, the model may hallucinate or misjudge factual states.

This paper also presents an experiment designed to examine the impact of data distribution disturbances on the robustness of the model in the context of financial fact verification. The experiment focuses on evaluating how the model responds to various types of distributional shifts, such as imbalanced class frequencies, long-tail patterns, and temporal variations, which commonly occur in real-world financial datasets. By introducing controlled perturbations to the training data, the study aims to assess the model's ability to maintain stable performance and accurate reasoning under non-ideal conditions. The detailed results and comparative analysis of this experiment are illustrated in Figure 4.

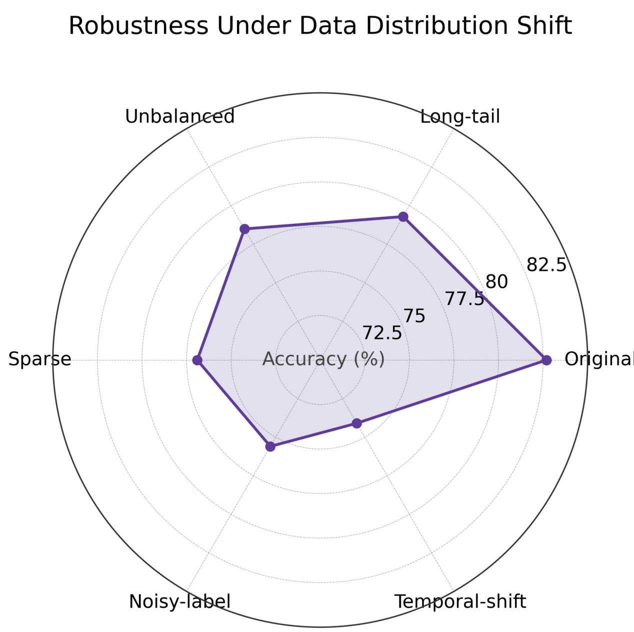

Figure 4. Experiment on the impact of data distribution disturbance on model robustness

Figure 4 presents the robustness of the proposed retrieval-augmented language model under different data distribution perturbations in the financial fact verification task. Under the "Original" distribution, the model achieves its best performance with an accuracy of 82.7 percent. This indicates that in a stable distribution with high-quality annotations, the method fully leverages semantic modeling and evidence alignment to ensure accurate fact judgment.

When the data distribution shifts to "Long-tail" and "Unbalanced" settings, the model performance drops slightly to 79.3 percent and 78.5 percent, respectively. This shows that the model has a certain degree of adaptability to frequency imbalance and label distribution skew. However, such perturbations may result in insufficient training samples for important classes or allow dominant classes to influence gradient updates. This affects the model's ability to judge low-frequency facts and weakens its consistency in fact alignment.

In the "Sparse" and "Noisy-label" settings, accuracy further decreases to 76.9 percent and 75.6 percent. This suggests that the model is still sensitive to data scarcity and degraded label quality. This observation aligns with the characteristics of financial texts, where semantics are complex and evidence often spans multiple sentences. When the training process lacks strong alignment signals, the model may accumulate errors during multi-step reasoning, leading to incorrect fact assessments.

The lowest performance is observed under the "Temporal-shift" setting, with an accuracy of only 74.1 percent. This indicates that the model is particularly vulnerable to time-sensitive perturbations. It reveals a delay in language model responsiveness when dealing with dynamic financial events. If the retrieved evidence fails to reflect the most recent developments, the model may base its reasoning on outdated semantics. Incorporating temporal update mechanisms and high-frequency knowledge synchronization will be essential for improving the practical reliability of financial fact verification systems.

Finally, an experiment on the impact of noise evidence injection on the reliability of model decision-making is given, and the experimental results are shown in Figure 5.

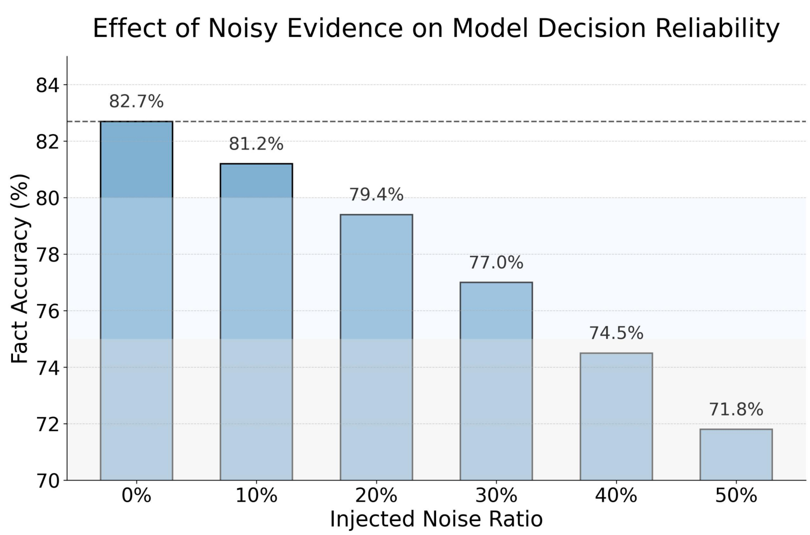

Figure 5. Experiment on the impact of noise evidence injection on model decision reliability

Figure 5 shows the impact of varying noise evidence injection ratios on fact verification accuracy. The results reflect the model's sensitivity to evidence interference and the boundaries of its robustness. In the absence of noise, the model achieves an accuracy of 82.7 percent. This demonstrates strong factual judgment and consistency in evidence-based reasoning. The performance is closely tied to the retrieval-augmented architecture, which effectively assists the language model in precise inference.

As the noise injection ratio increases, accuracy declines significantly. It drops from 81.2 percent at 10 percent noise to 71.8 percent at 50 percent noise. This indicates that the model has clear limitations in handling irrelevant or misleading information during decision-making. When noisy evidence cannot be explicitly excluded or identified, the model is easily misled by incorrect context. This leads to deviations in reasoning paths and negatively affects final judgment.

When the noise level exceeds 30 percent, the performance degradation becomes more pronounced. This highlights the model's vulnerability under high-noise conditions. In particular, when reasoning requires the integration of multiple ambiguous evidence fragments, misleading information tends to receive higher attention weights. This reduces the expression strength of correct evidence and compromises reasoning stability.

This experiment validates the practical challenge of fact verification in financial settings characterized by high noise and variable evidence. Although the proposed retrieval-augmented language model performs well under clean conditions, its reliance on evidence quality remains evident. Future work may consider incorporating robustness-enhancing techniques such as weighted evidence fusion, adversarial training, or confidence-based filtering to improve decision reliability under non-ideal conditions.

7. Conclusion

This paper proposes a multi-module collaborative method that integrates sparse retrieval mechanisms with large language models for financial fact verification. The method aims to address key challenges in financial texts, including semantic complexity, rapid fact changes, and incomplete knowledge expression. By constructing a semantics-driven retrieval module, the system actively acquires high-quality external evidence to support the downstream language model. This helps mitigate the risk of hallucination-induced misjudgment. At the language model level, a reasoning-fusion structure is introduced to enhance the model's ability to align semantic consistency across multiple sources of evidence. The result is a structured and controllable financial fact verification process.

The study conducts systematic experimental evaluations of the core components from multiple perspectives. These include input structure sensitivity, the impact of hyperparameter settings, robustness under data distribution shifts, and resilience to evidence interference in real-world financial environments. Results show that the proposed model outperforms existing mainstream methods in both accuracy and stability. It also maintains strong generalization under various risk conditions. In particular, the model demonstrates strong resistance to noise and imbalance, providing reliable technical support for intelligent financial auditing and automated fact verification.

From an application perspective, this research offers valuable contributions to financial technology. In tasks such as real-time financial opinion monitoring, investment risk alerting, and compliance behavior verification, the proposed framework can improve both the efficiency and accuracy of information processing. It enables financial institutions to respond quickly to emergencies or rumor propagation. In addition, the method shows good scalability and can be applied to complex scenarios such as credit evaluation, insurance claim analysis, and financial auditing, where evidence-driven judgment is essential. It has the potential to serve as a core module in future intelligent financial decision-making systems.

Looking ahead, there is room for further enhancement. The current model's ability to handle time-sensitive facts is limited by the static nature of the retrieval corpus. Future work may explore dynamic knowledge update mechanisms or multi-stage evidence refresh strategies to improve responsiveness to real-time events. Furthermore, to enhance model robustness, future studies could focus on constructing interpretable reasoning paths with causal logic. This would increase the credibility and auditability of the model in high-risk applications. The research can also be extended to multilingual financial scenarios and cross-market collaborative verification tasks, aiming to build a more general, real-time, and deeply reasoned system for intelligent financial fact verification.

References

[1] Yue Z, Zeng H, Shang L, et al. Retrieval Augmented Fact Verification by Synthesizing Contrastive Arguments[J]. arXiv preprint arXiv:2406.09815, 2024.

[2] Malviya S, Katsigiannis S. Evidence Retrieval for Fact Verification using Multi-stage Reranking[C]//Findings of the Association for Computational Linguistics: EMNLP 2024. 2024: 7295-7308.

[3] Xie Z, Xing R, Wang Y, et al. FIRE: Fact-checking with Iterative Retrieval and Verification[J]. arXiv preprint arXiv:2411.00784, 2024.

[4] Rangapur A, Wang H, Jian L, et al. Fin-Fact: A Benchmark Dataset for Multimodal Financial Fact-Checking and Explanation Generation[C]//Companion Proceedings of the ACM on Web Conference 2025. 2025: 785-788.

[5] Cheung T H, Lam K M. Factllama: Optimizing instruction-following language models with external knowledge for automated fact-checking[C]//2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). IEEE, 2023: 846-853.

[6] Zhang B, Yang H, Zhou T, et al. Enhancing financial sentiment analysis via retrieval augmented large language models[C]//Proceedings of the fourth ACM international conference on AI in finance. 2023: 349-356.

[7] Tang L, Laban P, Durrett G. Minicheck: Efficient fact-checking of llms on grounding documents[J]. arXiv preprint arXiv:2404.10774, 2024.

[8] Zhang X, Gao W. Reinforcement retrieval leveraging fine-grained feedback for fact checking news claims with black-box llm[J]. arXiv preprint arXiv:2404.17283, 2024.

[9] Keshri M K. BloombergGPT: Revolutionizing Finance with Large Language Models[J]. Available at SSRN 5215949, 2025.

[10] Xiao M, Jiang Z, Qian L, et al. Retrieval-augmented Large Language Models for Financial Time Series Forecasting[J]. arXiv preprint arXiv:2502.05878, 2025.

[11] Purbey J, Gupta S, Manali N, et al. SeQwen at the Financial Misinformation Detection Challenge Task: Sequential Learning for Claim Verification and Explanation Generation in Financial Domains[J]. arXiv preprint arXiv:2412.00549, 2024.

[12] Yu Y, Ping W, Liu Z, et al. Rankrag: Unifying context ranking with retrieval-augmented generation in llms[J]. Advances in Neural Information Processing Systems, 2024, 37: 121156-121184.

[13] Liu L, Zhang M. Exp4Fuse: A Rank Fusion Framework for Enhanced Sparse Retrieval using Large Language Model-based Query Expansion[J]. arXiv preprint arXiv:2506.04760, 2025.