Proceedings of the 4th International Conference on Management Research and Economic Development
DOI: 10.54254/2754-1169/2026.33665

Sentiment Analysis of Stock Market Live Streaming Content Based on Large Language Models

Runxin Pang[1], Hanzhe Zhang[2*]

[1]*School of Artificial Intelligence and Data Science, University of International Business and Economics, Beijing, China*
[2]*Faculty of Business and Economics, The University of Melbourne, Parkville, Australia*
**Corresponding Author. Email: hzhang3760@unimelb.student.edu.au*

Abstract. Stock market live streaming has become an important channel for investors to access market information, but such content is often emotionally charged and loosely structured, potentially affecting viewers' emotions and trading decisions, thereby indirectly disturbing market stability. To address the need for automated analysis of massive volumes of live streaming text, this study introduces Large Language Models (LLMs) for sentiment analysis and information mining. We collected 7,912 text segments from 20 live sessions on the Tiantian Fund platform and manually annotated financial entities, state descriptions, and sentiment tendencies. The Qwen2.5 model was fine-tuned with LoRA via the LLaMA-Factory framework. Results show that on entity-containing content, the fine-tuned model achieves significantly higher text similarity and ROUGE scores; however, on noisy text without entities, performance declines, indicating that automated processing still faces challenges such as overfitting and semantic drift. Overall, LLMs show the potential to process unstructured financial live streaming text in batch, offering new technical references for precise market regulation and investor education.

Keywords: Large Language Models, Stock Market Live Streaming, Model Fine-tuning, Sentiment Analysis

1. Introduction

Investors increasingly rely on stock market live streaming for real-time information. Such content, however, tends to carry strong emotional tones. These emotions can influence viewers' own feelings and trading choices, which in turn may create indirect risks to market stability. Live streaming texts are often noisy, informal, and filled with non-standard terminology. Traditional dictionary-based or shallow machine learning methods therefore struggle to analyze them. The sheer daily volume of live streams makes manual or rule-based processing impractical. Automation would help, but it encounters difficulties: subtle emotion expressions, inconsistent entity references, missing context, and noise that interferes with model training. Large language models (LLMs), with their strong contextual understanding and reasoning, offer a way forward [1]. Most existing financial NLP research has concentrated on well-structured documents such as news, earnings reports, and conference calls. Live streaming remains largely unexamined. To address this gap, we use LLMs for

DOI: 10.54254/2754-1169/2026.33665

entity recognition and sentiment analysis on stock live streaming texts. We build a domain-specific annotated dataset, efficiently fine-tune the model, and systematically test it on massive, noisy live streaming content, with the goal of offering a practical approach for automated financial live stream mining.

2. Related works

2.1. LLMs in NLP

After pre-training on large corpora, Large Language Models (LLMs) demonstrate powerful capabilities in solving Natural Language Processing (NLP) tasks and exhibit abilities such as in-context learning and step-by-step reasoning that smaller models lack. Powerful linguistic modeling, contextual learning, and word embedding techniques enable LLMs to better accomplish various NLP tasks including sentiment analysis, information retrieval, and similarity measurement [1, 2].

For instance, the BERT model proposed in 2018 can achieve effective named entity recognition after fine-tuning [3]. As an important NLP breakthrough, BERT inspired many advanced models, and later LLMs such as ChatGPT further enhanced various NLP tasks, allowing modern LLMs to handle multiple tasks in a unified framework [4]. As another core NLP task, sentiment analysis benefits from LLMs' strong context modeling ability. Studies using prompt engineering for conversational emotion alignment have achieved clear improvements on multiple datasets [5, 6], and LLM-based emotion recognition both advances NLP and delivers more natural interactive experiences.

2.2. LLMs in financial applications

LLMs have been widely applied in economics and finance due to their strong text and multimodal understanding capabilities. Although traditional dictionary and machine learning methods have been used to process unstructured financial data, LLMs show better performance with stronger contextual reasoning. For example, ChatGPT effectively processes financial reports and disclosures to support investor decisions [7], while the multimodal FinTral model achieves excellent results on financial NLP, stock prediction, and corporate disclosure tasks [8].

Current practical applications of financial LLMs also cover multimodal financial chart recognition and visual data analysis [9, 10], as well as the construction and optimization of intelligent trading agents for quantitative investment strategies [11, 12]. Nevertheless, existing studies mainly focus on news and conference calls, leaving financial live streaming under-explored without sufficient analysis of its unstructured characteristics.

3. Dataset and preprocessing

3.1. Data source and annotation standards

The stock market live streaming text data adopted in this study is collected from the Tiantian Fund platform, whose live content covers market trend analysis, investment strategy interpretation and financial product promotion. We selected 20 live sessions conducted in March, May and September 2024, totaling 7,912 text entries after sentence segmentation by commas.

During manual annotation, we labeled product promotions and noise information as empty, and annotated other meaningful sentences across six labels: industry, primary entity, secondary entity,

DOI: 10.54254/2754-1169/2026.33665

time, state, and sentiment. Based on text characteristics, experimental requirements, and annotation quality, we selected four labels for training:

• Primary Entity: Categories including upstream and downstream industry cycles, industrial operating conditions, national policy orientation, macroeconomic trends, segmented sector performance and irrelevant content.

• Secondary Entity: Refined object content such as financial indicators and market quotations, which serves as the core reference for state description.

• State: The current state of the described entity, serving as a description of its specific situation, closely linked to the secondary entity.

• Sentiment: Quantified market emotional tendency through scoring ranging from -2 to +2, The detailed definition of each sentiment category is presented in Table 1.

Table 1. Sentiment scoring standard

Score	Sentiment Category	Description
+2	Strongly Positive	Strong recommendation, high optimism
+1	Positive	Recommendation, favorable prospects
0	Neutral	No clear sentiment
-1	Negative	Non-recommendation, perceived risks
-2	Strongly Negative	Strong non-recommendation, high risk

3.2. Data preprocessing

After text annotation, we transformed the data into question-answer pair format suitable for LLM inference and fine-tuning. We adjusted the annotation results to construct prompt-response pairs, splitting texts with multiple entities into separate annotation entries, resulting in 8,090 annotated entries from 7,912 texts. Among these, 6,218 entries showed no entity appearance, while 1,872 contained entities, aligning with the characteristic of live streaming content having high noise with less meaningful content.

The dataset was randomly split at a 9:1 ratio to construct training and test sets: the detailed partition, including the distribution of entity-containing and non-entity samples, is shown in Table 2.

Table 2. Dataset partition statistics

Dataset	Total	With Entities	Without Entities
Training Set Set	7281	1663	5618
Test Set	809	209	600

4. Experiments and results

4.1. Experiments settings and model

Before formal fine-tuning, this study conducts zero-shot inference on the training set to evaluate the original model's basic performance, which serves as a baseline for subsequent effect comparison. During the inference phase, manual annotation information is masked, and standardized prompt words are added before each text. After multiple rounds of iteration and optimization, the final prompt covers task definitions, annotation specifications, contextual constraints and reasoning

DOI: 10.54254/2754-1169/2026.33665

examples, enabling the LLM to fully understand the annotation requirements via prompt engineering.

We conduct all experiments on an NVIDIA GeForce RTX 4090 GPU using the Qwen-2.5-7B-Instruct model. We perform inference with the vLLM framework for efficiency, setting temperature=0.0 and repetition penalty=1.2 for stable outputs. The model is fine-tuned via LLaMA-Factory with LoRA [13] to reduce computation costs.

4.2. Results

For result evaluation, we used the M3E-Base text embedding model to convert text results into vector representations, comparing similarity between manual annotations and inference results.

Table 3. Text similarity before and after fine-tuning

Dataset	With Entities	Without Entities
Training Set (Before) Set	0.7559	0.9876
Test Set (After)	0.8035	0.9247

As Table 3 shows, the similarity score for entity-containing texts rose from 0.7559 to 0.8035, an increase of about 5%. For texts without entities, the score dropped from 0.9876 to 0.9247, a decline of roughly 6%. This pattern indicates overfitting: the fine-tuned model seems to treat noise as if it were meaningful information.

We also applied ROUGE metrics [14] because our annotation task is essentially generating short summaries that describe entities, states, and sentiment. On entity-labeled content, the fine-tuned model performed substantially better. Relative to the training set, the test set ROUGE-1, ROUGE-2, and ROUGE-L F1 scores improved by 63%, 190%, and 78%, respectively. These gains show that the model became more capable of summarizing the original text and correctly identifying entities, states, and emotional tones.

For non-entity content, however, the accuracy of correctly recognizing the absence of entities fell from 0.9580 (training) to 0.7367 (test). This further confirms overfitting and might reflect problems with dataset quality or limited context length.

5. Conclusion and future work

Our findings show that fine-tuning helps LLMs recognize entities and detect sentiment in financial live streaming text. For text segments that contain entities, average similarity rises from 0.7559 to 0.8035. ROUGE-1, ROUGE-2, and ROUGE-L F1 scores also go up, by 63%, 190%, and 78%, respectively. However, performance degrades on non-entity noisy samples. Similarity drops from 0.9876 to 0.9247. The correct rejection rate falls from 0.9580 to 0.7367. These results indicate overfitting and sensitivity to noise. Despite these challenges, our findings still suggest that LLMs can be adapted for massive, unstructured financial live streaming content. Future work will focus on expanding multi-source datasets, optimizing fine-tuning strategies, and extending to multimodal scenarios.

References

[1] Zhao, W. X., Zhou, K., Li, J., & Wang, H. (2023). A survey of large language models. arXiv Preprint, arXiv: 2303.18223. https: //doi.org/10.48550/arXiv.2303.18223

DOI: 10.54254/2754-1169/2026.33665

[2] Raiaan, M. A. K., Mukta, M. S. H., Fatema, K., et al. (2024). A review on large language models: Architectures, applications, taxonomies, open issues and challenges. IEEE Access, 12, 26839–26874.
[3] Devlin, J., Chang, M.W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proc NAACL-HLT, 4171–4186.
[4] Qin, L., Chen, Q., Feng, X., et al. (2026). Large language models meet NLP: A survey. Front. Comput. Sci., 20(11), 2011361. https: //doi.org/10.1007/s11704-025-50472-3
[5] Lei, S., Dong, G., Wang, X., et al. (2023). InstructERC: Reforming emotion recognition in conversation with multi-task retrieval-augmented large language models. arXiv preprint arXiv: 2309.11911. https: //doi.org/10.48550/arXiv.2309.11911
[6] Wen, J., Tu, G., Li, R., Jiang, D., & Zhu, W. (2023). Learning more from mixed emotions: A label refinement method for emotion recognition in conversations. Transactions of the Association for Computational Linguistics, 11, 1485–1499.
[7] Dong, M.M., Stratopoulos, T.C., Wang, V.X. (2024). A scoping review of ChatGPT research in accounting and finance. International Journal of Accounting Information Systems, 55, 100715.
[8] Bhatia, G., Nagoudi, E. M. B., Cavusoglu, H., et al. (2024). FinTral: A family of GPT-4 level multimodal financial large language models. Find. ACL, 13064–13087.
[9] Wang, Z., Li, Y., Wu, J., et al. (2023). FinVis-GPT: A multimodal large language model for financial chart analysis. arXiv. Preprint. https: //arxiv.org/abs/2308.01430
[10] Gan, Z., Lu, Y., Zhang, D., et al. (2024). MME-Finance: A Multimodal Finance Benchmark for Expert-level Understanding and Reasoning. Proc. ACM MM, 4323–4334.
[11] Ding, H., Li, Y., Wang, J., et al. (2024). Large language model agent in financial trading: A survey. arXiv Preprint, arXiv: 2408.06361. https: //doi.org/10.48550/arXiv.2408.06361
[12] Lee, J., Stevens, N., & Han, S. C. (2025). Large language models in finance (FinLLMs). Neural Comput. Appl., 37(30), 24853–24867.
[13] Hu, E. J., Shen, Y., Wallis, P., & Allen-Zhu, Z. (2022). LoRA: Low-rank adaptation of large language models. Proc. ICLR. https: //doi.org/10.48550/arXiv.2106.09685
[14] Lin, C. Y. (2004). ROUGE: A package for automatic evaluation of summaries. Proceedings of the ACL Workshop on Text Summarization Branches Out, 74-81. https: //aclanthology.org/W04-1013/