SCIENTIFIC PUBLICATIONS OF THE STATE UNIVERSITY OF NOVI PAZAR
SERIES A: APPL. MATH. INFORM. AND MECH. Vol. 18, 1, 2026, 1–16

Architecture and Evaluation of a Multi-Agent System for Market Monitoring

Ulfeta A. Marovac, Irena R. Vodenska, Admir K. Hamzagić, Dalila A. Pramenković

Abstract: The contemporary financial environment is characterized by a rapidly increasing volume of heterogeneous data, which necessitates automated solutions for their integration and analysis in the context of market dynamics. In this paper, a multi-agent system for automated monitoring and analysis of capital markets is presented, developed using the CrewAI framework and combining structured financial data with sentiment analysis of financial news. The proposed system integrates numerical data with textual content from relevant news sources, enabling the identification of information with a potential impact on stock price movements. The system architecture comprises two specialized agents: MarketScanner, responsible for analyzing market indicators and generating concise analytical reports, and NewsAggregator, which collects relevant news and performs sentiment analysis using large language models and the FinBERT model for financial text processing. The system was evaluated using data from Apple, Bank of America, and Exxon Mobil. Experimental results indicate that the integration of a multi-agent architecture with sentiment analysis supports systematic examination of market-relevant information. Sentiment classification accuracy reached above 80% in selected cases when evaluated against expert annotations. The findings suggest that sentiment–market relationships vary across companies and depend on the specific market context.

Keywords and phrases: multi-agent systems, stock market analysis automation, CrewAI, information filtering, structured reports.

1 Introduction

Contemporary stock markets generate extremely large volumes of data, while financial media continuously publish information that may have a significant impact on stock price

Manuscript received January 30, 2026; accepted April 30, 2026

Ulfeta A. Marovac (ORCID 0000-0001-7232-3755) is with the Deparment of Technical Science, State University of Novi Pazar, Serbia, umarovac@np.ac.rs

Irena R. Vodenska (ORCID 0000-0003-1183-7941) is with the Administrative Sciences Department, Metropolitan College, Boston University, USA, vodenska@bu.edu

Admir K. Hamzagić (ORCID 0009-0009-1113-7428) is with the Department of Economics Science, State University of Novi Pazar, Serbia, ahamzagic@np.ac.rs

Dalila A. Pramenković is an MSc student at the Deparment of Technical Science, State University of Novi Pazar, Serbia, pramenkovicdalila@gmail.com

https://doi.org/10.46793/SPSUNP2601.01M

movements. Effective investment decision-making and reliable recognition of market trends therefore require the integrated analysis of structured financial data and unstructured textual content from diverse news sources. However, manual analysis of such sources becomes impractical in environments characterized by rapid market changes.

Financial news provides insights into macroeconomic indicators, corporate performance, regulatory developments, and geopolitical events that may influence market behavior. While continuous monitoring of such sources can support early identification of potential risks and opportunities, it also introduces challenges related to identifying information that is truly relevant to market movements. This highlights the need for systematic approaches capable of filtering relevant news content and linking it to observable market outcomes.

In this context, automated systems have become increasingly important for the efficient monitoring, analysis, and interpretation of market-relevant information in near real time. Approaches that integrate numerical market data with the analysis of textual news content are particularly valuable, as they enable more advanced forms of decision support in financial markets.

Despite significant advances in sentiment analysis and stock market prediction, several limitations persist in existing research and practical applications. A significant portion of the existing literature focuses either on numerical market data or on textual sentiment derived from financial news, frequently treating these data sources in isolation. Although hybrid models that combine price information with sentiment signals have demonstrated improved predictive performance, they are typically evaluated as standalone experiments rather than as components of integrated, operational systems for continuous market monitoring.

Another critical limitation concerns interpretability. Many sentiment analysis approaches, including domain-specific models such as FinBERT, produce sentiment labels or probability scores without providing explicit explanations of the underlying reasoning. This lack of transparency reduces their practical usefulness in real-world financial decision-making, where understanding why a particular sentiment signal is generated is as important as the signal itself.

Furthermore, existing research largely emphasizes model-centric evaluations, with limited attention given to system-level architectures capable of coordinating multiple analytical processes. Although recent studies have explored the use of large language models and multi-agent frameworks for financial analysis, most existing approaches remain primarily focused on trading execution or strategy optimization. Frameworks such as TradingAgents [17] emphasize collaborative decision-making aimed at maximizing profitability, often at the expense of structured information filtering and interpretability. Consequently, these systems provide limited support for continuous market monitoring and analyst-oriented reporting, highlighting the need for agent-based architectures that focus on integrating heterogeneous financial data sources and generating transparent, explanatory outputs.

This research addresses these gaps by focusing on the design and evaluation of a multi-agent system that integrates heterogeneous financial data sources, emphasizes interpretability through reasoning-enabled sentiment analysis, and supports automated market monitoring and structured reporting.

This paper presents a multi-agent system for the automated monitoring, analysis, and interpretation of stock market dynamics, developed using the CrewAI framework and relying on a combination of structured financial data and unstructured financial news content. Building upon recent advances in sentiment analysis and multi-agent architectures, the proposed approach is designed to address the limitations of existing trading-oriented systems by focusing on continuous market monitoring, structured information filtering, and interpretability rather than direct trade execution.

The system integrates structured market data obtained from reliable financial sources with real-time media content in order to identify news events that exhibit a verifiable relationship with short-term stock price movements. By coordinating specialized agents responsible for market analysis and news-based sentiment interpretation, the system enables a systematic linkage between qualitative media signals and quantitative market behavior.

Unlike many existing approaches that emphasize model-centric evaluation or isolated sentiment prediction, the proposed system adopts a system-level perspective aimed at analyst-oriented decision support. The effectiveness of the approach is evaluated by analyzing the alignment between sentiment signals generated by different models and observed market movements, as well as by assessing the interpretability of the generated analytical outputs.

A key contribution of the system lies in integrating media-sourced content with real market data to examine news events that exhibit potential relationships with short-term stock price movements. This approach supports investment decision-making by automating the processing of large volumes of information and generating clear, structured reports for end users.

The remainder of this paper is organized as follows. Section 2 provides an overview of related work in the field of sentiment analysis in finance, including classical and hybrid approaches, as well as recent methods based on large language models and multi-agent systems. Section 3 describes the architecture of the proposed multi-agent system, with particular emphasis on the roles of individual agents and their coordination within the CrewAI framework. Section 4 presents the data sources used in the study and the methodology of the experimental evaluation. The experimental results and their discussion are given in Section 5. Finally, Section 6 summarizes the conclusions of the study and outlines directions for future research.

2 Related Work

The field of automated stock market prediction has evolved from basic statistical approaches to more advanced methods that integrate heterogeneous data sources. Early studies examining the impact of news on stock prices relied on manually constructed dictionaries and traditional machine learning techniques to establish relationships between textual patterns and price movements. Li et al. [9] demonstrated that multi-dimensional sentiment spaces, constructed using the Harvard psychological and Loughran–McDonald financial dictionaries and combined with Support Vector Machines (SVM), significantly outperform traditional bag-of-words models in predicting daily stock returns. Despite these improvements, early approaches were limited in their ability to capture the contextual and semantic nuances of financial language.

Beyond dictionary-based sentiment representations, several studies explored more structured and network-oriented approaches to modeling the relationship between news and market behavior. Curme et al. [2] introduced a coupled network framework linking financial news sentiment with market returns, demonstrating that coupling sentiment networks with price networks improves return predictability. Similarly, Piskorec et al. [14] analyzed the cohesiveness of financial news and showed that financial news cohesiveness strongly correlates with market volatility, highlighting the importance of collective narrative structures rather than isolated sentiment signals.

Further advancements incorporated deep learning techniques to improve sentiment extraction from financial text. Souma et al. [15] demonstrated that neural network–based models outperform traditional lexicon-based approaches in capturing market-relevant sentiment from news articles. A comprehensive evaluation by Mishev et al. [12] systematically compared sentiment analysis methods in finance, ranging from lexicon-based techniques to Transformer architectures, providing empirical evidence of the progressive gains achieved through increasingly contextual language models.

With the advancement of natural language processing techniques, hybrid models that integrate structured numerical data with unstructured textual signals have become increasingly prominent. Domain-specific language models for finance, such as FinBERT, have demonstrated improved capability in capturing financial terminology and contextual dependencies compared to general-purpose language models, thereby bridging the gap between qualitative sentiment analysis and quantitative forecasting. One example of this approach is provided by Gu et al. [6], who proposed the FinBERT-LSTM model for classifying financial news at the market, industry, and individual stock levels, achieving high predictive accuracy by combining textual indicators with historical price data.

Recent research has shifted beyond simple sentiment polarity toward more fine-grained emotion analysis in order to better capture market volatility. Vuong and Gauch [16] employed LLaMA 3.1 to preprocess and filter social media content, demonstrating that emotion-aware analysis—particularly the detection of fear, joy, and surprise—significantly improves the identification of high-volatility price movements. In a related study, Bhat and Jain [1] showed that the emotional tone of financial news headlines alone, analyzed using distilled language models such as DistilRoBERTa-base, can achieve predictive performance comparable to models that rely on traditional financial indicators.

Efficiency and scalability have also emerged as important considerations in financial sentiment analysis. Kim et al. [8] demonstrated that using concise news summaries instead of full-length articles for FinBERT-based sentiment analysis significantly reduces computational cost and data volume while maintaining predictive accuracy. These findings are consistent with the observations of Pennington [13], who argues that Transformer-based models employing multi-head self-attention mechanisms are particularly effective at capturing subtle shifts in market sentiment expressed in financial news headlines. Recent survey work on LLM agents in finance highlights a shift from standalone prediction models toward multi-agent systems for market analysis and decision support, while emphasizing ongoing challenges in interpretability and data integration [4].

Recent studies have additionally explored the application of multi-agent systems (MAS) in financial analysis. Prior work demonstrated the feasibility of agent-based architectures powered by large language models and semantic search to support automated market monitoring and decision support in the cryptocurrency domain [10], motivating the development of more generalizable multi-agent frameworks for continuous financial market analysis. Frameworks such as TradingAgents [17] utilize specialized agents powered by large language models, including roles such as fundamental analysts, sentiment analysts, and risk managers, which engage in structured interactions to evaluate market conditions. By leveraging prompting strategies such as ReAct, these systems aim to enhance explainability and collaborative decision-making in trading-oriented scenarios.

3 System Architecture

The proposed system is based on a modular multi-agent architecture whose primary goal is the integration of structured stock market data and unstructured textual content from financial news. The system architecture is designed to enable parallel processing of numerical and textual information, followed by their subsequent synthesis for the analysis of market dynamics and the identification of news items with a potential impact on stock prices. The overall architecture of the proposed multi-agent system is illustrated in Figure 1.

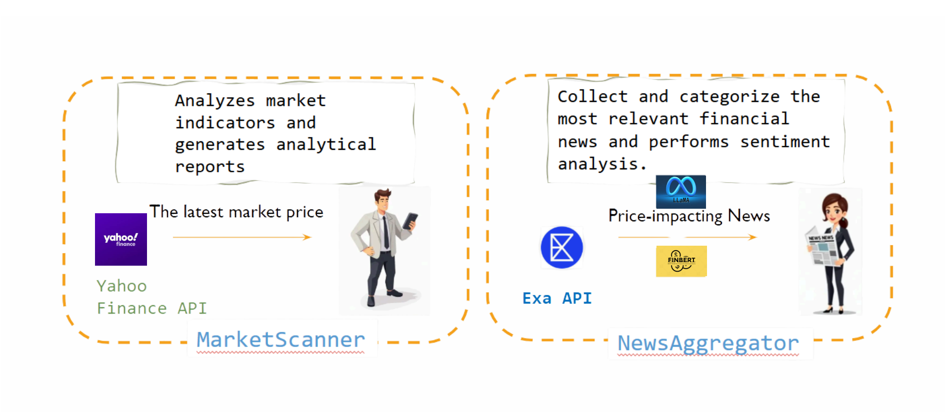

Figure 1. Architecture of the proposed multi-agent system for market monitoring

The multi-agent system consists of two specialized agents, MarketScanner and NewsAggregator, which operate as independent yet coordinated components of the system. Their collaboration enables the combination of quantitative market analysis with qualitative analysis of financial news content.

The MarketScanner agent is responsible for processing structured financial data. Its primary function is the automated analysis of stock price movements, trading volumes, and percentage changes for selected financial instruments, identified by their ticker symbols (e.g., AAPL). Based on the collected data, the agent generates concise and structured reports that provide an overview of current market trends and significant changes. This agent represents the quantitative component of the system and establishes the foundation for understanding the current market state.

The NewsAggregator agent focuses on processing unstructured textual data from financial news sources. Its task is to automatically collect the most recent and relevant news related to a given stock within a defined time window and to analyze this content in order to identify the dominant sentiment. The agent produces structured reports containing key information for each news item, including the headline, publication date, short summary, and estimated sentiment. Depending on the system configuration, sentiment analysis may rely on a traditional financial text processing model or on large language models with contextual reasoning capabilities.

Interaction between the agents is achieved through the exchange of structured data and coordinated task execution. While the MarketScanner provides up-to-date information on market movements, the NewsAggregator contributes qualitative signals derived from news content. By combining these complementary sources of information, the system enables a more comprehensive analysis of market conditions compared to approaches that rely exclusively on a single type of data.

The system architecture was evaluated using data from three companies representing different market sectors: Apple Inc. (AAPL), Bank of America Corporation (BAC), and Exxon Mobil Corporation (XOM). This selection allows the assessment of system behavior across diverse market contexts and contributes to evaluating its overall applicability.

4 Data Sources and Methodology

This study relies on two main data sources: structured stock market data and unstructured textual data from financial news. Numerical data on stock prices and trading volumes were obtained from the Yahoo Finance platform [18], which represents a reliable and widely used source of historical and real-time market information. For each company, daily closing prices (Close) and trading volume (Volume) were collected.

The 24-hour percentage return was computed from closing prices as:

$$R_t^{(c)} = \frac{Close_t^{(c)} - Close_{t-1}^{(c)}}{Close_{t-1}^{(c)}} \times 100,$$

where $Close_t^{(c)}$ denotes the official closing price of company c on trading day t.

To enable alignment with three sentiment categories (Positive, Neutral, Negative), market movement was discretized separately for each company using the standard deviation of daily returns computed over the observed dataset.

Let $\sigma_R^{(c)}$ denote the standard deviation of $R_t^{(c)}$ for company c. Market movement was categorized as follows:

$$\text{Positive movement: } R_t^{(c)} > \sigma_R^{(c)},$$

$$\text{Negative movement: } R_t^{(c)} < -\sigma_R^{(c)},$$

$$\text{Neutral movement: } -\sigma_R^{(c)} \leq R_t^{(c)} \leq \sigma_R^{(c)}.$$

Textual data were collected using the Exa semantic search tool [5], which enables the retrieval of relevant financial news based on the semantic meaning of queries rather than solely on keyword matching. The search process combined company identifiers (ticker symbols) with terms related to market forecasts, price movements, and performance evaluation.

Financial news articles were collected from verified and credible financial and informational sources, including cnn.com, bloomberg.com, marketbeat.com, investing.com, and finance.yahoo.com. The collection and initial filtering of news articles were performed automatically using the proposed multi-agent system, with the NewsAggregator agent responsible for identifying and selecting relevant news for each observed company. To ensure the relevance and quality of the analyzed data, news selection was performed automatically using the Exa semantic search tool in combination with filtering rules implemented within the NewsAggregator agent. The following selection criteria were applied:

- the news item must be directly related to the observed company,
- the content must be complete and informative, excluding short notifications without analytical context,
- the news item must contain sufficient contextual information for sentiment classification,
- the publication time must correspond to the observed period of market trend analysis.

The analysis was conducted for three companies from different market sectors: Apple Inc. (AAPL), Bank of America Corporation (BAC), and Exxon Mobil Corporation (XOM). Sentiment analysis was conducted using two experimental configurations in order to enable a direct comparison between traditional and contemporary methods for financial text processing.

The overall multi-agent architecture remained unchanged across experiments. The difference between configurations concerned exclusively the sentiment analysis module implemented within the NewsAggregator agent. Financial news articles related to the selected stocks were retrieved using the Exa semantic search tool, which enables semantic retrieval based on contextual similarity rather than simple keyword matching.

The multi-agent system was executed on a daily basis during the observation window (June 24–July 8, 2025). At each execution, the NewsAggregator agent retrieved financial news published within the preceding three-day window. The semantic retrieval process was based on a structured query template combining the stock ticker with financial-event and investment-related keywords. The general query structure was defined as follows:

```
(("TICKER stock" AND
  (forecast OR analysis OR "price target" OR upgrade OR downgrade
   OR earnings OR revenue OR guidance OR acquisition OR merger
   OR lawsuit OR "product launch" OR dividend OR buyback OR CEO
   OR outlook))
 OR
 ("Should I buy TICKER stock"
  OR "TICKER stock investment analysis"
  OR "TICKER stock forecast"
  OR "TICKER stock pros and cons"
  OR "Is TICKER a good investment"))
AND
(site:marketbeat.com OR site:finance.yahoo.com
 OR site:cnn.com OR site:investing.com
 OR site:bloomberg.com)
```

The same query structure was applied across all experimental runs, with the placeholder `TICKER` replaced by `AAPL`, `BAC`, or `XOM`. This rolling collection strategy ensured continuous incorporation of recently published content while preserving temporal proximity between news events and market reactions. The final datasets for each experimental configuration represent the aggregation of daily retrieval outputs across the full observation period.

For the first configuration, 37 articles were collected for AAPL, 36 for BAC, and 38 for XOM. For the second configuration, 26 articles were collected for AAPL, 33 for BAC, and 42 for XOM. Throughout the results section, N denotes the number of evaluated news articles for the corresponding company and experimental configuration.

The complete dataset used in this study, consisting of analyzed news articles with publication dates, URLs, and sentiment labels, is available from the authors upon reasonable request.

Although both datasets were gathered within the same temporal window and from comparable financial sources, they do not contain identical news items. However, a substantial overlap exists between the two samples due to the shared collection period and identical query structure.

No manual post-collection removal was performed beyond the automated selection criteria defined in the system. Variations in the number of collected news articles across companies reflect differences in semantic retrieval outcomes and the availability of relevant news during the observed period rather than predefined sample size constraints.

The first approach is based on the application of the FinBERT model[7], a pre-trained BERT model specialized for natural language processing tasks in the financial domain. In this experiment, FinBERT was used exclusively for sentiment classification of financial news into three categories: positive, neutral, and negative, without generating additional explanations for the assigned labels.

The second approach employs large language models (LLMs)[11], specifically the llama3-70b-8192 model. The model was used in a zero-shot setting without fine-tuning. The temperature parameter was set to 0.3 to produce stable and consistent outputs, while the maximum number of generated tokens was limited to 1500. No additional parameter tuning was applied.

Within the same three-class sentiment framework (positive, neutral, negative), the model assigns a sentiment label to each news article and additionally generates a brief summary and an explicit reasoning explaining the assigned sentiment.

This extended output structure enables a more comprehensive examination of the relationship between news sentiment, model-generated reasoning, and short-term stock price dynamics, thereby supporting a deeper contextual interpretation of financial narratives.

The LLaMA-based sentiment analysis was performed using a structured prompt that required the model to generate a title, summary, company-specific sentiment label, and a brief justification. The prompt structure was defined as follows:

```
Read the following financial news article and provide:

1. The title of the article.
2. A brief summary (1-2 sentences).
3. The entity (company) that is the subject of sentiment.
4. The overall sentiment (Positive, Neutral, or Negative).
5. A short reasoning explaining why this sentiment applies.

Respond in the following format:
- Title: ...
- Summary: ...
- Subject: ...
- Sentiment: ...
- Reasoning: ...

Text:
"""[Article Text]"""
```

Sentiment–market alignment was evaluated by comparing the aggregated sentiment signals of articles collected on day t with the categorized market movement for the same trading day. News articles were aligned at the daily aggregation level. For each system execution day t, all news articles retrieved during that run were collectively associated with the corresponding market data of trading day t. Non-trading days (weekends and market holidays) were aligned with the next available trading day.

Coordination among agents was implemented using the CrewAI framework [3], which enables the organization and synchronization of agent activities through clearly defined task assignments. Within the proposed system, CrewAI manages the execution of data collection, content filtering, and sentiment analysis processes, thereby ensuring modularity, extensibility, and reproducibility of the experimental workflow.

To establish the expert ground truth for sentiment labels and to evaluate the quality of generated summaries and justifications, a structured expert review procedure was implemented.

The initial sentiment annotation and quality assessment were performed by the third author. Subsequently, all annotations were reviewed by the second author, who has domain expertise in economics and financial analysis. In cases of uncertainty or disagreement, final decisions were reached through consultation and consensus between the two authors.

As the annotation procedure was based on expert labeling followed by review and consensus-based correction, rather than independent parallel annotation, inter-annotator agreement metrics were not calculated. Evaluation was conducted using classification accuracy as the primary performance metric. Accuracy was computed as the proportion of correctly aligned sentiment–market pairs relative to the total number of evaluated observations:

$$Accuracy = \frac{\text{Number of correct predictions}}{N},$$

where N denotes the number of evaluated news articles for the corresponding company and experimental configuration.

5 Results and Discussion

The experimental evaluation was conducted on three publicly traded companies from different market sectors: Apple Inc. (AAPL), Bank of America Corporation (BAC), and Exxon Mobil Corporation (XOM). For each company, financial news articles were collected and analyzed using the proposed multi-agent system.

Two sentiment analysis approaches were evaluated: a traditional financial NLP model (FinBERT) and a large language model–based approach (LLaMA). The performance of both approaches was assessed with respect to expert-labeled sentiment annotations and their alignment with short-term stock price movements.

Manual expert annotation was used as ground truth, with each news article assigned a sentiment label in the context of the corresponding stock. These expert annotations served as the reference for evaluating sentiment classification accuracy.

In addition to sentiment classification, the LLaMA-based approach was further evaluated in terms of the quality of generated summaries and the correctness of its reasoning, as manually assessed by domain experts. This additional evaluation provides a broader perspective on the interpretability and analytical capabilities of the proposed system. To ensure a consistent experimental setting, both sentiment analysis approaches were evaluated within the same time period and under identical system conditions, although the underlying datasets were not identical due to differences in retrieval outcomes. The evaluation encompassed several task-level aspects, including sentiment classification, alignment between news sentiment and short-term market movements, as well as, specifically for the LLaMA-based approach the quality of generated summaries and the correctness of model reasoning.

Sentiment classification performance was evaluated by comparing model predictions with manually assigned expert sentiment labels for each analyzed news article. A direct comparison was performed between the FinBERT model and the LLaMA3-based approach.

The results are summarized in Table 1, which reports sentiment classification accuracy for the three analyzed companies. For Apple Inc. (AAPL), the LLaMA3-based approach, achieved an accuracy of 80.77%, outperforming FinBERT, which reached 67.57% agreement with expert annotations. This indicates a clear advantage of the reasoning-enabled model for sentiment interpretation in this case.

Table 1. Sentiment classification accuracy comparison between LLaMA3 and FinBERT

	LLaMA3		**FinBERT**	
Company	**N**	**Accuracy (%)**	**N**	**Accuracy (%)**
Apple Inc. (AAPL)	26	80.77	37	67.57
Bank of America (BAC)	33	75.76	36	83.33
Exxon Mobil Corp. (XOM)	42	78.57	38	60.53

For Bank of America (BAC), a different pattern was observed. While LLaMA3 achieved an accuracy of 75.76%, FinBERT slightly outperformed it with 83.33%, suggesting that sentiment expressed in banking-related news may be more consistently captured by domain-specific financial language models.

In the case of Exxon Mobil Corporation (XOM), the LLaMA3-based approach again demonstrated superior performance, achieving 78.57% accuracy compared to 60.53% obtained by FinBERT.

Overall, the results presented in Table 1 indicate that the LLaMA3-based approach provides more stable sentiment classification performance across different market sectors, whereas FinBERT exhibits stronger performance only in specific, domain-consistent contexts. These findings suggest that reasoning-enabled language models are better suited for handling heterogeneous and narrative-rich financial news content.

In addition to sentiment classification, the LLaMA3-based approach was evaluated with respect to the quality of generated summaries and the correctness of its reasoning, emphasizing the interpretability and explanatory capabilities of the model.

Reasoning accuracy reflects how well the generated explanations justify the assigned sentiment within the financial context of each news article, while summary accuracy measures the extent to which the automatically generated summaries capture the key information of the original content. Both aspects were manually assessed by domain experts. The evaluation results, summarized in Table 2, indicate consistently high performance across all analyzed companies. For Apple Inc. (AAPL), the model achieved particularly strong results in both reasoning correctness and summary reliability. Comparable performance was observed for Bank of America (BAC) and Exxon Mobil Corporation (XOM), demonstrating that the LLaMA-based approach consistently provides meaningful and interpretable explanations alongside accurate summaries.

Table 2. Reasoning and summary accuracy of the LLaMA3-based approach

Company	**N**	**Reasoning Accuracy (%)**	**Summary Accuracy (%)**
Apple Inc. (AAPL)	26	88.46	100.00
Bank of America (BAC)	33	87.88	90.91
Exxon Mobil Corp. (XOM)	42	76.19	78.57

Overall, these results support the suitability of reasoning-enabled large language models for analytical decision-support scenarios in the financial domain, where transparency and interpretability are essential.

The temporal association between expert-labeled news sentiment and short-term stock price movements was analyzed by examining sentiment–market alignment over a three-day window. Specifically, sentiment was compared with categorized market returns corresponding to the system execution day (summary-level alignment), as well as with returns one and two days prior to the observed price change. The results are summarized in Table 3. Summary-level sentiment refers to alignment computed using the market return corresponding to the system execution day, where sentiment labels of articles collected on that day were evaluated against the categorized market movement. The temporal analysis was conducted on the dataset corresponding to the second experimental configuration (LLaMA-based sentiment analysis), using expert-validated sentiment labels.

Table 3. Temporal association between expert-labeled news sentiment and stock price trends (three-class framework)

Time reference	**AAPL**	**BAC**	**XOM**
Summary-level sentiment	57.69% (N=26)	42.42% (N=33)	73.81% (N=42)
Same day	59.09% (N=22)	40.74% (N=27)	60.71% (N=28)
One day before	54.55% (N=22)	52.63% (N=25)	37.93% (N=29)
Two days before	50.00% (N=18)	36.84% (N=19)	45.00% (N=20)

Note: N denotes the number of evaluated sentiment–market pairs for each company and temporal offset. N varies due to exclusion of non-trading days without corresponding market data.

The first row ("Summary-level sentiment") reflects alignment based on the market return corresponding to the system execution day. If the execution day fell on a non-trading day, the most recent available trading day was used.

The temporal offset analyses (“Same day”, “One day before”, and “Two days before”) were conducted independently from the summary-level analysis. For each offset, sentiment was compared strictly with the corresponding trading day return. If market data were unavailable for a given reference day due to market closure, the corresponding observation was excluded from the analysis rather than artificially realigned. Overall, the results do not indicate a consistent temporal decay effect across companies. While AAPL exhibits a gradual decline in alignment for more distant news, BAC demonstrates a reverse tendency, with slightly higher alignment observed two days before the price change. XOM shows intermediate and variable behavior across temporal offsets.

To assess whether differences across temporal offsets were statistically significant, Fisher’s exact test was applied separately for each company. Comparisons between the same-day, one-day, and two-day offsets did not reveal statistically significant differences ($p > 0.05$ in all cases). Specifically, for the same-day versus two-day comparison, the results were as follows: AAPL: p = 0.75; BAC: p = 0.55; XOM: p = 0.38. These findings indicate that the observed temporal variations should be interpreted descriptively given the limited sample sizes.

These findings suggest that sentiment–market alignment may be sensitive to company-specific dynamics rather than uniformly driven by temporal proximity. Larger samples and longer observation windows would be required to more robustly assess potential temporal effects.

The alignment between predicted news sentiment and short-term stock price movements was evaluated for both sentiment analysis approaches within the three-class framework. Alignment was defined as the agreement between the sentiment label (positive, neutral, negative) and the corresponding categorized market movement. Market movement was measured using the 24-hour return corresponding to the system execution day on which the news articles were collected.

As shown in Table 4, model performance differs across companies. LLaMA3 achieves higher sentiment–market alignment for XOM, while FinBERT performs slightly better for AAPL. For BAC, the two approaches exhibit comparable performance.

Table 4. Sentiment–market alignment comparison between LLaMA3 and FinBERT (three-class framework)

Model	**AAPL**	**BAC**	**XOM**
LLaMA3	46.15% (N=26)	39.39% (N=33)	64.29% (N=42)
FinBERT	54.15% (N=37)	38.89% (N=36)	34.21% (N=38)

To assess whether these differences were statistically significant, Fisher’s exact test was applied separately for each company. The differences were not statistically significant for AAPL ($p > 0.05$) or BAC ($p > 0.05$). However, for XOM the difference was statistically significant ($p \approx 0.02$), indicating stronger sentiment–market alignment of the LLaMA-based approach for this company.

Overall, the findings suggest that sentiment–market alignment depends on company-specific dynamics rather than on a clear superiority of one modeling approach. The results highlight the importance of contextual and sectoral factors in financial sentiment analysis.

In this study, news articles were collected and aggregated at the daily level, considering all items published within a 24-hour period. As a result, the analysis does not explicitly distinguish between news released during and outside trading hours. While this approach reflects real-world information availability, future work may explore more fine-grained temporal alignment with intraday market activity.

6 Conclusions

This paper presented a multi-agent system for automated monitoring and analysis of stock market dynamics, designed to integrate structured financial data with sentiment analysis of financial news. By combining quantitative market indicators with qualitative signals extracted from textual content, the proposed system enables systematic examination of sentiment–market relationships in short-term trading contexts. The system architecture, implemented within the CrewAI framework, demonstrates the feasibility of coordinating specialized agents to support continuous market monitoring and structured analytical reporting. The experimental evaluation indicates that large language models provide advanced reasoning and summary generation capabilities, contributing to improved interpretability of financial news. However, sentiment–market alignment results vary across companies. While the LLaMA-based approach demonstrates a statistically significant advantage for one company, no significant differences were observed for the others. These findings suggest that model performance depends on company-specific dynamics rather than on a consistent superiority of a particular modeling approach.

Temporal analysis indicates that alignment patterns differ across offsets; however, statistical testing did not confirm consistent temporal decay effects within the evaluated sample. This highlights the importance of cautious interpretation when assessing short-term sentiment–market relationships.

Despite these findings, several limitations should be acknowledged. The evaluation was conducted on a limited number of companies and within a restricted temporal window, which may constrain the generalizability of the results. In addition, the system relies on a predefined set of data sources and sentiment models and does not explicitly incorporate broader macroeconomic variables or longer-term market dependencies. Furthermore, the statistical analysis was performed on relatively small samples, which may limit the power to detect subtle temporal or model-specific effects. These considerations suggest that the reported results should be interpreted as exploratory rather than definitive.

Future work will focus on extending the system to a larger number of companies and a longer observation period in order to further examine the stability and general applicability of the obtained results. In addition, the inclusion of additional data sources and alternative sentiment analysis models may be considered to improve the robustness of the system. Such enhancements would contribute to further validation of the proposed approach without altering the core system architecture. Future work may explore finer-grained temporal analysis to better capture the relationship between the timing of news publication and market reactions.

References

[1] R. BHAT, B. JAIN, *Stock Price Trend Prediction using Emotion Analysis of Financial Headlines with Distilled LLM Model*, in Proceedings of the 17th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '24), Association for Computing Machinery, New York, NY, USA, 2024, pp. 67–73.

[2] C. CURME, H. E. STANLEY, I. VODENSKA, *Coupled network approach to the predictability of financial market returns and news sentiments*, International Journal of Theoretical and Applied Finance, vol. 18, no. 7, 2015.

[3] CREWAI, Official documentation, available at: `https://docs.crewai.com`, last accessed Jan. 29, 2026.

[4] Y. DONG, F. WU, K. ZHANG, Y. DAI, S. ZHANG, W. YE, S. CHEN, Z.-Q. CHENG, *Large language model agents in finance: A survey bridging research, practice, and real-world deployment*, Findings of the Association for Computational Linguistics: EMNLP, vol. 2025, pp. 17889–17907, 2025.

[5] EXA AI, available at: `https://docs.exa.ai`, last accessed Jan 22, 2026.

[6] W. GU, Y. ZHONG, S. LI,C. WEI,L. DONG, Z. WANG, C. YAN, *Predicting Stock Prices with FinBERT-LSTM: Integrating News Sentiment Analysis*, in Proceedings of the 2024 8th International Conference on Cloud and Big Data Computing (ICCBDC '24), Association for Computing Machinery, New York, NY, USA, 2024, pp. 67–72.

[7] A. H. HUANG, H. WANG, Y. YANG, *FinBERT: A large language model for extracting information from financial text*, Contemporary Accounting Research, vol. 40, no. 2, pp. 806–841, 2023.

[8] J. KIM, H.-S. KIM, S.-Y. CHOI, *Forecasting the S&P 500 Index Using Mathematical-Based Sentiment Analysis and Deep Learning Models: A FinBERT Transformer Model and LSTM*, Axioms, vol. 12, no. 9, Art. no. 835, 2023.

[9] X. LI, H. XIE, L. CHEN, J. WANG, X. DENG, *News impact on stock price return via sentiment analysis*, Knowledge-Based Systems, vol. 69, pp. 14–23, 2014.

[10] U. MAROVAC, D. PRAMENKOVIĆ, A. HAMZAGIĆ, *Application of a multi-agent system for monitoring market trends on the example of cryptocurrencies*, in Proceedings of the 15th International Conference on Information Society and Technologies (ICIST 2025), Mar. 9–12, Kopaonik, Serbia, 2025.

[11] META AI, Meta LLaMA 3, available at: `https://ai.meta.com/blog/meta-llama-3/`, last accessed Jan. 29, 2026.

[12] K. MISHEV, A. GJORGJEVIKJ, I. VODENSKA, L. T. CHITKUSHEV, D. TRAJANOV, *Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers*, IEEE Access, vol. 8, pp. 131662–131682, 2020.

[13] A. PENNINGTON, *Stock Market News Sentiment Analysis and Trend Prediction Using Transformer Models*, Journal of Computer Science and Software Applications, vol. 5, no. 8, 2025.

[14] M. PISKOREC, N. ANTULOV-FANTULIN, P. K. NOVAK,I. MOZETIC, M. GRCAR,I. VODENSKA, T. SMUC, *Cohesiveness in Financial News and Its Relation to Stock Market Volatility*, Scientific Reports, vol. 4, Art. no. 5038, 2014.

[15] W. SOUMA, I. VODENSKA, H. AOYAMA, *Enhanced news sentiment analysis using deep learning methods*, Journal of Computational Social Science, pp. 1–14, 2019.

[16] A. VUONG S. GAUCH, *Predicting Stock Price Movement with LLM-Enhanced Tweet Emotion Analysis*, arXiv preprint arXiv:2510.03633, 2025.

[17] Y. XIAO, E. SUN,D. LUO, W. WANG, *TradingAgents: Multi-agents LLM Financial Trading Framework*, arXiv preprint arXiv:2412.20138, 2024.

[18] YAHOO FINANCE, available at: `https://finance.yahoo.com/`, last accessed Jan. 22, 2026.