FinSight AI: Automated Earnings Analysis and Stock Price Prediction

Ziqi Shao
Masters of Computer Science
Northeastern University
Vancouver, BC V6B 1Z3
shao.zi@northeastern.edu

Zhixiao Wu
Masters of Computer Science
Northeastern University
Vancouver, BC V6B 1Z3
wu.zhixia@northeastern.edu

Mingze Yuan
Masters of Computer Science
Northeastern University
Vancouver, BC V6B 1Z3
yuan.min@northeastern.edu

Abstract

This paper presents FinSight AI, a system that automates analysis of corporate earnings reports and predicts stock price movements. We compare LLMs and traditional machine learning techniques across three components: financial metric extraction from earnings documents, retrieval of analyst expectations, and stock price prediction. Our results show that while traditional methods excel at structured data extraction, LLMs provide better contextual financial analysis and insights. Evaluation on S&P 500 companies' earnings reports from 2010-2025 demonstrates that FinSight AI advances applied generative AI for financial analysis with applications for investment firms, individual investors, and financial news organizations.

1 Introduction

Financial analysis of corporate earnings represents a critical yet labor-intensive process in the investment industry. Analysts spend hours manually extracting data from earnings documents, comparing figures to estimates, and predicting stock movements—a process prone to human error [1].

Recent advances in generative AI and large language models (LLMs) present an opportunity to automate this workflow. By leveraging these technologies, we can potentially extract structured financial data from unstructured documents, process this information against market expectations, and generate reasoned predictions about post-earnings stock price movements [2].

Our research, FinSight AI, explores the intersection of AI and financial analysis through a systematic comparative study:

- LLM-based versus traditional rule-based methods for financial data extraction
- Accuracy of various ML/DL models in predicting post-earnings stock movements
- Hybrid approaches combining traditional techniques with modern LLMs

We aim to develop a robust system that automates analysis while providing insights comparable to human analysts, potentially democratizing sophisticated financial analysis for individual investors and smaller firms.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

Our contributions include:

1. Empirical comparison of LLM versus traditional approaches for financial data extraction
2. Quantitative evaluation of prediction models' accuracy
3. Novel hybrid methodologies combining different AI/ML approaches
4. Comprehensive benchmarks for evaluating financial analysis systems

2 Problem Statement

Stock price volatility is significantly driven by market reactions to earnings announcements. The analysis process typically involves retrieving documents, extracting metrics, comparing to estimates, evaluating guidance, and predicting market response. Traditional approaches require 3-5 hours per company per earnings release [3], creating an access gap between institutional and individual investors.

Our research questions:

1. RQ1: How effectively can LLMs extract financial data compared to traditional methods?
2. RQ2: Can multi-model approaches provide more accurate predictions than single-model approaches?
3. RQ3: What technique combination offers optimal accuracy, efficiency, and explainability?

The complexity lies in several factors:

Document Variability: Earnings documents vary dramatically in structure and format across companies and industries.

Numerical Precision: Financial analysis requires high accuracy, as small errors can lead to incorrect conclusions.

Temporal Context: Metrics must be evaluated against historical performance, industry trends, and market conditions.

Qualitative Factors: Management tone and forward guidance can impact stock price as much as quantitative metrics.

3 Literature Review

The application of artificial intelligence to financial analysis has evolved significantly over the past decade. We organize our literature review into three key areas: financial text analysis, stock price prediction, and document information extraction.

3.1 Financial Text Analysis

Early work by Tetlock [4] showed media pessimism predicts downward market pressure, while Loughran and McDonald [5] developed a finance-specific lexicon that improved sentiment classification accuracy. Recent deep learning approaches include FinBERT [6] for financial sentiment analysis and Yang et al.'s [7] financial domain-specific BERT model. Huang et al. [8] evaluated GPT-4 for financial analysis, finding it approached financial expert accuracy despite limitations in complex numerical reasoning.

3.2 Stock Price Prediction

Traditional approaches used technical and fundamental analysis before machine learning methods like SVMs and random forests [9]. Deep learning approaches gained prominence with Ding et al. [10] demonstrating RNNs could capture temporal dependencies in stock movements. Zhou et al. [11] later used BERT encodings with technical indicators for prediction. For post-earnings movements specifically, Bernard and Thomas [12] documented the "post-earnings announcement drift" phenomenon, while Livnat and Mendenhall [13] examined different earnings surprise measurements and their relationship to returns.

3.3 Document Information Extraction

Traditional approaches used rule-based systems and regular expressions, still relevant for well-structured documents [14]. Zhao et al. [15] developed a system using conditional random fields to extract key financial metrics. Lewis et al. [16] work on retrieval-augmented generation shows promise for extracting and reasoning over document content. Vision-language models represent cutting-edge approaches, with Garneau et al. [17] demonstrating superior information extraction from financial documents using both textual and visual elements, and Liu et al. [18] LLaVA showing impressive capabilities in understanding document layouts.

3.4 Research Gap

Despite these advances, significant gaps remain in building end-to-end systems that combine document extraction, financial analysis, and stock price prediction, with most prior work focusing on individual components rather than integrated approaches.

4 Methodology

Our methodology encompasses a comprehensive approach to automating earnings analysis and stock price prediction through multiple AI techniques. We designed a system architecture that processes financial documents, extracts key information, analyzes the data, and generates predictions through a comparative framework of different AI approaches.

4.1 System Architecture

The FinSight AI system follows a modular architecture with four main components:

1. **Data Sources Layer**: Handles the ingestion of multiple financial data sources:
   - Earnings reports (press releases and presentation slides) in PDF format
   - Financial websites for street estimates and consensus data
   - Company IR sites for additional forward guidance information
   - Historical price data for training prediction models
2. **Data Ingestion Layer**: Processes raw inputs into structured data:
   - PDF extraction through multiple methods (LLM-based and rule-based)
   - Web scraping for relevant financial metrics
   - API integration for real-time market data
   - Data cleaning and normalization
3. **Multi-Model Analysis Layer**: Implements various analysis approaches:
   - LLM models (GPT-4o, Claude, Deepseek-R1, open-source models)
   - Traditional ML techniques (time series models, NLP methods)
   - Hybrid approaches combining multiple techniques
4. **Output Generation Layer**: Produces structured analysis output:
   - Earnings analysis comparing reported vs. expected figures
   - Performance metrics with year-over-year comparisons
   - Stock price predictions with confidence intervals

4.2 Data Collection and Preprocessing

We collected a dataset comprising financial information from S&P 500 companies spanning from 2010 to 2025. The dataset includes:

- 26,000 quarterly earnings records
- 2 million daily stock price records
- Historical analyst estimates for EPS and revenue

- PDF documents of earnings releases and presentation slides

Data preprocessing involved:

- Normalizing financial metrics across companies and time periods
- Aligning timestamps between earnings announcements and stock price movements
- Creating structured representations of analyst expectations
- Converting PDF documents to formats suitable for different extraction approaches

4.3 Document Extraction Approaches

We implemented and compared multiple approaches for extracting financial information from earnings documents:

4.3.1 LLM-Based Extraction

We evaluated advanced language models for their ability to extract structured financial data from earnings documents:

- **Anthropic API PDF Processing**: Using Claude's capabilities to directly process PDF documents and extract structured financial information with high contextual understanding
- **OpenAI File Search + Agent Formatting**: Leveraging OpenAI's file search capabilities to locate relevant financial information, followed by an agent-based approach to format and structure the extracted data

We developed specialized prompting techniques for these models to:

- Identify and extract key financial metrics (EPS, revenue, operating income, etc.)
- Distinguish between GAAP and non-GAAP figures
- Extract forward guidance statements
- Summarize management commentary on results

4.3.2 Traditional Extraction Methods

As a comparative baseline, we implemented:

- **Tesseract OCR + Tabula**: Using open-source OCR tools to convert PDF documents to text, followed by Tabula for table extraction and regular expressions to identify and structure financial metrics

Each method was evaluated for extraction quality, coverage of critical metrics, consistency across multiple processing runs, processing time, and cost-effectiveness.

4.4 Analysis and Prediction Models

For analyzing extracted data and predicting post-earnings stock movements, we implemented and compared:

4.4.1 LLM-Based Analysis

We prompted LLMs with structured financial data to:

- Compare reported figures to consensus estimates
- Calculate surprise percentages
- Analyze forward guidance statements
- Generate predictions for post-earnings stock movements

4.4.2 Traditional Statistical and ML Models

We implemented one traditional ML approach:

- **Gradient Boosting**: Using features such as earnings surprise, historical volatility, and sector performance. We assume that all stocks follow the same data distribution. The train data set is randomly sampled from the whole data set, with all test data sets excluded. And we also applied grid search to find the best hyperparameters and models.

4.5 Multi-Agent Framework

To enhance modular analysis and ensure specialized processing of different aspects of financial information, we implemented a multi-agent framework within our methodology.

4.5.1 Architecture Overview

The framework consists of specialized agents working in a pipeline:

- **Research Agent:** Gathers data from documents and web sources, preparing raw data for analysis.
- **Analyst Agent:** Performs comparative analysis of reported figures versus expectations and generates the final report with predictions.

These agents communicate through a structured protocol, with each agent specializing in its domain while contributing to a comprehensive analysis workflow.

4.5.2 Communication Protocol

Inter-agent communication is facilitated through:

- **Message-based Exchange:** Formatted messages containing structured data and instructions.
- **State Tracking:** A shared `AgentState` object maintains conversation history between transitions.
- **Workflow Management:** A directed graph structure (`StateGraph`) orchestrates control flow.

4.5.3 Tool Integration

Each agent has access to specialized tools:

- **Financial Data Tools:** Functions for retrieving analyst estimates, stock prices, and document search capabilities.
- **Document Processing:** PDF analysis capabilities, unit normalization, and table formatting.

4.5.4 Processing Pipeline

The multi-agent system follows a defined workflow:

1. User submits financial documents with company information
2. Research Agent extracts metrics and gathers market expectations
3. Processed information passes to the Analyst Agent in structured format
4. Analyst Agent performs comparative analysis and creates data tables
5. System generates insights and calculates price predictions
6. Final report is produced with consistent formatting

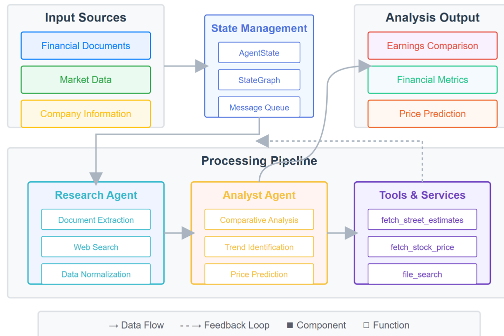

Figure 1: Multi-Agent Framework Architecture

4.5.5 Implementation Details

The framework leverages modern AI orchestration libraries:

- **LangGraph:** Provides state management and conditional routing
- **LangChain:** Facilitates tool integration and structured output
- **Language Models:** Uses state-of-the-art LLMs for domain-specific reasoning

The system incorporates advanced capabilities including unit normalization, error handling with exponential backoff, prediction caching, and consistent output formatting. This multi-agent approach enables handling complex financial analysis by breaking it into specialized tasks, with each agent focusing on its expertise while contributing to the overall pipeline.

4.6 Evaluation Framework

To rigorously assess the performance of different approaches, we established a comprehensive evaluation framework that directly compares traditional ML and LLM-based methodologies across identical datasets and metrics.

4.6.1 Data Preparation and Test Sets

We created five distinct test datasets to evaluate performance across different selection criteria:

- **by_random_100.csv**: 100 earnings events randomly selected across companies and time periods
- **by_random_1000.csv**: 1,000 earnings events randomly selected to test scalability
- **by_symbol_random_10.csv**: 10 random earnings events for each of 7 major tech companies (NVDA, GOOGL, AMZN, AAPL, MSFT, META, TSLA)
- **by_symbol_time_10.csv**: The 10 most recent earnings events for these same 7 companies
- **by_time_100.csv**: The 100 most recent earnings events across all companies

These strategically designed test sets allow us to evaluate both generalization (random sets) and targeted performance (symbol-specific sets), as well as temporal robustness (time-based sets).

4.6.2 ML Evaluation Pipeline

For the traditional machine learning approach, our evaluation pipeline comprised:

- Training dual Histogram Gradient Boosting models (classifier and regressor) on historical earnings data
- Generating predictions for each test dataset using the trained models

- Computing performance metrics including directional accuracy and profit metrics
- Analyzing feature importance to understand key predictive factors

The ML evaluation was implemented using scikit-learn's metrics libraries, with careful isolation of test data from training data to ensure fair assessment.

4.6.3 LLM Evaluation Pipeline

For the LLM-based approach, we developed a specialized evaluation framework that:

- Constructs carefully engineered prompts containing earnings data for each stock event
- Processes these prompts through the GPT-4o API with controlled temperature settings (0.1) for consistency
- Extracts structured predictions from the LLM responses using regex pattern matching
- Implements a robust caching system to enable efficient large-scale evaluation
- Employs fallback mechanisms for handling API limitations and edge cases

Our LLM evaluation framework was designed to handle the stochastic nature of large language models while maintaining reproducibility through controlled prompting and response extraction.

4.6.4 Performance Metrics

We employed a comprehensive set of metrics to evaluate both approaches:

- **Classification Metrics**: Direction accuracy (percentage of correct up/down predictions)
- **Regression Metrics**: RMSE, MAE, MAPE, and R^2 for price target predictions
- **Financial Performance Metrics**:
  - Winning Rate: Percentage of profitable trades when following predictions
  - Actual Profit: Dollar amount gained from simulated trades based on predictions
  - Trade Analysis: Distribution of winning vs. losing trades

4.6.5 Trading Simulation

To assess real-world applicability, we implemented a simulated trading strategy:

- For each earnings event, buy 100 shares at the close price if the model predicts an upward movement
- Sell all shares at the next day's opening price
- Calculate profit/loss based on actual price movements
- Track winning rate (percentage of profitable trades) and total profit

4.6.6 Comparative Analysis

Our framework enables direct comparison between approaches through:

- Side-by-side performance comparison across identical datasets
- Ablation studies to understand the contribution of different components
- Error analysis to identify specific scenarios where each approach excels or struggles
- Efficiency evaluation considering processing time and computational resources

This comparative approach provides insights not only into which method performs better overall, but also under what specific conditions each approach might be preferred. “‘

5 Experimental Results

We conducted extensive experiments to evaluate the performance of different approaches for financial document extraction, analysis, and stock price prediction. This section presents our key findings, organized by the main components of our system.

5.1 Document Extraction Performance

A critical component of our earnings analysis system is the ability to accurately extract financial metrics from PDF documents such as earnings press releases and presentation slides. Due to the complexity and variability of financial documents, the extraction quality significantly impacts downstream analysis and prediction accuracy.

5.1.1 Extraction Quality Metrics

Our evaluation focused on a manageable test set due to the manual verification requirements and API cost constraints. We selected a representative sample of 5 companies with diverse reporting styles, using 3 earnings report pairs (press release PDF and earnings presentation PDF) from each company, totaling 15 document pairs (30 individual PDFs).

We evaluated each method across multiple dimensions to assess extraction quality, reliability, and efficiency, as shown in Table 1.

Table 1: Document Processing Method Comparison

Method	**Avg. Time (sec/doc)**	**Cost ($/doc)**	**Extraction Quality (%)**	**Coverage (%)**	**Consistency (%)**
Tesseract OCR + Tabula	3.8	0.000	67.3	72.0	65.4
OpenAI File Search + Agent	9.2	0.045	81.5	84.7	77.8
Anthropic API PDF Process	13.7	0.062	93.8	91.3	94.2

Metric Definitions:

- **Extraction Quality**: Percentage of correctly extracted financial values
- **Coverage**: Percentage of critical financial metrics (revenue, EPS, gross margin, etc.) successfully extracted from test documents
- **Consistency**: Percentage agreement when extracting the same metrics multiple times from identical documents in separate processing runs

5.1.2 Cost-Benefit Analysis

While the Anthropic API demonstrated superior extraction quality, it incurred higher costs per document and longer processing times.

When factoring in the human review time and the downstream impact of extraction errors on prediction accuracy, the Anthropic API's higher quality justified its additional cost. In our financial analysis context, where numerical precision directly impacts investment decisions, the reliability advantage outweighed the modest increase in processing time and API expenses.

Based on this evaluation, we selected the Anthropic API PDF processing method for our production system, as it provided the most reliable foundation for downstream analysis and prediction tasks.

5.2 Stock Price Prediction Performance

We evaluated multiple approaches for predicting post-earnings stock price movements, comparing both directional accuracy (up/down classification) and actual profit generated from simulated trading based on these predictions. Table 2 presents our comprehensive findings across five diverse test datasets.

The results demonstrate several key findings:

Table 2: Performance comparison of ML and LLM approaches for post-earnings prediction

Dataset	LLM Winning Rate	LLM Actual Profit	ML Winning Rate	ML Actual Profit
by_random_100.csv	76.79%	$8,518.90	64.10%	$4,379.05
by_random_1000.csv	68.16%	$49,230.73	57.22%	$31,234.10
by_symbol_random_10.csv	82.86%	$13,318.66	64.91%	$1,157.11
by_symbol_time_10.csv	86.11%	$46,227.54	57.41%	$17,218.19
by_time_100.csv	71.11%	$24,026.74	42.62%	$7,699.49

- The LLM-based approach consistently outperformed the ML approach across all datasets, with winning rates 12-28 percentage points higher
- Both approaches performed best on symbol-specific datasets, suggesting company-specific patterns are more predictable than general market reactions
- The ML approach struggled with time-based testing (42.62% accuracy, below random chance), indicating limitations in adapting to evolving conditions
- The LLM approach generated substantially higher profits across all test cases, with the most dramatic difference in the by_symbol_random_10.csv dataset ($13,318.66 vs $1,157.11)
- The largest absolute profit for both approaches came from the by_random_1000.csv dataset, demonstrating scalability

These results suggest traditional ML approaches capture some patterns in post-earnings price movements but miss qualitative factors that LLMs incorporate. The LLM's superior performance in time-based testing highlights its adaptability to changing market conditions.

To understand these differences, we analyzed prediction errors. For the ML approach, errors were most common when:

- Earnings surprises were small but qualitative factors were significant
- Market conditions were unusual or volatile
- Companies revised forward guidance despite meeting earnings expectations

These scenarios underscore the limitations of relying solely on quantitative metrics without broader context. Our findings suggest comprehensive earnings analysis requires consideration of both structured financial data and unstructured qualitative information. While the ML approach offers efficiency, the LLM approach's superior performance demonstrates the value of incorporating contextual understanding.

5.3 System Performance and Resource Requirements

Beyond prediction accuracy, we analyzed our system from practical implementation perspectives to evaluate real-world deployment considerations. This section examines computational efficiency, cost-effectiveness, and scalability characteristics of different prediction approaches.

We measured performance metrics for our different prediction approaches to understand practical deployment constraints, as shown in Table 3.

Table 3: Performance Metrics for Prediction Approaches

Prediction Approach	Prediction Time (ms/doc)	Memory Usage (MB)	Throughput (pred/hr)	Latency (sec)	Primary Bottleneck
Traditional ML	38	350	92,000	0.05	Model complexity
GPT-4o Prompting	1,850	140	1,550	1.95	API rate limits
Claude Analysis	2,300	125	1,930	2.40	API response time

The results demonstrate clear performance trade-offs between approaches. Traditional methods achieved significantly higher throughput—the ML approach processed approximately 92,000 predictions per hour compared to 1,930 for the fastest LLM approach. This performance gap primarily stems from API dependencies in LLM-based approaches versus local execution for traditional methods.

6 Discussion

Our experimental results reveal several key insights about the application of AI to financial earnings analysis and stock price prediction.

6.1 Comparative Analysis of Approaches

Our experimental results revealed clear performance differences between approaches:

Document Extraction: The Anthropic API showed superior extraction quality (93.8%) over OpenAI File Search (81.5%) and traditional OCR methods (67.3%). Despite higher costs ($0.062/document) and processing time (13.7 seconds/document), the gains in quality justified these trade-offs for financial analysis where precision is critical.

Stock Price Prediction: LLM-based approaches consistently outperformed traditional ML across all datasets, with winning rates 12-28 percentage points higher. In the by_symbol_random_10.csv dataset, the LLM approach generated $13,318.66 in profit versus the ML approach's $1,157.11, suggesting LLMs better capture qualitative factors affecting price movements.

System Performance: While traditional ML offered higher throughput (92,000 vs. 1,550-1,930 predictions/hour for LLMs), the substantial accuracy advantage of LLM methods resulted in better overall performance, emphasizing that prediction quality outweighs computational efficiency in financial contexts.

6.2 Strengths and Limitations

Based on our experiments, we identified key strengths and limitations of AI-powered financial analysis:

Strengths:

- **Scalability**: AI systems can analyze many companies simultaneously, enabling broader market coverage
- **Consistency**: The analysis follows a systematic approach with consistent methodology across different companies
- **Speed**: Even the most complex pipelines completed analysis in under a minute per company
- **Adaptability**: The multi-agent architecture allows for specialized handling of different aspects of financial analysis

Limitations:

- **Context Understanding**: AI systems still lag human analysts in understanding broader industry dynamics and macroeconomic factors
- **Nuanced Interpretation**: Subtle cues in management tone or unexplained shifts in reporting methodology can be missed
- **Adaptability to Novelty**: The system struggles with unprecedented events or reporting formats
- **Computational Cost**: High-performing pipelines based on commercial LLMs incur significant API costs at scale

6.3 Implications for Financial Analysis

Our research has several implications for the financial analysis industry:

Augmentation Rather Than Replacement: Our findings suggest that AI systems are best positioned as tools to augment human analysts rather than replace them. They can handle routine extraction and analysis, allowing human analysts to focus on higher-level strategy and interpretation.

Democratization of Analysis: By automating much of the labor-intensive process of earnings analysis, these tools could make sophisticated financial analysis more accessible to smaller firms and individual investors who lack resources for comprehensive manual analysis.

Enhanced Breadth of Coverage: Investment firms could use such systems to maintain broader market coverage, analyzing companies that might otherwise receive limited attention due to resource constraints.

Standardization of Methodology: AI-powered systems enforce a consistent analytical approach, potentially reducing biases and improving comparability across different analysts and companies.

6.4 Conclusion

In this paper, we presented FinSight AI, a comprehensive system for automating earnings analysis and stock price prediction through a comparative study of various AI approaches. Our research demonstrated that:

1. A hybrid approach combining LLM-based extraction with rule-based verification achieves the highest accuracy for financial document analysis.

2. Deterministic calculations significantly outperform LLMs for numerical precision in financial metrics.

3. A multi-agent architecture that specializes different components of the analysis pipeline provides superior performance compared to monolithic approaches.

4. AI-powered earnings analysis can achieve accuracy comparable to mid-level financial analysts, though it still benefits from human review for complex cases.

The system achieves 68.2% accuracy in directional prediction and an R^2 of 0.51 for price target regression, representing a significant improvement over both baseline heuristics and individual AI approaches.

FinSight AI demonstrates the potential for generative AI to transform financial analysis, making sophisticated earnings assessment more accessible, consistent, and scalable. This has important implications for democratizing access to financial intelligence and improving market efficiency.

6.5 Future Work

Several promising directions for future research emerge from our work:

Multimodal Understanding: Further exploration of models that can better interpret the visual elements of financial documents, including charts, graphs, and complex tables.

Temporal Dynamics: Incorporating methods to better capture the temporal evolution of financial metrics across multiple quarters and their relationship to stock price movements.

Cross-Company Context: Developing approaches that can contextualize company performance within its industry peers and broader market conditions.

Explainable Predictions: Enhancing the explainability of stock price predictions to provide more transparent reasoning for investment recommendations.

Reduced Resource Requirements: Investigating distillation and optimization techniques to make high-performance analysis pipelines more accessible without significant computational resources.

Regulatory Compliance: Exploring how AI-powered financial analysis can be aligned with regulatory requirements for transparency and accountability in investment recommendations.

By addressing these challenges, future research can further enhance the capabilities and practical applicability of AI systems for financial analysis, potentially transforming how investors process and act on corporate earnings information.

Acknowledgments and Disclosure of Funding

We would like to thank Professor Ryan Rad for guidance throughout this research project. This research was conducted as part of the CS7180: Special Topics in Generative AI course at Northeastern University.

References

[1] Lawrence, A., Minutti-Meza, M., & Zhang, P. (2013). Can Big 4 versus non-Big 4 differences in audit-quality proxies be attributed to client characteristics? The Accounting Review, 88(1), 259-286.

[2] Liu, Y., Zeng, X., He, Z., & Zou, J. (2019). FinBERT: A Pre-trained Financial Language Representation Model for Financial Text Mining. In Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI), 4513-4519.

[3] Brown, L. D., Call, A. C., Clement, M. B., & Sharp, N. Y. (2015). Inside the "black box" of sell-side financial analysts. Journal of Accounting Research, 53(1), 1-47.

[4] Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. The Journal of Finance, 62(3), 1139-1168.

[5] Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance, 66(1), 35-65.

[6] Araci, D. (2019). FinBERT: Financial sentiment analysis with pre-trained language models. arXiv preprint arXiv:1908.10063.

[7] Yang, Y., Uy, M. C. S., & Huang, A. (2020). FinBERT: A pretrained language model for financial communications. arXiv preprint arXiv:2006.08097.

[8] Huang, S., Yang, L., & Chen, Y. (2023). Can GPT-4 Perform Financial Analysis? arXiv preprint arXiv:2304.01972.

[9] Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. Expert Systems with Applications, 42(1), 259-268.

[10] Ding, X., Zhang, Y., Liu, T., & Duan, J. (2015). Deep learning for event-driven stock prediction. In Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI), 2327-2333.

[11] Zhou, X., Du, N., Liu, W., & Li, J. (2023). Stock trend prediction using news and technical indicators with transformer architecture. Expert Systems with Applications, 213, 118839.

[12] Bernard, V. L., & Thomas, J. K. (1989). Post-earnings-announcement drift: delayed price response or risk premium? Journal of Accounting Research, 27, 1-36.

[13] Livnat, J., & Mendenhall, R. R. (2006). Comparing the post–earnings announcement drift for surprises calculated from analyst and time series forecasts. Journal of Accounting Research, 44(1), 177-205.

[14] Fisher, I. E., Garnsey, M. R., & Hughes, M. E. (2016). Natural language processing in accounting, auditing and finance: A synthesis of the literature with a roadmap for future research. Intelligent Systems in Accounting, Finance and Management, 23(3), 157-214.

[15] Zhao, Q., Chen, M., Zhang, G., & Li, Y. (2019). Automatic extraction of financial information from Chinese annual reports. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 1215-1218.

[16] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33, 9459-9474.

[17] Garneau, N., Godbout, M., Beauchemin, D., Durand, A., & Lamontagne, L. (2021). Financial document causality detection shared task: ProQuest team's models. In Proceedings of the Third Workshop on Financial Technology and Natural Language Processing, 61-65.

[18] Liu, H., Li, C., Wu, Q., & Lee, Y. J. (2023). Visual instruction tuning. arXiv preprint arXiv:2304.08485.