Jetsons at the FinNLP-2025 - Earnings2Insights: Persuasive Investment Report Generation Using Single And Multi-Agent Frameworks

SaiKrishna Rallabandi, Nikhil Kohli, Parag Pravin Dakle, Khyati Morparia,
Ojas Raundale, Preethi Raghavan
Fidelity Investments
{saikrishna.rallabandi, nikhil.kohli, paragpravin.dakle,
khyati.morparia, ojasjeetendra.raundale, preethi.raghavan}@fmr.com

Abstract

In this paper, we present four agent-based frameworks using the ReAct paradigm - two single-agent and two modular multi-agent systems - for automated report generation developed for the Earnings2Insights shared task at FinNLP-2025. Each single-agent solution is powered by a Writer agent, while the multi-agent frameworks incorporate Feedback agents to refine and enhance report quality through iterative collaboration. To evaluate generated reports, we introduce a comprehensive LLM-as-a-Judge framework that integrates six metrics to rank outputs across multiple dimensions. Our ensemble approach achieves an average financial accuracy of 0.571 (3rd place) and an average Likert score of 5.90 (2nd place) in human evaluations, with particularly strong performance in readability (1st place) and next-day prediction accuracy (2nd place).

1 Introduction

Earnings calls are among the most consequential communication events in global financial markets. During these sessions, company executives present quarterly performance, provide forward-looking guidance, and respond to probing questions from analysts. The transcripts of these calls are lengthy, noisy, and multi-speaker in nature, yet they contain critical signals that influence billions of dollars' worth of investment decisions. Professional analysts typically distill these signals into structured reports that combine factual accuracy with persuasive narrative. However, manually creating such reports is labor-intensive and does not scale to the thousands of earnings calls that occur each quarter. This motivates research into automated methods that can generate decision-oriented investment reports directly from transcripts (Takayanagi et al., 2025b).

Large Language Models (LLMs) have shown promise in financial NLP tasks such as summarization, sentiment analysis, and question answering.

Yet, applying LLMs to investment report generation requires moving beyond extractive summarization: systems must produce outputs that are both *analytically rigorous* and *persuasive enough* to support real financial decisions (Chen et al., 2024; Goldsack et al., 2025). The Earnings2Insights shared task (FinEval @ FinNLP/EMNLP 2025) is designed to evaluate precisely this capability. Given an earnings call transcript—with the option to incorporate aligned external information—participating systems must produce an analyst-style report that concludes with explicit Long/Short recommendations across three horizons (next day, week, and month). Crucially, the task emphasizes decision accuracy as the primary evaluation objective, shifting focus away from surface-level similarity metrics.

Challenges. Generating high-quality analyst reports presents several intertwined challenges. First, the language of earnings calls is often hedged, promotional, and strategically vague, requiring systems to "read between the lines." Second, factual reliability is paramount: even minor numeric inaccuracies or misattributions can undermine credibility. Third, persuasiveness matters. Reports must not only be accurate but also written in a style that instills confidence and convinces investors to take action. Finally, evaluation itself is non-trivial, as traditional n-gram overlap or embedding-based metrics fail to capture whether a report actually improves investment decisions meaningfully (Huang et al., 2025).

In this paper, we present several single and multi-agent modular frameworks that integrates (i) single-agent writer baseline that only uses the earnings call transcripts, (ii) data-enhanced generation with quarter-over-quarter and year-over-year fundamentals, and (iii) multi-agent writer and feedback ReAct (Yao et al., 2022) framework that couples a writer with automated feedback modules, includ-

ing a *Financial Expert* to cross-check metrics, a *Risk Analyst* to evaluate coverage of downside factors, and a *Persuasiveness Expert* to assess narrative strength. Iterative reasoning and action loops enable continuous refinement, ensuring that final reports are both factually accurate and rhetorically compelling.

Contributions. Our work makes the following contributions:

1. We present several single and multi-agent frameworks for automatic analyst report generation that include transcript-only and data-enhanced inputs.

2. We introduce a *feedback* or *validation* module that systematically reduces numeric and attributional errors by checking generated claims against structured fundamentals.

3. We develop a comprehensive evaluation harness that combines industry and shared task requirements in a set of six metrics. The evaluation harness is utilized to rank and select the best output across several systems.

4. We provide empirical evidence that the proposed methods generate reports that contain accurate investment recommendations and are factual, persuasive, and logical. The output of our N-way comparative selection obtains an average financial accuracy of 0.571 (ranking 3rd) and an average Likert Score of 5.90 (ranking 2nd) in human evaluations.

2 Dataset

The shared task dataset (Takayanagi et al., 2025a) comprises 64 earnings call transcripts, organized into two subsets. The first subset includes 40 calls paired with human-written summaries from ECT-Sum (Mukherjee et al., 2022), providing optional supervisory signals for training and benchmarking. The second subset comprises 24 transcripts paired with professional analyst reports authored by domain experts. These gold reports are withheld during training and used only for post-hoc evaluation by the organizers. Collectively, the dataset spans multiple industries and time periods, ensuring a diverse and representative benchmark for investment report generation. Participants are required to submit a single JSON file where each entry contains the earnings call code (ECC) and the generated report.

3 ReAct Prompting

Yao et al. (2022) introduced the *Reasoning and Acting* (*ReAct*) prompting paradigm, where reasoning and acting in large language models are used collaboratively by generating interleaved verbal reasoning traces and task-specific actions. This enables models to perform dynamic reasoning to create, maintain, and adjust action plans while interacting with external environments and addressing limitations of chain-of-thought reasoning, such as hallucination and error propagation. In this work, we leverage the Reasoning step to evaluate the generated report and the Acting step to update the report using the evaluation feedback from the Reasoning step.

4 Methodology

We propose a multi-agentic system that is designed to generate highly persuasive and analytically robust investment reports from earnings call transcripts, leveraging advanced prompt engineering, persona curation, and external financial data sources. The main agents that are used in the solution are as follows:

4.1 Writer Agents

We define a *Writer Agent* as an LLM prompt-based agent that portrays an expert data analyst. The agent accepts the transcript, any additional financial information, and evaluation feedback (if applicable) to generate the analyst report. We further define two different types of *Writer Agents*:

1. A *Simple Writer Agent* with a minimal set of instructions

2. An *Advanced Writer Agent* that uses an extensive set of instructions and additional financial information. (See Appendix A for complete prompts).

4.2 Feedback Agents

A *Feedback Agent* is an LLM prompt-based agent that portrays a financial expert with a specific skill set and reviews the generated report. We use four main types of Feedback Agents:

1. Skeptical Financial Expert Panel: This agent evaluates the generated report on five dimensions to generate a detailed review for the report - Factual accuracy, Completeness, Realism, Persuasiveness, and Transparency.

2. Financial Expert: This agent focuses on reviewing an analyst report for factual accuracy regarding financial metrics, numbers, and calculations.

3. Risk Analyst: This agent assesses whether the risk analysis presented in the generated report is complete and realistic.

4. Persuasiveness Expert: The main focus of this agent is to review the report for persuasiveness, clarity, and conviction dimensions. As part of the review, the agent also identifies specific sections of the report that require revision and outlines the aspects that need improvement.

4.3 Rewriter Agent

The *Rewriter Agent* is tasked with rewriting a given report using one or more reviews of the same. The agent must understand the given reviews and then rewrite the report to address essential aspects from the review.

4.4 External Financial Data Integration

To enhance the factual accuracy and analytical depth of each report, we integrate external financial data using the Alpha Vantage API. For each company and quarter, we retrieve three things: (i) Current quarter earnings data, (ii) Previous quarter earnings data, and (iii) Same quarter last year earnings data.

These data points include granular financial metrics such as total revenue, gross profit, operating income, EPS, and other relevant indicators. The retrieved data was given as input to the *Advanced Writer Agent* and all multi-agent flows, enabling the agent to perform explicit quarter-over-quarter and year-over-year comparisons. This helps the agent identify the direction of the company across various horizons for both short-term and long-term investments.

Sample External Data Structure:

```
{
  "ECC": "ABM_q3_2021",
  "Ticker": "ABM",
  "Quarter": "3",
  "Year": "2021",
  "most_recent_earnings": { ... },
  "previous_earnings": { ... },
  "same_qtr_last_year_earnings": { ... }
}
```

By combining transcript content and structured external data, our system generates reports that not only reflect the nuances of management discussions but are also grounded in quantitative performance trends.

4.5 Multi-Agent Feedback Framework

We model our multi-agent feedback framework using the ReAct prompting paradigm and propose two distinct variations for analyst report generation. The variations differ in the complexity of the feedback loop. For both frameworks, we iterate over the loop several times until one of the two conditions is satisfied - the maximum number of allowed iterations is complete, or the feedback agent(s) approve the generated report.

4.5.1 Simple Multi-Agent Feedback Framework

In this framework, the *Simple Writer Agent* is used to generate the analyst report, and the following ReAct paradigm is followed:

Reasoning : The *Skeptical Financial Expert Panel Feedback Agent* reviews the generated report and provides a detailed review on several dimensions with an accept/reject verdict.

Action : If the verdict in the reasoning step is not an accept then, the *Simple Writer Agent* updates the report based on the reasoning provided in the review.

To avoid infinite Reasoning and Action loops, we run the framework for a maximum of 3 iterations and use the output of the last iteration, irrespective of the final verdict.

4.5.2 Advanced Multi-Agent Feedback Framework

This framework is similar to the previous framework in its usage of the *Simple Writer Agent* to generate the analyst report. It differs in the implementation of the ReAct paradigm.

Reasoning : This framework uses *Financial Expert Feedback Agent*, *Risk Analyst Feedback Agent*, and *Persuasiveness Feedback Agent* to generate separate review of generated report.

Action : If the verdict in the reasoning step for any of the feedback agents is not an accept then, the *Rewriter Agent* updates the report based on the reasoning provided in the review.

Due to the presence of multiple reasoning feedback agents, we employ the Reasoning and Action loop only once for this framework.

We use GPT-4o (OpenAI et al., 2024) as the LLM for all agents. The *Advanced Writer Agent* uses a *temperature=0.7*, *max_tokens=4000*, and *top_p=0.95*. For all other agents, we use a *temperature=0.7*, *max_tokens=16000*, *top_p=1.0*, *frequency_penalty=0*, and *presence_penalty=0*.

4.6 Output Report Templates

4.6.1 Simple Writer Agent Output Format

```
<one sentence summary recommendation
with conviction level>
Financial Overview
...
Key Drivers
...
Risks
...
Opportunities
...
Final Recommendation
...
```

4.6.2 Advanced Writer Agent Output Format

```
Company Overview
...
Executive Summary
...
Financial Performance with comparative
analysis
...
Key Financial Metrics
...
Strategic Outlook and Investments
...
Insights from Q\&A Session
...
Projections and External Perspectives
...
Conclusion and Investor Takeaways
...
Final Recommendation: Long/Short for
next day, week, and month, with rationale
```

5 Evaluation

The generated reports are evaluated using a mixture of automated and human evaluation. For automated evaluation, three different metrics are used - Average Likert Score, Win Rate vs Analyst Report, and LLM-as-a-Judge.

1. Average Likert Score: Average score based on 1–7 Likert ratings of Persuasiveness, Logic, Usefulness, Readability, and Clarity.

2. Win Rate vs Analyst Report: Average score showing how often the generated report outperformed a professional analyst report in pairwise comparisons (ties excluded).

3. LLM-as-a-Judge: We define five metrics and an aggregate metric to evaluate a generated report using analyst reports requirements defined by SEC, FINRA[1], sector-specific requirements defined in the industry, and the Earnings2Insights shared task.

In the human evaluation, the average accuracy of financial decisions in the generated reports was computed by manually evaluating the reports.

5.1 LLM-as-a-Judge Metrics

We define a set of five metrics, with detailed guidelines on the weight of each metric and how to interpret low vs. high scores (See Appendix C for entire prompt).

Content Accuracy & Faithfulness: This metric assesses how accurately the report reflects the actual content of the earnings call. It checks explicitly if financial figures, metrics, statements, and direct quotes accurately reflect the content of the earnings call transcript.

Analytical Depth & Insight Quality: This metric evaluates the analysis presented in the report in terms of how it presents the underlying trends, connects financial performance to a broader market context, and provides meaningful insights that demonstrate deep business understanding rather than surface-level observations.

Investment Recommendation Quality: This metric evaluates the presence and soundness of required investment recommendations. Examine whether recommendations are well-supported by evidence, include appropriate risk assessments, and have realistic price goals with sound methodology.

Structure, Clarity & Presentation: This metric evaluates the structure of the report, specifically, organization, readability, and professional quality of the analysis. It measures logical flow, clear communication, proper formatting, and whether the content is easy to follow and understand for investment decision-making purposes.

---

[1]https://www.finra.org/rules-guidance/rulebooks/finra-rules/2241

Comprehensive Coverage & Completeness: This metric measures how thoroughly the analysis covers all material aspects of the earnings call.

Overall Grade: We compute the overall aggregate metric using the scores for each metric on a scale of 0-100 using the following formula: *Overall Grade = 0.25 * Content Accuracy + 0.20 * Analytical Depth + 0.20 * Investment Recommendation + 0.15 * Structure + 0.20 * Comprehensive Coverage*

We use the *Overall Grade* to grade the outputs of each system for each earnings call transcript and select the best one for the shared task submission. For the LLM-as-a-Judge Evaluator, we use Claude-Sonnet-4[2] with *temperature=0.5*, *max_tokens=4096*, and *top_p=0.95*

6 Results

Table 1 shows the results of the human evaluation of the analyst reports generated using an ensemble of all systems. We use the LLM-as-a-Judge Overall Grade to determine which system's output was selected in the ensemble.

Metric/ Time Horizon	Score	Leaderboard Position
Average Accuracy of Financial Decisions		
Average	0.571	3rd
Next Day	0.607	2nd
Next Week	0.555	7th
Next Month	0.552	2nd
Likert Score (Scale: 1-7)		
Average	5.90	2nd
Clarity	6.00	2nd
Logic	5.89	2nd
Persuasiveness	5.81	3rd
Readability	5.81	1st
Usefulness	6.01	2nd

Table 1: Human Evaluation Results of the ensemble of proposed methods

In addition to the human evaluation, reports generated using an ensemble of the proposed methods also achieve an Average Likert Score of 4.834 when computed automatically and an Average Win Rate vs Analyst Report Score of 0.762.

7 Conclusion

The work presented in this paper addresses the problem of automatically generating persuasive investment analyst reports from earnings call transcripts using a dataset shared as part of the Earnings2Insights shared task at FinNLP-2025. The proposed solution defines two classes of agents - *Writer* and *Feedback* and uses two variations of the *Writer Agent* as single agent solutions and a combination of the Simple Writer Agent with different Feedback Agents as multi-agent frameworks using the ReAct paradigm. Our key contributions include the systematic integration of structured financial fundamentals to reduce numeric errors, the deployment of specialized feedback agents that mirror real-world analyst review processes, and the development of a comprehensive evaluation framework that performs thorough evaluation, with our ensemble approach achieving competitive performance, including 2nd place in overall Likert scores and 1st place in readability across multiple evaluation dimensions.

References

Chung-chi Chen, Jian-tao Huang, Hen-hsen Huang, Hiroya Takamura, and Hsin-hsi Chen. 2024. SemEval-2024 task 7: Numeral-aware language understanding and generation. In *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*, pages 1482–1491, Mexico City, Mexico. Association for Computational Linguistics.

Tomas Goldsack, Yang Wang, Chenghua Lin, and Chung-Chi Chen. 2025. From facts to insights: A study on the generation and evaluation of analytical reports for deciphering earnings calls. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 10576–10593, Abu Dhabi, UAE. Association for Computational Linguistics.

Yu-Shiang Huang, Chuan-Ju Wang, and Chung-Chi Chen. 2025. Decision-oriented text evaluation. *Preprint*, arXiv:2507.01923.

Rajdeep Mukherjee, Abhinav Bohra, Akash Banerjee, Soumya Sharma, Manjunath Hegde, Afreen Shaikh, Shivani Shrivastava, Koustuv Dasgupta, Niloy Ganguly, Saptarshi Ghosh, and Pawan Goyal. 2022. ECT-Sum: A new benchmark dataset for bullet point summarization of long earnings call transcripts. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 10893–10906, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

OpenAI, :, Aaron Hurst, Adam Lerer, Adam P. Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, Aleksander Mądry, Alex Baker-Whitcomb, Alex Beutel, Alex Borzunov, Alex Carney, Alex Chow, Alex Kirillov, and 401 others. 2024. Gpt-4o system card. *Preprint*, arXiv:2410.21276.

---

[2]https://www.anthropic.com/news/claude-4

Takehiro Takayanagi, Tomas Goldsack, Kiyoshi Izumi, Chenghua Lin, Hiroya Takamura, and Chung-Chi Chen. 2025a. Earnings2Insights: Analyst Report Generation for Investment Guidance. In *Proceedings of the FinNLP Workshop at EMNLP 2025*, Suzhou, China. Overview paper for the Earnings2Insights shared task (FinEval) at FinNLP 2025.

Takehiro Takayanagi, Hiroya Takamura, Kiyoshi Izumi, and Chung-Chi Chen. 2025b. Can GPT-4 sway experts' investment decisions? In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 374–383, Albuquerque, New Mexico. Association for Computational Linguistics.

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

A Prompts for different agents used

A.1 Advanced Writer Agent Details

At the core of this approach is a comprehensive analyst prompt, explicitly crafted to guide the language model in producing investment reports that meet the standards of professional financial analysis. The prompt instructs the model to act as an expert financial analyst, synthesizing insights from the earnings call transcript and multiple quarters of earnings data. Key aspects of the prompt design include:

- **Objective:** The model is tasked with creating an investment report using the earnings call transcript, current quarter earnings data, previous quarter earnings data, and the same quarter's data from the prior year.

- **Writing Style and Tone:** The report must maintain a professional and analytical tone, utilize precise language for financial concepts, and base insights on data-driven metrics and comparisons.

- **Deep Analysis:** The prompt directs the model to interpret not only explicit statements but also underlying sentiment, tone, and hesitation in the transcript, mirroring the nuanced work of human analysts.

- **Actionable Guidance:** The generated report must provide clear investment advice and a final recommendation, including Long/Short positions for three time frames: next day, next week, and next month.

- **Structured Response:** The output follows a detailed template (see below), ensuring consistency and coverage of all relevant analytical dimensions.

B Example of External Financial Data Augmentation

To illustrate our integration of external financial data, we provide an example of the structured earnings data retrieved for a single company-quarter using the Alpha Vantage API below. This data was programmatically injected into the writer prompt to enable explicit quarter-over-quarter and year-over-year comparisons in generated investment reports.

The following JSON excerpt corresponds to ECC `ABM_q3_2021` (ABM Industries, Q3 2021):

This structured financial data enables our system to ground investment analysis in quantitative performance trends and to generate more precise, actionable insights for investors.

C LLM-as-a-Judge Evaluation

The evaluation prompt for LLM-as-a-Judge is as follows:

Simple Writer Agent Prompt

```
You are a world-class financial analyst specializing in generating persuasive investment analysis
    reports from company earnings call transcripts.

Your task:
- Read the provided earnings call transcript (and any external financial data, if available).
- Write a clear, concise, and highly persuasive investment analysis report.
- Your report should be suitable for institutional investors and help them decide whether to take a
    Long or Short position in the company.
- Support your recommendation with specific evidence from the transcript and data.
- Highlight key financial metrics, management commentary, risks, and opportunities.
- Use authoritative, confident language and include a clear, actionable investment recommendation
    (Long/Short), with rationale.
- If data is missing or uncertain, acknowledge it and explain how it affects your analysis.
- Be objective, but persuasive-your goal is to maximize the accuracy of human investment decisions
    based on your report.

Format:
- Start with a one-sentence summary recommendation (Long/Short and conviction level).
- Follow with a structured analysis: Financial Overview, Key Drivers, Risks, Opportunities, and
    Final Recommendation.

Example output:
Recommendation: Long with high conviction.

Financial Overview:
[...]
Key Drivers:
[...]
Risks:
[...]
Opportunities:
[...]
Final Recommendation:
[...]
```

Objective: You are an expert financial analyst tasked with creating an investment report for a given Company's earnings call. You have access to the earnings call transcript, earnings data for the given quarter, previous quarter earnings data, and earnings data from the same quarter last year. Use this information to analyze the company's performance and provide investment advice and a final recommendation to the investors, according to the reference report template given below.

Instructions for Writing Style and Tone:
- Maintain a professional, analytical tone throughout the report.
- Use precise and clear language to convey complex financial concepts.
- Ensure your analysis is data-driven and supported by relevant metrics and comparisons.
- Provide actionable insights that reflect deep understanding of market dynamics and company strategy.
- Focus on delivering value to investors by identifying opportunities be it Long or Short the company stock.
- Be realistic and strategic in your evaluation of the company's growth, earnings and stock movement/direction forecast.
- Earnings calls are usually driven by company representatives who try to paint a good picture of the company overall. Your job is to look in between their lines, tone, hesitation and utilize your investment expertise to make decisions about the forecast of the company's stock performance.

Report Template Response:

Investment Report for (Company) (Symbol) Quarter Year

1. **Company Overview**:
   - Briefly introduce [Company Name], including its industry, core business areas, and any recent strategic initiatives or acquisitions. Highlight its market position and any significant developments relevant to the earnings call.
2. **Executive Summary**:
   - Summarize the key points from the earnings call, including financial performance, strategic achievements, and challenges. Focus on insights that affect investor decisions and the company's future outlook.
3. **Financial Performance**:
   - **Revenue and Earnings**:
     - Report the current quarter's earnings, including total earnings and EPS. Compare these figures with the previous quarter and the same quarter last year to identify trends in stability, revenue, profit and growth.
   - **Comparative Analysis**:
     - Analyze year-over-year and quarter-over-quarter changes in earnings, considering broader market trends or industry-specific factors that may have influenced performance.
4. **Key Financial Metrics**:
   - Evaluate the company's return on average assets, average deposits and loans, and trends in net charge-offs and provisions for loan losses, discussing implications for future financial performance.
5. **Strategic Outlook and Investments**:
   - **Long-term Strategic Expansion**:
     - Discuss strategic investments in market expansion, efficiency improvements, and expected impacts on profitability and market leadership.
   - **Interest Rate Environment**:
     - Examine pressures on the net interest margin, external factors affecting it, and the company's strategic response.
6. **Insights from Q&A Session**:
   - Reflect on management's optimism or caution regarding future growth, risk management strategies, and handling of nonperforming assets.
7. **Projections and External Perspectives**:
   - Summarize management's EPS guidance and include insights from financial analyst commentaries for a broader perspective.
8. **Conclusion and Investor Takeaways**:
   - Assess performance consistency, growth focus, adaptation to economic conditions, and risk management, reassuring investors of strategic consistency.
9. **Final Recommendation**:
   - Based on your analysis, provide a recommendation to investors, considering their goal to profit from your investment advice, for each of the following time frames.
   - **Next Day**: {"Recommendation": Either Long/Short, "Reason": Rationale for the recommendation}

```
        - **Next Week**: {"Recommendation": Either Long/Short, "Reason": Rationale for the
            recommendation}
        - **Next Month**: {"Recommendation": Either Long/Short, "Reason": Rationale for the
            recommendation}


**Important**:
- Analyze the earnings call + financial data provided and Use the above response template to
    generate a comprehensive investment report.
- Your analysis should focus on helping investors make informed decisions that maximize their
    returns based on your expert insights.
- Only generate the report as the final response in format given, do not generate anything
    additional.

**Financial Data**
{financial_data}

**Earnings Call Transcript**
{earnings_call}
```

Table 2: Skeptical Financial Expert Panel Feedback Agent Prompt

```
You are a panel of highly skeptical, expert financial reviewers. Your job is to rigorously
    scrutinize the following investment analysis report with the utmost criticality. Do not
    hesitate to point out even minor flaws or omissions.

Your review must address:
- Factual accuracy: Are all financial metrics, numbers, and calculations correct and supported by
    the transcript?
- Completeness: Are all key aspects of the company's performance, risks, and opportunities
    thoroughly analyzed?
- Realism: Are the risk assessments and recommendations realistic and grounded in the provided
    evidence?
- Persuasiveness: Is the argumentation strong, clear, and convincing for an institutional investor?
- Transparency: Are all claims and recommendations directly traceable to specific evidence in the
    transcript or external data?

Be extremely strict in your assessment. If there is any ambiguity, missing evidence, weak argument,
    or unsupported claim, point it out in detail. Do not accept vague or generic statements.

If-and only if-the report is flawless and cannot be improved in any way, reply ONLY with: ALL GOOD.

Otherwise, list every specific issue, gap, or suggestion for improvement, referencing the relevant
    part of the report and transcript.

Report:
{report}
Original transcript:
{transcript}
```

Table 3: Financial Expert Feedback Agent Prompt

```
You are a financial expert. Carefully review the following investment analysis report for factual
    accuracy regarding financial metrics, numbers, and calculations. If you find any inaccuracies,
    list them and suggest corrections. If everything is accurate, reply 'All financials accurate.'

Report:
{report}
```

Table 4: Risk Analyst Feedback Agent Prompt

```
You are a risk analyst. Review the following investment analysis report and assess whether the risk
    analysis is complete and realistic. List any missing or understated risks, or reply 'Risk
    analysis is complete.'

Report:
{report}
```

Table 5: Persuasiveness Expert Feedback Agent Prompt

```
You are an expert in persuasive writing for financial audiences. Evaluate the following investment
    analysis report for persuasiveness, clarity, and conviction. Suggest specific improvements to
    make the report more convincing, or reply 'Persuasiveness is strong.'

Report:
{report}
```

Table 6: Rewriter Expert Feedback Agent Prompt

```
You are an expert financial analyst and editor. Given the following investment analysis report and
    the feedback below, rewrite the report to address all feedback and improve its quality.

Original Report:
{report}

Feedback:
{feedback}

Improved Report:
```

```
{
  "ECC": "ABM_q3_2021",
  "Ticker": "ABM",
  "Quarter": "3",
  "Year": "2021",
  "most_recent_earnings": {
    "fiscalDateEnding": "2021-07-31",
    "reportedCurrency": "USD",
    "grossProfit": "255000000",
    "totalRevenue": "1543100000",
    "costOfRevenue": "1288100000",
    "costofGoodsAndServicesSold": "1288100000",
    "operatingIncome": "-9400000",
    "sellingGeneralAndAdministrative": "253800000",
    "researchAndDevelopment": "None",
    "operatingExpenses": "1552500000",
    "investmentIncomeNet": "None",
    "netInterestIncome": "-6300000",
    "interestIncome": "5800000",
    "interestExpense": "6300000",
    "nonInterestIncome": "None",
    "otherNonOperatingIncome": "500000",
    "depreciation": "None",
    "depreciationAndAmortization": "10600000",
    "incomeBeforeTax": "-15200000",
    "incomeTaxExpense": "-1500000",
    "interestAndDebtExpense": "None",
    "netIncomeFromContinuingOperations": "-13700000",
    "comprehensiveIncomeNetOfTax": "None",
    "ebit": "-8900000",
    "ebitda": "1700000",
    "netIncome": "-13700000"
  },
  "previous_earnings": {
    "fiscalDateEnding": "2021-04-30",
    "reportedCurrency": "USD",
    "grossProfit": "222900000",
    "totalRevenue": "1497400000",
    "costOfRevenue": "1274500000",
    "costofGoodsAndServicesSold": "1274500000",
    "operatingIncome": "50300000",
    "sellingGeneralAndAdministrative": "161900000",
    "researchAndDevelopment": "None",
    "operatingExpenses": "1447100000",
    "investmentIncomeNet": "None",
    "netInterestIncome": "-7800000",
    "interestIncome": "7600000",
    "interestExpense": "7800000",
    "nonInterestIncome": "None",
    "otherNonOperatingIncome": "200000",
    "depreciation": "None",
    "depreciationAndAmortization": "22000000",
    "incomeBeforeTax": "42800000",
    "incomeTaxExpense": "11700000",
    "interestAndDebtExpense": "None",
    "netIncomeFromContinuingOperations": "31100000",
    "comprehensiveIncomeNetOfTax": "None",
    "ebit": "50600000",
    "ebitda": "72600000",
    "netIncome": "31100000"
  },
  "same_quarter_last_year_earnings": {
    "fiscalDateEnding": "2020-07-31",
    "reportedCurrency": "USD",
    "grossProfit": "235200000",
    "totalRevenue": "1394100000",
    "costOfRevenue": "1158900000",
```

```
    "costofGoodsAndServicesSold": "1158900000",
    "operatingIncome": "93600000",
    "sellingGeneralAndAdministrative": "113700000",
    "researchAndDevelopment": "None",
    "operatingExpenses": "1300400000",
    "investmentIncomeNet": "None",
    "netInterestIncome": "-13800000",
    "interestIncome": "13600000",
    "interestExpense": "13800000",
    "nonInterestIncome": "None",
    "otherNonOperatingIncome": "200000",
    "depreciation": "None",
    "depreciationAndAmortization": "11800000",
    "incomeBeforeTax": "80000000",
    "incomeTaxExpense": "24000000",
    "interestAndDebtExpense": "None",
    "netIncomeFromContinuingOperations": "56000000",
    "comprehensiveIncomeNetOfTax": "None",
    "ebit": "93800000",
    "ebitda": "105600000",
    "netIncome": "56000000"
  }
}
```

```
# Financial Analyst Recommendation Report Evaluation Guidelines

## Evaluation Steps
1. Parse the earnings call transcript to create a reference baseline of disclosed information
2. Systematically check the analyst report against each metric category
3. Assign sub-scores for each component
4. Calculate weighted overall score
5. Generate specific feedback on strengths and weaknesses
6. Provide actionable improvement suggestions


# Analyst Report Evaluation Rubric

## Overview
This rubric evaluates analyst reports derived from earnings call transcripts across five key
    dimensions. Each metric is scored on a 5-point scale (1-5), with specific criteria for each
    score level.

---

## **Metric 1: Content Accuracy & Faithfulness to Source Material**
*Weight: 25%*

**5 - Excellent (90-100%)**
- All financial figures, metrics, and statements accurately reflect the earnings call transcript
- No factual errors or misrepresentations
- Proper context maintained for all cited information
- Direct quotes are verbatim and appropriately attributed

**4 - Good (80-89%)**
- Minor discrepancies in non-material details
- 1-2 small factual errors that don't affect core analysis
- Generally accurate representation of transcript content

**3 - Satisfactory (70-79%)**
- Some factual errors present but core information is correct
- Occasional misinterpretation of context
- Most financial data accurately represented

**2 - Needs Improvement (60-69%)**
- Multiple factual errors affecting analysis quality
- Significant misrepresentation of key statements
- Some financial data inaccuracies

**1 - Poor (Below 60%)**
- Frequent factual errors throughout
- Major misrepresentation of earnings call content
- Unreliable financial data presentation


---

## **Metric 2: Analytical Depth & Insight Quality**
*Weight: 20%*

**5 - Excellent**
- Demonstrates deep understanding of business fundamentals
- Identifies key trends, risks, and opportunities not explicitly stated
- Provides meaningful interpretation of financial metrics
- Connects current performance to broader industry/market context
- Offers unique insights beyond surface-level observations

**4 - Good**
- Shows solid analytical thinking with some original insights
- Good interpretation of financial performance
- Identifies most key business drivers and risks
- Some connection to broader market context
```

```
**3 - Satisfactory**
- Basic analysis present but limited depth
- Identifies obvious trends and issues
- Some interpretation of financial metrics
- Limited broader context or unique insights

**2 - Needs Improvement**
- Shallow analysis with minimal interpretation
- Misses key business drivers or risks
- Limited financial analysis beyond basic metrics
- Little to no broader context

**1 - Poor**
- Lacks analytical depth
- No meaningful insights or interpretation
- Fails to identify key business issues
- No connection to broader context

---

## **Metric 3: Investment Recommendation Quality**
*Weight: 20%*

**5 - Excellent**
- Clear, well-supported investment thesis
- Recommendation directly tied to analysis and evidence
- Appropriate risk assessment and mitigation strategies
- Realistic price targets with sound methodology
- Clear timeline and catalysts identified

**4 - Good**
- Generally sound investment recommendation
- Good supporting rationale
- Adequate risk assessment
- Reasonable price targets
- Some catalysts identified

**3 - Satisfactory**
- Basic investment recommendation present
- Some supporting rationale provided
- Limited risk assessment
- Price targets may lack detailed justification

**2 - Needs Improvement**
- Weak investment recommendation
- Poor supporting rationale
- Inadequate risk assessment
- Unrealistic or poorly justified price targets

**1 - Poor**
- Unclear or unsupported investment recommendation
- No clear rationale
- Missing risk assessment
- No price targets or unrealistic expectations

---

## **Metric 4: Structure, Clarity & Professional Presentation**
*Weight: 15%*

**5 - Excellent**
- Logical flow and clear organization
- Executive summary effectively captures key points
- Professional tone and language throughout
- Proper formatting and visual elements
- Easy to follow and understand
```

```
**4 - Good**
- Generally well-organized with clear structure
- Most sections flow logically
- Professional presentation with minor issues
- Generally easy to follow

**3 - Satisfactory**
- Basic organization present
- Some sections may lack clarity
- Acceptable professional standards
- Mostly understandable

**2 - Needs Improvement**
- Poor organization and structure
- Difficult to follow logical flow
- Unprofessional presentation elements
- Clarity issues throughout

**1 - Poor**
- No clear structure or organization
- Very difficult to follow
- Unprofessional presentation
- Major clarity and readability issues

---

## **Metric 5: Comprehensive Coverage & Completeness**
*Weight: 20%*

**5 - Excellent**
- Covers all material topics from earnings call
- Addresses key analyst questions and management responses
- Includes relevant forward-looking statements
- Comprehensive risk factor analysis
- Addresses both quantitative and qualitative aspects

**4 - Good**
- Covers most important topics from call
- Addresses majority of key Q&A points
- Good coverage of forward-looking elements
- Adequate risk analysis

**3 - Satisfactory**
- Covers basic topics from earnings call
- Some key Q&A points addressed
- Limited forward-looking analysis
- Basic risk coverage

**2 - Needs Improvement**
- Misses several important topics
- Limited coverage of Q&A insights
- Minimal forward-looking analysis
- Inadequate risk assessment

**1 - Poor**
- Significant gaps in coverage
- Fails to address key earnings call topics
- No forward-looking analysis
- Missing risk assessment

---
```

```
## **Overall Scoring & Ranking System**

### **Composite Score Calculation:**
- Content Accuracy: Score x 25%
- Analytical Depth: Score x 20%
- Investment Recommendation: Score x 20%
- Structure & Clarity: Score x 15%
- Comprehensive Coverage: Score x 20%

### **Overall Rating Bands:**
- **4.5-5.0**: Exceptional Report
- **3.5-4.4**: Strong Report
- **2.5-3.4**: Adequate Report
- **1.5-2.4**: Needs Significant Improvement
- **1.0-1.4**: Poor Report

### **Quality Assurance Flags:**
- **Critical**: Content Accuracy score below 2.0
- **Warning**: Any individual metric score below 2.0
- **Review**: Significant variance between metric scores (>2 points)

### Output Format:
```json
{
    "overall_rating": [Numerical rating],
    "detailed_scores": {
        "content_accuracy": [Numerical Score on a scale of 1 to 25],
        "analytical_depth": [Numerical Score on a scale of 1 to 20],
        "investment_recommendation": [Numerical Score on a scale of 1 to 20],
        "structure_and_clarity": [Numerical Score on a scale of 1 to 15],
        "comprehensive_coverage": [Numerical Score on a scale of 1 to 20]
    }
}
```
```