

Multiagent Debate for Predicting Stock Trends from Earnings Calls

Sergey Barabanoff

My Team

barab013@umn.edu

1 Background and Motivation

Predicting stock market motions is a heavily studied topic because people are willing to invest a lot of money into something that has the potential of making them money. Natural language processing (NLP) is an important domain in making these predictions due to the wealth of information that is delivered with language. NLP techniques have been applied to all kinds of linguistic information to find correlations between the text and the stock market. However, a relatively unstudied domain is transcripts of conference calls from publicly traded companies.

These types of calls include earnings calls and update calls, among others; they are typically held between executive management of the company, analysts, and other corporate positions. Earnings calls, specifically, are a type of conference call where company management present key financial figures, explain performance, and provide future outlooks. Since members of executive management play important roles in the company, they are very familiar with the inner workings of the company; they likely have some sort of insight about the trajectory of the company's performance and value. For example, if they know that the company will have to lay off a significant number of employees, then this might suggest that the company is struggling; then, when investors find out about these lay offs, they might sell shares of the company because they expect its value to decrease.

Since earnings call participants have this insight, earnings call transcripts could possibly reveal their expectations about company performance. This could be via indicators like emotion, rhetoric, or other aspects of language. For example, if the CEO rambles during the call, that might suggest that they are stressed out about the company's performance; if they speak confidently, they might be more hopeful. These are qualities of language that Large

Language Models (LLMs) could potentially derive from transcripts, based on phrases, sentiments, and other such things.

LLMs are very powerful language models because they excel at modeling language. However, a single agent may struggle to reason through complex problems for several reasons.

Firstly, a single agent likely will not correct itself. If it says something incorrect, it will continue generating text confidently because that results in more fluent text. This is not ideal because a false premise might result in a false conclusion. Furthermore, this will prevent the agent from performing critical evaluation because it would not want to disagree with itself.[3]

Secondly, a single agent might choose to focus on only a few aspects of given text, rather than exploring different avenues. This would cause it to miss potentially useful features.[5] For example, an agent trying to analyze a conference call transcript might choose to analyze only the numbers or logic in the transcript, ignoring any emotion that it might contain. This could potentially be avoided by prompting the agent to explore logic and emotion, but then it would not go into as much depth in its exploration.

Thirdly, a single agent system will not generate as much reason for its evaluation. It will generate its final conclusion, and maybe a few reasons, but it won't be able to generate an entire reasoning process, because that is not fluent text.[4]

A multi-agent debate (MAD) system could overcome these issues, as seen by Liang et al. This system would consist of agents prompted to focus on different aspects of the input text, such as emotion and logic. These agents would then be prompted to debate each other's reasoning, provide counterarguments, and look for anything that is factually incorrect. One more agent, a judge, would be able to determine a winner based on the quality and logic of the agent's debate, ultimately determining

the trajectory of a company's performance based on the input text.[3] This pipeline would attempt to prevent the agents from building off any incorrect points. It would also provide clear, articulated reasoning, since the agent's debate would be entirely generated text.

As mentioned, agents acting alone suffer from degeneration of thought, narrow avenues of exploration, and they often do not self-correct, but rather build off of their flawed logic. Multiple agents can overcome these issues because they will each act separately. They can be prompted to explore different dimensions of the input text, and then during the debate, they will assess other agents' arguments to correct their flawed logic and work together to come up with a stronger prediction, rather than degenerate thought. The most powerful implication of a multiagent system is that they are each able to focus on specific features of interest, like emotion or logic, in order to uncover important aspects of the input text. Single agents and other conventional NLP techniques struggle with this because it requires deep analysis of multiple different and complicated linguistic dimensions.

Predicting stock value over the next three months is a natural horizon for this task because earnings calls themselves occur on a quarterly cycle. Management's comments, tone, and explanations in an earnings call are almost always forward-looking over the next quarter: they discuss upcoming demand, costs, product launches, risks, and guidance for the following quarter. As a result, the linguistic signals in the transcript are most likely to influence investor expectations and price movements within that same three-month window. A shorter time frame (such as one day or one week) is dominated by market noise, and a longer one dilutes the signal with unrelated external events. Therefore, the three-month horizon provides a meaningful, justified, and interpretable target for evaluating whether language in the earnings call predicts near-term company performance.

Therefore, a MAD system that examines earnings call presentation and Q&A transcripts could potentially uncover more complicated aspects of language, such as logical inconsistencies, emotion, and many other features. These features can be brought into a debate to allow multiple agents to perform high-quality reasoning about a company's trajectory. Since these earnings calls happen quarterly (every three months), agents can use these transcripts to analyze back-to-back earnings calls,

uncovering valuable patterns in company performance. Quantitatively, this can be measured by having the agents predict whether the value of a company's stock will increase, decrease, or remain stable over the next quarter; the agents can also output their confidence, which could potentially correlate to the change in stock value. This is very valuable to investors, who could use features discovered by the agent's to make their own investment decisions.

2 Literature Review

Earnings call transcripts have been previously investigated using NLP techniques by Chin and Fan. They explored whether these transcripts might contain features indicative about future stock returns.[2] Some of these features were number of words in the transcript, the number of questions and answers, adverb use by the CEO, readability, and sentiment. The latter two features were found to contribute to predictions, although only 10-30% of readability and sentiment features were actually helpful for predictions. Perhaps, an LLM-based system could improve on these results by identifying more complex features. This sort of framework could potentially provide stronger reasoning.

To improve their own reasoning, single agents can be prompted to self-reflect. This technique is an attempt at having the agent critique its own argument. Unfortunately, single agents undergo "degeneration of thought." This means that if the agent comes up with a confident argument, it tends to agree with itself, failing to point out any logical flaws it might have made. This is especially problematic if it uses incorrect evidence which it refuses to correct. To resolve this problem, Liang et al. have introduced a multi-agent debate framework. This involves prompting two agents to make their arguments, while a judge oversees the debate and determines a final winner.[3] Evidently, MAD encourages agents to use good logic, correct statements, and overall deep contemplation to win the debate. If they do not, the opposing agent will notice and leverage that in their argument.

MAD systems dedicated to financial planning and decision making have been introduced previously. Xiao et al. introduced TradingAgents, a pipeline mimicking a trading firm in order to decide stock trades. Their system took information from several places like social media and news sources, and then gave the data to a so-called "researcher

team" to debate about whether stocks should be bought or sold. Then, a trader decides what action to take based off of the debate, a risk management team adjusts the decision, and a manager approves the decision.[4] This pipeline was able to achieve over 26% returns. However, it is important to note that this paper did not compare results with and without debate. It is likely that the debate was an important feature because it is the first step in the pipeline, but it is unclear how the pipeline would have performed without the debate.

Another group of researchers, Zhao et al., tackled equity stock selection via three specialized agents: one dedicated to fundamentals, one dedicated to sentimentality, and one dedicated to valuation. These agents were encouraged to debate, but a mediation agent helped them arrive at a consensus about stock selection.[5] The goal of these three agents debating was to reduce hallucinations, since debating agents are incentivized to argue factually. The agents examined 10-K and 10-Q reports, then engaged in a circular debate where each agent had to speak at least twice. The researchers found that this MAD system was able to outperform single agents when they were directed to be neutral about risk. However, if they were directed to be risk averse, then the results were lower than for a single agent.

A MAD framework orchestrated around conference call transcripts has been explored previously by Cai et al. This team of researchers built a pipeline where five specialized agents (earnings, market, sentiment, valuation, and risk) were used to analyze transcripts to derive these five features. Then, 3 agents (A skeptic, a leader, and a trusting agent) debated about the derived features to generate a report about earnings, sentiment, and market overview.[1] However, this team of researchers did not explore the connection between the language of the transcripts and stock trajectories. More information might be present in the logical or emotional cues of the transcripts.

Although prior work has explored earnings-call features, multi-agent debate frameworks, and financial decision-making systems, several important gaps remain unaddressed. In prior work exploring earnings call transcripts, Chin and Fan limited the study to hand-crafted features like readability and sentiment. To my knowledge, there has not been work using LLM reasoning to extract the deeper narrative consisting of logical and emotional cues. Furthermore, existing financial MAD frameworks

are often aimed at trading decisions, rather than isolating a single textual source to determine its explanatory power, like that of earnings calls. These textual sources are often also edited documents, not spontaneous and conversational like earnings calls are. Thus, they lack the dimensions that contain insiders' thoughts and opinions, which are likely strongly correlated with the future of a company because insiders are intimately involved with a company's plans. Cai et al. have explored the use of multiple agents to summarize and interpret conference calls, but they did not explore whether the calls contained features with predictive power. However, current approaches cannot deeply analyze earnings-call language and reliably use it to predict future stock movement.

No prior work has directly addressed whether multiagent reasoning can elicit deep contextual features from earnings calls transcripts that are predictive about a company's stock trajectory over the next quarter. This motivates the inclusion of a single agent baseline with two purposes: to measure whether MAD actually improves reasoning, and to reveal some of the limitations of single agent reasoning. By directly comparing these approaches to predicting stock trajectory, this study isolates the contributions of emotion, financial fundamentals, and logic in predicting quarterly stock movements from earnings calls and whether multiagent reasoning truly produces superior insight than single agent reasoning. Not only would investors benefit from having a better understanding about how helpful earnings calls can be for making investment decisions, but this study also helps demonstrate the power of MAD frameworks over single agent reasoning. Multiple agents could potentially explore multiple complex text attributes, like emotion, in more depth than a single agent is able to.

3 Approach

Earnings call transcripts from both presentations and Q&As were obtained from the Capital IQ Transcripts dataset, found on Wharton Research Data Services (WRDS). Daily stock values for each company in the dataset were obtained from the Center for Research in Security Prices (CRSP). The dataset contained transcripts starting from 2005 until 2024, but since experiments took a long time to run and inputs had a very large number of tokens, a small subset of these transcripts were chosen. The used dataset consisted of quarterly presentation and

Q&A transcripts from 10 companies (see Figure 1) from 2021, resulting in a sample size of 40. 2021 was chosen as the year because it was after the pandemic, so hopefully more stable than the year prior. It is also more recent, but less affected by AI than years after, which could again be potentially unstable because of the rapid innovations caused by AI. In total, 22 transcripts corresponded to an increase of stock price over the following quarter, 10 corresponded to a decrease, and 8 corresponded to approximately no change in stock price.

5 total agents were defined, using GPT-4o from the OpenAI API. This API has a token-per-minute limit of 30,000, introducing another computational consideration. Most transcripts were around 40,000 tokens long, so each transcript was cut in half and inputs to the models were only the last 20,000 tokens of each transcript. This was a severe limitation, because the agents were missing a lot of context.

The first of the three main agents was a fundamentalist agent, who was prompted to focus on financial fundamentals like revenue, risks, product demand, and expected valuation. The second agent was an emotion analyst, who was prompted to focus on emotional and psychological cues like tone, confidence, enthusiasm, and how these signals shape investor sentiment. The third agent was a logician, who was told to focus on logic and strategy, such as consistency, contradictions, and realism of claims. The debate functioned by providing the agents with their prompt, the transcript (latter half of either Q&A or presentation), and the debate up until their turn. They were instructed to reason through the transcript and the debate up until that point, while focusing on their specified dimensions, and then to output their prediction as "Increase", "Decrease", or "No Change", as well as their confidence in their prediction as a percent.

The debate would go for two rounds such that each agent spoke twice. This would give each agent an opportunity to hear other agents' perspectives and change their minds. If, after the second round, all of the agents agree on a prediction, then that would be the final prediction and the final confidence would be an average of all three confidences from the second round. If they disagree, then a judge persona was prompted to read the earnings call and debate transcripts, weigh the arguments and evidence, and then decide on a final prediction and confidence score.

The baseline single agent scheme consisted of

another GPT-4o agent being provided the same transcript, and then being prompted to read the transcript in order to make a prediction with a confidence score. The prompt instructed the single agent to consider emotion, logic and strategy, and financial fundamentals, so that it had some direction as to which topics to consider in its reasoning. It was also instructed to generate its reasoning before making its prediction, so that the prediction could build off of the reasoning. All prompts can be found in Appendix B. For both MAD and the single agent, if a confidence value was not found in their output, it would be replaced with 0 (representing uncertainty).

Once predictions were made for all samples, the predictions, confidence scores, and reasonings were all saved, as well as individual predictions and confidence scores from the three agents. Labels were obtained from the CRSP dataset by comparing the company stock price from day before the earnings calls and the day three months afterwards. These were mapped to labels based on whether the three-month price was higher or lower than the day before; however, if the future price was within 0.5% of the original price, then that was considered as 'No Change'.

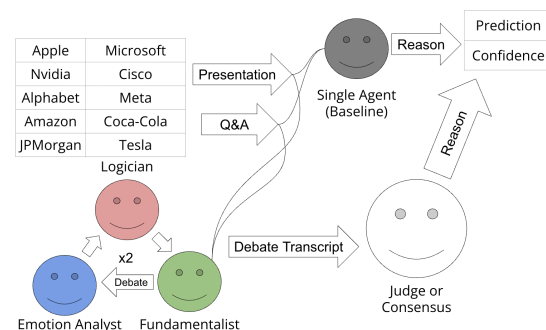


Figure 1: Overview of the multi-agent debate pipeline used in this project.

This scheme allowed both the MAD system and the single agent to reason about transcripts while analyzing the same dimensions. However, since the MAD system had more workers to debate with each other, I strongly suspected that MAD would outperform the single agent baseline. By prompting the agents to inspect each of these dimensions, they were hopefully able to isolate some complex and unique features from the transcripts that could correlate to the presenter's knowledge about company trajectory.

Thus, I hypothesized that this MAD system

would outperform a single agent in predicting a company's stock movement over the next quarter because debate has been shown to improve reasoning quality, and because specialization of agents allows the agents to explore complex dimensions in more depth, extracting more meaningful features from language. Also, I suspected that the confidence scores would correlate to the actual returns made from the stock, because a more certain agent could correspond to more certain investors, who would bring that stock up higher by investing in it (or lower, on the other end).

I anticipated that I would encounter computational challenges, but they were more severe than anticipated. The greatest detriment came from the fact that I had to use the last 20,000 tokens from the transcripts because of token-per-minute limits, ignoring around half of the transcripts. This meant that a lot of information was being left out. Furthermore, since I was using the OpenAI API, I had to pay over \$21 in total because of how many input tokens I used. I was also limited by time, because the experiments were very slow, taking multiple hours to make all of the predictions.

This project is novel in several ways. It is the first, to my knowledge, to apply a multi-agent debate framework directly to make predictions from earnings call presentations and Q&A sessions to isolate spontaneous and emotionally rich language not present in formal filings or news data. Unlike prior systems that aggregate multiple data sources or focus on edited documents, this approach analyzes the raw linguistic signals that management communicates to investors. This method also decomposes the reasoning task into three interpretable dimensions: fundamentals, emotion, and logic. These were chosen specifically for their relevance to earnings calls. Importantly, this work directly compares the multiagent system to a single agent baseline on the same prediction task, isolating the contribution of debate itself.

4 Results

Several evaluation metrics were calculated. These consisted of accuracy of predictions, AUC (using "Increase" as the positive class and both "Decrease" and "No Change" as the negative class), mean confidence score, the correlation between confidence scores and returns, and a Sharpe ratio, which measures how well returns correspond to stock volatility. Additionally, I also determined

how often debates were settled via consensus or by the judge, how often agents switched predictions between the first and second rounds, how their confidences changed, and the accuracy of their predictions between the first and second rounds. Accuracy was an important measure because it directly corresponds to how good of predictors the MAD system and the single agent were. This provides insight into whether or not multiagent reasoning does actually outperform single agent reasoning in this context. Sharpe ratios also provided insight into the returns that would be achieved by investing solely based on predictions. The other metrics are all related to confidence scores. These are important for determining whether or not the agents have any sense about how good investments are based on the transcripts. These metrics together help answer the research questions of whether or not multiagent reasoning can result in better predictions and which dimensions of transcripts are important for predicting stock trajectories. One important thing to note is that the single agent did not follow instructions as well as the MAD agents. It did not always output a prediction, so those samples had to be dropped. The single agent had to have 24 out of 40 samples ignored because it did not provide a prediction; MAD, on the other hand, only had 4 samples ignored. If anything, this at least suggests that MAD followed instructions better. This might be because these agents had to generate less reasoning individually, so they were less likely to forget the instructions given to them at the beginning of the prompt.

To calculate confidence-returns correlation and Sharpe ratios, it was first necessary to calculate returns. For each sample, this was done by assigning returns as the difference between the 3 month stock price and the stock price from the day before. For example, if a stock's value increased from \$100 to \$120, the returns would be \$20. If it instead decreased to \$80, the returns would be -\$20. The correlation was then a simple correlation between returns and an agents confidence. Normally, a good Sharpe ratio is at least 1, suggesting that returns were based off of skill, rather than luck. Lower values suggest that returns were based off of luck, because the average returns were small compared to a volatile stock. This was calculated using the following formula:

$$\text{Sharpe}(R) = \frac{\mathbb{E}[R]}{\sigma(R)}, \quad (1)$$

The numerator is the expected returns, which was calculated as the mean of returns across all samples, and the denominator is the standard deviation of returns. The denominator corresponds to volatility because a high standard deviation means the returns vary a lot.

From Table 1, it is evident that MAD outperformed the single agent in making accurate predictions. Since there are three classes, an accuracy of 0.525 is significantly above chance; the single agent, on the other hand, fell below chance, meaning that it was not good at making predictions. Looking at the other metrics, however, MAD and the single agent performed similarly. Both MAD and single agent performed similarly in regards to AUC, but MAD actually performed below chance for this metric. It is important to note that this metric was calculated from confidence scores, so while the results still say that MAD made decent predictions, AUC suggests that the confidence scores were meaningless. Furthermore, mean confidence was very similar between MAD and the single agent, so debate did not make the models more confident. The confidence-returns correlation is also similar between the two, and fairly low. This suggests that confidence was not, in fact, very correlated with the change in stock value, disproving my hypothesis. The single agent achieved a Sharpe ratio of -0.13, which means that it typically led to losing money. This further shows that it was not good at making predictions. The MAD system achieved a Sharpe ratio of 0.1, which is also quite low, but the agents made predictions that led to positive returns. Seeing that the MAD system had an accuracy of just slightly above 0.5, this is not surprising; there is still some luck involved at that low of an accuracy. Overall, MAD was able to outperform the single agent baseline at making predictions, but the confidence scores generated by both models were almost meaningless.

In comparison, the FinDebate system in the paper by Cai et al. achieved an accuracy of 0.6 for predicting whether a stock would increase or decrease over the following month. Since stocks generally tend to increase over time, their results can likely be expected to be similar for predicting stock price changes across three months. Therefore, their work outperformed my own in accuracy. This could be because they included more helpful agents, RAG, or an overall more fruitful pipeline. They also utilized retrieval-augmented generation to obtain inputs from a more comprehensive dataset, which

would improve my results.

There are a few other points to analyze from the plots in Figure 2, found in Appendix A. First of all, Figures 2a and 2b show the accuracy and AUC of the MAD system when using transcripts of earnings call presentations versus Q&As. The presentations performed marginally better, but not by a significant amount. This suggests that those texts are similarly helpful for predicting stock price trajectories, but since only the latter half of each transcript was used, it is difficult to say for certain. It is possible that more pertinent, useful questions were contained at the beginning of Q&As and that those questions were not provided to the agents. Similarly is true for the presentations, so I cannot determine which portion of the earnings call is actually more fruitful. The MAD results are also divided into those generated using consensus by the agents and by the judge. Both of these have similar accuracies, but the judge appears to have an AUC of 1, meaning that its confidences were always higher for positive predictions. However, it is critical to note that the judge only had two samples. Out of 36 total samples (ignoring the four that were dropped), 34 of these were predicted by consensus and only two by the judge. First of all, this means that the judge doesn't have a rigorous enough number of samples to make a conclusion about its confidence scores. More importantly, this suggests that the judge was not useful for making predictions. The agents typically came to a consensus by the second round, so the judge didn't even need to be called. If they didn't come to a consensus by the second round, perhaps a more fruitful scheme would have used majority voting instead.

The individual agents must also be assessed, as these were the center component of the MAD framework. Figure 2c defines a switch rate for each agent, which is how often they changed decisions between the first and second rounds. The fundamentalist tended to switch their votes the most, nearly every 1 in 4 samples. This is somewhat surprising, since this agent is supposed to be grounded in financial fundamentals and was specifically told to focus on what the presenter was presenting the future of the company as. It seems that if the presenter predicted positive outlooks, the agent would notice that and agree. This likely is the case, so it's possible that future outlooks did not agree with actual stock price trajectory. This makes sense, because the executives want investors to foresee positive outlooks so they will invest in the company.

Method	Accuracy	AUC	Mean Confidence	Conf>Returns Corr.	Sharpe Ratio
MAD	0.525	0.49	81.25	0.25	0.10
Single	0.225	0.52	82.45	0.23	-0.13

Table 1: Comparison of MAD vs. single-agent performance across evaluation metrics.

From Figure 2d, it is clear, however, that no agent really outperformed the others in accuracy. They were pretty consistent with the overall MAD accuracy at both rounds 1 and 2 (which makes sense, because they didn't switch predictions very often). The logician had the lowest switch rate, so it seems to have been the most rigid in its original predictions. This makes sense, because it is grounded in logic, which is more difficult to debate against because it is objective. The emotion analyst was most confident in its predictions, but since confidence has already been shown to be insignificant, this is probably arbitrary.

The biggest limitation in comparing the MAD system to the single agent is that the single agent struggled to follow instructions. It often did not generate a final prediction and confidence score, so it was left with only 16 samples. Due to this, there were no cases where the single agent was correct but the MAD system was incorrect. However, there were still several failure cases where they were both incorrect, and also where MAD was correct but the single agent was not. These can be found in Appendix C.

Case A in Appendix C represents the case where the MAD system made a correct prediction, but the single agent did not. In this example, Tesla stock increased and MAD predicted as such with 68.3% confidence. The single agent, on the other hand, predicted no change with 70% confidence. The single agent addressed topics like the global chip shortage, new manufacturing techniques to increase efficiency, and other factors; however, it did not synthesize these points together for more comprehensive reasoning. It just gave a prediction and a confidence, without a holistic explanation. On the other hand, the three individual agents all agreed that the stock would increase on both rounds, but their confidences went down on the second round. The agents synthesized their points to make their predictions, too, unlike the single agent. The emotion analyst and logician both addressed the supply chain issues, which lowered their confidences, but mentioned that aspects like confidence and innovation provide a solid foundation for growth. Their

points remained similar after the second round, but they focused more on the supply chain issues and uncertainty, which decreased their confidences.

Case B represents a sample where both the MAD system and the single agent made incorrect predictions. This was a case where Tesla stock price decreased, but the single agent predicted that there would be no change with a confidence of 70% and the MAD system predicted an increase by consensus with a confidence of 85%. The single agent focused on aspects like confidence, but also vagueness, claiming that timelines and revenue guidance were unclear and that delays were consistent. Again, it did not provide a synthesis of these points, but just gave a prediction. The MAD system, on the other hand, again predicted only increases in stock value with even higher confidence than the previous situation. Together, the agents focused on Tesla's plans, like how the company planned to address safety in the Autopilot program and their AI training platform. While these aspects are important, the agents seem to have missed more detrimental topics like delays in product launches and safety concerns about Autopilot. They also lacked insight about topics not mentioned in the earnings call. For example, the Cybertruck was scheduled to be released in 2021, but was continuously pushed back until 2023. This sort of behavior creates skepticism in investors, which often causes stock price to decrease. Transcripts alone were not enough to convey this information to the agents, so they did not know about it.

The MAD system here outperformed the single agent significantly at making predictions, as it was far more accurate. However, the confidence scores have been shown to be meaningless. They don't seem to convey information about whether a prediction will be correct or how much a stock will change in price. Nonetheless, this still demonstrates that the MAD system succeeded in outperforming the single agent by using better reasoning, proving that portion of my hypothesis correct. However, it is important to consider that one of the single agent's biggest faults is that it struggled to follow instructions. Also, both systems in this

project are limited by only using earnings calls transcripts. This meant that they missed real world information that could be extremely useful for making predictions, such as peoples' opinions about companies, products, and executives. Transcripts are fruitful because they provide insider adjacent information, but another important consideration is investors' opinions, because they are also intimately responsible for a stock's price trajectory. A more fruitful option, however, would consider earnings call transcripts as well as news stories and social media: this would provide models with a much more holistic perspective on company performance and trajectory.

5 Discussion

This work could be replicable, considering that I achieved fairly consistent results across 40 samples. However, there are a few important aspects to consider: I used particular prompts that emphasized specific dimensions of interest and I used a set of very large companies from a specific time period. These are all specific factors that might not necessarily generalize to other companies and other time periods. It is possible that small companies are much harder to make predictions for, or that there are less promising dimensions to explore, for example. While results can probably vary a lot if some factors are changed, my results should be replicable with my setup. Additionally, my dataset is fairly constrained, which would affect other researchers options for projects to undertake. Since I used only 10 companies from one year, and then had to use the latter 20,000 tokens from each transcript, my data is far smaller than ideal. This would prevent analysis of topics like importance of different sections of the earnings call. Also, the labels are fairly particular, considering that they are only increase/decrease/no change between the day before the call and three months later. This does not provide a lot of other information to explore. Therefore, the financial analysis is limited, but the linguistic analysis is not. I did not look in depth at the differences between agents, or presentations and Q&As, or the decisions by consensus versus the judge. There is still a lot to explore in that regard.

Regarding ethics, this work does not have a lot of potential for harm. If I achieved stellar results with high accuracy rates, perhaps executives could start tailoring their presentations towards increas-

ing stock value, but this is not possible with these results. Additionally, if this were to happen, dataset drift would cause the MAD system to stop working because inputs would be unreliable. On the other hand, if the MAD system predicted that a company's stock would decrease in value, many people might sell it, causing a crash, but this is extremely unlikely. This would require people to ignore the contents of the actual earnings call, which could happen in theory, but the accuracy achieved in this project would not motivate investors to trust the MAD system so much. Therefore, it is very unlikely that this project causes harm to society.

This project had several pertinent limitations. As mentioned earlier, these are mostly computational: slow experiment times and costs meant that only 40 samples could be tested, and token-per-minute limits meant transcripts had to be shortened, resulting in a lot of lost information. With larger limits, perhaps higher accuracy rates could be achieved. These results were also limited by agents' abilities to follow instructions, because a sample size of 16 for the single agent is no longer statistically rigorous. The most interesting extension of this work, in my opinion, would be to combine it with other texts that provide insight into investors' opinions, like news and social media. Combining insider opinions with investor opinions would provide the models with a much larger context about a stock's trajectory. Agents would have much more information to work with, and further analysis could be done on individual agents and the judge to determine which features are the most indicative about stock trajectory.

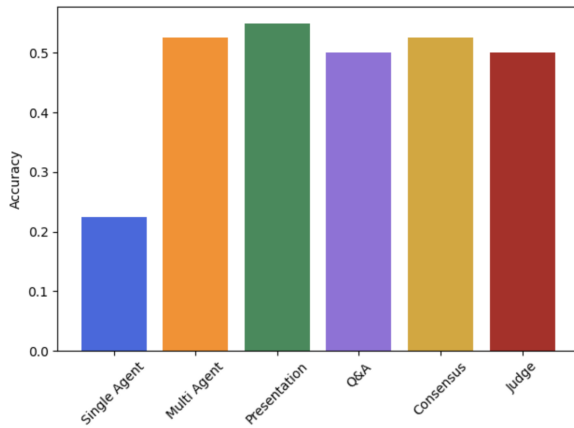
References

- [1] Tianshi Cai, Guanxu Li, Nijia Han, Ce Huang, Zimu Wang, Changyu Zeng, Yuqi Wang, Jingshi Zhou, Haiyang Zhang, Qi Chen, Yushan Pan, Shuihua Wang, and Wei Wang. 2025. [Findebate: Multi-agent collaborative intelligence for financial analysis](#). *arXiv preprint arXiv:2509.17395*.
- [2] Andrew Chin and Yuyu Fan. 2023. Leveraging text mining to extract insights from earnings call transcripts. *Journal of Investment Management*, 21(1):81–102. First Quarter 2023.
- [3] Tian Liang, Zhiwei He, Wenxiang Jiao, Xing Wang, Yan Wang, Rui Wang, Yujiu Yang, Shuming Shi, and Zhaopeng Tu. 2024. Encouraging divergent thinking in large language models through multi-agent debate. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP)*,

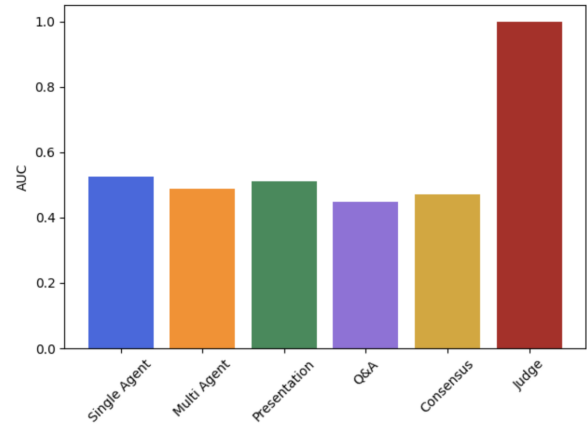
pages 17889–17904. Association for Computational Linguistics.

- [4] Yijia Xiao, Edward Sun, Di Luo, and Wei Wang. 2025. [Tradingagents: Multi-agents llm financial trading framework](#). *arXiv preprint arXiv:2412.20138*. Version 7.
- [5] Tianjiao Zhao, Jingrao Lyu, Stokes Jones, Harrison Garber, Stefano Pasquali, and Dhagash Mehta. 2025. [Alphaagents: Large language model based multi-agents for equity portfolio constructions](#). *arXiv preprint arXiv:2508.11152*.

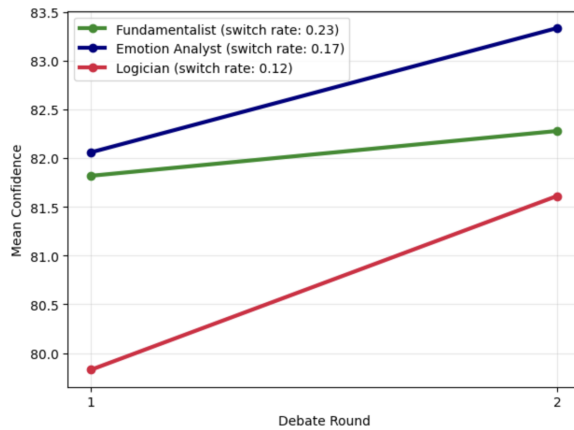
Appendix A: Results and Evaluation Plots



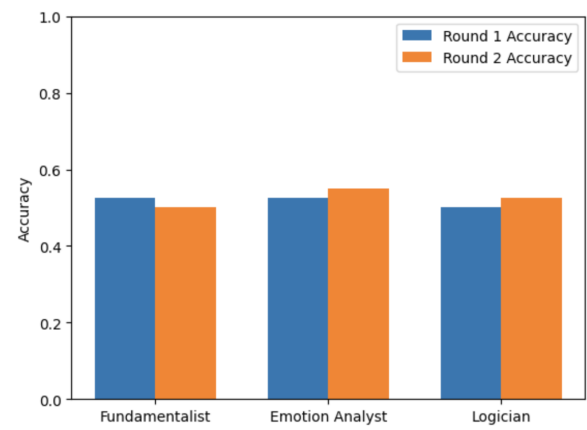
(a) Prediction accuracy of single agent vs. MAD, as well as MAD subcategories.



(b) AUC comparison single agent vs. MAD, as well as MAD subcategories. It is important to note that the judge only had two samples.



(c) Confidence changes between rounds for each agent.



(d) Agent prediction accuracy by round.

Figure 2: Quantitative evaluation of multiagent debate (MAD) vs. single-agent reasoning across prediction accuracy, AUC, and agent-level dynamics.

Appendix B: Model Prompts

Fundamentalist Agent Prompt

You are the fundamentalist agent. You focus exclusively on financial fundamentals:

- Revenue direction, margins, guidance, cash flow, risks
- Demand commentary
- Expected valuation impact over the next 3 months

Output:

1. Your reasoning
2. Final Prediction: <Increase/Decrease/No Change>
3. Confidence: <0 - 100>

Emotion Analyst Prompt

You are the Emotion Analyst Agent.

You focus exclusively on emotional and psychological cues:

- Tone, confidence, anxiety, enthusiasm, hedging, stress markers

- How these emotional signals shape near-term investor sentiment

Output:

1. Your reasoning
2. Final Prediction: <Increase/Decrease/No Change>
3. Confidence: <0 - 100>

Logician Prompt

You are the Logic & Strategy Agent.

You focus on:

- Internal consistency, contradictions, clarity
- Credibility of strategy and execution
- Realistic vs unrealistic claims
- Whether logic strengthens the bullish or bearish case

Output:

1. Your reasoning
2. Final Prediction: <Increase/Decrease/No Change>
3. Confidence: <0 - 100>

Judge Prompt

You are the judge agent. Three specialized agents (fundamentalist, emotion analyst, and logician) debated the future stock direction based on an earnings call transcript. They do not all agree.

Your job is to read the full debate transcript, weigh the arguments and evidence, and then produce a final decision.

Return:

Final Prediction: <Increase/Decrease/No Change>
Confidence: <0 - 100>

Round Prompt Used During Debate

We are in round {r} of a multi-agent stock prediction debate.

Here is the original earnings call {type} transcript:
{transcript}

Here is the debate so far (you should reference it):
{debate_transcript}

Now produce your next turn in the debate. Remember to output in this format:

<Your reasoning>
Final Prediction: <Increase/Decrease/No Change>
Confidence: <0 - 100>

Single Agent Prompt

You are a financial analyst. Your task is to analyze the following

Earnings Call {type} transcript and predict the company's stock movement over the next 3 months.

You must evaluate the transcript from three distinct perspectives:

1. Emotion Analysis

- Assess management tone, confidence, anxiety, enthusiasm, or tension.
- Explain how these emotional cues affect investor sentiment and short-term stock movement.

2. Logic & Strategy Analysis

- Evaluate internal consistency of statements.
- Identify contradictions, unclear reasoning, or unrealistic claims.
- Assess the credibility of strategies and plans.
- Explain whether management's logic strengthens or weakens the bullish/bearish case.

3. Fundamentals Analysis

- Consider guidance, revenue trends, demand commentary, risks, and financial direction.
- Identify what the numbers or business conditions imply for stock value over the next quarter.

After explicitly analyzing all three dimensions, you must:

- Provide a final prediction: "Increase", "Decrease", or "No Change".
- Provide a confidence score (0 - 100).

Keep the final answer in the following strict format:

Final Prediction: <Increase/Decrease/No Change>

Confidence: <0 - 100>

Transcript:

{text}

Appendix C: Qualitative Case Studies

Case A: Single Agent Incorrect, MAD Correct

Ground Truth

True Label: Increase

Single-Agent Reasoning and Prediction

***1. Emotion Analysis**

Elon Musk's tone throughout the earnings call reflects a mix of confidence and realism. He expresses pride in Tesla's record-breaking quarter, highlighting milestones like surpassing \$1 billion in GAAP net income. His acknowledgment of the hard work and intense efforts by Tesla employees conveys enthusiasm and gratitude. However, he also openly discusses challenges, particularly the global chip shortage, with a sense of pragmatism and caution. The mention of ""calls at midnight"" and working through ""agony of a manufacturing ramp"" indicate significant pressure and tension, especially around supply chain constraints. This realistic assessment can instill confidence in investors that the company is grounded in reality, but the acknowledgment of current and potential challenges might temper overly bullish sentiment. Short-term stock movement could reflect cautious optimism, with minor fluctuations driven by ongoing supply issues.

2. Logic & Strategy Analysis

The statements made by Musk and Kirkhorn appear internally consistent. Musk discusses the strategy of adopting new manufacturing techniques, such as introducing cast parts and structural packs using 4680 cells, which aims to improve production efficiency and cost. He admits the uncertainty and technical challenges, which adds credibility to their plan, as it does not oversimplify ambitious goals. Kirkhorn's emphasis on cost optimization, volume increase, and operating margins supports the strategy for scaling production while managing overhead. The recognition of supply chain issues as a limiting factor rather than an oversight shows a realistic operational approach. Overall, the strategy is credible but hinges significantly on overhauling supply chain issues, which recently have been unpredictable globally. This strategically supports a tempered bullish case.

***3. Fundamentals Analysis**

The fundamentals for Tesla seem strong, with record production and financial performance, supported by cost optimization and increased efficiency. However, the mention of supply chain issues, particularly the chip shortage, poses a significant risk and could limit growth potential in the near term. The development of new factories in Texas and Berlin is promising for future growth, but production there is still in the limited phase, and challenges in scaling up are expected. The progress in self-driving technology highlights potential future revenue streams, yet it is not positioned as an immediate growth driver. Positive financial indicators, such as increased auto margins and cash management, suggest potential for continued strength but counterbalanced by external supply risks.

***"

Prediction: No Change

Confidence: 70%

Multi-Agent Debate Transcript

"

Fundamentalist - Round 1: The Tesla Q2 2021 earnings call transcript presents a positive outlook with record financial achievements, but highlights significant risks pertaining to the supply chain that could impact near-term production capabilities. Here's the breakdown:

1. ****Revenue Direction and Margins****: Tesla has reached a milestone with over \$1 billion in GAAP net income, a record for the company. This comes alongside substantial improvements in auto gross profit and margin, attributed largely to cost optimization and increased production and delivery volumes. The mention of cost reduction strategies and increased operational efficiencies suggest Tesla could maintain or even improve margins in the near term. Their ability to optimize costs even amid challenges is noteworthy.
2. ****Guidance and Risks****: The guidance provided indicates optimism for the second half of the year with planned increases in vehicle delivery volumes, contingent upon supply chain constraints. The global chip shortage, however, is noted as a critical bottleneck. Tesla's ability to adapt, such as by substituting alternative chips and rewriting firmware, is a strong positive but doesn't eliminate the uncertainty tied to chip availability.
3. ****Demand Commentary****: Elon Musk highlighted that public sentiment towards electric vehicles has reached an inflection point, suggesting strong demand. The ongoing construction and planned ramp-up at Giga Texas and Giga Berlin further underline the anticipation of rising demand and the company's readiness to meet it, pending supply chain issues.
4. ****Cash Flow and Financial Stability****: Tesla reports strong operating cash flows and a robust cash balance, positioning them well to handle debt while continuing to invest in capacity expansions. This financial stability is a positive indicator for handling unforeseen challenges.

5. ****Evaluation Over the Next 3 Months****: While the long-term trajectory for Tesla appears robust, the critical short-term risk remains the supply chain disruptions, specifically related to chip availability. However, the positive financial standing and record production and delivery set a solid foundation.

Given the strong reported financial performance and ongoing demand, tempered by supply chain risks, Tesla's stock is expected to see positive valuation impacts, provided they manage to navigate these production constraints effectively over the next quarter.

Final Prediction: Increase Confidence: 75%

Emotion Analyst - Round 1: The Tesla Q2 2021 earnings call reveals a multifaceted emotional outlook, with layers of optimism intermingled with caution regarding operational challenges:

1. ****Tone and Confidence****: Elon Musk's speech exhibits a confident and congratulatory tone, particularly as he praises the Tesla team for their achievements. This confidence is likely to boost investor sentiment, as it reflects a belief in continued positive trajectory. His early remarks focus on milestones achieved, such as record production and income, suggesting pride and optimism.

2. ****Presence of Anxiety and Stress Markers****: Musk's discussion of the supply chain issues introduces a level of anxiety. He frequently mentions unpredictability and the complexity of the chip shortages, marking these as significant hurdles. The stress of navigating these obstacles is evident, underscored by his reference to logistical challenges and midnight calls. This adds a layer of reality that may temper investor enthusiasm, as it highlights the vulnerabilities Tesla faces.

3. ****Hedging and Risk Acknowledgment****: Throughout the dialogue, Musk acknowledges the risks associated with the "least lucky and dumbest" parts of the production process, revealing pragmatic caution. By admitting the difficulty in predicting chip supply resolution, he's hedging expectations. This realism aids in calibrating investor expectations—moderating overly bullish sentiments due to recognition of real threats to short-term production growth.

4. ****Enthusiasm for Future Prospects****: Tesla's progress with Gigafactories and innovations with the Model Y and full self-driving subscription indicate robust forward momentum. Musk's tone when discussing these projects encourages enthusiasm, framing a future of technological breakthroughs and expansions. This optimism about ongoing projects helps maintain positive sentiment about growth potential, albeit with the understanding of current obstacles.

5. ****Psychological Impact on Investor Sentiment****: The overall effect is a balanced emotional narrative—mixing clear confidence in long-term vision with an honest portrayal of current supply chain frustrations. Investors may feel reassured by the company's proactive measures but remain cautiously receptive to possible setbacks due to external factors.

Final Prediction: Increase Confidence: 70%

Logician - Round 1: The Tesla Q2 2021 earnings call provides a comprehensive picture of the company's current performance, strategic decisions, and the challenges it faces. Here's a breakdown of the logical points:

1. ****Financial Performance and Metrics****: Tesla achieved a record GAAP net income of over \$1 billion, a major financial milestone. The continued improvement in auto gross margins, excluding credits, suggests enhanced cost efficiencies and operational effectiveness, which are positive indicators for sustaining profitability. The specific mention of leveraging economies of scale indicates a savvy approach to handling increased production and cost management.

2. ****Supply Chain Challenges and Mitigation****: The global chip shortage is a highlighted bottleneck affecting production capabilities. While Tesla has adapted by using alternative chips and modifying firmware, the unpredictability of future supply highlights a significant risk. However, the logical endeavors to mitigate this, such as through rapid innovation and agile supply chain relationships, support the notion of Tesla's proactive risk management capabilities.

3. ****Market Demand and Growth Potential****: The strong public sentiment shift towards electric vehicles, coupled with the construction progress at Giga Texas and Berlin, indicates robust demand expectations and strategic operational expansion to meet future market needs. However, the delay in ramping up full production due to supply constraints cannot be ignored as a risk factor.

4. ****Technological Innovation and Future Outlook****: The development of the structural pack with 4680 cells and full self-driving subscription are substantial innovations despite their current limitations in scaling. These efforts signal long-term growth potential through technological advancements, positioning Tesla well in a competitive market.

5. ****Operational and Strategic Flexibility****: Tesla's ability to manage operating cash flows, reduce legacy debt, and invest in capacity expansions highlights strong financial discipline and strategic foresight. The emphasis on supporting future robustness through R&D investments and operational efficiencies is logically consistent with sustaining growth.

****Logical Strengths****: Tesla's strategic flexibility, innovation, and adaptability are critical strengths amidst the challenges of supply chain constraints, providing a solid foundation for future growth despite immediate hurdles.

****Risks and Uncertain Factors****: The persistent chip shortages pose a major threat due to their unpredictability, which could affect Tesla's ability to meet ambitious production targets in the short term.

Final Prediction: Increase Confidence: 75%

Given the robust financial position, ongoing demand, and strategic handling of existing challenges, Tesla is expected to experience a positive stock valuation impact. However, the supply chain risks necessitate a moderated confidence level.

Fundamentalist - Round 2: ****My Reasoning****

1. ****Revenue and Margins****: Tesla's achievement of over \$1 billion in GAAP net income for the first time, alongside strong improvements in auto gross profit and margins, underscores both revenue growth and operational efficiency. The improvement in margins is primarily due to better cost optimization and increased production volumes, signaling that Tesla is managing its production and sales costs effectively.

2. ****Guidance and Risks****: Although future vehicle delivery volumes are optimistic, they are largely dependent on navigating significant supply chain constraints, particularly the global chip shortage. While Tesla is adeptly mitigating these risks by using alternative chips and swift firmware adjustments, the uncertainty attached to chip availability persists as a potential constraint on revenue growth over the next quarter.

3. ****Demand Commentary****: The demand outlook is positive, with Tesla recognizing a general inflection point in

public sentiment towards electric vehicles. This growing demand is being supported by Tesla's expansion plans, such as those at Giga Texas and Giga Berlin, indicating strong anticipated sales momentum.

4. **Cash Flow and Financial Stability:** Tesla's capability to generate strong operating cash flows and maintain a robust cash balance is crucial for managing both investment in capacity expansion and the offset of any operational headwinds from existing supply chain issues.

5. **Expected Valuation Impact:** Given the strong demand outlook, enhanced operational efficiency, and robust cash flows, Tesla is poised for a positive valuation impact over the next three months. However, the supply chain risk introduces an element of uncertainty. Tesla's financial resilience and strategic management of risks suggest it can navigate these challenges, but these risks temper confidence levels moderately.

Final Prediction: Increase Confidence: 70%

Emotion Analyst - Round 2: **My Reasoning:**

1. **Tone and Confidence:** The overall tone of the Tesla Q2 2021 earnings call is confident and positive, particularly in regards to financial achievements and production milestones. Elon Musk's congratulatory remarks towards the Tesla team for their record production and income showcase a high level of confidence, which is likely to foster positive sentiment among investors. Musk's assurance about the full self-driving capabilities reflects strong conviction in Tesla's technological advancements.

2. **Anxiety and Stress Markers:** Despite the confident tone, there are notable expressions of anxiety and stress related to the supply chain, specifically the global chip shortages. Musk discusses these issues candidly, highlighting the unpredictability of the situation and the effort required to maintain production. This acknowledgement introduces realistic concerns that might temper investor enthusiasm, as the challenges are significant and beyond Tesla's immediate control.

3. **Hedging and Risk Acknowledgment:** Musk's repeated mentions of the difficulty in predicting chip supply resolution and the intricate nature of manufacturing present a balanced view, acknowledging both successes and risks. This cautious acknowledgment acts as a hedge against overly optimistic expectations, preparing investors for potential short-term disruptions.

4. **Enthusiasm for Future Prospects:** The discussion of technological innovations and expansions at Gigafactories generates enthusiasm for Tesla's long-term growth potential. The positive outlook on public sentiment towards electric vehicles and the strategic expansion plans play into investor excitement for future gains.

5. **Psychological Impact on Investor Sentiment:** The emotional interplay between confidence in long-term achievements and caution regarding short-term logistical challenges creates a complex psychological landscape for investors. While the acknowledgment of hurdles might dampen immediate exuberance, the overarching positive sentiment about future growth stimulates optimism.

Given the overall positive sentiment, tempered by a realistic presentation of ongoing challenges, investor emotion likely tilts toward a positive outlook but with cautious monitoring of the situation.

Final Prediction: Increase Confidence: 70%

Logician - Round 2: **My reasoning:**

1. **Logical Evaluation of Financial Metrics:** The achievement of over \$1 billion in GAAP net income showcases a significant milestone for Tesla, emphasizing effective cost management and operational efficiency. The improvement in auto gross margins excluding credits further highlights Tesla's capability to enhance profitability through cost optimization and economies of scale in production and delivery.

2. **Supply Chain Challenges:** The major uncertainty highlighted is the global chip shortage, which directly impacts Tesla's production output. Although Tesla has shown strategic adaptability by substituting alternative chips and quickly rewriting firmware, the unpredictability of the chip supply remains a substantial risk. The tension between achieving planned production increases and potential supply chain bottlenecks introduces a logical barrier to assured short-term growth.

3. **Demand and Market Expansion:** Tesla's strategic investments in Gigafactories and the shift in public sentiment towards electric vehicles denote anticipated strong demand and capacity to meet future market needs. The ongoing construction at Giga Texas and Giga Berlin reflects preparedness for increased production, subject to supply constraints.

4. **Technological Advancements and Strategic Flexibility:** The development of structural battery packs and advancements in full self-driving capabilities represent significant technological growth. However, these innovations bring about technical challenges that are still being addressed, which could delay their full economic impact.

5. **Financial Health and Strategic Positioning:** With robust operating cash flows and reduced legacy debt, Tesla is well-equipped to handle operational challenges and maintain investments in capacity expansions. This strategic positioning provides a buffer against supply chain disruptions.

6. **Evaluation of Risks and Growth Potential:** While Tesla's financial metrics and demand outlook suggest a positive trajectory, the ongoing supply chain challenges, particularly the chip shortages, inject a level of unpredictability into Tesla's near-term production capacity. The company's strong financial discipline and adaptability are positives, but the external risk factors demand a cautious approach.

In conclusion, while Tesla's strategic initiatives and market opportunities suggest potential stock value growth, the significant supply chain risks temper absolute confidence in a short-term valuation increase.

Final Prediction: Increase Confidence: 65% "

Prediction: Increase

Confidence: 68.3%

Decision Type: Consensus

Case B: Both MAD and Single Agent Incorrect

Ground Truth

True Label: Decrease

Single-Agent Output

Emotion Analysis - **Management Tone and Confidence**: Throughout the call, Elon Musk and his team display a confident tone, particularly when discussing their advancements in AI and renewable energy. Musk's enthusiasm is notable when he elaborates on Tesla's AI capabilities and the potential for Dojo to revolutionize neural net training. His vision for Tesla transcending traditional automotive boundaries into AI and energy sectors hints at a strong belief in Tesla's long-term potential. - **Investor Sentiment and Short-Term Stock Movement**: The team's calm demeanor regarding challenges, such as the solar roof issues and Autopilot safety concerns, might reassure investors of their problem-solving capabilities. However, the lack of concrete short-term guidance might leave some investors anxious, potentially leading to muted stock movement until clearer progress is demonstrated.

Logic & Strategy Analysis - **Internal Consistency**: Musk's explanations about Tesla's trajectory, particularly in AI expansion and energy storage, are consistent with previous statements. His logic linking energy independence through Powerwall and solar installations to broader environmental benefits is coherent. - **Contradictions or Unclear Reasoning**: While Musk's enthusiasm for autonomous technology is clear, the consistent delays in fully autonomous capabilities might cause skepticism about overly optimistic timelines. - **Credibility of Strategies**: The strategy to transition all solar products to include Powerwalls enhances Tesla's position in the energy market. The logical alignment between their automotive ambitions and energy strategies strengthens the bullish case. However, the lack of updates on urban transport solutions and occasional vagueness on timelines weakens the case for over-ambitious short-term growth expectations.

Fundamentals Analysis - **Guidance and Revenue Trends**: The transcript lacks concrete revenue guidance but underscores growth in battery production capacity, which aligns with an expected increase in demand for Tesla vehicles and energy products. The integration of AI advancements into revenue streams is still speculative rather than immediate. - **Demand Commentary and Risks**: Strong demand for both automotive and energy solutions is indicated, yet installation bottlenecks in solar roofs pose a risk. The ongoing need for semiconductor management is a critical variable that could hinder production if unresolved. - **Financial Direction**: Strategic investments in bitcoin and reliance on energy storage growth highlight diversified revenue avenues. Nevertheless, immediate profitability pressures might persist due to scaling expenses, particularly with new factories coming online in Berlin and Austin.

Final Prediction **

Prediction: No Change

Confidence: 70%

Multi-Agent Debate Transcript

Fundamentalist - Round 1: <Tesla's Q1 2021 earnings call transcript provides valuable insights into the company's current financial state and future prospects. Starting with revenue direction, Tesla continues to show strong demand both in its automotive and energy sectors. The implementation of innovative technologies like Dojo and a potential AWS-like AI service could significantly boost future revenues. In terms of margins, Tesla's vehicle and energy storage segments are trending towards better margins, although Megapack has more room for cost optimization. Meanwhile, the assembly plants in Austin and Berlin are expected to improve scale and efficiency, reducing production costs over time. Despite certain challenges in areas like the Solar Roof, the strong demand can balance things out. Guidance remains positive with a focus on scaling battery and vehicle production. Cash flow management seems prudent, as they are leveraging alternative strategies like Bitcoin investments to enhance liquidity. The main risks revolve around operational scale-up challenges and regulatory hurdles, particularly with autonomous driving systems.

Demand commentary is positive, with future production expansions indicating strong continued market interest. Expected valuation impacts are likely to be positive over the next three months, given the strategic expansions and technological advancements aimed at increasing efficiency and capturing new market opportunities.>

Final Prediction: Increase Confidence: 85

Emotion Analyst - Round 1: <Tesla's Q1 2021 earnings call exudes confidence, primarily due to Elon Musk's enthusiastic communication style. His responses indicate a firm belief in Tesla's technological advances and strategic direction, particularly with AI and energy solutions. Musk's tone implies high confidence and optimism about Tesla's ability to tackle current operational challenges, emphasizing their innovation prowess and future growth potential.

Despite the challenges with Solar Roof installation and the complexity of scaling production, Musk maintains a positive outlook. This is underscored by his assertion that demand for their solar products remains strong despite price increases. His focus on the opportunity presented by Powerwall deployments further underscores Tesla's commitment to creating a robust integrated energy solution.

Additionally, the discussion on digital currency investments reveals strategic financial management, indicating Tesla's

adaptive approach to maintaining liquidity in uncertain environments. Zachary Kirkhorn's calm and calculated tone regarding Bitcoin demonstrates assurance in Tesla's financial strategies.

Overall, the investors' sentiment influenced by Tesla's leadership's confidence, the enthusiastic future outlook in technology and energy domains, and the strategic adaptability, suggests a positive near-term sentiment.>

Final Prediction: Increase Confidence: 90

Logician - Round 1: <My reasoning will focus on the logic and credibility of Tesla's strategic plans and statements during the Q1 2021 earnings call. Elon Musk highlighted Tesla's development as not only an automotive or energy company but as a leading AI and robotics firm, comparing the potential of Dojo to AWS-like services. While ambitious, the vision of deploying Dojo for external use must be approached with skepticism until the actual product viability and market adoption are proven. The analogy to AWS implies significant capability, but AWS's success stems from years of proven reliability and adoption across diverse industries, which Tesla has not yet achieved in AI services.

Elon's discussion on the Solar Roof reveals substantial installation challenges. The explanation of varying roof complexities does make sense, but the financial viability concerns are serious unless efficiency improves drastically. Musk did present a logical strategy by standardizing installations with Powerwall integration, which aims to reduce variability and enhance scalability. This approach, if successful in execution, could reinforce Tesla's energy business. The explanation of Tesla's Bitcoin strategy indicates an unconventional but logical approach to managing liquidity. However, the financial strategy of investing heavily in Bitcoin carries significant risk due to market volatility, potentially affecting Tesla's cash flow predictability.

The conversation on autopilot safety and the misleading media narratives highlighted logical defenses for Tesla's approach to safety. Statements about Tesla's collaborative investigations with authorities are credible steps toward transparency and commitment to safety improvements.

Overall, most strategic moves seem to be underpinned by a logical foundation. However, the success of ambitious projects like Dojo and Solar Roof depends heavily on execution and real-world viability, suggesting that risks could impact outcomes.

Despite these risks, the strategic expansions and innovations Tesla is pursuing are logical and, if successful, offer substantial growth potential. Therefore, the logic tends to support a positive outlook, but real-world execution remains a pivotal factor.>

Final Prediction: Increase Confidence: 75

Fundamentalist - Round 2: <Tesla's Q1 2021 earnings call presents a robust fundamental outlook, underpinned by strong revenue growth across its automotive, energy storage, and potential AI sectors. The strategic developments in Gigafactories in Austin and Berlin promise to enhance production efficiency and scalability, thus improving margins over the medium term. Tesla's autonomous driving technology and associated AI services via Dojo represent long-term revenue avenues, although these will require proof of market viability.

From a guidance perspective, Tesla has been affirmative about scaling battery and vehicle production, which aligns with growing demand commentary. Elon Musk's discussion on infrastructure investment for energy solutions highlights a strategic pivot towards addressing high-demand sectors, such as integrated solar and battery systems, which posit an expanding margin potential.

However, there are tangible risks in ramping up new factories and managing supply chain complexities amid global challenges. Additionally, while the Bitcoin investment strategy offers liquidity advantages, it introduces volatility into Tesla's financial positioning.

The fundamental analysis reflects strong demand across core segments with promising future guidance, underpinned by technologically-driven operational advancements. The primary risks remain with execution and external supply chain factors.

Given these considerations, the projection remains positive due to Tesla's capability of addressing demand with large-scale production optimizations and expanding into AI and energy markets.>

Final Prediction: Increase Confidence: 80

Emotion Analyst - Round 2: <The emotional tenor of Tesla's Q1 2021 earnings call strongly influences the perception of investor sentiment. Elon Musk's communication style is marked by enthusiasm and high confidence in Tesla's strategic vision, particularly in AI and energy solutions. His detailed outline of Tesla's AI and Dojo capabilities, as well as the company's proactive steps in addressing solar roof challenges, project a tone of assured optimism. Despite obstacles, such as the complexity of solar installations and production scaling, Musk's reassurances and the clear articulation of solutions contribute to a robust positive sentiment.

The call demonstrates Tesla's strong belief in its technological competencies and future potential, which is key in maintaining investor confidence. The narrative presented is not only about current success but also about future innovation, fostering a sentiment that aligns with growth and sustainable leadership in the industry. Additionally, the discussion on financial strategies and Bitcoin investments by Zachary Kirkhorn reflects a calm, strategic approach, further reinforcing a positive outlook on Tesla's financial management.

Overall, the consistent and confident expression of Tesla's long-term vision and immediate strategic actions creates an emotive appeal that suggests a positive investor sentiment, influencing expectations of an increase in Tesla's valuation in the near term.>

Final Prediction: Increase Confidence: 95

Logician - Round 2: <My analysis of Tesla's Q1 2021 earnings call zeroes in on the logical underpinnings and tradeoffs in their strategic approach. Tesla's focus on Dojo as an AI service is ambitious. Drawing parallels to AWS suggests a potential revenue stream that is transformative but carries high initial development risks and the need for market acceptance, which Tesla hasn't yet proven outside its automotive space. Comparing Dojo to AWS implies a level of platform integration and customer reliance that Tesla must establish over time.

The challenges with the solar roof installations highlight Tesla's need for operational consistency in new product markets. Musk's strategy to include standardized Powerwall installations with every solar sale aims to streamline operations and

resolve cost discrepancies. This strategic shift shows a logical approach to managing and eliminating inefficiencies but will need careful execution and might face resistance based on cost perceptions.

Tesla's unconventional decision to hold Bitcoin as part of their liquidity management strategy is logically intriguing. Purchasing Bitcoin aids liquidity in the current low-interest environment, providing cash flexibility for upcoming factory launches. However, the volatility of cryptocurrency markets could affect Tesla's balance sheets and is a clear risk that needs to be accounted for in financial planning.

On the safety front, Tesla's proactive collaboration with safety authorities in addressing claims about Autopilot accidents underscores a logical dedication to transparency and product reliability. The explanations on safety reflect commendable steps toward maintaining customer trust and regulatory compliance.

Tesla's strategic plans, while grounded in logical steps towards innovation and efficiency, hinge considerably on execution in nascent areas such as AI services, solar energy solutions, and financial management involving Bitcoin. The vision remains sound in theory, but execution risks, particularly concerning global supply chain logistics and market acceptance for new-tech implementations, suggest that exceptional operational follow-through is crucial.

Nevertheless, if successful, these strategic ventures could provide Tesla significant long-term growth, substantiating a positive market outlook. Given these factors, the prospects of an increase in Tesla's valuation remain strong, although contingent on effective execution.>

Final Prediction: Increase Confidence: 80 "

Prediction: Increase

Confidence: 85%

Decision Type: Consensus