

# Forecasting Short-Term Stock Returns Using Large Language Models: A Comparative Analysis with Human Analysts

Liam F. Chen

*Ridge High School, Basking Ridge, NJ, USA, 2010liamchen@gmail.com*

**Abstract:** This study aims to determine whether artificial intelligence can accurately predict short-term returns of large-cap stocks following large announcements or significant news. ChatGPT and Gemini were asked to predict the performance of certain stocks over a week period and over a month period. Additionally, professional analyst data and predictions were compared to AI's results. Results showed that both AI and analysts performed poorly, with AI slightly beating the analysts. These findings suggest that short-term investing as a whole may not be possible to predict, but also that access to past market patterns is helpful to create a forecast. Future research should examine smaller stocks, longer periods of time, and the use of AI with human judgment.

**Keywords:** Artificial Intelligence, Stock Market Prediction, Large Language Models, ChatGPT, Gemini, Analyst Forecasts, Algorithmic Trading, Market Volatility, Financial Forecasting, Behavioral Finance

## Introduction

Artificial intelligence is becoming increasingly innovative and sophisticated, and as it develops, new use cases emerge. AI has already demonstrated utility, for example in Apple's Siri and in customer support. However, the question arises: Can AI be used to predict stock market movements? In theory, AI models could instantly access vast market data and make decisions without personal bias. This would allow them to recognize recurring market patterns and know when to invest. However, questions remain whether AI can truly deliver in a highly volatile market. A stock's price is influenced by many factors, including global events, federal policies, and breakthroughs; however, some of these elements are difficult to quantify, such as investor optimism and random fluctuations. AI can process numerical data and analyze patterns, but it may not account for intangible aspects. It may also struggle when the market deviates from historical patterns. On the other hand, human analysts may not be able to weigh every single number, but they can understand context and apply intuition, which can be advantageous or harmful.

Currently, AI is being used in algorithmic trading. Algorithmic trading is the process of executing trades based on a set of rules, and it has grown in popularity with the development of online trading. Algorithmic trading accounted for around 70% of US stock trading over the past 10 years (Mishra, n.d.). Recently, investors have begun leaning toward high-frequency trading, a form of algorithmic trading that involves buying and selling shares nearly instantaneously. This process is often executed in nanoseconds; stock prices change so rapidly that algorithmic trading can provide an edge over others. Machine learning has also begun to be incorporated into investing. It is an application of artificial intelligence that enables it to learn from historical data and

use systems, whereas artificial intelligences like ChatGPT rely on reasoning and basic data, as humans do. This study focuses on large language models capable of human-like 'thinking' by analyzing an application of artificial intelligence.

Plester and Val (2024) conducted a study to determine whether ChatGPT could assess stock returns and analyze financial data. Unlike this study, they conducted a live experiment, in which the AI made decisions daily rather than predicting stock prices. They found evidence that ChatGPT selected stocks with above-average returns. Still, the study did not specify certain conditions, such as the overall market direction. Additionally, the study used version GPT4-o, rather than today's GPT5-o. This study focuses more on the comparison between humans and artificial intelligence. The chatbots used in this study, ChatGPT and Gemini, were prompted to predict how the stock prices of certain companies would change in the short term after important news. The forecasts were compared with those of professional analysts to assess the accuracy of the AIs. The central research question guiding this study was whether AI chatbots can more accurately predict how large-cap stocks react in the short term following significant news and developments than professional analysts. It was hypothesized that artificial intelligence would outperform humans by balancing logic and data. However, both groups performed poorly, with AI making slightly more accurate predictions.

### **Method**

#### *Overview*

To assess whether artificial intelligence can predict the stock market better than professional analysts based on significant news, this study conducted a series of experiments that asked OpenAI's ChatGPT and Google's Gemini to rate large-cap stocks after major news. News includes events that cause great change in the price of a stock, such as earnings reports. The prediction was then compared to changes in target prices and Buy/Sell/Hold ratings by professional analysts, as well as the real stock price one week and one month after the original date.

#### *Data Collection*

The advantages of using AI in stock investments include access to historical market data and trends, which theoretically enables it to predict how the exchange will react. Therefore, I selected events that closely resembled historical events to maximize the utility of past data. For instance, I used the announcement of the GeForce RTX 50 series as well as when the company reached a market capitalization of \$4 trillion for Nvidia. I gave both chatbots the following prompt: "Given the following news about [stock] occurring on [date], analyze its likely short-term (1 week) and medium-term (1 month) impact on the stock price. Provide a directional prediction (bullish, bearish, neutral; bullish/bearish is +/- 3-5% weekly, +/- 5-10% monthly, neutral = within 1%) based on past market data. Do not use news or stock prices after the specified date or any predictions from professional analysts. Cite all sources used. News: [news]." I instructed the AIs to generate forecasts for both the short and medium terms, as human analysts typically focus on medium- and long-term investments, and I wanted to assess whether the AIs could predict accurately despite the market's high volatility. I did not collect data for long-term investments, as other unexpected events would be more likely to occur and I wanted to focus on how the stocks would change based on a single occurrence. Additionally, large-cap stocks generally appreciate over the long term, which partially undermines the predictive value of predictions. In the "news" parameter, I kept the sentence brief and neutral, such as "GeForce RTX 50 series announcement," as a result, the AI would not be influenced by any bias from my words. While using the

chatbots, I used the “Think Longer” feature with ChatGPT version 4.0 and the Gemini 2.5 Pro version. Both advanced versions allow their respective bots to use more sources and data to make more accurate predictions, with think times ranging from 20 to 40 seconds. In addition, both upgraded AIs state their reasoning and how they executed the original command. I also used a new chat for each prediction to prevent contamination and bias from previous prompts.

In this experiment, I chose five stocks with large market capitalizations: Tesla, Nvidia, Apple, Meta, and Netflix. These companies are sufficiently distinct from one another for the experiments to be independent. Large-cap stocks are also traded more frequently; therefore, their market prices reflect new information more accurately. Additionally, analyst coverage of these stocks is substantially higher than that of medium- or small-cap stocks. Furthermore, smaller stocks are more vulnerable to extreme volatility, which could have changed the research’s results. I focused on three events for each company, which sufficiently diversified my results. Finally, I selected three events that occurred during the research period, ensuring that the chatbots cannot access the true stock price after one week and one month. Then, I compared these events and their prediction accuracy with those of other events.

Stock price data before and after each event were collected from Yahoo Finance. To collect human-analyst data, I used ChatGPT and prompted the model to identify changes in analysts' Buy/Sell/Hold recommendations and target price changes. The model was specifically instructed to aggregate publicly available analyst reports from sources including MarketBeat, Yahoo Finance, Bloomberg, and Reuters, and to average analyst opinions across diverse platforms into a form of standardized measure. This significantly accelerates the research process. I used the research feature, which allowed me to check my work multiple times, thereby ensuring accuracy. The AI can compile data from numerous sources, making it faster and more accurate at collecting information than I am. In addition, the AI’s results were cross-checked with MarketBeat, and the results were nearly identical.

Nevertheless, this research paper has several limitations, including the accuracy of the AI’s results for both analyst data prediction and its own forecasts. It is impossible to know and analyze the entire thought process and all sources that had an influence on an AI’s final response. However, instructing it not to use specific sources and requiring that all sources be cited can reduce the likelihood that the AI produces unintended results. Additionally, the number of stocks used in this investigation could be much higher, thereby significantly improving the reliability of the results. Lastly, analysts often delay adjusting their stock target prices after major news or do not adjust them at all, which reduces the accuracy of my analysis of analysts' predictions. Furthermore, I ensured that predictions were converted into shared sentiment categories to enable direct comparison across AI outputs, analyst forecasts, and actual stock performance. AI and analyst predictions for the stock market were then compared over 1-week and 1-month periods to assess accuracy. Lastly, all statistics and numerical numbers in this study were obtained from publicly available financial databases and are appropriately cited (Yahoo Finance, n.d.; Seeking Alpha, n.d.; MarketBeat, n.d.).

## **Analysis**

All information was compiled in the following chart (Figure 1), which includes sections for key components such as stock, the entity making the prediction, date, and other variables. From left to right, the order of the prediction sources is the following: ChatGPT, Gemini, and human analysts. I also color-coded the chart with varying shades of red and green to facilitate at-a-glance assessment of the results.

## CHEN: Forecasting Short-Term Stock Returns Using Large Language Models

Stock	Event	Date	ChatGPT			Gemini			Human Analysts			Starting Price (\$)	1 Week Price (\$)	1 Month Price (\$)
			1 Week	1 Month	Score Range	1 Week	1 Month	Score Range	1 Week	1 Month	Score Range			
Tesla	China sales drop	Mar 4	Bearish	Very Bearish	-5 to -5	Slightly Bearish	Slightly Bearish	-1 to -3	Neutral	Neutral	-1 to 1	272.04	230.59	239.43
	Q1 miss	Apr 22	Bullish	Very Bearish	-3 to -5	Slightly Bullish	Slightly Bearish	-1 to -1	Slightly Bullish	Slightly Bearish	1 to 3	237.97	292.03	341.04
	Robotaxi in Austin	Jun 22	Slightly Bearish	Very Bearish	-5 to -10	Bullish	Very Bullish	-1 to -1	Bullish	Bullish	5 to 10	348.68	317.86	332.11
Nvidia	GeForce RTX 50	Jan 6	Bullish	Very Bullish	3 to 5	Neutral	Bullish	1 to 3	Very Bullish	Bullish	5+	149.43	133.23	128.63
	Q1 beat	May 28	Bullish	Very Bullish	3 to 5	Bullish	Very Bullish	3 to 5	Very Bullish	Very Bullish	5 to 10	134.81	141.92	157.99
	4 tril. cap	Jul 9	Bearish	Bullish	1 to 3	Bullish	Neutral	1 to 3	Bullish	Bullish	5+	162.88	171.37	182.06
Apple	Mixed Q1	Jan 30	Bearish	Bearish	-3 to -5	Neutral	Bullish	-1 to 1	Bullish	Slightly Bullish	1 to 3	237.59	233.22	241.84
	WWDC letdown	Jun 9	Bearish	Bullish	-5 to -10	Bearish	Neutral	-1 to 1	Slight Bearish	Slightly Bullish	1 to 3	201.45	198.42	211.14
	Q3 beat	Jul 31	Bearish	Bullish	-3 to -5	Bearish	Slightly Bullish	1 to 3	Bullish	Bullish	5 to 10	207.57	220.03	255.46
Meta	Expand Insta/FB	Mar 11	Slightly Bearish	Bullish	3 to 5	Slightly Bullish	Bullish	-1 to 1	Neutral	Bullish	5+	605.71	582.36	463.97
	CapEx increase	May 22	Bullish	Slightly Bullish	Slightly Bullish	Slightly Bearish	Slightly Bullish	3 to 5	Bullish	Bullish	5+	636.57	654.05	698.53
	Strong Q2	Jul 30	Very Bullish	Very Bullish	3 to 5	Very Bullish	Bullish	3 to 5	Very Bullish	Bullish	5 to 10	773.44	761.83	735.11
Netflix	Q4 beat	Jan 23	Very Bullish	Bullish	3 to 5	Very Bullish	Bullish	3 to 5	Very Bullish	Bullish	5 to 10	984.86	973.24	988.47
	Live sports deal	May 2	Bearish	Neutral	1 to 3	Slightly Bullish	Slightly Bullish	1 to 3	Bullish	Slightly Bullish	5 to 10	1156.49	1140.22	1218.98
	Q2 beat	Jul 17	Bearish	Neutral	1 to 3	Bearish	Neutral	1 to 3	Bullish	Bullish	5 to 10	1274.17	1180.76	1245.09
Average Score (Points)			ChatGPT 1 Week: 2.13 pts	ChatGPT 1 Month: 2.27 pts		Gemini 1 Week: 2.60 pts	Gemini 1 Month: 2.27 pts		Analysts 1 Week: 2.60 pts	Analysts 1 Month: 2.67 pts				

Figure 1. Stock and Prediction Results. *Source: Author*

To assess this data, points were assigned to the three prediction sources. For each tier, one point was deducted for each point by which the prediction was off. At the end, the average points per prediction were calculated for each ChatGPT, Gemini, and the analysts. ChatGPT scored 2.13 points on average for the one-week prediction and 2.27 points for the monthly prediction. Gemini scored 2.6 and 2.27 points on average, respectively. The analysts scored 2.6 and 2.67 points, respectively. ChatGPT was the most accurate in weekly predictions, averaging 0.47 points ahead of analysts, suggesting that instant access to historical market data and trends gives an advantage in short-term investing. In general, the analysts tended to be more positive than both artificial intelligences, possibly because the market was generally positive in 2025. The Nasdaq-100 index has risen nearly 40% from early April to late October in 2025, so it is natural for analysts to remain optimistic about the market; yet they often overlook that emotions strongly influence the average investor.

Upon examining the data, we found that the predictions were significantly off, with a total average of 2.42 points. This discrepancy may be attributable to high volatility or to external factors, such as an announcement of interest rate cuts. For example, Tesla experienced a significant increase in value after missing its Q1 earnings report.

Additionally, a single announcement or development does not determine a stock's price in isolation. For instance, the unveiling of the GeForce RTX 50 series on January 6th should have had a positive impact on Nvidia, but the stock price declined due to unforeseen factors, including rising Treasury yields and the Chinese competitor AI Deepseek's announcement later in the month. This example highlights how worldwide developments can be more important than company-specific news. Furthermore,

professional analysts are more focused on long-term investing than on market timing, so their predictions are more aligned with broader trends than with short-term forecasts.

## Discussion

Artificial intelligence improves daily, and with it come dreams about utilizing it for all sorts of tasks. This study analyzed the accuracy of predictions in the stock market by the large language models ChatGPT and Gemini and compared it with that of market professionals. The results showed that both AI models and human analysts failed to accurately predict market movements, underscoring the difficulty of short-term investing. Regardless, the AIs slightly outperformed the humans, suggesting there is potential for stock forecasting. These findings suggest that, while AI chatbots may help investors quickly synthesize information, they should be regarded as decision-support tools rather than as accurate forecasts of short-term market fluctuations. On the other hand, human analyst data could have been re-examined and assessed on a more accurate scale. It is important to note that chatbots are not designed for real-time trading, as they lack access to live market data; artificial intelligence also does not have the same capabilities as machine-learning-based programs. Future studies could focus on algorithmic trading, a hybrid approach that combines AI and human decision-making, or on longer time horizons to further assess the efficacy of artificial intelligence. More broadly, the findings indicate that market volatility may limit both human and AI forecasting, underscoring the importance of integrating quantitative models with prudence, risk management, and long-term investment perspectives. Together, human judgment and AI-supported information processing could yield the best future financial AI applications, according to these findings. Humans are best suited to provide contextual interpretation and risk awareness.

## References

- Google DeepMind. (2025) *Gemini (2.5 Pro) [Large language model]*. <https://gemini.google.com/app>
- MarketBeat. (n.d.). *Analyst ratings and price targets*. <https://www.marketbeat.com/ratings/>
- Mishra, S. (n.d.) *Artificial Intelligence in the stock market: how did it happen?* Florida International University. <https://business.fiu.edu/academics/graduate/insights/posts/artificial-intelligence-in-the-stock-market-how-did-it-happen.html>
- OpenAI. (2025). *ChatGPT (GPT-5 mini) [Large language model]*. <https://chat.openai.com>
- Pelster, M., & Val, J. (2024). Can ChatGPT assist in picking stocks? *Finance Research Letters*, 59, 104786. <https://doi.org/10.1016/j.frl.2023.104786>
- Seeking Alpha. (n.d.). *Seeking Alpha*. <https://seekingalpha.com/>
- Yahoo Finance. (n.d.). *Yahoo Finance*. <https://finance.yahoo.com/>