

Is Stock Forecasting Viable with ChatGPT? Investigating Asset Management with Beginner Traders

Anderson Lai¹ and Sean McGrath[#]

¹Chadwick School

[#]Advisor

ABSTRACT

There is a data disparity between institutional traders' stock analysis tools and the rudimentary analysis tools of beginner traders; in recent years, experts have theorized that ChatGPT might help bridge this gap. While some AI advocates note AI's capabilities over human intelligence and its ability to understand complex financial concepts, critics of AI assert that its lack of humanistic value raises concerns surrounding the reliability of ChatGPT in providing financial advice for an emotionally driven market. This paper examines whether ChatGPT4 can aid in stock selection for beginner stock traders (<1 year of experience) and help them outperform the stock market. The study applied a mixed-methodology analysis using a focus group and mock stock portfolios on MarketWatch, a financial news website with trading simulation features. The results suggested that AI-supported portfolios outperformed non-AI-supported portfolios, which par market baselines from major indexes like SPY and DJI. Additionally, AI-supported participants were more likely to report an increased understanding of the market than non-AI-supported participants, albeit with the risk of inaccuracies in ChatGPT's advice and technical difficulties of ChatGPT prompting. The study concludes that ChatGPT can serve as a financial assistant and educational tool to empower beginner investors to outperform the market.

Introduction

Background

During the latter half of the 20th century, the increase of investments in chip and microprocessor industries, defined as the "microelectronics revolution," facilitated the growth of machine learning and artificial intelligence. By leveraging the era of computing power and mass digitization, several wealth management groups invested immense sums of money to create predictive algorithms and automate the trading process for profit maximization; hence, the term "high-frequency trading" describes this process of instantaneously trading large volumes of stock (Lin, 2014). Mathematical models over vast data sets have empowered professional quantitative traders (quants) to identify trading opportunities and buy and sell securities at an unprecedented level, allowing large institutions to excel. As a result, the data from institutional traders' exclusive stock analysis tools overpowers the non-professional, individual trader who manages their investment portfolios with rudimentary technology.

In the year 2019, ESCP business school professors Andreas Kaplan and Michael Haenlien defined artificial intelligence (AI) as a "system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation." Hence, the interpretative ability of AI has prompted new paths and possibilities in the financial sector for institutional traders and personal investors alike. Advances in AI have led scholars and businessmen like Harvard professor Graham Allison

and Google Chairman Eric Schmidt to claim that “AI technologies will be drivers of future economic growth and national security” (Gilli et al., 2020). Notably, with the establishment of ChatGPT in November of 2022, this generative artificial intelligence and its ability to output humanlike dialogue accumulated a valuation of \$86 billion. Given ChatGPT’s exponentially more accessible and user-friendly nature compared to professional forms of interpretive software, such a craze has raised questions about the ability of ChatGPT to match the capabilities of professional algorithms and operate as an effective stock-forecasting tool for individual investors. Consequently, critics have hinted towards the ability of a consumer-level AI form such as ChatGPT to act as a financial tool for individual investors (Knutsson, 2023).

Gap in the Knowledge and Current Question

The gap in professional knowledge regarding ChatGPT stock prediction remains in the experimental analysis of AI aid for beginner stock traders. Since professional studies have not explored ChatGPT’s abilities for the average consumer, this gap in ChatGPT’s capabilities for novice traders is pertinent. To elaborate, previous studies have focused on ChatGPT capabilities in a broader tier, focusing on more general reasoning tasks or identifying technical terms. Moreover, professional research has focused on unlocking new strategies for stock forecasting that remain exclusive to the quant traders and stock analysts within the most elite circles of the stock trading community. Thus, focusing on beginner, inexperienced traders explores this gap in prior research methods. As a result, a gap persists in exploring ChatGPT’s specific stock picking and asset allocation abilities. Additionally, since ChatGPT is primarily used as a consumer-level artificial intelligence for composing emails or summarizing articles, ChatGPT can cater most to the average American investor, drawing a connection between ChatGPT and the beginner investor. Such aims ultimately beg the question: To what extent can ChatGPT aid in picking stocks for beginner stock traders to outperform the American stock market?

Literature Review

Optimism Surrounding ChatGPT

Regarding AI as an alternative method to predicting the stock market, it is critical to consider the capabilities of AI in human intelligence. Specifically, machine learning software such as IBM’s Deep Blue chess-playing program in 1997 began to succeed and proved more effective than humans. Likewise, beyond the ability for AI to outperform in purely algorithm-based tasks with a limited number of actions, AI has also exhibited the potential to outperform in tasks requiring a level of dexterity, precision, and an advanced ability to account for simultaneous activity: Google’s adaptive AI DeepMind in 2010 quickly mastered video-game operations and was able to outperform the top professionals in various multiplayer games (Gilli et al., 2020). In being highly practical and rational, AI has had the capabilities and potential to outperform human intelligence in its processing rates and execution of specific algorithm-based tasks for decades (Haenlein & Kaplan, 2019). Given AI’s reputation for offering unprecedented processing performance, the integration of natural language processing models has been applied to various fields, specifically in healthcare. In a 2023 study conducted by Vanderbilt University medical expert Douglas Johnson, ChatGPT-generated responses were directly compared to the medical responses of 33 physicians, ultimately concluding that ChatGPT-generated responses were generally accurate (Johnson et al., 2023). Since ChatGPT can leverage internet sources and extrapolate conclusions within highly complex and sensitive fields, Johnson and other researchers concluded that incorporating ChatGPT into healthcare can significantly elevate the accessibility of medical information for healthcare providers and patients.

Paralleling this application of ChatGPT to field-specific dilemmas, numerous researchers have also explored the ChatGPT model's ability to understand complex financial concepts compared to human financial experts. Specifically, by comparing ChatGPT responses and human responses in the face of financial dilemmas, economist Chen Ren from Hefei University applied a dataset of 7165 financial questions and identified that the ChatGPT responses demonstrated higher levels of professionalism, accuracy, and efficiency than most expert-generated responses (Ren et al., 2023). His conclusion: ChatGPT can dissect complex financial dilemmas at an unprecedented level beyond human scope. Beyond AI's ability to solve financial dilemmas, ChatGPT can also understand complex concepts in the financial field that are already known and not known, revealing both an advanced understanding of finance and the ability to derive unknown knowledge inductively. Extending upon Ren's conclusion in terms of ChatGPT's aptitude for finance, Wenzlaff and Spaeth, who are economists from the University of Hamburg, conducted their study regarding the ability of the ChatGPT model to define and understand domain-specific knowledge such as crowdfunding, alternative finance, and community finance; hence, they ultimately gathered that ChatGPT was capable of accurately answering all questions to which knowledge is available online and can extrapolate conclusions for answers that are not known (Wenzlaff & Spaeth, 2022). Therefore, Ren's research on ChatGPT's effectiveness in unpacking complex finance situations and Wenzlaff and Spaeth's analysis on ChatGPT's power to fill gaps in knowledge via inductive reasoning collectively express the optimism surrounding ChatGPT's potential effectiveness in real-life financial situations.

Skepticism Surrounding Thinking Models

Nonetheless, despite ChatGPT's ability to outperform humans algorithmically and understand finance-specific concepts to an equivalent degree, the utter lack of humanistic value and emotional intelligence raises concerns surrounding the reliability of ChatGPT in providing financial advice, especially given the impact of emotion on the behavior of the market. In Ohio State University professor Angus Fletcher's logical proof on the inability of computers to read or write literature, Fletcher asserted that the reliance of AI on the CPU's Arithmetic Logic Unit, which performs its computations using symbolic logic devoid of causal reasoning, severely inhibits the ability for the AI to interpret creative works. Since causal reasoning is also necessary to produce the narrative components of literature, Fletcher theorized that the artificial preference for symbolic logic over causal reasoning prevents the computer from dissecting and assembling literary elements that do not necessarily follow a singular and logical progression (Fletcher, 2021). In other words, AI struggles with the abstract and underlying messages within creative literature due to its highly systematic way of analyzing information. Similarly, applying such theories to stock prediction and diagnosis tasks, the absence of causal reasoning would make understanding market dynamics and the role of human-related market adjustments impossible for the ChatGPT AI, which would inhibit the stock-picking abilities of the ChatGPT software. To ground the importance of investor sentiment in dictating market prices, Wang and Zhao's study from Cambridge International School analyzed the impact of news events on investor sentiment in the Chinese market. It concluded that investor sentiment prediction models had high accuracy, showing a strong correlation between human emotions and stock pricing (Wang & Zhao, 2020). Since human emotion is rapidly volatile and may not always reflect the tone of the news event to its absolute purest nature, ChatGPT's AI text recognition may fail and produce a slight deviation from accurate projections. Regarding the ChatGPT thinking model, the absence of inductive reasoning skills within the thinking model may hinder its ability to understand the human psychology and intangible features of the market, requiring more than merely technical data analysis to make stock predictions.

In addition to the inability to fully comprehend the role of human emotion in the market, experimental data has confirmed that ChatGPT also needs more quant-level, state-of-the-art methods in terms of accuracy and reliability. Notably, by applying ChatGPT and GPT-4 to 25 different analytical natural language processing tasks with over 48,000 prompts, Wroclaw University's Jan Kocoń concluded that ChatGPT's outputs tend to

be more inaccurate with more difficult reasoning problems, and such ChatGPT-generated responses were inaccurate and inferior to state-of-the-art methods, raising concerns about the capability of ChatGPT as a genuinely beneficial predictive AI. Kocoń and other artificial intelligence specialists asserted that ChatGPT was a "master of none" but a "jack of all trades," which synthesizes the underlying concerns surrounding the use of ChatGPT (Kocoń, 2023). In fact, given the limited knowledge of the stock market and minimal financial literacy of the average consumer, the inaccuracy and overall inferiority of ChatGPT to more modern and field-specific algorithms raises concerns about ChatGPT's stock-picking ability to beat the likes of professional traders; however, this information still does not overrule the idea of ChatGPT as a supplementary tool and a powerful finance research tool.

ChatGPT as a Research Tool

Regarding ChatGPT being an advantageous financial research tool, University of Dublin professors Dowling and Lucey explored whether ChatGPT has any place or role in the financial field by comparing independent ChatGPT research to ChatGPT research supplemented by user expertise. Specifically, when Dowling and Lucey experimentally tested ChatGPT's synthesis abilities in research, they concluded that ChatGPT can create "plausible-seeming research studies for well-ranked journals," which is especially important in considering the advantages of ChatGPT in mass data analysis and its ability to note trends. Hence, Dowling and Lucey's data prove that ChatGPT is valuable and effective in research and understanding the complexities of the market; however, the authors note that the quality of its insights became much more acceptable when supplemented by the addition of private data and researcher expertise, which casts doubt upon its utter autonomy in the stock market (Dowling & Lucey, 2023).

To expand upon private data implementation in Dowling and Lucey's experiments, Marquette University economics professor David Krause evaluated the benefits and limitations of ChatGPT as a private firm investment analysis research tool. By applying ChatGPT to various private corporate contexts of different sizes, Krause concluded that machine learning techniques can extract insights from unstructured data. However, the author asserts that AI models may struggle to fully understand the "nuances of a private company's business model or competitive landscape," given the lack of access to data and mass information from private firms. Hence, as different companies present data differently, this gap in data and the unstructured nature of the data remains highly problematic for accurate stock prediction (Krause, 2023). Thus, by asserting Krause's argument that ChatGPT *must* be used in tandem with other means of research and effort, such assertions frame ChatGPT as a purely supplementary tool in finance, a field where unstructured data and partial data are common.

Given the need to add private data, the learning curve to operate the ChatGPT interface in finance may be too intimidating to use to the AI's utmost potential. Lewis University professors Ahangar and Fietko utilize empirical examples to test the limits of ChatGPT and conclude that ChatGPT can supply financial insights only when broken down with specific sub-questions, implying the need for additional training of the ChatGPT bot to effectively identify stock picks (Ahangar & Fietko, 2023). Therefore, if a beginner investor cannot extract this data or properly input the data, ChatGPT may require an overwhelming amount of expertise in both the consumer's prompting skills and the consumer's knowledge of investor data. Moreover, if a degree of expertise in finance is required to achieve accurate and more reliable results, such ChatGPT tools will not effectively guide beginner stock market investors with minimal knowledge; hence, the ChatGPT may only be useful for financial experts.

Methods

Taking inspiration from common approaches in the field of finance and AI, an appropriate approach involves a mixed-methodology analysis that assigns a focus group to develop and maintain mock stock portfolios on MarketWatch, which is a financial news website with trading simulation features.

Basis of Study

Within the topic of ChatGPT in finance, the most common methodologies involve the application of mock portfolios, specified datasets for asset allocation, and the leveraging of ChatGPT's in-context learning in the priming phase. To rephrase, researchers in the exploration of the financial applications of ChatGPT often utilize a fake portfolio with fake money to compare with broader market metrics. Economics professors Ko and Lee at Seoul National University examined the efficacy of ChatGPT in portfolio management concerning asset allocation and diversification of selective asset classes. By gathering the performances after simulating the ChatGPT's investments over a year, Ko and Lee could compare the performance metrics of the ChatGPT selections to randomly selected assets (Ko & Lee, 2023). Additionally, in a paper performed by the Department of Informatics at Athens University of Economics, Loukas analyzes the application of GPT-3.5 in banking by training the ChatGPT bot with a real-life dataset of customer service responses, specifying the description of the task and available classes for selection (Loukas, 2023). Ergo, the author's use of ChatGPT in-context learning and machine learning techniques represent a fundamental step in creating finance-specific ChatGPT 'bots.'

Recruitment and Organization

A focus group was deemed the best approach for implementing Ko and Lee's mock portfolio concept with Loukas' restricted stock selection pool. To recruit beginner investors for the focus group, the researcher emailed the entire high school junior and senior class from a private school in Southern California. Empowered by the ease of access to the researcher's student body, the researcher chose to examine high school students, who were mostly 17-18 years old and in the latter half of their high school careers, to reflect the nature of beginner traders with limited knowledge of the financial space. Deeming individuals with less than a year of experience as 'beginner traders,' the researcher filtered the total respondents of an outreach email based on the condition that the participant had less than a year of stock market trading history and consented to the study (See Appendix "Item 2" for interest form contents). The focus group was randomly selected and assigned to varying degrees of treatment: no ChatGPT aid, partial ChatGPT aid, and total ChatGPT aid. In this context, the members with no ChatGPT involvement simulated their investments based on their efforts and best judgment, while those with partial aid consulted the ChatGPT for financial advice. Lastly, the members with total ChatGPT aid utilized ChatGPT to allocate their assets entirely and referred to ChatGPT as their financial manager.

Then, alluding to Loukas' previous studies about ChatGPT's prompting system, the researcher created standardized prompts, which were essential to allow ChatGPT to evaluate specific stocks effectively and maintain consistency across subjects. The use of the key phrases "long-term success" and "stability" to emphasize security and mitigate trading risk in the prompts naturally limited the ChatGPT's financial advice to exclusively focus on large-cap stocks from various sectors of the U.S. market (see Appendix "Item 1" for standardized prompts); the intention behind this limitation was to minimize risks for novice investors who might otherwise engage in risky, instantaneous trading. Moreover, this approach aimed to overlook the intangible advantages of professional traders by focusing on purely low-volatility assets.

Data Management

On the platform MarketWatch, the researcher hosted a “Virtual Stock Exchange Game,” where all participants joined and made daily trades in their corresponding mock portfolios. Each participant started with \$100,000, which is a generally agreed-upon initial investment number, in their mock portfolios and utilized the same interfaces for uniformity.

Moreover, the three portfolios were compared to one another and common stock market indicators, including the major market indices like SPDR S&P 500 ETF Trust and Dow Jones Industrial Average. This comparison aims to provide a baseline for the performance of the more significant market, allowing the researcher to evaluate the extent to which beginner traders beat the market.

Alternatively, to account for the nuances within each focus group, multiple rounds of discussion were employed to gather subjective feedback from participants concerning their experiences, decision-making processes, and perceptions of ChatGPT. In an effort to maximize the quality of the data, small focus group discussions of seven individuals with representation of every treatment group were held. The discussion questions aimed to uncover AI's implications on the trading experience, decision-making, and skill development for beginner traders. (see Appendix “Item 3” for content). By combining the mock portfolio system with these focus group discussions, the researcher can better evaluate the extent to which AI is an appropriate tool for beginner traders.

Results & Discussion

Performance Comparison

Only eight of the nineteen portfolios garnered a net profit over the three weeks (15 trading days), and the overall performance of the collective group indicated a minuscule increase in portfolio values across the board, featuring a mean final net worth for all portfolios of \$100,023.47 and a mean profit margin of 0.02347%. Table 1 presents a detailed net worth and total returns ranking, segmented by treatment type. After the study, the first-place portfolio and last-place portfolio, as expressed in Table 1, display that the net worths of the mock portfolios ranged from a net return of \$8,580.33 for a percent increase of 8.58% to a loss of \$7,944.12 for a percent change of -7.94%. Notably, AI-supported portfolios dominated twelve out of the top thirteen spots, outperforming all but one non-AI-supported trader. While this lone independent portfolio had the greatest total returns of all portfolios, the other five independent portfolios occupied the majority of the lowest quartile, representing the lowest returns of all portfolios.

Table 1. *Net Worth Rankings Identified by Treatment Type.* Two participants were excluded from the data due to their withdrawal from the study.

Place	Treatment Type	Net Worth	Trades	Total Returns
1	Independent	\$108,580.33	9	\$8,580.33
2	Full AI	\$106,255.79	21	\$6,255.79
3	Full AI	\$102,460.86	14	\$2,460.86
4	Partial AI	\$102,123.46	44	\$2,123.46

5	Partial AI	\$101,655.88	3	\$1,655.88
6	Partial AI	\$100,747.35	6	\$747.35
7	Full AI	\$100,662.78	12	\$662.78
8	Full AI	\$100,008.54	1	\$8.54
9	Partial AI	\$99,805.98	3	-\$194.03
10	Full AI	\$99,544.45	17	-\$455.55
11	Full AI	\$99,536.82	1	-\$463.18
12	Full AI	\$99,299.86	10	-\$700.14
13	Full AI	\$98,906.97	8	-\$1,093.03
14	Independent	\$98,013.08	12	-\$1,986.92
15	Independent	\$94,958.54	11	-\$5,041.47
16	Independent	\$94,712.48	34	-\$5,287.52
17	Independent	\$94,559.76	102	-\$5,440.24
18	Full AI	\$94,247.71	10	-\$5,752.29
19	Independent	\$92,055.88	17	-\$7,944.12

Figure 1 below visualizes the overall mean performances of each treatment group. Evident in the steadier growth in net worth of the partial AI and full AI portfolios compared to the independent mean net worth, portfolios managed with full AI assistance maintained more stability and less dramatic fluctuations in net worth, albeit with minimal overall growth.

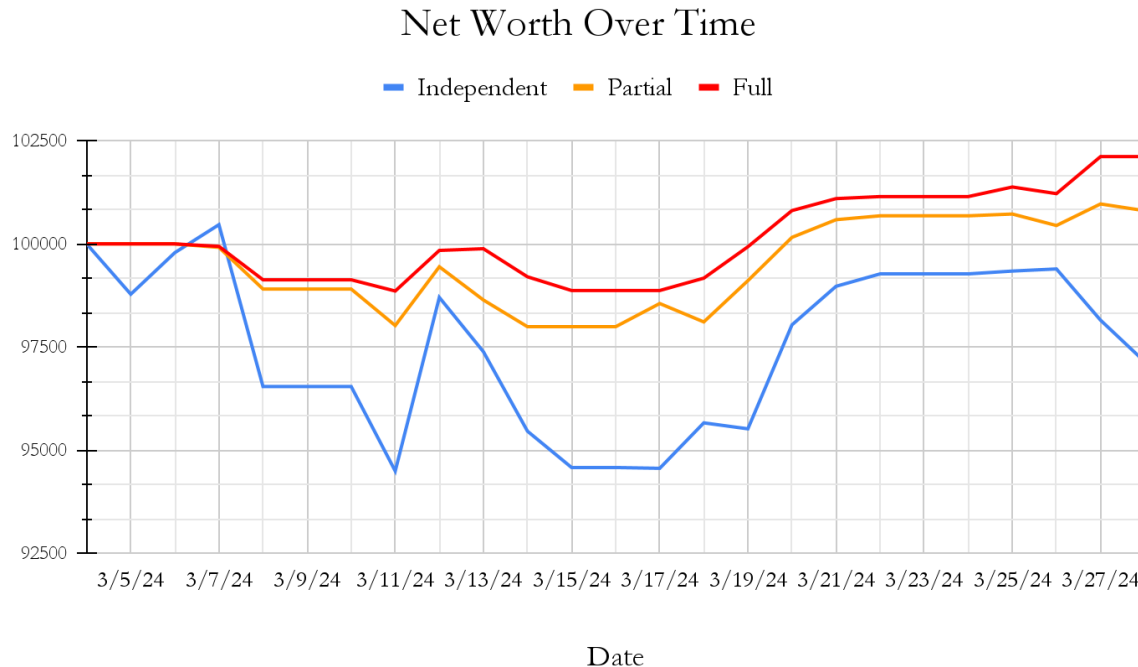


Figure 1. Mean Net Worth of Different Treatment Level Portfolios Over Time. Two participants were excluded from the data due to their withdrawal from the study.

Notably, the largest uptick for all three treatment levels occurred on March 11th, the immense growth of around 4.45% by the independent portfolios dwarfed growth rates of the partial AI-supported portfolio at 1.45% and full AI-supported portfolios at 0.99%. Such evidence in favor of the anti-AI side highlights a potential tradeoff: more significant profits can be achieved by independent traders in exchange for the long-term stability gained by partial and full AI investments.

However, the mean net worths of the AI-supported portfolios reveal the merits of long-term stability: greater long-term profit and less likelihood of net loss. As expressed in Figure 1 on the following page, the average net worth of portfolios with full AI was the greatest at \$102,117.69, while the average net worth of portfolios with partial AI support was second greatest at \$100,806.03. Alternatively, the average net worth of portfolios without AI support garnered only \$97,146.68, which denotes a 2.85332% decrease in portfolio net worth throughout the period. These averages express that the portfolios without AI support were the only treatment group with a decline in portfolio net worth through the period.

Context of the Market

Additionally, acknowledging the quantitative data in the context of the grander market, the mean net worths of the AI-supported portfolios prevail. Figure 2 displays a time-series analysis of percent change across the three-week trading simulation, including the performances of major indexes and every mean portfolio performance. Representing the three most-followed stock market indices in the United States, the S&P500 (500 of the largest companies), the Nasdaq Composite (majority of stocks on Nasdaq), and the Dow Jones Industrial Average (30 prominent companies) all create a baseline for the performance of the market due to their close ties with the overall growth rate of a nation's economy (Dempsey, 2017).

Percent Change of Indexes and Mean Portfolio Performances

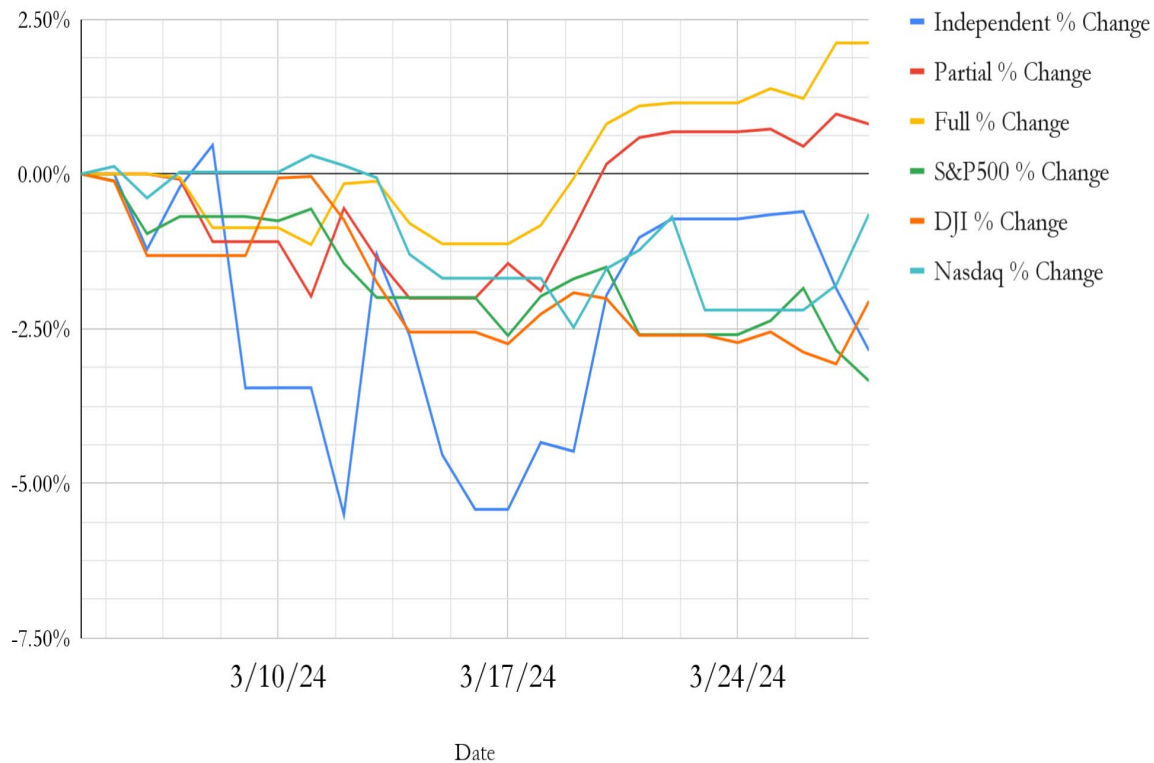


Figure 2. *Percent Change of Indexes and Mean Portfolio Performances Over Time.* The researcher generated this visualization from collected data and metrics about indices from MarketWatch historical data. Percent change was calculated based on overall percent change from each day to 3/4/24 (day one of the study).

Given the decline of the three indexes during the three weeks of simulated data, the simulated data of the AI-supported beginner traders greatly deviates from indicators of broader market performance, as partial AI-supported and full AI-supported traders produced a positive 0.81% and 2.12% overall profit respectively. On the other hand, paralleling the net losses by S&P500's percent change of -3.34% and DJI's percent change of -2.05%, the overall percent changes of the independent traders, which was deployed as the control group for the study, produced a percent change of -2.85%; hence, such evidence supports the notion that AI support can allow a beginner trader to 'beat' the market. Ultimately, this comparative graph underscores that the fully AI-managed portfolios experienced lower volatility patterns and generated larger returns than the other traders and major indices.

Thematic Analysis

With the researcher leading discussions through a series of prompts (see Appendix "Item 3" for questions), each online and in-person discussion continued for approximately 30 minutes each and was held within two weeks of closing the portfolios. The contents of these qualitative results were then categorized into common themes, and four major themes emerged: 1) improved understanding of the market; 2) sense of distrust in AI; 3) technical

issues with the AI interface; and 4) positive, perceived effectiveness of the AI-generated financial advice. Table 2 presents the proportion of participants, segmented by treatment level, with these insights.

Table 2. *Percentages of Participants Reporting Common Insights By Treatment Group.* This visualization was generated by the researcher from collected data. “N/A” was inputted due to insufficient insight.

Theme	Improved Understanding of Market	Distrust in AI	Technical Issues	Perceived Effectiveness (Positive Experience)
Full AI	62.5%	25%	25%	75%
Partial AI	100%	66%	50%	66%
Non-AI	42.9%	N/A	N/A	N/A

Regarding improved understanding of the stock market and trading, a higher proportion of participants with full AI or partial AI support reported an improved understanding of the market. Specifically, one participant with partial AI support reported improved knowledge of trading strategies: “[AI] let me know if the companies I wanted to invest in would be beneficial...consulting AI helped me choose between investing in major tech companies or the general U.S. market.” In contrast, individuals with non-AI support only reported educational value gained in the general dynamics of stock trading or reflection about the trading experience itself, netting responses such as a greater “familiarity with the trading software” and the “habit of checking the market every morning.” Thus, since all individuals with partial AI support reported more genuine learning experiences, the focus provided by the partial AI support in composition with human intuition may maximize the educational value of a mock portfolio experience, providing an alternative perspective largely unexplored in previous academic discussions and a point of future inquiry.

On the other hand, a higher proportion of partial AI participants reported distrust in AI in comparison to full AI participants; this phenomenon can be explained by their freedom to conduct individual research and only utilize the AI as a consultant, not as a driver of their portfolio. To elaborate, those with partial assistance expressed reservations about the reliability of the AI’s advice; one participant with partial aid noted, “I never felt safe with what it was saying. I had to make the AI play a character, so the trust in its advice was always in question.” Alluding to the stigma around ChatGPT and consumer AI, the participant’s sense of insecurity could potentially induce underutilization of AI capabilities. In contrast, the majority of participants with full assistance expressed minimal distrust. One from the Full AI group stated, “I didn’t really know what I was doing, but the AI gave me a recommendation based on its own mathematical models. I feel confident because AI articulates...mathematical graphing trends better than many individuals.” Evident in this participant’s blind faith, excessive reliance on AI may also inhibit the user’s financial literacy. Thus, finding this balance between the underutilization of AI advice and over-reliance on AI is an area for future research.

Individuals with partial assistance reported difficulties in extracting clear guidance from ChatGPT due to the “vagueness of its advice” and “confusing wording.” Moreover, two AI users’ specific feedback highlighted that ChatGPT’s legal content warning system completely blocked their access to specific or useful financial advice. Such notions underscore the importance of recognizing the technological expertise required to overcome the learning curve of ChatGPT; hence, this feedback questions the practicality of ChatGPT for all beginner traders.

Portfolio Composition and Perceived Effectiveness

Additionally, perceived effectiveness also exhibited variability in participant experiences, which manifested in the composition of the portfolios. Most participants, with full assistance, highlighted ChatGPT's ability to select a diverse and stable portfolio. For instance, an individual with full assistance commented, "It selected good stocks for me to invest in and gave me a diverse portfolio. I thought the AI proposed a well-informed and safe trading strategy." This principle of a "diverse portfolio" is manifested in the composition of the portfolio: every portfolio completely supported by ChatGPT evenly distributed its funds across investments and diversified into at least four different industries.

Conversely, participants with partial assistance felt that ChatGPT did not significantly contribute to their trading success: "AI didn't really help me find stocks...I used my previous knowledge to think of good companies and used AI as a final checkbox." Since the portfolio composition was dominated by commonly traded companies in the news, the diversification of these portfolios was significantly poorer, often having one or two large technology companies with over 40% portfolio holdings.

Similarly, independent traders naturally invested in mainstream stocks and had poor diversification of their assets; however, independent traders also invested in riskier assets. Independent traders often chose stocks based on their prevalence in the media; for instance, one independent trader commented, "I invested in Tesla and TSMC because they were on the news a lot." Notably, however, the independent traders invested in more volatile stocks without historic success; one independent trader explained, "I traded based on the philosophy, 'no guts, no glory.'" Hence, independent traders generally were more willing to take risks for greater profit.

Cases of Interest

Two cases of interest represent the extremes of AI implementation and merit in-depth analysis: the most successful independent trader and the least successful full-AI-supported trader. Although these case studies do not imply concrete conclusions, they provide pertinent insights for further investigation.

In order to achieve the highest profit of all portfolios, the most successful independent trader engaged in various high-risk investments with little diversification. Particularly, the independent trader's most notable profit was in LUNR, a space exploration company with a high upside but no previous historical success. As the trader describes, the trader's trading strategy implemented many "high-risk, high-reward" investments; LUNR later rose by 29.93% to achieve a profit of \$6,219.70 over the study period. In conversation with the AI-supported portfolios with much more stable investments, the successful independent trader's performance highlights a dimension of the market that the ChatGPT AI fails to recognize: humans' emotional influences on the market. Aligning with Fletcher's concern about the inability of AI to practice high emotional intelligence, AI's need for granular instruction and algorithmic development to make stock selections leads to a preference for broader ETFs and more mainstream companies with historic success, limiting the potential for astonishing profit. In contrast, the independent has a higher willingness to take risks despite its higher volatility and poorer odds of success. Therefore, while AI can select traditionally safer investments for beginner investors, AI may lack the capacity to fully capitalize on occasional market opportunities that require a nuanced understanding of human emotion and behavior.

Alternatively, the least successful AI trader, who gathered a net loss of -\$5,752.29, accredited their failure to an error made by the ChatGPT AI that misidentified a stock. The trader explained that the AI described Nvidia, a computing technology company, as a "stable and secure investment." The trader invested a significant portion of his cash into Nvidia before Nvidia's stock price dropped immensely. In fact, according to MarketWatch finance writer Philip Doorn, Nvidia should be identified as a growth stock in the S&P500, observing Nvidia as a top-performing stock but with lofty expectations and a higher likelihood of considerable volatility (Doorn, 2024). Thus, because the AI erroneously identified Nvidia, the considerable loss of this trader, which

was 2% lower than the S&P500, exemplifies AI's failure to interpret complex market dynamics that a human investor might understand intuitively, affirming Kocóń's concerns about the accuracy of ChatGPT's outputs. These results indicate that while ChatGPT can provide valuable data and analysis, its inaccuracies make it unreliable as a standalone decision-making tool without human oversight and contextual understanding.

Implications

These qualitative findings underscore a complex relationship between user experience and the perceived reliability of AI in financial decision-making. Participants who fully integrated ChatGPT assistance tended to appreciate its systematic approach, reporting enhanced trust and perceived effectiveness. In stark contrast, those with partial assistance faced usability challenges and relied significantly on their own knowledge rather than the AI's outputs. This disparity suggests that while AI can provide substantial support in data processing and initial decision-making, its efficacy may still necessitate considerable human oversight, particularly for users with limited technological acumen or those requiring tailored financial advice.

While the qualitative data suggests some correlation between ChatGPT and perceived effectiveness and educational value, the quantitative results suggest that ChatGPT greatly improves the outcomes for novice traders in a simulated stock trading environment. This finding aligns with the optimism noted in the literature regarding AI's ability to successfully create data-based trading strategies despite occasional inaccuracies.

Combining AI's consultation of data-rich stocks with the user's personal risk tolerance, such notions favor a hybrid decision-making model for increasing the likelihood of 'beating' the market. Given the preference for AI to suggest data-rich stocks with more prevalence in the news and longer legacies, AI's trading options are greatly reduced due to the dependence on this data for identifying a successful company. To address the data gap identified by Ahangar and Fietko in their study on ChatGPT prompting in finance, human oversight would naturally expand the range of stock selection to maximize opportunities in more emotionally driven moments in the market.

In addition, the experimental data asserts ChatGPT's capacity as an educational or foundational tool for beginner stock traders, as interaction with the ChatGPT conversational AI greatly aided the 'financial education' aspect of their mock trading experience. While such implementation of AI does not directly apply the ChatGPT bot to the real market, the activity of trading mock portfolios with AI assistance has educational merit. Due to the ability of ChatGPT to build strong foundational knowledge in stock trading, ChatGPT is a powerful tool for beginner traders to clarify doubts, explain concepts, and provide real-time guidance; thus, ChatGPT has applications as an instructional tool for self-learners and educational institutions.

Although AI accurately predicted the success of data-rich companies and struggled with high-volatility stocks like low-cap and growth stocks, such assertions highlight a need for further experimental data to confirm this hypothesis. Ergo, the notion that collaboration between AI and human intuition may, albeit with limitations, provide a starting point for empowering beginner traders to outperform the market's conditions, mitigating their losses and enhancing their wins.

Limitations

The methodology employed in this study has limitations that merit consideration. Primarily, the small sample size and the demographic characteristics of the focus group participants, predominantly high school juniors and seniors, introduce a potential bias that may not accurately reflect the broader population of novice investors.

Additionally, the study's duration of three weeks was relatively short, limiting the ability to observe the long-term efficacy of ChatGPT assistance in portfolio management.

Another significant limitation is the use of a mock portfolio rather than real investment scenarios. This artificial setting may not fully capture the emotional and psychological factors influencing investment decisions in a real-market environment, as the artificial setting may greatly inflate the risk tolerance of non-AI investors, skewing the outcomes.

Furthermore, due to the standardized prompting design, the scope of ChatGPT's advice was also confined to mostly large-cap stocks, which does not provide a comprehensive view of the market's opportunities. This restriction, while minimizing exposure to volatile assets for inexperienced investors, also limits the insights into AI's adaptive capabilities and broader applicative potential in financial decision-making.

Future Research

The findings from this research present a hopeful foundation for integrating AI tools like ChatGPT into the field of stock trading, particularly for novice investors. Despite previous concerns surrounding the inaccuracies of AI and its inability to address all psychological factors in the market, the AI portfolios were able to 'beat' the market in practice, presenting some evidence of AI's ability to democratize the market. Nonetheless, to truly validate this notion, future studies should focus on expanding the magnitude and the demographic scope of research participants, including more experienced traders and investors of different ages.

Additionally, evaluating 'ChatGPT literacy'—or the degree to which familiarity with navigating and utilizing AI tools affects trading success—could provide deeper insights into the educational and practical prerequisites for maximizing AI's potential in trading. Moreover, as AI technology evolves, its capability to integrate and interpret web-wide data in real-time will likely enhance its utility in domain-specific fields beyond novice trading. Future research should explore the deployment of new AI models that can navigate and utilize up-to-date market data and news, potentially overcoming current limitations and legal barriers associated with automated stock picking.

By replicating this study with these expanded parameters and incorporating AI systems that can access and analyze a broader array of real-time data, researchers can better understand the conditions under which AI tools most effectively support traders. By bridging the gap between novice and expert traders while paving the way for the next generation of AI-enhanced financial decision-making tools, future studies can affirm the hybrid-thinking model as a possible means of social mobility.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

- Dempsey, M. (2017). A short history of stock markets. *Stock Markets and Corporate Finance*, 13–65. https://doi.org/10.1142/9781786343277_0002
- Philip van Doorn. (2024, April 15). *Nvidia tops updated S&P 500 growth stock screen. Here's the rest of the list.* MarketWatch; MarketWatch. <https://www.marketwatch.com/story/nvidia-is-the-no-1-s-p-500-growth-stock-in-this-screen-see-which-other-companies-joined-or-left-the-list-ab455826?mod=philip-van-doorn>
- Dowling, M., & Lucey, B. M. (2023). ChatGPT for (Finance) research: The Bananarama Conjecture. *Finance Research Letters*, 53, 103662–103662. <https://doi.org/10.1016/j.frl.2023.103662>

- Fletcher, A. (2021). Why Computers Will Never Read (or Write) Literature: A Logical Proof and A Narrative. *Narrative*, 29(1), 1–28. <https://doi.org/10.1353/nar.2021.0000>
- Gilli, A., Gilli, M., Leonard, A.-S., & Stanley-Lockman, Z. (2020). Understanding the revolution: artificial intelligence, machine learning and big data. In “*NATO-Mation*”: *Strategies for Leading in the Age of Artificial Intelligence* (pp. 17–24). NATO Defense College. <http://www.jstor.org/stable/resrep27711.9>
- Haenlein, M., & Kaplan, A. (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review*, 61(4), 5-14. <https://doi.org/10.1177/0008125619864925>
- Johnson, D., Goodman, R., Patrinely, J., Stone, C., Zimmerman, E., Donald, R., Chang, S., Berkowitz, S., Finn, A., Jahangir, E., Scoville, E., Reese, T., Friedman, D., Bastarache, J., van der Heijden, Y., Wright, J., Carter, N., Alexander, M., Choe, J., Chastain, C., ... Wheless, L. (2023). Assessing the Accuracy and Reliability of AI-Generated Medical Responses: An Evaluation of the Chat-GPT Model. *Research square*, rs.3.rs-2566942. <https://doi.org/10.21203/rs.3.rs-2566942/v1>
- Knutsson, K. (2023, June 12). *Would you trust AI to help you invest your money and manage your portfolios?* Fox News; Fox News. <https://www.foxnews.com/tech/trust-ai-invest-money-manage-portfolios>
- Kocoń, J., Igor Cichecki, Oliwier Kaszyca, Kochanek, M., Dominika Szydło, Baran, J., Julita Bieleniewicz, Marcin Gruza, Janz, A., Kamil Kanclerz, A. Kocoń, Bartłomiej Koptyra, Wiktoria Mieleszczenko-Kowszewicz, Piotr Miłkowski, Oleksy, M., Piasecki, M., Łukasz Radliński, Wojtasik, K., S. Woźniak, & Przemysław Kazienko. (2023). Chat GPT: Jack of all trades, master of none. *Information Fusion*, 99, 101861–101861. <https://doi.org/10.1016/j.inffus.2023.101861>
- Ko, Hyungjin and Lee, Jaewook, Can Chatgpt Improve Investment Decision? From a Portfolio Management Perspective. Available at SSRN: <https://ssrn.com/abstract=4390529> or <http://dx.doi.org/10.2139/ssrn.4390529>
- Krause, D. (2023). ChatGPT and Other AI Models as a Due Diligence Tool: Benefits and Limitations for Private Firm Investment Analysis. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4416159>
- Lin, Tom C. W., The New Financial Industry (March 30, 2014). 65 *Alabama Law Review* 567 (2014), Temple University Legal Studies Research Paper No. 2014-11, Available at SSRN: <https://ssrn.com/abstract=2417988>
- Loukas, L., Ilias Stogiannidis, Prodromos Malakasiotis, & Vassos, S. (2023, August 28). *Breaking the Bank with ChatGPT: Few-Shot Text Classification for Finance*. ResearchGate; unknown. https://www.researchgate.net/publication/373451297_Breaking_the_Bank_with_ChateGPT_Few-Shot_Text_Classification_for_Finance
- MarketWatch*. (2024, April 23). MarketWatch. <https://www.marketwatch.com/>
- Ren, C., Lee, S.-J., & Hu, C. (2023). Assessing the efficacy of ChatGPT in addressing Chinese

financial conundrums: An in-depth comparative analysis of human and AI-generated responses. *Computers in Human Behavior Artificial Humans*, 1(2), 100007–100007. <https://doi.org/10.1016/j.chbah.2023.100007>