Robo-Advising Meets Large Language Models: Educating Investors on Alpha and Beta of Mutual Funds and Stocks *

Fangzhou Lu
University of Hong Kong

March 17, 2025

Abstract

We utilize a robo-advising setup to investigate the impact of financial literacy on investors, specifically focusing on the concepts of beta and alpha. Partnering with a large brokerage firm, we study the effects of educating investors about beta and alpha in mutual funds and stocks. To enhance investors' understanding, we integrate a back-end LLM (ChatGPT/DeepSeek) system with the robo-advisor. Our findings reveal that after learning about beta, investors show a preference for investing in ETFs or index funds rather than directly in active mutual funds or individual stocks. A one-click automatic enrollment mechanism has a much larger impact than a self-assembled portfolio, but the gap is much smaller for LLM-supported investors. This suggests that an LLM-based chatbot exerts a long-term improvement on financial literacy. LLM-based chatbots also generate a spillover effect, prompting investors to pursue international diversification even when these options are not initially recommended by the robo-advisors. Overall, investors who receive this financial education benefit significantly, with their monthly returns increasing by 40 basis points and an average monthly gain of 750 RMB.

JEL Codes: D1, E62, H20

Keywords: Financial literacy, Robo-advising, LLMs, ChatGPT, Factor investment

*Many thanks to my dissertation committee members Jennifer Carpenter, Jonathan Parker, Antoinette Schoar, and Robert Whitelaw for their guidance and advice, and to Daniel Greenwald, Kose John, Maarten Meeuwis, Christopher Palmer, Adrien Verdelhan, Haoxiang Zhu and seminar participants at MIT for helpful comments and suggestions.

Robo-Advising Meets Large Language Models: Educating Investors on Alpha and Beta of Mutual Funds and Stocks

Abstract

We utilize a robo-advising setup to investigate the impact of financial literacy on investors, specifically focusing on the concepts of beta and alpha. Partnering with a large brokerage firm, we study the effects of educating investors about beta and alpha in mutual funds and stocks. To enhance investors' understanding, we integrate a back-end LLM (ChatGPT/DeepSeek) system with the robo-advisor. Our findings reveal that after learning about beta, investors show a preference for investing in ETFs or index funds rather than directly in active mutual funds or individual stocks. A one-click automatic enrollment mechanism has a much larger impact than a self-assembled portfolio, but the gap is much smaller for LLM-supported investors. This suggests that an LLM-based chatbot exerts a long-term improvement on financial literacy. LLM-based chatbots also generate a spillover effect, prompting investors to pursue international diversification even when these options are not initially recommended by the robo-advisors. Overall, investors who receive this financial education benefit significantly, with their monthly returns increasing by 40 basis points and an average monthly gain of 750 RMB.

JEL Codes: D1, E62, H20

Keywords: Financial literacy, Robo-advising, LLMs, ChatGPT, Factor investment

1 Introduction

Financial literacy has long been recognized as a critical component of effective financial decision-making. As the complexity of financial markets and investment products continues to increase, the importance of educating investors about fundamental financial concepts has never been more pronounced (Campbell, 2006). In particular, understanding metrics such as beta and alpha is essential for evaluating investment risk and performance. Despite the recognized importance, there remains a significant gap in financial literacy among retail investors, which can lead to suboptimal investment choices and poorer financial outcomes. Recently, generative AI combined with machine learning has offered a new perspective to help potential portfolio managers determine the optimal portfolio, as big data fused with AI has played an increasingly significant role in the financial world (Goldstein, Spatt, and Ye, 2021, 2024). It can also potentially serve as a financial advisor for retail investors, aiding in overcoming financial literacy challenges in investing and other financial planning issues (Cong et al., 2019; D'Acunto and Rossi, 2021, 2023; D'Acunto and Rossi, 2023; Sheng et al., 2024).[1]

In this study, we leverage a robo-advising setup to explore the impact of financial literacy on investor behavior and performance. Specifically, we focus on educating investors about beta and alpha within the context of mutual funds and stocks. By partnering with a large brokerage firm in China, we are able to analyze the effects of targeted financial education on a substantial and diverse investor base. To enhance the educational experience, we integrate a back-end ChatGPT system with the robo-advisor, allowing investors to engage interactively with the educational content.[2]

To conduct the analysis, the robo-advisor systematically evaluates the investor's portfolio by calculating key metrics such as alpha and beta for each holding. It employs a multi-

[1] Also see Chak et al. (2022); Gargano and Rossi (2024).

[2] The ChatGPT chatbot was complemented by three other China-based LLMs, DeepSeek, Kimi, and Doubao, due to restrictions on accessing ChatGPT in mainland China. To enable the firm to use ChatGPT, a combination of an enterprise subscription, VPN, and a server in Singapore was employed. Consequently, DeepSeek served as a backup system during this process.

factor analysis, incorporating a comprehensive set of factors like size, value, and momentum, and computes statistical measures such as t-statistics to determine the significance of the alpha. This helps distinguish genuine manager skill from random fluctuations or factor loadings. The robo-advisor then aggregates these metrics to provide an overall portfolio assessment, computing weighted average alpha and beta, and benchmarking them against market indices to contextualize performance and volatility. Based on this analysis, the robo-advisor generates personalized recommendations aimed at optimizing the portfolio. These recommendations may include reallocating funds from underperforming assets, maintaining positions in high-performing assets, and introducing low-cost, diversified investment vehicles like ETFs or index funds. This data-driven approach helps investors achieve better returns, manage risk effectively, and minimize costs, thereby aligning their investment decisions with their financial goals and risk tolerance.

Specifically, to investigate the impact of financial literacy on investor behavior and performance, we aid this brokerage firm to conduct a Randomized Controlled Trial (RCT) involving 28,819 investors who expressed interest in participating. We randomly selected 14,409 of these investors to form the treatment group, while the remaining 14,410 investors constituted the control group. Both groups were similar in terms of investment portfolio size and personal characteristics, ensuring a balanced sample and similar pre-trends.

Our analysis yields several key insights into how financial education influences investor behavior. We find that after learning about beta, investors show a marked preference for investing in ETFs or index funds rather than directly in active mutual funds or individual stocks. This shift suggests a growing appreciation for the benefits of diversification and the cost efficiencies associated with passive investment strategies. Additionally, a small portion frequent traders who receive this education tend to construct their own factor portfolios, indicating a more sophisticated approach to portfolio management.

Most notably, our findings demonstrate that financial education has a significant positive impact on investment performance. Investors who receive education on beta and alpha

experience an increase in their monthly returns by 40 basis points, alongside an average monthly gain of 750 RMB. These results highlight the substantial financial benefits that can be realized through improved financial literacy. Moreover, this part of the results is largely consistent with the findings in D'Acunto et al. (2019) and Rossi and Utkus (2024), but significantly larger. This is partially due to the high volatility in the Chinese stock market and also partially due to the new LLM-based chatbot, which we will discuss in the following paragraph.

Furthermore, In the treatment group, investors received targeted financial education on the concepts of beta and alpha through our robo-advising platform, which was enhanced with various interactive learning tools. The treatment group was further divided into four subgroups to examine the heterogeneity of the intervention's effects. The first group is the LLM (ChatGPT/DeepSeek/Kimi/Doubao) Chatbot treatment group. This subgroup received financial education and personalized advice through a large language model (LLM)-powered chatbot, which provided detailed explanations and engaged in natural language interactions. The second is the without Chatbot group. This subgroup received financial education through static content of portfolio recommendation without interactive chatbot support. The third group is the Human Chat Assistant group. This subgroup had access to human financial advisors who provided real-time assistance and explanations via chat. The final group is the Normal Chatbot group. This subgroup used a standard rule-based chatbot that provided predefined responses and guidance based on a fixed set of rules.

Our findings indicated significant heterogeneity in the results across the different subgroups within the treatment group. Investors in the LLM Chatbot subgroup exhibited the highest advice-following rate and the most substantial improvements in investment outcomes. The interactive and natural language capabilities of the LLM-powered chatbot appeared to enhance investor engagement and comprehension, leading to better adherence to the robo-advisor's recommendations and superior portfolio performance. We conduct post usage survey and summarize the most frequently asked questions by investors to the Human/Robo

advisors and their responses. The most common reason for human assistants to be ineffective is the slow response time and distrust.[3] The most common reason for traditional robot assistants to be ineffective is the inaccurate response material that are "mechanical". This result sheds lights on the scenario when robo-advising is most effective.

We also investigate whether financial education has a lasting impact on investors' ability to manage their portfolios independently, particularly focusing on their understanding of alpha and beta. Through the randomized controlled trial, investors were further assigned either a one-click switch option to a recommended robo-advisor portfolio or a DIY (do-it-yourself) approach. The primary goal was to assess if investors could replicate the performance of the one-click portfolios in their DIY portfolios, indicating a solid grasp of alpha and beta. Our findings reveal that the LLM (ChatGPT/DeepSeek/Kimi/Doubao) chatbot significantly enhances investors' understanding and performance, narrowing the gap between one-click and DIY portfolios. While initial financial education is beneficial, the effectiveness of follow-up assistance, particularly from the ChatGPT chatbot, plays a crucial role in helping investors apply their knowledge effectively. This suggests that advanced AI-driven advisory tools can empower investors with the skills needed for long-term financial independence and literacy.

In addition to the primary findings, we observed significant heterogeneity in the effectiveness of financial literacy interventions across different investor characteristics. Younger investors, who are typically more tech-savvy, exhibited higher engagement and better investment outcomes when interacting with the robo-advisor, as evidenced by their increased advice-following rates and superior portfolio performance. Moreover, investors with larger portfolios demonstrated greater benefits from the detailed financial education, leveraging advanced strategies to enhance their investment returns more effectively. These variations underscore the importance of tailoring financial literacy programs to address the specific

[3]Greig et al. (2022) show that a combination of human and robo-advisors is key to gaining investor trust. Our results, however, differ, potentially due to the persistent insider trading and fraud in the Chinese capital market specifically, rather than a general ineffectiveness of human advisors in promoting trust in machines overall. Additionally, the younger generation in China is particularly tech-savvy, which may also contribute to our findings.

needs and preferences of different investor segments, thereby optimizing the impact of robo-advisory services.

One interesting puzzle we find is that investors in the treatment group also began diversifying their portfolios to include S&P 500 ETFs and other foreign ETFs, despite the robo-advisors initially recommending only reallocations within the Chinese A-share market ETFs. Further analysis shows significant increases in allocations to S&P500, US Factor Style Investment, and Chinese ADRs, suggesting that the robo-advising service effectively promotes diversification. However, this unexpected foreign diversification appears to be driven by LLM-based chatbot assistance, as this effect is only evident in the treatment subgroup with LLM-based chatbot. This indicates the potential of AI-driven assistance to extend and amplify traditional robo-advisor recommendations, leading to more diversified investment portfolios.

This study contributes to the literature on financial education by providing empirical evidence of its impact on investor behavior and performance. Our findings underscore the potential of integrating advanced technologies, such as AI-driven educational tools, within robo-advisory platforms to enhance investor learning. By equipping investors with a deeper understanding of key financial concepts, robo-advising can help them make more informed investment decisions and achieve better financial outcomes.

Our research is related to the literature on robo-advising as ChatGPT/DeepSeek/Doubao/Kimi or a modified version of the LLM model can potentially serve as robot advisors. D'Acunto et al. (2019) show that robot-advising can increase investors' portfolio diversification. Rossi and Utkus (2024) shows that Vanguard's robo-advisors leads to diversification. Our results are consistent with the previous results, but further find that robo-advisors with LLM tools can potentially help retail investors understand complicated financial concepts and choose better portfolios than robo-advisors combined with human advisors or standard chatbot. In our research, we also highlight the importance understanding beta and alpha, which was not previous studied.

This literature also contribute to the literature of mutual fund and factor investment. Lewellen and Lewellen (2022) show that when fund managers' portfolio firms stock price appreciates, there is significant and quantifiable fund inflow. Lewellen (2011) shows that institutions as a whole closely mimic the market portfolio, with pre-cost returns that have nearly perfect correlation with the value-weighted index and an insignificant CAPM alpha of 0.08% quarterly. Sialm and Tham (2016) show that firms brand affect mutual fund inflow and outflow in an irrational way. Sialm and Xu (2024) also use a China setting and study mutual fund's information acquisition process during trading suspension. Liberman et al. (2020) show that there is tax benefit of separating alpha from beta.

Our research is related to the emerging literature on LLM and AIGC more generally. Sheng et al. (2024) show that ChatGPT can help hedge fund managers increase their returns and outperform the ones that don't use ChatGPT. Cong et al. (2024) develop a general approach to analyze large scale text based data using neural network language processing and generative statistical models. Our paper show that LLMs can arm retail investors to understand complicated financial concepts and have lasting impact on investors' behavior.

Last but not least, the importance of financial literacy in shaping investor behavior and improving financial outcomes has been widely documented in the literature. Campbell (2006) shows that financial literacy plays a major role in shaping household financial outcomes in general. More specifically, Calvet et al. (2009) show that disposition effect is the most dominant behavioral biases when trading stocks, when investors take the fight or flight decisions based on mutual fund performance. We show that after investors learn regarding the difference between alpha and beta, they show less disposition effect and take the "fight" decision to reallocate their fund based on factor-adjusted alpha. Van Rooij et al. (2011) show that household typically have little knowledge of basic finance concepts beyond time value of money and inflation. We show that by leveraging technological advancements such as LLM models, robo-advisors can provide tailored financial education, helping investors to understand key financial concepts and make more informed decisions.

The remainder of the paper is organized as follows. Section II reviews the relevant of beta and alpha for investment. Section III describes the data and methodology used in our study. Section IV presents our empirical results. Section V concludes.

2 Financial Literacy: Beta and Alpha

Educating investors about the fundamental concepts of beta and alpha is crucial because it helps them discern the true sources of their investment returns. Beta measures a security's sensitivity to market movements and various factors, indicating how much of the investment's performance is due to general market risk and factor risks. In contrast, alpha represents the excess return of an investment relative to a benchmark, reflecting the portion of performance attributed to the mutual fund manager or investor's skill or strategy. By understanding these concepts, investors can more accurately assess whether their returns are due to market movements (beta) or the mutual fund managers and their own investment acumen (alpha), particularly when evaluating their fund managers' ability when purchasing mutual funds or their own skill when buying individual stocks.

This distinction is vital for improving financial literacy and investor performance. Often, investors may attribute their gains to personal skill when, in reality, these returns are primarily driven by market factors. By understanding that beta explains a significant portion of their returns, investors are less likely to overestimate their abilities and take undue risks. This awareness directly addresses overconfidence biases, particularly when it comes to stock investments. Recognizing that market exposure, rather than individual stock-picking prowess, drives their returns can lead to more prudent investment decisions, such as preferring diversified, low-cost ETFs or index funds that capture market returns efficiently.

In addition to addressing overconfidence biases, educating investors about beta and alpha also helps them correctly evaluate mutual fund managers. Investors often struggle to distinguish between a manager's genuine skill and performance driven by market movements. By

understanding alpha, investors can better assess whether a mutual fund manager is delivering true value beyond what can be attributed to market exposure. This knowledge empowers investors to make more informed decisions when selecting mutual funds, ensuring they are not overpaying for active management that does not deliver corresponding alpha. In essence, it enables investors to accurately evaluate their fund managers' ability when purchasing mutual funds or their own investment acumen (alpha) when it comes to buying their own stocks.

In comparison to typical robo-advisors that generate portfolio that rely on making tailor made suggestions based investors' age and risk preference, educating investors about beta and alpha might have a more profound and lasting impact. Behavioral biases often lead to suboptimal trading behaviors and require continuous effort to correct. However, by understanding beta and alpha, investors gain a foundational framework that inherently guides them toward more rational and data-driven decisions. This knowledge helps them recognize that outperforming the market consistently is challenging and often not due to skill but to exposure to market factors.

3 Institutional Details

3.1 Brokerage Firms

The brokerage firm we partner with has extensive market coverage and a strong reputation within the financial industry. It is one of the largest and most established brokerage firms in China, with over 200 branches spread across the country. This extensive network ensures that our study could reach a diverse and broad spectrum of investors, providing a solid foundation for our research on the impact of financial literacy.

The firm manages assets worth over 1 trillion RMB and serving more than 10 million clients, it stands as one of the top brokerage firms in China. This significant market share and large client base are essential for obtaining a representative sample for our study. The broad reach allows us to analyze the behaviors and outcomes of a varied group of investors,

enhancing the generalizability and validity of our findings. Additionally, the firm's strong reputation and credibility in the financial industry lend weight to our study, ensuring that our research is conducted within a reliable and trustworthy framework.

Furthermore, the firm's commitment to innovation and technology aligns well with our research objectives. By integrating a back-end LLM (ChatGPT/DeepSeek/Kimi/Doubao) system with their robo-advisor, we were able to provide an advanced and interactive educational experience for investors. This firm's willingness to adopt and implement cutting-edge technology demonstrated their dedication to improving client services and financial literacy. Their support and collaboration were instrumental in facilitating the seamless integration of our educational tools, thereby enhancing the overall effectiveness of our study.

3.2 RCT designs

To investigate the impact of financial literacy on investor behavior and performance, we conducted a Randomized Controlled Trial (RCT) involving 28,819 investors who expressed interest in participating. We randomly selected 14,000 of these investors to form the treatment group, while the remaining 14,819 investors constituted the control group. Both groups were similar in terms of investment portfolio size and personal characteristics, ensuring a balanced and fair comparison. The pilot program of the experiment is conducted since the beginning of December 2022, and went into full scale since December 2023 and ends on the end of November 2024.[4]

In the treatment group, investors received targeted financial education on the concepts of beta and alpha through our robo-advising platform. The education module provided

[4]For the randomization process, The brokerage firm use the Stata code automatically to re-randomized as follows. They first allocated small businesses to the treatment and control groups on the basis of a randomly generated number. Using this allocation, they then calculated the maximum and the average t-statistics on the differences in averages across the treatment and control groups for the following variables: investors' age, gender, years of schooling, total asset value, P/L in terms of percentage in the past year, If the maximum t-statistic for these variables was higher than 1.25 or the average t-statistic was higher than 0.35, we drew a new random number and allocated firms to the treatment and control groups on the basis of this new number. We repeated this process until the maximum t-statistic was 1.25 or lower and the average t-statistic was 0.35 or lower.

detailed explanations, examples, and personalized insights into how these metrics influence investment risk and performance. Investors in the control group continued to receive standard robo-advisory services without the additional financial literacy component.[5] Moreover, a subsample of the treatment group, was further enhanced with a back-end LLM system or other follow-up assistance for interactive learning.

Throughout the study period, we closely monitored the investment decisions and performance outcomes of both groups. The robo-advisor in the treatment group not only provided the usual portfolio analysis and recommendations but also incorporated educational content to help investors understand the rationale behind each suggestion. For instance, if an investor in the treatment group held mutual funds with varied alpha and beta values, the robo-advisor would explain the implications of these metrics and recommend adjustments based on their statistical significance and performance potential.

By comparing the outcomes of the treatment and control groups, we aimed to isolate the effect of financial literacy on investment behavior. Specifically, we measured changes in portfolio composition, trading activity, and overall investment performance. This RCT design allowed us to rigorously evaluate the causal impact of financial literacy on investor behavior and performance, providing robust evidence to support the integration of educational tools within robo-advisory platforms. The insights gained from this study highlight the potential benefits of enhancing financial literacy to help investors make more informed decisions and achieve superior financial outcomes.

To conduct the analysis, the robo-advisor begins by evaluating the investor's current portfolio holdings. It systematically assesses each mutual fund or stock within the portfolio by calculating key metrics such as alpha and beta, which are essential for understanding the performance and risk characteristics relative to the market. The robo-advisor utilizes a comprehensive set of factors, often numbering in the hundreds, to perform a multi-factor analysis. This analysis includes commonly used factors such as size (SMB), value (HML), and

[5]In the Appendix, we show how exactly our robo-advisor function to help investor understand the difference between alpha and beta.

momentum (MOM), among others. It also computes statistical measures, such as t-statistics, to determine the significance of the alpha, thereby distinguishing genuine manager skill from random fluctuations.

Once the individual metrics for each holding are calculated, the robo-advisor aggregates these metrics to provide an overall assessment of the portfolio. This involves computing an overall average alpha and beta for the entire portfolio, taking into account the proportion of each asset within the portfolio. The robo-advisor then benchmarks these aggregated metrics against market indices to provide context on the portfolio's performance and volatility.

Based on this detailed analysis, the robo-advisor generates personalized recommendations for the investor. These recommendations are aimed at optimizing the portfolio by suggesting actions such as reallocating funds from underperforming assets, maintaining positions in high-performing assets, and introducing low-cost, diversified investment vehicles like ETFs or index funds. The ultimate goal is to help the investor achieve better returns, manage risk more effectively, and minimize investment costs. Through this methodical and data-driven approach, the robo-advisor empowers investors to make informed decisions that align with their financial goals and risk tolerance.

3.3 LLMs and Robo-advisors

In the treatment group, investors received targeted financial education on the concepts of beta and alpha through our robo-advising platform, which was enhanced with various interactive learning tools. The treatment group was further divided into four subgroups to examine the heterogeneity of the intervention's effects:

ChatGPT/DeepSeek (LLM) Chatbot: This subgroup received financial education and personalized advice through a large language model (LLM)-powered chatbot, which provided detailed explanations and engaged in natural language interactions.[6] Without Chat-

[6]The ChatGPT chatbot was complemented by three other China-based LLM, DeepSeek, Kimi, and Doubao, due to restrictions on accessing ChatGPT in mainland China. To enable the firm to use ChatGPT, a combination of an enterprise subscription, VPN, and a server in Singapore was employed. Consequently,

bot: This subgroup received financial education through static content such as articles and videos, without interactive chatbot support. Human Chat Assistant: This subgroup had access to human financial advisors who provided real-time assistance and explanations via chat. Normal Chatbot: This subgroup used a standard rule-based chatbot that provided predefined responses and guidance based on a fixed set of rules. We also further divide each subgroup into two groups. The first group receive the benefit of automatic execution or one-click execution. The second group can only execute the recommended portfolio by doing it by themselves (DIY). This setup helps us to distinguish whether investors actually understand the logic behind the recommended portfolios. Throughout the study period, we closely monitored the investment decisions and performance outcomes of all subgroups in the treatment group, as well as the control group.

By comparing the outcomes of the treatment and control groups, as well as the different subgroups within the treatment group, we aimed to isolate the effect of financial literacy and the mode of delivery on investment behavior. Specifically, we measured changes in portfolio composition, trading activity, and overall investment performance. In conclusion, this RCT design allowed us to rigorously evaluate the causal impact of financial literacy on investor behavior and performance.

3.4 Summary Statistics

Table 1 presents the summary statistics for the characteristics of the investors and their portfolios, comparing the treatment and control groups. For each characteristic, the table reports the mean and standard deviation (in parentheses), along with the calculated difference between the treatment and control groups. The differences observed are not statistically significant. The average age of investors in the treatment group is 45.6 years, with a standard deviation of 10.2 years, compared to the control group, which has an average age of 46.1

DeepSeek served as a backup system during this process. These systems were integrated and switched periodically. Our data indicates no significant difference in investor performance when using either the ChatGPT or DeepSeek LLM.

years and a standard deviation of 10.5 years. This results in a difference of -0.5 years, which is not statistically significant. The average account age is 8.5 years in the treatment group, with a standard deviation of 5.2 years, while the control group has an average account age of 8.7 years and a standard deviation of 5.0 years. The difference here is -0.15 years, again, not statistically significant. Initial portfolio values are quite similar between the groups. The treatment group has an average initial portfolio value of $184.4 thousand with a standard deviation of $50.3 thousand, whereas the control group has an average value of $182.8 thousand with a standard deviation of $51.2 thousand, leading to a difference of $1.5 thousand, which is not statistically significant.

The gender distribution, measured as the male ratio, shows that males constitute 74% of the treatment group (standard deviation 0.44) and 76% of the control group (standard deviation 0.43), with a difference of -0.02, which is not statistically significant. Investors in the treatment group hold an average of 15.3 stocks (standard deviation 10.4), slightly higher than the control group, which holds an average of 14.8 stocks (standard deviation 10.7), resulting in a difference of 0.5 stocks, which is not statistically significant. The average number of mutual funds held is 5.6 (standard deviation 3.2) in the treatment group and 5.8 (standard deviation 3.1) in the control group, leading to a difference of -0.2, which is not statistically significant. Returns three years before the intervention show a negative average return of -8.4% (standard deviation 2.5) for the treatment group and -8.2% (standard deviation 2.6) for the control group, with a difference of -0.2%, which is not statistically significant. Finally, the average number of trades per month is 16.8 (standard deviation 24.1) in the treatment group, compared to 16.5 (standard deviation 26.2) in the control group, resulting in a difference of 0.3 trades per month, which is not statistically significant.

These summary statistics provide a comprehensive overview of the comparability between the treatment and control groups, highlighting the similarities and differences across various investor characteristics and portfolio attributes. The lack of statistically significant differences suggests that the two groups are comparable across these dimensions.

4 Results

We next present our main results. Table 2 shows the relationship between robo-advising and investment outcomes, categorizing the results for mutual fund investors, stock investors, and investors participating in both markets. The table includes key metrics such as Average Annual Return, Annualized CAPM Alpha, Portfolio Diversification (measured by the Herfindahl Index), and Investor Satisfaction (measured on a scale of 1 to 5). Statistical significance of the differences between the control and treatment groups is indicated by *, **, and *** for the 10%, 5%, and 1% significance levels, respectively. Panel A of Table 1 shows the results for mutual fund investors. The treatment group, which received robo-advising, achieved an Average Annual Return of 14.8%, compared to 10.2% for the control group, resulting in a statistically significant difference of 4.6 percentage points ($p < 0.01$). The Alpha for the treatment group was 0.4% while the control group had a CAPM Alpha of -2.1%, leading to a significant difference of 2.4 percentage points ($p < 0.01$). Furthermore, Investor Satisfaction was significantly higher in the treatment group, with an average score of 4.5 compared to 3.8 in the control group (difference = 0.7, $p < 0.05$).

For mutual fund investors, Figure 1 shows that there is a significant decrease of active mutual fund investment for the treatment group. While for the control group during the same time period, there is no difference in the share of active equity mutual fund investment. Figure 1 also shows that there is a significant increase in passive index fund and ETFs investments for the treatment group. For the control group, again, there is no change in the investment in passive index fund and ETFs investments. Figure 2 shows that there is a significant increase in Sharpe ratio and investor gain for investors that receive the treatment. But there is not much Sharpe ratio change and investor gain for the control group.

Panel B highlights the outcomes for stock investors. The treatment group showed an Average Annual Return of 14.5%, which is 4.1 percentage points higher than the control group's 10.4% return ($p < 0.05$). The Alpha for the treatment group was 0.36%, whereas the control group had an Alpha of -2.0%, resulting in a significant difference of 2.4 percentage

points ($p < 0.05$). Similar to mutual fund investors, the treatment group exhibited better Portfolio Diversification with a Herfindahl Index of 0.19 as opposed to 0.30 in the control group (difference = -0.11, $p < 0.01$). Investor Satisfaction also improved significantly, with the treatment group scoring 4.4 compared to 3.7 for the control group (difference = 0.7, $p < 0.05$).

For individual stock investors, Figure 4 shows that there is a significant decrease of active common equity investment for the treatment group. While for the control group during the same time period, there is no difference in the share of active common equity investment. Figure 4 also shows that there is a significant increase in passive index fund and ETFs investments for the treatment group, similar to the results of active mutual fund. For the control group, again, there is no change in the investment in passive index fund and ETFs investments. Figure 5 shows that there is a significant increase in Sharpe ratio and investor gain for common equity investors that receive the treatment. But there is not much Sharpe ratio change and investor gain for the control group.

Panel C provides the results for investors participating in both mutual funds and stocks. The treatment group outperformed the control group with an Average Annual Return of 14.7%, which is 4.4 percentage points higher than the 10.3% return of the control group ($p < 0.01$). The Alpha for the treatment group was 0.4%, compared to -2.0% for the control group, resulting in a significant difference of 2.4 percentage points ($p < 0.01$). Finally, Investor Satisfaction in the treatment group was significantly higher, with an average score of 4.5 compared to 3.7 in the control group (difference = 0.7, $p < 0.05$).

The results across all panels consistently show that robo-advising has a positive impact on investment outcomes. The treatment groups, which received robo-advising, exhibited higher Average Annual Returns and Alpha, better Portfolio Diversification, and greater Investor Satisfaction compared to the control groups. These findings suggest that robo-advising can significantly enhance investment performance and investor satisfaction, making it a valuable tool for both mutual fund and stock investors.

Our findings reveal that after learning about beta, investors tend to favor ETFs or index funds over mutual funds or individual stocks. This shift indicates a recognition that market exposure, rather than individual security selection, is the primary driver of their returns. Prioritizing the education of investors on the concepts of beta and alpha offers more substantial and enduring benefits. Understanding these concepts helps investors make more informed decisions, recognize the true sources of their returns, and avoid overestimating their abilities. This foundational knowledge leads to better financial outcomes and more disciplined investment strategies, ultimately enhancing overall investor performance. Additionally, it allows investors to more accurately evaluate mutual fund managers, ensuring they are making wise choices in their investment selections, whether evaluating their fund managers' ability when purchasing mutual funds or their own investment acumen (alpha) when it comes to buying their own stocks.

4.1 LLM-based robo-advisors

We next investigate the heterogeneous effect of the treatment based on different follow-up assistance after giving the initial recommendation, especially the LLM-based follow-up assistance. the Figure 3 shows the heterogeneous effect of the treatment based on different follow-up assistance after giving the initial recommendation for mutual fund investors. LLMs-based chatbot assistance induces the most economic significant increase in shift to passive index fund and ETFs. The results of without assistance at all is much weaker than LLMs-based chatbot but slightly stronger than normal chatbot and human based assistance. Figure 6 shows a similar result for equity investors. Panel A of Table 3 presents the difference-in-differences regression coefficients and t-statistics (in parenthesis) for various performance indicators, including Monthly Return, Monthly CAPM Alpha, and Portfolio Diversification (measured by the Herfindahl Index).

$$Performance_{it} = \alpha_t + \beta_T Robo_{it} + \delta_T Robo_i * Post + \gamma_t^{Charact} Charact_{it} + \lambda_T^{Charact} Charact_{it} * Post + \epsilon_{iT} \quad (1)$$

The treatment group (*Robo*, indicator variable) is defined as investors who intended to and actually received robo-advising services, while the control group includes those who intended to but did not receive such services. The post-treatment period (Post) is defined as months after December 2023. $Charact_{it}$ are various characteristics of the investors.

Overall, investors receiving robo-advising services showed significantly better performance outcomes. The treatment group achieved an Average Raw Return of 39 bps higher compared to the control group, with a t-statistic of 4.45. The Alpha for the treatment group was 0.21% higher, significantly better than the control group, with a t-statistic of 4.98. Portfolio Diversification improved significantly in the treatment group, as evidenced by a lower Herfindahl Index of -0.11, with a t-statistic of -3.03.

Next, we conduct the regression analysis in four subsamples of treatment group where we keep the control group the same.[7] Further breakdown shows that the LLM (ChatGPT/DeepSeek/Kimi/Doubao) Chatbot subgroup outperformed other groups with an Average Annual Return of 0.64% (t-statistic 6.14) and an Alpha of 0.43% (t-statistic 6.20), along with a significant improvement in Portfolio Diversification (-0.17, t-statistic -3.61). In contrast, groups without chatbot support, human chat assistants, and normal chatbots showed lower but still significant improvements in performance metrics compared to the control group.

The Human Chat Assistant subgroup also showed positive results, with high advice-following rates and improved investment outcomes, though not as pronounced as those in the LLM/ChatGPT subgroup. The personalized and real-time assistance provided by human advisors contributed to better understanding and execution of investment strategies.

[7] The results are qualitatively similar if we use matched control group for each subsamples of the treatment group.

The Normal Chatbot subgroup demonstrated moderate improvements, with advice-following rates and investment outcomes better than the control group but not as significant as those seen in the ChatGPT or Human Chat Assistant subgroups. The predefined responses of the rule-based chatbot were helpful but lacked the dynamic interaction and depth provided by the LLM-powered chatbot and human advisors. Lastly, the Without Chatbot subgroup showed the least improvement among the treatment subgroups. While the static educational content did enhance financial literacy to some extent, the lack of interactive support limited the investors' ability to fully understand and apply the concepts to their investment decisions.

Panel B of Table 3 examines investor reactions, including Average Response Time, Interaction Times, and Investor Sentiment. The LLM (ChatGPT/DeepSeek/Kimi/Doubao) Chatbot subgroup demonstrated the most efficient response time of 3.7 seconds on average (standard deviation of 1.4), with the highest interaction frequency of 14.3 times per month (standard deviation of 28.9) and a positive investor sentiment score of 0.25 (standard deviation of 0.37). Conversely, the human chat assistant subgroup had a much longer average response time of 64.8 seconds (standard deviation of 264.7), fewer interaction times (5.1, standard deviation of 12.4), and a negative investor sentiment score (-0.17, standard deviation of 0.46). The normal chatbot subgroup showed the quickest average response time of 1.2 seconds (standard deviation of 0.3) but had the lowest interaction frequency and investor sentiment score. In summary, the results consistently indicate that robo-advising, particularly when integrated with advanced LLM-based chatbots like ChatGPT, significantly enhances investment performance and investor satisfaction. The treatment groups exhibited higher average returns, better portfolio diversification, and greater investor sentiment, demonstrating the substantial benefits of financial education and advanced robo-advising tools.

Our findings indicated significant heterogeneity in the results across the different subgroups within the treatment group. Investors in the LLM (ChatGPT/DeepSeek/Kimi/Doubao) Chatbot subgroup exhibited the highest advice-following rate and the most substantial im-

provements in investment outcomes. The interactive and natural language capabilities of the LLM-powered chatbot appeared to enhance investor engagement and comprehension, leading to better adherence to the robo-advisor's recommendations and superior portfolio performance.

4.2 Do investors understand the difference between alpha and beta?

A further question we seek to investigate is: without the continuous assistance of a robo-advisor, could our financial education have a longer-term effect on investors' ability to manage their own portfolios? To address this question, in the randomized controlled trial in which investors were randomly assigned to receive one of several types of advisory assistance, half of the investors receive a one-click switch option to a recommended robo-advisor portfolio and the other half need to manage their portfolios themselves (DIY). The one-click option allows investors to seamlessly transition to a professionally managed portfolio by simply clicking an agreed button, while the DIY option requires investors to implement the recommended changes on their own.

The primary objective of this experiment is to assess whether the investors fully understand the concepts of alpha and beta, and how this understanding influences their long-term financial behavior and portfolio performance. By comparing the performance of one-click and DIY portfolios across different advisory methods, we can infer the extent to which investors grasp these fundamental financial concepts. If investors comprehend the distinction between alpha (a measure of risk-adjusted return) and beta (a measure of systematic risk), they should be able to replicate the performance of the one-click portfolios in their DIY portfolios. This understanding is crucial for long-term financial success, as it implies that investors would not need continuous assistance from robo-advisors, thereby becoming more self-sufficient and informed in their investment decisions.

We test these hypothesis and show the results in Table 4. Table 4 shows the effectiveness

of different advisory methods on investor portfolio performance, specifically comparing the performance of one-click robo-advisor portfolios with DIY (do-it-yourself) portfolios. The table presents the difference-in-differences regression coefficients and t-statistics for CAPM alpha, which serves as a measure of risk-adjusted return.

Across all groups, we observe that the one-click alpha tends to be higher than the DIY alpha. This indicates that the simplicity and ease of switching to a robo-advisor portfolio with a single click generally lead to better performance outcomes. However, this difference is not the primary focus of our analysis.

The key finding in this table is the significant impact of LLM (ChatGPT/DeepSeek/Kimi/Doubao) follow-up assistance on investor understanding and performance. For the ChatGPT chatbot group, the one-click alpha is 0.59%, and the DIY alpha is 0.58%, with an insignificant difference of just 1 bps per month. This suggests that investors who received follow-up assistance from the ChatGPT chatbot were able to effectively manage their portfolios themselves, achieving performance nearly identical to the one-click option. The high t-statistics (5.92 and 5.79) further underscore the robustness of these results.

In stark contrast, other types of follow-up assistance show a substantial gap between one-click and DIY alphas. For example, the group without any chatbot assistance shows a one-click alpha of 0.33% compared to a DIY alpha of just 0.03%, resulting in a significant difference of 30 bps. Similar patterns are observed for the Human Chat Assistant and Normal Chatbot groups, with differences of 25 and 18 bps, respectively.

These findings suggest that while initial financial education is beneficial, not all investors fully grasp the distinction between alpha and beta. As a result, they may rely more heavily on the straightforward one-click option, which leads to higher performance. However, the ChatGPT chatbot's ability to provide clear and effective follow-up assistance enables investors to apply their knowledge more effectively, thereby narrowing the performance gap between one-click and DIY portfolios.

In summary, the table demonstrates that the LLM (ChatGPT/DeepSeek/Kimi/Doubao)

chatbot not only enhances initial financial literacy but also significantly improves investors' ability to manage their portfolios independently, achieving performance levels comparable to the more automated one-click options. This underscores the potential of advanced AI-driven advisory tools to empower investors with both knowledge and practical portfolio management skills. The results from our tests also suggest that while initial financial education positively impacts investor behavior, the effectiveness of follow-up assistance plays a critical role in ensuring that investors internalize and apply their knowledge. Specifically, the LLM (ChatGPT/DeepSeek/Kimi/Doubao) chatbot demonstrates a significant capability in bridging the gap between education and practical application, enabling investors to achieve comparable performance in their DIY portfolios as in the one-click portfolios. This has profound implications for the long-term financial independence and literacy of investors, as those who fully understand the difference between alpha and beta can confidently manage their portfolios without constant reliance on advisory services.

4.3 Heterogeneity Based On Investor Characteristics

Table 5 presents the difference-in-differences regression coefficients and t-statistics (in parentheses) for various performance indicators, including Raw Return, Alpha, and Portfolio Diversification (measured by the Herfindahl Index), across different investor characteristics. The subgroups analyzed are based on age, account age, account size, and education level.

Overall, the results indicate significant heterogeneity in investment performance across different investor segments. Younger investors aged 18-35 achieved 0.48% higher Raw Returns, with a t-statistic of 5.12, and 0.29% higher monthly Alpha (t-statistic 5.45). This group also exhibited a notable improvement in Portfolio Diversification, as indicated by a Herfindahl Index of 0.13 lower (t-statistic -4.53). Investors aged 36-50 showed 0.38% higher Raw Returns (t-statistic 3.85) and 0.20% higher monthly Alpha (t-statistic 3.91), with a Herfindahl Index of 0.10 lower (t-statistic -2.82). The 51+ age group demonstrated 0.34% higher Raw Returns (t-statistic 3.20) and 0.14% higher monthly Alpha (t-statistic 3.18),

along with a Herfindahl Index of 0.08 lower (t-statistic -2.13).

Examining account age, investors with accounts aged 0-5 years exhibited the highest performance with 0.46% higher Raw Returns (t-statistic 5.01) and 0.27% higher monthly Alpha (t-statistic 5.22), with a Herfindahl Index of 0.12 lower (t-statistic -4.23). Those with 6-10 year-old accounts had 0.39% higher Raw Returns (t-statistic 3.91) and 0.21% higher monthly Alpha (t-statistic 4.13), with a Herfindahl Index of 0.10 lower (t-statistic -2.74). Investors with accounts older than 11 years showed 0.32% higher Raw Returns (t-statistic 3.02) and 0.14% higher monthly Alpha (t-statistic 3.11), with a Herfindahl Index of 0.08 lower (t-statistic -1.92).

Regarding account size, investors with low account sizes (<300k RMB) achieved 0.35% higher Raw Returns (t-statistic 3.42) and 0.18% higher monthly Alpha (t-statistic 3.53), with a Herfindahl Index of 0.08 lower (t-statistic -2.01). Medium-sized accounts (300k-1M RMB) exhibited 0.41% higher Raw Returns (t-statistic 4.52) and 0.22% higher monthly Alpha (t-statistic 4.63), with a Herfindahl Index of 0.11 lower (t-statistic -3.02). High account sizes (>1M RMB) showed 0.44% higher Raw Returns (t-statistic 4.82) and 0.24% higher monthly Alpha (t-statistic 4.92), with a Herfindahl Index of 0.12 lower (t-statistic -3.51).

Finally, analyzing education levels, investors with a high school education achieved 0.33% higher Raw Returns (t-statistic 3.12) and 0.16% higher monthly Alpha (t-statistic 3.22), with a Herfindahl Index of 0.07 lower (t-statistic -1.93). Those with a Bachelor's Degree exhibited 0.39% higher Raw Returns (t-statistic 4.01) and 0.21% higher monthly Alpha (t-statistic 4.12), with a Herfindahl Index of 0.10 lower (t-statistic -2.61). Investors with a postgraduate degree showed 0.42% higher Raw Returns (t-statistic 4.32) and 0.23% higher monthly Alpha (t-statistic 4.43), with a Herfindahl Index of 0.11 lower (t-statistic -2.93).

4.4 Investor Behavior

We hypothesize that after learning about beta and alpha, there should be behavioral change for the investors. Specifically, we hypothesize that investors would spend less time trading

after they switch to passive investment. Table 6 panel A presents the regression results for the impact of treatment on three variables: monitoring time, login times, and trading frequency, across three different time periods: one month, three months, and six months. The coefficients, along with their corresponding t-statistics, are reported for each variable and time period. In the one-month period, the coefficients for the treatment group are 0.05, 0.02, and 0.01 for monitoring time, login times, and trading frequency, respectively. These coefficients are positive but statistically insignificant, as indicated by the t-statistics (0.53, 0.25, and 0.11, respectively).

For the three-month period, the treatment group coefficients show a significant reduction in all three variables. The coefficients are -0.53, -0.65, and -0.50 for monitoring time, login times, and trading frequency, respectively, with corresponding t-statistics of -5.33, -4.38, and -3.51. These results indicate a significant negative impact of the treatment over the three-month period. In the six-month period, the results are similar to those observed in the three-month period, suggesting a sustained negative impact of the treatment. The treatment group coefficients are -0.55, -0.70, and -0.57 for monitoring time, login times, and trading frequency, with t-statistics of -5.31, -4.59, and -3.67, respectively.

In summary, while the one-month results indicate no significant effect of the treatment, the three-month and six-month results reveal a significant and sustained negative impact on monitoring time, login times, and trading frequency. This suggests that the treatment effectively reduces these activities over longer periods, with the magnitude of the effect remaining consistent between the three-month and six-month intervals.

Panel B shows the results when we estimate a triple difference-in-differences regression where we add an indicator LLM, that equals one if the treatment group has LLM-based follow-up assistance. The results show that when investors have LLM-based follow-up assistance, there is a significant marginal decrease in monitoring time, login time, and trading frequency. This suggests that LLM-based follow-up assistance indeed helps investors manage their portfolios more efficiently by reducing the need for constant oversight and potentially

improving their decision-making process.

4.5 Fund Characteristics

We also investigate the changes in investor's holding in terms of fund characteristics. Figure 2 shows that there is a significant decrease of average mutual fund management fee being charged for the treatment group but not for the control group. Table 7 presents the difference-in-differences regression coefficients and t-statistics (in parentheses) for various characteristics of mutual funds that investors purchase, including Expense Ratio, Turnover Ratio, Fund Size, and Beta. The analysis focuses on the impact of the robo-advising service on these mutual fund characteristics. The coefficient for the Expense Ratio in the treatment group is -0.35, with a t-statistic of -4.84, indicating a statistically significant reduction at the 1% significance level. This suggests that investors using the robo-advising service tend to choose funds with lower expense ratios. The Turnover Ratio for the treatment group is -9.18, with a t-statistic of -3.21, also significant at the 1% level, indicating that these investors prefer funds with lower turnover ratios after received robo advising, and these funds are ETFs that mechanically have lower turnovers.

The coefficient for Fund Size in the treatment group is 0.14, with a t-statistic of 0.11, which is not statistically significant. This implies that the robo-advising service does not have a significant impact on the size of the funds that investors choose. Similarly, the Beta coefficient for the treatment group is 0.03, with a t-statistic of 0.07, showing no significant change in the risk profile of the funds selected by investors using the robo-advising service. The constant terms for each characteristic are also reported. The Expense Ratio constant is 1.40, with a t-statistic of 8.21, significant at the 1% level, indicating that the baseline expense ratio for non-treated investors is relatively high. The Turnover Ratio constant is 38.35, with a t-statistic of 7.31, significant at the 1% level, showing a high baseline turnover ratio. The Fund Size constant is 2.83, with a t-statistic of 3.25, significant at the 5% level, and the Beta constant is 1.04, with a t-statistic of 20.42, significant at the 1% level , reflecting the

typical characteristics of funds chosen by non-treated investors. Overall, the results indicate that the robo-advising service significantly influences investors' preferences for mutual fund characteristics, particularly in reducing expense and turnover ratios, which can lead to better investment outcomes.

Table 8 presents the results of the difference-in-differences regression analysis examining the impact of the robo-advising service on investor account net flows, specifically focusing on overall net flow and ETF net flow in percentage terms. The table reports the coefficients and corresponding t-statistics (in parentheses) for the interaction term (Treat*Post), post-intervention period indicator (Post), and treatment group indicator (Treat). The results suggest that there is a significant increase (6.37%) in overall inflow into the treatment group's accounts compared to the control group's accounts, which implies a moderate increase in stock market participation among the treatment group. We also show that this inflow mainly goes to ETF investment, indicating a significant reallocation to ETF investment.

4.6 Oversea diversification and factor investment

One other interesting result we find is that, investors in the treatment group actually also start to diversify to S&P 500 ETFs and other foreign ETFs, even though the initial recommendation only consists of reallocating the Chinese A-share market ETFs.[8] Table 9 presents the results of the difference-in-differences regression analysis examining the impact of the robo-advising service on investors' allocation to S&P500, US Factor Style Investment, and Chinese ADRs. The table reports the coefficients and corresponding t-statistics (in parentheses) for the interaction term (Treated*Post), the post-intervention period indicator (Post), the treatment group indicator (Treat), and the constant.

For the allocation to S&P500, the coefficient for the interaction term (Treated*Post) is 1.52, with a t-statistic of 3.50, indicating a statistically significant increase at the 1% significance level. This suggests that the robo-advising service significantly increases the

[8]Although these are not directly recommended from the robo-advisors, the choices are available through investor's investment menu.

allocation to S&P500 by 1.52 percentage points. The coefficient for the treatment group (Treat) is 0.09, with a t-statistic of 0.19, which is not statistically significant, indicating no significant difference in allocation to S&P500 between the treatment and control groups before the intervention.

For the allocation to US Factor Style Investment, the coefficient for the interaction term (Treated*Post) is 2.04, with a t-statistic of 4.01, indicating a statistically significant increase at the 1% significance level. This implies that the robo-advising service significantly increases the allocation to US Factor Style Investment by 2.04 percentage points. The coefficient for the treatment group (Treat) is 0.12, with a t-statistic of 0.24, which is not statistically significant, indicating no significant difference in allocation to US Factor Style Investment between the treatment and control groups before the intervention.

For the allocation to Chinese ADRs, the coefficient for the interaction term (Treated*Post) is 1.81, with a t-statistic of 3.71, indicating a statistically significant increase at the 1% significance level. This suggests that the robo-advising service significantly increases the allocation to Chinese ADRs by 1.81 percentage points. The coefficient for the treatment group (Treat) is 0.10, with a t-statistic of 0.23, which is not statistically significant, indicating no significant difference in allocation to Chinese ADRs between the treatment and control groups before the intervention.

This result is puzzling given that the robo-advisor did not recommend foreign diversification directly. Therefore, we hypothesize that LLM-based chatbot may have recommended foreign diversification beyond what the brokerage firms initially recommended to them. Indeed, Panel B of Table 9 presents the results of the triple difference-in-differences regression, which includes an additional indicator for LLM (AI) interaction effects. The coefficient for the TreatedPostLLM interaction term is 3.48 for S&P500, 3.07 for US Factor Style Investment, and 4.71 for Chinese ADRs, all with t-statistics greater than 4, indicating statistically significant increases at the 1% significance level. This further suggests that LLM-based follow-up assistance significantly enhances the impact of the robo-advising service on these

allocations. At the same time, Treated*Post coefficients are not significant when controlling for Treated*Post*LLM, suggesting that the effect is concentrated in the subgroup that use LLM chatbots.

In conclusion, the results suggest that the introduction of LLM-based follow-up assistance significantly amplifies the impact of robo-advising services on foreign diversification, guiding investors towards greater allocations in S&P 500 ETFs, US Factor Style Investments, and Chinese ADRs. This highlights the potential of AI-driven assistance to enhance and extend the recommendations of traditional robo-advisors, leading to more diversified investment portfolios.

5 Robustness Checks

5.1 Beyond CAPM Alpha

In this robustness check, we examine the impact of different advisory methods on financial literacy using various factor models to compute alphas. Table A.1 presents the difference-in-differences regression coefficients along with their respective t-statistics for the Four Factor Alpha configurations. The alphas are calculated using the Fama-French Three Factor Model (Fama and French, 1992), the Three Factor Model with Momentum (Carhart, 1997), the Fama-French Five Factor Model (Fama and French, 2015), and the Five Factor Model with Momentum (FF Five + MOM Alpha).

The table highlights the comparative effectiveness of different advisory approaches: a ChatGPT-based robo-advisory chatbot, a traditional chatbot, a human chat assistant, and a control group of investors who intended to but did not receive the advisory service. The results across all alpha definitions consistently show that the ChatGPT-based chatbot provides the most significant positive impact on financial literacy.

The LLM (ChatGPT/DeepSeek/Kimi/Doubao) Chatbot group generates alphas of 0.44%, 0.43%, 0.41%, and 0.40% across the FF Three Alpha, FF Three + MOM Alpha, FF Five

Alpha, and FF Five + MOM Alpha, respectively, with highly significant t-statistics ranging from 5.70 to 6.27. These results clearly demonstrate the superior effectiveness of the ChatGPT-based chatbot in enhancing financial outcomes compared to both traditional methods and other chatbots.

Conversely, the group without any chatbot support shows lower, yet still significant, alphas ranging from 0.13% to 0.10%, with t-statistics between 1.99 and 2.25. This suggests that while traditional methods do have a positive impact, their effectiveness is markedly less than that of the ChatGPT-based advisory.

The Human Chat Assistant and Normal Chatbot groups also show significant positive alphas, but these are lower than those of the ChatGPT Chatbot group. The Human Chat Assistant group shows alphas ranging from 0.14% to 0.11% with t-statistics between 2.24 and 2.41, while the Normal Chatbot group shows alphas ranging from 0.12% to 0.09% with t-statistics between 2.03 and 2.29.

In summary, the robustness check confirms that the ChatGPT-based robo-advisor significantly outperforms both human and traditional chatbot advisors in improving financial outcomes, as indicated by the higher and more significant alphas across multiple factor models.

5.2 Alternative robo-advising model

We also conduct robustness checks based on the alternative robo-advising models that the brokerage firm applied during the pilot period. Our current empirical analysis is based on the sample where treated investors are advised to retain mutual funds or stocks that are performing well, using a simple rule of thumb to keep only those with positive alpha (after fees) and a t-statistic above 3. However, extensive literature has documented that past performance of stocks and mutual funds cannot reliably predict future performance, raising questions about whether the current model is the most beneficial for investors. Despite this, it is noteworthy that only a very small percentage of stocks or mutual funds (less than 10%)

are suggested to be kept by investors under this model.

Additionally, the brokerage firm experimented with alternative robo-advising models. One model applied a "one-size-fits-all" approach, suggesting that the entire portfolio be converted to passive ETFs. Another model involved a complex algorithm incorporating machine learning techniques and predictors such as past mutual fund performance, fund size, expense ratio, turnover ratios, manager tenure, active share, fund flows (for mutual funds), and neural network analysis for stocks to make keep-or-switch recommendations. Interestingly, the results from the 'one-size-fits-all' model, the machine learning model, and the simple rule of thumb based on retaining mutual funds and stocks with positive alpha (after fees) and a t-statistic above 3 were very similar.

6 Conclusion

In this study, we investigated the impact of financial literacy on investor behavior and performance by utilizing a robo-advising setup. Specifically, we focused on educating investors about the concepts of beta and alpha in mutual funds and stocks. Our partnership with a large brokerage firm enabled us to provide a comprehensive analysis of the effects of this financial education.

To enhance the educational experience, we integrated a back-end ChatGPT system with the robo-advisor. This integration allowed investors to interact with the system and deepen their understanding of beta and alpha, crucial metrics in evaluating investment risk and performance. Our findings reveal several significant outcomes from this educational intervention.

Firstly, investors who learned about beta demonstrated a noticeable shift in their investment preferences. They showed a preference for investing in ETFs or index funds rather than directly in mutual funds or individual stocks. This shift suggests that understanding beta helps investors appreciate the benefits of diversification and the lower costs associated

with passive investment vehicles.

Secondly, frequent traders exhibited behavior consistent with constructing their own factor portfolios. This indicates that financial literacy can empower more active investors to apply advanced investment strategies, potentially improving their portfolio management skills.

Most importantly, our study highlights the tangible financial benefits of financial education. Investors who received education on beta and alpha experienced a substantial improvement in their investment performance, with their monthly returns increasing by 40 basis points. Additionally, these investors achieved an average monthly gain of 750 RMB. These results underscore the value of financial literacy in enhancing investor outcomes.

Overall, our research contributes to the growing body of literature on financial education and its impact on investor behavior and performance. The findings emphasize the importance of integrating educational tools, such as ChatGPT, within robo-advisory platforms to facilitate investor learning. By equipping investors with a better understanding of key financial concepts, we can help them make more informed investment decisions and achieve better financial outcomes.

Future research could further explore the long-term effects of financial education on investor behavior and performance, as well as the potential benefits of educating investors on additional financial concepts. Additionally, examining the impact of financial literacy across different demographic groups could provide insights into tailoring educational interventions to meet the diverse needs of investors.

In conclusion, this study demonstrates that financial literacy, facilitated through innovative technologies like robo-advisors and AI-driven educational tools, can significantly enhance investor performance. By continuing to develop and integrate these educational resources, our research sheds light on how AI combined with big data can empower investors to achieve their financial goals and contribute to a more financially literate and resilient investment community (Goldstein et al., 2021, 2024).

References

Calvet, L. E., Campbell, J. Y., and Sodini, P. (2009). Fight or flight? Portfolio rebalancing by individual investors. *The Quarterly Journal of Economics*, 124(1):301–348.

Campbell, J. Y. (2006). Household finance. *The Journal of Finance*, 61(4):1553–1604.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1):57–82.

Chak, I., Croxson, K., D'Acunto, F., Reuter, J., Rossi, A. G., and Shaw, J. M. (2022). Improving household debt management with robo-advice. *NBER Working Paper*.

Cong, L. W., Liang, T., Zhang, X., and Zhu, W. (2024). Textual factors: A scalable, interpretable, and data-driven approach to analyzing unstructured information. *NBER Working Paper*.

Cong, L. W., Tang, K., Wang, J., and Zhang, Y. (2019). AlphaPortfolio: Direct construction through deep reinforcement learning and interpretable AI. *SSRN Working Paper*.

D'Acunto, F., Prabhala, N., and Rossi, A. G. (2019). The promises and pitfalls of robo-advising. *The Review of Financial Studies*, 32(5):1983–2020.

D'Acunto, F. and Rossi, A. G. (2023). Robo-advice: Transforming households into rational economic agents. *Annual Review of Financial Economics*, 15(1):543–563.

D'Acunto, F. and Rossi, A. G. (2021). New frontiers of robo-advising: Consumption, saving, debt management, and taxes. *Machine Learning in Finance: A Guide to Contemporary Practice, forthcoming*.

D'Acunto, F. and Rossi, A. G. (2023). It meets finance: financial decision-making in the digital era. *Handbook of Financial Decision Making*, pages 336–354.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.

Gargano, A. and Rossi, A. G. (2024). Goal setting and saving in the fintech era. *The Journal of Finance*, 79(3):1931–1976.

Goldstein, I., Spatt, C. S., and Ye, M. (2021). Big data in finance. *The Review of Financial Studies*, 34(7):3213–3225.

Goldstein, I., Spatt, C. S., and Ye, M. (2024). The next chapter of big data in finance. *The Review of Financial Studies*.

Greig, F., Ramadorai, T., Rossi, A. G., Utkus, S. P., and Walther, A. (2022). Algorithm aversion: Theory and evidence from robo-advice. *SSRN Working Paper*.

Lewellen, J. (2011). Institutional investors and the limits of arbitrage. *Journal of Financial Economics*, 102(1):62–80.

Lewellen, J. (2023). How many factors? *Working Paper*.

Lewellen, J. and Lewellen, K. (2022). Institutional investors and corporate governance: The incentive to be engaged. *The Journal of Finance*, 77(1):213–264.

Liberman, J., Sialm, C., Sosner, N., and Wang, L. (2020). The tax benefits of separating alpha from beta. *Financial Analysts Journal*, 76(1):38–61.

Rossi, A. G. and Utkus, S. (2024). The diversification and welfare effects of robo-advising. *Journal of Financial Economics*, 157:103869.

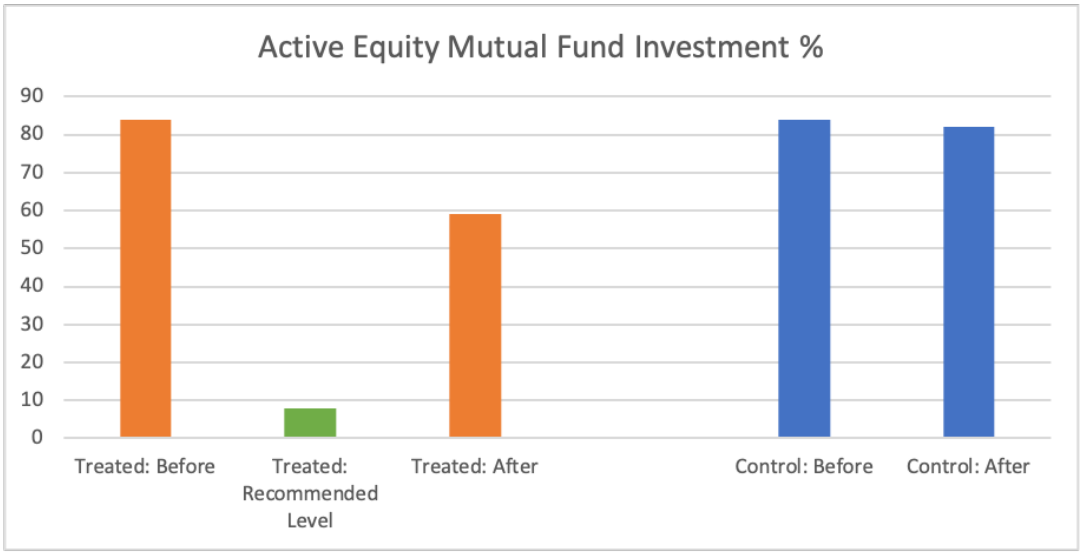
Sheng, J., Sun, Z., Yang, B., and Zhang, A. L. (2024). Generative AI and asset management. *SSRN Working Paper*.

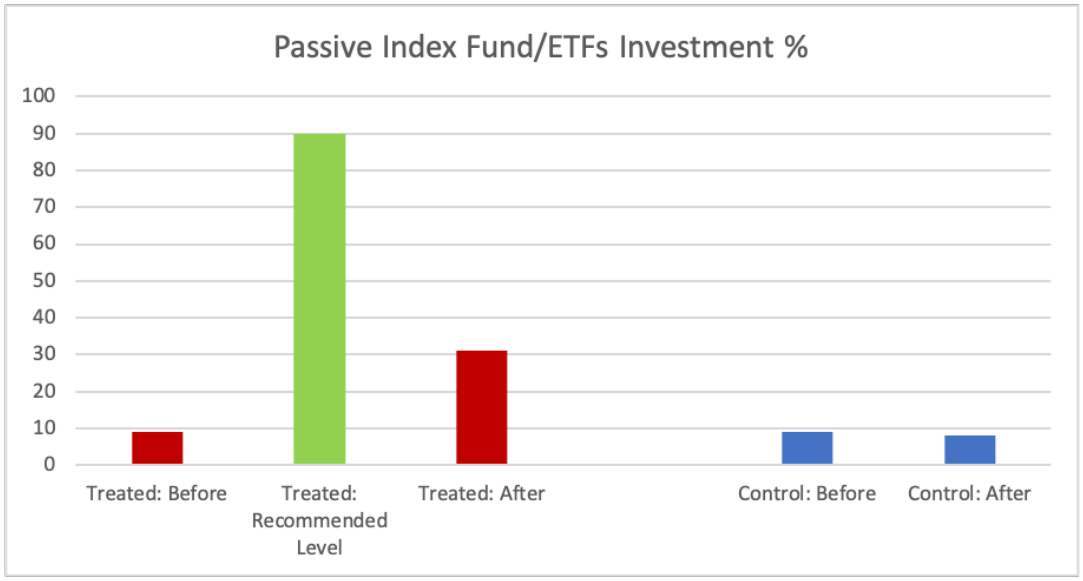
Sialm, C. and Tham, T. M. (2016). Spillover effects in mutual fund companies. *Management Science*, 62(5):1472–1486.

Sialm, C. and Xu, D. X. (2024). Information acquisition by mutual fund investors: Evidence from stock trading suspensions. *Working Paper*.

Van Rooij, M., Lusardi, A., and Alessie, R. (2011). Financial literacy and stock market participation. *Journal of Financial Economics*, 101(2):449–472.

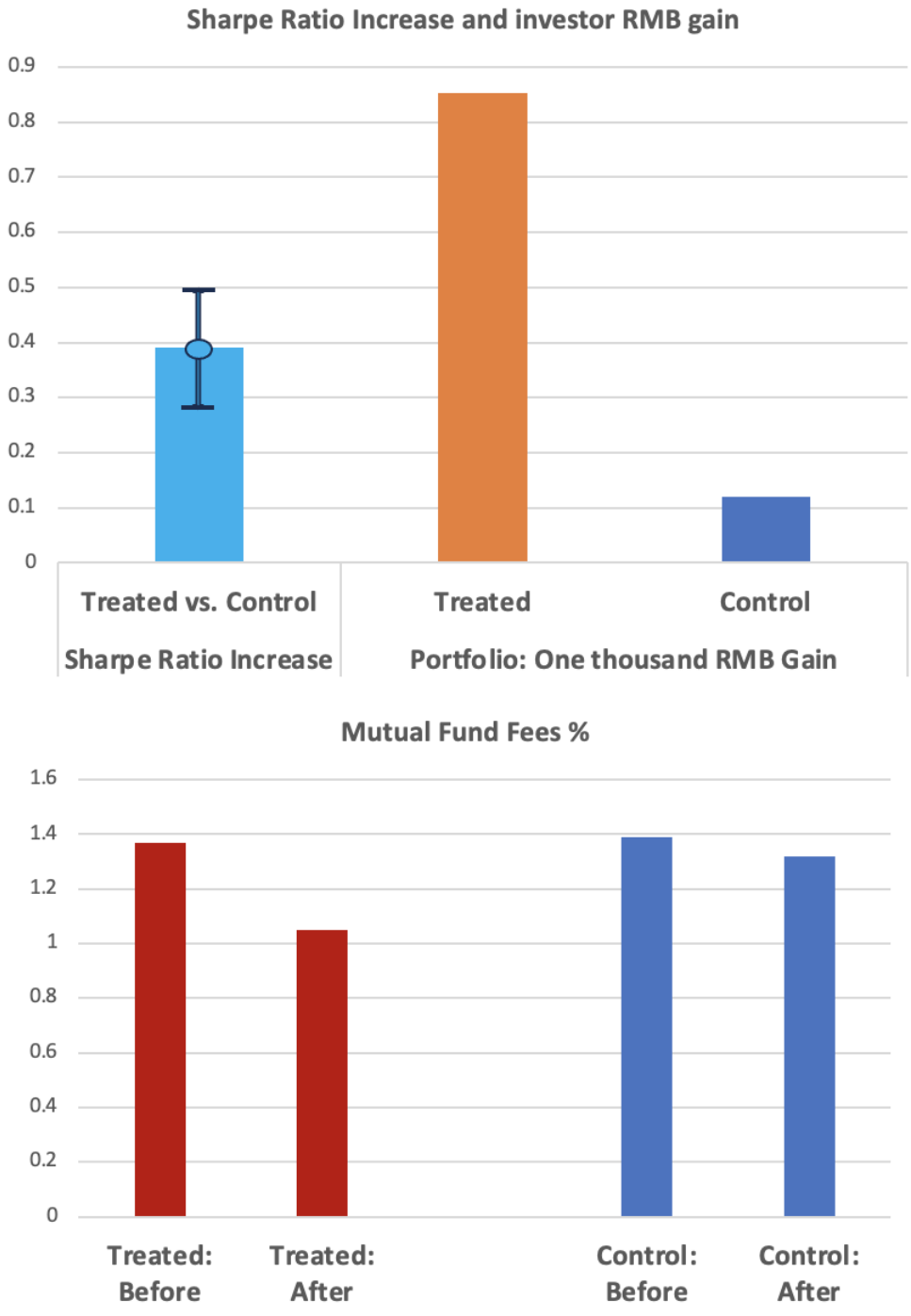
Figure 1: Robo-advising and portfolio reallocation for mutual fund investors





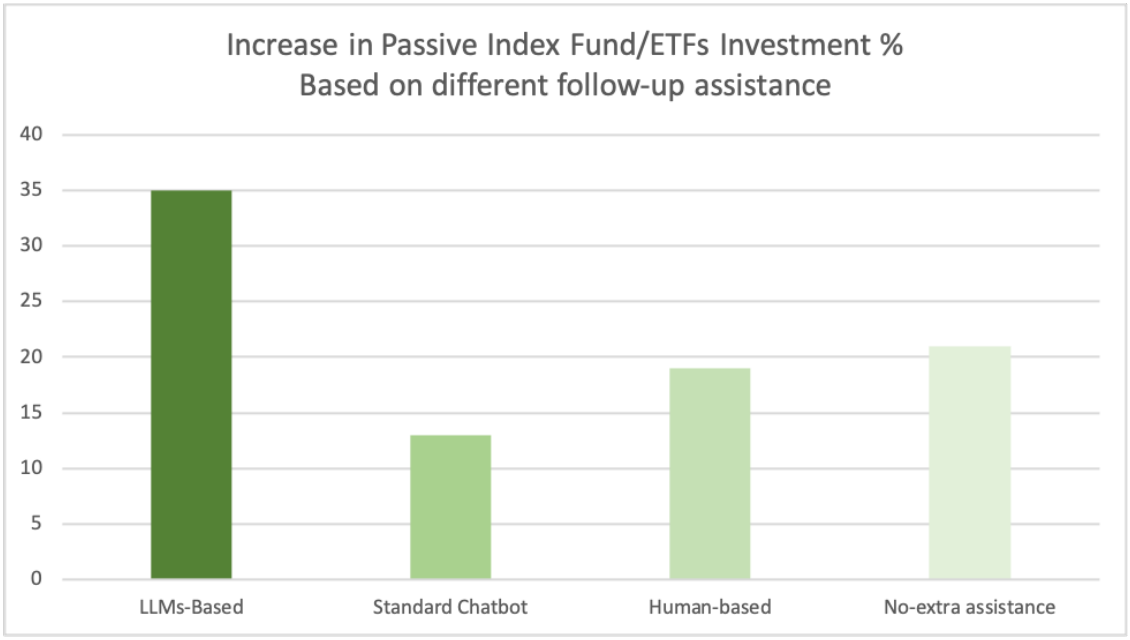
This Table shows the portfolio reallocation after robo-advising.

Figure 2: Robo-advising and investors' gain for mutual fund investors



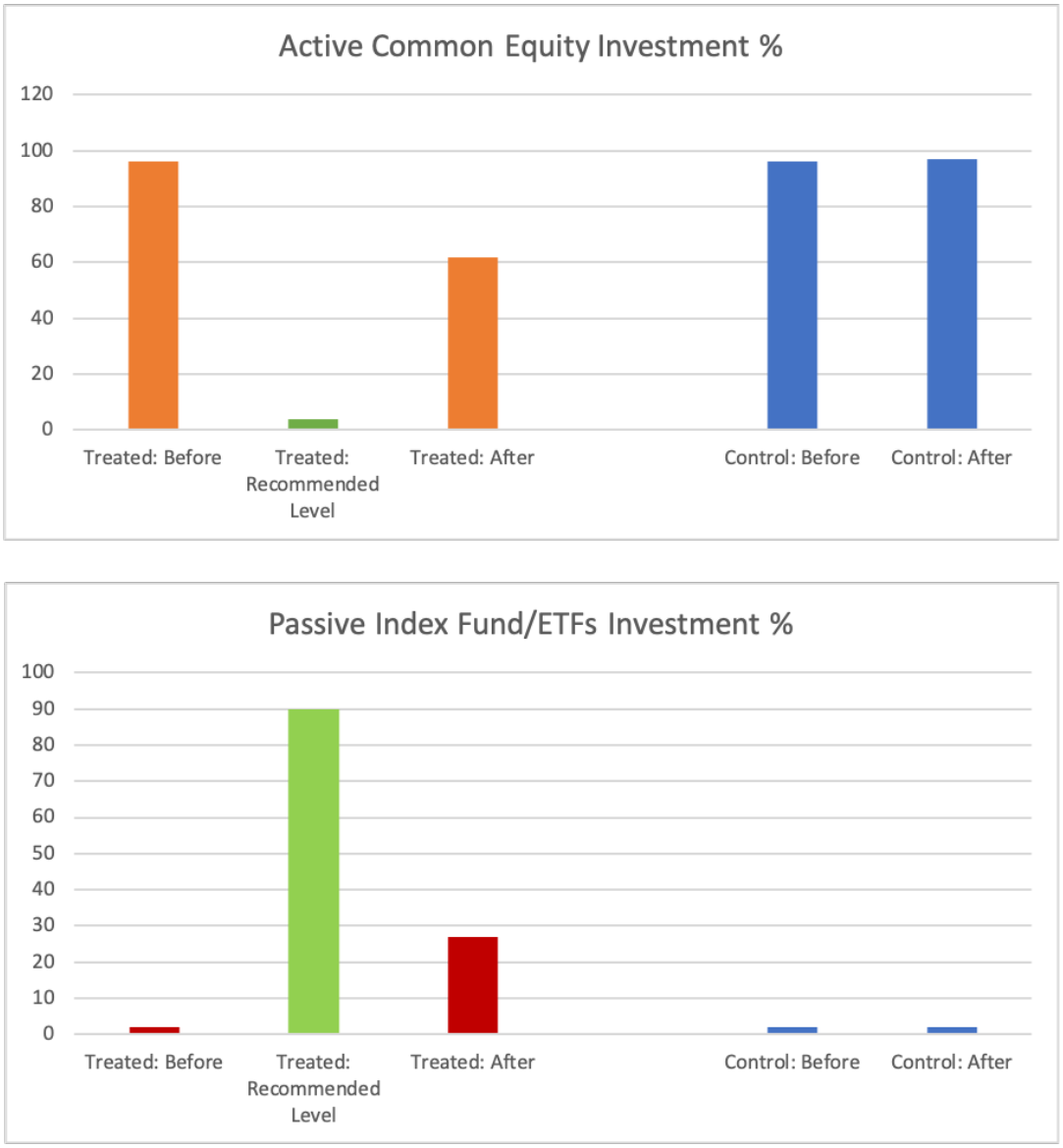
This Table shows the Sharpe ratio increase and monetary gain after robo-advising.

Figure 3: Heterogeneous effect based on different follow-up assistances for mutual fund investors



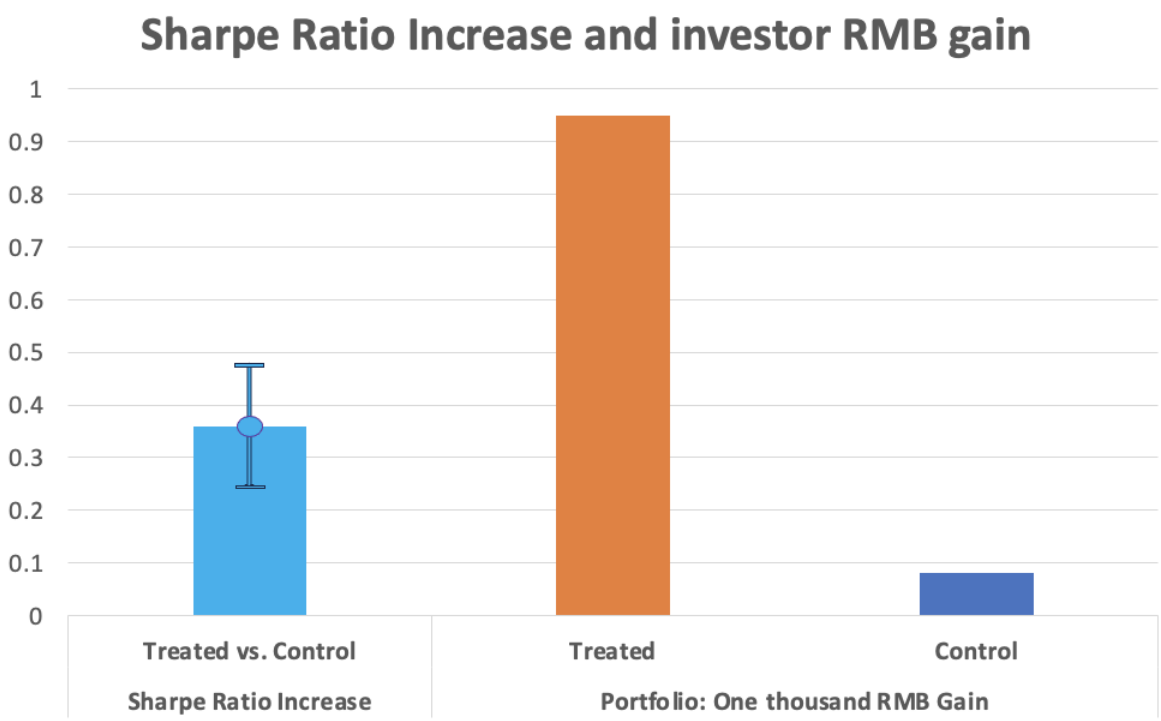
This Table shows the heterogeneous treatment effect based on different follow-up assistances.

Figure 4: Robo-advising and portfolio reallocation for stock investors



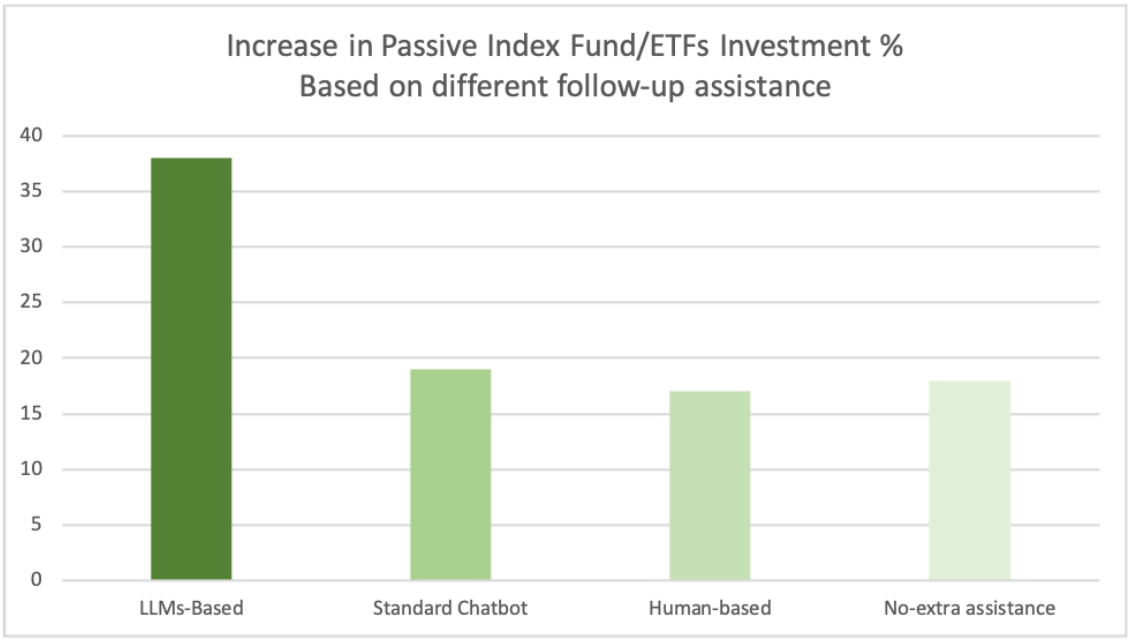
This Table shows the portfolio reallocation after robo-advising.

Figure 5: Robo-advising and investors' gain for stock investors



This Table shows the Sharpe ratio increase and monetary gain after robo-advising.

Figure 6: Heterogeneous effect based on different follow-up assistances for stock investors



This Table shows the heterogeneous effect based on different follow-up assistances.

Table 1: Summary Statistics

Summary statistics of the characteristics of the investors and their portfolios are presented in the table below. The table shows the means and standard deviations (in parentheses) for each characteristic before the treatment. Additionally, the mean differences between the treatment and control groups are calculated. T-tests for these mean differences are performed, and the results are marked with *, **, and *** to signify statistical significance at different levels. Specifically, * indicates significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level.

Characteristic	**Treatment Group**	**Control Group**	**Difference**
Age (years)	45.63 (10.21)	46.19 (10.54)	-0.56
Account ages (years)	8.56 (5.20)	8.71 (5.08)	-0.15
Initial Portfolio Value ($1000)	184.4 (50.39)	182.84 (51.21)	1.57
Gender (Male ratio)	0.74 (0.44)	0.76 (0.43)	-0.02
Number of Stocks held	15.38 (10.42)	14.82 (10.75)	0.56
Number of Mutual Funds held	5.63 (3.21)	5.89 (3.17)	-0.26
Returns Three Years before intervention (%)	-8.45 (2.58)	-8.27 (2.64)	-0.18
Average trade per month	16.87 (24.15)	16.54 (26.23)	0.33
N	14,409	14,410	

Table 2: Robo-advising and investment outcomes

This Table shows the relationship between robo-advising and investment outcomes separately for mutual fund investors only, stock investors, and investors in both markets. The table shows the metrics for the control and treatment groups, along with the calculated differences. The differences are marked with *, **, and *** to signify statistical significance at different levels. Specifically, * indicates significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level. The metrics include Average Annual Return, Alpha (CAPM Alpha), Portfolio Diversification (measured by the Herfindahl Index), and Investor Satisfaction (measured on a scale of 1 to 5). Panel A presents results for Mutual Fund Investors, Panel B for Stock Investors, and Panel C for Investors in both markets.

Panel A: Mutual Fund Investors

Metric	**Control Group**	**Treatment Group**	**Difference**
Average Annual Return (%)	10.24	14.89	4.65***
Annualized Alpha (%)	-2.03	0.41	2.44***
Investor Satisfaction (out of 5)	3.81	4.57	0.76**
N	2,849	2,849	

Panel B: Stock Investors

Metric	**Control Group**	**Treatment Group**	**Difference**
Average Annual Return (%)	10.41	14.53	4.12**
Annualized Alpha (%)	-2.04	0.36	2.40**
Portfolio Diversification (Herfindahl Index)	0.30	0.19	-0.11***
Investor Satisfaction (out of 5)	3.72	4.46	0.74**
N	6,381	6,381	

Panel C: Investors in both markets

Metric	**Control Group**	**Treatment Group**	**Difference**
Average Annual Return (%)	10.32	14.71	4.39***
Annualized Alpha (%)	-2.03	0.39	2.42***
Investor Satisfaction (out of 5)	3.77	4.51	0.74**
N	5,179	5,180	

Table 3: Comparison of Different Chatbot and Human Assistant Results on Financial Literacy Impact

Panel A shows the difference-in-differences regression coefficients δ_T in regression specification (1) and t-statistics (in parentheses) in the table. Treatment equals one if the investors intended to receive the robo-advising service and actually received the service. Control equals one if the investors intended to receive the robo-advising service but were not selected to receive the service. The sample period is from December 2022 to November 2024. Post equals one for observations (months) after December 2023. Raw Return is the monthly raw return. Alpha is the monthly CAPM alpha. Individual fixed effects and month fixed effects are included. Panel B shows the heterogeneous effect of different follow-up services after the initial treatment. Mean and standard deviations (in parentheses) are reported. The four subgroups are LLM-based Chatbot, Human Chat Assistance, Normal Chatbot, and No Chatbot. Response Time is the average response time in seconds. Interaction Times is the number of times investors interact with the chatbot or human live chat service per month. Investor Sentiment is scaled so that 1 is the highest and -1 is the lowest.

Panel A: Performance (Monthly)	**Raw Return (%)**	**Alpha (%)**	**Herfindahl Index**	**N**
Overall	0.39***	0.21***	-0.11***	691,656
	(4.45)	(4.98)	(-3.03)	
ChatGPT (LLM) Chatbot	0.65***	0.43***	-0.17***	172,914
	(6.14)	(6.20)	(-3.61)	
Without Chatbot	0.22**	0.13**	-0.09**	172,914
	(2.20)	(2.38)	(-2.10)	
Human Chat Assistant	0.25**	0.14**	-0.04*	172,914
	(2.16)	(2.47)	(-1.73)	
Normal Chatbot	0.21**	0.12**	-0.05*	172,914
	(2.17)	(2.29)	(-1.84)	
Panel B: Investor Reaction	**Response time**	**Interaction times**	**Investor sentiment**	
ChatGPT (LLM) Chatbot	3.71	14.36	0.25	
	(1.42)	(28.92)	(0.37)	
Human Chat Assistant	64.89	5.14	-0.17	
	(264.7)	(12.4)	(0.46)	
Normal Chatbot	1.24	1.31	-0.24	
	(0.31)	(10.62)	(0.27)	

Table 4: One-Click Robot advisor vs. DIY portfolio

This Table shows the difference-in-differences regressions coefficients and t-statistics (in parenthesis) in the Table. The Treatment equals one if the investors intend to receive the robo-advising service and actually received the robo-advising service. The Control equals one if the investor intend to receive the robo-adivisng service but was not selected to receive the service. The sample period is from December 2022 to November 2024. Post equals one for months after 2023 December. Alpha is the CAPM alpha. The treatment group is further sorted into two groups. One-click Alpha is the treatment alpha from investors who can switch to robo-advisor portfolio by simply click an agreed button. DIY alpha is the treatment alpha from investors who need to switch portfolios by themselves.

Performance (Monthly)	**One-click Alpha (%)**	**DIY Alpha (%)**	**Difference**
Overall	0.38***	0.19*	0.19
	(3.14)	(1.73)	(1.16)
ChatGPT (LLM) Chatbot	0.59***	0.58***	0.01
	(5.92)	(5.79)	(0.08)
Without Chatbot	0.33***	0.03	0.30**
	(2.67)	(0.32)	(2.07)
Human Chat Assistant	0.34***	0.09	0.25
	(2.82)	(0.79)	(1.55)
Normal Chatbot	0.28**	0.10	0.18
	(2.31)	(0.64)	(0.90)

Table 5: Heterogeneity in terms of investor characteristics

This table presents the difference-in-differences regression coefficients and their corresponding t-statistics (in parentheses) for various subgroups based on different investor characteristics. These characteristics include age, account age, account size, and education level. The table highlights the heterogeneity in investment performance, measured by monthly raw return, alpha, and the Herfindahl Index, across these different investor segments. Significance levels are indicated by asterisks, where *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Investor Segment	**Raw Return (%)**	**Alpha (%)**	**Herfindahl Index**
Age			
18-35	0.48***	0.29***	-0.13***
	(5.12)	(5.45)	(-4.53)
36-50	0.38**	0.20**	-0.10**
	(3.85)	(3.91)	(-2.82)
51+	0.34**	0.14**	-0.08*
	(3.20)	(3.18)	(-2.13)
Account Age			
0-5 years	0.46***	0.27***	-0.12***
	(5.01)	(5.22)	(-4.23)
6-10 years	0.39**	0.21**	-0.10**
	(3.91)	(4.13)	(-2.74)
11+ years	0.32**	0.14**	-0.08*
	(3.02)	(3.11)	(-1.92)
Account Size			
Low (<300k RMB)	0.35**	0.18**	-0.08
	(3.42)	(3.53)	(-2.01)
Medium (300k-1M RMB)	0.41***	0.22***	-0.11**
	(4.52)	(4.63)	(-3.02)
High (>1M RMB)	0.44***	0.24***	-0.12***
	(4.82)	(4.92)	(-3.51)
Education Level			
High School	0.33**	0.16**	-0.07*
	(3.12)	(3.22)	(-1.93)
Bachelor's Degree	0.39**	0.21**	-0.10**
	(4.02)	(4.23)	(-2.72)
Master's/PhD	0.43***	0.25***	-0.12***
	(4.72)	(4.83)	(-3.42)

Table 6: Login Time and Monitor Time

This table presents the difference-in-differences regression coefficients and their corresponding t-statistics (in parentheses) for different investor characteristics. Login time is the number of hours of login per day. Login time is the number of login per day. Trading Frequency is the number of trades per day. Significance levels are indicated by asterisks, where *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A			
Variables:	**Monitoring Time**	**Login Times**	**Trading Frequency**
One Month			
Treated*Post	0.05	0.02	0.01
	(0.53)	(0.25)	(0.11)
Three Month			
Treated*Post	-0.53***	-0.65***	-0.50***
	(-5.33)	(-4.38)	(-3.51)
Six Month			
Treated*Post	-0.55***	-0.70***	-0.57***
	(-5.31)	(-4.59)	(-3.67)
Panel B (Six Month Only)			
LLM (AI) Interaction effects			
Treated*Post*LLM	-0.62***	-0.92***	-0.80***
	(-4.82)	(-4.21)	(-4.29)
Treated*Post	-0.23***	-0.48***	-0.31***
	(-2.78)	(-2.95)	(-2.18)

Table 7: Comparing the mutual fund characteristics

This table presents the difference-in-differences regression coefficients along with their corresponding t-statistics (in parentheses) based on different characteristics of the funds that investors purchase. Significance levels are indicated by asterisks, where *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A Variables:	Expense Ratio (%)	Turnover Ratio (%)	Fund Size (billion RMB)	Beta
Treated*Post	-0.35***	-9.18***	0.14	0.03
	(-4.84)	(-3.21)	(0.11)	(0.07)
Treated	0.04	-0.36	0.02	-0.01
	(0.17)	(-0.40)	(0.09)	(-0.32)
Constant	1.40***	38.35***	2.83**	1.04***
	(8.22)	(7.30)	(3.25)	(20.42)
Panel B				
LLM (AI) Interaction effects				
Treated*Post*LLM	-0.51***	-13.63***	0.06	0.04
	(-3.69)	(-3.58)	(0.09)	(0.11)
Treated*Post	-0.13**	-5.38**	0.18	0.02
	(-2.41)	(-2.35)	(0.22)	(0.05)

Table 8: Individual Investor Account Netflow

This table shows the net flow of of deposits in or out of investors' accounts and the net flow in or out of their corresponding ETF accounts. Overall Funding Netflow is the percentage of inflow (positive) or outflow (negative) into the investor's account. ETF Netflow is the amount of inflow (positive) or outflow (negative) into ETFs, divided by the original capital in ETFs. A higher ETF Netflow indicates not only new ideposits into ETFs but also reallocated funds from stocks or mutual funds. The analysis includes individual fixed effects and month fixed effects. Significance levels are indicated by asterisks, where *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A:		
Variable	**Overall Funding Netflow (%)**	**ETF Netflow**
Treat*Post	6.37***	4.19***
	(5.12)	(4.75)
Treat	0.24	0.33
	(0.65)	(0.59)
Constant	0.11	0.14
	(0.09)	(0.11)
Panel B		
LLM (AI) Interaction effects		
Treat*Post*LLM	8.49***	6.31***
	(4.25)	(4.19)
Treat*Post	3.42***	1.02***
	(3.48)	(3.51)

Table 9: Spillover Effect: Oversea diversification

This table presents the difference-in-differences regression coefficients along with their corresponding t-statistics (in parentheses) for the impact of the robo-advising service on the allocation to S&P500, US Factor Style Investment, and Chinese ADRs in US and HK. The coefficients are in percentage (%). Statistical significance is denoted by *, **, and *** for the 10%, 5%, and 1% significance levels, respectively.

Panel A			
	Allocation to S&P500	Allocation to US Factor Style Investment	Allocation to Chinese ADRs
Treated*Post	1.52***	2.04***	1.81***
	(3.50)	(4.01)	(3.71)
Treat	0.09	0.12	0.10
	(0.19)	(0.24)	(0.23)
Panel B			
LLM (AI) Interaction effects			
Treated*Post*LLM	3.48***	3.07***	4.71***
	(4.84)	(5.35)	(4.90)
Treated*Post	-0.06	-0.25	-0.09
	(-0.11)	(-0.42)	(-0.23)

G Online Appendix

G.1 How robo-advising function

G.1.1 Example 1: Mutual funds

Consider an investor, Jane, who currently holds three mutual funds: Mutual Fund A, Mutual Fund B, and Mutual Fund C. Jane is curious about the performance of these funds and whether she should make any changes to her portfolio. She decides to use a robo-advisor for a detailed analysis and personalized recommendations.

Upon analyzing Jane's holdings, the robo-advisor provides the following insights:

Mutual Fund A: Alpha = 3%, Beta = 1.2, HML (High Minus Low) = 0.4, SMB (Small Minus Big) = 0.5, t-statistic = 1.8. Although Mutual Fund A has a positive alpha, the t-statistic below 2 indicates that the alpha is not statistically significant, suggesting that the outperformance may not be due to skill. Mutual Fund B: Alpha = -1%, Beta = 0.8, HML = -0.2, SMB = -0.1, t-statistic = -0.5. Mutual Fund B has a negative alpha and an insignificant t-statistic, confirming underperformance relative to its benchmark. Mutual Fund C: Alpha = 2%, Beta = 1.0, HML = 0.3, SMB = 0.2, t-statistic = 2.5. Mutual Fund C has a positive and statistically significant alpha, indicating that the fund manager has likely added value beyond market returns.

The robo-advisor calculates the overall portfolio's alpha and beta, adjusting for factor returns such as HML and SMB:

Overall Portfolio Alpha: 1.33% Overall Portfolio Beta: 1.0 (indicating market-level volatility) Overall Portfolio HML: 0.16 Overall Portfolio SMB: 0.2

Based on these insights, the robo-advisor makes the following recommendations to Jane:

Reevaluate Mutual Fund A: Despite its positive alpha, the t-statistic of 1.8 suggests that the alpha is not statistically significant. The robo-advisor recommends considering reallocating funds from Mutual Fund A to investments with more reliable performance metrics. Sell Mutual Fund B: Given its negative alpha and insignificant t-statistic, Mutual Fund B

is clearly underperforming. The robo-advisor advises selling this fund and reallocating the capital. Maintain Position in Mutual Fund C: With a positive and statistically significant alpha, Mutual Fund C demonstrates strong management performance. The robo-advisor suggests keeping this fund. Introduce Low-Cost ETFs or Index Funds: To enhance cost efficiency and maintain diversification, the robo-advisor recommends reallocating the funds from Mutual Fund B and potentially Mutual Fund A into low-cost ETFs or index funds that track broad market indices. For example, it suggests ETF X (Usually Hushen 300 ETFs, similar to S&P 500 ETFs), with a beta of 1.0, minimal exposure to HML and SMB, and very low fees, to capture market returns efficiently.

By following these recommendations, Jane can optimize her portfolio, ensuring she benefits from statistically significant high-alpha mutual funds while reducing costs and maintaining market exposure through ETFs or index funds. This balanced approach helps her achieve better returns while managing risk and minimizing fees.

G.1.2 Example 2: Equity

Consider an investor, John, who currently holds three individual stocks: Stock X, Stock Y, and Stock Z. John is interested in understanding the performance of these stocks and whether he should make any adjustments to his portfolio. He decides to use a robo-advisor for a detailed analysis and personalized recommendations.

Upon analyzing John's holdings, the robo-advisor provides the following insights:

Stock X:

Alpha = 4% Beta = 1.3 t-statistic = 1.9 Stock X has a positive alpha, indicating outperformance relative to its benchmark. However, the t-statistic is slightly below 2, suggesting that the alpha is not statistically significant, and the outperformance may not be due to skill. Stock Y:

Alpha = -2% Beta = 0.7 t-statistic = -1.2 Stock Y has a negative alpha, indicating underperformance relative to its benchmark. The t-statistic is also negative and not significant,

confirming that the stock is likely underperforming. Stock Z:

Alpha = 3% Beta = 1.0 t-statistic = 2.4 Stock Z has a positive and statistically significant alpha, indicating that the stock has likely provided returns above the market due to some stock-specific factors. The robo-advisor calculates the overall portfolio's alpha and beta based on the weighted average of the individual stocks:

Overall Portfolio Alpha: 1.67% Overall Portfolio Beta: 1.0 (indicating market-level volatility)

Based on these insights, the robo-advisor makes the following recommendations to John:

Reevaluate Stock X: Despite its positive alpha, the t-statistic of 1.9 suggests that the alpha is not statistically significant. The robo-advisor recommends monitoring Stock X closely and considering reallocating funds if better investment opportunities arise.

Sell Stock Y: Given its negative alpha and insignificant t-statistic, Stock Y is underperforming. The robo-advisor advises selling this stock and reallocating the capital to more promising investments.

Maintain Position in Stock Z: With a positive and statistically significant alpha, Stock Z demonstrates strong performance. The robo-advisor suggests keeping this stock in the portfolio.

Introduce Diversification Through ETFs: To enhance diversification and reduce individual stock risk, the robo-advisor recommends reallocating the funds from Stock Y into low-cost ETFs or index funds that track broad market indices. For example, it suggests ETF A (Usually Hushen 300 ETF, similar to S&P 500 ETFs), which has a beta of 1.0 and very low fees, to efficiently capture market returns.

By following these recommendations, John can optimize his stock portfolio by retaining stocks with statistically significant positive alpha, reducing exposure to underperforming stocks, and enhancing diversification with low-cost ETFs. This approach helps him achieve better returns while managing risk and minimizing fees.

Table A.1: Comparison of Different Chatbot and Human Assistant Results on Financial Literacy Impact (Robustness Checks)

This Table shows the difference-in-differences regressions coefficients and t-statistics (in parenthesis) in the Table. The Treatment equals one if the investors intend to receive the robo-advising service and actually received the robo-advising service. The Control equals one if the investor intend to receive the robo-advising service but was not selected to receive the service. Post equals one for months after 2023 December. Alphas are the Three Factor Alpha, Three Factor + Momentum Alpha, Five Factor Alpha, and Five Factor + Momentum Alpha.

Group	**FF Three Alpha (%)**	**FF Three + MOM Alpha (%)**	**FF Five Alpha (%)**	**FF Five + MOM Alpha (%)**
Overall	0.20***	0.19***	0.18***	0.17***
	(4.72)	(4.62)	(4.51)	(4.44)
ChatGPT (LLM) Chatbot	0.44***	0.43***	0.41***	0.40***
	(6.27)	(6.03)	(5.89)	(5.70)
Without Chatbot	0.13**	0.12**	0.12**	0.10**
	(2.25)	(2.21)	(2.04)	(1.99)
Human Chat Assistant	0.14**	0.13**	0.12**	0.11**
	(2.41)	(2.35)	(2.29)	(2.24)
Normal Chatbot	0.12**	0.12**	0.10**	0.09**
	(2.29)	(2.24)	(2.19)	(2.03)

Table A.2: Functionality Comparison of ChatGPT (LLM) Chatbot, Human Chat Assistant, and Normal Chatbot

Functionality Comparison of ChatGPT (LLM) Chatbot, Human Chat Assistant, and Normal Chatbot. Other Y variables: Number of question asked, positive and negative sentiment from the the investors' questions

Functionality	**ChatGPT (LLM) Chatbot**	**Human Chat Assistant**	**Normal Chatbot**
Explaining Beta	Comprehensive, contextual understanding (Uses advanced language models)	Good, but may lack consistency (Relies on human expertise)	Limited, basic explanations (Relies on predefined scripts)
Explaining Alpha	Detailed, data-driven insights (Access to large datasets)	Good, with practical examples (Uses personal experience)	Basic, often superficial (Limited to script accuracy)
Answering Financial Queries	Accurate, quick responses (High processing power)	Reliable, but slower (Depends on human availability)	Variable accuracy, slower (Limited by scripting)
User Interaction	High engagement, adaptive (Adapts to user input)	Personalized, but variable (Depends on human mood)	Standard responses, less engaging (Predefined interactions)
Learning and Improvement	Continuous learning, updates (AI model updates)	Static, based on human learning (Depends on individual growth)	Limited, no self-improvement (Requires manual updates)