AI, Investment Decisions, and Inequality

Alex G. Kim[1] David S. Kim[2] Maximilian Muhn[3]
Valeri V. Nikolaev[4] Eric C. So[5]

December 2024

Abstract

Using two large-scale experiments, we explore how generative AI shapes both overall performance and disparities in investment tasks. When investors are given AI summaries aligned with their sophistication, they become better at processing financial information and making investment decisions. Conversely, misaligned summaries generally have an adverse effect, suggesting AI's ability to benefit investors hinges on personalization of information. We also show AI's benefits accrue disproportionately to individuals with higher financial expertise, which stems from an inherent tradeoff between accessibility for less sophisticated investors and technical precision used by more sophisticated investors. Together, our findings suggest AI improves performance on investment tasks, on average, but also underscore the potential for these tools to widen rather than limit existing performance gaps.

Keywords: Generative AI, investment decisions, information processing, investor sophistication

JEL Codes: C45, C93, G12, G14, M41

[1]The University of Chicago, Booth School of Business. alex.kim@chicagobooth.edu
[2]Massachusetts Institute of Technology, Sloan School of Management. dskim@mit.edu
[3]The University of Chicago, Booth School of Business. maximilian.muhn@chicagobooth.edu
[4]The University of Chicago, Booth School of Business. valeri.nikolaev@chicagobooth.edu
[5]Massachusetts Institute of Technology, Sloan School of Management. eso@mit.edu

We are grateful to Irene Tan, who provided stellar research assistance. We appreciate insightful comments from Eddy Cardinaels, Emanuele Colonnelli, Jared Curhan, Ethan Mollick, Lilach Mollick, Drazen Prelec, Pietro Ramella, Kristi Rennekamp, Haresh Sapra, and workshop participants at the University of Chicago and MIT Behavioral Research Lab. The authors gratefully acknowledge financial support from the University of Chicago Research Support Center, the Fama-Miller Center for Finance Research, and the Stevens Doctoral Program at the University of Chicago Booth School of Business. This study's AEA RCT registry number is AEARCTR-0014588. The online appendix is available here.

1 Introduction

Generative artificial intelligence (AI) has been adopted at an unprecedented rate (e.g., Bick et al., 2024), eclipsing the historical uptake of computing and the internet. The financial industry has been at the forefront of this wave (e.g., Kelly and Xiu, 2023). Yet, there is little academic evidence on the effects of generative AI on individual investment decisions. Motivated by this sparsity and the proliferation of AI into the mainstream, we seek to address two central questions. First, do these tools enhance investors' ability to process information and make improved financial decisions? Second, does access to generative AI narrow or widen performance gaps across investors? We address these questions by performing two large-scale experiments that examine how generative AI shapes individual investors' information processing and investment choices.

Many see generative AI as a powerful democratizing force (e.g., Desai, 2023). By making specialized knowledge more accessible, AI helps users understand complex information that was previously outside the reach of non-experts (Dell'Acqua et al., 2023). Early research supports this optimistic view. Studies show that in tasks such as writing and programming, generative AI improves overall performance but also enables novice users to produce outputs that near expert quality, effectively leveling traditional skill-based performance gaps (e.g., Brynjolfsson et al., 2023; Noy and Zhang, 2023; Peng et al., 2023).

But financial decision-making presents unique challenges setting it apart from other domains where AI has proven to be democratizing. In writing or programming tasks, the AI's output largely represents the end goal. Financial decisions work differently, however. An AI-generated analysis of firm performance is not valuable on its own because investors must still be able to interpret the information to take appropriate action. Consider news about, e.g., uncertain tax positions, hedging transactions, or goodwill impairments, to name a few. Although AI can summarize and clearly explain this information, it becomes valuable only when investors understand how these concepts inform their decisions. Thus, our first hypothesis is that AI improves information processing in investment tasks, on average, but that this improvement hinges on AI alignment with an investor's financial sophistication.

Investment sophistication is also likely to shape how different investors benefit from AI. For some investors, AI likely acts primarily as a "filter". It helps them cut through boilerplate language and marketing jargon, zeroing in on value-relevant information. These investors can process more information faster. Yet for other investors investors, AI likely serves a different function—making complex information more digestible, often by simplifying it. However, this

simplification comes at a cost: when AI makes technical concepts more accessible, it is bound to omit subtle but potentially crucial details. For example, explaining revenue recognition in basic terms could obscure assumptions that sophisticated investors specifically seek out. To the extent that more sophisticated investors use AI as a filter and less sophisticated investors use AI as a simplifier, it is ambiguous ex ante whether AI widens or narrows disparities across investor groups. This ambiguity motivates our study.

Our experiments focus on a critical aspect of investment decision-making: the processing of earnings announcements. Specifically, we analyze how investors process earnings conference call transcripts, which provide rich information about firms' performance including summaries of past performance, outlook for the future, and dialogue between firm management, analysts, and investors. Our study examines real earnings call transcripts from publicly traded firms, while masking their identity.

The first experiment leverages AI summaries of earnings news—the most common application of AI in financial information processing (e.g., Ebert, 2023; Kim et al., 2024b). We recruit 1,800 participants and assess their sophistication in investing using a set of questions akin to FINRA's financial literacy test.[1] We then randomly assign each participant to receive either full earnings conference call transcripts (hereafter "raw transcript"; the control group), AI summaries written for sophisticated investors (hereafter "advanced summaries"; the first treatment group), or AI summaries written for less sophisticated investors (hereafter "simplified summaries"; the second treatment group).[2] Accordingly, we use a 2×3 experimental design that segments more versus less financially sophisticated participants and our three randomized conditions (see Section 2 for more details).

Our second experiment provides participants with flexibility in how they use generative AI. Instead of pre-written summaries, we provide participants with an AI assistant—a chatbot that is pre-loaded with the earnings call transcript. This allows us not only to measure their performance on investment tasks but also to track engagement with AI chatbots. For example, we can study how AI interactions (questions) vary with user sophistication. The control group, as before, works with raw transcripts of the earnings announcement conference call.

Both experiments examine two key aspects of investor behavior. First, we test how well

[1] We include a set of questions designed to assess individuals' basic financial knowledge required to process information released during earnings calls and make informed, rational investment decisions. Using the median participant score, we divide participants into a "more financially sophisticated" group and a "less financially sophisticated group."

[2] Each participant is required to separately evaluate two matched firms (same treatment condition). We use two firms per participant, rather than one, because the investment question at the end of the survey requires allocating capital between two stocks.

participants process information. Specifically, we test how well they can assess the general sentiment of the earnings announcement and forecast the direction of future earnings changes. In line with Chen et al. (2022), we compare the participants' judgments with the actual price response to the earnings announcement and whether subsequent earnings rose or fell. Second, we evaluate the impact of generative AI on participants' investment decisions. Participants allocate a hypothetical $1,000 between two stocks or choose not to invest and hold the initial endowment as cash, across both short-term (1-day) and long-term (1-year) investment scenarios. We assess the quality of these decisions by comparing the returns and Sharpe ratios of users' selected portfolios versus ex post optimal portfolios.

Our approach differs from typical laboratory experiments in that we use a large set of actual earnings announcements. We analyze releases from 2010 to 2022, carefully matching pairs of companies.[3] This design improves upon studies using hypothetical scenarios (Lian et al., 2018; Heeb et al., 2022) or limited firm samples (Bonsall et al., 2017). With actual market outcomes as our benchmark, our setup allows us to gauge whether participants make better decisions on average.

Our first main test shows that when AI summaries align with participants' level of financial sophistication, it significantly improves their decision-making. Alignment in our context means that more sophisticated participants receive the advanced AI summaries, and vice versa for less sophisticated participants. These aligned summaries boost the accuracy of earnings predictions by 7% for less sophisticated participants and 18% for more sophisticated ones. Similarly, we find that aligned AI summaries improve participants' ability to correctly assess whether stock prices were more likely to rise or fall in response to the earnings announcement.

We find that misaligned summaries do not help and sometimes hinder users' performance. Less sophisticated users are no better off from receiving the advanced summaries written for more sophisticated investors compared to receiving the raw transcript. Similarly, sophisticated users do not benefit from simplified summaries written for less sophisticated investors; they appear at least as well off as when processing the raw transcripts. This result speaks our motivating question of what specific aspects of human-AI interactions shape investment decisions. Specifically, our results suggest that tailoring AI outputs to align with an individual's financial sophistication is essential to generative AI having a positive impact. This result also spotlights large-scale personalization as a key differentiating factor, underscoring

[3]We select pairs that operate in the same four-digit SIC industry, and have similar market capitalization, book-to-market ratio, and profitability. We also force pairs to have the same earnings release date, which allows us to neutralize differences in industry or market performance that vary over time.

a fundamental distinction between generative AI and previous technological advances like computing and the internet in how they influence individual investment decisions.

Providing users with aligned summaries also raises the confidence participants express in their decision-making. This heightened confidence dovetails nicely with our earlier findings that aligned summaries improve performance in forecasting earnings and processing earnings news. However, sophisticated participants receiving misaligned summaries feel less confident, consistent with sophisticated investors being hampered by the absence of information that they might normally use to process earnings news. We also find that providing summaries intuitively helps participants make faster decisions. These findings are noteworthy in the context of investing, where investors seek to trade on information before it is fully reflected in stock prices.

A second key result of our paper is that AI summaries significantly shape performance on portfolio allocation tasks. Aligned AI summaries enhance participants' investment decisions across both the long- and the short-term horizons. Consistent with the idea that generative AI can widen access to financial markets, we find that less sophisticated participants make significantly better investment decisions with simplified summaries.

We also show that aligned summaries benefit participants more when the investment task is more complex. We conduct these tests by condensing our data to the earnings announcement level. For each announcement, we measure the difference in participants' average performance when treated with aligned summaries relative to raw transcripts. We then regress this difference on attributes of the firm and of the earnings announcement itself. We find that aligned AI summaries benefit participants more in smaller, more volatile firms and in instances when firms report losses or larger absolute earnings surprises. These findings suggest that an important channel through which AI shapes financial decision-making is by reducing the complexity of information processing.

Our findings on participants' investment decisions also reveal a paradox. While AI may make markets more accessible by improving the performance of less sophisticated investors, it simultaneously widens, rather than reduces, inequality across investor groups. For example, in our one-year portfolio allocation task, sophisticated participants receiving aligned summaries improve performance by more than 30% relative to the average performance, whereas the same estimate for less sophisticated participants falls below 10%. In the context of our experiment, this translates into an economically large difference in return performance: aligned AI increases one-year portfolio returns by 9.6% for more sophisticated investors and by 1.7% for less sophisticated ones, thereby widening the gap by 7.9%. Thus, our results are consistent with

generative AI amplifying pre-existing disparities in investment tasks by disproportionately benefiting more sophisticated investors.

We hypothesize that the widening performance gap across investor groups stems from an inherent trade-off–making AI summaries accessible to less sophisticated investors sacrifices technical precision. For example, whereas a simple summary might note "a reduction in profits," the advanced version specifies "operating margins declined due to higher input costs caused by supply chain disruptions"—creating a gap in the precision of the signals that participants receive. Consistent with this hypothesis, we find that the performance gap between more versus less sophisticated participants widens when more technical information is omitted from the simplistic summaries, such as discussion of R&D spending, share repurchases, and gaps between EBITDA vs. net profits. Thus, the efforts to make financial information more accessible via AI come at the cost of reduced precision, potentially limiting AI's ability to fully democratize financial decision-making.

Using results from our second experiment, we show that more versus less sophisticated users interact differently with AI chatbots. Specifically, we show that while these groups ask roughly the same number of questions, more sophisticated participants are more likely to ask for summaries that aggregate various components of earnings announcement news. Additionally, more sophisticated users are more likely to ask questions about specific financially oriented topics such as merger activities or capital expenditures. Less sophisticated participants, by contrast, are more likely to ask questions outside the scope of the earnings announcement.

Our second experiment also provides corroborating evidence that access to generative AI disproportionately benefits more sophisticated participants. Specifically, in revisiting the same tasks studied in the first experiment, more sophisticated participants exhibit greater gains from AI when assessing earnings news sentiment, predicting earnings changes, and forming investment portfolios. Combined with our earlier findings, the results suggest that differential ability to ask questions is another channel through which access to generative AI contributes to disparities across investor groups. Moreover, the consistency of our inferences across the first and second experiments suggests that the main findings of our paper are not due to a specific modality of human-AI interaction.

We conclude the paper by showing that our results are remarkably robust to different sampling and research design choices. For example, to address concerns that participants used outside information, we exclude participants who identified the company name through context clues, though less than 2% attempted identification and only 1% were correct. Additionally, we track experiment window activity and exclude participants with significant

off-task time (more than 10 seconds) to ensure engagement with provided materials only.

Our central contribution is to provide the first large-scale experimental evidence on how generative AI affects financial understanding and investment performance. While previous work has explored AI's ability to process financial information (e.g., Chen et al., 2022; Lopez-Lira and Tang, 2023; Croom, 2024; Kim et al., 2024b,a), we instead examine how AI interacts with human decision-making to shape participants' performance on common investment tasks. Our results indicate that access to AI not only helps investors make sense of earnings news, but it also aids them in forecasting firms' performance in terms of both earnings changes and stock returns. The positive impact of AI on the performance of less sophisticated participants is particularly notable given prior evidence that retail investors have traditionally relied little on fundamental analysis (Barber and Odean, 2008; Kaniel et al., 2008). Our results suggest that AI makes earnings news accessible to a broader population of investors.

Second, we demonstrate that generative AI's impact depends heavily on the match between its outputs and user expertise. While prior work has documented AI's average productivity impacts across domains (Brynjolfsson et al., 2023; Dell'Acqua et al., 2023; Noy and Zhang, 2023; Peng et al., 2023; Cui et al., 2024; Toner-Rodgers, 2024), our results reveal substantial heterogeneity based on personalization. We find that AI tools enhance performance only when aligned with the user's skill level—misaligned outputs can overwhelm more novice investors with complexity while failing to provide sufficient depth for more expert investors. This highlights how thoughtful personalization, rather than one-size-fits-all deployment, shapes whether and to what extent AI helps individuals' performance on investment tasks.

Finally, our evidence informs and challenges common narratives about AI leveling the playing field across user groups. While AI has shown democratizing effects in domains like writing and programming (Noy and Zhang, 2023; Peng et al., 2023; Dell'Acqua et al., 2023), we find financial decision-making presents unique challenges. Unlike tasks where AI outputs more likely approximate the end goal, in finance, success depends not just on receiving information but on effectively interpreting and translating it into action. Consistent with this distinction, we find AI can amplify rather than reduce expertise gaps by disproportionately benefiting more sophisticated investors. This finding adds important nuance to ongoing discussions about AI's role in financial markets and suggests that AI's impact on individual task performance depends on the extent to which AI outputs approximate the end goal.

2 The Experiment

This section describes the main experiment, which aims to assess investors' information processing capabilities and stock investment efficiency. In Section 2.1, we describe the overall experimental setup. In Section 2.2, we provide the descriptive statistics of the companies used in our analyses. Lastly, in Section 2.3, we elaborate on how we measure information processing and investment efficiency in our experiment.

2.1 Experiment Design

To assess the effect of generative AI on investors' information processing in the context of financial markets, our experiments focus on a critical aspect of investment decision-making: the processing of earnings announcements. Earnings announcements are a high-powered setting as investors need to process new and important information to make trading decisions, often under considerable pressure (e.g., Savor and Wilson, 2016; DellaVigna and Pollet, 2009). In addition, retail investors actively trade around earnings announcements (e.g., Kaniel et al., 2012), making it a suitable setting to understand how generative AI influences (relatively less sophisticated) investors' decision-making. More specifically, we focus on how investors process information from earnings announcement conference calls, which cover a variety of topics, including summaries of past performance, an outlook for the future, and a dialogue between firm management, analysts, and investors.

Our experiment follows a 2×3 between-subject design. For the first dimension, participants are divided into two groups based on their level of financial sophistication. For the second dimension, each participant is randomly assigned to review the raw transcripts of anonymized earnings conference calls (control group), advanced summaries written for more sophisticated participants, or simplified summaries written for less sophisticated participants.

Survey description Our survey begins by outlining the high-level purpose and scope of our experiment (understanding investors' decisions surrounding firm disclosures) as well as asking for the participants' consent. The survey then consists of three main parts: (i) assessing participants' financial expertise, (ii) information processing evaluation, and (iii) investment efficiency evaluation.

The first part of the survey assesses participants' level of financial expertise. We construct seven questions to evaluate whether participants have basic financial knowledge. We start with some questions that are adapted from the Financial Industry Regulatory Authority (FINRA)'s

survey to evaluate financial literacy (FINRA, 2009; Krische, 2019). We then add other newly developed questions to more directly examine whether respondents are likely to be able to utilize earnings news to inform their investment decisions. This section of the survey does not differ between treatment arms, but the answers from this part are used to divide the population into more versus less financially sophisticated groups.

The second part evaluates investors' ability to accurately process earnings announcement information for two firms. Forming beliefs about firms' future profitability is one of the primary objectives for investors processing financial information. As such, this part of the survey is designed to simulate how investors interpret earnings information and form beliefs about companies' prospects. We present information for two distinct firms (rather than only one) because participants are prompted to make an investment allocation decision between the two firms in the third part of the survey (more on this below). Each participant reviews the information and answers all questions for the first firm before proceeding to the second firm, where they repeat the same process.[4]

Participants in the control group receive a copy of an anonymized earnings conference call transcript, whereas participants in the treatment groups receive either an advanced summary or a simplified summary. (The assigned treatment condition is the same for both firms.) Participants are asked to review the provided material, assess its content, and then answer questions about their earnings expectations and the sentiment of the news. Specifically, we ask participants three questions about their earnings expectations and two questions about the nature of the news.

To assess participants' expectations about future earnings, we ask them to select a number between -5 (highly likely EPS will decrease in the following year compared to the current period) and $+5$ (highly likely EPS will increase in the following year compared to the current period). Participants then report their confidence level in these expectations on a scale from 0 (completely uncertain) to 1 (perfectly sure) and are also required to write a short narrative explaining the rationale behind their predictions.[5]

We next assess participants' ability to summarize the implications of the earnings announcement for concurrent stock prices. We do so by instructing participants to evaluate whether the sentiment is positive or negative, using a scale from $+5$ to -5. They also report

[4]We describe the selection of the firms in more detail in Section 2.2.

[5]We explicitly instruct the participants not to use any generative AI tool to prepare the narrative answers. We also note in the survey that if our algorithm detects an answer generated by an AI tool, the respondent's compensation may be negatively impacted.

their confidence in their sentiment assessment. Finally, as another secondary outcome variable, we also record how much time each participant spent answering these questions.

The third part of the experiment is an investment task. We ask participants to allocate $1,000, a hypothetical amount, to three different investment options: the stocks of the two companies they reviewed in the previous part or cash. They have to make this investment decision under two scenarios: a short-term and a long-term investment horizon. For the short-term horizon, participants are asked to form an investment portfolio that they must hold only for one day.[6] For the long-term investment question, they have an investment horizon of one year. Having completed their allocation plan, participants must also write a short rationale explaining their asset allocation decision. Similarly to the previous part of the survey, we also measure the time spent by the participants to make decisions.

Finally, to rule out that participants inadvertently identify the firm identities, we ask respondents whether they were able to identify the firm names from the anonymized documents at the end of the survey. We also ask participants whether they can guess the ultimate goal of the study. The full survey is provided in the Online Appendix.

Summary generation A key ingredient in our experiment is the use of AI-generated summaries. Our focus on AI summaries is motivated by the widespread proliferation of real-world institutional products that use AI to summarize earnings announcements, such as those provided by Bloomberg, AlphaSense, and Verity, among many others. For this task, we use OpenAI's `GPT-4o` and generate two versions of earnings call summaries: simplified and advanced.

To generate simplified summaries, we instruct the model to generate content targeted towards individuals with little financial knowledge. Further, the model is instructed to explain difficult financial terminologies and to omit excessively complex economic events that the target audience is less likely to comprehend. By contrast, to generate advanced summaries, we define the target audience as individuals with college-level financial knowledge and stock investing experience. We do not impose any restrictions on the content that the model can include in the advanced summaries. See the Online Appendix for the full prompts we use to generate these summaries.

For both types of summaries, we explicitly ask the model not to produce direct investment

[6]We select earnings conference calls that occur outside regular trading hours and assume that participants initiate their trades when the market opens in the most adjacent available trading day—either on the following day for after-hours calls or the same day for pre-market calls.

recommendations and not to include any external information from its knowledge base. We also instruct the model to produce narrative summaries rather than bullet-pointed documents. This ensures that the summaries and their corresponding raw transcripts are presented in the same format.[7]

Finally, we anonymize the names of people, regions, and products for both raw transcripts and AI-generated summaries. This mitigates the concern that participants may have prior knowledge about the companies and make predictions based on such knowledge.[8]

Recruiting We use Prolific's online survey platform to recruit the participants. We require that the participants reside in the United States, have English as their primary language, and have stock investment experience. Participants are randomly assigned to one of the two treatment groups or to the control group. We recruit 600 participants for each group, totaling 1,800 participants. Each participant is compensated $5 for a 20-minute survey.

To help gauge the generalizability of our findings, we provide demographic information about the participant pool in our sample in Table 1, Panel A. Among our respondents, 42.2% are male, 67.3% identify themselves as white, and 50% are 42 years old or older. In column (2), we provide the same demographic statistics from the 2023 U.S. Census data.[9] Table 1, Panel A shows that although our pool of respondents is generally comparable to the entire US population, males and minority ethnicities are slightly underrepresented.

Treatment assignment We randomly assign participants to one of three document types: a raw transcript, an advanced summary, or a simplified summary. Each participant receives two stocks to analyze, both derived from the same type of document (e.g., advanced summary) to maintain consistency in the exposure to the treatment. We randomize participants into one of the three groups via our survey tool Qualtrics.

We also divide the respondents into two groups based on their level of financial sophistication. As noted above, we rely on the participants' scores using the questions in

[7]We do not restrict the length of the output text. Specifically, we set `top_p` to be 1 and `temperature` to be 0 to reduce variation in model answers.

[8]To assess this concern, we ask the participants at the end of the survey whether they were able to identify the names of the companies used in the experiment. Overall, very few participants were successfully guessing the name of the company (e.g., from menus or context clues). Our results are robust when excluding these observations.

[9]Our participants are at least 18 years old, and the oldest participant is 75 years old. To match the distribution, we confine the population to the same age cohort.

the first part of our survey.[10] The maximum possible score is eight, and the minimum is zero. Participants' scores are not shared with them. Also, participants are informed that their scores on these questions will not affect their compensation.

We split the participants into two groups based on the median value of the actual score distribution. The test scores roughly follow a normal distribution with a median score of 4 and a mean score of 3.79. The standard deviation of participant scores is 2.01. The participants who scored four or above are classified as the "more" sophisticated group and vice versa.

Our survey took approximately 20.97 minutes to complete on average, and we considered participants who take more than 63 minutes to have timed out. After applying this time-out rule, our survey return rate was 79.8%. The remaining 20.2% of the respondents did not return the survey or exceeded the time limit. A potential concern is that the timed-out respondents were the least sophisticated participants who struggled to finish the survey within the allotted time. To address this, we compared the average financial knowledge scores of timed-out participants to those who completed the survey and found no statistically significant differences.[11]

2.2 Firm Selection

We provide participants with actual (anonymized) earnings call information. Using real firm data, as opposed to hypothetical earnings announcements, offers the key advantage of enabling us to assess whether participants make better decisions by benchmarking them with actual outcomes. Additionally, this approach more closely mirrors the investment tasks that individuals encounter in practice.

To select a pool of candidate earnings calls, we focus on stocks trading on the NYSE or Nasdaq during the 2010-2022 period and restrict the sample to earnings calls from the fourth fiscal quarter that occur outside trading hours. The latter filter prevents earnings call news from being reflected in stock prices before the end of the call. We also exclude stocks that did not have the market open within 24 hours of the call (e.g., announcements made on Friday evenings). Lastly, we exclude stocks in the finance industry (i.e., SIC codes 6000-6999) and small-cap stocks (i.e., lowest size quintile each year).

[10] When forming this score, we exclude the demographic question on the educational background and the question on investment experience, which only validates our initial recruiting condition of investment experience. Participants generally receive 1 point for each correct answer. However, the question about whether the participants regularly follow earnings calls has three possible scores (2: regularly follow, 1: followed at least once, and 0: never follow). We have seven questions in total.

[11] The mean score of the timed-out group is 4.85, while the mean score of the participants used in the analysis is 4.33. The difference between the two groups is not statistically significant at a conventional level.

From this candidate pool, we select 200 stocks to construct 100 economically similar pairs. We first group stocks based on economic characteristics. Within each group, firms must be in the same four-digit SIC industry code, release earnings on the same date (to control for industry- and macroeconomic conditions), fall into the same quartile of market capitalization, the same quartile of book-to-market ratio, and the same quartile of return-on-equity, and have earnings per share of the same sign. If a group has more than two stocks, we select a pair that exhibits the largest return gap on the next trading day to improve the differences in the communicated information, which arguably increases the power of our investment portfolio tests.[12]

Table 1, Panel B compares the characteristics of our sample against the broader population of stocks. We report the differences in means between the two groups for market capitalization, book-to-market ratio, ROE, and two measures of earnings surprise. The latter is based on either the random walk expectation or the analysts' consensus forecast. The characteristics of the stocks used in our experiment are generally not statistically different from those of the population. One exception is the book-to-market ratio, which is lower in our sample, consistent with our companies being more mature on average.

2.3 Measurements

We ask participants (i) to assess the overall sentiment of the earnings call transcript, (ii) to forecast the direction of future earnings changes, and (iii) to form short- and long-term investment portfolios. We then compare their responses to actual earnings and realized returns to evaluate their decisions' quality. This setup creates an important challenge. On average, earnings calls contain useful information about future earnings and returns (Frankel et al., 2022; Kim et al., 2024b). However, for any given firm (decision), it is not possible to attribute differences between predicted and actual outcomes solely to participants' information processing abilities. This is because earnings and returns are also driven by significant unanticipated factors not known to investors at the time of decision (e.g., related to information not discussed during the earnings call).Put differently, for any given earnings call, we do not know what the "correct" answer is even if we observe the realized value.

Our experimental design addresses this challenge by evaluating the decisions *on average*, using a relatively large sample of actual earnings release events (200 firms). This approach allows us to average out the effect of unanticipated events on participants' choices. Put

[12]In Section 6, we perform subsample analyses based on pairs with high and low return gaps. We find qualitatively similar results for both subsamples, which mitigates concerns that our findings stem from selecting pairs with distinct earnings news.

differently, this approach ensures that our measurements of participants' predictive accuracy and investment efficiency can be interpreted as unbiased estimates of their true information-processing abilities.[13]

Earnings prediction We compare the participants' answers with the actual values to assess the task accuracy. The accuracy of participant j's earnings prediction based on the conference call i is calculated as follows:

$$\begin{aligned} \text{Accuracy}_{ij} &= -1 \times |\mathbb{1}(\text{Increase})_i - \text{Scaled Answer}_{ij}| \\ &= -1 \times \left| \mathbb{1}(\text{Increase})_i - \left(\frac{\text{Answer}_{ij}}{10} + 0.5 \right) \right|, \end{aligned} \tag{1}$$

where $\mathbb{1}(\text{Increase})_i$ is an indicator that equals one when the firm's EPS corresponding to call i is larger than that of the following year and zero otherwise. As a participant's answer (Answer_{ij}) ranges from -5 to $+5$, we scale it with ten and add back 0.5 to have it distributed between 0 and 1. We multiply the absolute difference between $\mathbb{1}(\text{Increase})_i$ and participants' scaled answer by -1 such that higher values indicate greater accuracy.[14] When a participant's response moves toward the correct prediction, Accuracy_{ij} increases towards zero. If a participant makes an entirely incorrect prediction, Accuracy_{ij} takes a value of -1. As secondary outcome variables, we also obtain self-reported confidence (from 0 to 1) as well as the time taken (in seconds) to complete the corresponding answers.

In addition to evaluating participants' numerical answers, we also leverage a large language model and neural network to evaluate the quality of the participants' *narrative* answers. This approach leverages the model's ability to learn key economic patterns that explain changes in future earnings, thereby enabling it to assess the quality of participants' answers based on these learned patterns.

To identify the economic drivers of changes in future earnings, we use a large dataset of conference call transcripts and subsequent earnings changes. Specifically, we provide GPT-4o with information about whether earnings increased along with a conference call transcript and

[13] An analogy can be drawn to machine learning models: Even the best models trained on a given information set will not predict every single outcome correctly, but will outperform alternative approaches in larger samples on average. Likewise, our design focuses on aggregated performance across different treatment conditions over many earnings release events rather than individual accuracy on any single observation.

[14] We focus on the continuous scoring of participants' responses to ensure a sufficient variation in the dependent variables. However, our main results remain robust when employing a binary scoring approach, where a score of one is assigned if the direction of the outcome aligns with the direction of participants' responses, and zero otherwise.

instruct it to generate economic explanations for the observed change based on the transcript content. We then use vector representation (embeddings) of these generated texts as inputs to train a machine-learning model (see Online Appendix for detailed steps). This process allows the model to learn the economic narratives that influence changes in earnings.

Using this trained model, we evaluate the quality of participants' text responses. The model assigns a score (a softmax probability value transformed to be between 1 and -1) of 1 (-1) when a participant's answer is well-grounded economically to explain positive (negative) changes in future earnings. It assigns a score of zero when a participant's answer lacks economic substance, or is unrelated to earnings changes. Using these scores, we define the quality of textual responses as follows:

$$\text{Text Quality}_{ij} = -1 \times \left| \mathbb{1}(\text{Increase})_i - \text{score}\left\{ \hat{f}(c_{ij}) \right\} \right|, \tag{2}$$

where $\hat{f}(\cdot)$ represents the trained model, and c_{ij} denotes the numerical embedding of participant j's explanation for earnings call i. The Text Quality score ranges from -1 to 0, with higher values (closer to 0) indicating higher-quality text responses in predicting future earnings changes.[15]

Sentiment Along with earnings prediction, we also ask participants to assess whether the news has a positive or negative effect on the stock prices. As before, the answer options range from -5 to $+5$, and we use the same formula as Equation 1 to evaluate the accuracy. The only difference is that instead of $\mathbb{1}(\text{Increase})_{ij}$ being an indicator for EPS improvement, we use the sign of cumulative stock returns around the earnings announcement dates (from one day before to one day after the announcement) as a benchmark.[16] This approach is in line with Chen et al. (2022).

Investment efficiency In the third part of the survey, we ask participants to make an investment decision by allocating \$1,000 between the two firms or cash. We compute the relative investment weights w_A, w_B, and w_C by scaling the amount invested in stock A, stock B, and cash by \$1,000 ($w_A + w_B + w_C = 1$). Participants have to make separate asset allocation plans for short and long horizons. Short-term returns are measured from when the market

[15] For example, if a participant writes "Well, the earnings might increase because the CEO seems to be confident", the model assigns a moderately positive score (0.22). In contrast, if the participant explains "Based on strong growth in cash flows, its earnings are likely to improve," the model assigns a stronger positive score (0.57).

[16] $\mathbb{1}(\text{Increase})_{ij}$ takes a value of one when announcement returns are positive and zero otherwise.

opens on the trading day immediately following the earnings call until the market closes on the same day. We measure long-term returns as the cumulative buy-and-hold returns over 252 trading days immediately following the call.

We evaluate the efficiency of the investment based on the performance of this chosen portfolio relative to the hypothetical maximum realized returns and Sharpe ratios. In both investment scenarios, we calculate a participant j's realized portfolio (m) return and Sharpe ratio for a pair of stocks k:

$$\text{Portfolio Return}_{mjk} = r_A \times w_A + r_B \times w_B + 0 \times w_C, \tag{3}$$

$$\text{Sharpe Ratio}_{mjk} = \frac{r_A w_A + r_B w_B}{\sqrt{w_A^2 \sigma_A^2 + w_B^2 \sigma_B^2 + 2 w_A w_B \sigma_{AB}}}, \tag{4}$$

where σ_A and σ_B denote the standard deviations of stock A and B returns, respectively, and σ_{AB} is the covariance term.[17] The maximum portfolio returns (maximum Sharpe ratio) are computed as follows:

$$\max(\text{Portfolio Return})_{mk} = \max\{r_A, r_B, 0\}, \tag{5}$$

$$\max(\text{Sharpe Ratio})_{mk} = \max_{w_A, w_B, w_C} \frac{r_A w_A + r_B w_B}{\sqrt{w_A^2 \sigma_A^2 + w_B^2 \sigma_B^2 + 2 w_A w_B \sigma_{AB}}}. \tag{6}$$

We measure investment efficiency by the signed difference between the realized return (Sharpe ratio) of the participant's chosen portfolio and the maximum return (Sharpe ratio).

3 Does AI Improve Financial Decisions?

In this section, we examine whether AI-generated information can improve financial information processing and investment decisions. Sections 3.1 and 3.2 present the results of our first experiment. Section 3.3 examines cross-sectionally how our findings vary with firm characteristics.

3.1 Assessing Information Processing

Our first set of tests examines whether AI impacts investors' ability to process earnings announcement information through two key measures: earnings projections and news sentiment evaluation, as detailed in Section 2.1. We analyze how participants' performance

[17] For the short-term investment, we use 30-minute intervals to compute intraday return volatility; for the one-year investment, we use daily returns.

on these tasks varies based on both their assigned document type and financial sophistication level.

3.1.1 Empirical Design

We use the following main specification to assess the impact of our treatment arms on investors' information processing:

$$\begin{aligned}\text{Answer Quality}_{ij} = \beta_1 \text{Sophistication}_j + \beta_2 \text{Advanced Summary}_j + \beta_3 \text{Simplified Summary}_j \\ + \beta_4 \text{Sophistication}_j \times \text{Advanced Summary}_j \\ + \beta_5 \text{Sophistication}_j \times \text{Simplified Summary}_j + \zeta_i + \varepsilon_{ij}\end{aligned} \tag{7}$$

Answer Quality$_{ij}$ is the participant j's prediction accuracy (Equation 1) or the quality of the provided narratives (Equation 2) for earnings call i. Sophistication$_j$ is an indicator variable for financially sophisticated respondents. Advanced Summary$_j$ is an indicator variable that equals one when the participant j is assigned an advanced summary and zero otherwise. Simplified Summary$_j$ is defined analogously. As the difficulty of the earnings prediction can vary across earnings announcements, we control for earnings announcement (i.e., firm-by-quarter) fixed effects (ζ_i). Standard errors are clustered at the earnings announcement level.[18]

In the above specification, the baseline (i.e., intercept term) measures the performance of financially less sophisticated participants with raw earnings call transcripts (hereafter denoted as the baseline group). β_1 measures the relative performance of more sophisticated participants with raw transcripts relative to the baseline group. The coefficient β_2 represents the incremental effect of treating the less sophisticated group with advanced summaries, whereas β_3 is the effect of simplified summaries on the answer quality of the less sophisticated group. The incremental effect of advanced summaries on the more sophisticated group relative to the baseline group is given by $\beta_1 + \beta_2 + \beta_4$. Finally, the incremental effect of providing simplified summaries to the more sophisticated group is $\beta_1 + \beta_3 + \beta_5$. Figure 1 shows how the regression coefficients describe the various incremental effects.

3.1.2 Estimation Results

Earnings prediction Table 2, Panel A, column (1) shows that more financially sophisticated participants (Sophistication$_j$ = 1) make more accurate earnings forecasts

[18]We examine the robustness of our findings by examining two alternative specifications. First, we include interaction terms between Advanced Summary$_j$ and earnings announcement fixed effects, as well as between Simplified Summary$_j$ and earnings announcement fixed effects. Second, we address potential within-participant correlation by clustering standard errors at the participant level. Our results remain qualitatively similar.

FIGURE 1. **Regression Design**

	More Sophisticated	Less Sophisticated
Raw Transcript	β_1	0
Advanced Summary	$\beta_1 + \beta_2 + \beta_4$	β_2
Simplified Summary	$\beta_1 + \beta_3 + \beta_5$	β_3

Notes: This figure summarizes the interpretation of coefficients in our main specification in Equation (1).

compared to less sophisticated individuals on average. The magnitude of the coefficient is sizeable (0.027, t-statistic = 3.85), which is about 5.66% of the absolute mean dependent variable. Column (2) examines the average effect of two summary types on the prediction accuracy and indicates that advanced summaries improve average accuracy, whereas simplified summaries do not.

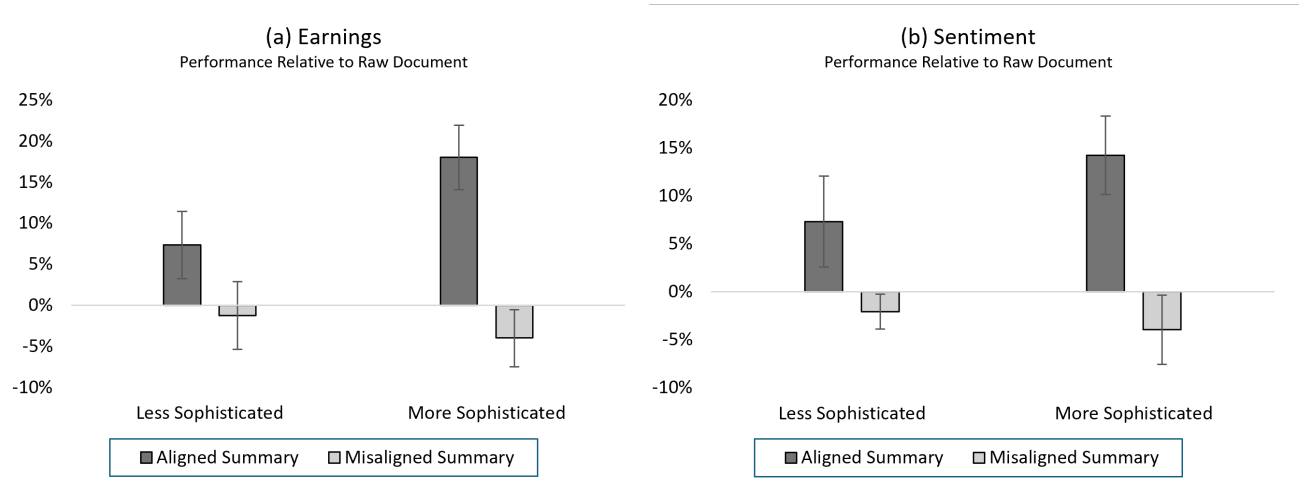
Column (3) of Table 2, Panel A documents the importance of aligning AI summaries with the financial sophistication of the participants. When summaries are aligned with user sophistication, prediction accuracy improves significantly. For less sophisticated investors, receiving simplified summaries enhances prediction accuracy (0.035, t-statistic = 2.94), while receiving advanced summaries reduces it (-0.006, t-statistic = -0.50), although the latter effect is statistically insignificant. Similarly, more sophisticated participants benefit from advanced summaries, but misaligned summaries (i.e., simplified summaries) harm their performance. Specifically, the interaction of Sophistication with Simplified Summary yields a significantly negative coefficient (-0.054), whereas its interaction with Advanced Summary results in a positive and highly significant coefficient (0.092, t-statistic = 5.29).

Columns (4) to (6) indicate that the informativeness of participants' written answers in earnings prediction tasks increases when they are provided with aligned summaries. Recall that participants provide a narrative rationale of at least 50 words, which we evaluate using a trained ANN model. The results broadly mirror the findings of our previous analysis. Sophisticated users produce more informative answers than less sophisticated users, and both groups benefit from AI summaries designed to match their financial sophistication.

Sentiment We next evaluate participants' ability to judge whether the earnings news will have a positive or negative impact on stock prices (i.e., the news sentiment) by substituting cumulative stock returns around the announcement for the EPS-based benchmark. Therefore, these tests parallel our earnings prediction tests but focus on sentiment evaluation accuracy. Table 2, Panel B, shows that the aligned summaries also improve participants' ability to

process the implications of the earnings news for stock price movements. Consistent with our results in Panel A, financially sophisticated participants, on average, outperform their less sophisticated counterparts (column (1)), demonstrating higher accuracy in recognizing the sentiment of the transcript. Advanced AI summaries significantly enhance accuracy on average (coefficient = 0.032, t-statistic = 3.17), while simplified summaries provide only a marginal improvement (coefficient = 0.010, t-statistic = 1.75). Importantly, we find that advanced summaries substantially benefit more sophisticated participants, while simplified summaries often hinder their performance. In contrast, less sophisticated participants gain more from simplified summaries but derive limited benefits from advanced summaries.

Figure 2. **Information Processing and AI Summaries**



Notes: This figure illustrates the effect of providing AI-generated summaries to the experiment participants. We follow Figure 1 and scale the cumulated coefficients by the absolute average of the dependent variable. Then, we subtract the respective outcome of raw transcripts for each group of participants. Error bars indicate 90% confidence intervals.

Visual illustration Figure 2 illustrates our main findings from Table 2 by plotting the incremental effects of aligned and misaligned summaries for each group of participants, based on column (3) in each panel (e.g., $\beta_2 + \beta_4$ indicates the incremental effects of aligned summary for more sophisticated participants). To enhance interpretation, we scale the corresponding cumulative coefficients by the absolute mean of the dependent variable. Panel A of Figure 2 shows that the aligned summaries increase prediction accuracy by approximately 7% for less sophisticated participants and 18% for more sophisticated participants. In contrast, misaligned summaries fail to provide similar benefits and may even degrade performance. Moreover, the incremental gains from aligned summaries are notably larger

for more sophisticated participants, a finding we examine further in Section 4.

Panel B of Figure 2 visualizes the results for the sentiment evaluation task based on Panel B of Table 2. Aligned summaries improve accuracy by approximately 7% for less sophisticated participants and 14% for more sophisticated participants, whereas the opposite holds for misaligned summaries.

Taken together, our findings in Table 2 and Figure 2 suggest that AI-generated summaries significantly improve investors' ability to process financial information, but only when these summaries are aligned with the user's level of financial sophistication. This underscores the potential of generative AI as a transformative tool in financial decision-making, provided it is deployed thoughtfully. Conversely, indiscriminate application of AI-generated summaries without consideration for the user's level of sophistication can diminish decision quality, highlighting the risks of mismatched AI outputs.

Confidence and response time We now turn to our secondary outcomes, focusing on how AI-generated summaries shape participants' self-reported confidence and the time they spend making decisions. Table 3 shows that AI summaries raise participants' confidence in their decisions while lowering the time they spend on investment-related tasks.

Panel A focuses on changes in participants' confidence based on AI-generated summaries. Participants rate their confidence in answering earnings prediction and sentiment assessment tasks on a scale from 0 (entirely uncertain) to 1 (entirely certain). Intuitively, columns (1) and (4) of Panel A indicate that financially sophisticated participants consistently report higher confidence levels. More importantly, columns (2) and (5) show that providing participants with AI-generated summaries boosts confidence. Notably, the coefficient on Advanced Summary exceeds that on Simplified Summary, indicating that aligned summaries enhance confidence despite their greater complexity. Interestingly, columns (3) and (6) indicate that higher confidence is primarily driven by the *less* sophisticated participants (when all variables are considered). In contrast, sophisticated participants become *less* confident when exposed to simplified summaries, consistent with the notion that misaligned information hinders decision-making. These patterns hold across both earnings prediction and sentiment assessment tasks, underscoring the importance of aligning AI output with users' financial expertise.

Panel B of Table 3 shows that AI summaries also improve the speed of financial decision-making in the earnings prediction task. Based on column (2), we estimate that aligned summaries reduce decision time for sophisticated participants by 75.62 seconds, or a 21.80% decrease relative to the average time. However, we do not observe the adverse effects for

misaligned summaries seen in Panel A—both types of summaries reduce decision times regardless of their alignment with user sophistication. Similar results emerge for sentiment assessment tasks, though the effects are statistically weaker.[19]

3.2 Assessing Investment Efficiency

In this section, we examine whether the improved information processing abilities translate into better asset allocation decisions.

3.2.1 Empirical Design

Recall that we ask participants to allocate \$1,000 across the two stocks and cash. To analyze the effect of AI summaries, we estimate the following at the participant level:

$$\begin{aligned}\text{Decision Quality}_{kj} = {} & \beta_1 \text{Sophistication}_j + \beta_2 \text{Advanced Summary}_j + \beta_3 \text{Simplified Summary}_j \\ & + \beta_4 \text{Sophistication}_j \times \text{Advanced Summary}_j \\ & + \beta_5 \text{Sophistication}_j \times \text{Simplified Summary}_j + \zeta_k + \varepsilon_{kj}\end{aligned} \tag{8}$$

$\text{Decision Quality}_{kj}$ measures portfolio performance of participant i (for the pair of stocks k) based on either portfolio returns or Sharpe ratios, as detailed in Section 2.3.[20]

3.2.2 Estimation Results

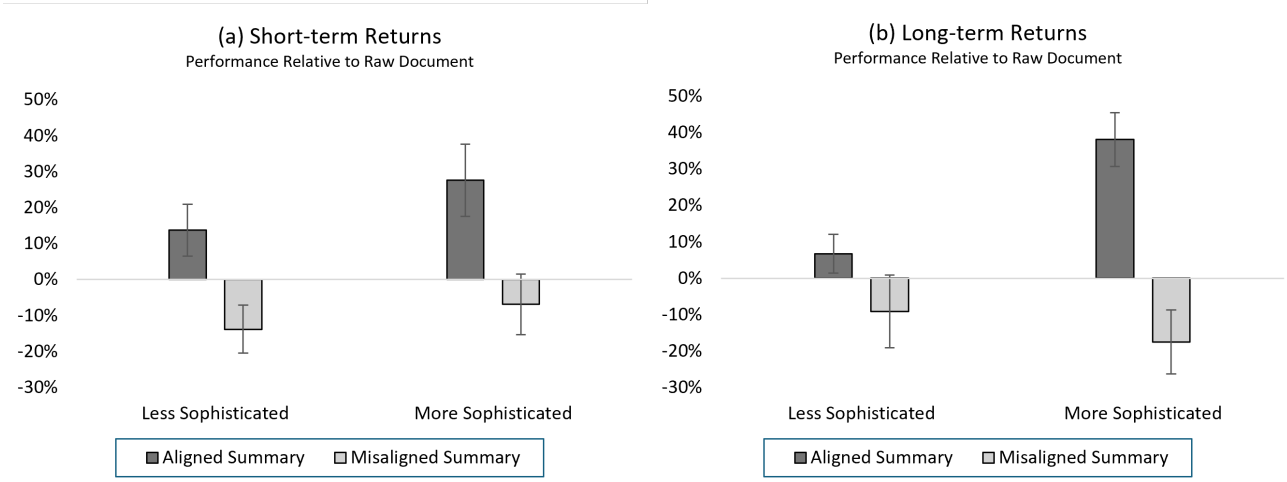
Short-term investment Table 4, Panel A focuses on short-term investment performance.[21]. As expected, sophisticated participants outperform their counterparts both in terms of portfolio returns (column (1)) and Sharpe ratios (column (6)). Providing AI-generated summaries alone does not significantly improve overall investment performance (columns (2) and (4)). However, when we account for the alignment between summary type and participants' financial sophistication (columns (3) and (6)), a clearer pattern emerges. Specifically, sophisticated participants benefit substantially from advanced summaries, achieving higher performance. Conversely, their performance deteriorates when based on simplified summaries, likely due to a lack of nuanced information. For less sophisticated participants, the opposite is true: they gain more from simplified summaries but do not experience similar benefits from advanced summaries.

[19] This may be because sentiment-related questions appear after earnings prediction tasks, reducing the need for participants to revisit the entire document.

[20] In untabulated tests, we include the interactions of $\text{Advanced Summary}_j$ with firm pair fixed effects and the interactions of $\text{Simplified Summary}_j$ with firm pair fixed effects. The results are qualitatively consistent.

[21] Participants are instructed to form a portfolio when the market opens on the following trading day and hold it for one day

Figure 3. **Portfolio Performance**



Notes: This figure illustrates the effect of providing AI-generated summaries to the participants. We follow Figure 1 and scale the cumulated coefficients by the absolute average dependent variable. Then, we subtract the outcome of raw transcripts for each group of participants. Error bars indicate 90% confidence intervals.

Long-term investment We next extend our analysis by examining the long-term investment portfolio of the participants. The results, which are presented in Table 4, Panel B, closely align with our earlier findings. More sophisticated participants who read advanced summaries achieve 9.6% higher returns in their annual portfolio compared to those who rely on raw transcripts. Less sophisticated participants who read simplified summaries still see a modest 1.7% improvement in long-term performance relative to raw transcripts. In contrast, misaligned summaries negatively impact portfolio performance. We visually summarize the results in Figure 3. The findings reinforce the critical role of alignment between AI-generated summaries and users' sophistication in enhancing long-term investment outcomes.

Short- vs. long-term investments Distinguishing between short- and long-term investment objectives requires a nuanced understanding of financial concepts, which might pose challenges for survey participants. Nevertheless, in the Online Appendix, we demonstrate that participants appear to understand these distinctions. First, participants use different portfolio weights for short- vs. long-term investments. Furthermore, when short-term weights are applied to long-term scenarios, the portfolio returns significantly deteriorate, particularly among the sophisticated participants (and vice versa). These results suggest that participants, especially those with greater financial sophistication, are capable of tailoring their investment strategies to align with different time horizons, highlighting their ability to adapt AI-based decisions to varying contexts.

3.3 When Does AI Improve Financial Decisions?

A distinguishing feature of our study is that we use financial data from 200 actual publicly traded companies. This feature allows us to perform cross-sectional analyses to identify the economic circumstances under which AI-generated information provides the greatest value to investors. We explore these circumstances here.

To do so, we quantify the value of AI-generated summaries for a given earnings announcement based on participants' average performance when treated with aligned summaries relative to raw transcripts. We then examine how this measure varies with firm and earnings announcement characteristics.

For brevity, we defer the results to Table OA-1 of the Online Appendix. The table shows that aligned AI benefits participants more when the task is more complex. Across the four dependent variables, we observe positive and statistically significant coefficients on Earnings Volatility, Loss, and Return Volatility, as well as negative coefficients on Size and Book-to-Market. These findings consistently indicate that AI provides greater value to investors when they analyze firms with greater uncertainty (e.g., Kim et al., 2024b). Furthermore, the negative effect of the book-to-market ratio suggests that AI value increases when firms' assets are more challenging to quantify and evaluate. This is consistent with the marginal benefit of AI adoption growing as investors face higher processing costs, i.e., operate in more opaque and complex information environments.

4 Does AI Widen Information Inequality?

In the previous section, we show that both more and less sophisticated participants benefit from AI-generated summaries when these outputs are aligned with their level of financial sophistication. Building on these findings, we now turn to a critical question: does the availability of AI-generated information narrow or widen the pre-existing gap between these two types of investors?

To motivate our analysis, we first introduce a simple theoretical model in Section 4.1, which highlights the ambiguity in how AI adoption affects knowledge-driven disparities. We then empirically test this relation in Section 4.2, where we analyze how the disparity changes when participants are provided with AI assistance. In Section 4.3, we evaluate how well our experimental data aligns with the predictions of the model. Finally, in Section 4.4, we explore the economic determinants of the knowledge-driven disparities.

4.1 A Motivating Model

We introduce a stylized model to analyze how the knowledge-driven disparity among investors may evolve after adopting AI technology. Within the model, and consistent with our experiment, AI technology simplifies complex corporate disclosures, making them more accessible to investors at the expense of being less detailed.

We assume a spectrum of investors indexed by $i \in [0, 1]$ whose sophistication (ability) is $a_i \in [0, \bar{a}]$. Without loss of generality, we assume a_i is increasing in i. Investors process a document D disclosed by a company. D's complexity is denoted by C and its informativeness is $I(C)$, which is an increasing function, $\frac{\partial I(C)}{\partial C} > 0$. Intuitively, more informative disclosures tend to be more complex. A user's utility from processing D increases in its informativeness and decreases in the processing cost. The cost varies across investors: it increases in complexity but decreases with an investor's sophistication.

To demonstrate that AI has an ambiguous effect on the knowledge gap, it suffices to impose simple functional relations among the variables. Accordingly, we assume that $I(C) = \mu C$ where μ is a positive constant.[22] We further assume that investor i's processing cost takes the form $Q_i(C) = \frac{C^2}{2a_i}$, which follows Sims (2003) and Maćkowiak and Wiederholt (2009) among others. Under these assumptions, investor i's utility from an investment decision that requires processing a disclosure D with complexity C is given by (users will not process disclosures that yield negative utility):

$$u(I(C), C, a_i) = \max\left[\mu C - \frac{C^2}{2a_i}, 0\right] \tag{9}$$

Consider an AI technology that can adjust the complexity of a given disclosure to optimize investors' utility by providing summaries. Suppose the original disclosure, denoted as D_R, has a complexity of C_R. Intuitively, the optimal level of complexity C^* generated by AI varies depending on the user's sophistication. The optimal complexity is determined by first-order condition:

$$\frac{\partial u_i}{\partial C_i} = \mu - \frac{C_i}{a_i}. \tag{10}$$

An interior solution implies $C_i^* = \mu a_i$, which denotes the user-specific optimal level of complexity. The investor's utility evaluated at the optimum becomes $u(I(C_i^*), C_i^*, a_i) = \frac{\mu^2 a_i}{2}$.

[22] The results are similar when we assume a concave relation, i.e., $\frac{\partial^2 I(C)}{\partial C^2} < 0$.

Note that one typically expects the summary's complexity to be lower compared to the original document, i.e., $C_i^* \leq C_R$.[23] Therefore, investors whose comprehension level is higher than $\frac{C_R}{\mu}$ do not require summaries but rather use the raw document. One could imagine, however, that AI-based output exceeds the complexity of the original disclosure (e.g., AI can place the information into a broader context or offer a sophisticated analysis of the disclosed information). We return to this important case later.

The above has an important implication: adopting AI technology increases investors' participation in financial markets. In particular, prior to the adoption of the technology, less able investors will refrain from financial markets since

$$u_i(I(C_R), C_R, a_i) = \mu C_R - \frac{C_R^2}{2a_i} < 0, \tag{11}$$

or equivalently if their ability $a_i < \frac{C_R}{2\mu}$. However, with the new technology, the utility of each investor becomes $\frac{\mu^2 a_i}{2} > 0$, and hence all investors participate in financial decision making.

Define the function $\Delta(a_i) \equiv u_i(I(C_i^*), C_i^*, a_i) - u_i(I(C_R), C_R, a_i)$ to measure the incremental utility obtained from the technology's summarization. It is straightforward to show that:

$$\Delta(a_i) = \begin{cases} 0 & \text{if } \frac{C_R}{\mu} < a_i \leq \bar{a} \\ \frac{\mu^2 a_i}{2} - \left(C_R\mu - \frac{C_R^2}{2a_i}\right) & \text{if } \frac{C_R}{2\mu} < a_i \leq \frac{C_R}{\mu} \\ \frac{\mu^2 a_i}{2} & \text{if } 0 \leq a_i \leq \frac{C_R}{2\mu} \end{cases} \tag{12}$$

which implies that all investors weakly benefit from new technology, i.e., $\Delta(a_i) \geq 0$. This indicates that the overall welfare gain from improved decision-making is:

$$\int_0^1 \Delta(a_i) a_i' di > 0. \tag{13}$$

where a_i' denotes partial derivative.

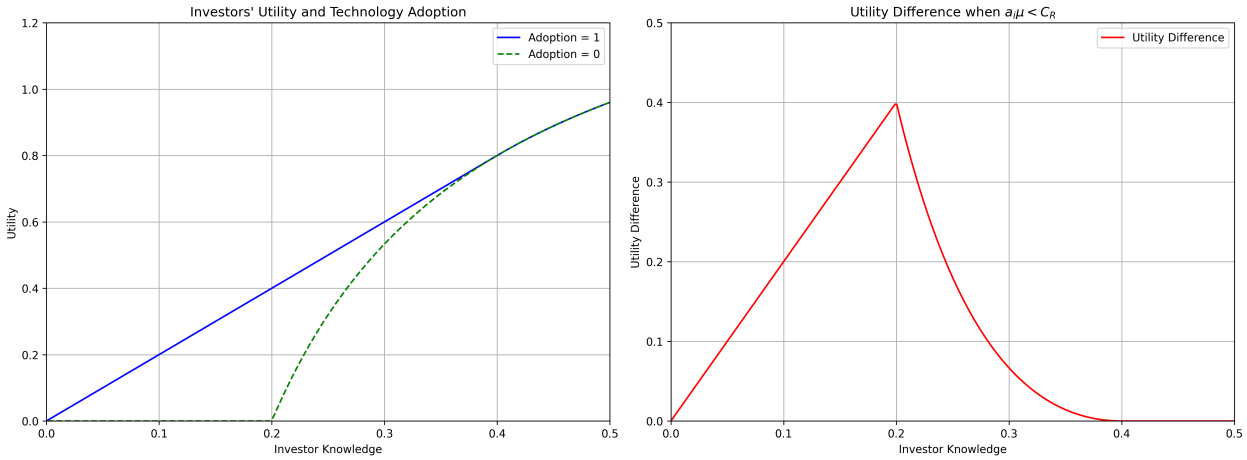
Interestingly, whether $\Delta(a_i)$ increases or decreases with a_i is ambiguous:

$$\frac{\partial \Delta(a_i)}{\partial a_i} = \begin{cases} 0 & \text{if } \frac{C_R}{\mu} < a_i \leq \bar{a} \\ \frac{\mu^2}{2} - \frac{C_R^2}{2a_i^2} & \text{if } \frac{C_R}{2\mu} < a_i \leq \frac{C_R}{\mu} \\ \frac{\mu^2}{2} & \text{if } 0 \leq a_i \leq \frac{C_R}{2\mu} \end{cases} \tag{14}$$

[23] We later relax this assumption and discuss its effect.

Figure 4 illustrates the investors' utility before (dotted line) and after (solid line) adopting the technology, assuming $\mu = 2$ and $C_R = 0.8$. The technology increasingly benefits users with low ability by making participation incentives compatible ($a_i < \frac{C_R}{2\mu} = 0.2$). However, conditional on participation, the benefit starts to decrease in users' sophistication ($0.2 < a_i < 0.4$). The technology does not make a difference for highly sophisticated users ($a_i > 0.4$).

Figure 4. **Investors' Utility After the Treatment**



Notes: This figure visualizes the investors' utility when $\mu = 2$ and $C_R = 0.8$. The left panel shows the utility function values. The solid line represents the utility after adopting the technology, and the dotted line represents the utility before adopting the technology. The right panel shows the net utility gain from the technology when $0 < a_i < 0.4$.

The above suggests that whether a group of investors benefits relatively more from the technology adoption remains an empirical question. Ceteris paribus, if one increases complexity C_R, the range of a_i with a strictly positive $\Delta(a_i)$, $\left[0, \frac{C_R}{\mu}\right]$, expands. Holding the entire distribution of ability constant, it becomes more likely that high-ability investors benefit more from the technology than low-ability investors.

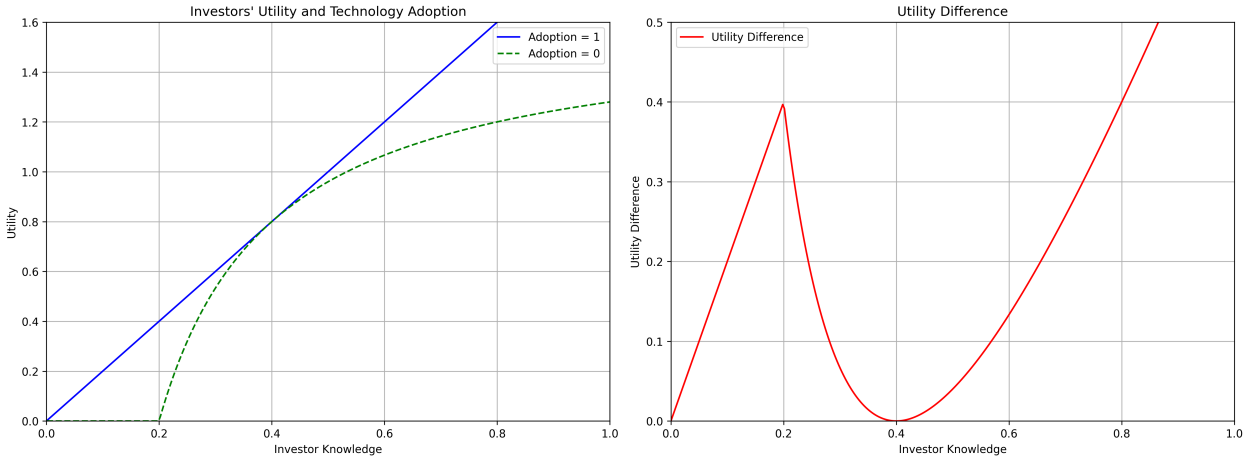
Recall that the above discussion considers the case when $C_i^* < C_R$ because our first experiment confines its summaries to the content of conference call transcript only. In practice, and in line with our second experiment, the model's output complexity can exceed that of the transcript. In particular, when answering user's questions, AI can search its knowledge base, synthesize its knowledge with the content of the earnings call, and generate economic insights not directly following from the transcript. This may result in $C_i^* > C_R$ and $I(C_i^*) > I(C_R)$. In such case, the utility increment realized by the users is $\frac{\partial \Delta(a_i)}{\partial a_i} = \frac{\mu^2}{2} - \frac{C_R^2}{2a_i^2} > 0$ for $a_i > \frac{C_R}{\mu}$. In other words, the knowledge gap increases in a_i for investors with high sophistication, further widening the knowledge-driven disparity between the sophisticated

and unsophisticated investors.

Figure 5 illustrates this extension using the same numerical setup as Figure 4 and assuming $\bar{a}$ is nonbinding. For $a_i < 0.4$, $\Delta(a_i)$ is identical to Figure 4. However, for $a_i > 0.4$, the informational value investors derive from new technology increases with their level of sophistication.

In sum, our simple setup illustrates that the effect of AI adoption need not equalize the informational playing field among investors and, in fact, can be highly non-monotonic, making it an empirical question.

Figure 5. **When AI Can Access External Knowledge**



Notes: This figure visualizes the investors' utility when $\mu = 2$ and $C_R = 0.8$. We also lift the constraint on C_i^* that $C_i^* < C_R$. The left panel shows the utility function values. The solid line represents the utility after adopting the technology, and the dotted line represents the utility before adopting the technology. The right panel shows the net utility gain from the technology.

4.2 Empirical Results

We empirically test whether the aligned with user sophistication AI summaries reduce or widen the knowledge gap. Disparities among investors will widen if more sophisticated participants are to gain more from AI adoption compared to less sophisticated ones. To evaluate this within our regression framework, we test whether $\beta_2 + \beta_4 - \beta_3$ from equation 1 is statistically different from zero (see Figure 1).[24]

[24]The baseline difference between the two groups without AI treatment is β_1. The post-treatment difference is derived by comparing $\beta_1 + \beta_2 + \beta_4$ (for sophisticated participants treated with advanced summaries) and β_3 (for less sophisticated participants treated with simplified summaries). The change in disparity is given by $\beta_2 + \beta_4 - \beta_3$.

We find that performance disparities increase after AI treatment, as evidenced by our results in Table 5. For the earnings prediction task, the performance gap between more vs. less sophisticated participants is 0.051 (based column (3) of Table 2, Panel A), representing 10.69% of the average prediction accuracy. Similarly, for the sentiment assessment task, the performance gap is 0.033 (from column (3) of Table 2, Panel B), or 6.90% of the average classification accuracy. Both values are positive, with statistically significant t-statistics of 2.99 and 3.05, respectively.

For the short-term investment task, the performance disparity between participants of differing financial sophistication is even more pronounced following the AI treatment, with $\beta_2 + \beta_4 - \beta_3 = 0.004$ (t-statistic = 1.83), which represents 13.79% of the average performance gap. A similar pattern emerges from the Sharpe ratios (column (5)). For long-term return prediction tasks (columns (6) and (7)), we observe consistent results, which we translate into one-year portfolio returns. For more sophisticated investors, aligned AI increases one-year portfolio returns by 9.6% but for less for less sophisticated ones, it increases portfolio returns by 1.7%, thereby widening the gap by 7.9%.

In sum, our empirical results suggest that AI-generated summaries, while enhancing overall information processing and decision-making for each group, exacerbate disparities between sophisticated and less sophisticated participants. Although AI is expected to democratize financial markets, our results highlight that it disproportionately benefits those with greater financial sophistication, thereby widening the knowledge-based performance gap between investor groups.

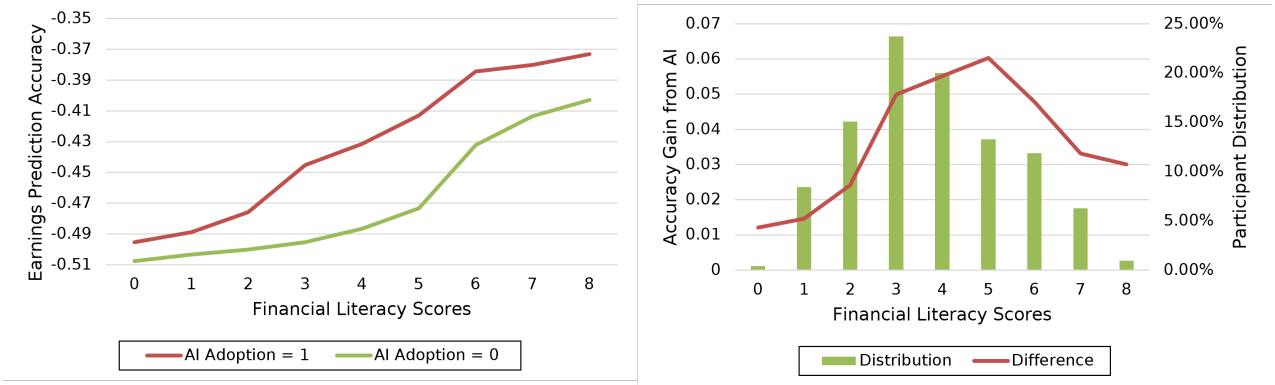
4.3 Is The Effect Monotonic?

In this section, we test our model's prediction that the benefit from AI adoption is non-monotonic in investor sophistication. Recall that for less sophisticated investors, an incremental benefit of AI summaries is expected to increase with the level of sophistication, whereas the opposite is true for investors with relatively high sophistication. The positive relationship for less sophisticated investors reflects their inability to extract value from raw earnings call transcripts without AI. In the absence of AI, these investors effectively bypass earnings call information in their decision-making. However, since AI summaries simplify complex disclosures, these investors begin incorporating such information into their decisions. Within this less sophisticated group, individuals are better able to take advantage of AI-generated summaries as their sophistication grows, resulting in greater incremental gains.

In contrast, the negative relationship among the more sophisticated investors is driven

by their baseline ability to effectively process earnings transcripts. For these individuals, the additional value provided by AI declines as their existing capacity to interpret information is already relatively high. Consequently, the incremental benefit from AI-generated summaries decreases with sophistication.

Figure 6. **Information Processing and AI Summaries**



Notes: This figure displays the model fit with our experimental data. The left panel shows the average accuracy of the participants who received each score in predicting earnings. AI Adoption = 0 denotes the accuracy when the participants are treated with raw transcripts. AI Adoption = 1 denotes the accuracy when the participants are treated with summaries. For participants whose score is lower than 4, we use simplified summaries to compute the average accuracy. We use advanced summaries to compute the average accuracy for those with scores above 3. The right panel shows the net accuracy gain from the treatment (Difference between AI Adoption=1 and AI Adoption=0), along with the score distribution of participants.

Our experimental data supports the inverse U-shaped relationship discussed above. Figure 6 displays the average earnings prediction accuracy for different levels of sophistication. The left panel shows the accuracy for participants treated with raw transcripts vs. the aligned AI-generated summaries.[25] Among less sophisticated participants (scores below 4), average accuracy remains slightly below -0.5, which is equivalent to random guessing. However, the accuracy rapidly picks up as sophistication scores approach 4. In contrast, when treated with aligned summaries, participants' average accuracy increases monotonically.

The right panel of Figure 6 illustrates the incremental accuracy gains from AI treatment and the distribution of participants by sophistication score. In line with the theoretical prior discussed above (also see Figure 4), we observe an inverse U-shape relation between informational gains and investor ability (scores). Consistent with our model, among less sophisticated investors (i.e., Score $\leq$ 3), those with relatively higher scores benefit more

[25]For participants with a score of less than 4, we take the average when treated with simplified summaries. We take the average prediction accuracy when treated with advanced summaries for participants whose scores are higher than or equal to 4.

from AI summaries. Conversely, among more sophisticated participants, higher scores are associated with smaller gains. Most participants in our sample fall within the lower end of the inverse U-shape, indicating that many fail to derive benefits from raw transcripts alone. In the aggregate, this distribution highlights the "unleveling" effect of AI, whereby its advantages disproportionately favor specific investor groups.

4.4 What Drives Disparities?

We next explore the mechanisms driving the disparities between more and less sophisticated participants. We hypothesize that AI benefits sophisticated participants more when simplified summaries sacrifice precision.

Methodology To study the sources of disparities, we first measure the *Wedge*—a call-level difference in AI-related gains between more and less sophisticated participants. To formalize this measure, we introduce mutually exclusive indicators $I_j^{k,d}$. Here $k \in (s, us)$ denotes either sophisticated (s) or less sophisticated participants (us), while $d \in (as, ss, r)$ refers to receiving an advanced summary (as), a simplified summary (ss), or a raw transcript (r). For example, when a participant j is classified as sophisticated and she is treated with an advanced summary, $I_j^{s,as} = 1$, and zero otherwise. We also use AQ_{ij} to denote one of the answer quality metrics for participant j for earnings call i. Accordingly, the wedge for each earnings call i is defined as:

$$\text{Wedge}_i = \underbrace{\left(\frac{\sum_j \text{AQ}_{ij} I_j^{s,as}}{\sum_j I_j^{s,as}} - \frac{\sum_j \text{AQ}_{ij} I_j^{s,r}}{\sum_j I_j^{s,r}}\right)}_{\text{Informational Gain of More Sophisticated People}} - \underbrace{\left(\frac{\sum_j \text{AQ}_{ij} I_j^{us,ss}}{\sum_j I_j^{us,ss}} - \frac{\sum_j \text{AQ}_{ij} I_j^{us,r}}{\sum_j I_j^{us,r}}\right)}_{\text{Informational Gain of Less Sophisticated People}} \quad (15)$$

AQ_{ij} is participant j's answer quality about earnings call i.

As a second step, we measure earnings announcement characteristics that are likely to be associated with a loss of precision in simplified summaries. These characteristics are likely to be important in decision-making but are less likely to be retained in simplified summaries. The choice of these characteristics is supported by our content analysis of the differences between simple and advanced summaries (see Online Appendix) and includes: (1) the absolute value of the difference between EBITDA and net income (scaled by the total assets); (2) the absolute value of accruals (scaled by the total assets) to proxy for accounting complexities in predicting cash flows (Ball and Nikolaev, 2022); (3) the absolute amount of extraordinary items (scaled by total assets) to proxy for non-recurring components of earnings; (4) the number of product

segments to proxy for the operational complexity; (5) R&D scaled by total assets to proxy for the innovation intensity; and (6) proxies for financing activities commonly omitted from simplified summaries, namely, the dividend payout ratio and the number of shares repurchased.

We then estimate the following regression to study the determinants of Wedge.

$$\text{Wedge}_i = \gamma Z_i + \delta_t + \varepsilon_i \tag{16}$$

Z_i is a vector of earnings call characteristics described above. We also use year-quarter fixed effects control for macroeconomic trends. Continuous variables are winsorized at the 1% and 99% levels to mitigate the influence of outliers.

Results Table 6 shows that *Wedge* increases when the omitted information is likely to be essential for decision-making. Specifically, it grows with a higher difference between EBITDA and net income, higher accruals magnitude, more complex revenue compositions (e.g., more product segments), a higher R&D intensity, and greater share repurchases and dividend payout. In terms of economic magnitudes, using the results in column (1), a one standard deviation increase in the difference between EBITDA and net income is associated with a 0.26 standard deviation increase in Wedge related to earnings prediction ($= \frac{0.023 \times 3.341}{0.296}$). Similarly, a one standard deviation increase in R&D investment is associated with a 0.27 standard deviation increase in Wedge ($\frac{0.022 \times 3.885}{0.296}$). We find consistent results across all four columns. These results suggest that disparities become more pronounced when the information omitted from simplified summaries plays a critical role in financial decision-making.

5 Interactive AI Assistance

In this section we discuss the results from our second experiment, which involves interactions with an AI assistant (chatbot). This design provides participants with more flexibility in how they use generative AI to process earnings news.

Experiment Design Instead of pre-written summaries, we provide the treatment group with an AI assistant pre-loaded (as a part of the prompt) with a specific earnings call transcript. The treatment group interacts with the chatbot, which is pre-configured to provide information about the company's earnings release, while the control group works directly with the raw transcript. This experiment allows us to not only measure participants' performance on investment tasks but also to track how they engage with AI. For example, we can observe

the questions they ask the AI chatbot and study how the nature of their questions varies with investor sophistication.[26]

We instruct participants to answer our survey questions by using the chatbot to understand the disclosed information. Similar to our main experiment, the control group is presented with the raw transcript only. The questions contained in the survey are identical to those we use in our main experiment. Additionally, we collect the questions that the participants ask and the answers generated by the AI chatbot. We use 400 participants with 200 participants assigned to the control group and the rest is the treatment group. We then estimate the following regression:

$$\begin{aligned}\text{Answer Quality}_{kj} = \beta_1 \text{Sophistication}_j + \beta_2 \text{AI Chatbot}_j \\ + \beta_3 \text{Sophistication}_j \times \text{AI Chatbot}_j + \zeta_k + \varepsilon_{kj}\end{aligned} \tag{17}$$

Answer Quality is defined as previously and is measured at the earnings call level for earnings prediction and sentiment assessment and at the stock pair level for portfolio performance. AI Chatbot$_j$ is an indicator that equals one when participant j is treated with an AI chatbot and zero otherwise. We include earnings call or pair-level fixed effects (ζ_k) and cluster standard errors at the earnings call (pair) level.

A critical concern related to human-AI interactions is that participants might directly ask the chatbot about the expected direction of future earnings or instruct it to suggest a portfolio allocation. This creates two challenges. First, our objective is to understand whether an AI model can help investors *process* information and not to examine whether the model can successfully predict future corporate outcomes or directly make decisions. Second, as our sample contains earning releases from 2010 to 2022 (which falls into the training period of GPT-4o) and, since the model has access to rich textual information, may identify the firm and hence may suffer from the look-ahead bias (Sarkar and Vafa, 2024). We tackle these challenges by including several features to ensure the model focuses on information processing rather than providing direct prediction tasks. First, using a system prompt, we explicitly

[26] We also provide participants with some basic information about the nature of an earnings conference call, but we do not provide them with the actual raw transcript. Specifically, we provide the following prompt in the survey. "Assume you are an investor in the U.S. stock market. You will be provided with the interactive AI Chatbot "FinChat," which contains information on the transcript of company earnings announcements conference calls. During these calls, management provides an overview of the financial results (e.g., earnings, sales), discusses key business developments, and often offers forward-looking guidance. They also answer questions from analysts and investors. These calls can be highly informative and help investors make more informed decisions. We will ask you questions about the earnings calls of two companies (Company A and Company B). Aside from the details provided in these calls, these two companies are generally comparable and operate in the same industry."

instruct the chatbot not to answer questions about information not included in the transcript. Second, when a user directly asks a survey question, the model is also instructed not to provide answers (we collect instances where the participants attempt to ask the AI assistant the survey questions). Finally, our survey instructs the participants not to ask the survey questions.[27]

What Questions Do Investors Ask? The content and quality of participants' questions reveal significant differences in how investors engage with the AI chatbot.[28] Table 7 shows that more sophisticated participants ask higher-quality questions, resulting in better engagement with the chatbot. Despite asking a similar quantity of questions, sophisticated participants are more likely to ask for summaries or specific financial insights, such as merger activity or capital expenditures. They also demonstrate a clear pattern of starting with general summaries and following up with targeted detailed questions. In contrast, less sophisticated participants often stop after a single interaction and are more likely to directly ask survey questions despite instructions to avoid this. Overall, our results indicate that more sophisticated participants take greater advantage of AI, asking higher-quality questions and generating higher-quality output. This reinforces the emergence of knowledge-driven disparities in financial decision-making.

The Effect of the AI Assistant Table 8 shows that sophisticated participants consistently outperform less sophisticated participants on investment tasks when using the AI chatbot. Intuitively, the coefficient on Sophistication is positive, indicating that more sophisticated people perform better than less sophisticated ones when working with raw transcripts. More importantly, less sophisticated participants exhibit a deterioration in their performance when they rely on the AI assistant as follows from the significantly negative coefficient on Chatbot (β_2). This is consistent with the inability to ask relevant or effective questions. [30]

[27] We also note that the participants' compensation might be negatively impacted when they fail to comply with these instructions. We manually inspect whether the participants cohere to the instructions and exclude questions that are identical to the survey questions.

[28] We received a total of 1,617 messages from the 200 participants of the treatment group, yielding eight questions per participant on average. We classify the questions into five categories: (i) questions that ask to summarize certain content; (ii) questions that ask specific financial issues (e.g., risks, financial metrics); (iii) questions that are identical to the survey questions[29]; (iv) questions that cannot be answered using the information presented in the transcript; (v) other invalid questions (questions with less than three words with answers less than ten words). 269 questions were identical to the survey questions, 194 questions were not answered because the model could not find the answers within the transcript, 194 questions asked the model to summarize certain content, 1,067 questions asked about specific financial information and answered successfully by the model, and 87 questions were invalid for other reasons.

[30] This result is likely exacerbated by the absence of raw transcripts, leaving less sophisticated participants unable to engage meaningfully with the chatbot.

Most importantly, Table 8 documents that more sophisticated participants derive substantial benefits from the chatbot. The coefficient on Sophistication × Chatbot, (i.e., β_3) is positive and statistically significant across all four investment-related tasks in our study. In terms of economic significance, the chatbot improves the information processing and investment efficiency of sophisticated participants by approximately 25% relative to the average. This result underscores the value of AI interactions for those with a foundation of financial knowledge.

Taken together, Table 8 highlights that the benefits from AI assistance depend heavily on users' financial sophistication. When used effectively, the AI assistant provides valuable insights that enhance investor decision-making. However, without the ability to ask the right questions, less sophisticated users are unable to make informed inferences leading to poorer economic outcomes.

6 Robustness Checks

To enhance the robustness of the findings of our main experiment, we perform a battery of robustness checks. These tests address possible uses of generative AI to answer survey questions, participant's ability to identify the company from anonymized transcripts, and the use of other software to complete the survey. We discuss these analyses in more detail in the Online Appendix.

7 Conclusion

Using two large-scale experiments with actual financial data from 200 publicly traded firms, we study the effect of generative AI on financial information processing and decision-making. Our analysis yields two central results. First, there is a significant improvement in both information processing ability and the quality of investment decisions following AI adoption. However, this effect holds (for both sophisticated and unsophisticated groups of users) as long as AI output is aligned with user sophistication. Second, AI widens the knowledge-driven disparities between sophisticated and unsophisticated participants.

These results contribute to our understanding of AI's role in financial markets. First, our evidence shows that generative AI significantly enhances both financial comprehension and investment performance, making earnings information more accessible to a broader investor base. Second, our finding that AI's effectiveness critically depends on the alignment between its outputs and user expertise highlights the fundamentally value of personalization in shaping

AI's impact on investment decisions. Third, we challenge prevailing narratives about AI's democratizing effects by showing that in financial markets, where successful outcomes require both understanding and strategic action, AI can widen performance gaps by disproportionately benefiting sophisticated investors. Our results reveal that AI's potential to democratize access to financial markets is more nuanced than previously thought, particularly in contexts where technical domain specific knowledge is critical.

References

Ball, R. and V. V. Nikolaev (2022). On earnings and cash flows as predictors of future cash flows. *Journal of Accounting and Economics 73*(1), 101430.

Barber, B. M. and T. Odean (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies 21*(2), 785–818.

Bick, A., A. Blandin, and D. J. Deming (2024). The rapid adoption of generative ai. *National Bureau of Economic Research*.

Bonsall, S. B., A. J. Leone, B. P. Miller, and K. Rennekamp (2017, mar). A plain english measure of financial reporting readability. *Journal of Accounting and Economics 63*(2), 329–357.

Brynjolfsson, E., D. Li, and L. R. Raymond (2023). Generative ai at work. *National Bureau of Economic Research*.

Chen, Y., B. T. Kelly, and D. Xiu (2022). Expected returns and large language models. *Available at SSRN*.

Croom, J. (2024). Interactivity and illusions of ability: The effect of generative ai on investor judgments. *Available at SSRN*.

Cui, Z. K., M. Demirer, S. Jaffe, L. Musolff, S. Peng, and T. Salz (2024). The effects of generative ai on high skilled work: Evidence from three field experiments with software developers. *Available at SSRN 4945566*.

Dell'Acqua, F., E. McFowland III, E. R. Mollick, H. Lifshitz-Assaf, K. Kellogg, S. Rajendran, L. Krayer, F. Candelon, and K. R. Lakhani (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of ai on knowledge worker productivity and quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper* (24-013).

DellaVigna, S. and J. M. Pollet (2009). Investor inattention and friday earnings announcements. *The Journal of Finance 64*(2), 709–749.

Desai, M. (2023). What the finance industry tells us about the future of ai. *Harvard Business Review*.

Ebert, L. (2023). The rise of ai assistants: Hedge fund managers and chatgpt. *BNP Paribas Capital Introduction Team*. Accessed online: https://www.bnpparibas.com.

FINRA (2009). Financial capability in the united states: National survey—questionnaire and respondent-level data.

Frankel, R., J. Jennings, and J. Lee (2022). Disclosure sentiment: Machine learning vs. dictionary methods. *Management Science 68*(7), 5514–5532.

Heeb, F., J. F. Kölbel, F. Paetzold, and S. Zeisberger (2022, sep). Do investors care about impact? *The Review of Financial Studies 36*(5), 1737–1787.

Kaniel, R., S. Liu, G. Saar, and S. Titman (2012). Individual investor trading and return patterns around earnings announcements. *The Journal of Finance 67*(2), 639–680.

Kaniel, R., G. Saar, and S. Titman (2008). Individual investor trading and stock returns. *The Journal of Finance 63*(1), 273–310.

Kelly, B. and D. Xiu (2023). Financial machine learning. *Foundations and Trends® in Finance 13*(3-4), 205–363.

Kim, A., M. Muhn, and V. Nikolaev (2024a). Financial statement analysis with large language models. *arXiv preprint arXiv:2407.17866*.

Kim, A., M. Muhn, and V. V. Nikolaev (2024b). Bloated disclosures: can ChatGPT help investors

process information? *Chicago Booth Research Paper* (23-07), 2023–59.

Krische, S. D. (2019). Investment experience, financial literacy, and investment-related judgments. *Contemporary Accounting Research 36*(3), 1634–1668.

Lian, C., Y. Ma, and C. Wang (2018, sep). Low interest rates and risk-taking: Evidence from individual investment decisions. *The Review of Financial Studies 32*(6), 2107–2148.

Lopez-Lira, A. and Y. Tang (2023). Can ChatGPT forecast stock price movements? return predictability and large language models. *arXiv preprint*.

Maćkowiak, B. and M. Wiederholt (2009). Optimal sticky prices under rational inattention. *American Economic Review 99*(3), 769–803.

Noy, S. and W. Zhang (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science 381*(6654), 187–192.

Peng, S., E. Kalliamvakou, P. Cihon, and M. Demirer (2023). The impact of ai on developer productivity: Evidence from github copilot. *arXiv preprint arXiv:2302.06590*.

Sarkar, S. K. and K. Vafa (2024). Lookahead bias in pretrained language models. *Available at SSRN*.

Savor, P. and M. Wilson (2016). Earnings announcements and systematic risk. *The Journal of Finance 71*(1), 83–138.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics 50*(3), 665–690.

Toner-Rodgers, A. (2024). Artificial intelligence, scientific discovery, and product innovation.

Table 1. Demographics and Descriptives

Panel A. Demographics

	Our Survey	Census
	(1)	(2)
Observations	1,800	237,784,222
Gender		
Male	42.2%	49.6%
Female	57.8%	50.4%
Race		
White	67.3%	60.5%
Black	15.8%	12.1%
Asian	6.3%	6.0%
Mixed	6.7%	12.8%
Other	3.9%	8.6%
Age		
First Quartile	32	32.5
Median	39	39.2
Mean	42.4	46.3
Fourth Quartile	50	52.5

Panel B. Descriptive Statistics and Representativeness

	Used Stocks	Population	Difference	t-stat
	(1)	(2)	(3)	(4)
Market Capitalization	10,794	15,555	-4,761	(-0.76)
Book-to-Market	0.267	0.359	-0.092***	(-3.21)
ROE	-0.089	0.012	-0.101	(-1.50)
SUE_{rw}	0.000	-0.004	0.004	(0.49)
$SUE_{analyst}$	0.300	0.102	0.198	(0.60)
N	200	8,672		

Notes: In Panel A, we provide the demographic characteristics of our survey respondents. The first column displays the demographic information of our participants. We exclude participants who do not provide consent to sharing their demographic information when computing the ratios. The second column displays the demographic information obtained from the 2023 US Census data. We include a population older than 18 years and younger than 75 years, as the oldest participant in our experiment is 75 years old. Census data provides age data in five-year intervals and quartile values are imputed using the midpoint of each group. In Panel B, we provide descriptive statistics of the firms used in our sample and compare them to the universe of firms. Used Stocks indicates a sample of earnings announcements (stock-year) used in the experiment. Population indicates a full sample of earnings announcements (stock-year) after excluding small-cap stocks, finance industry, and stocks missing transcripts. Difference indicates the difference in means between two groups. t-stat indicates the t-statistics from the test of difference in means between two groups. Population is the average market capitalization in a million dollars for each group. Book-to-Market is the average book-to-market for each group. ROE is the average return on equity for each group. SUE_{rw} is the average earnings surprise for each group based on a random walk. $SUE_{analyst}$ is the average earnings surprise for each group based on analysts' consensus. t-values are reported in the parentheses. *, **, *** represent statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

Table 2. Information Processing

Panel A. Earnings Predictions						
Dependent Variable =	Accuracy			Text Quality		
	(1)	(2)	(3)	(4)	(5)	(6)
Sophistication	0.027***		0.014	0.055***		0.044**
	(3.85)		(1.19)	(5.35)		(2.36)
Advanced Summary		0.040***	-0.006		0.043***	-0.012**
		(4.45)	(-0.50)		(4.03)	(-2.30)
Simplified Summary		-0.007	0.035***		0.015	0.015*
		(-0.86)	(2.94)		(0.67)	(1.90)
Sophistication × Advanced Summary			0.092***			0.083***
			(5.29)			(5.01)
Sophistication × Simplified Summary			-0.054***			-0.056**
			(-3.69)			(-2.36)
Avg(Dependent Variable)	-0.477	-0.477	-0.477	-0.430	-0.430	-0.430
Call Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.592	0.595	0.607	0.645	0.646	0.663
N	3,600	3,600	3,600	3,600	3,600	3,600

Panel B. Sentiment Assessment			
Dependent Variable =		Accuracy	
	(1)	(2)	(3)
Sophistication	0.020***		0.012
	(2.82)		(0.91)
Advanced Summary		0.032***	-0.010*
		(3.17)	(-1.89)
Simplified Summary		0.010*	0.035**
		(1.75)	(2.54)
Sophistication × Advanced Summary			0.078***
			(4.18)
Sophistication × Simplified Summary			-0.054***
			(-2.99)
Avg(Dependent Variable)	-0.478	-0.478	-0.478
Call Fixed Effects	Yes	Yes	Yes
Adjusted R^2	0.631	0.630	0.637
N	3,600	3,600	3,600

Notes: Sophistication indicates a participant scoring four or above in the financial knowledge questions. Advanced Summary and Simplified Summary are indicators for summaries targeted at financially knowledgeable and less knowledgeable people, respectively. Panel A shows earnings prediction accuracy and Panel B shows the sentiment assessment accuracy of the participants. Standard errors are clustered by earnings calls. t-values are reported in the parentheses. *, **, *** represent statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

Table 3. Confidence and Timeliness

Panel A. Confidence						
Dependent Variable =	Earnings Prediction			Sentiment Assessment		
	(1)	(2)	(3)	(4)	(5)	(6)
Sophistication	0.103***		0.130***	0.078***		0.103***
	(12.26)		(8.17)	(9.32)		(6.21)
Advanced Summary		0.110***	0.088***		0.121***	0.099***
		(9.90)	(4.87)		(11.29)	(5.53)
Simplified Summary		0.013	0.067***		0.038***	0.091***
		(1.28)	(4.50)		(3.80)	(5.95)
Sophistication × Advanced Summary			0.040*			0.041*
			(1.68)			(1.82)
Sophistication × Simplified Summary			-0.114***			-0.110***
			(-5.96)			(-5.45)
Avg(Dependent Variable)	0.653	0.653	0.653	0.724	0.724	0.724
Call Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.064	0.061	0.122	0.048	0.066	0.114
N	3,600	3,600	3,600	3,600	3,600	3,600
Panel B. Timeliness						
Dependent Variable =	Earnings Prediction			Sentiment Assessment		
	(1)	(2)	(3)	(4)	(5)	(6)
Sophistication	41.29***		98.21***	-0.52		-6.92
	(3.38)		(2.81)	(-0.15)		(-1.14)
Advanced Summary		-75.62***	-46.13***		-14.27***	-22.47***
		(-3.42)	(-2.80)		(-3.57)	(-4.09)
Simplified Summary		-97.36***	-52.91***		-19.00***	-20.61***
		(-4.75)	(-3.51)		(-4.08)	(-2.96)
Sophistication × Advanced Summary			-62.11**			3.62
			(-2.20)			(0.43)
Sophistication × Simplified Summary			-93.22**			-16.62**
			(-2.24)			(-1.98)
Avg(Dependent Variable)	346.95	346.95	346.95	59.30	59.30	59.30
Call Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.014	0.021	0.025	0.005	0.012	0.006
N	3,600	3,600	3,600	3,600	3,600	3,600

Notes: Sophistication indicates a participant scoring four or above in the financial knowledge questions. Advanced Summary and Simplified Summary are indicators for summaries targeted at financially knowledgeable and less knowledgeable people, respectively. Standard errors are clustered by earnings calls. t-values are reported in the parentheses. Panel A measures the self-reported confidence of the participants. Panel B measures the time taken (in seconds) to complete each section of the survey. *, **, *** represent statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

Table 4. Investment Efficiency

Panel A. Short-Term Investments						
Dependent Variable =	Portfolio Returns			Sharpe Ratio		
	(1)	(2)	(3)	(4)	(5)	(6)
Sophistication	0.001		0.002*	0.272*		0.694*
	(1.20)		(1.70)	(1.95)		(1.77)
Advanced Summary		0.001	-0.004***		0.238	-0.861**
		(1.01)	(-3.38)		(1.40)	(-2.39)
Simplified Summary		0.002	0.004***		-0.501	0.550*
		(1.35)	(3.14)		(-1.58)	(1.75)
Sophistication × Advanced Summary			0.012***			2.226***
			(4.93)			(3.69)
Sophistication × Simplified Summary			-0.006***			-1.132***
			(-3.63)			(-2.77)
Avg(Dependent Variable)	-0.029	-0.029	-0.029	-4.412	-4.412	-4.412
Group Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.431	0.431	0.603	0.589	0.590	0.601
N	1,800	1,800	1,800	1,800	1,800	1,800
Panel B. Long-Term Investments						
Dependent Variable =	Portfolio Returns			Sharpe Ratio		
	(1)	(2)	(3)	(4)	(5)	(6)
Sophistication	0.019***		0.003	2.254**		1.287
	(2.93)		(0.53)	(2.37)		(1.55)
Advanced Summary		0.037***	-0.023		1.096**	-0.644
		(3.56)	(-1.50)		(2.21)	(-0.74)
Simplified Summary		-0.015	0.017**		-0.345	1.013**
		(-1.23)	(2.07)		(-0.80)	(2.31)
Sophistication × Advanced Summary			0.119***			2.350**
			(6.57)			(2.40)
Sophistication × Simplified Summary			-0.061***			-2.203***
			(-3.39)			(-2.80)
Avg(Dependent Variable)	-0.252	-0.252	-0.252	-12.275	-12.275	-12.275
Group Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.663	0.668	0.688	0.766	0.767	0.769
N	1,800	1,800	1,800	1,800	1,800	1,800

Notes: Sophistication indicates a participant scoring four or above in the financial knowledge questions. Advanced Summary and Simplified Summary are indicators for summaries targeted at financially knowledgeable and less knowledgeable people, respectively. Panel A shows short-term investment results and Panel B shows long-term investment results. Standard errors are clustered by earnings calls. t-values are reported in the parentheses. *, **, *** represent statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

Table 5. Does AI Widen Knowledge-Driven Disparity?

	Information Processing			Investment Efficiency			
	Earnings Prediction		Sentiment	Short-Term		Long-Term	
	Numeric	Textual	Numeric	Returns	Sharpe	Returns	Sharpe
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\beta_2 + \beta_4 - \beta_3$	0.051***	0.056***	0.033***	0.004*	0.815**	0.079***	0.693*
t-value	(2.99)	(3.79)	(3.05)	(1.83)	(2.41)	(2.89)	(1.92)

Notes: Based on our baseline empirical specification Answer Quality$_{ij}$ = β_1Sophistication$_j$ + β_2Advanced Summary$_j$ + β_3Simplified Summary$_j$ + β_4Sophistication$_j$ × Advanced Summary$_j$ + β_5Sophistication$_j$ × Simplified Summary$_j$ + $\zeta_i + \varepsilon_{ij}$, we report the F-test results on whether $\beta_2 + \beta_4 - \beta_3 > 0$. We report converted t-values and their statistical significance. Columns (1) and (2) report F-test results from Table 2, Panel A, columns (3) and (6). Column (3) reports F-test results from Table 2, Panel B, column (3). Columns (4) and (5) report F-test results from Table 4, Panel A, columns (3) and (6). Columns (6) and (7) report F-test results from Table 4, Panel B, columns (3) and (6). *, **, *** represent statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

Table 6. Why Does Wedge Arise?

Wedge of	Earnings Prediction (1)	Sentiment Assessment (2)	Short-Term Returns (3)	Long-Term Returns (4)
|EBITDA – Net Income|	3.341**	2.769**	0.734*	3.322**
	(2.40)	(2.23)	(2.01)	(2.28)
|Accruals|	1.870*	2.323*	0.802	1.991*
	(1.69)	(1.88)	(1.35)	(1.89)
|Extraordinary Items|	1.683	1.580	0.603*	1.688
	(0.95)	(0.87)	(1.82)	(1.23)
Product Segments	0.005*	0.005*	0.003	0.006*
	(1.75)	(1.83)	(1.58)	(1.84)
R&D	3.885**	3.537**	0.559**	3.761**
	(2.28)	(2.30)	(2.13)	(2.48)
Share Repurchase	1.523*	1.736*	0.768**	1.422
	(1.91)	(1.80)	(2.13)	(1.60)
Dividend Payout	0.056**	0.060**	0.007**	0.053*
	(2.31)	(2.26)	(2.28)	(1.71)
Avg(Dependent Variable)	0.045	0.042	0.008	0.040
Year-Quarter Fixed Effects	Yes	Yes	Yes	Yes
Adjusted R2	0.128	0.137	0.068	0.076
N	200	200	200	200

Notes: |EBITDA – Net Income| is the difference between EBITDA and net income scaled by the total assets. |Accruals| is the absolute value of accounting accruals scaled by the total assets. |Extraordinary Items| is the absolute amount of extraordinary items in the income statement scaled by total assets. Product Segment is the number of product segments. R&D is the R&D expense scaled by total assets. Share Repurchase is the shares repurchased scaled by total shares outstanding. Dividend Payout is the dividend payout ratio. *, **, *** represent statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively. Continuous variables are winsorized at the 1% and 99% levels. Standard errors are clustered at the industry level.

Table 7. Questions Asked by Participants

	Sophistication		Difference
	High	Low	
	(1)	(2)	(1) – (2)
Summarization (%)	0.102	0.063	0.038**
			(2.15)
Specific financial issues (%)	0.651	0.508	0.143***
			(3.69)
Total number of questions	9.201	8.645	0.556
			(0.95)
Identical to survey questions (%)	0.140	0.215	-0.075***
			(-2.70)
No answers (Out of scope) (%)	0.106	0.093	0.013
			(0.59)
Invalid question (%)	0.001	0.121	-0.120***
			(-7.42)

Notes: We analyze the content of questions asked by the participants to the model. Total number of questions is the average number of questions asked by a participant. Summarization denotes questions that require a summary of the earnings call or a specific part of the earnings call. Specific financial issues are questions that inquire about specific themes (e.g., earnings and risks). Identical to survey questions denote questions that overlap with the questions we ask in the experiment and the model refused to answer. No answers denote questions that cannot be answered using information contained in the transcript only. Invalid questions are the ones with less than 3 words, followed by a model's answer with less than 10 words. The variables are scaled by the total number of questions asked by each participant. *, **, *** represent statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

Table 8. ChatBot Implementation

Dependent Variable =	Info Processing (Accuracy) Earnings (1)	Sentiment (2)	Investment Efficiency (Returns) Short-Term (3)	Long-Term (4)
Sophistication	0.056*	0.061	0.002*	0.004
	(1.86)	(1.61)	(1.95)	(0.30)
Chatbot	-0.066**	-0.062*	-0.003***	-0.055**
	(-2.09)	(-1.95)	(-2.70)	(-2.21)
Sophistication × Chatbot	0.105***	0.122***	0.008***	0.089***
	(3.09)	(2.95)	(2.78)	(3.20)
Avg(Dependent Variable)	-0.477	-0.478	-0.029	-0.252
Group Fixed Effects	Yes	Yes	Yes	Yes
Adjusted R^2	0.562	0.593	0.522	0.608
N	800	800	400	400

Notes: We examine the effect of using the ChatBot on information processing and investment efficiency. Chatbot is an indicator that equals one when a participant is assigned an interactive Chatbot. Sophistication is an indicator that equals one when a participant's financial literacy score is higher than or equal to four. Continuous variables are winsorized at the 1% and 99% levels. Standard errors are clustered by earnings calls. *, **, *** represent statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively.