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 show how generative AI shapes both overall performance and disparities in performance among investors in investment tasks. When investors are given AI summaries aligned with their sophistication, they become better at both 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 narrow 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, 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.

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 of adoption (e.g., Kelly and Xiu, 2023), which makes it more surprising that there is sparse academic evidence about the effects of generative AI on individual investment decisions. Motivated by this sparsity and the proliferation of AI into mainstream, we seek to address two central questions. First, do these tools enhance investors' ability to process news and make financial decisions? Second, does access to generative AI narrow or widen performance gaps across investors? We address these questions through two large-scale experiments examining how generative AI shapes individual investors' information processing and investment choices.

Many see 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 like writing and programming, generative AI improves overall performance but also enables novice users to produce outputs that closely match 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 that set it apart from other domains where AI has proven democratizing. In writing or programming tasks, the AI's output represents - more or less - the end goal. Financial decisions work differently. An AI-generated analysis of firm performance isn't valuable on its own - investors must interpret this information and take appropriate action. Consider news about complex metrics like derivative revaluations or goodwill impairments. AI can summarize news and explain these concepts clearly. Yet this information only becomes valuable when investors understand how these metrics jointly inform their given task. Thus, our first hypothesis is that AI improves information processing in investment tasks on average, but that this improvement hinges on AI outputs aligning with an investor's financial sophistication.

Investment sophistication also shapes 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 may omit subtle but crucial details about firm value. For example, explaining revenue recognition patterns in basic terms could obscure timing issues 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. Throughout our tests, we mask the identity of the firm to prevent participants from using outside information to inform their decisions.

The first experiment leverages AI summaries of earnings news, currently the most common application of AI in financial information processing (e.g., Ebert, 2023; Kim et al., 2024b). We recruited 1,800 participants and first 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). We then 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 is designed to provide participants with more flexibility in how they use generative AI to process earnings announcement news. Instead of pre-written summaries, we provide participants with an AI assistant - a chatbot that is pre-loaded with the earnings announcement transcript. This second experiment allows us to not only measure their performance on investment tasks but also to track how they engage with AI chatbots. For example, we can observe the questions they ask to the AI chatbot and study how the nature of their questions varies with investor sophistication. The control group, as before, work with raw transcripts of the earnings announcement conference call.

[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."

Both experiments examine two key aspects of investor behavior. First, we test how well participants process information. Specifically, we test how well they can assess the overall sentiment of the earnings announcement and forecast the direction of future earnings changes. In line with Chen et al. (2022), we compare participants' judgments against 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 actual earnings announcements. We analyze releases from 2010 to 2022, carefully matching pairs of companies. 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. This design improves upon studies using hypothetical scenarios (Lian et al., 2018; Heeb et al., 2022) or limited firm samples (Bonsall et al., 2017), and allows us to evaluate decisions against actual market outcomes.

Our first main test shows that when AI summaries align with participants' level of financial sophistication, it significantly improves 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 earnings prediction accuracy 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 also find that misaligned summaries do not appear to help and sometimes hinder performance. Less sophisticated users are no better off from receiving the advanced summaries that are written for more sophisticated investors compared to receiving the raw transcript. Similarly, sophisticated users don't benefit from simplified summaries that are written for less sophisticated investors; they appear at least as well off as when processing the raw transcripts. This result speaks to one of our motivating questions about what specific aspects of human-AI interactions shape the implications for 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 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, we also find that 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 decisions faster on average. 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 likely 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 and facilitates investment tasks is by reducing the complexity of information processing.

Our findings on participants' investment decisions also reveal a striking paradox. While AI may make markets more accessible by improving the performance of less sophisticated investors, it may simultaneously widen, rather than narrow, 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: specifically 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 performance gaps in investment tasks by disproportionately benefiting more sophisticated investors.

We next provide evidence 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 that "costs increased," the advanced version specifies "operating margins declined due to higher input costs" - creating a gap in the precision of the signals that participants receive. Consistent with this idea, we find that the performance gap across more versus less sophisticated participants widens among earnings announcements where there is more technical information omitted from simplistic summaries such as R&D spending, share repurchases, and gaps between operating versus total profits. This suggests that efforts to make financial information more accessible through AI may 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 versus less sophisticated participants. Specifically, in revisiting the same tasks studied in the first experiment, we see that more sophisticated participants see greater gains from AI when assessing earnings news, forecasting future earnings changes, and forming investment portfolios. When combined with our earlier findings on participants' questions, the results of this section suggest that differential command of 'what questions to ask' is another channel through which access to generative AI increases pre-existing disparities across investor groups. Moreover, the consistency of our inferences across the first and second experiment suggest that the main findings of our paper are unlikely limited to a specific modality of human-AI interaction (i.e., AI summaries versus AI chatbots).

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 (over 10 seconds) to ensure engagement with provided materials only. These findings mitigate concerns that our results are sensitive to specific aspects of our main tests.

A central contribution is to provide some of the first systematic 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 to make sense of earnings news, it also aids them in forecasting firms' performance in terms of both earnings changes and stock returns. The positive impact of AI on less sophisticated participants' performance is particularly noteworthy 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 broader population of investors.

Second, we demonstrate that 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 that suggests that AI's impact on individual task performance likely varies with 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 experiment structure. 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

Our experiment aims 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. These 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). Also, research shows that retail investors actively trade around earnings announcements (e.g., Kaniel et al., 2012), making it a suitable setting to understand how generative AI influences investors' decision-making. We specifically focus on how investors process information from the earnings announcement conference call, which covers a variety of topics including summaries of past performance, outlook for the future, and 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 receive 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 decisions surrounding firm disclosures and investments) as well as asking for the participants' consent. The survey 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 the participants are equipped with 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 augment this set by adding other questions that are more specialized to our experiment's setting, which examine whether the respondents are likely able to utilize earnings news to inform their investment decisions. This section of the survey does not differ across treatment arms, but answers from this part are used to divide the population into more versus less financially sophisticated groups.

The second part aims to evaluate investors' ability to accurately process earnings announcement information. We ask participants to assess two distinct firms (instead of 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 answer all questions for the first firm before proceeding to the second firm, where they repeat the same process.[2] Participants in the control group receive a copy of an anonymized earnings conference call transcript, whereas participants in the treatment groups either receive an advanced summary or a simplified summary. Participants are prompted to review the provided material, assess its content, and then answer questions about their earnings expectations and the sentiment of the news.

Forming a belief about each firm's 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 a company's prospects. 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.[3]

We next assess participants' ability to summarize the implications of the earnings announcment 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

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

[3]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 can be negatively impacted.

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

The third part of the experiment is a mock investment exercise. We ask participants to allocate $1,000 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 investment and a long-term investment. For the short-term investment horizon question, participants are prompted to form an investment portfolio that they only hold for one day.[4] For the long-term investment question, they have an investment horizon of a year. After selecting their allocation plan, they must also write a short rationale for their asset allocation decisions. Similarly to the previous part of the survey, we collect the time spent by the participants to finalize their asset allocation decisions.

Finally, to rule out that participants inadvertently identify the firms they are assigned, we ask respondents whether they can identify the firms' names from the anonymized documents. We also ask participants whether they can guess the study's ultimate goal. The full survey is provided in the Online Appendix.

Summary Generation A key ingredient in our experiment is the use of summaries automatically generated by generative AI. We use AI summaries to approximate how an investor likely uses the technology to distill information from transcripts of earnings announcement conference calls, which are typically lengthy due to a mixture of value-relevant information and filler content. Additionally, our use of 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 conference 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

[4]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.

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

Finally, we anonymize the names of people, regions, and products for both the 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.[5]

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.[6] 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 Group 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.

[5]To verify this notion, 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.

[6]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.

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 the first part of our survey.[7] 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. In untabulated tests, we find that our main findings do not appear sensitive to the use of the above versus below median cutoff for more versus less sophisticated investors.

Our survey takes approximately 20.97 minutes to complete, and we consider participants who take more than 63 minutes to have timed out. Based on this time-out rule, our survey's return rate was approximately 79.8%. The remaining 20.2% of the respondents did not return the survey or timed out. One potential concern is that the timed-out respondents were the least sophisticated people and that they could not finish their survey in time. To mitigate this concern, we compare the average financial knowledge score of the timed-out participants and others. 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. These findings indicate that the timed-out participants are not systematically more or less sophisticated than those who successfully complete their survey.

2.2 Firm Selection

We provide participants with actual (anonymized) earnings call information. Unlike studying hypothetical earnings announcements, the key advantage of using real firm data is that it allows us to assess whether participants make better decisions by benchmarking it with ex-post outcomes. Additionally, studying how people process and act on real earnings announcement information better replicates common investment tasks used in practice.

To select a pool of candidate earnings calls, we focus on the fourth fiscal quarter earnings

[7]When forming this score, we exclude the demographic question on the educational background and the question on investment experience, which only validates our initial participation condition by Prolific. 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.

calls that happened after or before trading hours from 2010 to 2022 for ordinary stocks on the NYSE or Nasdaq. We focus on these stocks to alleviate concerns that earnings information might already be reflected in stock prices before the call concludes. For similar reasons, we 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) to reduce noise from low liquidity. Each participant evaluates a pair of firms that meet this criteria.

We randomly select 200 stocks, or equivalently 100 pairs, of stocks that are economically similar but where the information revealed in the earnings call is somewhat different. To construct these firm pairs, we first group stocks based on fundamental economic characteristics. Stocks within each group must operate in the same industry as defined by four-digit SIC codes, release earnings on the same date to control for industry- and macroeconomic-conditions, fall into the same quartile for market capitalization, book-to-market ratio, and return-on-equity, and have current EPS values that are either both positive or negative to control for asymmetric earnings news. For groups with more than two stocks, we choose two stocks that yield the largest return gap over the next trading day from the earnings announcement date, such that announcements are distinct in content and implications.[8]

To examine the representativeness of the stocks used in our experiment, Table 1, Panel B compares the characteristics of our selected stocks against the broader population of stocks. We report the difference in means between the two groups for market capitalization, book-to-market, ROE, and earnings surprise based on the random walk or analysts' consensus. Overall, the financial fundamentals of the stocks used in our experiment are generally comparable to those of the broader set of stocks. However, the average book-to-market of stocks in our sample is smaller than that of the broader universe.

2.3 Measurements

We ask participants to assess the overall sentiment of the earnings announcement, forecast the direction of future earnings changes, and form short- and long-term portfolios. We compare their responses to actual earnings and returns observed in the data to evaluate their predictions. These actual outcomes serve as useful benchmarks, as earnings announcements contain information that can help forecast future earnings and returns.

[8]In Section 5.2, we conduct 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.

However, exogenous shocks unrelated to earnings announcements (or missing contextual information) can also influence the predictability of earnings and returns following the earnings release. Indeed, for any given firm, it is challenging to attribute differences between predicted and actual outcomes solely to participants' information processing abilities, given the factors beyond the information provided in the earnings announcements drove the actual outcomes. However, prior literature shows that, on average, earnings calls contain information about future earnings and returns (Frankel et al., 2022; Kim et al., 2024b). Thus, we can address this issue using a large sample of actual earnings release events. Our large sample approach also allows us to average out noise and mitigate the influence of any single outlier event or industry-specific anomaly. Put differently, this approach ensures that our measurements of participants' predictive accuracy and investment efficiency can be unbiased estimates of their true information processing abilities (even if for any given participant and event, we do not know what is the "correct" answer given the provided information).

Earnings Prediction We compare the participants' answers with the realizations to assess the accuracy in real-world investment tasks. The accuracy of participant j's earnings prediction after reviewing the earnings 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.[9] 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

[9]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.

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 instruct it to generate economic explanations for the observed change based on the transcript content. We then use these outputs to train a machine-learning model on the embeddings of these texts (see Online Appendix for detailed steps). This process ultimately allows the model to learn the key economic events and patterns 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.[10]

Sentiment Along with earnings prediction, we ask participants to assess whether the news positively or negatively impacts the firm's stock price. Their answers again span from -5 to $+5$, similar to the earnings prediction. We use the same formula as Equation 1 to evaluate the accuracy. However, instead of $\mathbb{1}(\text{Increase})_{ij}$ being an indicator for EPS improvement, we use cumulative stock returns around the earnings announcement dates (from one day before to one day after the announcement) as a benchmark. This approach is conceptually consistent

[10]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).

with Chen et al. (2022). $\mathbb{1}(\text{Increase})_{ij}$ takes a value of one when announcement returns are positive and zero otherwise.

Investment Efficiency In the third part of the survey, we ask participants to make investment decisions between the two firms, or to hold cash. We measure investment efficiency by comparing the participants' answers with hypothetical maximum realized returns and Sharpe ratios. Participants provide their asset allocation plan for the stocks (stocks A and B) and cash holdings. Their allocation amounts add up to \$1,000. We compute 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 must provide their asset allocation plans for both short-term and long-term. Consistent with the information provided in the survey, we define short-term returns as the daily returns measured from when the market opens on the trading day immediately following the earnings conference call until it market closes on the same day. We measure long-term returns as the cumulative returns from when the market opens on the trading day immediately following the earnings conference call, extending up to 252 trading days after that.

We obtain participant j's realized portfolio (m) return and Sharpe ratio after reading the pair of earnings call information k in both investment scenarios.

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

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

σ_A and σ_B denote the standard deviation of stock A and B returns. σ_{AB} is a covariance of stock A and B returns. For the one-day investment, we use a 30-minute interval of intraday returns to compute return volatility; for the one-year investment, we use daily returns. We then assess the optimality of the participant's choice by comparing it against the theoretical maximum returns and Sharpe ratio, defined 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 define investment efficiency as the signed difference between the realized returns (Sharpe ratio) and maximum returns (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. In Section 3.1 and Section 3.2, we present the results from our first experiment. In Section 3.3, we examine how our main findings vary across different profiles of firms and earnings announcement characteristics.

3.1 Assessing Information Processing

Our first main tests examine whether generative AI impacts investors' ability to accurately 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 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 participant j's earnings prediction accuracy (Equation 1) or the quality of the provided narratives (Equation 2) after reading a document j about a firm's earnings information. Sophistication$_j$ is an indicator variable for financially sophisticated respondents. Advanced Summary$_j$ is an indicator variable that equals one when participant j is assigned an Advanced Summary and zero otherwise. Simplified Summary$_j$ is defined similarly. As the difficulty of earnings prediction might 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.[11]

In this specification, the coefficient on Sophistication (β_1) indicates the relative performance of financially more sophisticated participants when they receive raw earnings call

[11] We examine the robustness of our findings by implementing 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 to those obtained from the main specification.

Figure 1. **Regression Designs**

	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).

transcripts compared to less sophisticated participants when they receive raw earnings call transcripts (hereafter denoted as the baseline group). The coefficient on Advanced Summary (β_2) represents the incremental effect of treating the less sophisticated group with advanced summaries. The coefficient on Simplified Summary (β_3) is the effect of treating the less sophisticated group with simplified summaries. The treatment effect of advanced summaries to the more sophisticated group can be measured as $\beta_1 + \beta_2 + \beta_4$. Lastly, the treatment effect of providing simplified summaries to the more sophisticated group is $\beta_1 + \beta_3 + \beta_5$. Figure 1 summarizes the measurement of treatment effects using the regression coefficients.

3.1.2 Estimation Results

Earnings Prediction Table 2, Panel A, column (1) shows that more financially sophisticated people ($\text{Sophistication}_j = 1$) make more accurate earnings forecasts compared to less sophisticated people on average. The effect is sizable with the magnitude of the coefficient of 0.027 (t-value = 3.85), which is about 5.66% of the absolute mean dependent variable. Then, in column (2), we examine the average effect of two summaries on the prediction accuracy and find advanced summaries improve average accuracy while simplified summaries do not.

Column (3) of Table 2, Panel A highlights the importance of aligning AI summaries with participants' financial sophistication. When summaries are aligned with the sophistication of the investor, prediction accuracy improves markedly. For less sophisticated investors, receiving simplified summaries enhances prediction accuracy (0.035, t-value = 2.94), while receiving advanced summaries reduces it (-0.006, t-value = -0.50). Similarly, more sophisticated participants benefit from advanced summaries, but misaligned summaries (i.e., simplified summaries) hinder their performance. Specifically, the interaction of Sophistication with Simplified Summary yields a negative and significant coefficient (-0.054), whereas its interaction with Advanced Summary results in a positive and highly significant coefficient (0.092, t-value = 5.29).

Similarly, we find that the informativeness of participants' short answers in earnings

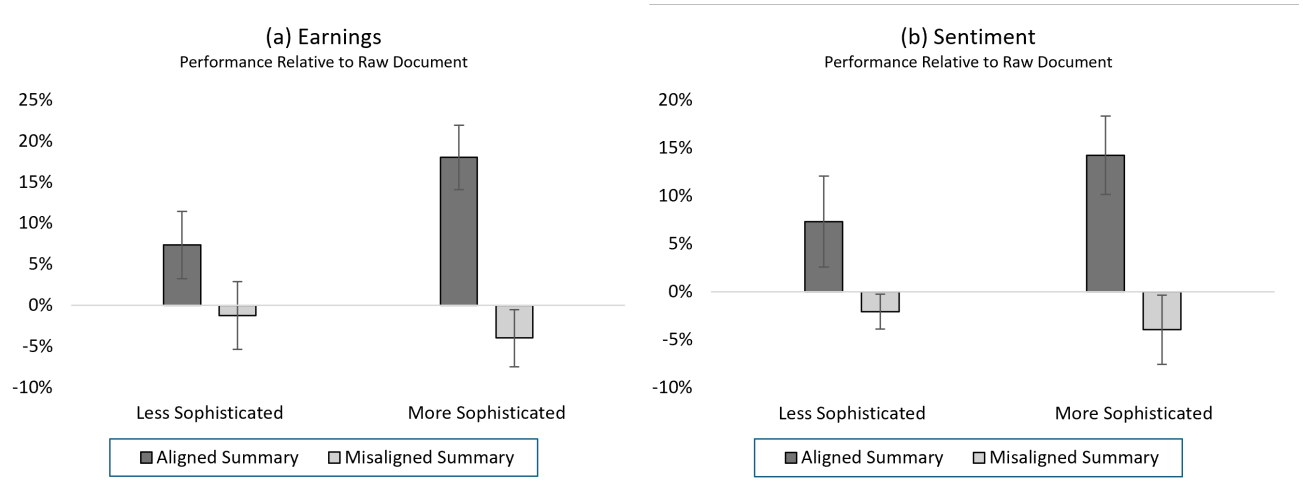
prediction tasks increases when they are provided with aligned summaries. As part of the survey, respondents provide a narrative answer (at least 50 words) explaining the rationale behind their prediction. Our trained artificial neural network assesses whether the participants' answers provide helpful information about the direction of future earnings changes following Eq 2. The results are shown in columns (4) to (6). Overall, the results are consistent with our previous analysis. Financially sophisticated people's answers are, on average, more informative than those of less sophisticated people. When financially sophisticated people are provided with advanced summaries, their narratives' quality improves. The same applies to the less sophisticated group and simplified summaries.

Sentiment Table 2, Panel B, shows that aligned summaries also improve participants' ability to process the implications of the earnings news for stock prices (i.e., the new sentiment). These tests parallel our earnings prediction tests, but here we use sentiment evaluation accuracy as the outcome variable. Consistent with our results in Panel A, we see that financially sophisticated participants outperform their less sophisticated counterparts on average (column (1)), demonstrating greater accuracy in evaluating the sentiment of financial news. Advanced AI summaries significantly enhance accuracy on average (coefficient = 0.032, t-value = 3.17), while simplified summaries provide only marginal improvements (coefficient = 0.010, t-value = 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 see limited benefits from advanced summaries.

Visualization In Figure 2, we visually display our main findings from Table 2. We do so by plotting the cumulative coefficients from columns (1) through (3) from Table 2 scaled by the absolute mean of each dependent variable. After subtracting the control group (raw transcript) outcomes, Panel A of Figure 2 shows that 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 delve into further in Section 4.

Panel B of Figure 2 visualizes the results for the sentiment evaluation task. Aligned summaries improve accuracy by approximately 7% for less sophisticated participants and 14% for more sophisticated participants, underscoring the importance of tailoring AI outputs to match user expertise.

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



Notes: This figure illustrates the effect of providing AI-generated summaries to 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.

Taken together, our findings in Table 2 and Figure 2 suggest that AI-generated summaries can significantly enhance 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 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 In Table 3, we provide evidence that AI summaries raise participants' confidence in their decisions while also lowering the time they spend on our investment tasks. In Panel A of Table 3, we examine changes in participants' confidence before and after receiving AI-generated summaries. Participants are asked to rate their confidence in answering earnings prediction and sentiment assessment tasks on a scale from 0 (entirely uncertain) to 1 (entirely certain).

Columns (1) and (4) of Panel A Table 3 show that, for both tasks, financially sophisticated participants consistently report higher confidence levels than their less sophisticated counterparts. Additionally, columns (2) and (5) show that providing participants with AI-generated summaries further boosts confidence overall. Notably, the coefficient for Advanced Summary exceeds that for Simplified Summary, indicating that tailored summaries

designed for more sophisticated users enhance confidence despite their greater complexity.

However, columns (3) and (6) of Table 3, Panel A provides evidence that higher reported confidence is primarily driven by *less* sophisticated participants when all variables are included in a single regression model. By 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.

The first three columns of Table 3, Panel B, show that AI summaries also significantly improve the speed of financial decision-making in the earnings prediction task. Using the coefficients in column (2), we estimate that advanced summaries reduce decision time for sophisticated participants by 75.62 seconds — a 21.80% decrease from the average. When all variables are included simultaneously in the regression (column (3)), we do not observe the asymmetric effects seen in timeliness. Both types of summaries reduce decision times for participants regardless of their alignment with user sophistication. Similar results emerge for sentiment assessment tasks, though the effects are statistically weaker.[12]

3.2 Assessing Investment Efficiency

In the previous section, we show that aligned generative AI outputs improve investors' ability to process earnings news. In this section, we examine whether this improvement ultimately translates into better financial portfolio management.

3.2.1 Empirical Design

We modify our earlier tests in Section 3.1.1 by asking participants to allocate $1,000 into three asset classes: the two stocks and cash. To analyze the impact of AI summaries, we estimate the following regression at the stock-pair level:

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

Answer Quality$_{kj}$ is either the difference between the maximum obtainable returns and the realized returns or the difference between the maximum obtainable Sharpe ratios and the

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

realized Sharpe ratios. We calculate the realized returns and Sharpe ratios, based on the allocation plans for both short-term (return on the day following the earnings announcement) and long-term investments (from one day after the earnings announcements to one year after the announcements).[13]

3.2.2 Estimation Results

Short-Term Investments Participants are asked to form a portfolio when the market opens on the following trading day after the earnings call (which was held when the market was closed). Participants can invest in either one of the two stocks or hold their endowment as cash.

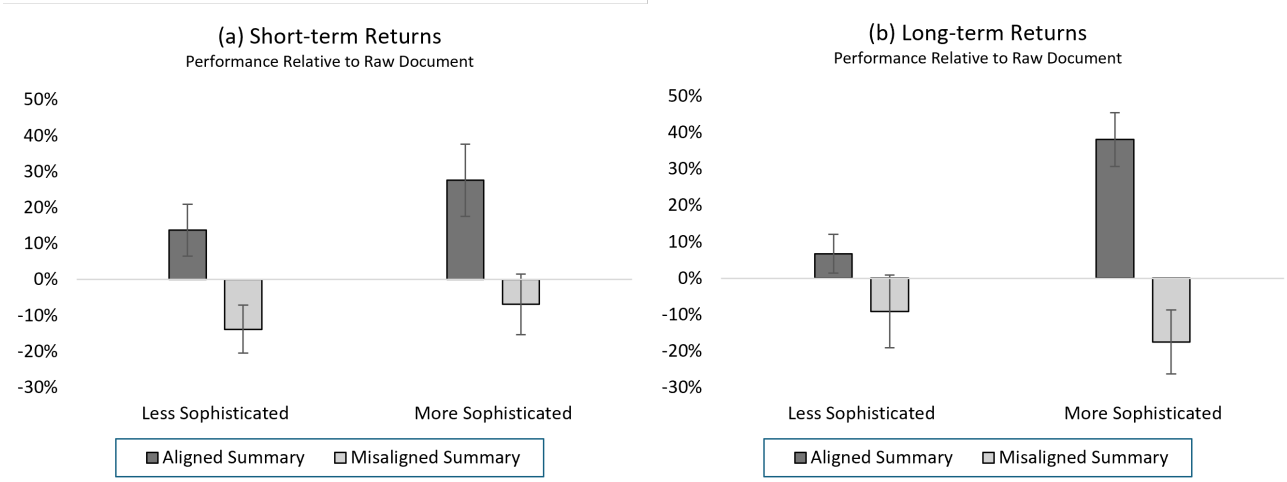
Table 4, Panel A shows that, in line with our earlier findings, financially sophisticated participants outperform their less sophisticated counterparts in terms of portfolio returns on average (column (1)). Providing AI-generated summaries alone does not significantly improve overall investment profitability (column (2)). However, when we account for the alignment between summary type and participants' financial sophistication (column (3)), a clearer pattern emerges. Sophisticated participants benefit substantially from advanced summaries, achieving higher profitability. Conversely, their performance deteriorates when given 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.

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

Difference between Short- and Long-Term Investments Distinguishing between short- and long-term investments requires a nuanced understanding of financial concepts,

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

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.

which might pose challenges for survey participants. Nevertheless, our analysis in the Online Appendix demonstrates that participants appear to understand these distinctions reasonably well. Specifically, participants allocate significantly different weights for their short- and long-term investments. Furthermore, when short-term allocation weights are applied to long-term scenarios, the long-term portfolio returns deteriorate significantly, particularly among the more financially 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 decision-making across 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 conditions under which AI-generated information provides the greatest value to investors.

3.3.1 Empirical Strategy

We estimate the effect of AI-generated summaries for each announcement by measuring the difference in participants' average performance when treated with aligned summaries relative to raw transcripts. For simplicity, we introduce an indicator $I_j^{k,d}$ where $k \in (s, us)$

where s denotes a sophisticated participant i.e., $\text{Sophistication}_j = 1$, and us denotes a less sophisticated participant, i.e., $\text{Sophistication}_j = 0$, and $d \in (as, ss, r)$ where as denotes an advanced summary, ss a simplified summary, and r a raw transcript. $I_j^{k,d}$ equals one when a participant of type k is assigned a document type d and zero otherwise. For example, when a participant j is classified as a sophisticated participant and is treated with an advanced summary, then $I_j^{s,as} = 1$. AQ_{ij} is the answer quality of participant j for earnings call i. Then, the effect of AI for each earnings call i is defined as:

$$\text{AI Effect}_i = \underbrace{\left(\frac{\sum_j \text{AQ}_{ij}\left(I_j^{s,as} + I_j^{us,ss}\right)}{\sum_j \left(I_j^{s,as} + I_j^{us,ss}\right)}\right)}_{\substack{\text{Average Answer Quality} \\ \text{Given Aligned Summaries}}} - \underbrace{\left(\frac{\sum_j \text{AQ}_{ij}\left(I_j^{s,r} + I_j^{us,r}\right)}{\sum_j \left(I_j^{s,r} + I_j^{us,r}\right)}\right)}_{\substack{\text{Average Answer Quality} \\ \text{Given Raw Transcripts}}} \tag{9}$$

AI Effect_i is a call-level measure (i.e., one per earnings announcement) and the first term is the averaged Answer Quality when participants are treated with summaries aligned with their level of financial sophistication (i.e., when $\text{Sophistication}_j = 1$, then we use $\text{Advanced Summary}_j = 1$ and $\text{Sophistication}_j = 0$, then we use $\text{Simplified Summary}_j = 1$). The second term is the averaged Answer Quality when participants are treated with raw transcripts. We use all four measures of Answer Quality (i.e., earnings prediction, sentiment assessment, short-term portfolio returns, and long-term portfolio returns). We then regress AI Effect_i on several firm-level characteristics.

$$\text{AI Effect}_i = \gamma \mathbf{X}_i + \delta_t + \vartheta_l + \varepsilon_i \tag{10}$$

$\mathbf{X}_i$ is a vector of call-level (i.e., firm-year-level) characteristics. We expect AI summaries to be relatively more informative when a firm's information environment is more complex and opaque (Kim and Nikolaev, 2023; Kim et al., 2024a). Therefore, we include firm size and book-to-market ratio to account for the difficulty in valuation. We also include earnings volatility, an indicator for operating loss, earnings surprise, and return volatility to proxy for the complexity of information. In addition to firm-level characteristics, we include year-fixed effects (δ_t) to control for unobserved macroeconomic trends and industry-fixed effects measured at the two-digit SIC codes (ϑ_l) to control for time-invariant industry-level characteristics. Continuous variables are winsorized at 1% and 99% levels to mitigate the effect of outliers. Standard errors are clustered at the firm level.

3.3.2 Results

Table OA-1 of our Online Appendix shows that aligned summaries benefit participants more when the investment task is more likely complex. Across the four dependent variables, we consistently 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 indicate that AI-generated information provides greater value to investors when firms exhibit volatile earnings, report losses, and experience high stock return volatility (e.g., Kim et al., 2024b). Furthermore, AI-generated summaries provide more benefits on average when firms have a lower book-to-market ratio, suggesting that their value increases when firms are more challenging to evaluate. This is consistent with the marginal benefit of AI adoption grows as investors face higher processing costs 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 groups?

To address this question, we first introduce a simple motivating model in Section 4.1, which highlights the theoretical and empirical ambiguity of AI adoption's effect on 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 motivating model. Finally, in Section 4.4, we explore the determinants that drive changes in this knowledge-driven gap, shedding light on the economic mechanisms underlying these disparities.

4.1 A Motivating Model

We introduce a simple model explaining under which economic conditions the knowledge-driven disparity among investors will be exacerbated after adopting AI technology. Specifically, within the context of our experiment, this model considers how AI technology can simplify complex corporate disclosures, potentially making them more accessible at the expense of losing some information.

We assume a continuum of investors indexed by $i = 1, 2, \cdots, N$ whose comprehension level is characterized by a_i, which is distributed between 0 and $\bar{a}$. We also assume that when $i < j$, $a_i < a_j$. A disclosure D's complexity is denoted by C and its informativeness, $I(C)$, increases in complexity, i.e., $\frac{\partial I(C)}{\partial C} > 0$. Intuitively, this assumption implies that more informative disclosures are, in general, more complex in their content. The utility of processing the disclosure is increasing in its informativeness and decreasing in the processing cost to interpret it. The processing cost is heterogeneous across investors. It increases in complexity but decreases with an investor's ability (i.e., comprehension level).

To illustrate our point, we impose some simple functional relations among the variables. First, we assume that $I(C) = \mu C$ ($\mu > 0$).[14] Second, we formulate investor i's processing cost $Q_i(C)$ as $Q_i(C) = \frac{C^2}{2a_i}$. We use a quadratic cost function following Sims (2003) and Maćkowiak and Wiederholt (2009). Investor i's utility of processing a disclosure of complexity C is

$$u_i(I(C), C, a_i) = \max\left[\mu C - \frac{C^2}{2a_i}, 0\right] \tag{11}$$

Consider a language technology that can adjust the complexity of a given disclosure to optimize investors' utility by providing summaries. The model finds the optimal level of complexity C^* that matches the user's comprehension level. Our assumption that $I(C) = \mu C$ implies a trade-off between complexity and informativeness in documents. The original disclosure, denoted as D_R, has a complexity of C_R. For each investor i, the technology optimizes C_i by solving the following first-order condition:

$$\frac{\partial u_i}{\partial C_i} = \mu - \frac{C_i}{a_i}. \tag{12}$$

Equating the first-order condition to zero leads to $C_i^* = \mu a_i$, which denotes the customized optimal level of complexity for investor i. In this case, the investor's utility becomes $\frac{\mu^2 a_i}{2}$. Also note that the summary's complexity level should be lower than that of the original document, i.e., $C_i^* \leq C_R$.[15] Therefore, investors whose comprehension level is higher than $\frac{C_R}{\mu}$ do not require summaries but rather use the raw document.

This leads to an interesting observation that adopting technology increases investors' use of complex yet helpful information in the stock market. Before the adoption of the technology,

[14] For computational brevity, we assume a simple linear relation between C and $I(C)$. However, the results are consistent even when we assume a concave relation, i.e., $\frac{\partial^2 I(C)}{\partial C^2} < 0$.

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

investors use the disclosed information only when

$$u_i(I(C_R), C_R, a_i) = \mu C_R - \frac{C_R^2}{2a_i} > 0, \tag{13}$$

which implies a participation constraint $a_i > \frac{C_R}{2\mu}$. This indicates that investors below certain threshold of comprehension level do not use the information for their investment as their net benefit of document interpretation is always negative. However, after adopting the technology, the utility of each investor becomes $\frac{\mu^2 a_i}{2} > 0$, and all investors benefit from using the information, increasing the use of information in the market as a whole.

We define the function $\Delta(a_i)$ as the net utility increase obtained from the technology's summarization:

$$\begin{aligned}\Delta(a_i) &= u_i(I(C_i^*), C_i^*, a_i) - u_i(I(C_R), C_R, a_i) \\ &= \begin{cases} 0 & \text{if } \frac{C_R}{\mu} < a_i \leq \bar{a} \\ \frac{\mu^2 a_i}{2} - \left(C_R\mu - \frac{C_R^2}{2a_i}\right) & \text{if } \frac{C_R}{2\mu} < a_i \leq \frac{C_R}{\mu} \\ \frac{\mu^2 a_i}{2} & \text{if } 0 \leq a_i \leq \frac{C_R}{2\mu} \end{cases}\end{aligned} \tag{14}$$

This function implies that all investors weakly benefit from the technology, i.e., $\Delta(a_i) \geq 0$. This indicates that the overall utility of society increases after adopting the technology:

$$\int_0^{\bar{a}} \Delta(a_i) da_i > 0. \tag{15}$$

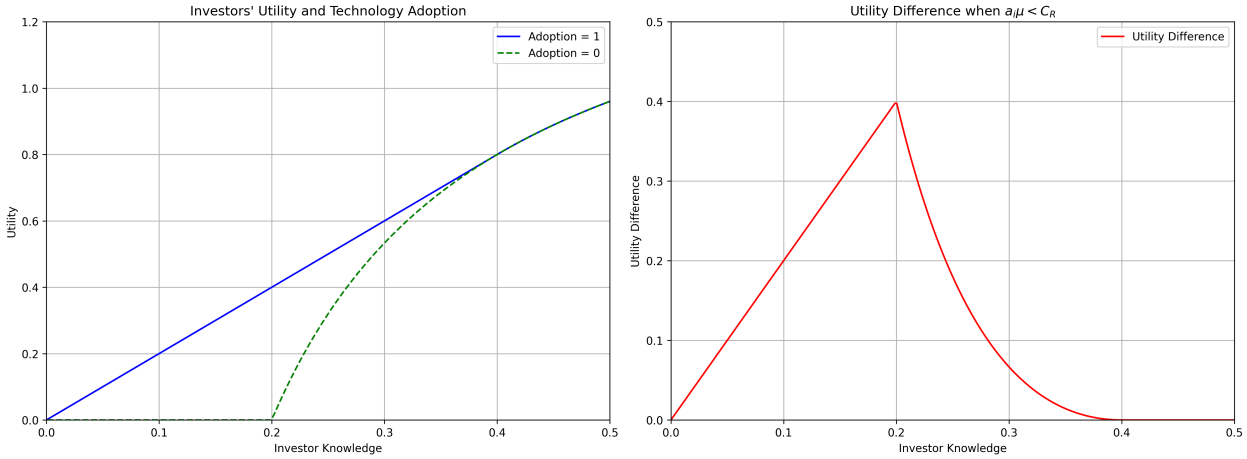
However, whether $\Delta(a_i)$ increases or decreases with a_i is not trivial and depends on the relative value of a_i.

$$\frac{\partial \Delta(a_i)}{\partial a_i} = \begin{cases} 0 & \text{if } \frac{C_R}{\mu} < a_i \leq \bar{a} \\ \frac{\mu^2}{2} - \frac{C_R^2}{2a_i^2} & \text{if } \frac{C_R}{2\mu} < a_i \leq \frac{C_R}{\mu} \\ \frac{\mu^2}{2} & \text{if } 0 \leq a_i \leq \frac{C_R}{2\mu} \end{cases} \tag{16}$$

Therefore, $\Delta(a_i)$ is constant in a_i when $a_i > \frac{C_R}{\mu}$, decreases in a_i when $\frac{C_R}{2\mu} < a_i < \frac{C_R}{\mu}$, and increases in a_i when $a_i < \frac{C_R}{2\mu}$. Intuitively, the technology widens the knowledge-driven disparity when the user's comprehension level is low. The technology reduces the gap when the user's comprehension level is in the middle range. The technology does not make a difference when the user already has a very high level of comprehension.

Figure 4 displays the investors' utility before (dotted line) and after (solid line) adopting the technology. We assume $\mu = 2$ and $C_R = 0.8$. As predicted by the model, when $a_i < \frac{C_R}{2\mu} = 0.2$, investors with high comprehension levels benefit relatively more than investors with low comprehension levels. When $\frac{C_R}{2\mu} = 0.2 < a_i < \frac{C_R}{\mu} = 0.4$, investors with low comprehension levels benefit relatively more than investors with high comprehension levels. When $a_i > 0.4$, there is no difference before and after the technology adoption. Therefore, whether a certain group benefits relatively more from the technology remains an empirical question.

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 $0 < a_i < 0.4$.

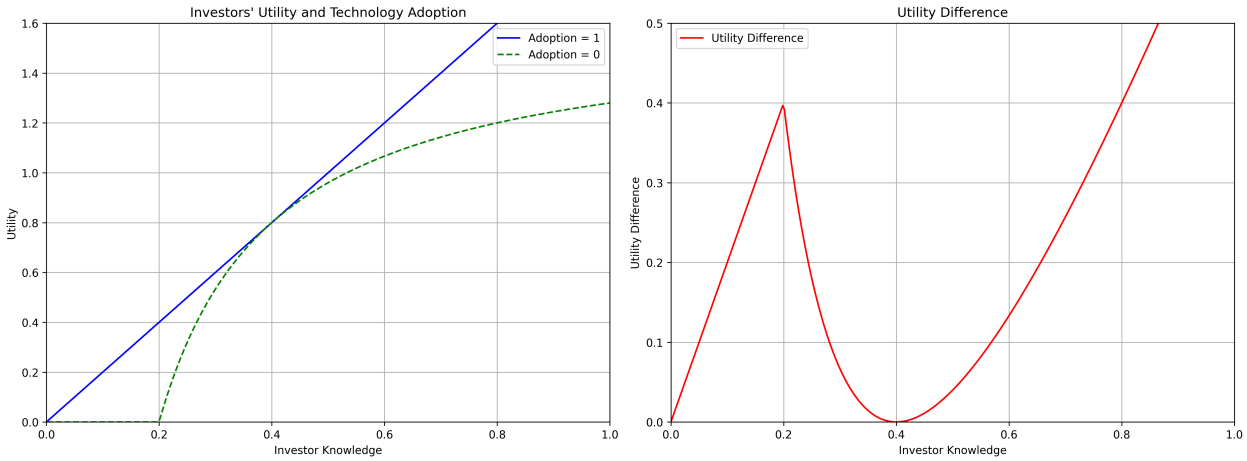
Holding everything else constant, if we increase the raw disclosure's complexity C_R, we find that the support of a_i that yields a positive $\Delta(a_i)$, $\left[0, \frac{C_R}{\mu}\right]$, expands. The support of investors whose $\Delta(a_i)$ increases with ability, $\left[0, \frac{C_R}{2\mu}\right]$, also increases. Holding the entire distribution of ability constant, it becomes more likely that high-ability investors benefit more from the technology than low-ability investors.

Note that, we restrict C_i^* to be lower than C_R. This is because in our experiment, we instruct the model to confine its answers to the content provided in the conference call transcript only. In reality, however, the model can search its knowledge base and synthesize it with the content provided in the earnings call. Doing so, it can generate economic insights that are not present in the transcript, allowing C_i^* and $I(C_i^*)$ to become larger than C_R and $I(C_R)$. In such case, $\frac{\partial \Delta(a_i)}{\partial a_i} = \frac{\mu^2}{2} - \frac{C_R^2}{2a_i^2}$ when $a_i > \frac{C_R}{\mu}$. Therefore, $\Delta(a_i)$ is increasing in a_i when investors are highly sophisticated, further widening the knowledge-driven disparity between

the sophisticated and unsophisticated investors.

Using the same numerical setup as Figure 4, we also plot the case where the restriction on C_i^* is removed (we assume that $\bar{a}$ is sufficiently large) in Figure 5. As expected, when $a_i < 0.4$, $\Delta(a_i)$ is identical to Figure 4. However, when $a_i > 0.4$, sophisticated investors do not use the raw transcript anymore but also rely on the technology. Therefore, more sophisticated investors can benefit relatively more if the model can access external sources of information and augment the transcript's content.

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 AI-generated summaries widen the knowledge gap by examining their treatment effect when aligned with users' financial sophistication. If the knowledge disparity widens, the treatment effect should be larger for more sophisticated participants compared to less sophisticated ones (and vice versa if it narrows). Within our regression framework (e.g., equation 1), we evaluate whether the expression $\beta_2 + \beta_4 - \beta_3$ is statistically different from zero (see Figure 1).[16]

[16]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 $\beta_1 + \beta_2 + \beta_4 - \beta_3$.

The performance disparity increases after AI treatment, as evidenced by our results in Table 5. For the earnings prediction task, the statistic measuring the performance gap between more and less sophisticated participants has a value of 0.051 (from column (3) of Table 2, Panel A), representing 10.69% of the average prediction accuracy. Similarly, for the sentiment assessment task, the statistic is 0.033 (from column (3) of Table 2, Panel B), or 6.90% of the average classification accuracy. Both values are positive, statistically significant, and associated with t-values of 2.99 and 3.05, respectively. In the short-term return prediction tasks, the performance disparity between participants of differing financial sophistication becomes even more pronounced after treatment, with $\beta_2 + \beta_4 - \beta_3 = 0.004$ (t-value = 1.83). A similar pattern emerges when we benchmark against the Sharpe ratio (column (5)). For long-term return prediction tasks (columns (6) and (7)), we observe consistent results. Specifically, in terms of one-year portfolio returns, the availability of AI increases the performance gap between the two investor groups by approximately 7.9%.

Overall, our empirical results suggest that AI-generated summaries, while enhancing overall information processing and decision-making, exacerbate disparities between sophisticated and less sophisticated participants. Although AI has the potential to democratize financial insights, our results highlight a trade-off: it disproportionately benefits those with greater financial sophistication, thereby widening the performance gap between investor groups.

4.3 Model Fit

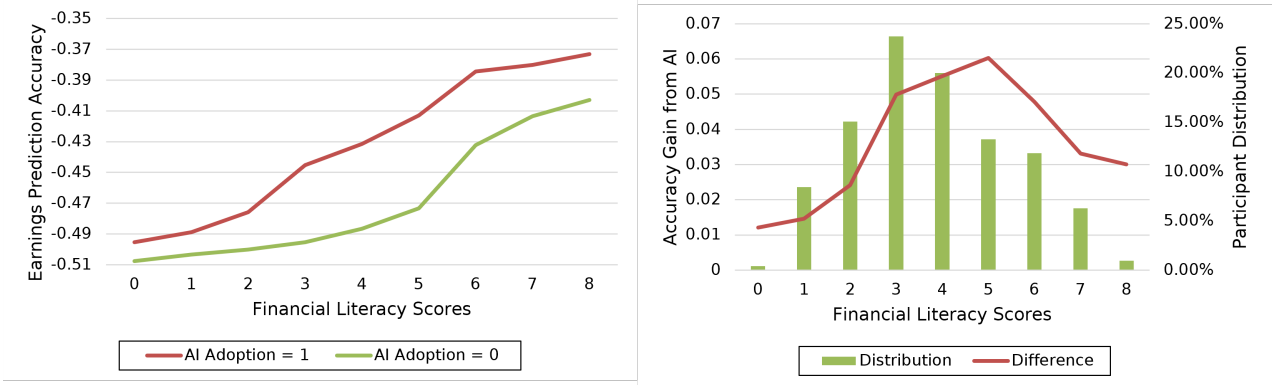
In this section, we examine whether the experimental data fits the model described in Section 4.1. For investors in the lower range of investor sophistication (i.e., investor knowledge 0-0.2 in Figure 4), incremental benefits of AI increase as their knowledge level increases. The opposite is true for investors in the higher range of investor sophistication (i.e., investor knowledge 0.2-0.4 in Figure4). This pattern creates an inverse U-shaped relationship between investor sophistication and incremental gains from AI.

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 largely ignore earnings call information in their decision-making. However, when AI simplifies complex disclosures, these investors begin incorporating such information into their decisions. Within this group, relatively more sophisticated individuals are better able to use AI-generated summaries to derive actionable insights, resulting in greater incremental gains for them.

In contrast, The negative relationship for more sophisticated investors is driven by their

baseline ability to process raw transcripts effectively. For these individuals, the additional value provided by AI diminishes as their existing capacity to interpret information is already high. Consequently, the marginal utility of AI-generated summaries decreases with increasing sophistication in this group.

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 predicted by the model. Figure 6 displays participants' average earnings prediction accuracy across different levels of sophistication. The left panel shows the average accuracy for participants treated with raw transcripts and for those treated with AI-generated summaries aligned to their sophistication level.[17] Among less sophisticated participants (scores below 4), average accuracy remains slightly below -0.5, which is equivalent to random guessing, when processing raw transcripts. Accuracy improves rapidly as scores approach 4. In contrast, when treated with aligned summaries, participants' average accuracy increases monotonically, consistent with the left panel of Figure 4.

The right panel of Figure 6 illustrates the net accuracy gains from AI treatment and the distribution of participants by sophistication score. As in Figure 4, we observe an inverse U-shape relation between informational gains and investor ability (scores). Consistent with our

[17]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.

model, among less sophisticated investors (i.e., Score ≤ 3), those with relatively higher scores benefit more 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. This distribution highlights the "unleveling" effect of AI, whereby its advantages disproportionately favor specific investor groups.

4.4 When Does the AI Increase Unleveling?

We explore the mechanisms driving the unleveling effect by examining how AI's information processing widens disparities between more and less sophisticated participants. Specifically, we identify the conditions under which AI contributes more to informational gains for sophisticated participants.

4.4.1 Methodology

Advanced vs. Simplified Summaries The disparity between advanced and simplified summaries stems from differences in the information they provide. To explore the informational differences between advanced and simplified summaries, we systematically analyze the content omitted from simplified summaries compared to advanced ones. Advanced summaries contain detailed financial insights, while simplified summaries are designed for readers with minimal financial literacy. Our goal is to identify the key types of information present in advanced summaries that are consistently missing from simplified summaries.

Using a standardized LLM prompt, we extract details included in advanced summaries but omitted from simplified summaries. For each earnings announcement, we input both the advanced and simplified summaries into an LLM. We instruct the model to identify any details mentioned in the advanced summary that are not mentioned in the simplified summary. If such details are missing, the model is instructed to respond with "n/a." By using the same standardized prompt for all pairs, we ensure consistency in how the model identifies omitted content. After obtaining the differences for all comparisons, we aggregate these results into batches of 20 excerpts each. For each batch, we prompt the model to identify recurring topics or types of information that frequently appear in the advanced summaries but are consistently missing in the simplified summaries. Throughout this process, the model is instructed to avoid firm-specific details. By iterating over multiple batches (10 sets of 20 excerpts each), we produce multiple lists of commonly missing topics. Finally, we consolidate the lists by asking the model to reconcile overlapping or synonymous concepts. This process yields a list of

items that are commonly missing in simplified summaries compared to advanced summaries, which includes EBITDA (earnings before interest, tax, depreciation, and amortization), cash flow projections, revenue mix or composition, dividend policies, share repurchase, recurring revenue items, and details of innovations.[18]

Regression Estimation To quantify the unleveling effect, we define *Wedge* as the call-level difference in informational gains between more and less sophisticated participants. Using the same definition of indicators as in Section 3.3.1, we define Wedge as follows.

$$\text{Wedge}_i = \underbrace{\left(\frac{\sum_j \text{AQ}_{ij} I_j^{s,as}}{\sum_j I_j^{s,as}} - \frac{\sum_j \text{AQ}_{ij} I_j^{s,r}}{\sum_j I_j^{s,r}}\right)}_{\text{Informational Gain of More Sophisticated People}} - \underbrace{\left(\frac{\sum_j \text{AQ}_{ij} I_j^{us,ss}}{\sum_j I_j^{us,ss}} - \frac{\sum_j \text{AQ}_{ij} I_j^{us,r}}{\sum_j I_j^{us,r}}\right)}_{\text{Informational Gain of Less Sophisticated People}} \quad (17)$$

AQ_{ij} is participant j's answer quality about earnings call i. We then estimate the following regression to study the determinants of Wedge.

$$\text{Wedge}_i = \gamma Z_i + \delta_t + \varepsilon_i \quad (18)$$

Z_i is a vector of earnings call characteristics informed by the content analysis. We include the absolute value of the difference between EBITDA and net income scaled by the total assets to capture the magnitude of incremental information conveyed by EBITDA in addition to the bottom-line earnings. We also include the absolute value of accounting accruals scaled by the total assets since accruals are a strong predictor of future cash flows (Ball and Nikolaev, 2022). Similarly, to model recurring and non-recurring components of earnings, we include the absolute amount of extraordinary items in the income statement scaled by total assets. To proxy for the complexity in revenue composition, we include the number of product segments. In addition, the amount of R&D expense scaled by total assets proxies for the innovation intensity. Lastly, we add variables related to financing activities that are commonly excluded from simplified summaries, such as the dividend payout ratio and the number of shares repurchased scaled by total shares outstanding. Year-quarter fixed effects control for macroeconomic trends. Continuous variables are winsorized at the 1% and 99% levels to mitigate the influence of outliers.

[18] To ensure the consistency of the results, we repeat this experiment 10 times and confirm that the list is not sensitive to repetitions.

Results Table 6 shows that *Wedge* increases when omitted information is critical for decision-making. Specifically, Wedge grows with larger differences between EBITDA and net income, higher accrual magnitudes, more complex revenue compositions (e.g., more product segments), higher R&D investment, greater share repurchases, and higher dividend payout ratios. 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 unleveling is more pronounced when the information omitted from simplified summaries plays a critical role in financial decision-making.

5 Additional Experiment and Robustness

In Section 5.1, we discuss the results from a second experiment involving participants interacting with AI chatbots. In Section 5.2, we design tests to mitigate concerns that our results are sensitive to particular aspects of our experimental design.

5.1 Integrating an Interactive AI Agent

Our second experiment is designed to provide participants with more flexibility in how they use generative AI to process earnings announcement news. Instead of pre-written summaries, we provide the treatment group with an AI assistant – a chatbot pre-loaded with the earnings call transcript. Hence, 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 their performance on investment tasks but also to track how they engage with AI chatbots. For example, we can observe the questions they ask the AI chatbot and study how the nature of their questions varies with investor sophistication.

Experiment Design In the second experiment, we have only one treatment group and one control group. The treatment group is provided with a chatbot that can answer questions about the company's earnings release (i.e., the earnings call transcript is contained in the model's system prompt). 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.[19]

We instruct participants to answer our survey questions by consulting the chatbot. 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 are also able to collect the questions asked by the participants and the answers generated by the model. We use 400 participants with 200 participants assigned to the control group and 200 participants assigned to the treatment group. We then estimate the following regression:

$$\begin{aligned}\text{Answer Quality}_{kj} = \beta_1 \text{Sophistication}_j + \beta_2 \text{AI Chatbot}_j \\ + \beta_3 \text{Sophistication}_j \times \text{AI Chatbot}_j + \zeta_k + \varepsilon_{kj}\end{aligned} \tag{19}$$

Answer Quality is defined at the earnings call level for earnings prediction and sentiment assessment and at the 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 about allowing human-AI interactions is that participants might directly ask for the answers to the questions (e.g., asking for the chatbot's prediction about future earnings or returns). We face two empirical challenges if such use cases are prevalent and participants directly follow the model's predictions. First, our estimates may be contaminated by the model's ability to predict earnings and returns. Note that the focus of the second experiment is to understand whether an AI model can help investors process information but not to examine whether the model can successfully predict future corporate outcomes. Second, as our sample contains earning releases from 2010 to 2022 (which falls into the training period of GPT-4o) and the model has access to rich textual information, the model's answers might suffer from look-ahead bias (Sarkar and Vafa, 2024).

To prevent direct reliance on chatbot-generated answers (e.g., asking for the chatbot's prediction about future earnings or returns), we implement some features to ensure the model

[19]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."

focuses on facilitating information processing rather than providing predictions. First, using a system prompt that is not visible to users, we explicitly instruct the chatbot to answer questions within the scope of the transcript provided to the model. Therefore, our model refuses to answer if a user asks about information not included in the transcript. Second, when a user asks for a direct answer to the survey questions, the model is also configured not to produce answers. We also not only prevent the model from directly providing the answers to the questions but also collect the extent to which the participants attempt to do so. Finally, in the survey instructions, we warn the participants not to ask questions identical to what we ask in the survey.[20]

What Questions Do Investors Ask? The content and quality of participants' questions reveal significant differences in how investors engage with the AI chatbot. By analyzing the content of the participants' questions, we find evidence suggesting that the participants' level of financial sophistication is positively associated with the quality of questions they ask to the AI model.

We received a total of 1,617 messages from the 200 participants of the treatment group, yielding 8 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[21]; (iv) questions that cannot be answered using the information presented in the transcript (these questions include when participants ask the model to make some economic judgments and when participants inquire about information that is not present 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.

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,

[20]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.

[21]We add a disclaimer that the participants should not ask the survey questions directly to the model and that the model will refuse to answer such questions. However, some participants still asked these questions directly to the model.

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, often stop after a single interaction and are more likely to directly ask survey questions despite instructions to avoid this. Additionally, less sophisticated participants are more likely to ask questions identical to the survey questions.

Overall, our results highlight a key disparity: more sophisticated participants exhibit better command of working with AI, asking higher-quality questions that yield more actionable insights. This enhanced interaction may contribute to the widening performance gap between the groups as more sophisticated participants can extract greater value from the chatbot by effectively leveraging its capabilities. Investment sophistication thus appears to be positively correlated with the ability to generate high-quality signals through AI, further reinforcing the knowledge-driven disparities in financial decision-making.

Performance Using the AI Chatbot Table 8 shows that sophisticated participants consistently outperform less sophisticated participants when using the AI chatbot. Specifically, the coefficient on Sophistication, β_1, is positive and sometimes marginally significant, consistent with more sophisticated people performing better than less sophisticated people with raw transcripts.

Less sophisticated participants, however, perform worse when using the chatbot. In Table 8, the coefficient on Chatbot (i.e., β_2), is significantly negative across all specifications, indicating that less sophisticated individuals struggle with the chatbot. This is likely because they lack the skills to ask relevant or effective questions. This result is potentially exacerbated by the absence of raw transcripts, leaving less sophisticated participants unable to engage meaningfully with the chatbot.

Perhaps most importantly, Table 8 provides evidence that more sophisticated participants derive substantial benefits from the chatbot. The coefficient on Sophistication $\times$ Chatbot, (i.e., β_3) is positive and statistically significant across all of the investment 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 strong foundation in financial knowledge.

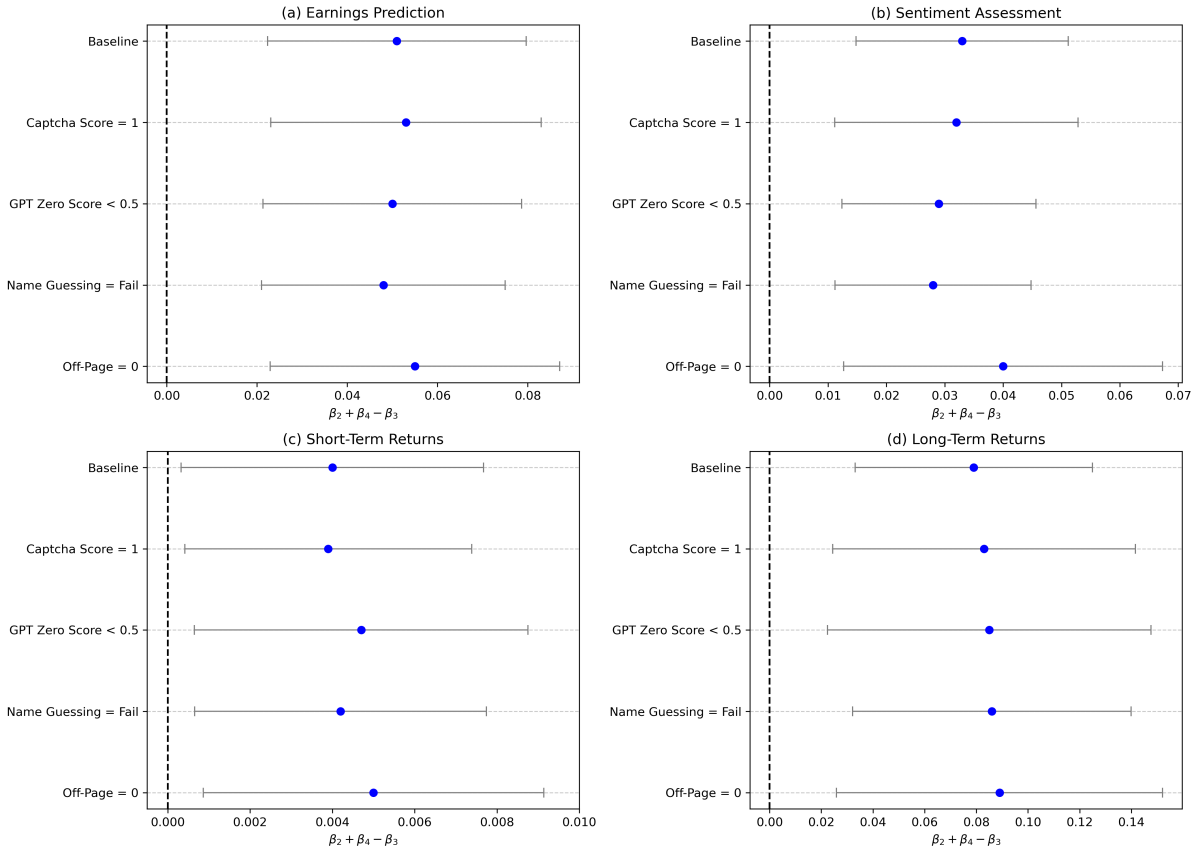
Taken together, Table 8 highlights that the benefits of an AI chatbot depend heavily on users' financial sophistication. When used effectively, the chatbot provides sharp, actionable insights that enhance decision-making. However, without the ability to generate high-quality

prompts, less sophisticated users may misinterpret or misuse the information, leading to poorer economic outcomes. This finding reinforces our previous result that aligning AI tools with users' capabilities to ensure equitable benefits.

5.2 Robustness Checks

To enhance the robustness of the findings of our main experiment, we perform a battery of robustness checks. For brevity, Figure 7 presents the values of $\beta_2 + \beta_4 - \beta_3$ and their 90% confidence intervals under these alternative approaches, using earnings prediction accuracy, sentiment assessment, short-term portfolio returns, and long-term portfolio returns as dependent variables.

Figure 7. **Robustness**



Notes: This figure displays the values of $\beta_2 + \beta_4 - \beta_3$ and their 90% confidence intervals under alternative specifications.

First, to mitigate the concern that some of our online participants are automatic survey bots rather than actual humans, we implement a Captcha tool at the beginning of the survey

and collect the scores. We repeat our analyses using participants whose Captcha score is one (highly likely to be humans). Second, to account for the possibility that participants used generative AI tools to complete their surveys, we employed GPT Zero, a tool designed to detect text likely generated by large language models. We investigate all short narrative answers submitted by the participants and exclude participants with a GPT Zero score higher than 0.5 for at least one answer. Third, although we make our best effort to anonymize the transcripts and their summaries, participants might have identified company names using context clues, such as product or service descriptions. To validate this, we ask participants if they can identify the company's name during the task. Less than 2% answered this question, and only 1% of them submitted the correct answer. We exclude participants who answered this question (either correctly or incorrectly) and repeat the analysis. Lastly, to address concerns about participants using external tools or resources, we analyze the total time spent with the experiment window closed. Having a closed window makes it more likely that the participants try to use information or tools not provided in our experiment. We repeat the analysis using only participants whose total off-task time is less than ten seconds. The results are robust under all of these alternative specifications.

6 Conclusion

Using two large-scale experiments with actual financial data from 200 publicly traded firms, we study the impact of generative AI technology for investors. Our findings highlight two central effects: (i) a significant improvement in information processing and investment efficiency following AI adoption and (ii) an "unleveling" effect, where AI widens the knowledge-driven gap between sophisticated and unsophisticated participants.

This study makes three key contributions to our understanding of AI's role in financial markets. First, we provide novel evidence that generative AI significantly enhances both financial comprehension and investment performance, making earnings information more accessible to a broader investor base. Second, we demonstrate that AI's effectiveness critically depends on the alignment between its outputs and user expertise, highlighting how personalization fundamentally shapes AI's impact on investment decisions. Third, we challenge prevailing narratives about AI's democratizing effects by showing that in financial markets, where success requires 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 assumed, particularly in contexts where strategic application of information is crucial.

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.

Kim, A. and V. V. Nikolaev (2023). Context-based interpretation of financial information. *Chicago Booth Research Paper* (23-08).

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.