# AI Democratization, Return Predictability, and Trading Inequality *

Anne Chang†

Xi Dong ‡

Xiumin Martin§

Changyun Zhou**

First Version: August 17, 2023; This Version: September 12, 2024

* We would like to thank Elizabeth Blankespoor, Danling Jiang, Vitaly Meursault, Steven Poser and participants at Monash University, Australian National University, and "AI Era in Finance" Symposium 2024 for comments and suggestions.

† Zicklin School of Business, Baruch College. E-mail: Yanru.Chang@baruch.cuny.edu

‡ Zicklin School of Business, Baruch College. E-mail: Xi.Dong@baruch.cuny.edu

§ Olin School of Business, Washington University in Saint Louis. E-mail: xmartin@wustl.edu

** Southwestern University of Finance and Economics. E-mail: zhoucy@swufe.edu.cn

# AI Democratization, Return Predictability, and Trading Inequality

First Version: August 17, 2023; This Version: September 12, 2024

**Abstract**

We conduct the first analysis on the impact of democratized AI (ChatGPT) on the trading activities of investors by leveraging a dataset of long textual information spanning 19 years of earnings calls. We have three key findings. First, *AI-sentiment* generated by ChatGPT strongly predicts returns for up to 12 months, while traditional human-dictionary-based sentiment yields little predictability. Second, before the arrival of ChatGPT, short sellers traded in alignment with *AI-sentiment* within two weeks following earnings calls, while retail traders did not. Following the widespread deployment of ChatGPT, there was a significant 65-fold increase in retail-trader alignment with *AI-sentiment*, whereas the alignment of short-sellers with *AI-sentiment* may have weakened. Third, stock experiencing increased retail traders' alignment with *AI-sentiment* also witnessed a significant decrease in bid-ask spreads. An exogenous variation in AI availability due to ChatGPT outages led to notable reduction in retail-AI alignment and reversal of narrowed bid-ask spreads, further supporting the causal role of AI. Overall, the study suggests that democratizing AI has the potential to level the playing field and bridge the information gap between privileged and ordinary investors.

JEL: G10, G11, G14

Keywords: Return Predictability, Artificial Intelligence AI, ChatGPT, AI democratization, Open AI, Outage, Machine Learning ML, Earnings Conference Calls, Retail Investors, Short Sellers, Households, Anomalies, Financial Market Fairness, Financial Literacy, Information Asymmetry

*"I think AI very naturally, …, will raise the floor in a big way. I hope AI can reduce inequality and more than that, I hope it can dramatically lift up the floor in the world."*—Sam Altman, OpenAI CEO

*"Market forces won't naturally produce AI products and services that help the poorest."*—Bill Gates, Microsoft Cofounder

*"The rich get richer…"*—Eric Schmidt, Former CEO of Google, when commenting on the role of AI.

## Introduction

The Financial market regulator, Security Exchange Committee (SEC), has long been focused on protecting ordinary investors. It has instituted extensive disclosure requirements for public companies to ensure important information is made available to all investors. Chief among these is Regulation Fair Disclosure (Reg FD), which mandates companies to make crucial information like earnings calls accessible to the general investing public in real-time. The ideal of the ongoing initiative to democratize access to all sources of information is to level the playing field. Yet, critics question whether such openness achieves its intended purpose. Indeed, managers can reiterate old information (Frankel, Jennings, and Lee), or, worse, public disclosures have the potential to increase information advantage of sophisticated investors over retail investors, as predicted by the mosaic theory (Levine and Cheynel 2016). Furthermore, ordinary investors can barely compete with sophisticated ones in utilizing such information, given their inferior ability in timely processing and trading on long and sophisticated financial textual information. The democratization of Artificial Intelligence (AI), with ChatGPT at its center, introduces fresh tension to this debate. On the one hand, supporters of democratized AI boast it as revolutionary technology for parsing through general narratives, implying a potential to narrow information gap between retail and institutional investors. The evidence from Kim, Muhn and Nokolaev (2024) provides some support to this argument in that they find generative AI is more effective in identifying firms' (non-)financial information. However, the actual prowess of these democratized AI in navigating financial text information like earnings calls remains

unknown given the limitations of large language models (LLMs) in processing long text,[1] and that LLMs are not trained to forecast financial market returns. On the other hand, AI literacy may introduce a new layer of disparity within the investing community: investors proficient in leveraging AI technologies stand to significantly enhance their financial standing, while those lacking such expertise risk falling further behind, casting doubt on whether democratizing AI can level the playing field for all investors. In this study, we first aim to examine whether and how ChatGPT predicts returns from analyzing earning calls, which informs us whether such democratized AI generates investment insights that are actionable especially for ordinary investors. Armed by this evidence, we then examine how AI and its democratization shapes retail investors' trading behavior and their welfare, relative to privileged investors such as short sellers.

To examine these research questions, we concentrate on the most relevant form of democratized AI where the usage of AI, rather than just its *source code*, has been made easily accessible to the general public.[2] Specifically, we employ ChatGPT, arguably the most advanced and accessible AI tool available, to extract sentiment information from earnings calls, termed as *AI-sentiment*. Earnings calls are chosen as the focal point due to Reg FD's mandate for their public accessibility. Furthermore, they are the longest form of textual information fed to ChatGPT in contemporaneous studies for return predictability analysis. The capability differences in timely processing and/or trading on long and jargon-laden textual information among AI, privileged, and ordinary investors are likely to be most pronounced. The effectiveness of LLMs in processing long financial text information for investment insights is also untested (see Footnote 1).

[1] Modern LLMs are predominantly implemented with Transformers (Vaswani et al. 2017), which apply self-attention to encode input contexts. Its information processing ability may deteriorate as text length increases(e.g., Dai et al. 2019). Intuitively, the possible connections between words exponentially grow as the number of words increase. To mitigate this issue, Open AI claims to adopt mechanisms similar to sparse attention, which selectively focus on parts of the input sequence rather than the entire sequence to reduce computational complexity (Child et al. 2019). However, it might only consider immediate neighbors and specific distant tokens based on predefined patterns, such as skipping a fixed number of tokens. For example, Liu et al. (2023) find that language model performance is highest when relevant information occurs at the very beginning or end of its input context. This suggests that although extended-context models can process longer input contexts, they may not effectively reason over the information within its context window. In the same vein, Sun et al. (2021) found that longer context improves prediction of only a few tokens and Krishna et al. (2022) found that long-context neural generation even degenerates in modestly-sized Transformer language models because models fail to properly condition on long context.

[2] Seger, Ovadya, Siddarth, Garfinkel, and Dafoe (2023) identifies democratizing AI usage as the foremost kind of AI democratization. By focusing on open usage, our studies differ from general LLM/AI studies that use open-, or even closed-, source models that ordinary people are not capable of using.

Therefore, coupled with their spontaneous and intricate nature, earnings calls provide a unique testing ground for understanding the potential for AI to democratize information processing and level the playing field for all investors.

We first examine the capability of *AI-sentiment* to forecast returns across multiple horizons, offering a nuanced perspective on its predictive power. We then delve into the trading behaviors of two distinct investor classes—short sellers, which empirical evidence suggests being the most informed sophisticated traders (Chen, Da, and Huang 2018), and retail investors, a.k.a. ordinary investors. We investigate how their trades align with *AI-sentiment* before and after the wide deployment of ChatGPT, which offers insights into the shifting dynamics in trading on "equally accessible" information following AI democratization.

Our study has several distinctive features. First, we provide, to our knowledge, the first analysis on how democratized AI shapes human trading. We do so at both low (daily to yearly) and high frequency (intraday) levels. Second, we are the first to use ChatGPT to comprehensively analyze the long textual information of earnings calls for return predictability. Our analysis preserves the full contextual information of earnings calls, given the backdrop that understanding context is the central advantage distinguishing AI from the technologies before the age of AI. We enable such analysis by employing the gpt-3.5-turbo-16k model that allows analyzing each transcript in one single feed, costing around $18,300 to analyze the entire transcripts data.[3] In contrast, contemporaneous ChatGPT studies predominantly use the gpt-3.5-turbo model that only manages 4,096 tokens, while over 90% of earnings call transcripts exceed that token limit.[5] Third, utilizing

[3] The 16k model supports up to 16,385 tokens. GPT4 can also handle long text but it is far more costly.

[5] Therefore, exiting studies either analyze other, short forms of text, or split long texts into multiple parts and feed them separately to the model. Two contemporaneous studies focus on using ChatGPT for return predictability based on different forms of financial information. Lira-Lopez and Tang (2023) documents daily return predictability using ChatGPT on Raven Pack headlines. Headlines on average are 76 words long. Chen, Tang, Zhou, and Zhu (2023) analyze headlines and frontpage news from the Wall Street Journal (WSJ) and forecast time-series aggregate monthly market and macroeconomy conditions. The average length of articles in WSJ typically spans 400 to 1,200 words, and has exhibited a downward trend since 2000 (Madrigal, 2013). In contrast, earnings calls average 6,896 words during our study period, which are an order or orders of magnitude longer than news headlines and frontpage news. Other GPT studies that do not focus on return predictability include Kim, Muhn, and Nikolaev (2023a), Kim, Muhn, and Nikolaev (2023b), and Jha, Qian, Weber, and Yang (2024). In particular, Jha et al. (2024) focuses on extracting firm real investments from chunked call transcripts. Given the anomaly literature has shown that investments can predict returns, Jha et al. show their investment score provides superior return predictability information beyond the traditional investment return predictor from Fama and French (2015).

call audio recordings, we precisely pinpoint the day and intraday timing of earnings calls to calculate returns. Prior studies predominantly use the coarser two or three day returns around an earnings call, and identify call days using an imprecise *calendar* date of earnings. This accurate timing allows us to precisely identify the returns after the call day, and attribute intraday returns to those related to earnings announcements/press releases (before call), those related to calls (during or after call). All these improvements allow for a clean test of return predictability and its magnitude, which helps evaluate how much of the investment insights generated by ChatGPT are actionable for retail investors vs. those privileged investors capable of high-frequency trading around the clock. Fourth, we provide two methods to guard against look-ahead bias. The most effective way to rule out look-ahead bias is to conduct tests beyond the model's training window (Sarkar and Vafa 2024). Therefore, we examine a "post-final knowledge" period between October 2021 and December 2023 as the *API version* of gpt-3.5-turbo-16k is limited to the knowledge data up to September 2021. This model differs from GPT4 or GPT4o, which updates their knowledge data overtime making a truly out-of-sample test infeasible.[6] Additionally, we exploit the knowledge cutoff feature of gpt-3.5-turbo-16k model, which allows us to restrict GPT to only use knowledge that was available up to the day before each earnings call is held.

In our first set of tests, we examine the return predictability based on sentiment measures extracted from a comprehensive sample of earnings call transcripts, totaling 19 years (2005-2023) and 280,554 transcripts. To gauge the effectiveness of AI, we employ a benchmark sentiment measure established by Loughran and McDonald (LM 2011). This human dictionary-based sentiment, referred to as *HD-sentiment*, is generated by counting the positive and negative words. It proxies for basic actionable investment insights that human investors may gather from call transcripts, as their fundamental comprehension of positive and negative words generally aligns with the LM human dictionary. In our regressions, we also control for the return predictability of earnings surprises (*SUE*) and various other anomaly variables that can predict returns.

[6] See https://platform.openai.com/docs/models/gpt-4-and-gpt-4-turbo

First, we use an event-time regression approach to examine ChatGPT's predictability on the daily returns following the trading day when the conference call is held (referred to as the event day). Given conference calls usually occur outside regular trading hours, limiting immediate retail investor trading, we reserve the event day later for a higher frequency investigation. We find the return predictability of *HD-sentiment* is weak both in magnitude and horizon for the full sample, with effect notably disappearing after the first three days post-earnings calls. In contrast, *AI-sentiment* displays strong return predictability after controlling for *HD-sentiment*. The magnitude of the return predictability gradually increases as the return horizon lengthens, and plateaus at the half-year horizon, but remains significant even through to an entire year. This contrasting pattern between *AI-sentiment* and *HD-sentiment* holds for the post-final knowledge subsample (October 2021 to December 2023). Remarkably, in this short subsample, *AI-sentiment* can still significantly predict returns over a quarter in the future, despite the substantial loss of observations. In sharp contrast, *HD-sentiment* stops displaying return predictability at any horizon. Overall, *AI-sentiment* displays long-run return predictability many months after earnings calls, suggesting substantial actionable investment insights that democratized AI could offer to ordinary investors, even if these investors cannot trade fast.

Second, we strengthen the point that the AI insights are actionable by using a calendar time portfolio approach. At the end of every month, we sort stocks by *AI-sentiment, HD-sentiment*, and *SUE* based on the most recent available earnings calls and announcements. Since earnings are only reported quarterly, investors have, on average, 1.5 months to react to the *AI-sentiment* signal. We consider various factor models to compute alphas. Our results suggest that long-short portfolios based on *AI-sentiment* yield an alpha up to 1% per month equal-weighted and 60 bps (capped-) value-weighted. In contrast, long-short portfolios sorted on *HD-sentiment* generate no alpha, while those sorted on *SUE* generate strong alpha, consistent with the post-earnings-announcement drift (PEAD) phenomenon. In a double-sort test, after controlling for the return predictability of *SUE* and *HD-sentiment, AI-sentiment* still generates strong alphas, supporting AI's capability to unearth information not encapsulated by earnings surprises and other

anomaly variables. Finally, the long-short *AI-sentiment* portfolio yields a 9-month cumulative alpha over 4%.

Third, we then zoom in to the higher frequency returns during the event day of the earnings call. Earnings calls are a unique information set where the playing ground for ordinary investors and privileged traders are particularly uneven. Indeed, around 65% of earnings calls are held outside regular trading hours. While sophisticated investors often have the capability to trade during these hours at high frequency, many retail investors may not have equal market access for trading on time sensitive information like earnings calls.[7] We find that *AI-sentiment* and *HD-sentiment* are both significantly related to the high-frequency return over the interval before the call, exhibiting magnitudes two to three times larger than those observed during and after the call—an insight unattainable without the precise timing we obtain. The relatively large magnitude of the relation to before-call returns is unrelated to return predictability but simply suggests the call text indeed echoes some earlier information already disclosed through press releases or other news as earnings surprises are controlled for. Notwithstanding, the regression coefficients are still economically large for returns during and after the call. This suggests that both *HD-* and *AI-sentiment* measures may provide investment insights for the privileged traders who possess high-frequency trading capabilities. Our findings of large coefficients both before and during the call suggest that prior studies examining returns around earnings calls cannot legitimately examine return predictability, as they suffer from substantial look-ahead bias. Indeed, only the intraday interval after the call is the cleanest interval for studying intraday return predictability. The magnitude of the regression coefficient over this interval is considerable. However, *AI-sentiment*'s ability to forecast the long-term returns, excluding any return predictability during the event day is three times larger, suggesting that AI not only provides investment guidance over the high-frequency intraday interval, where privileged traders may prevail, but also delivers substantially more

[7] Retail brokers only provide National Best and Bid Offer (NBBO) prices during regular trading hours. The vast majority of them provide substantially increased bid-ask spreads that are not NBBO or do not allow trading during extended trading hours.

substantive actionable investment insights over the long term, which is more relevant for ordinary investors who are likely to trade more slowly.

Fourth, we explore suggestive evidence on the difference that may make *AI-sentiment* deliver greater return predictability than *HD-sentiment*. We find that *AI-sentiment*'s response to high and low *HD-sentiment* is muted. Therefore, when corporate executives use many positive or negative words in earnings calls, such extreme language does not convey extreme news, and AI sees through that. Combined with the evidence *AI-sentiment* predicts returns after controlling for *HD-sentiment*, our results suggest that ChatGPT employs a more nuanced analysis, not overly influenced by the frequency of positive or negative words that are subject to the manipulation of corporate executives (Cao, Jiang, Yang, and Zhang 2023), potentially indicating its ability to discern underlying sentiments more effectively.

In our second set of results, we examine how short selling and retail trading are aligned with sentiment measures. We first examine the dozen years before the wide deployment of ChatGPT on January 1, 2023, a phase we term as the pre-democratization period.[8] Our findings reveal a pronounced tendency among short sellers to align their trading with *AI-sentiment* immediately on the earnings call day and for the subsequent ten days, even after controlling for earnings surprises and various other anomaly variables. Retail investors show no significant reaction to *AI-sentiment*. Given the limited accessibility of ChatGPT and other AI models to the general public in the pre-democratization period, the results suggest that short sellers are informed traders that may utilize alternative information sources or even proprietary AI/ML models for earnings call analysis well before AI democratization. Their trading decisions, which correlate with the insights akin to those of ChatGPT, reflect the anticipated behavior of short sellers as archetypal arbitrageurs with privileged technology and information access. In contrast, retail investors are unable to learn such insights, suggesting that before the democratization of AI, the democratization of information implemented years ago may have not effectively leveled the playing field for information access.

[8] ChatGPT, released on November 30, 2022, achieved rapid growth, reaching 100 million monthly active users by January 2023. This surge signals a heightened public engagement with ChatGPT and its transition to widespread use.

We then analyze shifts in short selling and retail trading surrounding ChatGPT's widespread deployment around January 1, 2023. Post the deployment, retail investors' alignment with *AI-sentiment* in the ten days following earnings calls increases 65-fold compared to their long-term average alignment with *AI-sentiment* during the pre-democratization years, consistent with a paradigm shift. In contrast, short sellers display no significant change or can even weaken their alignment with *AI-sentiment*, suggesting that they no longer enjoy a clear information advantage by trading on these AI insights. We then employ a difference-in-differences (DiD) framework in which we address many firm and firm-trader related endogeneity concerns on trading by comparing the difference in retail and short trading responses to AI in the same firm, controlling for firm-trader or time-trader fixed effects. We show that retail-AI alignment increases significantly relative to short selling after AI democratization, suggesting a robust shift in retail trading behavior toward greater alignment with *AI-sentiment* post-democratization. To further reduce the endogeneity concerns of the relation between AI democratization and trading, we use ChatGPT outages as exogenous shocks to identify changes in the availability of ChatGPT for retail investors. If AI democratization aids retail investors, then unexpected outages would reduce the post-democratization increase in retail investors' alignment with AI. We find that the increased retail-AI alignment indeed significantly decreases with outages occurring during the trading day of an earnings call.

Finally, if AI democratization enhances retail investors' ability to process information, it should reduce information asymmetry and improve market liquidity. Therefore, we examine the relation between AI democratization and information asymmetry. Using a propensity score matching-difference in differences (PSM-DiD) framework, we find a significant decrease in bid-ask spreads for stocks with more increased retail-AI alignment after AI democratization. This effect is mitigated during ChatGPT outages.

Overall, our results suggest that long and complex information (earnings calls) contain long-term investment insights actionable for ordinary investors. Democratized AI (ChatGPT) can decode these insights. Informed traders had already harnessed these insights before AI democratization, whereas retail investors had not, suggesting that the information asymmetry between the few privileged investors (akin to insiders) and ordinary investors may have exacerbated rather than mitigated despite the democratization of

information access for many years. AI democratization may allow retail investors to truly benefit from the democratization of information access, a purpose long advocated by regulators. Therefore, democratization of information demands democratization of the tools to analyze the information. Our results reveal the financial market consequences that illuminate the debate on whether, and how, AI should be democratized and made accessible to all.[9] Our findings underscore the necessity of "open" the *usage, rather than just the source codes,* of AI to enable a more inclusive pursuit of financial wisdom. These findings are critical for understanding the intricate interplay among technology, regulation, and market dynamics.

## Literature and Contribution

We make four contributions. First, our study is informative to the financial market regulators who are particularly concerned about the wellbeing of disadvantaged groups of market participants. Our evidence suggests that, while years of regulatory push on democratizing access to all kinds of information, such as to earnings calls, is necessary, it is not sufficient to bridge this gap. By lowering information processing costs, AI democratization shows greater promise in achieving this goal. Furthermore, our findings potentially alleviate concern about the technological advancement of OpenAI—often deemed as the start of a new industrial revolution—might benefit sophisticated investors to the detriment of ordinary investors.

Second, we shed light on whether and how the state-of-the-art democratized AI, ChatGPT, can indeed generate valuable investment insights. We are the first to connect ChatGPT's information processing results with human trading, and by employing AI to comprehensively analyze the return predictability of a central, and unique piece of information—earnings calls. We show that democratized AI generates return predictability over extended horizons that significantly transcends the narrow earnings call window that the

[9] One central part of the debate is that whether AI should be open to all, or should it be cloistered, reserved for a select few? OpenAI, initially starting as a non-profit with a mission to advance digital intelligence in a way that could benefit humanity as a whole, has transitioned towards a more closed, profit-oriented model. This transition has led to criticisms, such as those from Elon Musk, who sued OpenAI for straying from its original, altruistic mission towards a profit-driven approach. These contrast underscore different attitudes towards opening AI. Another central part of the debate is the difference between making AI accessible to and easily usable by the general public and open-sourcing AGI. As the then chief scientist of OpenAI Ilya Sutskever argued with Elon Musk: "As we get closer to building AI, it will make sense to start being less open. The Open in OpenAI means that everyone should benefit from the fruits of AI after its built, but it's totally OK to not share the science." In contrast, other models like XAi's Grok and Meta's Llama 2 are accessible to everyone.

earnings call literature focuses on. This provides far more convincing evidence of actionable insights for the slow-moving, ordinary investors. Existing studies either overlook these investors or imply that they are unlikely beneficiaries of AI-driven strategies. The substantial length (i.e. richer context) and acute time sensitivity of earnings calls create an ideal laboratory for exploring the dynamics between AI and human participants with unequal information processing and trading capabilities.

Third, our study contributes to the ongoing discussion on the impact of recent technological advancements in data collection on financial markets. Zhu (2019) demonstrates that real-time data, such as satellite images and foot track increase price informativeness, particularly among firms where sophisticated investors have strong incentive to collect such information. These findings suggest that big data potentially benefits sophisticated investors to the detriment of retail investors. In contrast, the evidence from our study indicates that, unlike big data, ChatGPT benefits retail investors, likely due to its democratizing effects.

Fourth, our study relates to the extensive literature on conference calls. Earlier research on closed calls demonstrates their informativeness to market participants (Frankel et al. 2009; Mayew and Venkatachalam 2012). Subsequent studies indicate that Reg FD, by mandating public access, reduces information asymmetry (Brown, Hillegeist, and Lo 2004). Our findings extend this literature by suggesting that following the democratization of information access initiated by Reg-FD, sophisticated investors' information advantage may increase, rather than reduce, due to their abundant resources such as leveraging on AI/ML technology to process information. Moreover, extant research shows that retail investors utilize information from conference calls (Huang, Martin, Tang and Xie 2024). Our study further extends this research highlighting the value of AI in enhancing retail investors' understanding of conference calls. We also show that the investment insights in earnings calls are only decodable by AI or a few sophisticated investors before AI democratization and why studies based on human-dictionary sentiment analysis would conclude nonexistence of valuable information in calls.

# 1. Sample Construction and Descriptive Statistics

## 1.1. Earnings Calls, Their Precise Timing, and Relevant Returns

We obtain 280,554 transcripts of conference calls spanning 2005-2023 from S&P Capital IQ, extracting dates and firm identifiers to link each firm-quarter observation to the databases of CRSP, Compustat, and Trade and Quote (TAQ). Our sample is confined to common stocks on the NYSE, AMEX, and Nasdaq.

Data vendors including COMPUSTAT, I/B/E/S, and Seeking Alpha provide the *calendar* day of earnings without specifying whether the earning is reported after hours. Their reported dates can also differ (DellaVigna and Pollet 2009). Although I/B/E/S includes the timing of earnings releases, it does so with very limited scope and questionable reliability (Berkman and Truong 2008). Prior research either predominantly identifies earnings call event days using the imprecise calendar date of earnings, or resorts to analyzing a broader window—typically a two-day or three-day return period surrounding the earnings announcement date to mitigate the imprecision of timing (e.g., Hirshleifer, Lim and Teoh 2009). When matching returns with an earnings call that occurs after hours (i.e., after 4 p.m.) on calendar day *d*, simply matching this call with the day *d* return would be problematic, as the trading day timing convention for returns in CRSP starts from 4 p.m. of the previous calendar day *d*-1 to 4 p.m. of the current day *d*. Contemporaneous studies using ChatGPT to analyze conference calls also use this timing convention. Once we precisely identify the timing of calls, as we will show in Section 3, we find that three-fourths of the intraday return associated with the sentiment information extracted by ChatGPT from each call is accrued before or during the call. This makes it difficult to make any clean statement on return predictability over the intraday or the two- to three-day windows around the call, provided that one does not have precise timing of the call.

To ensure the precise timing of earnings calls, we retrieve audio recordings of these calls from S&P Capital IQ. Since batch downloading is unavailable for both transcripts and audios, we develop a Python script to automate the retrieval of URLs for both transcripts and audios, simplifying the download process. We then identify the timing of calls from the audios. While S&P Capital IQ provides the start time of each call, it does not include call end time. We determine the duration of earnings calls by analyzing the length of audio recordings, then add each call's duration to the start time to derive the end time. Considering that our data comprises U.S.-based firms, we standardize the start and end times to Eastern Standard Time (EST).

For an earnings call that takes place after hours, we use the next *trading* day instead of the current calendar day as the event day *d*.

Our primary analysis centers on 173,229 earnings call transcripts from 2010 to 2023, hereafter referred to as the "full sample," as prior years before 2010 have insufficient data for reliable results. We focus on earnings calls for their regularity and detailed disclosure of firm performance, including management's discussion on past and future outcomes. Other types of conference calls like special calls are excluded due to their irregular nature and narrower focus. During our sample period, the average conference call transcript comprises 6,896 words, approximately 9,195 tokens. We also examine a "post-final knowledge sample" between October 2021 and December 2023 because the knowledge base of the API version of gpt-3.5-turbo-16k is limited to data available up to September 2021.

### 1.2. AI Sentiment

We utilize the gpt-3.5-turbo-16k API model that supports up to 16,385 tokens, far surpassing the 4,096-token limit of gpt-3.5-turbo, to analyze each transcript in one single feed and preserve the contextual information of the entire call. We process a total of 280,554 call transcripts from 2005 to 2023 through the model, costing around $18,300.

Our analysis includes two samples: a "full sample" from 2010 to 2023, and a "post-final knowledge sample" from October 2021 to December 2023. The rationale for establishing the post-final knowledge sample stems from the model's training data cut-off in September 2021, implying it does not possess any information beyond September 2021. This separation allows us to assess GPT's predictive capability of exclusive information that was truly unknown to it at the time of analysis. To further guard against look-ahead bias, we design the prompt to limit the model's knowledge to data available up to a certain date, thereby avoiding the incorporation of subsequently available information. The prompt design rules out look-ahead bias while allowing information to be updated precisely before each call.

To enhance the model's reasoning and problem-solving abilities, we implement a chain-of-thought prompt as follows. First, the model is strategically structured around distinct roles such as "system," "user," and "assistant" to facilitate the API model's comprehension of the definitions and institutional context

surrounding earnings call transcripts. Additionally, the model is configured to simulate the perspective of a stock market trader, enhancing the relevance and financial acumen of its responses. Additionally, we set the model temperature to zero to enhance accuracy and diminish the occurrence of model hallucinations. Subsequently, we instruct the API model to evaluate the sentiment embedded in each earnings call transcript and generate a corresponding sentiment score on a scale of -10 to 10. This approach ensures that the model's analysis is both contextually informed and aligned with financial market insights. We use the sentiment score given by the model as our measure for *AI-sentiment*. *AI-sentiment* is standardized for regression analysis.

Prompt:

"role": "system", "content": *Forget all your previous instructions. You are a stock market trader with experience in both fundamental analysis and technical analysis. Knowledge cutoff: {(date)}.*

"role": "user", "content": *Can you define the concept of conference call transcript?*

"role": "assistant", "content": *A conference call transcript is a written document that accurately records the spoken dialogue from a multi-party telephone meeting. A conference call usually starts with managers' presentation. The starting point of presentation usually starts with one word 'Presentation' in one line with no other words in that line. The presentation is then followed by question-and-answer session where analysts and investors ask questions. The question-and-answer session usually starts with 'Question and Answer' in one line with no other words in that line.*

"role": "user", "content": *I will give you a text of conference call transcript. Describe the sentiment of the text on a scale of -10 to 10. Here, -10 means most negative and 10 means most positive and 0 means neutral.*

"role": "assistant", "content": *Sure. Please give me the text of conference call transcript you want me to analyze.*

"role": "user", "content": *{transcript text}.*

### 1.3. Short Selling

To develop a direct measure of short selling activity, we utilize short sale volume data from Financial Industry Regulatory Authority, Inc. (FINRA), a self-regulatory organization that provides security-level

aggregate short-sale volume data on a daily basis.[10] We construct the short volume measure, denoted as $SV_{i,d}$, which is the aggregate number of shares of stock *i* reported in the short sell trading during regular trading hours on day *d* (scaled by share outstanding). The validity of this measure and FINRA short volume data are supported by prior research (Blocher, Dong, Ringgenberg, Savor 2023; Wang, Yan, Zheng 2019). We construct a detrended abnormal short-seller holding measure. This detrended holding measure can be backed out from the short volume measure. Specifically, let $SH_d$ denote the short-seller holding (scaled by share outstanding) of a stock on day *d*. The abnormal short-seller holding measure, defined as the holding detrended by its one-year (252 trading days) moving average can be expressed as follows:

$$ShortSelling_d = SH_d - \frac{(SH_{d-1} + SH_{d-2} + \cdots + SH_{d-252})}{252}$$

$$= SH_d - SH_{d-1} + \left(\frac{251}{252}SH_{d-1} - \frac{251}{252}SH_{d-2}\right) + \left(\frac{250}{252}SH_{d-2} - \frac{250}{252}SH_{d-3}\right) + \cdots + \left(\frac{1}{252}SH_{d-251} - \frac{1}{252}SH_{d-252}\right)$$

$$= SV_d + \frac{251}{252}SV_{d-1} + \frac{250}{252}SV_{d-2} + \cdots + \frac{1}{252}SV_{d-251} \quad (1)$$

where the last row of the formula provides a way to use short volume to compute short-seller holding. We winsorize stock-level holding data by year-quarter at the 1% and 99% percentiles to mitigate the impact of extreme values. In all the regressions, we multiply *ShortSelling* by 100000 to ease interpretation.

### 1.4. Retail Trading

Retail trading data is obtained from the TAQ database. As outlined by Boehmer, Jones, Zhang, and Zhang (2021), the Regulation National Market System (Reg NMS) mandates that a broker/dealer must provide a minimal price improvement over the National Best Bid or Offer (NBBO) for retail orders. We utilize this price improvement criterion to distinguish marketable orders placed by retail investors from those by institutional investors. Specifically, a transaction is classified as a retail purchase if its sub-penny price lies between 60 and 100 basis points, and as a retail sale if the sub-penny price is between 0 and 40 basis points.

[10] See https://www.finra.org/finra-data/browse-catalog/short-sale-volume-data and https://www.finra.org/finra-data/browse-catalog/short-sale-volume-data/monthly-short-sale-volume-files.

To calculate net buying activity by retail investors, denoted as $NB_{i,d}$, we subtract the volume of sales from the volume of purchases of stock $i$ on day $d$ (scaled by share outstanding). Like the short selling measure, we also employ a detrended, abnormal retail holding measure to proxy for the level of retail demand rather than the changes of demand, in line with the noise trader models that relate the level of demand to noise trader sentiment (e.g., De Long, Shleifer, Summers, and Waldmann 1990; Shleifer and Vishny 1997). Specifically, let $RH_d$ denote the retail holding (scaled by share outstanding) of a stock in time $d$, and the abnormal retail holding measure, defined as the holding detrended by its one-year (252 trading days) moving average can be expressed as follows:

$$RetailTrading_d = RH_d - \frac{(RH_{d-1} + RH_{d-2} + \cdots + RH_{d-252})}{252}$$

$$= RH_d - RH_{d-1} + \left(\frac{251}{252}RH_{d-1} - \frac{251}{252}RH_{d-2}\right) + \left(\frac{250}{252}RH_{d-2} - \frac{250}{252}RH_{d-3}\right) + \cdots + \left(\frac{1}{252}RH_{d-251} - \frac{1}{252}RH_{d-252}\right)$$

$$= NB_d + \frac{251}{252}NB_{d-1} + \frac{250}{252}NB_{d-2} + \cdots + \frac{1}{252}NB_{d-251} \quad (2)$$

where the last row of the formula provides a way to use retail net buy to compute retail holding. We winsorize stock-level holding data by year-quarter at the 1% and 99% percentiles to mitigate the impact of extreme values. In all the regressions, we multiply *RetailTrading* by 100000 to ease interpretation.

## 1.5. Bid-Ask Spread

We measure bid-ask spread using the following formula:

$$BidAskSpread_{i,d} = \frac{Ask_{i,d} - Bid_{i,d}}{(Ask_{i,d} + Bid_{i,d})/2} \quad (3)$$

which is the difference between the ask price and the bid price, scaled by the mid-point of the bid and ask prices for stock $i$ on day $d$. Bid and ask price data are obtained from CRSP. We winsorize stock-level bid-ask spread by year-quarter at the 1% and 99% percentiles to mitigate the impact of extreme values. In all the regressions, we multiply *BidAskSpread* by 100000 to ease interpretation.

## 1.6. Other Measures

To evaluate the effectiveness of AI versus traditional human-curated dictionaries in analyzing sentiment from earnings call transcripts, we employ the Loughran-McDonald (LM) dictionary for sentiment analysis. This dictionary categorizes words as either negative or positive. Accounting for word negation, we count the occurrences of negative (*NEG*) and positive (*POS*) words within each transcript (for stock $i$ and day $d$), calculating *HD-sentiment* using the formula:

$$HD\text{-}sentiment_{i,d} = \frac{POS_{i,d} - NEG_{i,d}}{POS_{i,d} + NEG_{i,d}} \quad (4)$$

Standardized unexpected earnings (*SUE*) are derived based on a random walk model, specifically as the year-over-year difference in earnings per share (*EPS*), adjusted for the quarter-end price. *AI-sentiment*, *HD-sentiment* and *SUE* are standardized for regression analysis.

Regarding other control variables: *Beta* represents the market beta relative to the CRSP equal-weighted return index, estimated over the past three years before the end of month t-1. *BM* is the ratio of book value of equity to market value of equity. *MVE* denotes the total market value of common equity. Momentum (*Mom12m*) is defined as the cumulative return from month t-12 to t-2.

### 1.7. Summary Statistics

Table 1 presents summary statistics for the variables used in our full sample (Panel A) and post-final knowledge sample (Panel B). *AI-sentiment* in our full sample is generally positive, with a mean (median) of 5.62 (6.00) within the range of -10 to 10. Its standard deviation is 1.83. *HD-sentiment* is relatively neutral with a mean (median) of 0.25 (0.26) within the range of -1 to 1. This contrast implies that the *AI-sentiment* may excel at pinpointing positive sentiment within earnings call transcripts, while the *HD-sentiment* may be more conservative, reflecting subtler shifts in sentiment. Additionally, the mean (median) *ShortSelling* is 0.22 (0.11), while the mean (median) value of *RetailTrading* is 0.0002 (-0.0004). The significantly higher holding volume of short sellers underscores their more pronounced positions compared to that of retail investors.

## 2. Return Predictability of AI

This section focuses on the return predictability of sentiment measures extracted by AI from earnings calls. In Section 2.1, we apply an event-time regression approach to examine the return predictability of *AI-*

*sentiment*. In Section 2.2, we construct calendar time portfolio sorts to evaluate the predictive ability of *AI-sentiment* on both monthly return and cumulative monthly return.

## 2.1. Event-Time Study

We first examine the predictability of *AI-sentiment* and *HD-sentiment* on the daily returns using an event-time regression as follows:

$$R_{i,d+T} = b + \alpha AI\text{-}sentiment_{i,d} + \beta HD\text{-}sentiment_{i,d} + \gamma SUE_{i,d} + \delta x_{i,d} + \varepsilon_{i,d} \quad (5)$$

where the dependent variable, $R_{i,d+T}$, is stock $i$'s cumulative daily returns on the $T$ trading days after the event day $d$ of an earnings call. If an earnings call takes place after regular trading hours (i.e., after 4 p.m.) on day $d$, we use the next trading day as the event day $d$. This adjustment accounts for the fact that average investors are unable to trade after hours and must wait until the subsequent trading day to respond. $AI\text{-}sentiment_{i,d}$ is the *AI-sentiment* score of stock $i$'s earnings call on day $d$ computed by gpt-3.5-turbo-16k API model. $HD\text{-}sentiment_{i,d}$ is the human dictionary sentiment score of stock $i$'s earnings call on day $d$ computed by LM dictionary. $SUE_{i,d}$ is earnings surprise derived based on a random walk model. $AI\text{-}sentiment_{i,d}$, $HD\text{-}sentiment_{i,d}$ and $SUE_{i,d}$ are all standardized in the regression. $x_{i,d}$ is the vector of aforementioned control variables. We include year-quarter fixed effects and cluster standard errors by firm and year-quarter. We multiply the regression coefficients by 100 to ease interpretation. A significantly positive coefficient $\alpha(\beta)$ implies that *AI-sentiment* (*HD-sentiment*) positively predicts future returns.

We present the return predictability of *AI-sentiment* and *HD-sentiment* in Table 2 using both full sample (Panel A) and post-final knowledge sample (Panel B). Panel A column (1) shows that the market quickly reacts on the next trading day after an earnings call is held (coefficient = 0.13, t-statistic = 7.54). This result indicates that one unit standard deviation increase in *AI-sentiment* on earnings call event day correlates with an average return increase of 0.13% on the next trading day, after controlling for *HD-sentiment*. Furthermore, the predictive power of *AI-sentiment* on returns amplifies over time, and plateaus at the six-month mark (coefficient = 0.99, t-statistic = 5.42), but maintains statistical significance up to one year post-earnings call (coefficient = 1.04, t-statistic = 3.71). Specifically, one unit standard deviation uplift in *AI-sentiment* on the earnings call event day is associated with an average return enhancement of 1.04% over

the following one year. In stark contrast, the return predictability of *HD-sentiment* demonstrates relatively weak return predictability both in terms of magnitude and horizon, with its effect notably disappearing after the first three days post-earnings calls. Even within this three-day window, the magnitude of *HD-sentiment*'s return predictability is merely one-third of that observed for *AI-sentiment*, with coefficients ranging from 0.032 to 0.058 and t-statistics from 1.71 to 2.09 across columns (1)-(3). In Panel B, we replicate the same regression using the post-final knowledge sample and the contrasting pattern between *AI-sentiment* and *HD-sentiment* holds. *AI-sentiment*'s predictability of returns still extends up to a quarter (coefficient = 0.75, t-statistic = 2.88), despite the substantial loss of observations. *HD-sentiment* fails to predict returns across all examined time horizons. Notably, the coefficients of *HD-sentiment* turn significantly negative in columns (9) and (10), indicating that human dictionary even predicts long-term returns in the wrong direction, leading to negative returns one year post-earnings call (coefficient = -2.25, t-statistic = -3.13).

Our tests substantiate the long-run return predictability of *AI-sentiment* after the earnings call event day, implying that investors can extract significant investment insights from earnings calls by utilizing democratized AI. This advantage of AI democratization holds true even for those investors unable to trade after hours and is especially crucial in the current era with broadly accessible AI technologies.

## 2.2. Calendar Time Portfolios

We further reinforce the actionability of AI insights by employing a calendar time portfolio approach. At the end of each month, we sort stocks into monthly decile portfolios based on *AI-sentiment*, *HD-sentiment*, and *SUE* from the most recent available earnings calls. We use both equal-weighted and value-weighted portfolios where market values are capped at upper 80 percent following Jensen, Kelly, and Pedersen (2023). The *High* (*Low*) portfolio comprises stocks in the highest (lowest) decile. The *High-Low* portfolio is the long-short portfolio.

Table 3, Panel A reports the CAPM alphas for *AI-sentiment*, *HD-sentiment*, and *SUE* portfolios. *High AI-sentiment* portfolio has an equal-weighted CAPM alpha of 0.48% (t-statistic = 2.89), and the long-short

*AI-sentiment* portfolio CAPM alpha is 1.01% (t-statistic = 3.06). In contrast, *HD-sentiment* portfolios do not yield significant CAPM alphas, whether in the *High* (*Low*) or the long-short configurations.

Furthermore, the long-short *SUE* portfolio yields a CAPM alpha of 1.57% (t-statistic = 3.61), consistent with the PEAD phenomenon. The equal-weighted CAPM alpha of long-short *AI-sentiment* portfolio is close to that of long-short *SUE* portfolio, indicating that *AI-sentiment* is as a valid signal as earnings surprises to identify stocks that will outperform or underperform the market. Additionally, despite similar abnormal returns of long-short portfolios, *AI-sentiment* and *SUE* diverge in their performance within the *High* or *Low* portfolios. Specifically, the *Low* portfolio based on *SUE* sorts significantly generates a negative abnormal return of -2.47% (t-statistic = -3.91), indicating that *SUE* identifies stocks that underperform the market due to negative earnings surprises. Conversely, the *High* portfolio based on *AI-sentiment* sorts significantly generates a positive abnormal return of 0.48% (t-statistic = 2.89), suggesting an effective capture of stocks likely to outperform based on positive market sentiment analyzed by AI. This contrast suggests that *AI-sentiment* may exploit positive market information more effectively than *SUE*, particularly in identifying stocks for superior performance. When we use value-weighted portfolios, the observed patterns persist. Long-short *AI-sentiment* (*SUE*) portfolio generates a CAPM alpha of 0.59% (0.91%) with t-statistic of 2.55 (1.84). *High AI-sentiment* (*Low SUE*) portfolio generates a CAPM alpha of 0.29% (-1.66%) with t-statistic of 2.43 (-2.86).

In Table 3, Panel B, we compute the Fama-French three-factor model alphas (FF3), Carhart four-factor model alphas (Carhart 4), and Fama-French five-factor model alphas (FF5) for *AI-sentiment* portfolios. Long-short *AI-sentiment* portfolio alphas range from 0.33% (value-weighted portfolio in Carhart four-factor model) to 0.83% (equal-weighted portfolio in five-factor model) and most are statistically significant despite a weaker significance level for Carhart four-factor models. Furthermore, all *High AI-sentiment* portfolios generate significantly positive abnormal returns, while the *Low AI-sentiment* portfolios do not yield any significant abnormal returns, consistent with the pattern observed in Table 3, Panel A.

In Table 3, Panel C, we further perform sequential double sorts on *HD-sentiment* (*SUE*) and *AI-sentiment*. At the end of each month, we sort stocks into terciles first based on *HD-sentiment* (*SUE*) and

then on *AI-sentiment*. We report the average CAPM alphas (CAPM), Fama-French three-factor model alphas (FF3), and Fama-French five-factor model alphas (FF5) for *AI-sentiment* across *High*, *Medium*, and *Low HD-sentiment* (*SUE*) portfolios. Long-short *AI-sentiment* portfolio alphas range from 0.37% to 0.65% with t-statistics ranging from 2.21 to 3.38. These results confirm the robust alpha generation capability of the long-short *AI-sentiment* portfolio even when controlling for *HD-sentiment* and *SUE*, thereby strengthening the argument that AI insights are actionable and extend beyond those offered by earnings surprises alone.

Table 3 Panel A-C focus on monthly returns generated by portfolio sorts. As shown in Section 2.1, AI-sentiment predicts one-year cumulative daily returns post-earnings call. We now explore if this long-term return predictability persists in portfolio sort analysis. Table 4 examines the cumulative monthly CAPM alphas for long-short AI-sentiment, HD-sentiment, and SUE portfolios. An equal-weighted long-short AI-sentiment portfolio yields significant cumulative CAPM alphas, rising from 1.68% over 2 months (t-statistic = 3.55) to 4.33% over 9 months (t-statistic = 3.93). Conversely, the long-short HD-sentiment portfolio does not generate significant long-term CAPM alphas, consistent with its muted long-term return predictability observed in Section 2.1. The contrasting pattern holds in value-weighted portfolio sorts. A value-weighted long-short AI-sentiment generates a 9-month CAPM alpha of 4.16% (t-statistic = 4.84), while the long-short HD-sentiment portfolio shows no significant long-term effect. Additionally, an equal-weighted long-short SUE portfolio yields a 9-month CAPM alpha of 3.68% (t-statistic = 3.61), while the value-weighted long-short SUE portfolio generates no significant long-term CAPM alpha. Collectively, Table 4 presents the long-term alpha generation capability of long-short AI-sentiment portfolio and is reconciled with the long-term return predictability of AI-sentiment demonstrated in Section 2.1.

To shed light on what difference may make *AI-sentiment* generate greater return predictability than *HD-sentiment*, we segment *HD-sentiment* into ten distinct groups and calculate the average values of both *AI-sentiment* and *HD-sentiment* within each *HD-sentiment* group. *AI-sentiment* and *HD-sentiment* are standardized for comparability. Figure 1 displays a positive linear relationship between *AI-sentiment* and *HD-sentiment*. However, *AI-sentiment* exhibits notable stability across the spectrum of *HD-sentiment*,

revealing a low sensitivity to variations in *HD-sentiment*. This steadiness is observed irrespective of *HD-sentiment* being negative, zero, or positive, indicating that changes in human-derived sentiment do not significantly influence the AI-generated sentiment scores. Specifically, whereas a human might interpret extensive use of positive or negative language during earnings calls as indicative of exceptionally positive or negative news, our findings suggest that AI, exemplified by ChatGPT, discerns beyond the superficial narrative and remains unaffected by the extremities of sentiment expressed during these communications.

## 3. High-Frequency Return Predictability of AI

Earnings calls are mostly held on the same day as earnings announcements. However, it is unclear whether the market reacts to earnings calls or other concurrent events like earnings announcements/press releases. In this section, we refine the return predictability tests in Section 2 by identifying the high-frequency daily returns.

To disentangle intraday returns to those related to earnings announcements/press releases (before call) and those related to calls (during or after call), we create return measures of three intervals within an earnings call date: before, during, and after the call, which are more refined than conventional daily returns. These granular intervals offer a more precise assessment of daily returns, enabling us to filter out the noise from simultaneous intra-day events.

We match each observation to the trading price from TAQ database. TAQ provides granular trading data at second-level or millisecond-level which covers both intraday and after-hour periods, helping us identify the real-time learning of the market. Specifically, for each firm that holds an earnings call on a particular date $d$, we first match this observation with several prices: 1) the last volume-weighted average trading price before the earnings call starts on day $d$ ($P^i_{lastbefore,d}$); 2) the last volume-weighted average trading price during the call on day $d$ ($P^i_{lastduring,d}$); 3) the closing price on day $d$-1 ($P^i_{close,d-1}$) and day $d$ ($P^i_{close,d}$) from CRSP. $d$-1 is the last trading day prior to earnings call day $d$. If a call is held after hours on day $d$, $P^i_{close,d}$ is based on the next trading day as $d$ because average investors are unable to trade after

hours and must wait until the subsequent trading day to respond. We then created three high-frequency returns: before-call return ($r^i_{before,d}$), during-call return ($r^i_{during,d}$), and after-call return ($r^i_{after,d}$) as below:

$$r^i_{before,d} = \frac{P^i_{lastbefore,d}}{P^i_{close,d-1}} - 1 \tag{6}$$

$$r^i_{during,d} = \frac{P^i_{lastduring,d}}{P^i_{lastbefore,d}} - 1 \tag{7}$$

$$r^i_{after,d} = \frac{P^i_{close,d}}{P^i_{lastduring,d}} - 1 \tag{8}$$

We present the high-frequency return predictability of *AI-sentiment* and *HD-sentiment* in Table 5, Panel A (full sample) and Panel B (post-final knowledge sample).11 Panel A columns (1)-(3) suggest that *AI-sentiment* and *HD-sentiment* are both significantly and positively related to high-frequency returns over the intervals before, during, and after the call. We also add the cumulative return from earnings call start to market close (i.e., during and after call) in column (4), which can be regarded as the summation of columns (2) and (3). During- and after-call returns predicted by *AI-sentiment* (coefficients = 0.60, t-statistic = 6.84) account for nearly 30% of the overall daily return, meaning the rest of the daily return is explained by before-call information. The relatively large magnitude of the relation to before-call returns does not indicate return predictability, but rather suggests that earnings call information indeed echoes information previously disclosed through press releases or other news channels as earnings are controlled for. Nonetheless, the regression coefficients are still economically large for returns during and after the call. This supports our conclusion that market immediately reacts once the earnings call starts, suggesting both *AI-sentiment* and *HD-sentiment* provide important investment insights for some privileged traders who possess high-frequency and after-hour trading capabilities.

[11] We miss some observations when calculating interval returns for two reasons. First, some calls have missing audio recordings and we cannot identify their start and end times to compute interval returns. Second, we need to link firm-level identifiers to each call using WRDS's ccmxpf_linktable. However, over half of observations in ccmxpf_linktable having missing permno, causing some calls to have missing permno. Thus, when we further use TAQ's permno-symbol link to match the permno of calls with symbol from TAQ trading price, those calls with missing permno cannot be matched to TAQ and have missing interval returns.

Furthermore, as shown in Table 2, *AI-sentiment*'s ability to forecast the long-term returns after the call (coefficient = 1.04, t-statistics = 3.71) is nearly three times greater than that for intraday after-call return (coefficient = 0.37, t-statistics = 5.04), suggesting that AI not only provides investment guidance in the high-frequency interval but also delivers more substantive actionable investment insights over the long term, which is particularly relevant for ordinary investors who are likely to trade slowly. In Table 5, Panel B, the results remain similar when we repeat the same regression using post-final knowledge sample. Interestingly, *HD-sentiment* shows significant predictability for during- and after-call returns whether we use full sample (coefficient = 0.32, t-statistics = 5.62) or post-final knowledge sample (coefficient = 0.41, t-statistics = 4.03). This finding suggests that, although *HD-sentiment* does not predict long-term returns, it is effective in forecasting high-frequency returns within the earnings call event day. Collectively, with sharply attributed high-frequency returns as a clean setting, our findings highlight that the potential of *AI-sentiment* lies in both long-term and high-frequency intraday return predictability.

# 4. Short Selling, Retail Trading, and AI

Our previous findings suggest that *AI-sentiment* provides superior return predictability over *HD-sentiment*. To further explore whether the earnings call information uncovered by *AI-sentiment* affects investors' trading dynamics, we delve into the daily trading behaviors of short sellers (the privileged investors) and retail investors (the ordinary investors) around earnings calls. In Section 4.1 and 4.2, we investigate how short selling and retail trading align with *AI-sentiment* around earnings calls, respectively. In Section 4.3, we explore the shifts in short selling and retail trading surrounding the wide deployment of ChatGPT and study the causal effect of AI democratization on trading. In Section 4.4, we examine the relation between retail trading and *AI-sentiment* when there are exogenous ChatGPT outages.

## 4.1. Short Selling

We first examine how short selling aligns with sentiment measures. Unlike the full sample, which ends in 2023, the sample period applied in Section 4.1 and 4.2 is 2010-2022, a phase we term the pre-democratization period. ChatGPT, released on November 30, 2022, achieved rapid growth, reaching 100

million monthly active users by January 2023.[12] This surge signals a heightened public engagement with ChatGPT and its transition to widespread use. Thus, we designate January 1, 2023, as the date when ChatGPT's deployment became widespread.

Table 6 reports the alignment between short selling and *AI-sentiment* around earnings calls in the pre-democratization period. We regress *ShortSelling*, a detrended abnormal short-seller holding measure constructed in Section 1.3, on *AI-sentiment*. We examine short selling on the earnings call event day and within 5, 10, and 21 trading days post-earnings calls. This relatively short window enables us to observe investor response to AI insights while minimizing the influence of long-term factors. Like the previous tests, we add *HD-sentiment* as a benchmark. We also control for *SUE* and other anomaly variables to rule out the possibility that short sellers simply trade on those public information signals. We include year-quarter fixed effects and cluster standard errors by firm and year-quarter. We multiply *ShortSelling* by 100000 to ease interpretation. A significantly negative coefficient for *AI-sentiment* indicates that short-selling activity increases as *AI-sentiment* decreases, suggesting that short sellers trade in the correct direction based on the earnings call information deciphered by AI.

Column (1) shows that short sellers demonstrate a pronounced reaction to *AI-sentiment* on the same day when an earnings call is held (coefficient = -706.9, t-statistic = -2.02), controlling for *HD-sentiment*, *SUE* and other anomaly variables. This means one standard deviation decrease in *AI-sentiment* is associated with 0.7% increase in abnormal short-seller holding. This suggests that short sellers increase short positions immediately on earnings call days if *AI-sentiment* is more negative. Interestingly, short sellers' alignment with *HD-sentiment* is stronger than that of *AI-sentiment* (coefficient = -2696.3, t-statistic = -6.24). This pattern suggests that short selling is aligned with both *AI-sentiment* and *HD-sentiment* of earnings calls. This is plausible since short sellers, as informed traders, have the capability to extract signals from various sources, including both advanced AI technologies and basic human judgment.

[12] See https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/.

Column (2)-(4) present how short selling aligns with *AI-sentiment* in the 5, 10, and 21 trading days following earnings calls. Short sellers continue to adjust their positions in alignment with *AI-sentiment* for up to one month following earnings calls (coefficient = -15370.6, t-statistic = -1.79). The amplifying magnitude of the correlation between short selling and *AI-sentiment* across columns (1) to (4) suggests that short sellers keep increasing positions after earnings call days, implying a long-term trend instead of a temporary adjustment.

These findings imply that short sellers' trading aligns with the information AI extracts from earnings calls. Given that ChatGPT was not widely accessible to the general investing public in the pre-democratization period, these findings suggest that short sellers are informed participants, with superior information sources or processing capacities such as proprietary AI models to analyze earnings calls.

## 4.2. Retail Trading

We then examine how retail trading aligns with sentiment measures. We follow the same regression and setting used in Section 4.1 and replace the dependent variable with *RetailTrading*, a detrended abnormal retail holding measure created in Section 1.4. To facilitate interpretation, we multiply *RetailTrading* by 100000. A significantly positive coefficient for *AI-sentiment* indicates that retail trading activity increases with *AI-sentiment*, suggesting that retail trading aligns with the insights derived from AI analysis of earnings calls.

Table 7 presents the alignment between retail trading and *AI-sentiment* around earnings calls in the pre-democratization period. In contrast to the pronounced patterns of short selling related to *AI-sentiment* observed in Table 6, retail investors exhibit reaction only to *HD-sentiment* but no significant reaction to *AI-sentiment* on earnings call days and during the one month after earnings calls. The negative and significant coefficients on *HD-sentiment* indicate that retail investors trade even against human understanding. These findings suggest that prior to the democratization of AI, unlike the privileged short sellers, retail investors may lack the sophistication required to distill actionable insights from earnings, and earlier efforts to democratize information may not have fully leveled the playing field for information access.

## 4.3. AI Democratization in 2023

We next analyze shifts in short selling and retail trading surrounding ChatGPT's wide deployment. As discussed in Section 4.1, we set January 1, 2023 as the date of ChatGPT's wide deployment. We define the pre- (post-) democratization periods as before (after) ChatGPT's wide deployment date. In Section 4.3.1, we examine how short-sellers-AI alignment and retail-AI alignment change around AI democratization using both full sample (2010-2023) and a shorter sample (2022-2023). In Section 4.3.2, we narrow the sample to 2022-2023 and implement a difference-in-differences (DiD) framework to address firm and firm-trader related endogeneity concerns and compare the difference in retail and short trading alignment with AI.

*4.3.1. Pre- and Post-AI Democratization*

Table 8 reports how short selling (Panel A and C) and retail trading (Panel B and D) are aligned with *AI-sentiment* before and after ChatGPT's wide deployment, where Panel A and B present full sample results and Panel C and D present 2022-2023 sample results. *After* is a dummy variable that equals 1 for post-democratization period, and 0 for pre-democratization period. We regress *ShortSelling* (*RetailTrading*) on *AI-sentiment* and *AI-sentiment*After*. Following Section 4.1 and 4.2, we focus on short selling (retail trading) on earnings call days and over 5, 10, and 21 trading days after. We include *HD-sentiment*, *SUE*, and other anomaly variables used in previous tests as controls. We include year-quarter fixed effects and cluster standard errors by firm and year-quarter. Our variable of interest is *AI-sentiment*After*, which captures the changes of short-sellers-AI alignment (retail-AI alignment) after AI democratization. A significantly negative (positive) coefficient for *AI-sentiment*After* means an increase in short sellers' (retail investors') alignment with *AI-sentiment* after AI democratization.

In Table 8, Panel A, the coefficient of *AI-sentiment*After* is insignificant across all horizons using the full sample, meaning that after AI democratization, short sellers have no significant change in alignment with *AI-sentiment*. Conversely, as shown in Table 8, Panel B, retail trading exhibits a significantly heightened reaction to *AI-sentiment* during the one month after earnings calls in the post-democratization period. To further explore the economic magnitude of short selling's and retail trading's response to AI democratization, we calculate a ratio that compares the changes in trading alignment with *AI-sentiment*.

Specifically, we divide the coefficient on *AI-sentiment*After*, which measures changes in trading behavior following AI democratization, by the coefficient on *AI-sentiment*, which represents the average trading alignment with *AI-sentiment* before AI democratization. We report these ratios in the bottom rows of Table 8, Panels A and B. In Table 8, Panel A, short-sellers-AI alignment shows no significant increase across all trading horizons, with ratios only ranging from 1.1 to 1.2. Conversely, in Table 8, Panel B, retail-AI alignment in the post-democratization period generally increases compared to their reaction before AI democratization. Notably, within ten days after earnings calls, retail investors' alignment with *AI-sentiment* after AI democratization (coefficient = 264.8) increases 65-fold compared to their long-term average alignment with *AI-sentiment* before AI democratization (coefficient = 4.05). Additionally, on the earnings call day, this ratio has a negative sign due to the negative coefficient of *AI-sentiment* (coefficient = -1.08) and positive coefficient of *AI-sentiment*After* (coefficient = 27.4). The flipping sign suggests that retail investors on average trade in the wrong direction (although insignificant) before AI democratization and is corrected after AI democratization.

Given a 13-year pre-democratization period may contain noisy events such as COVID-19 and meme stock frenzy that could simultaneously affect trading, we narrow the sample to 2022-2023 and define pre-democratization period as January to December 2022 and the post-democratization period as January to December 2023. Intuitively, the narrower the test window around the wide deployment event, the more likely it is that trading changes are related to ChatGPT's deployment rather than other confounding factors. In Table 8, Panel C and D, the contrast still holds when we use the short sample in 2022-2023. The significantly negative coefficients of *AI-sentiment* on short selling (retail trading) across column (1)-(4) suggest that short sellers' (retail investors') trading is strongly aligned with (against) *AI-sentiment* in the pre-democratization period in the year 2022. The significantly positive coefficients of *AI-sentiment*After* in Table 8, Panel D suggest that despite being "dumb" before AI democratization, retail investors trade in the correct direction after AI democratization. Interestingly, in Table 8, Panel C, the positive coefficient of *AI-sentiment*After* on short selling suggests after AI democratization, short sellers become "less smart"

and no longer enjoy a clear information advantage by trading on these AI insights, despite the change being weakly significant.

These patterns indicate that retail investors who are unable to engage with AI strategies before AI democratization begin to trade with AI afterward, while short sellers who previously use AI strategies do not exhibit increased alignment with AI post-democratization. The findings suggest that democratized AI usage enables retail investors to align their trading decisions more closely with the actionable insights extracted by AI from earnings calls. In contrast, short sellers show no significant increase in alignment with AI democratization, as they already possess privileged information access and advanced technologies for analyzing earnings calls before AI democratization.

*4.3.2. Difference-in-Differences*

Section 4.3.1 highlights the significant increase in retail-AI alignment, contrasted with the muted change in short-sellers-AI alignment, from pre- to post-AI democratization periods. The contrasting responses of retail investors and short sellers are consistent with the hypothesis that more democratized usage of AI tools benefits retail investors by aligning their trading decisions more closely with AI-generated insights, while simultaneously diminishing the informational edge traditionally held by short sellers.

To rigorously test this hypothesis, we develop a difference-in-differences (DiD) framework to address firm- and trader-related endogeneity concerns. The framework compares the differences in retail and short trading alignment with AI within the same firm, thereby mitigating potential endogeneity issues arising from unobserved firm characteristics. Specifically, we employ the following regression:

$$StandardizedTrading_{i,g,d} = \beta_0 + \beta_1 AI\text{-}sentiment_{i,d} + \beta_2 AI\text{-}sentiment_{i,d} * Treat_g^{IsRetail} + \beta_3 AI\text{-}sentiment_{i,d} * After_d + \beta_4 AI\text{-}sentiment_{i,d} * Treat_g^{IsRetail} * After_d + \delta x_{i,g,d} + \alpha_{i,g} + \gamma_{d,g} + \varepsilon_{i,g,d} \quad (9)$$

where the dependent variable is the standardized trading from trader $g$ on stock $i$ on earnings call day $d$. We categorize retail investors as the treatment group, given their increased alignment with *AI-sentiment* post-democratization, and short sellers as the control group, due to their comparatively subdued alignment. We stack the retail investor-stock and short seller-stock samples during 2022-2023 to form a unified dataset.

Since each firm holds earnings calls on a quarterly basis, our sample is structured at the stock-trader-quarter level. This includes both retail trading (treatment group) and short selling (control group) for each stock in each quarter.

To ensure comparability between retail trading and short selling, we first reverse the sign of *ShortSelling* so that a positive correlation between *ShortSelling* and *AI-sentiment* reflects short selling is aligned with *AI-sentiment*. We then standardize both *ShortSelling* and *RetailTrading* to create a unified dependent variable *StandardizedTrading*. $Treat^{IsRetail}$ is a dummy variable that equals 1 if the stock in a given quarter is traded by retail investors and 0 if traded by short sellers. *After* is a dummy variable that equals 1 for post-democratization period (January to December 2023), and 0 for pre-democratization period (January to December 2022). The variable of interest is *AI-sentiment*Treat*$^{IsRetail}$**After*. A positive (negative) coefficient for *AI-sentiment*Treat*$^{IsRetail}$**After* indicates an increase (decline) in retail-AI alignment after AI democratization compared to that of short sellers. $x_{i,g,d}$ is the vector of control variables including *HD-sentiment*, *SUE*, *Beta*, *BM*, *MVE*, and *Mom12m*. We include stock-trader fixed effects ($\alpha_{i,g}$) to control for time-invariant characteristics specific to each stock-trader group combination, and year-quarter-trader fixed effects ($\gamma_{d,g}$) to control for time-specific effects that are unique to each trader group. We do not include $Treat_g^{IsRetail}$, $After_d$, and $Treat_g^{IsRetail} * After_d$ in the regression as they are absorbed by these fixed effects. We cluster standard errors by firm and year-quarter.

Table 9 presents the results of the DiD test examining the relation between AI democratization and trading activities. The negative coefficients for *AI-sentiment*Treat*$^{IsRetail}$ suggest that retail investors are less sensitive to *AI-sentiment* compared to short sellers in the pre-democratization period. The negative coefficients for *AI-sentiment*After* suggest that short selling is less aligned with *AI-sentiment* after AI democratization, consistent with the pattern documented in Table 8 Panel C. The coefficients for *AI-sentiment*Treat*$^{IsRetail}$**After* are positive and significant across all trading horizons, indicating that retail trading aligns more closely with *AI-sentiment* following AI democratization compared to short selling. Specifically, in column (1), the coefficient for *AI-sentiment*Treat*$^{IsRetail}$**After* is 0.09 (t-statistic = 2.68),

meaning that retail trading increases by 0.09 standard deviations more than short selling in alignment with a unit change in *AI-sentiment* after AI democratization. This finding confirms a robust shift in retail trading behavior towards greater alignment with AI insights than short selling post-democratization, suggesting that AI democratization bridges the information gap between privileged short sellers and general retail investors.

Moreover, we repeat this analysis for each half-year period between 2021 and 2023. Specifically, we replace *After* with a series of half-year dummy variables starting from 2021HY2 to 2023HY2 in the regression, with the first half of 2021 (2021HY1) serving as the reference period. This allows us to benchmark each subsequent half-year's coefficient against 2021HY1. Figure 2 plots retail-AI alignment relative to short-sellers-AI alignment over time. Retail-AI alignment remains stable during 2021-2022, but shows a significant increase from 2023 onwards. These results help rule out the possibility of pre-existing trends before AI democratization and suggest a robust increase in retail-AI alignment caused by AI democratization shocks.

### 4.4. ChatGPT Outages

To further reduce the endogeneity concerns of the relation between AI democratization and trading, we exploit ChatGPT outages as exogenous shocks to identify changes in the availability of AI tools for retail investors. If AI democratization benefits retail investors in processing earnings call information, unexpected ChatGPT outages would diminish this effect. Specifically, we incorporate the outage variable into the DiD framework discussed in Section 4.3.2. We extract outage events from OpenAI's official reports, detailing the date, start and end time, type (web or API errors), and severity (major or partial outage).[13][14] Since ChatGPT, and therefore data on its outages, becomes available only after AI democratization, we name the variable *After&Outage* to denote that the measurement of outages pertains exclusively to the post-democratization context. *After&Outage* is a continuous variable representing the duration of outages

[13] See https://status.openai.com/history.

[14] While OpenAI's official reports record outage events before 2023, these outages are specific to earlier models that are not widely accessible to most investors before AI democratization and are thus irrelevant in our analysis.

occurring on the same day as an earnings call. We focus on major outages caused by ChatGPT web errors, as these outages are more severe and more likely to affect retail investors who primarily use the web version rather than the more complex API. We require the outage start time to be after the end time of the earnings call. For multiple outages occurred within the day, we sum the total duration. The variable of interest is the interaction term *AI-sentiment*Treat*$^{IsRetail}$**After&Outage* which captures whether the increased retail-AI alignment relative to short-sellers-AI alignment after AI democratization is muted when ChatGPT experiences outages. We use the 2022-2023 sample period, include firm-trader and year-quarter-trader fixed effects, and cluster standard errors by firm and year-quarter.

Table 10 presents how retail-AI alignment changes relative to short-sellers-AI alignment when ChatGPT outages occur after AI democratization. The coefficients of *AI-sentiment*Treat*$^{IsRetail}$**After* are positively significant, consistent with the results in Table 9 that AI democratization leads to increased retail-AI alignment relative to that of short sellers. In contrast, the coefficients for *AI-sentiment*Treat*$^{IsRetail}$**After&Outage* are negative and significant across all time horizons. For example, in column (1), the coefficient for *AI-sentiment*Treat*$^{IsRetail}$**After* is 0.091 (t-statistics = 2.69), while the coefficient for *AI-sentiment*Treat*$^{IsRetail}$**After&Outage* is -0.075 (t-statistics = -3.74). This represents a reduction of over two-thirds (0.075 / 0.091) in the relation between AI democratization and retail trading for each additional hour of ChatGPT outages. This significant decrease highlights that the retail-AI alignment is notably diminished during ChatGPT outage periods. Overall, our identification strategies reveal that while retail investors align more with *AI-sentiment* after AI democratization, retail-AI alignment significantly decreases when ChatGPT outages occur on the same day as an earnings call. This suggests that ChatGPT outages, which unexpectedly limit retail investors' ability to utilize AI insights, reduce the alignment between retail trading and *AI-sentiment*. These findings further strengthen the causal relationship between AI democratization and changes in retail trading behavior.

## 5. Information Asymmetry and AI

AI democratization provides actionable trading insights to uninformed investors. As sophisticated investors lose some of their informational edge over uninformed investors, the risk of uninformed investors engaging

in unfavorable trades (e.g., buying overpriced stocks or selling underpriced ones) is reduced. This decrease in information asymmetry alleviates adverse selection, encouraging trading and enhancing market liquidity (Glosten and Milgrom 1985; Kyle 1985). Therefore, if AI democratization enhances retail investors' ability to process information, it should reduce information asymmetry and improve market liquidity.

We examine the relation between AI democratization and information asymmetry. We measure information asymmetry using bid-ask spread and conduct a propensity score matching-difference in differences (PSM-DiD) framework as follows:

$$BidAskSpread_{i,d} = \beta_0 + \beta_1 Treat_i^{AI} + \beta_2 Treat_i^{AI} * After_d + \delta x_{i,d} + \alpha_d + \varepsilon_{i,d} \quad (10)$$

where *BidAskSpread* is the difference between the daily bid and ask price scaled by the average of bid and ask price of stock $i$ on earnings call day $d$. We define the treatment and control groups based on the change in retail-AI alignment after AI democratization. Stocks with more increased retail-AI alignment after AI democratization experience more AI-informed retail trading, leading to greater price informativeness and, consequently, reduced information asymmetry. Specifically, we run the regression (9) on a stock-by-stock basis and define each stock's retail-AI alignment using the coefficient of *AI-sentiment*Treat*$^{IsRetail}$**After*. A higher coefficient of *AI-sentiment*Treat*$^{IsRetail}$**After* means a greater increase in retail-AI alignment after AI democratization compared to that of short sellers. *Treat*$^{AI}$ is a dummy variable that equals 1 if a stock's retail-AI alignment is above the median and 0 otherwise. The treatment and control groups are balanced using propensity score matching.[15] *After* is a dummy variable that equals 1 for post-democratization period (January to December 2023), and 0 for pre-democratization period (January to December 2022). The variable of interest is *Treat*$^{AI}$**After*. A negative (positive) coefficient for *Treat*$^{AI}$**After* indicates a decrease (increase) in information asymmetry for high retail-AI alignment stocks after AI democratization relative to low retail-AI alignment stocks. $x_{i,d}$ is the vector of control variables including *AI-sentiment*, *HD-sentiment*, *SUE*, *Beta*, *BM*, *MVE*, and *Mom12m*. We include year-quarter fixed effects ($\alpha_d$) and cluster

[15] We employ nearest neighbor propensity score matching. Table IA1 shows a significant difference between the unmatched treatment and control groups. After matching, the differences between the treatment and control groups become statistically insignificant, suggesting that the groups are well-balanced.

standard errors by firm and year-quarter. We do not include $After_d$ in the regression as it will be absorbed by the year-quarter fixed effects.

Table 11 presents the results of the PSM-DiD test examining the relation between AI democratization and information asymmetry. We report the bid-ask spread on earnings call days and its moving average over 5, 10, and 21 trading days after. The positive coefficients for $Treat^{AI}$, while mostly insignificant, suggest that high retail-AI alignment stocks have greater information asymmetry. Conversely, the coefficients for $Treat^{AI}$*$After$ are negative, suggesting a significant decrease in information asymmetry for stocks with more increased retail-AI alignment after AI democratization. In column (4), the coefficient for $Treat^{AI}$*$After$ is -25.7 (t-statistic = -3.71), while the coefficient for $Treat^{AI}$ is 15.0 (t-statistic = 0.82). This means after AI democratization, information asymmetry of high retail-AI alignment stocks decreases nearly two-fold (25.7 / 15.0) related to low retail-AI alignment stocks.

We further incorporate ChatGPT outages into the PSM-DiD framework. If AI democratization reduces information asymmetry, unexpected ChatGPT outages would weaken this effect. We measure ChatGPT outages using *After&Outage* as mentioned in Section 4.4. A positive (negative) coefficient for $Treat^{AI}$*$After\&Outage$ suggests an increase (decrease) in information asymmetry for high retail-AI alignment stocks after AI democratization relative to low retail-AI alignment stocks when ChatGPT outages occur.

Table 12 presents the relation between AI democratization and information asymmetry when ChatGPT outages occur. The negative coefficients for $Treat^{AI}$*$After$ are consistent with the results in Table 11 that AI democratization reduces information asymmetry for stocks with high retail-AI alignment during non-outage period. In contrast, the positive coefficients for $Treat^{AI}$*$After\&Outage$ suggest that information asymmetry for high retail-AI alignment stocks notably increases during ChatGPT outages after AI democratization. Specifically, in column (1), the coefficient for $Treat^{AI}$*$After$ is -27.1 (t-statistic = -1.80), while the coefficient for $Treat^{AI}$*$After\&Outage$ is 67.6 (t-statistic = 4.01). This indicates that for each additional hour of ChatGPT outages, bid-ask spread increases by 2.5-fold (67.6 / 27.1) for high retail-AI alignment stocks which initially benefit from AI democratization. Overall, these findings suggest that while

AI democratization lowers information asymmetry for stocks with high retail-AI alignment, ChatGPT outages limit the access of AI tools to retail investors, thereby increasing information asymmetry.

## 6. Conclusion

Reg FD aims to democratize investment information, but its effectiveness is debated due to challenges like average investors' limited capacity to utilize complex information promptly. The advent of AI democratization introduces a new layer of tension to this discourse. We conduct the first analysis of how democratized AI impacts human traders' use of information. Using ChatGPT, the most widespread and advanced AI technology, to analyze the longest sample of 19-year earnings calls, we discover that ChatGPT-generated *AI-sentiment* exhibits a strong predictability of stock returns over a 12-month period, after controlling for human-dictionary based *HD-sentiment* and standardized unexpected earnings. A long-short trading strategy based on the latest month's *AI-sentiment* yields a monthly alpha of 1%.

In the years before ChatGPT exists, short selling is notably aligned with *AI-sentiment* in the two weeks following earnings calls, whereas retail trading showed little alignment. After ChatGPT's wide deployment, retail investors' alignment with *AI-sentiment* has increased 65-fold compared to that before, while short-sellers-AI alignment can even weaken. ChatGPT outages, which diminish retail investors' ability to utilize AI insights, reduce retail-AI alignment, further supporting the causal role of AI. Furthermore, there is a significant decrease in information asymmetry for stocks with high retail-AI alignment after AI democratization. These findings highlight the potential of AI in mining actionable insights from earnings calls that align with the knowledge of informed traders. Collectively, our findings suggest that the democratization of information access demands democratization of AI to effectively level the playing field, underscoring the transformative impact of "open" AI (usage) to bridge the information gap between the privileged and ordinary investors.

## References


Barber, B.M. and Odean, T., 2000. Trading is hazardous to your wealth: The common stock investment performance of individual investors. The Journal of Finance, 55(2), pp.773-806.

Basu, S. and Xiang, Z., 2022. Why don't analysts always value earnings conference calls. Available at SSRN.

Beckmann, L., Beckmeyer, H., Filippou, I., Menze, S. and Zhou, G., 2024. Unusual financial communication-evidence from ChatGPT, earnings calls, and the stock market (January 15, 2024). Available at SSRN.

Berkman, H. and Truong, C., 2009. Event day 0? After-hours earnings announcements. Journal of Accounting Research, 47(1), pp.71-103.

Blankespoor, E., deHaan, E. and Marinovic, I., 2020. Disclosure processing costs, investors' information choice, and equity market outcomes: A review. Journal of Accounting and Economics, 70(2-3), p.101344.

Blocher, J., Dong, X., Ringgenberg, M.C. and Savor, P.G., 2021. Short covering. Available at SSRN 2634579.

Boehmer, E., Jones, C.M. and Zhang, X., 2008. Which shorts are informed?. The Journal of Finance, 63(2), pp.491-527.

Boehmer, E., Jones, C.M., Zhang, X. and Zhang, X., 2021. Tracking retail investor activity. The Journal of Finance, 76(5), pp.2249-2305.

Brochet, F., Kolev, K. and Lerman, A., 2018. Information transfer and conference calls. Review of Accounting Studies, 23, pp.907-957.

Brown, S., Hillegeist, S.A. and Lo, K., 2004. Conference calls and information asymmetry. Journal of Accounting and Economics, 37(3), pp.343-366.

Cao, S., Jiang, W., Yang, B. and Zhang, A.L., 2023. How to talk when a machine is listening: Corporate disclosure in the age of AI. The Review of Financial Studies, 36(9), pp.3603-3642.

Chen, Y., Da, Z. and Huang, D., 2019. Arbitrage trading: The long and the short of it. The Review of Financial Studies, 32(4), pp.1608-1646.

Chen, J., Tang, G., Zhou, G. and Zhu, W., 2023. ChatGPT, stock market predictability and links to the macroeconomy. Available at SSRN 4660148.

Cheynel, E. and Levine, C.B., 2020. Public disclosures and information asymmetry: A theory of the mosaic. The Accounting Review, 95(1), pp.79-99.

Child, R., Gray, S., Radford, A. and Sutskever, I., 2019. Generating long sequences with sparse transformers. arXiv preprint arXiv:1904.10509.

Christophe, S.E., Ferri, M.G. and Hsieh, J., 2010. Informed trading before analyst downgrades: Evidence from short sellers. Journal of Financial Economics, 95(1), pp.85-106.

Clinch, G. and Li, W., 2022. Short-Selling and Information Arrival around Earnings Announcements: Evidence from Regulation SHO. The Accounting Review, 97(4), pp.237-258.

Cohen, L., Lou, D. and Malloy, C.J., 2020. Casting conference calls. Management Science, 66(11), pp.5015-5039.

Da, Z., Engelberg, J. and Gao, P., 2011. In search of attention. The Journal of Finance, 66(5), pp.1461-1499.

Dai, Z., Yang, Z., Yang, Y., Carbonell, J., Le, Q.V. and Salakhutdinov, R., 2019. Transformer-xl: Attentive language models beyond a fixed-length context. arXiv preprint arXiv:1901.02860.

DellaVigna, S. and Pollet, J.M., 2009. Investor inattention and Friday earnings announcements. The Journal of Finance, 64(2), pp.709-749.

DeLong, J.B., Shleifer, A., Summers, L.H. and Waldmann, R.J., 1990. Noise trader risk in financial markets. Journal of Political Economy, 98, 703-738.

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

Glasserman, P., Lin, C., 2023. Assessing look-ahead bias in stock return predictions generated by gpt sentiment analysis. arXiv preprint arXiv:2309.17322.
Glosten, L.R. and Milgrom, P.R., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14(1), pp.71-100.
Hansen, A.L. and Kazinnik, S., 2023. Can ChatGPT decipher Fedspeak?. Available at SSRN.
Hirshleifer, D., Lim, S.S. and Teoh, S.H., 2009. Driven to distraction: Extraneous events and underreaction to earnings news. The Journal of Finance, 64(5), pp.2289-2325.
Hollander, S., Pronk, M. and Roelofsen, E., 2010. Does silence speak? An empirical analysis of disclosure choices during conference calls. Journal of Accounting Research, 48(3), pp.531-563.
Huang, M., Martin, X., Tang, Y. and Xie, Y., 2022. Scapegoating: Uncontrollable Factors Disclosure and Disposition Effect. Available at SSRN 4043925.
Huang, A.H., Wang, H. and Yang, Y., 2022. FinBERT: A large language model for extracting information from financial text. Contemporary Accounting Research, 40(2), pp.806-841.
Huang, X., Jorion, P. and Schwarz, C., 2024. Who is minding the store? Monitoring and competition in retail trade execution. Available at SSRN.
Hvidkjaer, S., 2008. Small trades and the cross-section of stock returns. The Review of Financial Studies, 21(3), pp.1123-1151.
Jensen, T.I., Kelly, B. and Pedersen, L.H. 2023. Is there a replication crisis in finance?. Journal of Finance, 78: 2465-2518.
Jha, M., Qian, J., Weber, M. and Yang, B., 2024. ChatGPT and corporate policies (No. w32161). National Bureau of Economic Research.
Kaniel, R., Saar, G. and Titman, S., 2008. Individual investor trading and stock returns. The Journal of Finance, 63(1), pp.273-310.
Kaniel, R., Liu, S., Saar, G. and Titman, S., 2012. Individual investor trading and return patterns around earnings announcements. The Journal of Finance, 67(2), pp.639-680.
Kelley, E.K. and Tetlock, P.C., 2013. How wise are crowds? Insights from retail orders and stock returns. The Journal of Finance, 68(3), pp.1229-1265.
Kelley, E.K. and Tetlock, P.C., 2017. Retail short selling and stock prices. The Review of Financial Studies, 30(3), pp.801-834.
Kim, A., Muhn, M., Nikolaev, V., 2023a. From transcripts to insights: Uncovering corporate risks using generative AI. ArXiv preprint arXiv:2310.17721.
Kim, A.G., Muhn, M. and Nikolaev, V.V., 2023b. Bloated disclosures: Can ChatGPT help investors process information. Chicago Booth Research Paper, (23-07).
Krishna, K., Chang, Y., Wieting, J. and Iyyer, M., 2022. Rankgen: Improving text generation with large ranking models. arXiv preprint arXiv:2205.09726.
Kyle, A.S., 1985. Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pp.1315-1335.
Li, E.X.,Tu, Z. and Zhou, D. 2023 The promise and peril of Generative AI: Evidence from ChatGPT as sell-side analysts. Available at SSRN.
Liu, N.F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F. and Liang, P., 2023. Lost in the middle: How language models use long contexts. arXiv.
Lopez-Lira, A. and Tang, Y., 2023. Can chatgpt forecast stock price movements return predictability and large language models. Available at SSRN.
Loughran, T. and McDonald, B., 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance, 66(1), pp.35-65.
Madrigal, A.C., 2013. The remarkable decline in the Wall Street Journal's long-form journalism. The Atlantic.
Matsumoto, D., Pronk, M. and Roelofsen, E., 2011. What makes conference calls useful? The information content of managers' presentations and analysts' discussion sessions. The Accounting Review, 86(4), pp.1383-1414.

Mayew, W.J. and Venkatachalam, M., 2012. The power of voice Managerial affective states and future firm performance. The Journal of Finance, 67(1), pp.1-43.
Sarkar, S.K., and Vafa, K., 2024. Lookahead bias in pretrained language models. Available at SSRN.
Seger, E., Ovadya, A., Siddarth, D., Garfinkel, B. and Dafoe, A., 2023, August. Democratising AI: Multiple meanings, goals, and methods. In Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society (pp. 715-722).
Shleifer, A. and Vishny, R.W., 1997. The limits of arbitrage. The Journal of Finance, 52, 35-55.
Sun, S., Krishna, K., Mattarella-Micke, A. and Iyyer, M., 2021. Do long-range language models actually use long-range context. arXiv preprint arXiv2109.09115.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is all you need. Advances in neural information processing systems, 30.
Wang, X., Yan, X.S. and Zheng, L., 2020. Shorting flows, public disclosure, and market efficiency. Journal of Financial Economics, 135(1), pp.191-212.
Zhu, C., 2019. Big data as a governance mechanism. The Review of Financial Studies, 32(5), pp.2021-2061.

## Table 1 Summary Statistics

This table presents summary statistics of variables using full sample (Panel A) and post-final knowledge sample (Panel B). *AI-sentiment* is the sentiment score of earnings call transcripts computed by gpt-3.5-turbo-16k model. *HD-sentiment* is the sentiment score of earnings call transcripts computed by using Loughran-McDonald (LM) dictionary. *SUE* is the difference between reported earnings and expected earnings, assuming earnings follow a seasonal random walk trend. *Beta* is the market beta with respect to the CRSP equal-weighted return index. *BM* is the book value of equity over the market value of equity. *MVE* is the total market value of common equity. *Mom12m* is 12-month momentum computed as the cumulative return from month t-12 to t-2. *ShortSelling* is abnormal short-seller holding calculated in Section 1.3. *RetailTrading* is abnormal retail holding calculated in Section 1.4. *BidAskSpread* is the difference between the daily bid and ask price scaled by the average of bid and ask price.

*Panel A: Full Sample*

| Variable | N | Mean | P1 | P50 | P99 | Std |
|---|---|---|---|---|---|---|
| AI-sentiment | 81484 | 5.62 | -2.00 | 6.00 | 8.00 | 1.83 |
| HD-sentiment | 81484 | 0.25 | -0.33 | 0.26 | 0.67 | 0.22 |
| SUE | 81484 | 0.06 | -0.33 | 0.00 | 0.53 | 3.92 |
| Beta | 81484 | 1.24 | 0.16 | 1.18 | 2.92 | 0.56 |
| BM | 81484 | 0.51 | -0.52 | 0.42 | 2.38 | 0.68 |
| MVE | 81484 | 14.21 | 9.45 | 14.26 | 19.03 | 2.05 |
| Mom12m | 81484 | 0.18 | -0.79 | 0.09 | 2.60 | 0.83 |
| ShortSelling | 82228 | 0.22 | 0.01 | 0.11 | 2.48 | 0.49 |
| RetailTrading | 85015 | 0.0002 | -0.0180 | -0.0004 | 0.0329 | 0.0111 |
| BidAskSpread | 80745 | 0.0033 | 0.0000 | 0.0008 | 0.0410 | 0.0086 |

*Panel B: Post-Final Knowledge Sample*

| Variable | N | Mean | P1 | P50 | P99 | Std |
|---|---|---|---|---|---|---|
| AI-sentiment | 18590 | 5.11 | 0.00 | 5.00 | 8.00 | 2.20 |
| HD-sentiment | 18590 | 0.28 | -0.29 | 0.30 | 0.69 | 0.21 |
| SUE | 18590 | 0.13 | -0.47 | 0.00 | 0.94 | 6.20 |
| Beta | 18590 | 1.27 | 0.14 | 1.19 | 3.05 | 0.59 |
| BM | 18590 | 0.50 | -0.61 | 0.38 | 2.81 | 0.80 |
| MVE | 18590 | 14.21 | 9.10 | 14.31 | 19.26 | 2.24 |
| Mom12m | 18590 | 0.05 | -0.88 | -0.02 | 2.22 | 0.67 |
| ShortSelling | 19289 | 0.28 | 0.02 | 0.13 | 3.44 | 0.58 |
| RetailTrading | 19920 | -0.0004 | -0.0363 | -0.0005 | 0.0404 | 0.0138 |
| BidAskSpread | 18326 | 0.0044 | 0.0000 | 0.0011 | 0.0492 | 0.0098 |

## Table 2 Return Predictability of *AI-sentiment* and *HD-sentiment*

This table presents the return predictability of *AI-sentiment* and *HD-sentiment* using full sample (Panel A) and post-final knowledge sample (Panel B). *AI-sentiment* is the sentiment score of earnings call transcripts computed by gpt-3.5-turbo-16k model. *HD-sentiment* is the sentiment score of earnings call transcripts computed by using Loughran-McDonald (LM) dictionary. *SUE* is the difference between reported earnings and expected earnings, assuming earnings follow a seasonal random walk trend. *Beta* is the market beta with respect to the CRSP equal-weighted return index. *BM* is the book value of equity over the market value of equity. *MVE* is the total market value of common equity. *Mom12m* is 12-month momentum computed as the cumulative return from month t-12 to t-2. *AI-sentiment*, *HD-sentiment* and *SUE* are standardized for regression. We include year-quarter fixed effect. Standard errors are two-way clustered by firm and year-quarter. T-statistics are reported in parentheses. *, **, and *** represent significance at the 10%, 5% and 1% levels, respectively.

*Panel A: Full Sample*

| | (1) [d+1] | (2) [d+1 d+2] | (3) [d+1 d+3] | (4) [d+1 d+5] | (5) [d+1 d+10] | (6) [d+1 d+21] | (7) [d+1 d+42] | (8) [d+1 d+63] | (9) [d+1 d+126] | (10) [d+1 d+252] |
|---|---|---|---|---|---|---|---|---|---|---|
| AI-sentiment$_{i,d}$ | 0.13*** | 0.15*** | 0.18*** | 0.19*** | 0.24*** | 0.35*** | 0.46*** | 0.60*** | 0.99*** | 1.04*** |
| | (7.54) | (5.72) | (5.36) | (4.89) | (4.19) | (4.95) | (4.53) | (4.74) | (5.42) | (3.71) |
| HD-sentiment$_{i,d}$ | 0.032* | 0.051* | 0.058** | 0.045 | 0.0083 | 0.034 | 0.17 | 0.12 | 0.020 | -0.26 |
| | (1.71) | (1.97) | (2.09) | (1.21) | (0.14) | (0.45) | (1.33) | (0.79) | (0.08) | (-0.75) |
| SUE$_{i,d}$ | -0.027 | -0.030 | -0.034 | -0.012 | 0.11 | 0.040 | 0.31* | 0.17 | -0.26 | -0.43 |
| | (-0.95) | (-0.75) | (-0.68) | (-0.20) | (1.16) | (0.28) | (1.92) | (0.76) | (-0.74) | (-0.67) |
| Beta | 0.050 | 0.093 | 0.15* | 0.14 | 0.024 | 0.050 | 0.20 | 0.49 | 1.23 | 2.88*** |
| | (1.14) | (1.63) | (1.75) | (1.09) | (0.12) | (0.16) | (0.40) | (0.86) | (1.61) | (2.68) |
| BM | 0.043 | 0.058 | 0.068 | 0.11* | 0.11 | 0.074 | 0.18 | 0.26 | 0.80 | 0.54 |
| | (1.29) | (1.36) | (1.41) | (1.89) | (1.23) | (0.49) | (0.75) | (0.72) | (1.41) | (0.52) |
| MVE | 0.017* | 0.013 | 0.0081 | 0.012 | -0.018 | -0.0098 | -0.041 | -0.051 | -0.019 | -0.17 |
| | (1.81) | (1.03) | (0.57) | (0.57) | (-0.63) | (-0.19) | (-0.48) | (-0.39) | (-0.11) | (-0.53) |
| Mom12m | -0.048** | -0.062 | -0.11** | -0.18*** | -0.16** | 0.0017 | -0.014 | 0.17 | 0.086 | -0.50 |
| | (-2.03) | (-1.47) | (-2.64) | (-3.29) | (-2.49) | (0.01) | (-0.09) | (0.63) | (0.28) | (-0.94) |
| Year-Quarter FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Cluster | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr |
| Sample Period | 10-23 | 10-23 | 10-23 | 10-23 | 10-23 | 10-23 | 10-23 | 10-23 | 10-23 | 10-23 |
| N | 81484 | 81484 | 81484 | 81484 | 81484 | 81484 | 81484 | 81484 | 81484 | 81484 |
| r2 | 0.0043 | 0.0045 | 0.0049 | 0.0051 | 0.0048 | 0.0040 | 0.0038 | 0.0041 | 0.0034 | 0.0034 |

*Panel B: Post-Final Knowledge Sample*

| | (1)<br>[d+1] | (2)<br>[d+1 d+2] | (3)<br>[d+1 d+3] | (4)<br>[d+1 d+5] | (5)<br>[d+1 d+10] | (6)<br>[d+1 d+21] | (7)<br>[d+1 d+42] | (8)<br>[d+1 d+63] | (9)<br>[d+1 d+126] | (10)<br>[d+1 d+252] |
|---|---|---|---|---|---|---|---|---|---|---|
| AI-sentiment$_{i,d}$ | 0.13** | 0.18** | 0.23** | 0.24** | 0.28 | 0.53*** | 0.66*** | 0.75** | 0.70* | 0.65 |
| | (3.04) | (2.77) | (2.48) | (2.37) | (1.86) | (3.66) | (4.05) | (2.88) | (2.00) | (1.06) |
| HD-sentiment$_{i,d}$ | 0.0081 | 0.060 | 0.071 | 0.017 | -0.017 | 0.011 | 0.053 | -0.54 | -1.43** | -2.25** |
| | (0.15) | (0.69) | (0.71) | (0.14) | (-0.09) | (0.05) | (0.14) | (-1.57) | (-2.70) | (-3.13) |
| SUE$_{i,d}$ | -0.018 | -0.027 | 0.026 | 0.022 | 0.32 | -0.040 | 0.097 | 0.18 | -0.22 | -0.41 |
| | (-0.32) | (-0.39) | (0.26) | (0.20) | (1.18) | (-0.19) | (0.26) | (0.45) | (-0.35) | (-0.45) |
| Beta | -0.016 | -0.0086 | 0.036 | 0.0032 | -0.20 | -0.32 | -0.55 | 0.30 | 1.44 | 3.38** |
| | (-0.17) | (-0.08) | (0.16) | (0.01) | (-0.37) | (-0.45) | (-0.36) | (0.20) | (1.23) | (2.35) |
| BM | 0.14* | 0.17* | 0.10 | 0.14 | 0.20 | 0.32* | 0.72* | 1.21** | 2.05* | 3.42** |
| | (2.03) | (2.12) | (1.14) | (1.31) | (1.38) | (2.23) | (1.96) | (2.61) | (2.23) | (2.51) |
| MVE | 0.012 | -0.017 | -0.040 | -0.015 | -0.061 | 0.023 | 0.16 | 0.46 | 0.88** | 1.46** |
| | (0.51) | (-0.51) | (-1.17) | (-0.24) | (-0.90) | (0.19) | (0.69) | (1.85) | (3.07) | (2.86) |
| Mom12m | 0.015 | 0.037 | 0.0044 | -0.092 | -0.063 | 0.100 | -0.048 | 0.39 | 2.32** | 2.01** |
| | (0.27) | (0.51) | (0.05) | (-0.91) | (-0.36) | (0.22) | (-0.08) | (0.46) | (2.85) | (2.41) |
| Year-Quarter FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Cluster | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr |
| Sample Period | 21Oct-23Dec | 21Oct-23Dec | 21Oct-23Dec | 21Oct-23Dec | 21Oct-23Dec | 21Oct-23Dec | 21Oct-23Dec | 21Oct-23Dec | 21Oct-23Dec | 21Oct-23Dec |
| N | 18590 | 18590 | 18590 | 18590 | 18590 | 18590 | 18590 | 18590 | 18590 | 18590 |
| r2 | 0.0038 | 0.0044 | 0.0048 | 0.0055 | 0.0060 | 0.0051 | 0.0051 | 0.0062 | 0.0092 | 0.010 |

## Table 3 Portfolio Sorts

This table presents the alphas for monthly calendar-time equal-weighted and value-weighted single sorts portfolios using CAPM (Panel A), Fama-French three-factor (FF3) model, Carhart four-factor (Carhart4) model, and Fama-French five-factor (FF5) model (Panel B), and equal-weighted and value-weighted double sorts portfolios using CAPM, Fama-French three-factor (FF3) model, and Fama-French five-factor (FF5) model (Panel C). The sample period is 2010-2023. *AI-sentiment* is the sentiment score of earnings call transcripts computed by gpt-3.5-turbo-16k model. *HD-sentiment* is the sentiment score of earnings call transcripts computed by using Loughran-McDonald (LM) dictionary. *SUE* is the difference between reported earnings and expected earnings, assuming earnings follow a seasonal random walk trend. T-statistics are reported in parentheses. *, **, and *** represent significance at the 10%, 5% and 1% levels, respectively.

*Panel A: CAPM Alphas of Single Sorts Portfolios*

| | Equal Weighted | | | Value Weighted (80p) | | |
|---|---|---|---|---|---|---|
| | AI-sentiment | HD-sentiment | SUE | AI-sentiment | HD-sentiment | SUE |
| High | 0.48*** | 0.03 | -0.21 | 0.29** | 0.19 | -0.09 |
| | (2.89) | (0.13) | (-0.43) | (2.43) | (1.62) | (-0.24) |
| Low | -0.52 | -0.17 | -2.47*** | -0.30 | 0.01 | -1.66*** |
| | (-1.48) | (-0.68) | (-3.91) | (-1.60) | (0.06) | (-2.86) |
| High-Low | 1.01*** | 0.20 | 1.57*** | 0.59** | 0.18 | 0.91* |
| | (3.06) | (1.02) | (3.61) | (2.55) | (1.02) | (1.84) |

*Panel B: Three-Factor, Carhart Four-Factor, and Five-Factor Alphas of Single Sorts Portfolios*

| | Equal Weighted | | | Value Weighted (80p) | | |
|---|---|---|---|---|---|---|
| | High | Low | High-Low | High | Low | High-Low |
| FF3 | 0.60*** | -0.20 | 0.80*** | 0.32*** | -0.15 | 0.47** |
| | (4.81) | (-0.77) | (2.78) | (3.08) | (-1.02) | (2.38) |
| Carhart4 | 0.54*** | -0.06 | 0.60** | 0.27*** | -0.06 | 0.33* |
| | (4.48) | (-0.24) | (2.28) | (2.68) | (-0.42) | (1.83) |
| FF5 | 0.63*** | -0.19 | 0.83*** | 0.34*** | -0.22 | 0.56*** |
| | (5.00) | (-0.76) | (2.85) | (3.24) | (-1.44) | (2.81) |

*Panel C: CAPM, Three-Factor, and Five-Factor Alphas of Double Sorts Portfolios*

| | | Equal Weighted AI-sentiment | | | | | | Value Weighted (80p) AI-sentiment | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | High | Medium | Low | High-Low | | | High | Medium | Low | High-Low |
| HD-sentiment | CAPM | 0.17 | -0.32 | -0.48* | 0.65*** | HD-sentiment | CAPM | 0.16* | -0.13* | -0.32** | 0.48** |
| | | (1.02) | (-1.58) | (-1.67) | (3.07) | | | (1.84) | (-1.74) | (-2.21) | (2.52) |
| | FF3 | 0.33** | -0.08 | -0.18 | 0.52*** | | FF3 | 0.13 | -0.11 | -0.25* | 0.37** |
| | | (3.41) | (-0.93) | (-1.02) | (2.78) | | | (1.58) | (-1.52) | (-1.84) | (2.21) |
| | FF5 | 0.38** | -0.03 | -0.17 | 0.55*** | | FF5 | 0.13 | -0.12 | -0.3** | 0.43** |
| | | (3.97) | (-0.34) | (-0.93) | (2.95) | | | (1.59) | (-1.64) | (-2.22) | (2.49) |
| SUE | CAPM | 0.16 | -0.46** | -0.5* | 0.65*** | SUE | CAPM | 0.17** | -0.24** | -0.28* | 0.46*** |
| | | (0.87) | (-2.21) | (-1.78) | (3.26) | | | (2.28) | (-2.32) | (-1.94) | (2.7) |
| | FF3 | 0.33** | -0.23** | -0.21 | 0.55*** | | FF3 | 0.17** | -0.2** | -0.21 | 0.38** |
| | | (3.27) | (-2.12) | (-1.18) | (3.06) | | | (2.27) | (-2.12) | (-1.6) | (2.5) |
| | FF5 | 0.4** | -0.18* | -0.2 | 0.6*** | | FF5 | 0.15* | -0.22** | -0.24* | 0.38** |
| | | (4) | (-1.66) | (-1.13) | (3.38) | | | (1.92) | (-2.24) | (-1.79) | (2.47) |

**Table 4 Long-Term Return of Portfolio Sorts**

This table presents the cumulative monthly CAPM alphas for monthly calendar-time equal-weighted and value-weighted single sorts portfolios. The sample period is 2010-2023. *AI-sentiment* is the sentiment score of earnings call transcripts computed by gpt-3.5-turbo-16k model. *HD-sentiment* is the sentiment score of earnings call transcripts computed by using Loughran-McDonald (LM) dictionary. *SUE* is the difference between reported earnings and expected earnings, assuming earnings follow a seasonal random walk trend. T-statistics are reported in parentheses. *, **, and *** represent significance at the 10%, 5% and 1% levels, respectively.

| | | (1)<br>[m+1 m+2] | (2)<br>[m+1 m+3] | (3)<br>[m+1 m+4] | (4)<br>[m+1 m+5] | (5)<br>[m+1 m+6] | (6)<br>[m+1 m+9] |
|---|---|---|---|---|---|---|---|
| Equal Weighted | AI-sentiment | 1.68*** | 2.31*** | 2.94*** | 3.61*** | 4.06*** | 4.33*** |
| | | (3.55) | (3.85) | (4.18) | (4.55) | (4.64) | (3.93) |
| | HD-sentiment | 0.36 | 0.51 | 0.55 | 0.54 | 0.37 | 0.50 |
| | | (1.38) | (1.60) | (1.57) | (1.43) | (0.86) | (0.78) |
| | SUE | 2.06*** | 2.90*** | 3.14*** | 3.30*** | 3.73*** | 3.68*** |
| | | (6.01) | (6.69) | (5.53) | (5.10) | (4.35) | (3.61) |
| Value Weighted (80p) | AI-sentiment | 0.99*** | 1.32*** | 2.04*** | 2.80*** | 3.31*** | 4.16*** |
| | | (2.85) | (2.92) | (3.91) | (4.80) | (5.01) | (4.84) |
| | HD-sentiment | 0.19 | 0.16 | 0.17 | 0.11 | 0.05 | 0.42 |
| | | (0.74) | (0.53) | (0.49) | (0.28) | (0.11) | (0.64) |
| | SUE | 0.46 | 0.65 | 0.59 | 0.37 | 0.36 | -0.34 |
| | | (1.05) | (1.20) | (0.91) | (0.49) | (0.40) | (-0.36) |

**Table 5 High-Frequency Return Predictability of *AI-sentiment* and *HD-sentiment***

This table presents the high-frequency return predictability of *AI-sentiment* and *HD-sentiment* using full sample (Panel A) and post-final knowledge sample (Panel B). *AI-sentiment* is the sentiment score of earnings call transcripts computed by gpt-3.5-turbo-16k model. *HD-sentiment* is the sentiment score of earnings call transcripts computed by using Loughran-McDonald (LM) dictionary. *SUE* is the difference between reported earnings and expected earnings, assuming earnings follow a seasonal random walk trend. *Beta* is the market beta with respect to the CRSP equal-weighted return index. *BM* is the book value of equity over the market value of equity. *MVE* is the total market value of common equity. *Mom12m* is 12-month momentum computed as the cumulative return from month t-12 to t-2. *AI-sentiment*, *HD-sentiment* and *SUE* are standardized for regression. We include year-quarter fixed effect. Standard errors are two-way clustered by firm and year-quarter. T-statistics are reported in parentheses. *, **, and *** represent significance at the 10%, 5% and 1% levels, respectively.

*Panel A: Full Sample*

| | (1)<br>[d-1 Close, d Call Start] | (2)<br>[d Call Start, d Call End] | (3)<br>[d Call End, d Close] | (4)=(2)+(3)<br>[d Call Start, d Close] |
|---|---|---|---|---|
| AI-sentiment$_{i,d}$ | 1.52*** | 0.22*** | 0.37*** | 0.60*** |
| | (13.95) | (7.08) | (5.04) | (6.84) |
| HD-sentiment$_{i,d}$ | 0.52*** | 0.060** | 0.26*** | 0.32*** |
| | (6.34) | (2.47) | (5.25) | (5.62) |
| SUE$_{i,d}$ | 0.32 | -0.030 | -0.19*** | -0.22*** |
| | (1.67) | (-0.97) | (-3.23) | (-3.28) |
| Beta | 0.43 | 0.017 | -0.081 | -0.073 |
| | (0.97) | (0.44) | (-0.71) | (-0.60) |
| BM | 0.18 | 0.011 | 0.086 | 0.096 |
| | (1.58) | (0.38) | (1.37) | (1.30) |
| MVE | -0.23*** | -0.026** | 0.13*** | 0.10** |
| | (-2.84) | (-2.34) | (3.67) | (2.43) |
| Mom12m | -0.13 | -0.089*** | -0.036 | -0.13** |
| | (-1.60) | (-3.38) | (-0.54) | (-2.47) |
| Year-Quarter FE | Yes | Yes | Yes | Yes |
| Cluster | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr |
| Sample Period | 10-23 | 10-23 | 10-23 | 10-23 |
| N | 31238 | 31238 | 31238 | 31238 |
| r2 | 0.017 | 0.0063 | 0.013 | 0.014 |

*Panel B: Post-Final Knowledge Sample*

| | (1)<br>[d-1 Close, d Call Start] | (2)<br>[d Call Start, d Call End] | (3)<br>[d Call End, d Close] | (4)=(2)+(3)<br>[d Call Start, d Close] |
|---|---|---|---|---|
| AI-sentiment$_{i,d}$ | 1.80*** | 0.21*** | 0.30* | 0.51** |
| | (8.35) | (5.01) | (1.91) | (3.20) |
| HD-sentiment$_{i,d}$ | 0.55*** | 0.073 | 0.33** | 0.41*** |
| | (3.67) | (1.45) | (3.10) | (4.03) |
| SUE$_{i,d}$ | 0.60*** | -0.032 | -0.23* | -0.26* |
| | (3.71) | (-0.53) | (-1.87) | (-1.92) |
| Beta | -0.24* | 0.14** | -0.30 | -0.15 |
| | (-2.12) | (2.52) | (-1.25) | (-0.60) |
| BM | 0.22* | -0.0083 | 0.12 | 0.11 |
| | (1.95) | (-0.20) | (1.09) | (0.97) |
| MVE | -0.22*** | -0.0044 | 0.31*** | 0.31*** |
| | (-5.41) | (-0.26) | (7.48) | (6.18) |
| Mom12m | 0.099 | -0.12 | -0.026 | -0.15 |
| | (0.78) | (-1.61) | (-0.20) | (-1.07) |
| Year-Quarter FE | Yes | Yes | Yes | Yes |
| Cluster | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr |
| Sample Period | 21Oct-23Dec | 21Oct-23Dec | 21Oct-23Dec | 21Oct-23Dec |
| N | 8624 | 8624 | 8624 | 8624 |
| r2 | 0.083 | 0.0054 | 0.019 | 0.019 |

## Table 6 Short Selling and *AI-sentiment* in Pre-Democratization Period

The table presents short selling around earnings calls in alignment with *AI-sentiment*, *HD-sentiment*, and *SUE* during the pre-democratization period in 2010-2022. Short selling (multiplied by 100000) is abnormal short-seller holding calculated in Section 1.3. *AI-sentiment* is the sentiment score of earnings call transcripts computed by gpt-3.5-turbo-16k model. *HD-sentiment* is the sentiment score of earnings call transcripts computed by using Loughran-McDonald (LM) dictionary. *SUE* is the difference between reported earnings and expected earnings, assuming earnings follow a seasonal random walk trend. Control variables include beta (*Beta*), book-to-market ratio (*BM*), market value (*MVE*), and momentum (*Mom12m*). We include year-quarter fixed effect. Standard errors are two-way clustered by firm and year-quarter. T-statistics are reported in parentheses. *, **, and *** represent significance at the 10%, 5% and 1% levels, respectively.

| | Short Selling$_{i,d}$ | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | [d] | [d+1:d+5] | [d+1:d+10] | [d+1:d+21] |
| AI-sentiment$_{i,d}$ | -706.9** | -3550.9** | -7263.8** | -15370.6** |
| | (-2.02) | (-1.02) | (-1.80) | (-1.79) |
| HD-sentiment$_{i,d}$ | -2696.3*** | -13329.3*** | -26549.5*** | -55517.0*** |
| | (-6.24) | (-7.79) | (-6.94) | (-6.13) |
| SUE$_{i,d}$ | 6581.2*** | 32497.9*** | 64702.2*** | 134756.8*** |
| | (5.86) | (6.31) | (6.23) | (4.37) |
| Control | Yes | Yes | Yes | Yes |
| Year-Quarter FE | Yes | Yes | Yes | Yes |
| Cluster | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr |
| Sample Period | 10-22 | 10-22 | 10-22 | 10-22 |
| N | 72895 | 72902 | 72896 | 72890 |
| r2 | 0.15 | 0.15 | 0.15 | 0.15 |

## Table 7 Retail Trading and *AI-sentiment* in Pre-Democratization Period

The table presents retail trading around earnings calls in alignment with *AI-sentiment*, *HD-sentiment*, and *SUE* during the pre-democratization period in 2010-2022. Retail trading (multiplied by 100000) is abnormal retail holding calculated in Section 1.4. *AI-sentiment* is the sentiment score of earnings call transcripts computed by gpt-3.5-turbo-16k model. *HD-sentiment* is the sentiment score of earnings call transcripts computed by using Loughran-McDonald (LM) dictionary. *SUE* is the difference between reported earnings and expected earnings, assuming earnings follow a seasonal random walk trend. Control variables include beta (*Beta*), book-to-market ratio (*BM*), market value (*MVE*), and momentum (*Mom12m*). We include year-quarter fixed effect. Standard errors are two-way clustered by firm and year-quarter. T-statistics are reported in parentheses. *, **, and *** represent significance at the 10%, 5% and 1% levels, respectively.

| | Retail Trading$_{i,d}$ | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | [d] | [d+1:d+5] | [d+1:d+10] | [d+1:d+21] |
| AI-sentiment$_{i,d}$ | 1.72 | 16.2 | 32.0 | 77.4 |
| | (0.40) | (0.76) | (0.77) | (0.91) |
| HD-sentiment$_{i,d}$ | -16.5*** | -81.1*** | -159.8*** | -336.1*** |
| | (-2.86) | (-2.86) | (-2.80) | (-2.83) |
| SUE$_{i,d}$ | 66.4*** | 325.6*** | 643.7*** | 1317.6*** |
| | (4.47) | (4.44) | (4.31) | (4.34) |
| Control | Yes | Yes | Yes | Yes |
| Year-Quarter FE | Yes | Yes | Yes | Yes |
| Cluster | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr |
| Sample Period | 10-22 | 10-22 | 10-22 | 10-22 |
| N | 75304 | 75302 | 75301 | 75299 |
| r2 | 0.031 | 0.030 | 0.030 | 0.029 |

## Table 8 AI Democratization and Trading

This table presents how short-sellers-AI alignment and retail-AI alignment change around AI democratization. Panel A (B) reports the alignment of short selling (retail trading) with *AI-sentiment* before and after AI democratization, using the full sample. Panel C (D) reports the alignment of short selling (retail trading) with *AI-sentiment* before and after AI democratization, using a short sample in 2022-2023. Short selling (multiplied by 100000) is abnormal short-seller holding calculated in Section 1.3. Retail trading (multiplied by 100000) is abnormal retail holding calculated in Section 1.4. *After* is a dummy variable equals one for periods subsequent to January 1, 2023. *AI-sentiment* is the sentiment score of earnings call transcripts computed by gpt-3.5-turbo-16k model. Control variables include *HD-sentiment, SUE, Beta, BM, MVE, and Mom12m. HD-sentiment* is the sentiment score of earnings call transcripts computed by using Loughran-McDonald (LM) dictionary. *SUE* is the difference between reported earnings and expected earnings, assuming earnings follow a seasonal random walk trend. *Beta* is the market beta with respect to the CRSP equal-weighted return index. *BM* is the book value of equity over the market value of equity. *MVE* is the total market value of common equity. *Mom12m* is 12-month momentum computed as the cumulative return from month t-12 to t-2. *AI-sentiment, HD-sentiment* and *SUE* are standardized for regression. We include year-quarter fixed effect. Standard errors are two-way clustered by firm and year-quarter. T-statistics are reported in parentheses. *, **, and *** represent significance at the 10%, 5% and 1% levels, respectively.

*Panel A: Short Selling Change in Full Sample*

| | Short Selling$_{i,d}$ | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | [d] | [d+1:d+5] | [d+1:d+10] | [d+1:d+21] |
| AI-sentiment$_{i,d}$ | -722.7* | -3628.2* | -7417.4** | -15588.5** |
| | (-1.98) | (-2.00) | (-2.05) | (-2.06) |
| AI-sentiment$_{i,d}$*After$_{d}$ | -820.0 | -4330.2 | -8017.0 | -17350.6 |
| | (-1.35) | (-1.33) | (-1.22) | (-1.23) |
| Control | Yes | Yes | Yes | Yes |
| Year-Quarter FE | Yes | Yes | Yes | Yes |
| Cluster | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr |
| Sample Period | 10-23 | 10-23 | 10-23 | 10-23 |
| N | 82228 | 82224 | 82195 | 82060 |
| r2 | 0.16 | 0.16 | 0.16 | 0.16 |
| Change in AI-sentiment (AI-sentiment*After/AI-sentiment) | 1.1 | 1.2 | 1.1 | 1.1 |

*Panel B: Retail Trading Change in Full Sample*

| | Retail Trading$_{i,d}$ | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | [d] | [d+1:d+5] | [d+1:d+10] | [d+1:d+21] |
| AI-sentiment$_{i,d}$ | -1.08 | 2.66 | 4.05 | 16.2 |
| | (-0.24) | (0.12) | (0.09) | (0.18) |
| AI-sentiment$_{i,d}$*After$_{d}$ | 27.4*** | 137.1*** | 264.8*** | 546.9** |
| | (2.77) | (2.82) | (2.72) | (2.63) |
| Control | Yes | Yes | Yes | Yes |
| Year-Quarter FE | Yes | Yes | Yes | Yes |
| Cluster | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr |
| Sample Period | 10-23 | 10-23 | 10-23 | 10-23 |
| N | 85015 | 85006 | 84983 | 84846 |
| r2 | 0.024 | 0.023 | 0.023 | 0.023 |
| Change in AI-sentiment (AI-sentiment*After/AI-sentiment) | -25.4 | 51.5 | 65.4 | 33.8 |

*Panel C: Short Selling Change in 2022-2023*

| | Short Selling$_{i,d}$ | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | [d] | [d+1:d+5] | [d+1:d+10] | [d+1:d+21] |
| AI-sentiment$_{i,d}$ | -2882.8*** | -14417.5*** | -29268.4*** | -61386.8*** |
| | (-3.62) | (-3.62) | (-3.76) | (-3.93) |
| AI-sentiment$_{i,d}$*After$_{d}$ | 2141.7* | 10406.7* | 21705.3* | 45546.3* |
| | (2.23) | (2.06) | (2.14) | (2.13) |
| Control | Yes | Yes | Yes | Yes |
| Year-Quarter FE | Yes | Yes | Yes | Yes |
| Cluster | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr |
| Sample Period | 22-23 | 22-23 | 22-23 | 22-23 |
| N | 17368 | 17357 | 17334 | 17203 |
| r2 | 0.21 | 0.21 | 0.21 | 0.21 |

*Panel D: Retail Trading Change in 2022-2023*

| | Retail Trading$_{i,d}$ | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | [d] | [d+1:d+5] | [d+1:d+10] | [d+1:d+21] |
| AI-sentiment$_{i,d}$ | -27.1** | -110.6* | -217.5* | -448.3* |
| | (-2.53) | (-1.96) | (-1.97) | (-1.94) |
| AI-sentiment$_{i,d}$*After$_{d}$ | 39.9** | 183.4** | 342.1* | 683.6* |
| | (2.66) | (2.37) | (2.21) | (2.04) |
| Control | Yes | Yes | Yes | Yes |
| Year-Quarter FE | Yes | Yes | Yes | Yes |
| Cluster | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr |
| Sample Period | 22-23 | 22-23 | 22-23 | 22-23 |
| N | 17961 | 17954 | 17932 | 17797 |
| r2 | 0.0072 | 0.0084 | 0.0084 | 0.0081 |

## Table 9 AI Democratization and Trading: DiD

This table presents the relation between AI democratization and retail trading activities relative to short selling in a DiD setting. The dataset includes both retail trading (treatment group) and short selling (control group) for each stock in each quarter. Standardized trading includes both retail trading and short selling. We reverse the sign of *ShortSelling* and standardize both *ShortSelling* and *RetailTrading*. $Treat^{IsRetail}$ is a dummy variable that equals 1 if the stock in a given quarter is traded by retail investors and 0 if traded by short sellers. *After* is a dummy variable equals one for periods subsequent to January 1, 2023. *AI-sentiment* is the sentiment score of earnings call transcripts computed by gpt-3.5-turbo-16k model. Control variables include *HD-sentiment, SUE, Beta, BM, MVE, and Mom12m. HD-sentiment* is the sentiment score of earnings call transcripts computed by using Loughran-McDonald (LM) dictionary. *SUE* is the difference between reported earnings and expected earnings, assuming earnings follow a seasonal random walk trend. *Beta* is the market beta with respect to the CRSP equal-weighted return index. *BM* is the book value of equity over the market value of equity. *MVE* is the total market value of common equity. *Mom12m* is 12-month momentum computed as the cumulative return from month t-12 to t-2. *AI-sentiment*, *HD-sentiment* and *SUE* are standardized for regression. We include firm-trader and year-quarter-trader fixed effect. Standard errors are two-way clustered by firm and year-quarter. T-statistics are reported in parentheses. *, **, and *** represent significance at the 10%, 5% and 1% levels, respectively.

| | Standardized $Trading_{i,g,d}$ | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | [d] | [d+1:d+5] | [d+1:d+10] | [d+1:d+21] |
| $AI\text{-}sentiment_{i,d}$ | 0.017 | 0.018 | 0.020* | 0.022* |
| | (1.64) | (1.77) | (1.95) | (2.18) |
| $AI\text{-}sentiment_{i,d}*Treat^{IsRetail}_{g}$ | -0.044* | -0.042* | -0.043* | -0.045* |
| | (-2.19) | (-2.08) | (-2.20) | (-2.30) |
| $AI\text{-}sentiment_{i,d}*After_{d}$ | -0.035* | -0.036** | -0.039** | -0.039** |
| | (-2.36) | (-2.42) | (-2.54) | (-2.63) |
| $AI\text{-}sentiment_{i,d}*Treat^{IsRetail}_{g}*After_{d}$ | 0.090** | 0.090** | 0.089** | 0.088** |
| | (2.68) | (2.59) | (2.61) | (2.69) |
| Control | Yes | Yes | Yes | Yes |
| Firm-Trader FE | Yes | Yes | Yes | Yes |
| Year-Quarter-Trader FE | Yes | Yes | Yes | Yes |
| Cluster | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr |
| Sample Period | 22-23 | 22-23 | 22-23 | 22-23 |
| N | 35188 | 35170 | 35125 | 34854 |
| r2 | 0.60 | 0.61 | 0.61 | 0.61 |

**Table 10 ChatGPT Outages, AI Democratization, and Trading: DiD**

This table presents the relation between AI democratization and retail trading activities relative to short selling when there are ChatGPT outages in a DiD setting. The dataset includes both retail trading (treatment group) and short selling (control group) for each stock in each quarter. Standardized trading includes both retail trading and short selling. We reverse the sign of *ShortSelling* and standardize both *ShortSelling* and *RetailTrading*. *After&Outage* is a continuous variable representing the duration of outages occurring on the same day as an earnings call. We focus on major outages caused by ChatGPT web errors and require the outage start time to be after the end time of the earnings call. For multiple outages occurred within the day, we sum the total duration. $Treat^{IsRetail}$ is a dummy variable that equals 1 if the stock in a given quarter is traded by retail investors and 0 if traded by short sellers. *After* is a dummy variable equals one for periods subsequent to January 1, 2023. *AI-sentiment* is the sentiment score of earnings call transcripts computed by gpt-3.5-turbo-16k model. Control variables include *HD-sentiment, SUE, Beta, BM, MVE, and Mom12m. HD-sentiment* is the sentiment score of earnings call transcripts computed by using Loughran-McDonald (LM) dictionary. *SUE* is the difference between reported earnings and expected earnings, assuming earnings follow a seasonal random walk trend. *Beta* is the market beta with respect to the CRSP equal-weighted return index. *BM* is the book value of equity over the market value of equity. *MVE* is the total market value of common equity. *Mom12m* is 12-month momentum computed as the cumulative return from month t-12 to t-2. *AI-sentiment*, *HD-sentiment* and *SUE* are standardized for regression. We include firm-trader and year-quarter-trader fixed effect. Standard errors are two-way clustered by firm and year-quarter. T-statistics are reported in parentheses. *, **, and *** represent significance at the 10%, 5% and 1% levels, respectively.

| | Standardized $Trading_{i,g,d}$ | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | [d] | [d+1:d+5] | [d+1:d+10] | [d+1:d+21] |
| $AI\text{-}sentiment_{i,d}$ | 0.017 | 0.018 | 0.020* | 0.022* |
| | (1.63) | (1.77) | (1.95) | (2.17) |
| $After\&Outage_{i,d}$ | -0.0099 | -0.0082 | -0.0084 | -0.016 |
| | (-0.64) | (-0.57) | (-0.59) | (-1.20) |
| $AI\text{-}sentiment_{i,d}*Treat^{IsRetail}{}_g$ | -0.044* | -0.041* | -0.043* | -0.045* |
| | (-2.19) | (-2.08) | (-2.19) | (-2.30) |
| $AI\text{-}sentiment_{i,d}*After_d$ | -0.036** | -0.037** | -0.039** | -0.039** |
| | (-2.37) | (-2.43) | (-2.55) | (-2.64) |
| $AI\text{-}sentiment_{i,d}*After\&Outage_{i,d}$ | 0.017* | 0.017* | 0.016* | 0.022** |
| | (2.07) | (2.14) | (2.12) | (2.43) |
| $Treat^{IsRetail}{}_g*After\&Outage_{i,d}$ | 0.10** | 0.095** | 0.096** | 0.078** |
| | (3.25) | (3.12) | (3.03) | (3.23) |
| $AI\text{-}sentiment_{i,d}*Treat^{IsRetail}{}_g*After_d$ | 0.091** | 0.090** | 0.090** | 0.088** |
| | (2.69) | (2.60) | (2.62) | (2.70) |
| $AI\text{-}sentiment_{i,d}*Treat^{IsRetail}{}_g*After\&Outage_{i,d}$ | -0.075*** | -0.075** | -0.075** | -0.064*** |
| | (-3.74) | (-3.27) | (-2.93) | (-3.86) |
| Control | Yes | Yes | Yes | Yes |
| Firm-Trader FE | Yes | Yes | Yes | Yes |
| Year-Quarter-Trader FE | Yes | Yes | Yes | Yes |
| Cluster | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr |
| Sample Period | 22-23 | 22-23 | 22-23 | 22-23 |
| N | 35188 | 35170 | 35125 | 34854 |
| r2 | 0.60 | 0.61 | 0.61 | 0.61 |

## Table 11 Information Asymmetry and AI Democratization: PSM-DiD

This table presents the relation between AI democratization and information asymmetry in a PSM-DiD setting. *BidAskSpread* (multiplied by 100000) is the difference between the daily bid and ask price scaled by the average of bid and ask price. $Treat^{AI}$ is a dummy variable that equals 1 if the stock's retail-AI alignment is above the median and 0 otherwise. Retail-AI alignment is the coefficients of *AI-sentiment*Treat*$^{IsRetail}$**After*, obtained by running regressions in Table 9 on a stock-by-stock basis. The treatment and control groups are balanced using propensity score matching. *After* is a dummy variable equals one for periods subsequent to January 1, 2023. Control variables include *AI-sentiment*, *HD-sentiment*, *SUE*, *Beta*, *BM*, *MVE*, *and Mom12m. AI-sentiment* is the sentiment score of earnings call transcripts computed by gpt-3.5-turbo-16k model. *HD-sentiment* is the sentiment score of earnings call transcripts computed by using Loughran-McDonald (LM) dictionary. *SUE* is the difference between reported earnings and expected earnings, assuming earnings follow a seasonal random walk trend. *Beta* is the market beta with respect to the CRSP equal-weighted return index. *BM* is the book value of equity over the market value of equity. *MVE* is the total market value of common equity. *Mom12m* is 12-month momentum computed as the cumulative return from month t-12 to t-2. *AI-sentiment*, *HD-sentiment* and *SUE* are standardized for regression. We include year-quarter fixed effect. Standard errors are two-way clustered by firm and year-quarter. T-statistics are reported in parentheses. *, **, and *** represent significance at the 10%, 5% and 1% levels, respectively.

| | Bid-Ask Spread$_{i,d}$ | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | [d] | [d+1:d+5] | [d+1:d+10] | [d+1:d+21] |
| $Treat^{AI}_i$ | 46.1* | 16.6 | 17.4 | 15.0 |
| | (2.31) | (0.88) | (0.92) | (0.82) |
| $Treat^{AI}_i$*$After_d$ | -26.2 | -19.6* | -22.3** | -25.7*** |
| | (-1.79) | (-2.26) | (-2.57) | (-3.71) |
| Control | Yes | Yes | Yes | Yes |
| Year-Quarter FE | Yes | Yes | Yes | Yes |
| Cluster | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr |
| Sample Period | 22Jan-23Dec | 22Jan-23Dec | 22Jan-23Dec | 22Jan-23Dec |
| N | 12502 | 12437 | 12425 | 12335 |
| r2 | 0.37 | 0.46 | 0.48 | 0.49 |

**Table 12 ChatGPT Outages, AI Democratization, and Information Asymmetry: PSM-DiD**

This table presents the relation between AI democratization and information asymmetry when there are ChatGPT outages in a PSM-DiD setting. *BidAskSpread* (multiplied by 100000) is the difference between the daily bid and ask price scaled by the average of bid and ask price. *After&Outage* is a continuous variable representing the duration of outages occurring on the same day as an earnings call. We focus on major outages caused by ChatGPT web errors and require the outage start time to be after the end time of the earnings call. For multiple outages occurred within the day, we sum the total duration. $Treat^{AI}$ is a dummy variable that equals 1 if the stock's retail-AI alignment is above the median and 0 otherwise. Retail-AI alignment is the coefficients of $AI\text{-}sentiment*Treat^{IsRetail}*After$, obtained by running regressions in Table 9 on a stock-by-stock basis. The treatment and control groups are balanced using propensity score matching. *After* is a dummy variable equals one for periods subsequent to January 1, 2023. Control variables include *AI-sentiment, HD-sentiment, SUE, Beta, BM, MVE, and Mom12m. AI-sentiment* is the sentiment score of earnings call transcripts computed by gpt-3.5-turbo-16k model. *HD-sentiment* is the sentiment score of earnings call transcripts computed by using Loughran-McDonald (LM) dictionary. *SUE* is the difference between reported earnings and expected earnings, assuming earnings follow a seasonal random walk trend. *Beta* is the market beta with respect to the CRSP equal-weighted return index. *BM* is the book value of equity over the market value of equity. *MVE* is the total market value of common equity. *Mom12m* is 12-month momentum computed as the cumulative return from month t-12 to t-2. *AI-sentiment*, *HD-sentiment* and *SUE* are standardized for regression. We include year-quarter fixed effect. Standard errors are two-way clustered by firm and year-quarter. T-statistics are reported in parentheses. *, **, and *** represent significance at the 10%, 5% and 1% levels, respectively.

| | Bid-Ask $Spread_{i,d}$ | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | [d] | [d+1:d+5] | [d+1:d+10] | [d+1:d+21] |
| $After\&Outage_{i,d}$ | -87.4*** | -10.2 | -16.0 | -65.4*** |
| | (-5.32) | (-0.17) | (-0.28) | (-5.15) |
| $Treat^{AI}_{i}$ | 46.1* | 16.6 | 17.4 | 15.0 |
| | (2.31) | (0.88) | (0.92) | (0.82) |
| $Treat^{AI}_{i}*After_{d}$ | -27.1 | -19.8 | -22.4* | -26.7*** |
| | (-1.80) | (-1.73) | (-2.05) | (-3.63) |
| $Treat^{AI}_{i}*After\&Outage_{i,d}$ | 67.6*** | 5.46 | 6.44 | 49.5* |
| | (4.01) | (0.10) | (0.13) | (2.31) |
| Control | Yes | Yes | Yes | Yes |
| Year-Quarter FE | Yes | Yes | Yes | Yes |
| Cluster | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr | Firm&YrQtr |
| Sample Period | 22Jan-23Dec | 22Jan-23Dec | 22Jan-23Dec | 22Jan-23Dec |
| N | 12502 | 12437 | 12425 | 12335 |
| r2 | 0.37 | 0.46 | 0.48 | 0.49 |

**Figure 1 A Bin-Scatter Plot of AI-sentiment vs. HD-sentiment**

This figure displays a bin-scatter plot comparing *AI-sentiment* against *HD-sentiment*. *AI-sentiment* is positioned on the y-axis, while *HD-sentiment* is plotted along the x-axis. *HD-sentiment* are segmented into 10 distinct bins. For each bin, we calculate the average values of both *AI-sentiment* and *HD-sentiment*. Both *AI-sentiment* and *HD-sentiment* have been standardized prior to analysis.

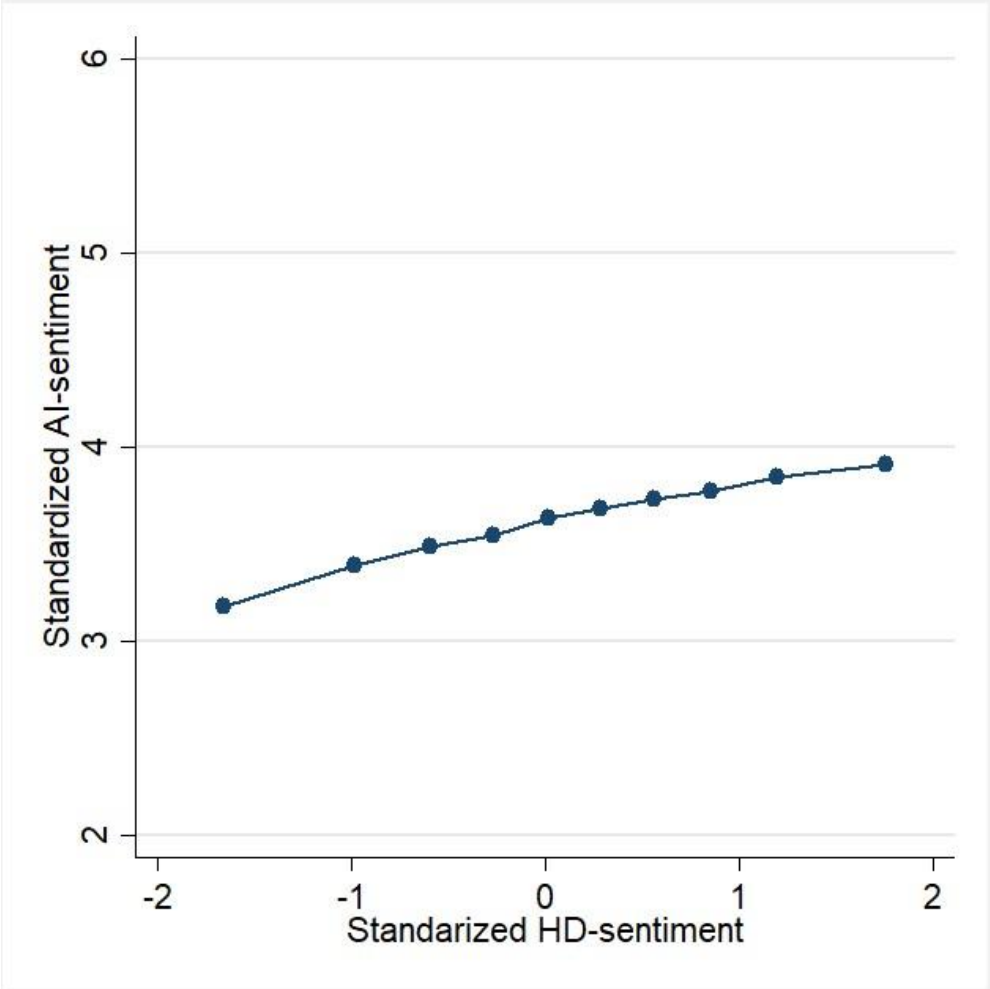


**Figure 2 Trend of Retail Investors' AI Sensitivity**

This figure plots retail-AI alignment relative to short sellers based on half-year period between 2021 and 2023 where the first half of 2021 is the reference.

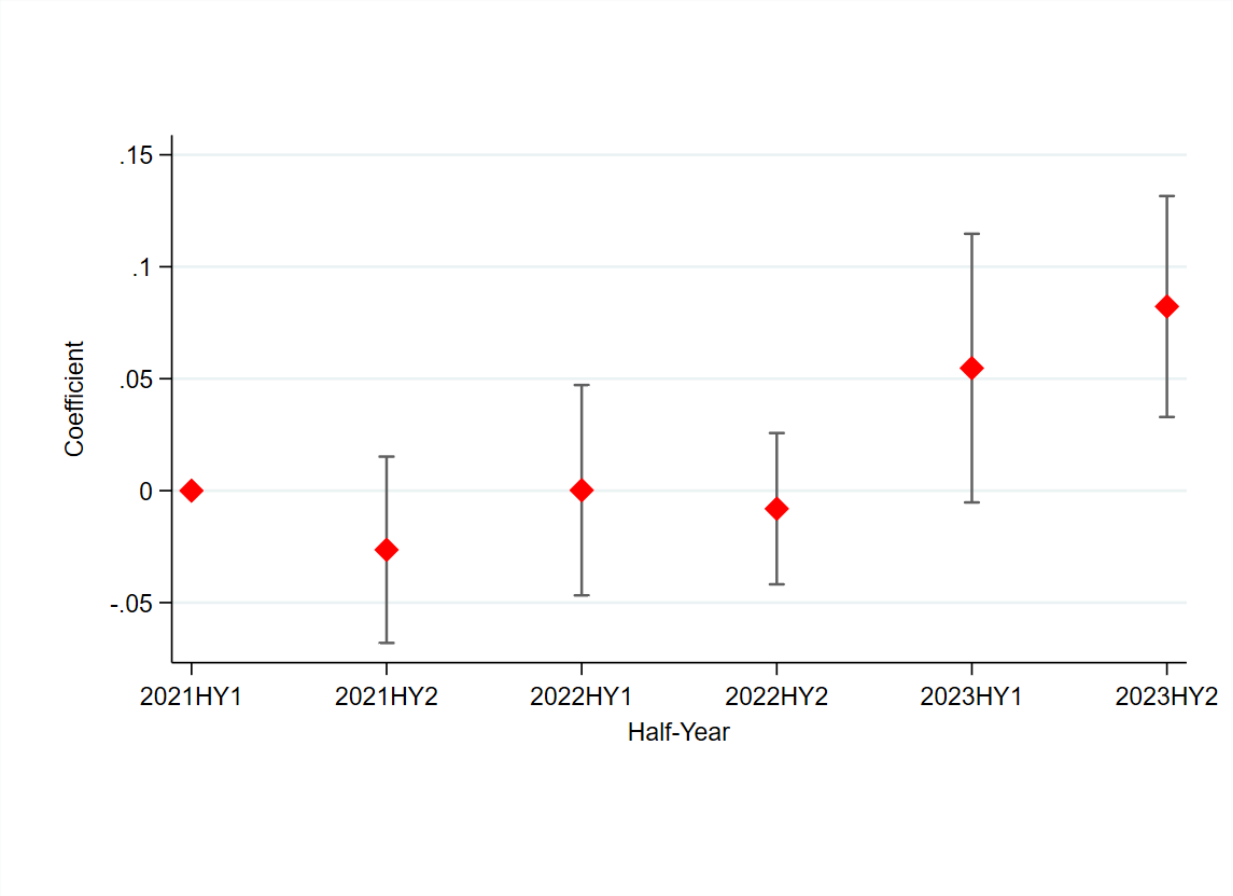


**Table IA1 Propensity Score Matching**

This table presents the summary statistics of variables before and after propensity score matching. *AI-sentiment* is the sentiment score of earnings call transcripts computed by gpt-3.5-turbo-16k model. *HD-sentiment* is the sentiment score of earnings call transcripts computed by using Loughran-McDonald (LM) dictionary. *SUE* is the difference between reported earnings and expected earnings, assuming earnings follow a seasonal random walk trend. *Beta* is the market beta with respect to the CRSP equal-weighted return index. *BM* is the book value of equity over the market value of equity. *MVE* is the total market value of common equity. *Mom12m* is 12-month momentum computed as the cumulative return from month t-12 to t-2. *AI-sentiment*, *HD-sentiment* and *SUE* are standardized. *, **, and *** represent significance at the 10%, 5% and 1% levels, respectively.

| | Unmatched | | | | Matched | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean Treated | Mean Control | Difference | t-statistic | Mean Treated | Mean Control | Difference | t-statistic |
| AI-sentiment | -0.0195 | 0.0120 | -0.0315* | -1.96 | -0.0199 | -0.0064 | -0.0135 | -0.74 |
| HD-sentiment | -0.0019 | 0.0012 | -0.0031 | -0.19 | -0.0023 | 0.0099 | -0.0123 | -0.69 |
| SUE | -0.0379 | 0.0233 | -0.0612*** | -3.81 | -0.0384 | -0.0385 | 0.0001 | 0.01 |
| Beta | 1.2654 | 1.2839 | -0.0185* | -1.92 | 1.2653 | 1.2648 | 0.0005 | 0.04 |
| BM | 0.4751 | 0.5153 | -0.0402*** | -3.18 | 0.4801 | 0.4713 | 0.0088 | 0.65 |
| MVE | 14.2620 | 14.1240 | 0.1380*** | 3.81 | 14.2630 | 14.2810 | -0.0180 | -0.45 |
| Mom12m | -0.0159 | -0.0354 | 0.0195** | 2.31 | -0.0198 | -0.0173 | -0.0025 | -0.28 |