AI (ChatGPT) Democratization, Return Predictability, and Trading Inequality

First Version: August 17, 2023; This Version: February 18, 2025

Abstract

We present the first analysis of democratized AI's (ChatGPT) role in investors' trading activities, leveraging a 19-year dataset of earnings call transcripts. We have four findings. First, we develop an AI-sentiment measure utilizing full transcript context. We find strong, long-lasting return predictability beyond six months, contrasting human-dictionary (HD) sentiment, which shows little or negative predictive power implying its inability to handle managerial biases. Second, over the prolonged period preceding ChatGPT's deployment, short selling had been aligned with AI-sentiment, while retail trading had not. Post-deployment, retail traders' alignment increases up to 23-fold, while short sellers' alignment can diminish. Consequently, stocks with higher retail-AI alignment witness significant bid-ask spread reductions. Third, exogenous ChatGPT outages notably reduce retail-AI alignment and reverse bid-ask spread improvements. Last, AI-sentiment provides investors with long-lasting return insights by leveraging text context and discerning between genuinely and excessively positive HD-sentiment. These findings highlight the critical role of AI democratization in leveling the playing field.

JEL: G10, G11, G14

Keywords: Artificial Intelligence, AI democratization, ChatGPT, Sentiment, Return Predictability, Machine Learning, Earnings Conference Calls, Retail Investors, Short Sellers, Households, Anomalies, Information Asymmetry, Trading Inequality

"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 dedicated to protecting ordinary retail investors, recognizing that household participation in financial markets is vital to far-reaching issues such as equality, wealth democratization, market liquidity and efficiency, societal risk-sharing, capital supply, and even economic growth. To this end, SEC has instituted extensive disclosure requirements for public companies, ensuring that important information is available to all investors. Chief among these is Regulation Fair Disclosure (Reg FD), which mandates companies to make crucial information—such as earnings calls—accessible to the general investing public in real time. The overarching goal of this ongoing initiative of democratizing information access is to level the playing field for all market participants. Additionally, other societal advances such as big data, digitalization, and improved formatting have further made more comprehensive and accurate information accessible to the general public.

Yet, the advent of Artificial Intelligence (AI), particularly its democratization following the successful widespread deployment of tools like ChatGPT, has transformational potential to impact the SEC's objectives. Such democratization is also expected to accelerate exponentially on a global scale with the emergence of low-cost models (e.g., DeepSeek), whose technological breakthroughs could potentially reduce the training costs of all future AI models by one or even two orders of magnitude. On the one hand, proponents of democratized AI hail it as a revolutionary technology for parsing general narratives, suggesting its potential to narrow the information gap between retail and privileged investors. However, the actual efficacy of AI-generated textual sentiment from financial text, such as earnings calls, remains uncertain. This uncertainty arises from the limits of large language models (LLMs) in processing long text,[1] the fact that LLMs are not trained to extract

[1] Modern LLMs primarily use Transformers (Vaswani et al. 2017) with self-attention to encode contexts. However,

sentiment to forecast financial market returns, earnings calls being central and unique information disclosure where human managers can manipulate and obfuscate information, and various AI issues such as hallucinations, fabrication of information, and over- or under-thinking. On the other hand, AI literacy may introduce a new layer of disparity within the investing community: if AI is truly a revolutionary tool as its developers claim, investors proficient in leveraging AI technologies stand to significantly enhance their financial standing, while those lacking such expertise risk falling further behind. Indeed, a common speculation is that sophisticated investors, such as hedge funds, are the primary adopters of AI in investment strategies. This disparity casts doubt on whether democratizing AI can level the playing field for all investors. In this study, we investigate the impact of AI on stock trading of privileged investors (the rich) and ordinary retail investors (the poor). Our aim is to elucidate AI's implications for equity within the capital market. We examine the most relevant form of democratized AI, ChatGPT, where the usage— rather than just its source code—has been made easily accessible to the general public at low costs. We utilize the large-token model of ChatGPT to extract comprehensive sentiment from earnings calls, ensuring the preservation of their full textual context. We study whether the AI-based call textual sentiment predicts future returns, how such sentiment shapes investors' trading behaviors, particularly retail ones, and what actionable insights AI sentiment could offer to ordinary investors. We focus on sentiment because it captures the overall assessment of a call. Unlike the myriad of complex financial signals grounded in sophisticated economic and financial theories, sentiment is arguably the simplest and most immediate indicator that ordinary investors can use to make their trading decisions. Our goal is to explore how this affects the information gap between rich and poor investors, and we might need to clarify that in the paper. Collectively, our results shed light on democratized AI's impact on investors' welfare.

To operationalize our empirical analysis, we employ the GPT-3.5-turbo-16k API model to extract sentiment information from earnings calls. We develop a sentiment measure, referred to

performance may degrade with increased text length due to the exponential growth of word connections (e.g., Dai et al. 2019). To combat this, OpenAI uses sparse attention mechanisms, focusing on segments of the input to reduce computational complexity (Child et al. 2019). These mechanisms often prioritize immediate neighbors and select distant tokens based on patterns, such as skipping a fixed number of tokens. Liu et al. (2023) found that model performance is highest when relevant information is at the beginning or end of the input, suggesting extended-context models may struggle to analyze longer inputs effectively. Similarly, Sun et al. (2021) found longer contexts improve prediction for only a few tokens, and Krishna et al. (2022) observed that long-context neural generation can degrade in modestly-sized Transformer models due to ineffective use of extended contexts.

as AI-sentiment. Our analysis features three key advantages. First, we process a comprehensive 19-year dataset of 280,554 call transcripts from 2005 to 2023, where the earnings call transcripts represent the longest text—averaging 7000 words per call—ever fed to ChatGPT in a single feed for direct return predictability analysis—significantly exceeding the textual length in contemporaneous studies with the same objective. Second, the 16k model we employ supports up to 16,385 tokens, far exceeding the 4,096-token limit of GPT-3.5-turbo commonly employed in existing studies. Although the model is expensive, with a total cost of $18,300, its capacity allows each transcript to be analyzed in one single feed, thereby preserving the contextual information of the entire call. Third, while GPT4 or later versions with more recent knowledge cutoff dates are utilized by several studies subsequent to ours, GPT 3.5 models provide a sufficiently long sample for out-of-sample testing, which is crucial for validating return predictability conclusions.

To assess the effectiveness of AI-sentiment, we benchmark AI-sentiment against the currently most widely-used textual sentiment measure based on bag of words from the seminal work of Loughran and McDonald (LM 2011). This human dictionary-based sentiment (and its variations), referred to as HD-sentiment, is derived by counting the positive and negative words. Earnings calls are chosen as the focal point due to their public accessibility mandated by Reg FD and their comprehensive, context-rich content. Conceptually, we would like to gauge the positivity (negativity) of managerial disclosures s(I,t) of earning calls. This signal can have three components: fundamentals v(I,t), bias u(I,t), and error e(I,t). Different methods for measuring sentiment vary in how much bias v or noise e that they introduce. GPT-based sentiment may have a different weights on all three components, resulting in different signal-to-noise ratio, than traditional HD sentiment. Indeed, prior research already suggests that the latter measure is either too noisy or prone to managerial positivity bias in information disclosures from earnings calls.

The structure of our empirical analysis is as follows. We first examine the capability of AI-sentiment to forecast returns across multiple horizons and different sample periods, offering insights into whether AI excels in extracting value-relevant information. We then analyze the trading behaviors of two distinct investor classes—short sellers, who empirical evidence suggests are the most informed sophisticated traders (Chen, Da, and Huang 2019), and retail investors, a.k.a. ordinary investors. We investigate how their trades align with AI-sentiment before and after AI democratization, which offers insights into the shifting dynamics of utilizing information in the age of AI democratization, and its implications for investor equality. Finally, we shed light on what unique investment insights AI-sentiment could offer ordinary investors.

In our first set of tests, we examine the return predictability of AI- and HD-sentiment. In our regressions, we also control the return predictability of various other anomaly variables known to predict returns. We have two key sets of results. First, we find the return predictability of HD-sentiment is weak in both magnitude and horizon for the full sample, with its effect notably disappearing after the first three days post-earnings calls. In contrast, AI-sentiment demonstrates strong return predictability, in the right direction, after controlling for HD-sentiment. The magnitude of AI-sentiment's return predictability is remarkably long-lasting, increasing as the return horizon extends and remaining significant for at least six months after earnings calls. This contrasting pattern between AI-sentiment and HD-sentiment is also evident in the period of October 2021 to December 2023, a period free of look-ahead bias. Remarkably, in this shorter subsample period, AI-sentiment can still significantly predict returns up to half a year ahead, despite the significantly reduced number of observations. In sharp contrast, HD-sentiment displays no short-term return predictability and even significantly predicts returns in the wrong direction over the long term. This is in line with the market-level pattern in Jiang, Lee, Martin, and Zhou (2019), which is consistent with managers on average being overly optimistic on calls. Overall, AI-sentiment—but not HD-sentiment—exhibits long-run return predictability long after earnings calls, suggesting substantial actionable investment insights that democratized AI could provide to ordinary investors, even if these investors are unable to execute trades quickly.

Second, to strengthen the point that the AI insights are actionable, especially for ordinary investors, we employ a calendar time portfolio approach. At the end of every month, we sort stocks based on AI-sentiment, HD-sentiment, and SUE according to the most recent available earnings calls and earnings announcements. Our results show that long-short portfolios based on AI-sentiment generate an alpha of around 1% per month for equal-weighted portfolios and 60 bps for (capped) value-weighted portfolios, with cumulative alphas of 4% and 3%, respectively, over six months. In a double-sort test, after controlling for the return predictability of HD-sentiment and SUE, long-short AI-sentiment portfolios continue generating strong alphas, supporting AI's capability to unearth information not captured by earnings surprises and other anomaly variables.

In our second set of results, we assess whether and to what extent short selling and retail trading are aligned with AI-sentiment. We first examine the thirteen years preceding the widespread deployment of ChatGPT on January 1, 2023—a period we refer to as the pre-democratization phase of AI—when democratization of access to earnings call information was underway, but AI tools like ChatGPT were not yet widely available.[2] Our findings reveal a pronounced tendency among short sellers to align their trading with AI-sentiment immediately on the earnings call day and over the subsequent ten days. Conversely, retail investors show no significant reaction to AI-sentiment. Given the limited accessibility of any AI models to the general public during the pre-democratization period, the results suggest that short sellers may have utilized alternative information sources or even proprietary AI/ML models for earnings call analysis—well before the democratization of AI. Their trading decisions, which correlate with insights similar to those generated by ChatGPT, reflect the anticipated behavior of short sellers as archetypal arbitrageurs with privileged access to technology and information. In contrast, retail investors were unable to derive such insights, suggesting that prior to the advent of AI democratization, the democratization of information access through Reg FD, although implemented years earlier, may have fallen short of achieving regulators' goal of leveling the playing field.

We then analyze shifts in short selling and retail trading surrounding the widespread deployment of ChatGPT around January 1, 2023. Post-deployment, retail investors' alignment with AI-sentiment within the ten days starting from the earnings call increases up to 23-fold, a strong conviction of a paradigm shift. In contrast, short sellers either do not significantly change or even weaken their alignment with AI-sentiment, suggesting that they no longer enjoy a clear information advantage from leveraging these AI insights and may therefore choose to scale back their reliance on such signals.[3] We then employ a difference-in-differences (DiD) framework to address various firm, firm-trader, or time-trader related endogeneity concerns. By comparing the difference between retail and short trading responses to AI within the same firm and controlling for firm-trader or time-trader fixed effects, we demonstrate that retail-AI alignment increases significantly relative to short selling after AI democratization, suggesting a robust shift in retail trading behavior toward

[2]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.

[3]This finding also alleviates the look-ahead bias concern as such bias implies all signals to future information are included. Short sellers would not scale back from trading on all signals if they still have advantages on some categories of information. They may increase the strength of trading on the signals that they have advantage (e.g., Dessaint, Foucault, and Frésard 2024).

greater alignment with AI-sentiment post-democratization.

We also use ChatGPT outages as exogenous shocks to identify changes in its availability for retail investors. If AI democratization aids retail investors, unexpected outages would diminish the post-democratization increase in retail investors' alignment with AI. Our findings confirm this conjecture. This result not only mitigates endogeneity concerns regarding the relationship between AI democratization and trading but also supports the notion that retail trading is indeed linked to the usage of AI. This linkage between retail investors and AI is supported by recent survey evidence where retail investors report that AI significantly improve their ability to process complex financial information efficiently (Blankespoor, Croom, and Grant 2024), as well as by the observations that AI-based investment recommendations are increasingly sold or discussed on social media.

Finally, if the democratization of AI enhances retail investors' ability to process information, it may reduce information asymmetry and improve market liquidity. To investigate this, 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 that experience greater retail-AI alignment following the democratization of AI. However, this effect is mitigated during ChatGPT outages, reinforcing the notion that the use of ChatGPT is likely the underlying cause of the reduced information asymmetry.

In our third set of results, we examine what unique insights AI-sentiment could provide to investors by benchmarking it against state-of-the-art technology and metrics. Our findings are twofold. First, we find that the return predictability of sentiment extracted by ChatGPT is significantly stronger and longer-lasting compared to FinBERT—a premier LLM used by finance professionals before the advent of ChatGPT. Unlike ChatGPT, FinBERT fails to predict return out-of-sample. Given that FinBERT is limited to understanding the context of short-text comparable to only a single sentence, our findings suggest that ChatGPT's long-run return predictability stems from its ability to leverage long textual context, thereby avoiding the pitfalls of drawing incomplete inferences from isolated sentences.

Second, we examine the differences and connections between AI-sentiment and HD-sentiment in capturing return-predictive information. Our analysis reveals several key insights. Firstly, when earnings calls contain extreme positive (or negative) language according to HD-sentiment, AI-sentiment tends to be less extreme. This indicates a potential discrepancy in the intensity of sentiment captured by the two measures. Next, we find that the return predictability of both sentiment measures is driven by two scenarios. In the shorter-term scenario, when both HD-sentiment and AI-sentiment are extremely positive, cumulative abnormal returns (CAR) immediately increase after the call and plateau at the two-month horizon. This suggests that when both HD and AI confirm the positive news, the information is quickly assimilated into prices. In the loner-term scenario, when HD-sentiment is extremely positive but AI-sentiment is extremely negative, CAR aligns with AI-sentiment, yielding large negative returns over the long term. This suggests that the market takes considerably longer to recognize and price in such negative information. ChatGPT appears less influenced by frequent extremely positive language in earnings calls, which corporate executives may strategically manipulate, knowing that machines are listening (Cao, Jiang, Yang, and Zhang 2023). These negative long-run returns primarily drive AI's long-term return predictability, indicating that ChatGPT's superiority for the long run stems from its ability to identify cases where excessively positive language in earnings calls conceals underlying bad news.

Overall, our findings suggest that democratized AI, such as ChatGPT, effectively extracts sentiment information from long and complex financial texts, such as earnings calls, uncovering long-term investment insights that are actionable for ordinary investors. The source of these insights lies in the broader context provided by lengthy earnings call transcripts. While informed traders had long harnessed such insights prior to AI democratization, retail investors had not. As a result, the information gap between the few privileged investors and ordinary investors may worsen, reinforcing the notion that "the rich get richer" which challenges the objective behind years of efforts to democratize information access. However, AI democratization may allow retail investors to effectively benefit from broader access to long and complex information. Therefore, regulatory pushes for democratization of information access and societal innovations of big data and improved or digitalized information reporting urgently necessitate the democratization of AI—the revolutionary tools to process information, which holds unprecedented potential to shift the balance between wealthy and ordinary investors. Our results reveal the financial market consequences that illuminate the debate on whether, and how AI should be democratized and made accessible to all—i.e., "Open AI."[4] Our findings underscore the necessity of "open" the usage (at low costs), rather than

[4]A central debate in AI is whether it should be open to all or reserved for a select few. OpenAI, initially a non-profit, has shifted toward a more closed, profit-oriented model, drawing criticism from figures like Elon Musk for deviating from its altruistic mission. This shift underscores varying attitudes toward open AI. Another key issue

just the source codes, of AI to enable a more inclusive pursuit of financial wisdom.

Literature and Contribution

We make three contributions. First, our study provides valuable insights from three aspects for financial market regulators who are particularly concerned about the welfare of disadvantaged market participants. First, our findings suggest that years of regulatory efforts to democratize information access, such as earnings calls, are insufficient to bridge the information gap between ordinary and privileged investors. Privileged investors like short sellers are well-positioned to correctly trade on complex financial information, while ordinary investors often lack the resources and tools necessary to process and act on such information effectively. However, by lowering information processing costs, AI and its democratization show great promise in addressing this gap. Our study alleviates concerns that AI tools like ChatGPT—widely regarded as signaling the start of a new industrial revolution—might benefit sophisticated investors at the expense of ordinary investors. Second, other societal advancements that provide more comprehensive and accurate information to investors may similarly allow privileged investors to gain further informational advantage, given their superior technology and resources to process information. Examples of such advancements include big data, enhanced public information access, and improved or digitalized information reporting format (e.g., images, audio, XBRL, Edgar). Indeed, as demonstrated by Zhu (2019), real-time data, like satellite imagery and foot traffic, increases price informativeness, especially for firms targeted by sophisticated investors. In contrast, our findings suggest that AI and its democratization present a unique opportunity to benefit retail investors, both directly and indirectly. Directly, providing low-cost, accessible AI tools for information processing can help level the playing field, enabling retail investors to analyze complex financial data, such as earnings calls. Indirectly, regulators could require firms to include AI-generated sentiment summaries in disclosures, while news vendors and brokers could incorporate such information into their services, as some have already begun to do. Additionally, private firms could develop AI-driven sentiment-linked index products and ETFs. Moreover, transformative cost-reduction technologies—such as those from DeepSeek—may signal an inflection point at which small firms can leverage inexpensive AI models to offer low-cost investment recom-

is the difference between making AI accessible to the general public and fully open-sourcing AGI. OpenAI's chief scientist Ilya Sutskever argued for reduced openness as AI development progresses to ensure everyone benefits, even if the science isn't shared. In contrast, models like XAi's Grok and Meta's Llama 2 remain more openly accessible.

mendations to retail investors. Third, for firm-disclosed textual information such as earnings calls, privileged investors may already extract nearly the maximum available insights, as such calls have inherent informational ceilings—limits on their maximum informational value—due to managerial obfuscation or strategic communication. Retail investors, however, often have minimal understanding of these calls. Our study suggests that AI democratization may raise the "floor" of information processing, by providing retail investors with tools to achieve a level of comprehension closer to that of short sellers. This aligns with Sam Altman's vision of "lifting up the floor in the world" through "open" AI. In sum, our findings shed light on the debate over whether AI democratization reduces barriers to critical financial insights or reinforces existing inequalities. Without easily usable AI tools at low costs, the sheer strength of the return predictability of AI-sentiment we document remains inaccessible to ordinary investors, suggesting that the information—and wealth—gap between privileged and ordinary investors may grow exponentially in the AI era, as some fear.

Second, we develop a call-based sentiment measure using ChatGPT, the crown jewel of democratized AI. To date, the fields of economics, finance, and accounting continue to rely heavily on the bag-of-words HD-sentiment, based on the extensively cited work of LM (2011) and its variations. Our findings reveal that ChatGPT-based AI-sentiment predicts returns far beyond the narrow earnings call window typically analyzed in the call literature, whereas traditional HD-sentiment often produces uninformative or even misleading insights at return horizons relevant to the slow-moving, retail investors. This highlights the potential of AI in delivering actionable insights, particularly for these investors. It suggests that it may be time for academic research to systematically adopt AI-based sentiment measures to complement bag-of-words HD-sentiment measures. Our AI-sentiment offers a unique and valuable measure for future research, meeting the persistent and substantial demand across disciplines to analyze or control textual sentiment information (evidenced by 6,079 Google Scholar citations of LM 2011 alone). Our sentiment analysis possesses three distinctive features. First, we demonstrate significant and long-lasting return predictability by analyzing a central and uniquely complex source of textual information—earnings calls. Earnings calls are, on average, orders of magnitude longer than the texts typically used in contemporaneous ChatGPT-based return predictability studies. They are also distinctive information, voluntarily disclosed by insiders, differing from textual information disseminated by newswire or media journalists outside the firm. AI may have different capabilities in analyzing these different types of information. Indeed, the sentiment studies based on other types of information document return predictability at far shorter horizons, lasting three days at the stock level and a month at the market level with/without out-of-sample statistical inference.[5] We employ a model capable of processing the entirety of this type of lengthy texts while preserving their full context. Our comprehensive study of sentiment analysis on such central textual information represents a crucial and complementary contribution to the field (see a recent NBER review by Eisfeldt and Schubert 2024). The return predictability based on a simple AI-sentiment measure also differs from that based on complex financial variables (e.g. anomalies), where sophisticated human economic and financial theories are already developed to link these variables to returns. [6][7] Second, our sentiment measure—derived from GPT-3.5 rather than GPT-4 or later models—offers a sufficiently long out-of-sample for statistical inference, mitigating look-ahead bias. Across current AI/LLM research related to predictability of various variables, conclusions are simply meaningless absent robust out-of-sample statistical inference. This is a first-order concern that underpins the fervent academic exploration of AI, shaping the foundation for both current and future studies. Yet many AI studies appear to run ahead of themselves, building narratives based on assumptions that are far from having conclusive evidence. Indeed, Levy (2024) shows that most AI studies in accounting and finance suffer from such look-ahead bias. Third, we shed light on the sources of both the short- and long-run return predictability of

[5]Two most related studies develop ChatGPT-based sentiment measures to predict returns using different financial texts. Earlier than our study, Lira-Lopez and Tang (2023) document return predictability, lasting up to three days, using RavenPack newswire headlines, which average 76 words. Subsequent to our study, Chen, Tang, Zhou, and Zhu (2024) analyze Wall Street Journal (WSJ) headlines and front-page news to forecast aggregate monthly market returns, where WSJ articles average 600 to 1,000 words. We complement these studies by analyzing earnings call transcripts that average 7,000 words. Furthermore, conference calls are voluntarily disclosed by insiders, and also features an question and answer section that is akin to the chain of thought format suitable for AI to reason. In contrast, newswire and media text information are predominantly produced by journalists outside the firm and also feature highly distilled news lacking chain of thought for AI reasoning. Finally, our cross-sectional predictions provide sufficient power for out-of-sample statistical inference, as a larger cross-sectional return data are available post-September 2021, compared to time-series market level predictions, which lack observations to do so at horizons starting from the monthly level.

[6]Broadly, a few studies employ ChatGPT to analyze earnings calls for economic variables or summarizing transcripts (e.g., Kim, Muhn, and Nikolaev, 2023a,b; Jha, Qian, Weber, and Yang, 2024). Unlike our study, these works neither develop a general sentiment measure, nor focus on return predictability. We complement their understanding on earnings calls by (1) comprehensively analyzing the return predictability of earnings transcripts from approximately 3,000 firms, compared with, for instance, the random sample of 360 firms studied by Kim, Muhn, and Nikolaev (2023b); (2) processing entire transcripts to preserve full contextual information, whereas their approach segments transcripts into sub-4,000-token sections.

[7]Another example of long-text information is 10-K reports. However, their substantial length makes it prohibitively costly for academic researchers to comprehensively evaluate ChatGPT's ability to analyze long-form text. Furthermore, the complexity of 10-Ks—combining unstructured information (text) and structured information (financial statement tables)—hinders clean attribution of AI's ability to process long-form text independently of tabular data. Kim, Muhn, and Nikolev (2024) focus exclusively on analyzing financial statement tables in 10-Ks using GPT-4. Their study focuses on predicting earnings and does not examine the textual components, including sentiment analysis.

AI sentiment, and link the long-run predictability to recent advance in the understanding of sticky beliefs.

Third, our study relates to the extensive literature on conference calls. Earlier research on earnings calls demonstrates their informativeness to market participants (Frankel et al. 2009). 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 Reg FD, sophisticated investors' information advantage may actually increase. This is due to their abundant resources such as leveraging AI/ML technology to process information more effectively. Moreover, extant research shows that retail investors utilize information from conference calls (Huang, Martin, Tang and Xie 2024). Our study further extends this research by highlighting the value of AI in enhancing retail investors' understanding of conference calls. We also show that the investment insights from earnings calls are only decodable by AI or a few sophisticated investors before AI democratization. This may explain why studies based on human-dictionary sentiment have concluded that there is no valuable information in these calls (LM) or even negative value (e.g., negative return predictability documented by Jiang et al. 2019).

1 Sample Construction and Descriptive Statistics

1.1 Earnings Calls, Their Precise Timing, and Relevant Returns

We obtain 280,554 transcripts of conference calls from S&P Capital IQ, spanning 2005-2023. Our sample is confined to common stocks on the NYSE, AMEX, and Nasdaq.

Data vendors such as COMPUSTAT and I/B/E/S typically report the calendar day of earnings announcements without specifying whether they occur after market hours. Moreover, these reported dates can vary (DellaVigna and Pollet 2009). Although I/B/E/S does provide some information on the timing of earnings releases, its coverage is limited and reliability questionable (Berkman and Truong 2008). Prior research has primarily relied on imprecise calendar dates or broader return windows— (typically a two-day or three-day periods) to mitigate the timing issue (e.g., Hirshleifer, Lim and Teoh 2009). When aligning returns with an earnings call that occurs after hours (i.e., after 4 p.m. on calendar day d), simply associating it with day d returns is problematic for return predictability analysis. The trading day timing convention in CRSP starts from 4 p.m. of the previous day (d-1) and ends at 4 p.m. of the current day (d). This convention determines the timing of daily returns. To the best of our knowledge, some contemporaneous studies using ChatGPT to analyze conference calls (e.g., Kim et al. 2023a, b) appear to use the calendar day timing. This approach is problematic for making clear statement about return predictability, which is the focus of our study. Research has shown that substantial returns are often realized around the window of earnings announcements (Engleberg, Mclean, and Pontiff 2018) or specifically during conference calls (Matsumoto, Pronk, and Roelofsen 2011). These contemporaneous returns can represent the bulk of the return predictability when calendar day timing is used, potentially leading to incorrect conclusions about return predictability. This issue is further exacerbated by the fact that 60% of earnings calls are during non-trading hours. To ensure precise timing of earnings call days and to assign them properly to corresponding trading days, we obtain the start time for each call from S&P Capital IQ. If an earnings call occurs after market hours, we designate the next trading day as the event day (d) instead of the current calendar day. We retrieve audio recordings from S&P Capital IQ, analyze their length, and add the duration to the start time to determine the end time of each earnings call.[8]

We observe the number of firms covered by the earnings calls data start from a few firms in 2005 and steadily increase steadily over time, particularly after 2010. In addition, the retail trading measures developed by Boehmer, Jones, Zhang, and Zhang (2021) are available only from 2010 onwards. Therefore, to ensure sufficient number of observations for generating reliable return predictability results and for linking to retail trading, we focus on the period from 2010 to 2023, encompassing 173,229 earnings call transcripts and provides a robust dataset for our return predictability and trading analyses. We referred to this sample as the "full sample" in our analyses. We focus on earnings calls due to their regular and comprehensive disclosure of firm performance, which includes management's discussion on both past and future outcomes. Other types of conference calls, such as special calls, are excluded due to their irregular nature and narrower focus (Tang 2013). During our sample period, the average conference call transcript comprises approximately 7,000 words, highlighting the extensive information typically conveyed through these calls.

We also examine a "post-final knowledge sample" from October 2021 to December 2023. Con-

[8]Since batch downloading is unavailable for both transcripts and audios, we developed a Python script to automate the retrieval of URLs for these resources, streamlining the download process.

sidering that the knowledge base of the API version of gpt-3.5-turbo-16k is constrained to data available up to September 2021, this "post-final knowledge sample" allows us to predict returns using only prior knowledge. This approach ensures a clear and untainted analysis, as it eliminates any potential influence from information that emerged after the model's training cut-off date.

1.2 AI Sentiment

We use GPT-3.5-turbo-16k model to generate the AI sentiment measure. As noted earlier, this model not only preserves the entire context of earnings calls, but also has the advantage of having a sufficiently long post-knowledge sample—unlike the GPT4 models—allowing for a truly out-of-sample test. This aspect is crucial for return predictability studies, as it ensures that our analysis is based on data and information available prior to the evaluation period, thereby enhancing the robustness and credibility of our findings.

To enhance the model's reasoning and problem-solving abilities, we follow the logic of "domain-specific prompting" to predefine the concept of conference call transcripts and configure the model to simulate the perspective of a financial expert. This approach ensures that the model's analysis is contextually grounded and aligned with the nuances of financial communication. 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, thereby enhancing the relevance and financial acumen of its responses. We also set the model temperature to zero to enhance accuracy and reduce the occurrence of model hallucinations. Next, 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 grounded in context and aligned with financial market insights. We use the sentiment score provided by the model as our measure of AI-sentiment, which 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 Retail Trading

Retail trading data are 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. 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 $RH_{i,d}$ denote the retail holding (scaled by share outstanding) of a stock i on day 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:

$$
\begin{aligned}
RetailTrading_{i,d} &= RH_{i,d} \; - \; \frac{RH_{i,d-1} + RH_{i,d-2} + \cdots + RH_{i,d-252}}{252} \\
&= RH_{i,d} \; - \; RH_{i,d-1} \; + \; \Big(\frac{251}{252} RH_{i,d-1} \; - \; \frac{251}{252} RH_{i,d-2}\Big) \\
&\quad + \Big(\frac{250}{252} RH_{i,d-2} \; - \; \frac{250}{252} RH_{i,d-3}\Big) \; + \; \cdots \; + \; \Big(\frac{1}{252} RH_{i,d-251} \; - \; \frac{1}{252} RH_{i,d-252}\Big) \\
&= NB_{i,d} \; + \; \frac{251}{252} NB_{i,d-1} \; + \; \frac{250}{252} NB_{i,d-2} \; + \; \cdots \; + \; \frac{1}{252} NB_{i,d-251} \qquad (1)
\end{aligned}
$$

where the last row of the formula provides a method to compute retail holding using retail net buy. 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.

1.4 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.[9] First, we construct a short volume measure, denoted as $SV_{i,d}$, which represents the aggregate number of shares of stock i that were reported to be sold short during regular trading hours on day d (scaled by share outstanding). The validity of this measure and the reliability of FINRA short volume data are supported by prior research (Blocher, Dong, Ringgenberg, Savor 2023; Wang, Yan, Zheng 2019). Second, we construct a detrended abnormal short-seller holding measure, which can be derived from the short volume measure. Specifically, let $SH_{i,d}$ denote the short-seller holding (scaled by share outstanding) of a stock i on day d. The abnormal short-seller holding measure, defined as the holding detrended by

[9] 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.

its one-year (252 trading days) moving average can be expressed as follows:

$$
\begin{aligned}
ShortSelling_{i,d} = SH_{i,d} \;-\; & \frac{SH_{i,d-1} + SH_{i,d-2} + \cdots + SH_{i,d-252}}{252} \\
= SH_{i,d} \;-\; & SH_{i,d-1} \;+\; \Big(\frac{251}{252}\,SH_{i,d-1} \;-\; \frac{251}{252}\,SH_{i,d-2}\Big) \\
& + \Big(\frac{250}{252}\,SH_{i,d-2} \;-\; \frac{250}{252}\,SH_{i,d-3}\Big) \;+\; \cdots \;+\; \Big(\frac{1}{252}\,SH_{i,d-251} \;-\; \frac{1}{252}\,SH_{i,d-252}\Big) \\
= SV_{i,d} \;+\; & \frac{251}{252}\,SV_{i,d-1} \;+\; \frac{250}{252}\,SV_{i,d-2} \;+\; \cdots \;+\; \frac{1}{252}\,SV_{i,d-251} \qquad (2)
\end{aligned}
$$

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

1.5 Bid-Ask Spread

We measure bid-ask spread using the following formula:

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

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:

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

We standardize AI-sentiment and HD-sentiment to be on the same scale for all subsequent regressions.

Regarding other control variables: 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. 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 and HD-sentiment are standardized with a mean of 0.00 and a standard deviation of 1.00 for direct comparison. In the full sample, AI-sentiment displays more extreme negative values (1st percentile: -4.15) and fewer extreme positive values (99th percentile: 1.30), indicating a negatively skewed distribution. In contrast, HD-sentiment exhibits less extreme negative values (1st percentile: -2.63) and higher positive extremes (99th percentile: 1.96), reflecting a positively skewed distribution. The positive skew in HD-sentiment likely results from the use of overly optimistic language, potentially linked to managerial efforts to influence perceptions by emphasizing extreme positive terms. In contrast, AI-sentiment, with its negative skew, appears less influenced by such linguistic manipulation, potentially providing a more objective assessment of sentiment in earnings call transcripts. These patterns suggest a discrepancy in the intensity of sentiment captured by the two measures, a pattern we examine more closely in Section 4.2.

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 short-selling volume underscores the more pronounced positions held by short sellers compared to retail investors. In the post-final knowledge sample, most variables are comparable to those in the full sample. However, there are notable differences, including a higher earnings surprise and lower 12-month momentum and abnormal retail trading.

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 for both monthly returns and cumulative monthly returns.

2.1 Event-Time Study

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

$$CAR_{i,d+T} = b + \alpha\, \textit{AI-sentiment}_{i,d} + \beta\, \textit{HD-sentiment}_{i,d} + \delta\, x_{i,d} + \epsilon_{i,d} \tag{5}$$

where the dependent variable, $CAR_{i,d+T}$, is stock i's cumulative abnormal returns (CAR) on the T trading days after the event day d of an earnings call. The abnormal returns are calculated following the DGTW method (Daniel, Grinblatt, Titman, and Wermers, 1997), which adjusts for size, book-to-market ratio, and momentum characteristics. Specifically, each stock's return is benchmarked against a portfolio of stocks with similar market capitalization (size), book-to-market ratios, and past six-month returns (momentum), allowing us to isolate the abnormal performance attributable to our sentiment measures.

$\textit{AI-sentiment}_{i,d}$ is the AI-sentiment score of stock i's earnings call on day d computed by gpt-3.5-turbo-16k API model. $\textit{HD-sentiment}_{i,d}$ is the human dictionary sentiment score of stock i's earnings call on day d computed by LM dictionary. $\textit{AI-sentiment}_{i,d}$ and $\textit{HD-sentiment}_{i,d}$ are both standardized in equation (5) and all subsequent regressions. $x_{i,d}$ is the vector of aforementioned control variables measured at the end of the fiscal quarter preceding d. 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 the full sample (Panel A) and the post-final knowledge sample (Panel B). Panel A column (1) shows that the market absorbs AI-sentiment information on the next trading day following an earnings call (coefficient = 0.13, t-statistic = 7.55). A one unit standard deviation increase in AI-sentiment on earnings call event day correlates with an average subsequent return increase of 0.13% on the next trading day, after controlling for HD-sentiment. Furthermore, the predictive strength of AI-sentiment on returns increases progressively and continues to be statistically significant for up to six months following the earnings call (coefficient = 0.99, t-statistic = 5.42).

In stark contrast, the return predictability of HD-sentiment demonstrates relatively weak return predictability in terms of both 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 0.031 to 0.057 and t-statistics from 1.69 to 2.06 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 = 2.89), despite the substantial loss of observations. HD-sentiment fails to predict returns across all examined time horizons. Notably, the coefficient on HD-sentiment turns significantly negative in columns (9), indicating that human dictionary even predicts long-term returns in the wrong direction, leading to negative returns six months after the earnings call (coefficient = -1.42, t-statistic = -2.64).

The evidence from our tests establishes the long-run return predictability of AI-sentiment following the earnings call event day. This indicates that investors can derive significant investment insights from earnings calls by utilizing democratized AI. Notably, this advantage of AI democratization is accessible even to investors who are unable to trade after hours, which is particularly crucial in the current era where AI technologies are broadly accessible.

2.2 Calendar Time Portfolios

We next employ a calendar time portfolio approach to assess the actionability of AI insights. 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 and hold the portfolio for one month. Given that earnings are reported quarterly, investors typically have on average 1.5 months to react to the AI-sentiment signal. This long-time window helps to further support that ordinary investors can have sufficiently long time to act on the investment insights provided by the AI-sentiment signal.

We use both equal-weighted and value-weighted portfolios where market values are capped at upper 80 percent following Jensen, Kelly, and Pedersen (2023). As traditional value-weighted portfolios have become excessively concentrated in a few mega-cap stocks in recent decades, we implement capped value to mitigate this mega-cap dominance, enhance diversification, and better align the portfolios with actionable investment opportunities across a broader range of stocks. Given our more recent sample, this approach accounts for the growing concentration of market capitalization in large firms, ensuring a more representative portfolio. The High (Low) portfolio comprises stocks in the highest (lowest) decile and the High-Low portfolio represents 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 0.42% (t-statistic = 2.51), and the long-short AI-sentiment portfolio CAPM alpha is 0.91% (t-statistic =2.79). In contrast, HD-sentiment portfolios do not yield significant CAPM alphas, regardless of whether they are in High (Low) or long-short configurations. Furthermore, the long-short SUE portfolio yields a CAPM alpha of 0.6% (t-statistic = 2.17), consistent with the PEAD phenomenon. The equal-weighted CAPM alpha of long-short AI-sentiment portfolio is comparable to that of long-short SUE portfolio. This indicates that AI-sentiment is a valid signal, akin to earnings surprises, for identifying stocks that will outperform or underperform the market. When we use value-weighted portfolios, long-short AI-sentiment portfolio still generates a CAPM alpha of 0.61% (t-statistic = 2.76), while both long-short HD-sentiment and SUE portfolios do not yield significant CAPM alphas.

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 0.36% (value-weighted portfolio in Carhart four-factor model) to 0.72% (equal-weighted portfolio in five-factor model) and are all statistically significant.

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 the AI-sentiment long-short portfolios across High, Medium, and Low HD-sentiment (SUE). Effectively, the averaged long-short AI-sentiment portfolio alpha captures the alphas attributable to AI-sentiment after controlling for the effect of HD-sentiment (SUE). We observe that long-short AI-sentiment portfolio alphas range from 0.39% to 0.67%, all of which are statistically significant with t-statistics ranging from 2.47 to 3.71. These results confirm the robust alpha generation capability of the long-short AI-sentiment portfolio, even when controlling for HD-sentiment and SUE. This strengthens the argument that AI-derived insights are actionable and extend beyond those offered by earnings surprises alone.

Table 3 Panel A-C focuses on monthly returns generated by portfolio sorts. As shown in Section 2.1, AI-sentiment predicts long-term cumulative daily returns post-earnings call. We now explore if this long-term return predictability persists in the portfolio sort analysis. Table 4 examines the cumulative monthly CAPM alphas for the long-short AI-sentiment, HD-sentiment, and SUE portfolios up to six months. An equal-weighted long-short AI-sentiment portfolio yields significant cumulative CAPM alphas, rising from 1.56% over 2 months (t-statistic = 3.66) to 4% over 6 months (t-statistic = 5.69). 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 as well. A value-weighted long-short AI-sentiment generates a 6-month CAPM alpha of 3.36% (t-statistic = 5.92), while the long-short HD-sentiment portfolio shows no significant long-term effect. Additionally, an equal-weighted long-short SUE portfolio yields a 4-month CAPM alpha of 1.11% (t-statistic = 2.00), 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, corroborating the long-term return predictability of AI-sentiment demonstrated in Section 2.1.

3 Retail Trading, Short Selling, and AI

Our previous findings suggest that AI-sentiment provides superior return predictability over HD-sentiment. Next, we contrast the daily trading behaviors between retail investors (the ordinary investors) and short sellers (the privileged investors) around earnings calls to explore whether the earnings call information uncovered by AI-sentiment affects investors' trading dynamics. In Section 3.1 and 3.2, we investigate how retail trading and short selling align with AI-sentiment around earnings calls, respectively, before the wide deployment of ChatGPT. In Section 3.3, we explore the shifts in retail trading and short selling surrounding the wide deployment of ChatGPT and study the causal effect of AI democratization on trading. In Section 3.4, we examine the relation between retail trading and AI-sentiment when there are exogenous ChatGPT outages. In Section 3.5, we examine the relation between AI democratization and information asymmetry.

3.1 Retail Trading

We first examine how retail trading aligns with sentiment measures in the pre-democratization period. ChatGPT, released on November 30, 2022, experienced rapid growth, reaching 100 million monthly active users by January 2023.[10] 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.

We regress RetailTrading, a detrended abnormal retail trading measure constructed in Section 1.3, on AI-sentiment. We examine retail trading on the earnings call event day and within 5, 10, and 21 trading days post-earnings calls. Based on our earlier evidence indicating that the return predictability may plateau as early as one quarter after the earnings call, examining the short windows within 21 trading days allows us to focus on investor responses that can leverage AI-driven investment insights. This approach helps minimize the influence of long-term factors.

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

Like the previous tests, we add HD-sentiment as a benchmark. We also control for other anomaly variables to rule out the possibility that retail traders simply trade on those public signals. We include year-quarter fixed effects and cluster standard errors by firm and year-quarter. We multiply RetailTrading by 100000 to ease interpretation. A significantly positive coefficient for AI-sentiment implies that retail trading activity increases with AI-sentiment, suggesting that retail trading aligns with the insights derived from AI analysis of earnings calls.

Table 5, Panel A presents the results. Retail investors exhibit negative reaction to HD-sentiment but no significant reaction to AI-sentiment on earnings call days and during the one month after earnings calls. These findings suggest that prior to the democratization of AI, retail investors may lack the sophistication required to distill actionable insights from earnings calls, and earlier regulatory efforts to democratize information access may not effectively help level the playing field.

3.2 Short Selling

We next examine how short selling aligns with sentiment measures. We follow the same regression and setting used in Section 3.1 and replace the dependent variable with ShortSelling, a detrended abnormal short-seller holding measure created in Section 1.4. To facilitate interpretation, we multiply ShortSelling by 100000.

Table 5, Panel B reports the results. 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. In contrast to the muted response of retail investors related to AI-sentiment observed in Table 5, Panel A, short sellers demonstrate a pronounced reaction to AI-sentiment on the day of earnings calls (coefficient = -706.9, t-statistic = -2.02), controlling for HD-sentiment and other anomaly variables. A one standard deviation increase in AI-sentiment is associated with 0.7% reduction in abnormal shorting. This suggests that short sellers reduce short positions immediately on earnings call days if AI-sentiment is more positive. 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.

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 after the earnings calls (coefficient = -15370.6, t-statistic = -1.79). The increasing magnitude of the correlation between short selling and AI-sentiment from columns (1) to (4) suggests that short sellers are progressively increasing positions after earnings calls, implying a long-term trend instead of a temporary adjustment. These findings imply that short sellers' trading aligns with the information that AI extracts from earnings calls. Given that ChatGPT was not available in the pre-democratization period, these findings suggest that short sellers are informed traders, with superior information sources or processing capacities such as proprietary AI models to analyze earnings calls.

3.3 AI Democratization

We next analyze shifts in short selling and retail trading surrounding ChatGPT's wide deployment. We define the pre- (post-) democratization periods as before (after) ChatGPT's wide deployment date of January 1, 2023. In Section 3.3.1, we examine how retail-AI alignment and short-sellers-AI alignment change around AI democratization using a sample of 2022-2023. In Section 3.3.2, we 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.

3.3.1 Trading and Pre- vs. Post-AI Democratization

Table 6 reports how retail trading (Panel A) and short selling (Panel B) are aligned with AI-sentiment before and after ChatGPT's wide deployment using 2022-2023 sample. 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 period 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 that the trading changes are attributable to ChatGPT's deployment rather than to other confounding factors. After is a dummy variable that equals 1 for post-democratization period, and 0 for pre-democratization period. We regress RetailTrading (ShortSelling) on AI-sentiment and AI-sentiment*After. Following Section 3.1 and 3.2, we focus on retail trading (short selling) on earnings call days and over 5, 10, and 21 trading days after. We include HD-sentiment 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 retail-AI alignment (short-sellers-AI alignment) after AI democratization. A significantly positive (negative) coefficient for AI-sentiment*After implies an increase in retail investors' (short sellers') alignment with AI-sentiment after AI democratization.

In Table 6, Panel A, the significantly negative coefficients of AI-sentiment on retail trading across column (1)-(4) suggest that retail investors' trading is strongly against AI-sentiment in the pre-democratization period in the year 2022. The positive and significant coefficient of AI-sentiment*After on retail trading suggests retail trading exhibits a significantly heightened reaction to AI-sentiment during the one month after earnings calls in the post-democratization period. The changes of retail response to AI-sentiment not only are large comparing to their pre-democratization levels in 2022, they also represent directional changes as retail traders change from trading against AI-sentiment in 2022 to trading with AI-sentiment in 2023. These retail trading changes are also large if we compare the coefficients on AI-sentiment*After with the long-term coefficients on AI-sentiment during the pre-democratization period over 2010-2022. The two types of coefficients are both positive, but the coefficients on AI-sentiment*After are an order of magnitude larger than the coefficients on AI-sentiment. For example, on the earnings call day, retail investors' alignment with AI-sentiment after AI democratization (coefficient = 39.9) increases 23-fold compared to their long-term average alignment with AI-sentiment before AI democratization documented in Table 5, Panel A (coefficient = 1.72). This paradigm shift suggests that despite being "dumb" before AI democratization, retail investors trade in the correct direction after AI democratization.

In contrast, in Table 6, Panel B, the significantly negative coefficients of AI-sentiment on short selling suggest that short sellers' trading is strongly aligned with AI-sentiment in the pre-democratization period. A one standard deviation increase in AI-sentiment corresponds to a 0.61 decrease in short selling over 21 trading days, more than 100% of its standardization (0.56). This substantial adjustment underscores the pronounced impact of AI-driven market signals on short selling, exemplifying a "rich get richer" dynamic. The positive and significant coefficient of AI-sentiment*After on short selling indicates that short sellers even weaken their alignment with

AI-sentiment after AI democratization, although the change is only weakly significant.

The findings suggest that democratization of AI enables retail investors, who are unable to engage with AI strategies before AI democratization, to align their trading decisions correctly and more closely with the actionable insights extracted by AI from earnings calls. In contrast, short sellers' alignment with AI diminishes, as they no longer enjoy a clear information advantage by trading on these AI insights and thus may choose to scale back their reliance on such signals.

3.3.2 Difference-in-Differences

Section 3.3.1 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:

$$
\begin{aligned}
&StandardizedTrading_{i,g,d+T} \\
&\quad = \beta_0 + \beta_1\, AI\text{-}sentiment_{i,d} + \beta_2\, AI\text{-}sentiment_{i,d} \times Treat_g^{IsRetail} + \beta_3\, AI\text{-}sentiment_{i,d} \times After_d \\
&\quad + \beta_4\, AI\text{-}sentiment_{i,d} \times Treat_g^{IsRetail} \times After_d + \delta\, x_{i,g,d} + \alpha_{i,g} + \gamma_{d,g} + \epsilon_{i,g,d} \qquad (6)
\end{aligned}
$$

where the dependent variable is the standardized trading of trader g on stock i in the T trading days following 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 unit of observation is 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 invert the sign of ShortSelling so that a positive correlation between ShortSelling and AI-sentiment reflects short selling is aligned with AI-sentiment. We then separately standardize ShortSelling and RetailTrading, and stack them into a unified dependent variable StandardizedTrading. $Treat^{IsRetail}$ is a dummy variable that equals 1 if the trading variable for a stock in a given quarter is retail trading and 0 if it is short selling. 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 $AI\text{-}sentiment \times Treat^{IsRetail} \times After$. A positive (negative) coefficient for $AI\text{-}sentiment \times Treat^{IsRetail} \times After$ 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.

Our stacked regression approach alleviates several endogeneity concerns. First, by stacking retail trading and short selling into one variable, we compare retail trading and short selling within the same firm. Second, we include stock-trader fixed effects ($\alpha_{i,g}$) to control for time-invariant characteristics specific to each stock-trader group combination. In other words, the time-invariant differences between retail traders and short sellers of the same firm are controlled for by this fixed effect. Third, we also include the year-quarter-trader fixed effects ($\gamma_{d,g}$) to control for time-specific effects that are unique to each trader group. These fixed effects control for the general differences between retail trading and short selling across all firms in a quarter. We do not include $Treat_g^{IsRetail}$, $After_d$, and $Treat_g^{IsRetail} \times After_d$ in the regression as they are absorbed by these fixed effects. We cluster standard errors by firm and year-quarter.

Table 7 presents the results of the DiD test examining the relation between AI democratization and trading activities. The negative coefficients for $AI\text{-}sentiment \times Treat^{IsRetail}$ suggest that retail investors are less sensitive to AI-sentiment compared to short sellers in the pre-democratization period. The negative coefficients for $AI\text{-}sentiment \times After$ suggest that short selling is less aligned with AI-sentiment after AI democratization, consistent with the pattern documented in Table 6, Panel B. The coefficients for $AI\text{-}sentiment \times Treat^{IsRetail} \times After$ 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 $AI\text{-}sentiment \times Treat^{IsRetail} \times After$ 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 1 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.

3.4 ChatGPT Outages

We exploit ChatGPT outages as exogenous shocks to identify changes in the availability of AI tools for retail investors. The purposes of this test are twofold. First, we further reduce the endogeneity concerns of the relation between AI democratization and trading. Second, we provide evidence supporting that retail response to AI-sentiment is indeed linked to the usage of ChatGPT. Intuitively, if AI democratization benefits retail investors in processing earnings call information, unexpected ChatGPT outages would diminish this effect as retail investors are not able to use the tool.

Specifically, we incorporate the outage variable into the DiD framework discussed in Section 3.3.2. 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).[11][12] Since ChatGPT, and therefore the 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

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

[12] 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.

post-democratization context. $After\&Outage$ 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, 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 $AI\text{-}sentiment \times Treat^{IsRetail} \times After\&Outage$ 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 8 presents how retail-AI alignment changes relative to short-sellers-AI alignment when ChatGPT outages occur after AI democratization. The coefficients of $AI\text{-}sentiment \times Treat^{IsRetail} \times After$ are significantly positive, consistent with the results in Table 7 that AI democratization leads to increased retail-AI alignment relative to that of short sellers. In contrast, the coefficients for $AI\text{-}sentiment \times Treat^{IsRetail} \times After\&Outage$ are negative and significant across all time horizons. For example, in column (1), the coefficient for $AI\text{-}sentiment \times Treat^{IsRetail} \times After$ is 0.091 (t-statistics = 2.69), while the coefficient for $AI\text{-}sentiment \times Treat^{IsRetail} \times After\&Outage$ 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 and support that retail response to AI-sentiment is linked to ChatGPT usage.

3.5 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:

$$BidAskSpread_{i,d+T} = \beta_0 + \beta_1\, Treat_i^{AI} + \beta_2\, Treat_i^{AI} \times After_d + \delta\, x_{i,d} + \alpha_d + \epsilon_{i,d} \tag{7}$$

where BidAskSpread is the difference between the daily bid and ask price scaled by the average of bid and ask price of stock i in the T trading days following 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 (6) on a stock-by-stock basis and define each stock's retail-AI alignment using the coefficient of $AI\text{-}sentiment \times Treat^{IsRetail} \times After$. A higher coefficient of $AI\text{-}sentiment \times Treat^{IsRetail} \times After$ 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.[13] 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 $Treat^{AI} \times After$. A negative (positive) coefficient for $Treat^{AI} \times After$ indicates a decrease (increase) in information asymmetry for high

[13]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.

retail-AI alignment stocks after AI democratization relative to low retail-AI alignment stocks. $x_{i,d}$ is the vector of control variables including HD-sentiment, SUE, Beta, BM, MVE, and Mom12m. We include year-quarter fixed effects (α_d) and cluster 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 9, Panel A 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 $Treat^{AI} \times After$ 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 $Treat^{AI} \times After$ 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 3.4. A positive (negative) coefficient for $Treat^{AI} \times After\&Outage$ 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 9, Panel B presents the relation between AI democratization and information asymmetry when ChatGPT outages occur. The negative coefficients for $Treat^{AI} \times After$ are consistent with the results in Table 9, Panel A that AI democratization reduces information asymmetry for stocks with high retail-AI alignment during non-outage period. In contrast, the positive coefficients for $Treat^{AI} \times After\&Outage$ 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 $Treat^{AI} \times After$ is -27.1 (t-statistic = -1.80), while the coefficient for $Treat^{AI} \times After\&Outage$ 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.

4 Understanding ChatGPT's Superior Performance

In this section, we explore what unique investment insights AI-sentiment could offer to ordinary investors. In Section 4.1, we examine whether AI-sentiment's return predictability arises from superior contextual understanding by comparing AI-sentiment with FinBERT-sentiment in predicting stock returns. In Section 4.2, we examine the differences and connections between AI-sentiment and HD-sentiment in capturing return-predictive information by performing double sorts on the two sentiment measures and comparing the return predictability across sorted groups.

4.1 Unique Sentiment Information from Long Text

We extend the event study analysis presented in Table 2 by incorporating FinBERT-sentiment into a direct comparison with AI-sentiment. Our analysis aims to assess whether GPT demonstrates superior return predictability over short-text models like FinBERT, particularly in interpreting sentiment within long and complex textual contexts.

As a fine-tuned BERT model specifically designed for financial text analysis, FinBERT can effectively capture sentiment within pre-defined financial contexts.[14] However, its reliance on a 512-token input restriction and limited capacity to process extended dependencies in long texts may constrain its ability to fully interpret nuanced sentiment. The 512 token limit implies that FinBERT is only able to understand the context within a sentence. By contrast, the GPT model we use, with a capacity of 16K tokens, is designed to analyze lengthy and intricate documents, leveraging broader contextual understanding to potentially identify sentiment patterns more effectively. Thus, by comparing AI-sentiment with FinBERT-sentiment, we examine whether ChatGPT's long-run return predictability stems from its ability to leverage long textual context.

The FinBERT model we employ is based on a training sample roughly aligned with the sample period of the trading sample of ChatGPT. Furthermore, unlike ChatGPT, we cannot request FinBERT to make predictions based on knowledge only up to the day before a specific call. Therefore,

[14] The FinBERT model used in this analysis is the version fine-tuned by ProsusAI, specifically designed for financial sentiment analysis. The model is available at https://huggingface.co/ProsusAI/finbert.

to make the comparison fair, we focus on the post-final knowledge sample instead of the full sample to eliminate the look-ahead bias in AI-sentiment.

Table 10 presents the return predictability of AI-sentiment and FinBERT-sentiment. AI-sentiment significantly predicts stock returns for up to six months post-earnings calls, whereas FinBERT-sentiment shows no predictive capability. Notwithstanding, the superior return predictability of AI-sentiment over FinBERT-sentiment still exists when applied to the full sample, as shown in Table IA2. These findings suggest that the long-text GPT model outperforms FinBERT in extracting sentiment information from the broader and more complex context of long earnings calls.

4.2 Differences and Connections between AI-sentiment and HD-sentiment

To understand ChatGPT's superior performance in return predictability over traditional human dictionary methods, we first segment HD-sentiment into ten distinct bins and calculate the average values of both AI-sentiment and HD-sentiment within each bin. To ensure comparability, both AI-sentiment and HD-sentiment are standardized. Figure 2 presents the bin-scatter plot. It 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 remains irrespective of HD-sentiment being negative, zero, or positive, suggesting that changes in human-derived sentiment do not significantly influence the AI-generated sentiment. 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 less affected by the extremities of sentiment expressed during these communications.

We then extend the event study by performing independent double sorts of earnings calls based on each call's AI-sentiment and HD-sentiment. In each calendar quarter, we sort AI-sentiment (HD-sentiment) into deciles and define the top three deciles as High_{AI} (High_{HD}) group and the bottom three as Low_{AI} (Low_{HD}) group. This process produces four groups of earnings calls: $\text{High}_{\text{AI}}\text{High}_{\text{HD}}$, $\text{Low}_{\text{AI}}\text{Low}_{\text{HD}}$, $\text{High}_{\text{AI}}\text{Low}_{\text{HD}}$, and $\text{Low}_{\text{AI}}\text{High}_{\text{HD}}$. Following Section 5.1, we use the post-final knowledge sample instead of the full sample to eliminate the look-ahead bias in AI-sentiment. We include other anomaly variables used in previous tests as controls. We include year-quarter fixed effects and cluster standard errors by firm and year-quarter.

Table 11 reports the return predictability of the four double-sorted groups. Following Section 2.1, the abnormal returns are calculated following the DGTW method. $\text{High}_{\text{AI}}\text{High}_{\text{HD}}$ significantly predicts returns from the next trading day (coefficient = 0.23, t-statistic = 2.00) till two months after the call (coefficient = 1.15, t-statistic = 2.68). This implies an earnings call with extremely positive AI-sentiment and HD-sentiment generates a positive 1.15% average two-month post-call return. However, $\text{Low}_{\text{AI}}\text{Low}_{\text{HD}}$ shows no significant predictive power as abnormal returns are insignificantly different from zero. These patterns indicate that when earnings calls exhibit positive news confirmed by both AI and human analysis, the information is quickly incorporated into prices.

$\text{High}_{\text{AI}}\text{Low}_{\text{HD}}$ shows no significant return prediction over any horizon, whereas $\text{Low}_{\text{AI}}\text{High}_{\text{HD}}$ behaves sharply differently. In column (1), the coefficient on $\text{Low}_{\text{AI}}\text{High}_{\text{HD}}$ is negative and significant (coefficient = -0.28, t-statistic = -2.11), indicating that earnings calls with extremely negative AI-sentiment and extremely positive HD-sentiment experience an average next-day return decrease of 0.28%. This negative market reaction suggests that CAR aligns with AI-sentiment when HD is extremely positive while AI strongly disagrees. However, the immediate reaction is far smaller than the long-term reaction. Indeed, CAR intensifies from two months post-call (coefficient = -1.88, t-statistic = -2.55) to six months post-call (coefficient = -3.70, t-statistic = -2.59), representing a 13-fold increase in magnitude compared to the short term effect. The results suggest that the market initially underreacts to the negative news concealed by extremely positive HD-sentiment. Over a prolonged period, the market gradually prices in this bad news, as discerned by AI.

Overall, these findings suggest that, while market reacts in short term when AI-sentiment and HD-sentiment are both positive, AI's long-term return predictability primarily arises when HD-sentiment is extremely positive while AI-sentiment is negative. This suggests that that the market takes considerably longer to recognize and price in such bad news identified by AI, highlighting ChatGPT's ability to detect situations where executives adopt an overly optimistic language in earnings calls that AI flags as bad news.

5 Conclusion

Regulators have long been championing the democratization of investment information to level the playing field, but its effectiveness is debated due to challenges like average investors' limited capacity to process complex information promptly. This era is also marked by advancements like the development of machine-processable information reporting formats and proliferation of big data. These changes almost immediately provide further advantages to privileged investors who already possess superior technology advantage to process the exploding volume of complex real-time information. Consequently, the already extreme wealth gap in society may be further exacerbated by such advancements, potentially counteracting regulators' primary objective of creating a more equitable market landscape.

The advent of AI democratization introduces a new layer of tension to this discourse. If handled improperly, the revolutionary power of AI could exponentially expedite the inequality among traders in the financial market, potentially accelerating the process of the rich getting richer in an unprecedented manner. This study conducts the first analysis of how democratized AI impacts human traders' use of democratized information, particularly contrasting sophisticated, information-privileged investors with ordinary retail investors. By employing ChatGPT, one of the most widespread and advanced AI technologies, to analyze a 19-year sample of earnings calls, we develop a long-text (average 7,000 words) AI-sentiment measure that preserves full transcript context. We show that AI-sentiment predicts significant returns persisting 6 months, while human-dictionary-based (HD) sentiment shows little or negative predictive power. The return predictability of AI remains still strong and long-lasting after controlling for human-dictionary based sentiment and standardized unexpected earnings.

In the years before the existence of ChatGPT, short selling is notably aligned with AI-sentiment in the two weeks following earnings calls, whereas retail trading showed little alignment. Following the widespread deployment of ChatGPT's, retail investors' alignment with AI-sentiment has substantially increased compared to the pre-ChatGPT era, even as short-sellers-AI alignment might have weakened. 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, which also reverses during ChatGPT outages.

We further shed light on what insights AI-sentiment could provide to ordinary investors. We find that ChatGPT's long-term return insights stems from leveraging long-text context and discerning between genuinely and excessively positive HD-sentiment.

These findings highlight the potential of AI to mine actionable insights from long and complex financial information such as earnings calls that align with the knowledge of informed traders. Collectively, our results suggest that in an era of democratizing information access and proliferating big data, democratizing AI—the tools for analyzing information—is crucial to leveling the playing field between privileged and ordinary investors, underscoring the importance of "open" AI (usage) in bridging the information gap between the privileged and ordinary investors.

Given the potential of AI-sentiment in delivering actionable insights, particularly for slow-moving, ordinary investors, our long-text AI-sentiment measure also provides a unique and valuable tool for future research. Our evidence suggests that it may be time for academic research to systematically adopt AI-based sentiment measures to complement LM-type HD-sentiment measures.

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.

Blankespoor, E., Croom, J. and Grant, S.M., 2024. Generative AI and investor processing of financial information. Available at SSRN 5053905.

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.

Daniel, K., Grinblatt, M., Titman, S. and Wermers, R., 1997. Measuring mutual fund performance with characteristic-based benchmarks. *The Journal of Finance*, 52(3), pp.1035-1058.

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.

Dessaint, O., Foucault, T. and Frésard, L., 2024. Does alternative data improve financial forecasting? the horizon effect. *The Journal of Finance*, 79(3), pp.2237-2287.

Eisfeldt, A.L. and Schubert, G., 2024. AI and Finance (No. w33076). National Bureau of Economic Research.

Engelberg, J., McLean, R.D. and Pontiff, J., 2018. Anomalies and news. *The Journal of Finance*, 73(5), pp.1971-2001.

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

Glasserman, P. and 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.

Jiang, F., Lee, J., Martin, X. and Zhou, G., 2019. Manager sentiment and stock returns. *Journal of Financial Economics*, 132(1), pp.126-149.

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. G., 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).

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

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.

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* arXiv:2109.09115.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is all you need. In *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.

Wang, Y., Liu, Q., Xu, J., Liang, T., Chen, X., He, Z., Song, L., Yu, D., Li, J., Zhang, Z., Wang, R., Tu, Z., Mi, H., and Yu, D. 2025. Thoughts are all over the place: on the underthinking of o1-Like LLMs. *arXiv preprint* arXiv:2501.18585.

Zhu, C., 2019. Big data as a governance mechanism. *The Review of Financial Studies*, 32(5), pp.2021-2061.

Figure 1 Trend of Retail Investors' AI Sensitivity

This figure presents 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. In regression (6), we replace After with a series of half-year dummy variables starting from 2021HY2 to 2023HY2 and plot the coefficients of each half-year dummy variable.

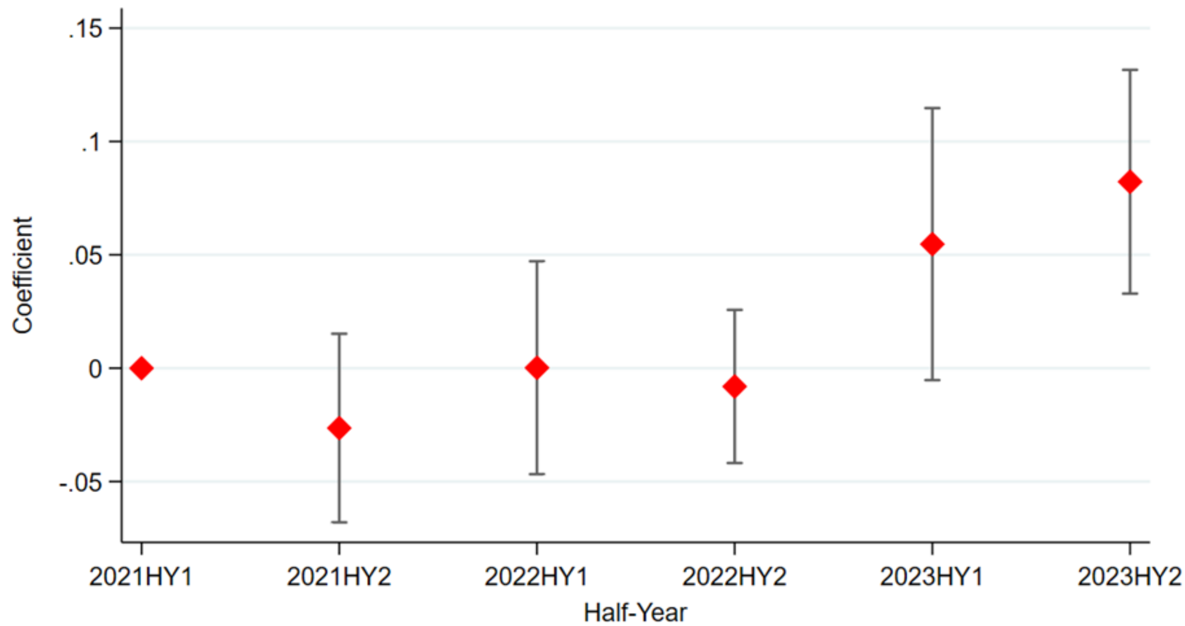

Figure 2 A Bin-Scatter Plot of AI-sentiment vs. HD-sentiment

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.

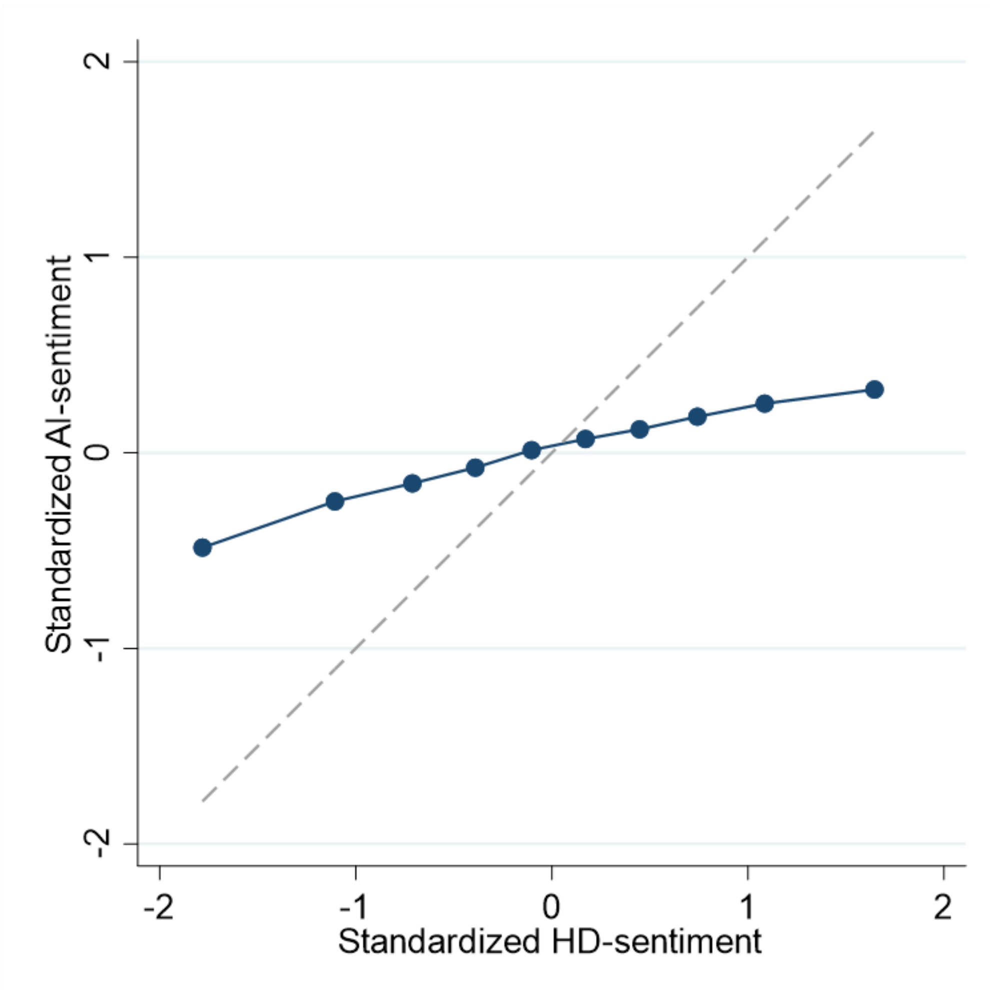

Table 1 Summary Statistics

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. AI-sentiment and HD-sentiment are both standardized (demeaned and with a standard deviation of one). 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 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 (standardized)	81536	0.00	-4.15	0.21	1.30	1.00
HD-sentiment (standardized)	81536	0.00	-2.63	0.08	1.96	1.00
SUE	81536	0.06	-0.33	0.00	0.53	3.92
Beta	81536	1.24	0.16	1.17	2.92	0.56
BM	81536	0.51	-0.52	0.42	2.38	0.68
MVE	81536	14.21	9.45	14.26	19.03	2.05
Mom12m	81536	0.18	-0.79	0.09	2.59	0.83
ShortSelling	82146	0.22	0.01	0.11	2.47	0.49
RetailTrading	85263	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 (standardized)	18589	0.00	-2.32	-0.05	1.31	1.00
HD-sentiment (standardized)	18589	0.00	-2.70	0.09	1.91	1.00
SUE	18589	0.13	-0.47	0.00	0.94	6.20
Beta	18589	1.27	0.13	1.19	3.05	0.59
BM	18589	0.50	-0.61	0.38	2.81	0.80
MVE	18589	14.21	9.10	14.31	19.26	2.24
Mom12m	18589	0.05	-0.88	-0.02	2.22	0.67
ShortSelling	19277	0.28	0.02	0.13	3.39	0.58
RetailTrading	20017	-0.0004	-0.0363	-0.0005	0.0403	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). Cumulative abnormal returns (CAR) are calculated following the DGTW method (Daniel et al., 1997). 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. Control variables include SUE, Beta, BM, MVE, Mom12m. 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 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 and HD-sentiment 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

	$\mathbf{CAR}_{i,[horizon]}$								
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
	[d+1]	**[d+1:d+2]**	**[d+1:d+3]**	**[d+1:d+5]**	**[d+1:d+10]**	**[d+1:d+21]**	**[d+1:d+42]**	**[d+1:d+63]**	**[d+1:d+126]**
AI-sentiment$_{i,d}$	0.13***	0.15***	0.18***	0.19***	0.24***	0.35***	0.45***	0.60***	0.99***
	(7.55)	(5.75)	(5.40)	(4.90)	(4.21)	(4.96)	(4.53)	(4.75)	(5.42)
HD-sentiment$_{i,d}$	0.031*	0.050*	0.057**	0.044	0.0077	0.032	0.17	0.12	0.026
	(1.69)	(1.95)	(2.06)	(1.19)	(0.13)	(0.43)	(1.33)	(0.79)	(0.11)
Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter FE	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
Sample Period	10-23	10-23	10-23	10-23	10-23	10-23	10-23	10-23	10-23
N	81536	81536	81536	81536	81536	81536	81536	81536	81536
r^2	0.0043	0.0045	0.0049	0.0051	0.0048	0.0040	0.0038	0.0041	0.0034

Panel B: Post-Final Knowledge Sample

	$\mathbf{CAR}_{i,[horizon]}$								
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
	[d+1]	**[d+1:d+2]**	**[d+1:d+3]**	**[d+1:d+5]**	**[d+1:d+10]**	**[d+1:d+21]**	**[d+1:d+42]**	**[d+1:d+63]**	**[d+1:d+126]**
AI-sentiment$_{i,d}$	0.13**	0.18**	0.23**	0.24**	0.28*	0.54***	0.66***	0.75**	0.71*
	(3.03)	(2.78)	(2.49)	(2.37)	(1.88)	(3.72)	(4.15)	(2.89)	(1.98)
HD-sentiment$_{i,d}$	0.0080	0.059	0.070	0.017	-0.021	0.0085	0.048	-0.54	-1.42**
	(0.15)	(0.68)	(0.70)	(0.13)	(-0.11)	(0.04)	(0.12)	(-1.58)	(-2.64)
Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter FE	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
Sample Period	21Oct-23Dec	21Oct-23Dec	21Oct-23Dec	21Oct-23Dec	21Oct-23Dec	21Oct-23Dec	21Oct-23Dec	21Oct-23Dec	21Oct-23Dec
N	18589	18589	18589	18589	18589	18589	18589	18589	18589
r^2	0.0038	0.0043	0.0048	0.0055	0.0059	0.0051	0.0051	0.0062	0.0092

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 analysis uses the full sample. 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. For single sorts portfolios, 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 call. For double sorts portfolios, at the end of each month, we sort stocks into terciles first based on HD-sentiment (SUE) and then on AI-sentiment. 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	0.42**	-0.06	-0.65	0.26**	0.10	-0.44
	(2.51)	(-0.27)	(-1.40)	(2.22)	(0.87)	(-1.47)
Low	-0.50	-0.24	-1.26***	-0.36**	-0.03	-0.73**
	(-1.42)	(-0.96)	(-2.79)	(-1.97)	(-0.18)	(-2.05)
High-Low	0.91***	0.18	0.60**	0.61***	0.13	0.30
	(2.79)	(0.96)	(2.17)	(2.76)	(0.74)	(0.87)

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	0.53***	-0.18	0.70**	0.29***	-0.21	0.50***
	(4.29)	(-0.72)	(2.50)	(2.88)	(-1.53)	(2.70)
Carhart4	0.47***	-0.05	0.52**	0.24**	-0.12	0.36**
	(4.01)	(-0.22)	(2.07)	(2.52)	(-0.96)	(2.28)
FF5	0.55***	-0.16	0.72**	0.30***	-0.26*	0.56***
	(4.43)	(-0.66)	(2.53)	(2.97)	(-1.83)	(2.99)

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**
Average across HD-sentiment-sorted portfolios	**CAPM**	0.19	-0.31	-0.48*	0.67***	0.18**	-0.14**	-0.30**	0.49***
		(1.18)	(-1.55)	(-1.70)	(3.24)	(2.35)	(-2.03)	(-2.27)	(2.74)
	FF3	0.35**	-0.08	-0.20	0.55***	0.15**	-0.12*	-0.24*	0.39**
		(3.60)	(-0.96)	(-1.12)	(3.01)	(2.07)	(-1.85)	(-1.94)	(2.47)
	FF5	0.40**	-0.03	-0.18	0.57***	0.15**	-0.13*	-0.29**	0.44***
		(4.13)	(-0.37)	(-0.99)	(3.12)	(2.07)	(-1.95)	(-2.33)	(2.78)
Average across SUE-sorted portfolios	**CAPM**	0.18	-0.27	-0.46*	0.64***	0.22**	-0.14	-0.37**	0.59***
		(1.09)	(-1.34)	(-1.68)	(3.38)	(2.90)	(-1.51)	(-2.80)	(3.71)
	FF3	0.35**	-0.04	-0.19	0.54***	0.22**	-0.09	-0.31**	0.52***
		(3.73)	(-0.45)	(-1.09)	(3.20)	(2.84)	(-1.21)	(-2.65)	(3.59)
	FF5	0.39**	0.02	-0.17	0.56***	0.20**	-0.09	-0.33**	0.53***
		(4.22)	(0.24)	(-0.98)	(3.35)	(2.53)	(-1.29)	(-2.84)	(3.54)

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 using the full sample. 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.

		CAPM Alpha$_{[horizon]}$				
		(1)	**(2)**	**(3)**	**(4)**	**(5)**
		[m+1:m+2]	**[m+1:m+3]**	**[m+1:m+4]**	**[m+1:m+5]**	**[m+1:m+6]**
Equal Weighted	**AI-sentiment**	1.56***	2.24***	2.90***	3.62***	4.00***
		(3.66)	(4.22)	(4.90)	(5.54)	(5.69)
	HD-sentiment	0.17	0.11	-0.08	-0.22	-0.49
		(0.71)	(0.35)	(0.23)	(0.55)	(1.16)
	SUE	1.01***	1.25***	1.11**	1.06*	1.01
		(2.84)	(2.70)	(2.00)	(1.68)	(1.48)
Value Weighted (80p)	**AI-sentiment**	1.10***	1.58***	2.17***	2.84***	3.36***
		(3.60)	(3.95)	(4.67)	(5.57)	(5.92)
	HD-sentiment	0.14	0.00	-0.14	-0.28	-0.46
		(0.60)	(0.01)	(0.41)	(0.74)	(1.05)
	SUE	0.25	0.33	0.13	-0.16	-0.57
		(0.63)	(0.73)	(0.26)	(0.30)	(0.94)

Table 5 Trading and AI-sentiment in Pre-Democratization Period

The table presents retail trading (Panel A) and short selling (Panel B) around earnings calls in alignment with AI-sentiment and HD-sentiment during the pre-democratization period in 2010-2022. Retail trading (multiplied by 100000) is abnormal retail holding calculated in Section 1.3. Short selling (multiplied by 100000) is abnormal short-seller 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. Control variables include SUE, Beta, BM, MVE, and Mom12m. 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 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. We include year-quarter fixed effect. AI-sentiment and HD-sentiment are standardized for regression. 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: Retail Trading

	Retail Trading$_{i,[horizon]}$			
	(1) **[d]**	**(2)** **[d+1:d+5]**	**(3)** **[d+1:d+10]**	**(4)** **[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)
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

Panel B: Short Selling

	Short Selling$_{i,[horizon]}$			
	(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)	(-2.04)	(-2.09)	(-2.12)
HD-sentiment$_{i,d}$	-2696.3***	-13329.3***	-26549.5***	-55517.0***
	(-5.63)	(-5.59)	(-5.59)	(-5.61)
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 6 AI Democratization and Trading

This table presents how retail-AI alignment and short-sellers-AI alignment change around AI democratization. Panel A (B) reports the alignment of retail trading (short selling) with AI-sentiment before and after AI democratization, using a short sample in 2022-2023. Retail trading (multiplied by 100000) is abnormal retail holding calculated in Section 1.3. Short selling (multiplied by 100000) is abnormal short-seller 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 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 and HD-sentiment 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: Retail Trading

	Retail Trading$_{i,[horizon]}$			
	(1) **[d]**	**(2)** **[d+1:d+5]**	**(3)** **[d+1:d+10]**	**(4)** **[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

Panel B: Short Selling

	Short Selling$_{i,[horizon]}$			
	(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

Table 7 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. StandardizedTrading includes both retail trading and short selling. We invert the sign of ShortSelling and separately standardize both ShortSelling and RetailTrading. $\text{Treat}^{\text{IsRetail}}$ is a dummy variable that equals 1 if the trading variable for a stock in a given quarter is retail trading and 0 if it is short selling. 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 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 and HD-sentiment 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$_{i,g,[\textbf{horizon}]}$			
	(1) **[d]**	**(2)** **[d+1:d+5]**	**(3)** **[d+1:d+10]**	**(4)** **[d+1:d+21]**
AI-sentiment$_{i,d}$	0.017	0.018	0.020*	0.022*
	(1.64)	(1.77)	(1.95)	(2.18)
AI-sentiment$_{i,d}$ × $\text{Treat}_g^{\text{IsRetail}}$	-0.044*	-0.042*	-0.043*	-0.045*
	(-2.19)	(-2.08)	(-2.20)	(-2.30)
AI-sentiment$_{i,d}$ × After_d	-0.035*	-0.036**	-0.039**	-0.039**
	(-2.36)	(-2.42)	(-2.54)	(-2.63)
AI-sentiment$_{i,d}$ × $\text{Treat}_g^{\text{IsRetail}}$ × 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 8 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. StandardizedTrading includes both retail trading and short selling. We invert the sign of ShortSelling and separately standardize 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. $\text{Treat}^{\text{IsRetail}}$ is a dummy variable that equals 1 if the trading variable for a stock in a given quarter is retail trading and 0 if it is short selling. 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 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 and HD-sentiment 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$_{i,g,[\text{horizon}]}$			
	(1) [d]	**(2)** [d+1:d+5]	**(3)** [d+1:d+10]	**(4)** [d+1:d+21]
AI-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-sentiment$_{i,d}$ $\times$ Treat$_g^{\text{IsRetail}}$	-0.044*	-0.041*	-0.043*	-0.045*
	(-2.19)	(-2.08)	(-2.19)	(-2.30)
AI-sentiment$_{i,d}$ $\times$ After$_d$	-0.036**	-0.037**	-0.039**	-0.039**
	(-2.37)	(-2.43)	(-2.55)	(-2.64)
AI-sentiment$_{i,d}$ $\times$ After&Outage$_{i,d}$	0.017*	0.017*	0.016*	0.022**
	(2.07)	(2.14)	(2.12)	(2.43)
Treat$_g^{\text{IsRetail}}$ $\times$ After&Outage$_{i,d}$	0.10**	0.095**	0.096**	0.078**
	(3.25)	(3.12)	(3.03)	(3.23)
AI-sentiment$_{i,d}$ $\times$ Treat$_g^{\text{IsRetail}}$ $\times$ After$_d$	0.091**	0.090**	0.090**	0.088**
	(2.69)	(2.60)	(2.62)	(2.70)
AI-sentiment$_{i,d}$ $\times$ Treat$_g^{\text{IsRetail}}$ $\times$ 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 9 Information Asymmetry, AI Democratization, and ChatGPT Outages: 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$\times\text{Treat}^{\text{IsRetail}}$ $\times$ After, obtained by running regressions in Table 7 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. 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. 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 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 and HD-sentiment 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: Information Asymmetry and AI Democratization

	Bid-Ask Spread$_{i,[horizon]}$			
	(1) [d]	**(2)** [d+1:d+5]	**(3)** [d+1:d+10]	**(4)** [d+1:d+21]
$\text{Treat}_i^{\text{AI}}$	46.1*	16.6	17.4	15.0
	(2.31)	(0.88)	(0.92)	(0.82)
$\text{Treat}_i^{\text{AI}} \times \text{After}_d$	-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	22-23	22-23	22-23	22-23
N	12502	12437	12425	12335
r2	0.37	0.46	0.48	0.49

Panel B: Information Asymmetry under ChatGPT Outages

	Bid-Ask Spread$_{i,[horizon]}$			
	(1) **[d]**	**(2)** **[d+1:d+5]**	**(3)** **[d+1:d+10]**	**(4)** **[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$_i^{\text{AI}}$	46.1*	16.6	17.4	15.0
	(2.31)	(0.88)	(0.92)	(0.82)
Treat$_i^{\text{AI}}$ $\times$ After$_d$	-27.1	-19.8	-22.4*	-26.7***
	(-1.80)	(-1.73)	(-2.05)	(-3.63)
Treat$_i^{\text{AI}}$ $\times$ 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	22-23	22-23	22-23	22-23
N	12502	12437	12425	12335
r2	0.37	0.46	0.48	0.49

Table 10 Return Predictability of AI-sentiment and FinBERT-sentiment

This table presents the return predictability of AI-sentiment and FinBERT-sentiment using post-final knowledge sample. Cumulative abnormal returns (CAR) are calculated following the DGTW method (Daniel et al., 1997). AI-sentiment is the sentiment score of earnings call transcripts computed by gpt-3.5-turbo-16k model. FinBERT-sentiment is the sentiment score of earnings call transcripts computed using a fine-tuned BERT model (FinBERT) tailored for financial text analysis. HD-sentiment is the sentiment score of earnings call transcripts computed by using Loughran-McDonald (LM) dictionary. Control variables include SUE, Beta, BM, MVE, Mom12m. 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 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, FinBERT-sentiment and HD-sentiment 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.

	$\mathbf{CAR}_{i,[horizon]}$								
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
	[d+1]	**[d+1:d+2]**	**[d+1:d+3]**	**[d+1:d+5]**	**[d+1:d+10]**	**[d+1:d+21]**	**[d+1:d+42]**	**[d+1:d+63]**	**[d+1:d+126]**
AI-sentiment$_{i,d}$	0.12**	0.15**	0.20**	0.20*	0.22	0.49***	0.56***	0.71***	0.83*
	(3.35)	(3.13)	(2.71)	(2.26)	(1.63)	(3.55)	(4.73)	(3.71)	(2.17)
HD-sentiment$_{i,d}$	-0.020	0.0087	0.0068	-0.054	-0.056	0.072	0.070	-0.34	-0.80
	(-0.22)	(0.08)	(0.05)	(-0.32)	(-0.25)	(0.22)	(0.13)	(-0.61)	(-1.02)
FinBert-sentiment$_{i,d}$	0.044	0.088	0.093	0.12	0.047	-0.12	-0.041	-0.41	-1.18
	(0.44)	(0.72)	(0.68)	(0.60)	(0.19)	(-0.35)	(-0.08)	(-0.68)	(-1.85)
Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter FE	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
Sample Period	21Oct-23Dec	21Oct-23Dec	21Oct-23Dec	21Oct-23Dec	21Oct-23Dec	21Oct-23Dec	21Oct-23Dec	21Oct-23Dec	21Oct-23Dec
N	18444	18444	18444	18444	18444	18444	18444	18444	18444
r2	0.0038	0.0045	0.0050	0.0059	0.0081	0.0068	0.0077	0.0092	0.011

Table 11 Return Predictability: Double Sort based on AI- and HD-sentiment

This table presents the return predictability estimates from regressions conducted within subgroups formed by double sorting on AI-sentiment and HD-sentiment using post-final knowledge sample. Each calendar quarter, AI-sentiment (HD-sentiment) is sorted into deciles, with the top three deciles defined as the AI High (HD High) group and the bottom three as the AI Low (HD Low) group. This classification creates four groups of earnings calls: AI High and HD High ($\text{High}_{\text{AI}}\text{High}_{\text{HD}}$), AI Low and HD Low ($\text{Low}_{\text{AI}}\text{Low}_{\text{HD}}$), AI High and HD Low ($\text{High}_{\text{AI}}\text{Low}_{\text{HD}}$), and AI Low and HD High ($\text{Low}_{\text{AI}}\text{High}_{\text{HD}}$). Cumulative abnormal returns (CAR) are calculated following the DGTW method (Daniel et al. 1997). Control variables include SUE, Beta, BM, MVE, and Mom12m. 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 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. 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.

	$\textbf{CAR}_{[horizon]}$								
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
	[d+1]	**[d+1:d+2]**	**[d+1:d+3]**	**[d+1:d+5]**	**[d+1:d+10]**	**[d+1:d+21]**	**[d+1:d+42]**	**[d+1:d+63]**	**[d+1:d+126]**
$\mathbf{High_{AI}High_{HD}}$	0.23**	0.39**	0.46***	0.40**	0.49**	0.83***	1.15***	0.97	0.65
	(2.00)	(2.45)	(2.71)	(2.02)	(2.12)	(2.83)	(2.68)	(1.52)	(0.64)
$\mathbf{Low_{AI}Low_{HD}}$	-0.20*	-0.30	-0.27	-0.18	-0.04	-0.24	-0.38	0.23	1.14
	(-1.78)	(-1.63)	(-1.42)	(-0.69)	(-0.09)	(-0.68)	(-0.68)	(0.40)	(1.09)
$\mathbf{High_{AI}Low_{HD}}$	-0.02	-0.13	-0.02	-0.07	0.04	0.09	-0.03	0.72	1.28
	(-0.21)	(-0.91)	(-0.08)	(-0.32)	(0.11)	(0.20)	(-0.05)	(0.86)	(0.99)
$\mathbf{Low_{AI}High_{HD}}$	-0.28**	-0.24	-0.24	-0.30	0.02	-0.66	-1.88**	-3.36***	-3.70***
	(-2.11)	(-1.43)	(-1.02)	(-1.02)	(0.03)	(-1.11)	(-2.55)	(-3.59)	(-2.59)

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 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 and HD-sentiment 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

Table IA2 Return Predictability of AI-sentiment and FinBERT-sentiment: Full Sample

This table presents the return predictability of AI-sentiment and FinBERT-sentiment using full sample. Cumulative abnormal returns (CAR) are calculated following the DGTW method (Daniel et al., 1997). AI-sentiment is the sentiment score of earnings call transcripts computed by gpt-3.5-turbo-16k model. FinBERT-sentiment is the sentiment score of earnings call transcripts computed using a fine-tuned BERT model (FinBERT) tailored for financial text analysis. HD-sentiment is the sentiment score of earnings call transcripts computed by using Loughran-McDonald (LM) dictionary. Control variables include SUE, Beta, BM, MVE, Mom12m. 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 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, FinBert-sentiment and HD-sentiment 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.

	$\mathbf{CAR}_{i,[horizon]}$								
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
	[d+1]	**[d+1:d+2]**	**[d+1:d+3]**	**[d+1:d+5]**	**[d+1:d+10]**	**[d+1:d+21]**	**[d+1:d+42]**	**[d+1:d+63]**	**[d+1:d+126]**
AI-sentiment$_{i,d}$	0.11***	0.12***	0.15***	0.16***	0.19***	0.30***	0.34***	0.50***	0.91***
	(6.97)	(5.42)	(5.21)	(4.56)	(3.57)	(4.62)	(3.67)	(4.14)	(5.05)
HD-sentiment$_{i,d}$	-0.0043	0.0074	0.016	0.00037	-0.066	-0.035	-0.042	-0.060	-0.064
	(-0.16)	(0.22)	(0.44)	(0.01)	(-0.92)	(-0.29)	(-0.24)	(-0.29)	(-0.20)
FinBERT-sentiment$_{i,d}$	0.063**	0.079**	0.072*	0.079	0.13	0.13	0.38*	0.32	0.16
	(2.39)	(2.15)	(1.68)	(1.32)	(1.59)	(0.84)	(1.86)	(1.19)	(0.39)
Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter FE	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
Sample Period	10-23	10-23	10-23	10-23	10-23	10-23	10-23	10-23	10-23
N	81278	81278	81278	81278	81278	81278	81278	81278	81278
r2	0.0044	0.0046	0.0050	0.0053	0.0055	0.0044	0.0044	0.0044	0.0034