Wisdom or Whims? Decoding Investor Trading Strategies with Large Language Models*

Shuaiyu Chen[†] Lin Peng [‡] Dexin Zhou[§]

September 29, 2024

Abstract

Using large language models, we analyze trading strategies expressed in over 77 million messages on a leading investor social media platform. We find that stocks experiencing bullish sentiment in technical analysis (TA) posts tend to have lower future returns and a higher likelihood of buy herding on Robinhood. In contrast, sentiment extracted from fundamental analysis (FA)-related posts positively predicts future returns. More intense TA posting is associated with less informative retail order flows, whereas FA posting is positively linked to flow informativeness. We further show that social media TA sentiment tends to contradict signals derived from a state-of-the-art AI-based technical strategy, and the profitability of the AI strategy largely stems from exploiting the TA sentiment. Our findings provide insights into the investment approaches of retail investors, the role of social media, and the interactions between different market players in the era of social media and AI-powered trading.

Keywords: AI, Large language models, Social media, Retail investors, Herding, Technical analysis, Fundamental analysis

---

*We thank Dacheng Xiu for his helpful comments and for generously providing the data of the machine-learning-based technical signals. We also appreciate suggestions from Zhi Da, Winston Dou, Clifton Green, Ron Kaniel, John Shim, and seminar participants at Purdue University, the University of Notre Dame, and Binghamton University.

[†]chen4144@purdue.edu, Mitch Daniels School of Business, Purdue University.

[‡]lin.peng@baruch.cuny.edu., Zicklin School of Business, Baruch College.

[§]dexin.zhou@baruch.cuny.edu., Zicklin School of Business, Baruch College.

1. Introduction

There has been renewed interest in the role of retail investors in financial markets. The advent of fintech brokerage platforms and social media sites like Reddit has spurred increased stock market participation by retail investors, raising important questions about their decision-making processes and the impacts of their trading activities.[1] Simultaneously, many of the fintech brokerages engage in payment for order flow (PFOF), which channels a significant portion of retail trades to sophisticated AI-powered high-frequency trading firms (e.g., Ernst and Spatt, 2022). These dynamics thus pose new regulatory challenges for policymakers concerning the impact of AI on financial markets, including, for example, how AI-powered traders interact with retail investors (Dou et al., 2024).

This paper seeks to provide insights into the behavior and impact of retail investors by analyzing the strategies they employ and their interactions with AI-powered strategies. Characterizing retail investors' investment strategies is challenging, as these investors represent a diverse population with heterogeneous social economic backgrounds.[2] A promising trend in recent literature is the use of surveys to gain insights into investor beliefs.[3] We propose an alternative approach by applying large language models (LLMs) to analyze rich, real-time investor social media data. This enables us to infer the trading strategies retail investors use, identify the factors influencing these strategies, and explore their implications for financial markets. We document substantial heterogeneity in retail investor strategies, and their varying levels of informativeness, and provide novel evidence on how retail investors interact with sophisticated AI-powered strategies.

Our social media dataset consists of over 77 million messages posted by nearly

[1]For example, while evidence shows that retail investors are noise traders vulnerable to behavioral biases (see, e.g., Barber and Odean, 2000; Kumar and Lee, 2006; Barber and Odean, 2008; Barber et al., 2022; Bryzgalova et al., 2023), another body of research suggests that retail orders can correctly predict future returns or provide liquidity (see, e.g., Kaniel et al., 2008; Kelley and Tetlock, 2013; Boehmer et al., 2021; Welch, 2022).

[2]E.g., Barber and Odean (2000) document a few dimensions of retail investor diversity.

[3]See, for example, Choi and Robertson (2020), Giglio et al. (2021), Chinco et al. (2022), Liu et al. (2022), Jiang et al. (2024b), and Laudenbach et al. (2024).

800,000 users on StockTwits, a leading investor platform, from 2012 to 2022. We leverage LLMs to decipher the trading strategies employed by investors on this platform.

We first focus on the extent to which StockTwits users engage in technical trading strategies, which aim to predict future outcomes based on past trends, reversals, cycles, and other identifiable price patterns.[4] We find that a considerable portion of posts mention technical rules—28% of the strategy-related messages reference technical trading. As expected, the use of TA is more common among investors who self-report using technical or momentum trading styles. However, we also observe significant time-series variation in an individual user's use of TA. Notably, we observe that the release of earnings announcements, analyst forecasts, and recommendations significantly reduces TA-related posts, suggesting that StockTwits users are more likely to rely on TA when other sources of information are scarce.

The prevalence of technical strategies among retail investors is particularly relevant in today's markets, given the rise of AI-driven trading. In particular, recent theoretical work by Dou et al. (2024) shows that the interaction between AI-powered sophisticated traders and naive, "information insensitive" investors, who rely solely on technical analysis, can have significant implications for price efficiency.[5]

Motivated by the model, we empirically investigate its implications for the interactions of these two types of traders. To capture AI-powered trading by sophisticated investors, we utilize the state-of-the-art AI-based technical signal from Jiang et al. (2023). First, we confirm that a long-short strategy based on this signal yields a significantly positive abnormal return of 10.2% per annum in our sample. More

[4]This strategy is deeply rooted in antiquity and human psychology and has been widely used throughout the history of financial markets (see Lo and Hasanhodzic 2010 for a review), making it an ideal candidate for analyzing the decision-making of retail investors.

[5]On September 8, 2023, Nasdaq announced SEC approval for launching the Dynamic Midpoint Extended Life Order (M-ELO), becoming the first exchange to offer an AI-powered order type. This move reflects a growing trend among major industry players like MetaTrader, Two Sigma, BlackRock, and J.P. Morgan, who are integrating AI into their trading processes. Policymakers, regulators, and financial market supervisors worldwide are focusing on understanding AI's application in trading, its impact on markets, and any unintended consequences. See SEC (2023) and Bagattini et al. (2023) for examples of regulatory efforts on the application of AI technologies in securities markets.

importantly, we document a significant *negative* correlation between the sentiment of the AI signal and the StockTwits TA sentiments. Given that both the AI signal and the StockTwits TA sentiment are grounded in technical strategies, this negative relationship raises important questions about how these two types of traders might interact.

We investigate this question by examining how the profitability of the AI signal varies with StockTwits TA sentiment. We find that the AI signal generates particularly large profits when trading *against* TA sentiment, with an annualized return of 13.40%, a 31% increase over the unconditional AI portfolio return. In sharp contrast, the AI signal is no longer profitable when trading in the same direction as the TA sentiment. These findings provide empirical support for a key feature of Dou et al. (2024): a major driver of AI profits is the exploitation of retail sentiment associated with the misuse of technical signals.

We further investigate a potential mechanism behind this misuse. As suggested by Stein (2009), when individual traders are unable to determine in real time how many others are using the same strategy and taking similar positions, it creates a coordination problem that can lead to overcrowding in the strategy and, eventually, crashes. Additionally, Barber et al. (2022) provide empirical evidence that retail buy herding can lead to price bubbles and subsequent crashes. Building on these insights, we examine whether the TA sentiment on StockTwits is associated with retail investor herding on technical strategies.

Specifically, we examine herding episodes on Robinhood, a leading zero-commission broker whose active user base grew rapidly from half a million in 2014 to 17.3 million in 2021. Following Barber et al. (2022), we identify retail herding episodes based on sharp increases in the number of new Robinhood investors in a given stock. We indeed find a strong positive relationship between StockTwits TA sentiment and buy-herding episodes. These results suggest that intense discussions of TA signals on online social media platforms may have contributed to overcrowding in the use these signals, thereby driving the observed herding episodes.

Our approach to identifying retail investment strategy usage extends beyond technical trading rules. We also examine retail investors' use of fundamental analysis, another key class of strategy that focuses on fundamental information such as earnings trends and firm valuation. We similarly apply LLMs to identify StockTwits messages that describe fundamental analysis (i.e., FA messages). We find that about 35% of strategy-related messages mention fundamental analysis, with these messages more frequently posted by self-identified fundamental investors.

Next, we assess the sentiment conveyed by the TA messages and evaluate their informativeness by associating the sentiment with future stock returns. While the average StockTwits sentiment correctly predicts one-week-ahead returns (Cookson et al., 2024b), we uncover substantial heterogeneity across different strategy categories. Specifically, the sentiment of TA messages *negatively* predicts future returns, whereas the sentiment of FA messages is a *positive* predictor. Strategy-related messages that are neither TA nor FA show no significant power in predicting future returns.

Finally, we show that StockTwits message activities and sentiments are not just a sideshow; they are significantly correlated with the aggregate retail trading activity of the corresponding stocks. To explore this, we examine retail order imbalances. Consistent with Boehmer et al. (2021) and Barber et al. (2023), we confirm that retail net buys positively predict one-week-ahead returns, validating the informativeness of retail order flows.

However, we find that the informativeness of retail order flows on a stock is virtually eliminated during periods of intensive TA usage on StockTwits. In contrast, retail informativeness exhibits higher informativeness when more discussions focus on fundamental analysis. This result suggests that the informativeness of retail order flows is more nuanced than previously understood—when retail investors heavily engage in discussions of technical patterns on social media platforms, their order flows tend to be dominated by noise trading. Thus, our results highlight the diversity of investment approaches employed by retail investors and their differential informativeness.

Our study contributes to several strands of literature. First, it extends the aforementioned literature on retail trading and its informativeness. Recent research has reignited interest in retail investors as zero-cost trading platforms have attracted a large number of new investors to financial markets (e.g., Barber et al., 2022; Welch, 2022; Eaton et al., 2022). However, the investment approaches and strategies employed by these retail investors have not been well understood. Our research advances this literature by directly extracting investor strategies from their own words and linking these strategies to their trades. We provide the novel finding that retail order flow informativeness is contingent on the types of dominant strategies as reflected on popular social media platforms and that retail investors' herding on technical signals creates profitable opportunities for sophisticated traders.

We also contribute to the literature on technical analysis.[6] Building on previous studies that use price patterns to predict future returns, an emerging set of new studies reexamines the potential effectiveness of using price and volume patterns to predict future returns (e.g., Han et al., 2013, 2016; Jiang et al., 2023; Murray et al., 2024). We provide new evidence on the heterogeneity of investors' ability to use such analysis—while AI can exploit the analysis to a great extent, retail investors may perform poorly.

Importantly, our documented disparity in investors' ability to benefit from technical analysis, and social media's potential role in amplifying this disparity, highlights critical concerns about the fairness and integrity of financial markets in the era of rapid AI adoption by sophisticated players. Striking the right balance between innovation and regulation will be key to harnessing the benefits of AI while mitigating its potential downsides (see, e.g., Dou et al. 2024). Regulators and exchanges must carefully consider how to address this imbalance and ensure adequate protection in this new environment. Our findings, which provide new evi-

[6] For earlier studies, see, for example, Brown and Jennings (1989); Jegadeesh (1991); Brock et al. (1992); Jegadeesh and Titman (1993); Blume et al. (1994); Lo et al. (2000); George and Hwang (2004).

dence on the determinants of retail investors' strategies and their interactions with AI-powered strategies, can inform policy considerations regarding the regulation of AI-powered trading, investor education, and investment social media platforms.

Another strand of recent research underscores the role of social networks in shaping retail investors' decisions and uses social media data as a lens to infer investor decision-making.[7] A closely related paper is that of Cookson et al. (2024b), who finds that user sentiment on investor social media platforms positively predicts one-day-ahead returns. Our key contribution is identifying significant heterogeneity in the informativeness of social media sentiment—while sentiment from posts focused on fundamental analysis *positively* predicts future returns, sentiment from technical analysis posts *negatively* predicts them, with this predictability lasting up to a week. We also show that these sentiments are strongly associated with investor herding episodes and the profitability of sophisticated investors, suggesting that social media sentiments play a potentially important role in shaping equilibrium prices and trading.

Finally, our study adds to an increasing number of papers that use LLMs to answer economics and finance questions (e.g., Korinek, 2023).[8] We show that such tools, when applied to rich social media data, provide powerful inferences that help us better understand investors' decision-making. Our paper also illustrates a novel, relatively fast, and cost-effective way of implementing LLMs—instead of purely relying on cutting-edge LLMs, one can first generate useful examples using state-of-

[7]Several studies utilizing data from online social networks have shown that the posting activity and message quality of retail investors help predict stock returns and trading volume (e.g., Antweiler and Frank, 2004; Chen et al., 2014). These studies also highlight how investor disagreement and echo chambers influence belief formation (e.g., Giannini et al., 2018, 2019; Cookson and Niessner, 2020; Cookson et al., 2023), how the dissemination of informative content can be affected (e.g., Chen and Hwang, 2022; Farrell et al., 2022; Bradley et al., 2024), and the role of investor horizon differences (Cookson et al. 2024a). Furthermore, there is growing interest in the skill and role of influencers in these networks (e.g., Coval et al., 2021; Kakhbod et al., 2023; Hirshleifer et al., Forthcoming). The review of social media and finance of Cookson et al. (2024c) summarizes this emerging line of research.

[8]For example, Jiang et al. (2024a) and Lopez-Lira and Tang (2023) use LLMs to predict future returns. Li et al. (2023) extract corporate culture from analyst reports. Jha et al. (2024) extract information related to corporate investments, and Eisfeldt et al. (2023) investigate which jobs are more replaceable with the advent of GPTs. Huang et al. (2024) use LLMs to examine the narratives on investor social media.

the-art (SOTA) LLMs and then use these examples to fine-tune a smaller language model.

2. Data

Our sample consists of US common stocks (SHRCD = 10 or 11) traded on the NYSE, AMEX, and NASDAQ for the period from 2012 to 2022. We obtain investor social media data from StockTwits, stock data from CRSP, and other accounting and financial statement variables from the merged CRSP-Compustat database.

2.1. StockTwits Data

StockTwits is a leading social media platform for retail investors to share their opinions and exchange ideas about stocks, ETFs, and Cryptos. Similar to Twitter, StockTwits users can post short messages, initially capped at 140 characters until May 8, 2019, and later expanded to 1,000 characters. Unique to StockTwits is its emphasis on financial markets, with users tagging their posts with "cashtags" (e.g., $TSLA) to indicate the ticker symbols of stocks mentioned in the message.

Using StockTwits API, we downloaded 157,674,830 messages posted by 948,867 users between 2012 and 2022.[9] We obtain detailed information at the message level, such as timestamps, content, and self-labeled sentiment "bullish" or "bearish." At the user level, we obtain characteristics such as trading style (Technical, Momentum, Fundamental, Value, Growth, and Global Macro), investment horizon (Day Trader, Swing Trader, Position Trader, and Long-Term Investor), and the level of trading experience (Novice, Intermediate, and Professional).

Following Cookson and Niessner (2020) and Cookson et al. (2024b), we perform several steps to select valid messages that are specific to a public firm and are posted by human users (as opposed to bots). Specifically, we first select messages that explicitly mention only one company. We then exclude users who post more

[9]The API is available at https://api.StockTwits.com/developers.

than 1,000 messages in a single day and exclude messages sourced from third-party platforms, which are likely redistributions of financial news or written by algorithms. Finally, we require both user identifier and username to be non-missing.

2.2. Identifying Trading Strategies From Messages

Leveraging large language models, we first decipher the strategies conveyed in the message content. Specifically, we identify messages related to technical and fundamental messages, two important types of investment strategies. We also attempt to identify messages that use other investment strategies. We use the same procedure to accomplish these three classification tasks. First, we leverage the cutting-edge large language model from OpenAI, GPT-4, to identify if a message contains an investment strategy (or TA/FA).[10] As with many social media messages, those on StockTwits tend to be short, with many abbreviated and colloquial words, and with many non-standard spelling. Thus, it is difficult to identify trading strategies purely based on a dictionary. Moreover, given that trading strategies are highly diverse, identifying messages containing trading strategies can be a highly challenging task.

While we find that GPT-4 has an excellent ability to identify these strategy-related messages, and those identifications tend to align with our own judgment, it is infeasible to use GPT-4 to classify all the messages in our sample due to the limited throughput and high costs. Thus, we use the examples generated by the cutting-edge GPT-4 model to fine-tune a smaller classification model.[11]

Next, we illustrate our classification procedure by identifying TA-related messages. We first randomly sample $20,000$ messages from the sample of messages.[12] We then ask the GPT-4 to determine whether the message entails technical trading

[10]Specifically, we use the GPT 4-Turbo model (gpt-4-0125-preview endpoint).

[11]This approach, first proposed by Hinton (2015), is widely known as knowledge distillation (KD) in the machine learning literature. See Gu et al. (2023) for a more recent review of this approach and its applications in LLMs.

[12]To achieve a more balanced sample, $10,000$ messages are sampled from users with a self-declared technical investment style and the other $10,000$ are from the other groups.

using the following prompt:

> You have a deep understanding of the language of social media and financial markets. Please analyze the message from an investor social media platform. Please parse the message along two dimensions. 1) Presence of technical analysis (0=no, 1=possibly, 2=likely). 2) if technical analysis is used, what is the technical indicator? (output the indicator or "" if you cannot locate it. If multiple signals exist, please separate by a comma) Output in JSON format: {"technical_analysis":, "technical_indicator": }.

We collect GPT's response for the $20,000$ messages. Table A.1 in the appendix provides a sample of positive and negative responses by GPT. Then, we use these responses to fine-tune a BERT model (henceforth TechBERT) to provide a prediction of whether a message uses technical analysis. Through cross-validation, we find that the fine-tuned TechBERT model can achieve an F1 score of 0.83, which indicates a high level of performance. Since BERT has a drastically smaller parameter count, we are able to run this model locally to provide a probabilistic prediction of whether each message contains technical trading.[13]

[Insert Figure 1 and Figure 2 near here]

We visualize TechBERT's classification results in Figures 1 and 2 across investor types depending on their self-labeled horizons and investment approaches. We find that TechBERT's classification prediction exhibits some desirable properties, as most of the messages fall either in the low probability region (i.e., $< 5\%$) or high probability region (i.e., $> 95\%$). This result shows that TechBERT's prediction is quite unambiguous.

[Insert Figure 3 near here]

[13] BERT has established itself as a state-of-the-art tool for many natural language processing tasks, including classification (Devlin et al., 2018). González-Carvajal and Garrido-Merchán (2020) show that BERT achieves superior performance compared to traditional natural language processing tools that do not rely on deep learning.

Figure 3 presents word cloud plots to exhibit the high-frequency unigram and bigram in the TA messages, respectively. In the unigram plot (Panel A), there are several striking patterns. First, technical messages often contain analyses of charts, consistent with the finding in Jiang et al. (2023). Second, we can see many familiar technical terms, such as resistance, support, and gap. In the bigram plot (Panel B), besides other common terms referring to technical signals, we also find terms related to horizons, such as short-term and next week. These terms inform us that our research design should focus on short-term returns at the weekly horizon.

As a second validation, we show that the reliance on technical signals is more pronounced for self-declared investor characteristics that are often associated with technical analysis. For example, as shown in Figure 1, the TechBERT model finds a higher fraction of TA usage among the self-declared day trader and swing trader groups compared to long-term investors. When we classify messages by the user-declared investment approach, Figure 2 shows a higher usage of technical analysis among self-declared technical and momentum traders compared to fundamental, value, growth, or global macro investors.

Notably, these figures also reveal the heterogeneity in the investment styles even within the same group of users. These results highlight the importance of conducting analyses at the message level to more accurately identify the specific approaches investors implement under different market conditions. The focus on the message-level investment strategy distinguishes our study from related work (e.g., Cookson and Niessner, 2020) and allows us to provide more granular and time-varying measurements of investment strategies.

We follow the same approach to identify the messages that contain an investment strategy and messages containing fundamental analysis. We only need to revise the prompt that we feed into GPT-4. Then, we fine-tune specialized BERT models to help identify strategy-related messages and fundamental analysis-related messages.[14] Overall, we find that approximately a quarter of all messages involve discussions related to trading strategies. TA and FA messages comprise roughly 36% and 44% of the strategy-related messages, respectively.

2.3. Other Variables

We follow Cookson and Niessner (2020) to assign a sentiment score to each StockTwits message.[15] Specifically, we randomly select 10,000 messages with a self-declared bullish or bearish label. Then, we train a sentiment classifier using the maximum entropy method. The classifier delivers a probabilistic prediction on whether a given message is bullish, and we apply this classifier to messages without sentiment labeling as bullish or bearish.[16]

The list of other stock variables and firm characteristics, as well as their construction, are listed in Table A.2.

Our final sample consists of 77,575,573 messages across 765,512 unique users for 5,872 stocks during the period of 2012 through 2022. Our subsequent analyses are conducted at the investor-stock-calender week level. Panel A in Table 1 reports summary statistics in the investor-stock-week sample. We also construct a stock-week level sample for additional tests. Panel B reports summary statistics for this sample.

[14] For overall strategy, we use the following prompt: *You have a deep understanding of the language of social media and financial markets. Please analyze the message from an investor social media platform. Please parse the message along two dimensions. 1) Presence of investment strategy (e.g., technical analysis, fundamental analysis, event-driven strategy, arbitrage strategy). If true, please answer 1, otherwise 0. 2) if a strategy is identified, please specify the strategy Output in JSON format: "has_strategy":, "strategy_type": .*

For fundamental analysis, we use the following prompt: *You have a deep understanding of the language of social media and financial markets. Please analyze the message from an investor social media platform. Please parse the message along two dimensions. 1) Presence of fundamental analysis (0=no, 1=possibly, 2=likely). 2) If fundamental analysis is used, select one of the following 15 topics that is most relevant: 'acquisitions-mergers','analyst-ratings','assets','bankruptcy','credit','credit-ratings','dividends','earnings','equity-actions','investor-relations','labor-issues','marketing','price-targets','products-services','revenues'. Output in JSON format: "fundamental_analysis":, "fundamental_topic":.*

[15] In StockTwits, users have the option to declare sentiment when posting a message. However, not all messages contain the self-declared sentiment flag.

[16] We use the messages with a bullish/bearish flag that are not in the training sample to conduct model validation. Our classifier achieves an F1 score of 0.9, which indicates the high accuracy of our model.

[Insert Table 1 near here]

3. Technical Analysis

Recent literature has highlighted the potential of technical analysis in generating profitable trading signals (e.g., Jiang et al., 2023). Moreover, Dou et al. (2024) argue that AI-powered algorithmic investors may exploit retail investors' misuse of technical rules. In this section, we analyze the tendency of retail investors to use technical analysis and how they interact with the technical signal extracted using state-of-the-art AI in Jiang et al. (2023).

3.1. Determinants of Retail Technical Usage

We first validate that self-reported technical and short-term investors tend to use more technical analysis in their investment decisions. We also investigate what market and information environments lead to higher usage of technical analysis. To that end, we estimate the following regression at the stock-investor-week level:

$$
\begin{aligned}
&\textit{Retail TA Usage}_{i,j,t+1} \\
=&\beta_1 \textit{Technical Investor}_{j,t} + \beta_2 \textit{Short-term Investor}_{j,t} + \beta_3 \textit{Professional Investor}_{j,t} \\
+&\beta_4 \textit{Earnings News}_{i,t} + \beta_5 \textit{Analyst News}_{i,t} + \gamma \boldsymbol{X_{i,t}} + FEs + \epsilon_{i,j,t+1}.
\end{aligned}
\tag{1}
$$

where the dependent variable is the TA usage by investor j on stock i in week $t+1$, measured as the percentage of investor j's messages classified as TA-related by our classification LLM (TechBERT).

We consider two factors potentially affecting retail TA usage: trader types and news releases. StockTwits users self-report their types for their investment approach, holding horizon, and experience. Technical Investor$_{j,t}$ is a dummy variable equal to one if investor j's self-reported approach is "*Technical*" or "*Momentum*". Short-term Investor$_{j,t}$ is a dummy variable equal to one if investor j's self-reported horizon is "*Day Trader*", "*Swing Trader*", or "*Position Trader*". Professional $\text{Investor}_{j,t}$ is a dummy variable equal to one if investor j's self-reported experience is "*Professional*". Earnings $\text{News}_{i,t}$ is an indicator of the release of earnings-related news on stock i in week t. Analyst $\text{News}_{i,t}$ is an indicator of the release of analyst forecasts and recommendations on stock i in week t. $X_{i,t}$ represents a vector of stock characteristics available at the beginning of week t, including (log) market capitalization, (log) book-to-market, asset growth, gross profits-to-asset, (log) number of analysts, (log) institutional ownership, the maximum daily return in the prior month, and abnormal turnover. Standard errors are clustered by investor and calendar week.

[Insert Table 2 near here]

The regression results are reported in Table 2. Column (1) shows that users with a self-declared technical investment approach and a short-term investment horizon tend to use technical analysis more frequently. Specifically, self-declared technical investors post 9.7% more TA messages on a stock in a given week than investors using other approaches, and short-term investors post 2.1% more than long-term investors. These results validate our LLM-based classification method and confirm the specialization of investment approaches among retail investors, as we would expect users with a self-declared technical approach and short horizon to use technical analysis more intensely, echoing the findings in Cookson et al. (2024a).[17] We also find that self-reported professional investors have a higher tendency to post messages with technical analysis. This is likely due to professional investors having a better knowledge of the technical approach. These patterns are also consistent with the graphical evidence presented in Figures 1 and 2.

Column (2) considers the release of firm-level news, including earnings news and analyst recommendations from financial media outlets such as the Wall Street Journal, CNBC, or Reuters. These variables capture the arrival of fundamental

[17] They show that short-horizon predictions on Motley Fool's CAPS forum tend to contain more technical vocabulary.

or value-relevant information flows. We find that StockTwits investors tend to use more technical analysis when there is less news, suggesting that technical analysis substitutes for fundamental analysis and other approaches when there is a lack of fundamental news. In column (3), we add investor fixed effects and remove the self-reported investment approach and holding horizon indicators, as both are very persistent at the investor level. The coefficients for fundamental and analyst news releases remain qualitatively similar to those observed in columns (1) and (2).

In summary, our analyses highlight significant heterogeneity in users' tendency to discuss technical signals on social media. We find substantial variation among investors, with self-declared technical-style investors, short-term investors, and professional investors driving most of the usage. Furthermore, the availability of news coverage, especially related to firm fundamentals, tends to reduce investors' need for technical analysis.

3.2. Retail Investor vs. AI in Technical Analysis

Technical signals have been widely adopted in algorithmic trading over the past 30 years. The popularity of technical analysis in quantitative hedge funds indicates the profitability of exploiting past price patterns. Since the seminal work of Jegadeesh and Titman (1993), the discovery of technical signals and their implications for market efficiency and economic theories has been a crucial topic in financial economics. The advances in AI and big data in the past decade have fueled more academic finance studies in this regard. For example, Jiang et al. (2023) recently proposed a powerful technical trading signal using machine learning techniques to analyze price and volume charts. The evidence suggests that skillful technical analysis generates excellent returns.

However, it remains questionable whether retail investors can apply technical analysis skillfully. We address this question by examining whether retail investors' technical analyses reach conclusions similar to those offered by AI and whether retail technical analysis strategies lead to profitable trades.

3.2.1. Do Retail Investors Agree with AI in Technical Analyses?

We first investigate whether retail investors using technical analysis tend to reach conclusions similar to those obtained by AI. Specifically, we estimate the following panel regression at the stock-investor-week level:

$$\begin{aligned} Sentiment_{i,j,t}^{type} =& \beta_1 AI\ Signal_{i,t} + \beta_2 Return_{i,t-1}^{1d} + \beta_3 Return_{i,t-1}^{1w} + \beta_4 Return_{i,t-1}^{1m} \\ &+ \beta_5 Return_{i,t-1}^{1q} + \beta_6 Return_{i,t-1}^{1y} + \beta_7 Earnings\ News_{i,t-1} \\ &+ \beta_8 Analyst\ News_{i,t-1} + \gamma \mathbf{X_{i,t}} + \delta_t + \eta_j + \epsilon_{i,j,t}, \quad type = TA, NonTA, \end{aligned} \tag{2}$$

where the dependent variable is the sentiment of investor j on stock i in week t, measured by the number of investor j's bullish and bearish messages. Specifically, sentiment is defined as $Sentiment_{i,j,t} = \frac{N_{i,j,t}^{Bullish} - N_{i,j,t}^{Bearish}}{N_{i,j,t}^{Bullish} + N_{i,j,t}^{Bearish}}$ (see Cookson and Niessner, 2020). We consider two measures of sentiment based on the messages classified into technical analysis (TA) and non-TA categories. We then calculate the TA and non-TA sentiment of investor j on stock i in week t, using j's technical and non-technical messages, respectively.

The key independent variable, $AI\ Signal_{i,t}$, is the stock-level estimate for the probability of a positive return in the subsequent week $t+1$, developed in Jiang et al. (2023) from training a convolutional neural network (CNN) on image data representing the price pattern over the preceding five days.[18]

[Insert Table 3 near here]

Table 3 reports the results of this regression analysis, with columns (1) and (2) corresponding to TA sentiment, and columns (3) and (4) corresponding to non-TA sentiment, respectively. In columns (2) and (4), the regressions include the cumulative returns over the past day, one week, one month, six months, and twelve months. All specifications include the control variables from Table 2, as well as investor and calendar week fixed effects.

[18] The signal is denoted as $CNN5d5p$ in Jiang et al. (2023) and is available until 2019. The authors thank Dacheng Xiu for generously providing the data.

We find that retail investors' sentiment is negatively associated with the AI return forecast, and this negative relationship is substantially stronger for TA sentiment. These results suggest that retail investors on StockTwits tend to disagree with AI predictions, especially those investors who rely on technical analysis.[19] Given that the AI signal is also based on past price movements, the sharp disagreement between StockTwits retail investors and AI highlights striking heterogeneity in investors' ability to interpret technical patterns.

3.2.2. Returns to Retail and AI Technical Strategies

In this subsection, we use retail investors on StockTwits as a lens to infer retail strategies and trading behaviors, and to investigate the extent to which sophisticated technical investors employing quantitative methods and AI technologies interact with retail investors.

Our analysis uses the AI signal from Jiang et al. (2023) as a proxy for sophisticated technical trading and the various StockTwits sentiment measures as proxies for retail technical trading strategies. We examine the profitability of retail TA sentiment and the AI signal, as well as their interactions, by forming univariate and double-sorted portfolios, respectively. We focus on weekly return predictability, motivated by the word cloud plot in Figure 3, which prominently highlights the phrase "next week." Additionally, content analyses of TA-related posts suggest that StockTwits users tend to focus on short-term horizons. Table 4 reports the results from univariate and double sorts.

[Insert Table 4 near here]

Panel A presents the benchmark results of univariate portfolios sorted by the TA sentiment and the AI signal, respectively. "Retail TA Bull" denotes the portfolio

[19] Beyond establishing this contemporaneous negative relationship between the TA sentiment and AI signal, Table A.3 in the appendix shows that TA sentiment is also significantly and negatively related to the lagged AI signal. In addition, as a robustness check, we measure investors' sentiment in Table A.4 based solely on their self-reported bullish and bearish labels in messages, yielding similar results.

of stocks with retail TA sentiment greater than the cross-sectional median. "Retail TA Bear" and "Bull–Bear" are defined similarly. "AI Buy" represents the portfolio of stocks with the AI signal greater than 0.5. Additionally, we create portfolios for "AI Sell" and "Buy-Sell". The portfolio returns are based on the average DGTW-adjusted returns of equal-weighted portfolios.

Column (3) shows that a long-short portfolio based on retail TA sentiment yields an abnormal return of -6.48% per annum. In contrast, column (6) indicates that a long-short portfolio based on the sentiment of non-TA messages generates a positive and significant return of 5.33% per annum.

[Insert Table 4 near here]

These results are confirmed in Figure 4, which shows that a one-dollar initial investment at the start of our sample period would grow to $1.30 by investing in bearish retail TA stocks, while the same one-dollar would decline to less than $0.60 if invested in bullish retail TA sentiment stocks. A long-short strategy of buying bearish retail TA stocks and shorting bullish retail TA stocks would turn a one-dollar investment into $2 by the end of the sample period. These findings suggest that StockTwits users have not been able to generate profitable signals using technical analysis. A skillful investor, therefore, could benefit by paying attention to StockTwits TA signals and trading contrary to them.

Column (9) in Panel A of Table 4 shows that a portfolio formed by the AI signal produces a positive and significant return of 10.20% per annum, consistent with the main findings in Jiang et al. (2023).

We next explore the asset pricing implications of the interaction between retail investors' TA usage and AI-powered trading strategy. To this end, we form two-way sorted portfolios based on the retail sentiment and the AI signal. In Panel B of Table 4, columns (1)–(4) present the average DGTW-adjusted returns of the four individual portfolios. The top portion of this panel focuses on the two-way portfolios sorted by the retail TA sentiment and AI signal.

Column (1) corresponds to a portfolio of stocks associated with bullish views from both the retail TA sentiment and the AI signal, which yields an abnormal return of -1.21% per year. Column (2) refers to stocks that are associated with bullish sentiment according to retail TA messages but are expected to have negative subsequent returns according to the AI signal, showing an annualized return of -8.17%. Stocks in column (3) are associated with bearish sentiment according to retail TA posts but are recommended by the AI as buying opportunities, with this portfolio generating an annualized return of 5.22%. Finally, column (4) refers to a portfolio with bearish views from both retail TA sentiment and the AI signal, yielding a return of 2.44% per year.

Columns (5) and (6) present the returns to long-short strategies trading on the interactions between the retail TA sentiment and the AI signal. Column (5) refers to a strategy that only follows the AI signal when it aligns with retail TA sentiment (i.e., Buy/Bull–Sell/Bear). Column (6) corresponds to a strategy that follows the AI signal only when it disagrees with retail TA sentiment (i.e., Buy/Bear–Sell/Bull). As indicated in column (6), the AI signal generates substantially higher profits when it disagrees with retail TA sentiment. The magnitude is significant, with an annualized return of 13.4%. In contrast, when the AI and TA sentiment agree, as shown in column (5), the strategy yields an insignificant negative return of -3.66% per year.[20]

We also perform a similar two-way sorted portfolio analysis based on the AI signal and the non-TA sentiments, with the results presented in the bottom portion of Panel B. Column (5) shows that a long-short portfolio that follows the AI signal only when it aligns with the non-TA sentiment yields the highest return among all the long-short portfolios, at 16.03% per year. In contrast, column (6) shows that a portfolio following the AI signal only when it disagrees with the non-TA sentiment yields an insignificant return of 4.7% per year.

Together, the evidence suggests that a key source of the profitability of the AI signal from Jiang et al. (2023) lies in its ability to capitalize on retail investors who

[20] Table A.5 in the appendix reports value-weighted portfolio returns, and our findings are qualitatively similar.

misinterpret technical signals.

3.3. *Retail Investors' Herding and Technical Analysis*

In the previous subsection, we document the poor performance of StockTwits users' technical analysis. To understand the potential reasons behind this phenomenon, we next consider herding—an important behavior of retail investors that has been documented in the literature (see, e.g., Wermers, 1999). A recent paper by Barber et al. (2022) finds that retail investors on the Robinhood platform tend to engage in attention-induced trading, with their buy herding often leading to significant reversals in the following period.

Following their study, we create an indicator for Robinhood buy herding events, $RH\ Herding_{i,t}$, as the top 10 stocks with the highest Robinhood user change ratio in week t, with a minimum of 100 users at the end of week $t-1$.[21] We then estimate the following stock-week panel regression, where the dependent variable is the indicator for Robinhood buy herding:

$$\begin{aligned} RH\ Herding_{i,t} =& \beta_1 Sentiment_{i,t}^{TA} + \beta_2 Sentiment_{i,t}^{NonTA} + \beta_3 Attention_{i,t} \\ & + \beta_4 Earnings\ News_{i,t} + \beta_5 Analyst\ News_{i,t} + \gamma \mathbf{X_{i,t}} + \delta_t + \eta_j + \epsilon_{i,t}. \end{aligned} \tag{3}$$

Our sample period for this regression analysis spans from May 2018 to August 2020, during which data on Robinhood user accounts is available.[22] Our key explanatory variables are the StockTwits users' sentiments on stock i in week t, as indicated by the number of bullish and bearish messages: $Sentiment_{i,t} = \frac{N_{i,t}^{Bullish} - N_{i,t}^{Bearish}}{N_{i,t}^{Bullish} + N_{i,t}^{Bearish}}$. We measure sentiment using technical analysis (TA)-related messages and non-TA messages separately. $Attention_{i,t}$ is a measure of StockTwits users' attention on stock i in week t, defined as the number of messages on stock i divided by the total number of messages across all stocks, i.e., $Attention_{i,t} = \frac{\#Messages_{i,t}}{\sum_i \#Messages_{i,t}}$ (in percentage

[21] Barber et al. (2022) create the indicator for Robinhood buy herding events based on daily changes in the number of Robinhood users. However, we focus on weekly Robinhood user changes because our empirical analysis is conducted on a weekly basis.

[22] The authors thank Xing Huang for generously providing the data of Robinhood user accounts.

points) (see, Cookson et al., 2024b). Earnings $\text{News}_{i,t}$ is an indicator of the release of earnings-related news articles on financial media outlets in week t. Analyst $\text{News}_{i,t}$ is an indicator of the release of analyst forecasts and recommendations in week t. $X_{i,t}$ represents a vector of stock characteristics available at the beginning of week t, including (log) market capitalization, (log) book-to-market, asset growth, gross profits-to-asset, (log) number of analysts, (log) institutional ownership, the maximum daily return in the prior month, abnormal turnover, and lagged returns over five horizons: one day, one week, one month, one quarter, and one year. All specifications include calendar week fixed effects. Standard errors are clustered by calendar week.

[Insert Table 5 near here]

Table 5 presents the results from this regression analysis. We find that retail TA sentiment is significantly and positively related to contemporaneous Robinhood buy herding. In contrast, retail non-TA sentiment is negatively, but insignificantly, associated with herding events.

Overall, our results indicate that the sentiment of StockTwits users employing technical analysis is strongly related to retail investors' herding behavior on the Robinhood platform. Given that Robinhood buy herding leads to contemporaneous price overreaction and subsequent negative returns, and that StockTwits bullish TA sentiment also predicts lower future returns, our evidence suggests that crowded trading on technical signals likely contributes to the herding behavior we observe. This is consistent with Stein (2009), who argue that investors' reliance on common, unanchored strategies can result in the crowding of a strategy.

4. Fundamental Analysis and Other Strategies

In this section, we extend our LLM-based method to identify messages related to fundamental analysis and other strategies (see Section 2.2 for details). Unlike technical analysis, which focuses on price and volume patterns, fundamental analysis emphasizes understanding firms' underlying performance and valuation. By contrasting these approaches with technical analysis, we offer new insights into the return predictability of retail investors' sentiment and the informativeness of retail order imbalances.

4.1. Distribution of Use of Various Strategies

We first conduct a descriptive analysis to examine the frequencies of retail investors' messages related to technical analysis (TA), fundamental analysis (FA), other strategies (OS), and those unrelated to any trading strategy (NS).

[Insert Table 6 near here]

Table 6, Panel A, includes messages by all users and reports the fraction of messages that belong to fundamental and non-fundamental categories, respectively. Under our classification, we find that 31% of all messages contain trading strategies. Within these messages, 62.9% (=19.5%/31%) are either fundamental or technical messages. Zooming in on FA/TA messages, a few interesting patterns emerge. We find that only a very small fraction of messages (roughly 0.4%) have been classified as both fundamental and technical analysis. This is intuitive because a typical StockTwits message is usually short, so retail investors are unlikely to discuss both technical and fundamental content in a single message.

In Panels B and C, we examine the messages posted by self-declared technical and fundamental investors, respectively. Although self-declared technical investors tend to use more technical analysis (17.5% of total messages), they also use a substantial amount of fundamental analysis, with 10% of their messages are FA related. Similarly, we note that self-declared fundamental investors also utilize technical analysis significantly, with 7.5% of their messages falling into the technical category.

We also extend the regression analysis from Table 2 to study the determinants of retail usage of fundamental analysis and other strategies. Table A.6 in the appendix shows that both strategies are used more frequently by self-declared professional investors compared to those with novice or intermediate investment experience, with usage 8.1% higher for fundamental analysis and 2.1% higher for other strategies. In contrast to technical analysis, self-declared technical and short-term investors tend to use fundamental analysis less frequently in their investment decisions. Specifically, self-declared technical investors post 4.6% fewer FA messages on a stock in a given week than investors using other approaches, and short-term investors post 1.9% fewer FA messages than long-term investors, consistent with the graphical patterns presented in appendix Figures 6 and 5.

Overall, these results highlight that investors often employ a diverse investment approach and are not strictly confined to methods associated with their self-declared investment approach.

4.2. Return Predictability with Sentiment of Various Strategies

In our previous analysis, we have shown that trades based on retail investors' technical analysis tend to generate negative returns. In contrast, nontechnical sentiment exhibits a positive relationship with future returns. Next, we examine the informativeness of sentiments of other types of retail investment strategies as reflected in StockTwits posts. Specifically, we estimate the following panel regressions at the stock-week level:

$$\begin{aligned} Return_{i,t+1} =& \beta_1 Sentiment_{i,t}^{type} + \beta_2 Attention_{i,t} + \beta_3 Earnings\ News_{i,t} \\ &+ \beta_4 Analyst\ News_{i,t} + \gamma \boldsymbol{X_{i,t}} + \delta_t + \eta_j + \epsilon_{i,t+1}, \end{aligned} \tag{4}$$

where the dependent variable is stock return in the next week. Our key explanatory variables, $Sentiment^{type}$, where $type = TA, FA, OS, NS$, represent the sentiments of the four message types: technical analysis (TA), fundamental analysis (FA), other strategies (OS), and non-strategy (NS). The other explanatory and control variables are the same as those in Table 5, and all specifications include calendar week fixed effects.

[Insert Table 7 near here]

The regression results are reported in Table 7. In column (1), we find that the sentiment of all messages on StockTwits positively predicts the next-week return, confirming the findings of Cookson et al. (2024b). In contrast, consistent with our portfolio sort results in Table 4, column (2) shows that TA sentiment negatively predicts returns in the following week, suggesting that retail investors' technical analysis generates negative returns. Furthermore, column (3) shows that FA sentiment positively predicts future returns, indicating that retail investors may be informed about certain aspects of firms' fundamentals. Column (4) examines the sentiment of other retail trading strategies (OS) and shows that OS sentiment is negatively associated with future returns, although this predictability is not significant. Finally, column (5) demonstrates that the sentiment of non-strategy-related messages is not informative about future returns, suggesting that these messages mainly reflect retail investors' emotions or "noisy" beliefs.[23]

Together, the evidence suggests substantial heterogeneity in the informativeness of the different strategies discussed on StockTwits, with fundamental analysis being highly informative, while technical analysis leads to significant losses for investors who follow such strategies.

4.3. StockTwits Sentiments and Retail Order Flow

Our results thus far show that StockTwits TA sentiments negatively predict future returns, while FA sentiments positively predict future returns. One might argue that investors may "talk the talk" but not "walk the talk," meaning that social media discussions among retail investors may not reflect their actual trading behaviors. Since StockTwits does not disclose its users' real trades from their brokerage

[23] Cookson et al. (2024a) show that retail investors' analyses exhibit large variations in predictive horizons. Thus, we also examine the predictive power of StockTwits message sentiments corresponding to various retail investment strategies for returns in weeks $t+2$, $t+3$, and $t+4$. Table A.7 in the appendix shows that the negative return predictability of TA sentiment is primarily concentrated within one week. In contrast, FA sentiment can positively and significantly predict returns for up to two subsequent weeks, consistent with the intuition that fundamental information is typically more persistent and lasts longer.

accounts, we cannot directly examine whether an investor's trades align with the messages they share on StockTwits. However, we can infer the representativeness of StockTwits sentiments in understanding retail investor beliefs by investigating whether these sentiments are directly related to the contemporaneous aggregate retail order imbalances.

We identify retail market orders and consider two alternative measures of retail market order imbalance (OIB): $OIB_{i,t}^{BJZZ}$, constructed following the algorithm in Boehmer et al. (2021), and $OIB_{i,t}^{BHJOS}$, based on the modified method in Barber et al. (2022). Our retail OIB data is available from 2012 to 2021. We estimate the following panel regression at the stock-week level:

$$\begin{aligned} OIB_{i,t} =& \beta_1 Sentiment_{i,t}^{type} + \beta_2 Attention_{i,t} + \beta_3 Earnings\ News_{i,t} \\ &+ \beta_4 Analyst\ News_{i,t} + \gamma \boldsymbol{X_{i,t}} + \delta_t + \eta_j + \epsilon_{i,t}. \end{aligned} \tag{5}$$

Panels A and B present regression results focusing on each of these two alternative OIB measures as the dependent variable.

[Insert Table 8 near here]

Panel A shows a positive and significant relationship between all four types of StockTwits sentiments and the Boehmer et al. (2021) OIB measure. Panel B demonstrates that this finding is even more significant for the modified retail OIB measure proposed by Barber et al. (2023), with the relationship being particularly strong for TA sentiment. These results indicate that retail investors tend to be net buyers of a stock when StockTwits sentiments are bullish. The evidence suggests that StockTwits users' sentiments, especially those revealed by TA messages, are representative of retail investor beliefs and are closely aligned with their trading decisions.

4.4. Retail Strategies and Retail Order Flow Informativeness

As shown in Boehmer et al. (2021), the aggregate retail order imbalance is informative and positively predicts future stock returns. Given that StockTwits sentiments can serve as a proxy for retail investor beliefs and the trading strategies they are likely employing, we next explore how these strategies influence the informativeness of retail order flows.

In particular, given the substantial differences between TA, FA, and other sentiments and their differential predictive power of future returns that we have documented, we investigate whether the informativeness of retail order flows also depends on the strategies that representative retail traders rely on. Specifically, we regress future returns on retail order imbalances, conditioning on the intensity of retail investors' discussions about TA, FA, OS, and NS, respectively, on StockTwits:

$$\begin{aligned} Return_{i,t+1} =& \beta_1 OIB_{i,t} + \beta_2 High\ Retail_{i,t}^{type} + \beta_3 OIB_{i,t} \times High\ Retail_{i,t}^{type} \\ &+ \beta_4 Attention_{i,t} + \beta_5 Earnings\ News_{i,t} + \beta_6 Analyst\ News_{i,t} \\ &+ \gamma \boldsymbol{X_{i,t}} + \delta_t + \eta_j + \epsilon_{i,t+1}, \end{aligned} \tag{6}$$

where $High\ Retail_{i,t}^{type}, type = TA, FA, OS, NS$ is a dummy variable equal to 1 if the percentage of messages related to an investment approach on stock i in week t (e.g., $\frac{\#TA\ messages}{\#messages}$ for technical analysis) is above the median in the cross-section of stocks. We hypothesize that the intensive use of technical analysis by retail investors likely reduces the informativeness of retail OIB. In contrast, when retail investors heavily discuss fundamental analysis on StockTwits, their trading flows are more informed, leading to a stronger positive relationship between retail OIB and future returns. Table 9 presents the results.

[Insert Table 9 near here]

In Panel A, we report the results based on the Boehmer et al. (2021) retail OIB. While column (1) confirms that retail OIB tends to be informative, consistent with Boehmer et al. (2021), column (2) shows that its informativeness is greatly decreased when StockTwits users heavily discuss technical analysis. Economically, the intense discussion of technical analysis strategies is associated with a reduction of more than 25% in the return predictability of retail OIB. When retail technical usage is low, the retail OIB exhibits stronger predictive power. These results suggest that retail orders are not always informative and that retail investors' heavy reliance on technical analysis strategies often leads to money-losing trades.

In contrast, column (3) shows that more intense discussions of fundamental analysis significantly *improve* the return predictability of retail OIB. When retail usage of fundamental analysis is low, the retail OIB nearly loses its predictive power. This stands in sharp contrast to the effect of retail technical analysis.[24]

Columns (4) and (5) show that messages not involving TA or FA, as well as those unrelated to trading strategies, do not significantly interact with order flow informativeness. In Panel B, we repeat our tests using the Barber et al. (2023) OIB measure and find consistent results.

Taken together, our evidence suggests that the strategies discussed on StockTwits and the sentiments conveyed by these social media platforms are closely linked to aggregate retail trading activities. Furthermore, retail investors tend to apply fundamental strategies skillfully, and the prevalence of such strategies may be one reason for the informativeness of aggregate retail order flows observed in previous studies. In contrast, the crowded use of technical trading strategies reduces the informativeness of aggregate retail order flow, creating profitable opportunities for sophisticated traders.

5. Conclusion

By applying large language models to analyze over 77 million messages posted on a popular social media platform for investors, we infer retail investors' beliefs and the trading strategies that they employ. We find that the sentiment expressed

[24] In a related paper, Farrell et al. (2022) show that the ability of retail order imbalances to predict stock returns increases with the publication of articles in Seeking Alpha.

in technical analysis-related messages incorrectly predicts future returns, while sentiment in fundamental analysis-related messages tends to be more informative. Moreover, periods of heightened discussion around technical analysis on the platform are associated with less informative retail order flows and a greater likelihood of herding behavior among Robinhood users.

This dichotomy highlights the heterogeneity in retail investors' investment approaches and their varying degrees of sophistication. Our findings suggest that the informativeness of social media sentiment and retail order flows is more nuanced than previously documented: both sentiment and order flows lose their predictive power when retail investors heavily rely on technical trading strategies, potentially leading to overcrowding.

Importantly, our study offers new insight into the interaction between retail investors and sophisticated traders who utilize powerful AI-based strategies. We find that a state-of-the-art, highly profitable AI-based technical analysis strategy derives much of its profit by trading against StockTwits technical sentiment. This direct supports the theory of Dou et al. (2024). The evidence also suggests that investors' differential ability to effectively utilize technical signals play a crucial role in determining winners and losers in financial markets.

Given that investor beliefs and strategies are often unobservable, alternative information from social media platforms can provide useful insights into investor decision making. For instance, our findings on the interaction between retail investors and AI-powered trading reveal a clear informational imbalance between retail investors and sophisticated AI-driven traders. As AI continues to reshape the investment landscape, and with social media potentially exacerbating this imbalance, concerns about market fairness and integrity become more pressing. Our results contribute to policy discussions on regulating AI in financial markets and overseeing investment-related social media platforms.

REFERENCES

Antweiler, Werner, and Murray Z Frank, 2004, Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards, *The Journal of Finance* 59, 1259–1294.

Bagattini, Giulio, Zeno Benetti, and Claudia Guagliano, 2023, Artificial Intelligence in EU Securities Markets, *ESMA50-164-6247. European Securities and Markets Authority* .

Barber, Brad M, Xing Huang, Philippe Jorion, Terrance Odean, and Christopher Schwarz, 2023, A (Sub)penny for Your Thoughts: Tracking Retail Investor Activity in TAQ, *The Journal of Finance* .

Barber, Brad M, Xing Huang, Terrance Odean, and Christopher Schwarz, 2022, Attention-Induced Trading and Returns: Evidence from Robinhood Users, *The Journal of Finance* 77, 3141–3190.

Barber, Brad M, and Terrance Odean, 2000, Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors, *The journal of Finance* 55, 773–806.

Barber, Brad M, and Terrance Odean, 2008, All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors, *The Review of Financial Studies* 21, 785–818.

Blume, Lawrence, David Easley, and Maureen O'hara, 1994, Market Statistics and Technical Analysis: The Role of Volume, *The Journal of Finance* 49, 153–181.

Boehmer, Ekkehart, Charles M Jones, Xiaoyan Zhang, and Xinran Zhang, 2021, Tracking Retail Investor Activity, *The Journal of Finance* 76, 2249–2305.

Bradley, Daniel, Jan Hanousek Jr, Russell Jame, and Zicheng Xiao, 2024, Place Your Bets? The Value of Investment Research on Reddit's Wallstreetbets, *The Review of Financial Studies* 37, 1409–1459.

Brock, William, Josef Lakonishok, and Blake LeBaron, 1992, Simple Technical Trading Rules and The Stochastic Properties of Stock Returns, *The Journal of Finance* 47, 1731–1764.

Brown, David P, and Robert H Jennings, 1989, On Technical Analysis, *The Review of Financial Studies* 2, 527–551.

Bryzgalova, Svetlana, Anna Pavlova, and Taisiya Sikorskaya, 2023, Retail Trading in Options and the Rise of the Big Three Wholesalers, *The Journal of Finance* 78, 3465–3514.

Chen, Hailiang, Prabuddha De, Yu Hu, and Byoung-Hyoun Hwang, 2014, Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media, *The Review of Financial Studies* 27, 1367–1403.

Chen, Hailiang, and Byoung-Hyoun Hwang, 2022, Listening in on Investors' Thoughts and Conversations, *Journal of Financial Economics* 145, 426–444.

Chinco, Alex, Samuel M Hartzmark, and Abigail B Sussman, 2022, A New Test of Risk Factor Relevance, *The Journal of Finance* 77, 2183–2238.

Choi, James J, and Adriana Z Robertson, 2020, What Matters to Individual Investors? Evidence from The Horse's Mouth, *The Journal of Finance* 75, 1965–2020.

Cookson, J Anthony, Chukwuma Dim, and Marina Niessner, 2024a, Disagreement on the Horizon, *Available at SSRN* .

Cookson, J Anthony, Joseph E Engelberg, and William Mullins, 2023, Echo Chambers, *The Review of Financial Studies* 36, 450–500.

Cookson, J Anthony, Runjing Lu, William Mullins, and Marina Niessner, 2024b, The Social Signal, *Journal of Financial Economics* 158, 103870.

Cookson, J Anthony, William Mullins, and Marina Niessner, 2024c, Social Media and Finance, *Available at SSRN 4806692* .

Cookson, J Anthony, and Marina Niessner, 2020, Why Don't We Agree? Evidence from A Social Network of Investors, *The Journal of Finance* 75, 173–228.

Coval, Joshua D, David Hirshleifer, and Tyler Shumway, 2021, Can Individual Investors Beat The Market?, *The Review of Asset Pricing Studies* 11, 552–579.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, 2018, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *CoRR* abs/1810.04805.

Dou, Winston Wei, Itay Goldstein, and Yan Ji, 2024, AI-Powered Trading, Algorithmic Collusion, and Price Efficiency, *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper* .

Eaton, Gregory W, T Clifton Green, Brian S Roseman, and Yanbin Wu, 2022, Retail Trader Sophistication and Stock Market Quality: Evidence from Brokerage Outages, *Journal of Financial Economics* 146, 502–528.

Eisfeldt, Andrea L, Gregor Schubert, and Miao Ben Zhang, 2023, Generative AI and Firm Values, Technical report, National Bureau of Economic Research.

Ernst, Thomas, and Chester S Spatt, 2022, Payment for Order Flow and Asset Choice, Technical report, National Bureau of Economic Research.

Farrell, Michael, T Clifton Green, Russell Jame, and Stanimir Markov, 2022, The Democratization of Investment Research and The Informativeness of Retail Investor Trading, *Journal of Financial Economics* 145, 616–641.

George, Thomas J, and Chuan-Yang Hwang, 2004, The 52-Week High and Momentum Investing, *The Journal of Finance* 59, 2145–2176.

Giannini, Robert, Paul Irvine, and Tao Shu, 2018, Nonlocal Disadvantage: An Examination of Social Media Sentiment, *The Review of Asset Pricing Studies* 8, 293–336.

Giannini, Robert, Paul Irvine, and Tao Shu, 2019, The Convergence and Divergence of Investors' Opinions Around Earnings News: Evidence from A Social Network, *Journal of Financial Markets* 42, 94–120.

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus, 2021, Five Facts about Beliefs and Portfolios, *American Economic Review* 111, 1481–1522.

González-Carvajal, Santiago, and Eduardo C Garrido-Merchán, 2020, Comparing BERT against Traditional Machine Learning Text Classification, *arXiv preprint arXiv:2005.13012* .

Gu, Yuxian, Li Dong, Furu Wei, and Minlie Huang, 2023, Knowledge Distillation of Large Language Models, *arXiv preprint arXiv:2306.08543* .

Han, Yufeng, Ke Yang, and Guofu Zhou, 2013, A New Anomaly: The Cross-Sectional Profitability of Technical Analysis, *Journal of Financial and Quantitative Analysis* 48, 1433–1461.

Han, Yufeng, Guofu Zhou, and Yingzi Zhu, 2016, A Trend Factor: Any Economic Gains from Using Information over Investment Horizons?, *Journal of Financial Economics* 122, 352–375.

Hinton, G, 2015, Distilling the Knowledge in A Neural Network, *arXiv preprint arXiv:1503.02531* .

Hirshleifer, David, Lin Peng, and Qiguang Wang, Forthcoming, News Diffusion In Social Networks and Stock Market Reactions, *The Review of Financial Studies* .

Huang, Linmei, Bill Qiao, and Dexin Zhou, 2024, Are Memes a Sideshow: Evidence from WallStreetBets, *Available at SSRN 4785481* .

Jegadeesh, Narasimhan, 1991, Seasonality in Stock Price Mean Reversion: Evidence from The US and The UK, *The Journal of Finance* 46, 1427–1444.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency, *The Journal of Finance* 48, 65–91.

Jha, Manish, Jialin Qian, Michael Weber, and Baozhong Yang, 2024, ChatGPT and Corporate Policies, Technical report, National Bureau of Economic Research.

Jiang, Jingwen, Bryan Kelly, and Dacheng Xiu, 2023, (re-) Imag (in) ing Price Trends, *The Journal of Finance* 78, 3193–3249.

Jiang, Jingwen, Bryan T Kelly, and Dacheng Xiu, 2024a, Expected Returns and Large Language Models, *Available at SSRN* .

Jiang, Zhengyang, Cameron Peng, and Hongjun Yan, 2024b, Personality Differences and Investment Decision-Making, *Journal of Financial Economics* 153, 103776.

Kakhbod, Ali, Seyed Mohammad Kazempour, Dmitry Livdan, and Norman Schuerhoff, 2023, Finfluencers, *Available at SSRN 4428232* .

Kaniel, Ron, Gideon Saar, and Sheridan Titman, 2008, Individual Investor Trading and Stock Returns, *The Journal of Finance* 63, 273–310.

Kelley, Eric K, and Paul C Tetlock, 2013, How Wise Are Crowds? Insights from Retail Orders and Stock Returns, *The Journal of Finance* 68, 1229–1265.

Korinek, Anton, 2023, Generative AI for Economic Research: Use Cases and Implications for Economists, *Journal of Economic Literature* 61, 1281–1317.

Kumar, Alok, and Charles MC Lee, 2006, Retail Investor Sentiment and Return Comovements, *The Journal of Finance* 61, 2451–2486.

Laudenbach, Christine, Annika Weber, Rüdiger Weber, and Johannes Wohlfart, 2024, Beliefs About the Stock Market and Investment Choices: Evidence from a Survey and a Field Experiment, *The Review of Financial Studies* Forthcoming.

Li, Kai, Feng Mai, Rui Shen, Chelsea Yang, and Tengfei Zhang, 2023, Dissecting Corporate Culture Using Generative AI–Insights from Analyst Reports, *Available at SSRN 4558295* .

Liu, Hongqi, Cameron Peng, Wei A Xiong, and Wei Xiong, 2022, Taming The Bias Zoo, *Journal of Financial Economics* 143, 716–741.

Lo, Andrew W, and Jasmina Hasanhodzic, 2010, *The Evolution of Technical Analysis: Financial Prediction from Babylonian Tablets to Bloomberg Terminals*, volume 96 (John Wiley & Sons).

Lo, Andrew W, Harry Mamaysky, and Jiang Wang, 2000, Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation, *The Journal of Finance* 55, 1705–1765.

Lopez-Lira, Alejandro, and Yuehua Tang, 2023, Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models, *arXiv preprint arXiv:2304.07619* .

Murray, Scott, Yusen Xia, and Houping Xiao, 2024, Charting by Machines, *Journal of Financial Economics* 153, 103791.

SEC, 2023, Conflicts of Interest Associated with the Use of Predictive Data Analytics by BrokerDealers and Investment Advisers, *Release Nos. 34-97990. U.S. Securities and Exchange Commission.* .

Stein, Jeremy C, 2009, Presidential Address: Sophisticated Investors and Market Efficiency, *The Journal of Finance* 64, 1517–1548.

Welch, Ivo, 2022, The Wisdom of the Robinhood Crowd, *The Journal of Finance* 77, 1489–1527.

Wermers, Russ, 1999, Mutual Fund Herding and The Impact on Stock Prices, *the Journal of Finance* 54, 581–622.

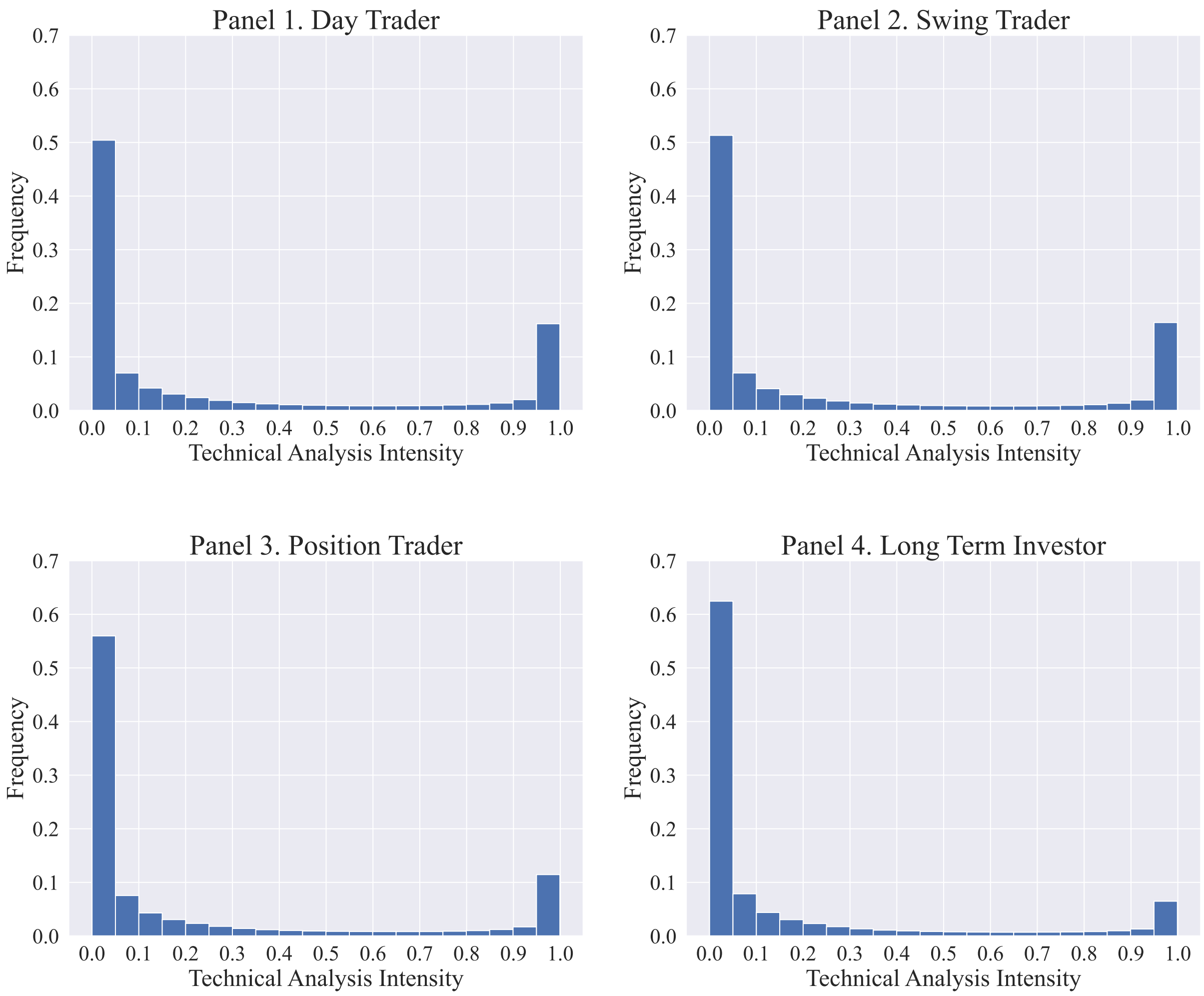


Fig. 1. Distribution of Messages across Technical Analysis Intensity, by Self-declared Investment Horizon

We apply a fine-tuned BERT model specializing in identifying messages pertaining to technical analysis on StockTwits messages. The model outputs a probabilistic score for each message (i.e., Technical Analysis Intensity). We then plot histograms of message frequency by technical analysis intensity within each self-declared horizon subgroup.

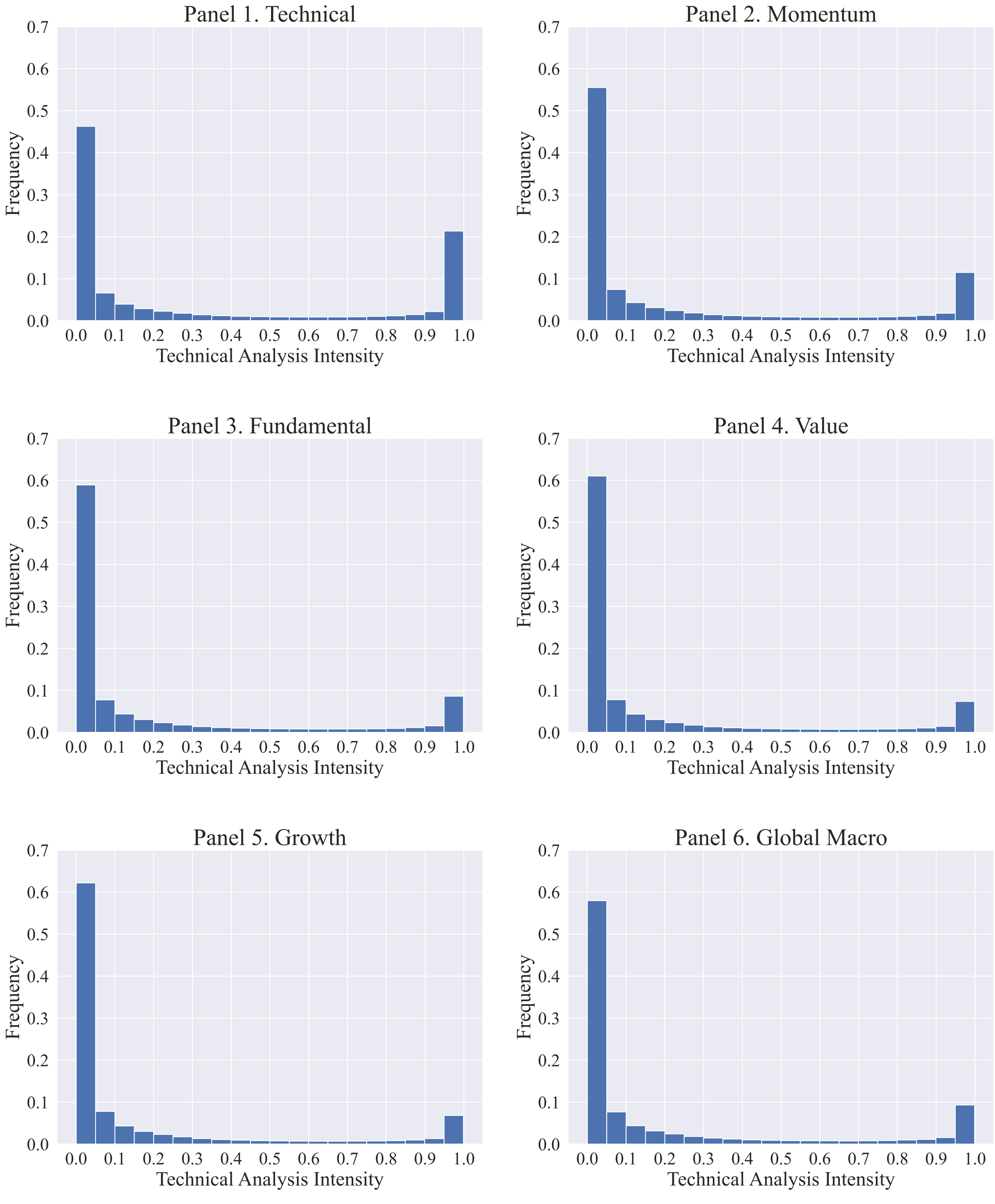
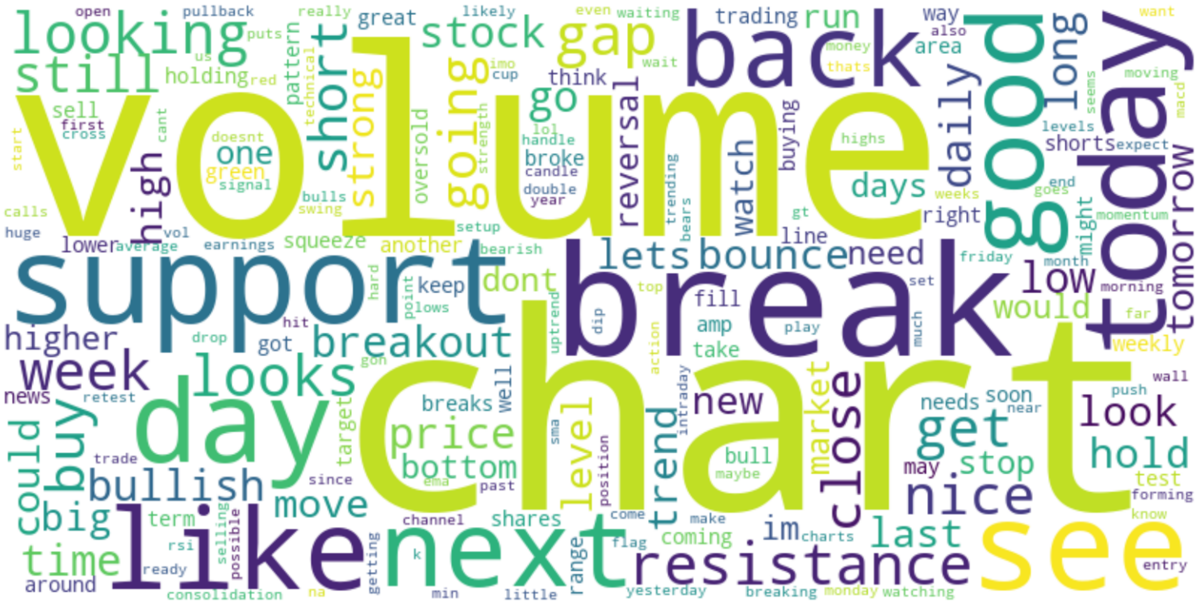


Fig. 2. Distribution of Messages across Technical Analysis Intensity, by Self-declared Investment Approach

We apply a fine-tuned BERT model specializing in identifying messages pertaining to technical analysis on StockTwits messages. The model outputs a probabilistic score for each message (i.e., Technical Analysis Intensity). We then plot histograms of message frequency by technical analysis intensity within each self-declared investment approach subgroup.

Panel A. Word cloud for unigram words in technical messages



Panel B. Word cloud for bigram phrases in technical messages

Fig. 3. Word Cloud of Technical Messages

We classify all messages with a technical analysis intensity of 0.95 or higher as technical messages. Panel A shows the word cloud for unigram terms in technical messages, and Panel B shows the word cloud for bigram phrases in technical messages.

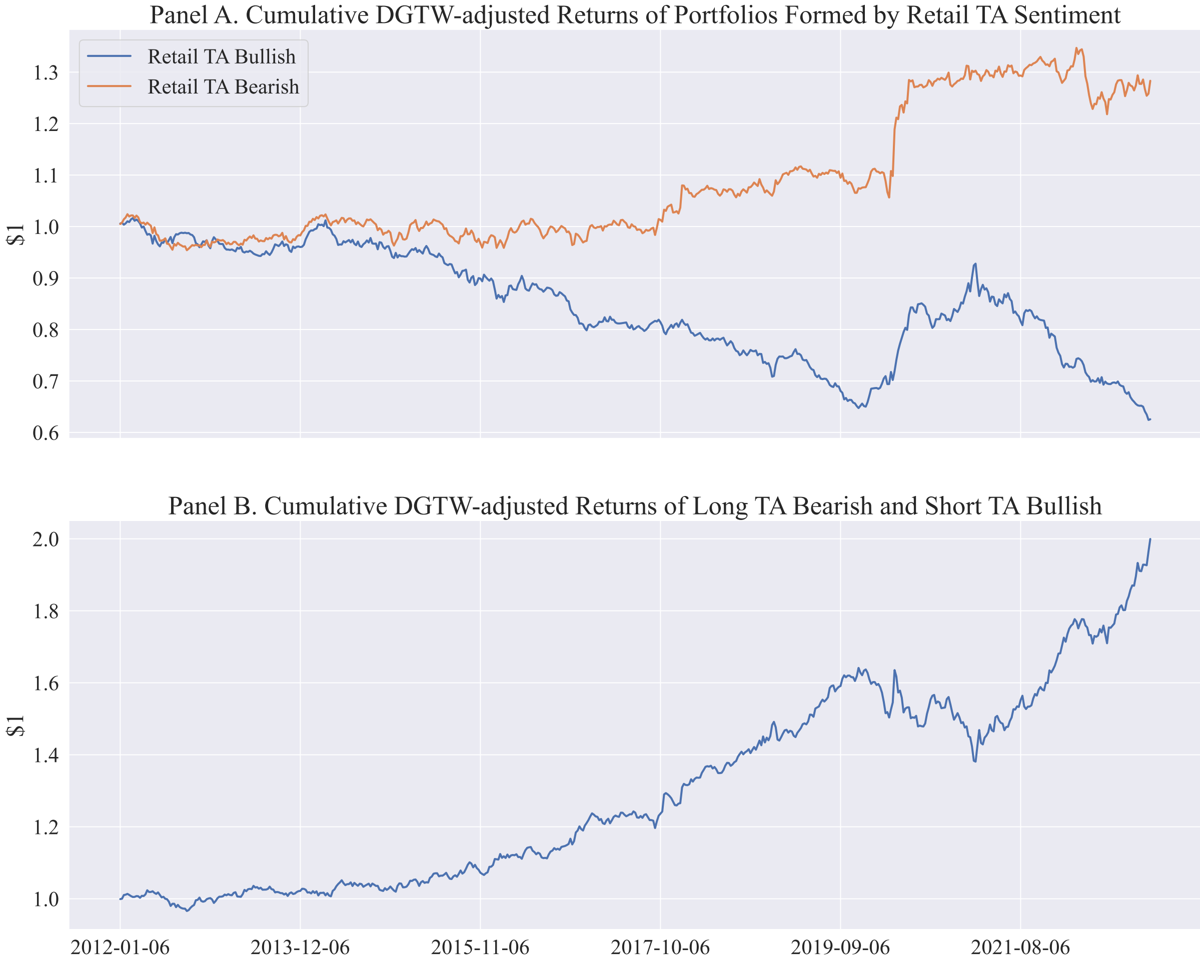


Fig. 4. Performance of Portfolios Sorted by Retail TA Sentiment

Panel A plots the cumulative DGTW-adjusted returns of equal-weighted portfolios sorted by retail TA sentiment. Retail TA Bullish denotes the portfolio of stocks with retail TA sentiment greater than its cross-sectional median. Retail TA Bearish portfolio is defined similarly. Panel B plots the cumulative DGTW-adjusted returns of the long-short strategy that goes long on the TA Bearish portfolio and short on the TA Bullish portfolio.

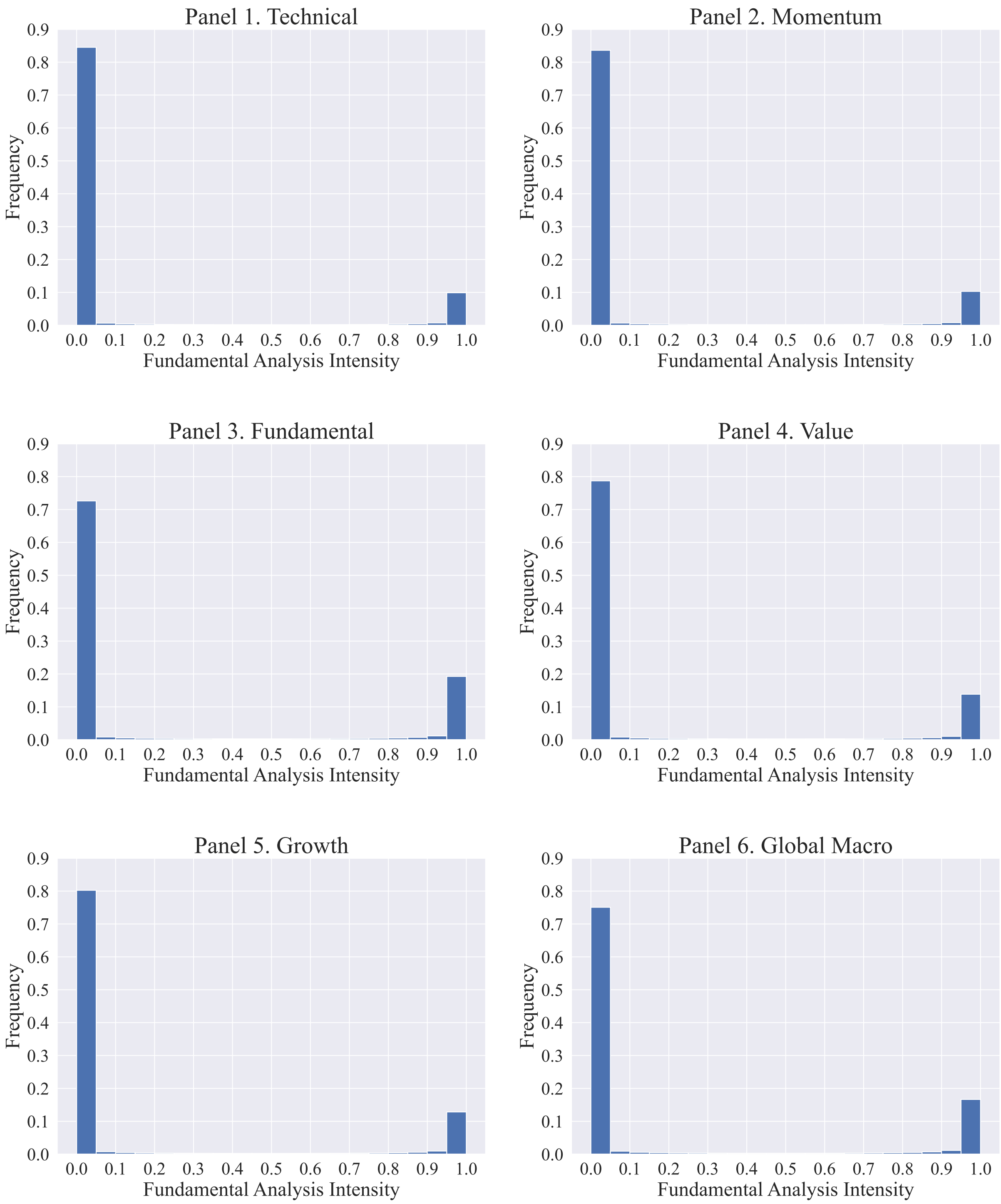


Fig. 5. Distribution of Messages across Fundamental Analysis Intensity, by Self-declared Investment Approach

We apply a fine-tuned BERT model specializing in identifying messages pertaining to fundamental analysis on StockTwits messages. The model outputs a probabilistic score for each message (i.e., Fundamental Analysis Intensity). We then plot histograms of message frequency by fundamental analysis intensity within each self-declared investment approach subgroup.

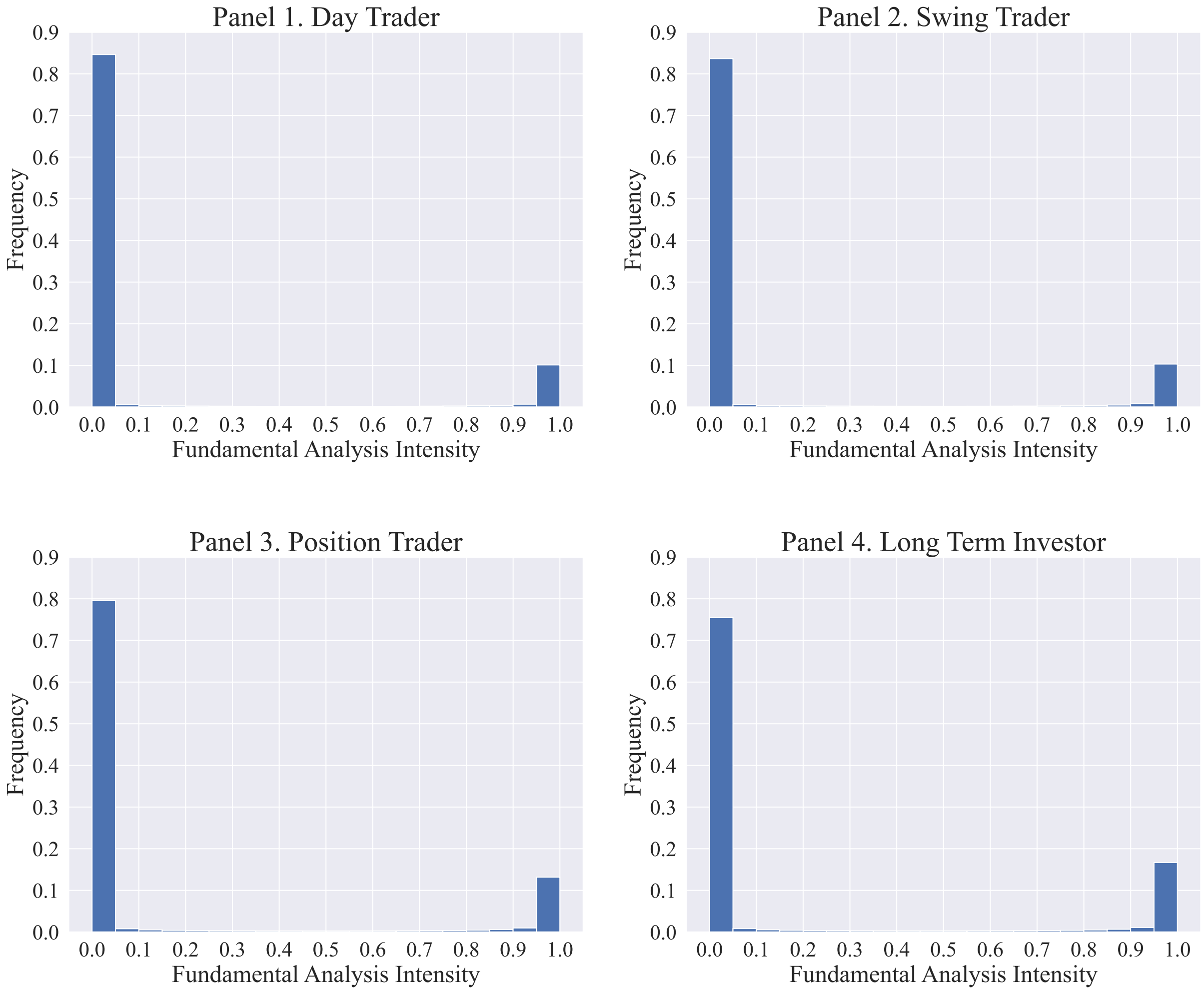


Fig. 6. Distribution of messages across Fundamental Analysis Intensity, by Self-declared Investment Horizon

We apply a fine-tuned BERT model specializing in identifying messages pertaining to fundamental analysis on StockTwits messages. The model outputs a probabilistic score for each message (i.e., Fundamental Analysis Intensity). We then plot histograms of message frequency by fundamental analysis intensity within each self-declared investment horizon subgroup.

Table 1: **Summary Statistics**

Panel A reports the summary statistics for variables in the investor-stock-week sample and panel B reports summary statistics for the stock-week sample. The sample period is from 2012 to 2022. See Table A.2 in the appendix for detailed variable definitions.

	N	Mean	Median	StdDev	10th	25th	75th	90th
Retail TA Usage$_{i,j,t}$	16,994,396	0.11	0.00	0.28	0.00	0.00	0.00	0.50
Sentiment$_{i,j,t}$	16,994,396	0.42	1.00	0.80	-1.00	0.00	1.00	1.00
Sentiment$^{TA}_{i,j,t}$	3,234,251	0.50	1.00	0.81	-1.00	0.00	1.00	1.00
Sentiment$^{NonTA}_{i,j,t}$	15,744,015	0.42	1.00	0.80	-1.00	0.00	1.00	1.00
Technical Investor$_{j,t}$	16,994,396	0.22	0.00	0.42	0.00	0.00	0.00	1.00
Short-term Investor$_{j,t}$	16,994,396	0.30	0.00	0.46	0.00	0.00	1.00	1.00
Earnings News$_{i,t}$	16,994,396	0.23	0.00	0.42	0.00	0.00	0.00	1.00
Analyst News$_{i,t}$	16,994,396	0.28	0.00	0.45	0.00	0.00	1.00	1.00
Log(Market Cap$_{i,t}$)	16,994,396	7.54	7.55	2.94	3.48	5.16	10.37	11.50
Log(Book-to-Market$_{i,t}$)	16,994,396	-1.62	-1.57	1.39	-3.46	-2.66	-0.65	0.16
Asset Growth$_{i,t}$	16,994,396	0.73	0.15	2.04	-0.21	-0.03	0.62	1.95
Gross Profit-to-Asset$_{i,t}$	16,994,396	0.14	0.16	0.35	-0.30	0.00	0.35	0.56
Log(1+#Analysts$_{i,t}$)	16,994,396	2.05	2.08	1.16	0.00	1.10	3.09	3.50
Log(Inst. Own$_{i,t}$)	16,994,396	-1.06	-0.59	1.08	-2.65	-1.59	-0.29	-0.10
MAX$_{i,t}$	16,994,396	0.12	0.08	0.13	0.03	0.04	0.15	0.28
Abnormal Turnover$_{i,t}$	16,994,396	0.04	-0.02	0.65	-0.64	-0.33	0.32	0.75
Return$^{1d}_{i,t}$	16,994,396	0.01	0.00	0.11	-0.05	-0.02	0.02	0.06
Return$^{1w}_{i,t}$	16,994,396	0.04	0.00	0.30	-0.13	-0.05	0.06	0.18
Return$^{1m}_{i,t}$	16,994,396	1.11	1.01	0.78	0.74	0.88	1.14	1.43
Return$^{1q}_{i,t}$	16,994,396	1.28	1.00	1.75	0.54	0.77	1.25	1.87
Return$^{1y}_{i,t}$	16,994,396	1.94	1.05	4.80	0.23	0.54	1.74	3.34
AI Signal$_{i,t}$	6,854,262	0.51	0.50	0.04	0.46	0.48	0.53	0.56

	N	Mean	Median	StdDev	10th	25th	75th	90th
$\text{Sentiment}_{i,t}$	627,514	0.56	0.61	0.43	0.00	0.33	1.00	1.00
$\text{Sentiment}_{i,t}^{TA}$	627,514	0.40	0.33	0.56	0.00	0.00	1.00	1.00
$\text{Sentiment}_{i,t}^{NonTA}$	627,514	0.54	0.62	0.46	0.00	0.33	1.00	1.00
$\text{Sentiment}_{i,t}^{FA}$	627,514	0.43	0.50	0.55	0.00	0.00	1.00	1.00
$\text{Sentiment}_{i,t}^{OS}$	627,514	0.30	0.00	0.56	0.00	0.00	1.00	1.00
$\text{Sentiment}_{i,t}^{NS}$	627,514	0.47	0.57	0.53	0.00	0.00	1.00	1.00
$\text{Earnings News}_{i,t}$	627,514	0.17	0.00	0.38	0.00	0.00	0.00	1.00
$\text{Analyst News}_{i,t}$	627,514	0.18	0.00	0.38	0.00	0.00	0.00	1.00
$\text{Log(Market Cap}_{i,t})$	627,514	7.14	7.22	2.46	3.75	5.29	9.02	10.53
$\text{Log(Book-to-Market}_{i,t})$	627,514	-1.11	-1.05	1.15	-2.61	-1.81	-0.34	0.27
$\text{Asset Growth}_{i,t}$	627,514	0.39	0.07	1.34	-0.18	-0.03	0.30	1.00
$\text{Gross Profit-to-Asset}_{i,t}$	627,514	0.21	0.21	0.33	-0.13	0.04	0.38	0.59
$\text{Log(1+\#Analysts}_{i,t})$	627,514	2.00	2.08	0.98	0.69	1.39	2.83	3.18
$\text{Log(Inst. Own}_{i,t})$	627,514	-0.68	-0.28	0.96	-2.10	-0.84	-0.09	0.00
$\text{MAX}_{i,t}$	627,514	0.08	0.05	0.08	0.02	0.03	0.09	0.16
$\text{Abnormal Turnover}_{i,t}$	627,514	-0.04	-0.08	0.60	-0.68	-0.37	0.24	0.61
$\text{Return}_{i,t}^{1d}$	627,514	0.00	0.00	0.05	-0.04	-0.01	0.02	0.04
$\text{Return}_{i,t}^{1w}$	627,514	0.00	0.00	0.10	-0.09	-0.04	0.04	0.09
$\text{Return}_{i,t}^{1m}$	627,514	1.01	1.00	0.21	0.81	0.92	1.08	1.20
$\text{Return}_{i,t}^{1q}$	627,514	1.03	1.01	0.41	0.65	0.84	1.16	1.37
$\text{Return}_{i,t}^{1y}$	627,514	1.17	1.04	1.11	0.36	0.67	1.39	1.93
$\text{AI Signal}_{i,t}$	371,382	0.50	0.50	0.04	0.46	0.48	0.53	0.56
$\text{OIB}_{i,t}^{BJZZ}$	456,086	-0.01	-0.00	0.15	-0.18	-0.08	0.07	0.16
$\text{OIB}_{i,t}^{BHJOS}$	435,328	-0.01	0.00	0.15	-0.17	-0.07	0.07	0.14
$\text{RH Herding}_{i,t}$	52,072	0.01	0.00	0.12	0.00	0.00	0.00	0.00

Table 2: **Determinants of Retail Investors' Technical Analysis Usage**

This table examines the determinants of retail investors' usage of technical analysis (TA), revealed by their TA-related messages posted on StockTwits from 2012 to 2022. We estimate the following panel regression at the stock-investor-week level:

$$\begin{aligned}&\textit{Retail TA Usage}_{i,j,t+1}\\ =&\beta_1 \textit{Technical Investor}_{j,t} + \beta_2 \textit{Short-term Investor}_{j,t} + \beta_3 \textit{Professional Investor}_{j,t}\\ +&\beta_4 \textit{Earnings News}_{i,t} + \beta_5 \textit{Analyst News}_{i,t} + \gamma \boldsymbol{X}_{i,t} + FEs + \epsilon_{i,j,t+1}.\end{aligned}$$

The dependent variable is the TA usage by investor j on stock i in week $t+1$, measured as the fraction of investor j's messages classified as TA-related by our classification LLM (TechBERT). We consider factors potentially affecting retail TA usage, including 1) investor types and 2) news releases. StockTwits users self-report their types for their investment approach, holding horizon, and experience. Technical Investor$_{j,t}$ is a dummy variable equal to one if investor j's self-reported approach is "*Technical*" or "*Momentum*". Short-term Investor$_{j,t}$ is a dummy variable equal to one if investor j's self-reported horizon is "*Day Trader*", "*Swing Trader*", or "*Position Trader*". Professional Investor$_{j,t}$ is a dummy variable equal to one if investor j's self-reported experience is "*Professional*". Earnings News$_{i,t}$ is an indicator of the release of earnings-related news on stock i in week t. Analyst News$_{i,t}$ is an indicator of the release of analyst forecasts and recommendations on stock i in week t. $X_{i,t}$ represents a vector of stock characteristics available at the beginning of week t, including (log) market capitalization, (log) book-to-market, asset growth, gross profits-to-asset, (log) number of analysts, (log) institutional ownership, the maximum daily return in the prior month, and abnormal turnover. Specifications in columns (1) and (2) include calendar week fixed effects and column (3) includes both calendar week and investor fixed effects. Standard errors are clustered by investor and calendar week, and associated t-statistics are reported in brackets, where *, **, and *** denote significance at levels 10%, 5%, and 1%.

	Retail TA Usage$_{i,j,t+1}$		
	(1)	(2)	(3)
Technical Investor$_{j,t}$	0.097***	0.097***	
	[19.92]	[19.92]	
Short-term Investor$_{j,t}$	0.021***	0.021***	
	[5.62]	[5.62]	
Professional Investor$_{j,t}$	0.051***	0.050***	
	[7.26]	[7.25]	
Earnings News$_{i,t}$		-0.010***	-0.005***
		[-12.28]	[-10.86]
Analyst News$_{i,t}$		-0.007***	-0.003***
		[-6.58]	[-7.08]
Stock Characteristics	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes
Investor FEs	No	No	Yes
Observations	16,994,396	16,994,396	16,818,446
R-squared	0.063	0.064	0.295

Table 3: **Do Retail Investors Agree with AI in Technical Analysis?**

This table examines how retail investors' sentiment revealed by their StockTwits messages is related to the technical trading signal generated by machine-learning models. Specifically, we estimate the following panel regression at the stock-investor-week level:

$$
\begin{aligned}
Sentiment^{type}_{i,j,t} =& \beta_1 AI\ Signal_{i,t} + \beta_2 Return^{1d}_{i,t-1} + \beta_3 Return^{1w}_{i,t-1} + \beta_4 Return^{1m}_{i,t-1} + \beta_5 Return^{1q}_{i,t-1} \\
&+ \beta_6 Return^{1y}_{i,t-1} + \beta_7 Earnings\ News_{i,t-1} + \beta_8 Analyst\ News_{i,t-1} \\
&+ \gamma \boldsymbol{X_{i,t}} + \delta_t + \eta_j + \epsilon_{i,j,t}, \quad type = TA, NonTA
\end{aligned}
$$

where the dependent variable is the sentiment of investor j on stock i in week t, measured by the number of investor j's bullish and bearish messages, $Sentiment_{i,j,t} = \frac{N^{Bullish}_{i,j,t} - N^{Bearish}_{i,j,t}}{N^{Bullish}_{i,j,t} + N^{Bearish}_{i,j,t}}$ (see, Cookson and Niessner, 2020). In columns (1) and (2), the sentiment is measured based on technical analysis (TA)-related messages, while columns (3) and (4) focus on the sentiment of non-TA messages. The key explanatory variable, $AI\ Signal_{i,t}$, is the stock-level estimate for the probability of a positive return in the subsequent week $t+1$, developed in Jiang et al. (2023) from training a convolutional neural network (CNN) on image data representing the price pattern over the preceding five days (denoted by $CNN5d5p$ in Jiang et al. (2023)). The data of this AI signal are available from 2012 to 2019. In columns (2) and (4), the regression specifications include the (cumulative) returns over the past one day, one week, one month, one quarter, and one year. See Table 2 for definitions of other explanatory and control variables. All specifications include both calendar week and investor fixed effects. Standard errors are clustered by investor and calendar week, and associated t-statistics are reported in brackets, where *, **, and *** denote significance at levels 10%, 5%, and 1%.

	$\text{Sentiment}^{TA}_{i,j,t}$		$\text{Sentiment}^{NonTA}_{i,j,t}$	
	(1)	(2)	(3)	(4)
AI $\text{Signal}_{i,t}$	-0.489***	-0.483***	-0.255***	-0.250***
	[-9.40]	[-9.41]	[-3.89]	[-3.81]
$\text{Return}^{1d}_{i,t-1}$		0.055**		0.024
		[2.40]		[0.83]
$\text{Return}^{1w}_{i,t-1}$		-0.013		-0.012
		[-1.03]		[-0.66]
$\text{Return}^{1m}_{i,t-1}$		0.012		0.008
		[1.55]		[0.89]
$\text{Return}^{1q}_{i,t-1}$		0.017***		0.031***
		[3.26]		[6.43]
$\text{Return}^{1y}_{i,t-1}$		-0.024***		-0.021***
		[-10.55]		[-7.07]
Earnings $\text{News}_{i,t-1}$	-0.046***	-0.045***	-0.054***	-0.053***
	[-8.78]	[-8.80]	[-7.88]	[-7.80]
Analyst $\text{News}_{i,t-1}$	-0.066***	-0.064***	-0.078***	-0.076***
	[-13.02]	[-13.11]	[-10.98]	[-10.90]
Stock Characteristics	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes
Investor FEs	Yes	Yes	Yes	Yes
Observations	1,531,977	1,531,977	6,108,652	6,108,652
R-squared	0.194	0.195	0.165	0.166

Table 4: **Performance of Retail TA and AI Strategies**

Panel A presents the average DGTW-adjusted returns of equal-weighted portfolios formed on retail TA sentiment and AI Signal. Retail TA Bull (Bear) denotes the portfolio of stocks with retail TA sentiment greater (less) than its cross-sectional median. Retail Bull–Bear indicates the return difference of the Retail Bull and Bear portfolios. We also form portfolios based on retail non-TA sentiment. AI Buy (Sell) is the portfolio of stocks with AI Signal greater (less) than 0.5. AI Buy–Sell indicates the return difference between the AI Buy and Sell portfolios. Panel B presents the average DGTW-adjusted returns of the two-way sorted portfolios in columns (1)–(4). Columns (5) and (6) focus on the two long-short strategies, one in which retail investors agree with AI (Buy/Bull–Sell/Bear), and the other in which retail investors disagree with AI (Buy/Bear–Sell/Bull). The average DGTW-adjusted returns are annualized in percentage and associated t-statistics are reported in brackets, where *, **, and *** denote significance at levels 10%, 5%, and 1%.

Panel A. Univariate Sorted Portfolios

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Retail TA			Retail Non-TA			AI Signal		
	Bull	Bear	Bull–Bear	Bull	Bear	Bull–Bear	Buy	Sell	Buy–Sell
$r^{DGTW-adj}$	-4.04	2.44	-6.48	0.58	-4.75	5.33	2.48	-7.72	10.20
	[-2.04]	[1.35]	[-3.50]	[0.36]	[-2.51]	[3.73]	[1.34]	[-4.02]	[5.64]

Panel B. Two-way Sorted Portfolios

	(1)	(2)	(3)	(4)	(5)	(6)
	Retail TA					
	Bull		Bear		Buy/Bull	Buy/Bear
AI Signal	Buy	Sell	Buy	Sell	– Sell/Bear	– Sell/Bull
$r^{DGTW-adj}$	-1.21	-8.17	5.22	2.44	-3.66	13.40
	[-0.53]	[-3.53]	[2.44]	[0.35]	[-0.52]	[5.01]
	Retail Non-TA					
	Bull		Bear		Buy/Bull	Buy/Bear
AI Signal	Buy	Sell	Buy	Sell	– Sell/Bear	– Sell/Bull
$r^{DGTW-adj}$	5.11	-4.72	-0.02	-10.92	16.03	4.70
	[2.39]	[-2.03]	[-0.01]	[-4.83]	[6.88]	[1.78]

Table 5: **Retail Herding and Technical Analysis**

This table shows that retail herding is positively correlated with the contemporaneous sentiment of technical analysis (TA) related messages on StockTwits. We estimate the following panel regression at the stock-week level:

$$
\begin{aligned}
RH\ Herding_{i,t} =& \beta_1 Sentiment^{TA}_{i,t} + \beta_2 Sentiment^{NonTA}_{i,t} + \beta_3 Attention_{i,t} \\
&+ \beta_4 Earnings\ News_{i,t} + \beta_5 Analyst\ News_{i,t} + \gamma \boldsymbol{X}_{i,t} + \delta_t + \eta_j + \epsilon_{i,t},
\end{aligned}
$$

where the dependent variable is an indicator of Robinhood herding events, identified as the top 10 stocks with the highest Robinhood user change ratio in week t and a minimum of 100 users at the end of week $t-1$ (see, Barber et al., 2022). The data on Robinhood user accounts is available from May 2018 to August 2020. Our key explanatory variables are the StockTwits users' sentiments on stock i in week t revealed by the number of bullish and bearish messages, $Sentiment_{i,t} = \frac{N^{Bullish}_{i,t} - N^{Bearish}_{i,t}}{N^{Bullish}_{i,t} + N^{Bearish}_{i,t}}$, where we measure sentiments using technical analysis (TA) related messages and non-TA messages separately. $\text{Attention}_{i,t}$ is a measure of StockTwits users' attention on stock i in week t, defined as the number of messages on stock i divided by the total number of messages across all stocks, i.e., $Attention_{i,t} = \frac{\#Messages_{i,t}}{\sum_i \#Messages_{i,t}}$ (in percentage points) (see, Cookson et al., 2024b). Earnings $\text{News}_{i,t}$ is an indicator of the release of earnings-related news in week t. Analyst $\text{News}_{i,t}$ is an indicator of the release of analyst forecasts and recommendations in week t. $\boldsymbol{X}_{i,t}$ represents a vector of stock characteristics available at the beginning of week t, including (log) market capitalization, (log) book-to-market, asset growth, gross profits-to-asset, (log) number of analysts, (log) institutional ownership, the maximum daily return in the prior month, abnormal turnover, and lagged returns over five horizons: one day, one week, one month, one quarter, and one year. All specifications include calendar week fixed effects. Standard errors are clustered by calendar week and associated t-statistics are reported in brackets, where *, **, and *** denote significance at levels 10%, 5%, and 1%.

	RH $\text{Herding}_{i,t}$		
	(1)	(2)	(3)
$\text{Sentiment}^{TA}_{i,t}$	0.003***		0.003***
	[2.97]		[3.08]
$\text{Sentiment}^{NonTA}_{i,t}$		-0.001	-0.001
		[-1.29]	[-1.62]
$\text{Attention}_{i,t}$	0.020***	0.020***	0.020***
	[5.97]	[5.96]	[5.96]
Earnings $\text{News}_{i,t}$	0.008***	0.008***	0.008***
	[4.78]	[4.76]	[4.78]
Analyst $\text{News}_{i,t}$	-0.003***	-0.003***	-0.003***
	[-2.69]	[-2.71]	[-2.73]
Stock Characteristics	Yes	Yes	Yes
Lagged Returns	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes
Observations	52,072	52,072	52,072
R-squared	0.011	0.011	0.011

Table 6: **Distribution of Messages by Retail Investment Strategies**

This table presents the distribution of messages by StockTwits users' investment strategies classified by our LLM approach. Specifically, a message can be related to technical analysis (TA), fundamental analysis (FA), other strategy (neither TA nor FA), and non-strategy. Numbers represent the fraction of total messages. Panel A includes the messages posted by all users in our StockTwits sample, while Panel B and C focus on the messages of self-reported technical and fundamental investors, respectively.

Panel A. All Investors

	Strategy	Non-strategy	Sum
	0.310	0.690	1
TA or FA	0.195		
Other Strategy	0.115		
	TA	Non-TA	Sum
FA	0.004	0.107	0.111
Non-FA	0.084	0.804	0.888
Sum	0.088	0.911	

Panel B. Self-reported Technical Investors

	Strategy	Non-strategy	Sum
	0.408	0.592	1
TA or FA	0.268		
Other Strategy	0.140		
	TA	Non-TA	Sum
FA	0.007	0.093	0.100
Non-FA	0.168	0.731	0.899
Sum	0.175	0.824	

Panel C. Self-reported Fundamental Investors

	Strategy	Non-strategy	Sum
	0.341	0.659	1
TA or FA	0.220		
Other Strategy	0.121		
	TA	Non-TA	Sum
FA	0.005	0.145	0.150
Non-FA	0.070	0.780	0.850
Sum	0.075	0.925	

Table 7: **Predicting Return with Retail Sentiments by Various Strategies**

This table examines the weekly stock return predictability with sentiments of various retail trading strategies. We estimate the panel regressions at the stock-week level,

$$\begin{aligned} Return_{i,t+1} = & \beta_1 Sentiment_{i,t}^{type} + \beta_2 Attention_{i,t} + \beta_3 Earnings\ News_{i,t} \\ & + \beta_4 Analyst\ News_{i,t} + \gamma X_{i,t} + \delta_t + \eta_j + \epsilon_{i,t+1}, \end{aligned}$$

where $Sentiment^{type}, type = TA, FA, OS, NS$, represent the sentiments of the four message types: technical analysis (TA), fundamental analysis (FA), other strategies (OS), and non-strategy (NS). See Table 5 for definitions of other explanatory and control variables. All specifications include calendar week fixed effects. Standard errors are clustered by calendar week and associated t-statistics are reported in brackets, where *, **, and *** denote significance at levels 10%, 5%, and 1%.

	Return$_{i,t+1}$ (%)					
	(1)	(2)	(3)	(4)	(5)	(6)
Sentiment$_{i,t}$	0.056*					
	[1.79]					
Sentiment$_{i,t}^{TA}$		-0.067**				-0.069**
		[-2.36]				[-2.54]
Sentiment$_{i,t}^{FA}$			0.050*			0.059**
			[1.94]			[2.33]
Sentiment$_{i,t}^{OS}$				-0.036		-0.033
				[-1.34]		[-1.28]
Sentiment$_{i,t}^{NS}$					-0.000	0.006
					[-0.00]	[0.28]
Attention$_{i,t}$	-0.111***	-0.115***	-0.113***	-0.115***	-0.114***	-0.115***
	[-3.05]	[-3.17]	[-3.12]	[-3.15]	[-3.14]	[-3.14]
Earnings News$_{i,t}$	0.089*	0.088*	0.088*	0.089*	0.089*	0.087
	[1.68]	[1.66]	[1.66]	[1.67]	[1.68]	[1.64]
Analyst News$_{i,t}$	-0.042	-0.043	-0.042	-0.043	-0.043	-0.042
	[-1.11]	[-1.15]	[-1.12]	[-1.14]	[-1.14]	[-1.12]
Stock Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Lagged Returns	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	627,060	627,060	627,060	627,060	627,060	627,060
R-squared	0.125	0.125	0.125	0.125	0.125	0.125

Table 8: **Retail Order Flows and Sentiments by Various Strategies**

This table examines the contemporaneous relation between the trading activity of retail investors and sentiments regarding technical analysis (TA), fundamental analysis (FA), other strategies (OS), and non-strategy (NS). We estimate the following panel regression at the stock-week level:

$$\begin{aligned} OIB_{i,t} =& \beta_1 Sentiment_{i,t}^{type} + \beta_2 Attention_{i,t} + \beta_3 Earnings\ News_{i,t} \\ &+ \beta_4 Analyst\ News_{i,t} + \gamma X_{i,t} + \delta_t + \eta_j + \epsilon_{i,t}, \quad type = TA, FA, OS, NS. \end{aligned}$$

The dependent variable is the retail order imbalance (OIB). We consider two alternative measures: $OIB_{i,t}^{BJZZ}$, constructed following the algorithm in Boehmer et al. (2021), and $OIB_{i,t}^{BHJOS}$, based on the modified method in Barber et al. (2022). Panels A and B present results focusing on each of these two alternative OIB measures as the dependent variable. Our OIB data is available from 2012 to 2020. See Table 5 for definitions of other explanatory and control variables. All specifications include calendar week fixed effects. Standard errors are clustered by calendar week and associated t-statistics are reported in brackets, where *, **, and *** denote significance at levels 10%, 5%, and 1%.

	$\text{OIB}_{i,t}^{BJZZ}$ (%)				
	(1)	(2)	(3)	(4)	(5)
$\text{Sentiment}_{i,t}^{TA}$	0.355***				0.282***
	[8.42]				[6.63]
$\text{Sentiment}_{i,t}^{FA}$		0.232***			0.163***
		[5.20]			[3.58]
$\text{Sentiment}_{i,t}^{OS}$			0.387***		0.322***
			[10.36]		[8.47]
$\text{Sentiment}_{i,t}^{NS}$				0.310***	0.236***
				[6.93]	[5.22]
$\text{Attention}_{i,t}$	0.449***	0.446***	0.449***	0.455***	0.466***
	[16.16]	[16.09]	[16.14]	[16.35]	[16.67]
$\text{Earnings News}_{i,t}$	-0.140**	-0.148**	-0.140**	-0.144**	-0.139**
	[-2.31]	[-2.45]	[-2.31]	[-2.37]	[-2.29]
$\text{Analyst News}_{i,t}$	0.102*	0.104*	0.102*	0.102*	0.105*
	[1.86]	[1.90]	[1.86]	[1.86]	[1.92]
Stock Characteristics	Yes	Yes	Yes	Yes	Yes
Lagged Returns	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes
Observations	456,086	456,086	456,086	456,086	456,086
R-squared	0.010	0.009	0.010	0.009	0.010

	$\text{OIB}^{BHJOS}_{i,t}$ (%)				
	(1)	(2)	(3)	(4)	(5)
$\text{Sentiment}^{TA}_{i,t}$	0.798^{***}				0.670^{***}
	[18.51]				[15.47]
$\text{Sentiment}^{FA}_{i,t}$		0.443^{***}			0.307^{***}
		[10.21]			[7.01]
$\text{Sentiment}^{OS}_{i,t}$			0.644^{***}		0.505^{***}
			[17.43]		[13.69]
$\text{Sentiment}^{NS}_{i,t}$				0.609^{***}	0.469^{***}
				[13.82]	[10.50]
$\text{Attention}_{i,t}$	0.614^{***}	0.604^{***}	0.608^{***}	0.621^{***}	0.647^{***}
	[17.07]	[16.89]	[16.94]	[17.18]	[17.68]
Earnings $\text{News}_{i,t}$	-0.018	-0.033	-0.020	-0.025	-0.017
	[-0.26]	[-0.50]	[-0.30]	[-0.38]	[-0.25]
Analyst $\text{News}_{i,t}$	0.088^{*}	0.091^{*}	0.088^{*}	0.088^{*}	0.094^{*}
	[1.78]	[1.83]	[1.76]	[1.76]	[1.88]
Stock Characteristics	Yes	Yes	Yes	Yes	Yes
Lagged Returns	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes
Observations	435,328	435,328	435,328	435,328	435,328
R-squared	0.018	0.017	0.018	0.017	0.019

Table 9: **Retail Strategies and Retail Order Flow Informativeness**

This table shows that the return predictive power of retail order imbalance is weaker (stronger) when StockTwits users heavily post messages about technical (fundamental) analysis. We estimate the following panel regression at the stock-week level:

$$Return_{i,t+1} = \beta_1 OIB_{i,t} + \beta_2 High\ Retail^{type}_{i,t} + \beta_3 OIB_{i,t} \times High\ Retail^{type}_{i,t} + \beta_4 Attention_{i,t}$$
$$+ \beta_5 Earnings\ News_{i,t} + \beta_6 Analyst\ News_{i,t} + \gamma X_{i,t} + \delta_t + \eta_j + \epsilon_{i,t+1}.$$

Panels A and B report results from the regressions in which $OIB_{i,t}$ is measured based on the procedure in Boehmer et al. (2021) and Barber et al. (2023), respectively. $High\ Retail^{type}_{i,t}, type = TA, FA, OS, NS$, is a dummy variable equal to 1 if the percentage of messages related to an investment approach on stock i in week t is above the median in the cross section of stocks. See Table 5 for definitions of other explanatory and control variables. All specifications include calendar week fixed effects. Standard errors are clustered by calendar week and associated t-statistics are reported in brackets, where *, **, and *** denote significance at levels 10%, 5%, and 1%.

	Return$_{i,t+1}$ (%)				
	(1)	(2)	(3)	(4)	(5)
OIB$^{BJZZ}_{i,t}$	0.485***	0.647***	0.192	0.447***	0.566***
	[5.10]	[4.95]	[1.41]	[3.90]	[4.96]
High Retail$^{TA}_{i,t}$		0.010			
		[0.27]			
OIB$^{BJZZ}_{i,t}$ × High Retail$^{TA}_{i,t}$		-0.401**			
		[-2.07]			
High Retail$^{FA}_{i,t}$			0.112***		
			[3.02]		
OIB$^{BJZZ}_{i,t}$ × High Retail$^{FA}_{i,t}$			0.603***		
			[3.20]		
High Retail$^{OS}_{i,t}$				-0.036	
				[-1.12]	
OIB$^{BJZZ}_{i,t}$ × High Retail$^{OS}_{i,t}$				0.100	
				[0.51]	
High Retail$^{NS}_{i,t}$					-0.095***
					[-2.76]
OIB$^{BJZZ}_{i,t}$ × High Retail$^{NS}_{i,t}$					-0.166
					[-0.86]
Attention$_{i,t}$	-0.121***	-0.121***	-0.121***	-0.120***	-0.114***
	[-4.26]	[-4.24]	[-4.24]	[-4.22]	[-4.05]
Earnings News$_{i,t}$	0.121**	0.121**	0.116**	0.121**	0.121**
	[2.20]	[2.20]	[2.12]	[2.20]	[2.19]
Analyst News$_{i,t}$	-0.025	-0.024	-0.022	-0.023	-0.021
	[-0.62]	[-0.62]	[-0.57]	[-0.59]	[-0.54]
Stock Characteristics	Yes	Yes	Yes	Yes	Yes
Lagged Returns	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes
Observations	455,729	455,729	455,729	455,729	455,729
R-squared	0.120	0.120	0.120	0.120	0.120

	Return$_{i,t+1}$ (%)				
	(1)	(2)	(3)	(4)	(5)
$\text{OIB}^{BHJOS}_{i,t}$	0.451***	0.553***	0.183	0.391***	0.619***
	[3.92]	[3.85]	[1.26]	[3.01]	[4.74]
$\text{High Retail}^{TA}_{i,t}$		0.021			
		[0.58]			
$\text{OIB}^{BHJOS}_{i,t} \times \text{High Retail}^{TA}_{i,t}$		-0.254			
		[-1.38]			
$\text{High Retail}^{FA}_{i,t}$			0.099***		
			[2.77]		
$\text{OIB}^{BHJOS}_{i,t} \times \text{High Retail}^{FA}_{i,t}$			0.548***		
			[3.06]		
$\text{High Retail}^{OS}_{i,t}$				-0.047	
				[-1.54]	
$\text{OIB}^{BHJOS}_{i,t} \times \text{High Retail}^{OS}_{i,t}$				0.164	
				[0.90]	
$\text{High Retail}^{NS}_{i,t}$					-0.085**
					[-2.55]
$\text{OIB}^{BHJOS}_{i,t} \times \text{High Retail}^{NS}_{i,t}$					-0.346*
					[-1.92]
$\text{Attention}_{i,t}$	-0.099***	-0.098***	-0.098***	-0.097***	-0.092***
	[-3.99]	[-3.95]	[-3.98]	[-3.93]	[-3.77]
$\text{Earnings News}_{i,t}$	0.095*	0.096*	0.091*	0.095*	0.095*
	[1.83]	[1.83]	[1.76]	[1.83]	[1.83]
$\text{Analyst News}_{i,t}$	-0.023	-0.023	-0.021	-0.021	-0.020
	[-0.65]	[-0.64]	[-0.60]	[-0.61]	[-0.56]
Stock Characteristics	Yes	Yes	Yes	Yes	Yes
Lagged Returns	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes
Observations	435,107	435,107	435,107	435,107	435,107
R-squared	0.135	0.135	0.135	0.135	0.135

6. Appendix A

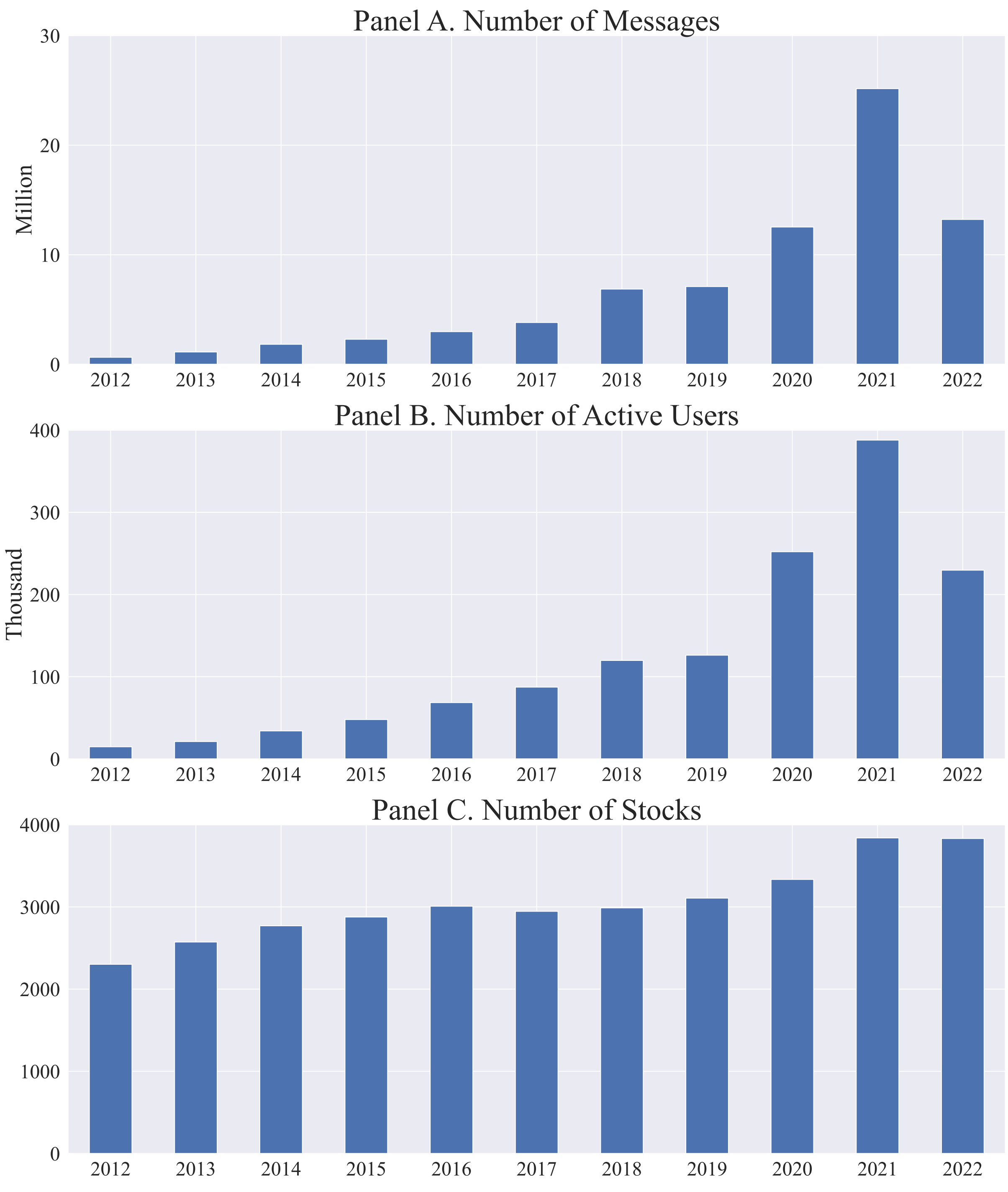


Fig. A.1. StockTwits Coverage

This figure plots the number of messages, active users, and stocks on StockTwits from 2012 to 2022.

Table A.1: **Examples of GPT Responses**

No.	Message	Ticker	Score	Indicators
1	$IOVA Biotechnology Company, Phase 2, Hammer, Support Line, Oversold, JMP Securities $38, Q4: Institutional Bought $77M, Sold $13M, Speculation Trade, Entry: Above $24	IOVA	2	Hammer, Support Line, Oversold
2	$CVS if it can hold firmly above $106 will signal entry at the close as well. Stops tight at $104	CVS	2	Support Level, Stop Loss
3	$RETA 10 wk SMA has caught up. $300 stock btw, Livermore's finest	RETA	2	10 wk SMA
4	RT @mentholatum $AAPL the oversold compression on AAPL will release... another $50 up day maybe.... when????.... Someday soon// Bold call	AAPL	2	Oversold Compression
5	$AAPL next retracement $100.36 which is 38.2% of the move down. should be coming within next hr	AAPL	2	Fibonacci Retracement
6	$ACOR Acorda Therapeutics (ACOR, $8.65) was this week's top stock market loser, declining -10%. Expect a Downtrend reversal	ACOR	1	Downtrend Reversal
7	$META Bout to break the big $100 level then breakdown further.	META	1	Breakdown
8	$SAIC Science Applications International Corporation (SAIC) has been systematically hitting all-time highs in the last 10 days. Science Applications International Corporation (SAIC) price climbed on Wednesday a 2.17% ending at $103.10 and marking the n	SAIC	1	All-time Highs
9	$PETS new retail shorts probably got in at 35 or lower, this will fly on short covering above $38.50ish when most down over 10%	PETS	1	Short Covering
10	10:27:29 AM Makes fresh HOD $CARA $19.55 +12.2% ON 1,400K VOL (ISW Pre-Market Watch/Scan)	CARA	1	HOD, Volume
11	$TSLA added more under $890 ... well it has been while since last time I played with TSLA... I just love how their earning growing and what ELON said... I still expect volatile days but worth to start adding... GL	TSLA	0	-
12	$MSFT Lmaooo you bears are dumb as shit. I sold all my Bitcoin to buy shares at $275 hand over fist.	MSFT	0	-
13	$MU I picked up some of the $25s for a punt...Company is undervalued massively...if they deliver, this soars > 15%.	MU	0	-
14	$ETSY at $13.66 - Sell Stock Market Alert sent at 10:14 AM ET #stocks	ETSY	0	-

Table A.2: **Variable Definitions**

Variable	**Definition**
Retail TA Usage$_{i,j,t}$	Fraction of total messages posted by investor j on stock i in week t that are classified as technical related.
Sentiment$_{i,j,t}$	The difference in the number of bullish and bearish messages to the sum of bullish and bearish messages on stock i posted by investor j in week t, $\frac{N^{Bullish}_{i,j,t}-N^{Bearish}_{i,j,t}}{N^{Bullish}_{i,j,t}+N^{Bearish}_{i,j,t}}$ following Cookson and Niessner (2020).
Sentiment$^{TA}_{i,t}$	Sentiment calculated using messages related to technical analysis.
Sentiment$^{NonTA}_{i,t}$	Sentiment calculated using non-technical messages.
Sentiment$^{FA}_{i,t}$	Sentiment calculated using messages related to fundamental analysis.
Sentiment$^{OS}_{i,t}$	Sentiment calculated using messages that are not related to other strategies.
Sentiment$^{NS}_{i,t}$	Sentiment calculated using messages that are not related to technical analysis, fundamental analysis, or other strategies.
Technical Investor$_{i,t}$	Dummy variable equal to one if investor i's self-reported approach is "Technical" or "Momentum".
Short-term Investor$_{i,t}$	Dummy variable equal to one if investor i's self-reported holding period is "Day Trader", "Swing Trader", or "Position Trader".
Professional Investor$_{i,t}$	Dummy variable equal to one if investor i's self-reported experience is "Professional".
Attention$_{i,t}$	A measure of StockTwits users' attention on stock i in week t, defined as the number of messages on stock i divided by the total number of messages across all stocks, i.e., $Attention_{i,t} = \frac{\#Messages_{i,t}}{\sum_i \#Messages_{i,t}}$ (Cookson et al., 2024b).
AI Signal$_{i,t}$	Jiang et al. (2023) weekly return forecast generated by applying a convolutional neural network (CNN) to price patterns over the past 5 days.
OIB$_{i,t}$	Retail marketable volume imbalance on stock i in week t following Boehmer et al. (2021) (BJZZ) or Barber et al. (2023) (BHJOS).
RH Herding$_{i,t}$	Indicator for top 1% of positive Robinhood user change ratio in week t and a minimum of 100 users at the end of week $t-1$ following Barber et al. (2022).
Earnings News$_{i,t}$	Indicator for the release of earnings-related news about stock i in week t.
Analyst News$_{i,t}$	Indicator for the release of analyst-related news about stock i in week t.
Max$_{i,t}$	Maximum return in the prior month.
Abnormal Turnover$_{i,t}$	Measure of abnormal trading volume, $\log(1+Turnover_{i,t}) - \log(1+\frac{1}{4}\sum_{h=1}^{4} Turnover_{i,t-h})$.
Market Capitalization$_{i,t}$	Market capitalization.
Book-to-Market$_{i,t}$	Ratio of book value to market value.
Asset Growth$_{i,t}$	Growth rate of annual total assets.
Gross Profits-to-Asset$_{i,t}$	Ratio of gross profits to total assets.
Number of Analysts$_{i,t}$	Number of IBES equity analysts covering stock i.
Institutional Ownership$_{i,t}$	Fraction of shares outstanding held by 13F institutional investors.

Table A.3: **Robustness: Sentiment and Lagged AI Signal**

We repeat Table 3 by regressing retail investors' sentiments on the lagged, rather than contemporaneous, AI Signal:

$$\begin{aligned}Sentiment^{type}_{i,j,t} =& \beta_1 AI\ Signal_{i,t-1} + \beta_2 Return^{1d}_{i,t-1} + \beta_3 Return^{1w}_{i,t-1} + \beta_4 Return^{1m}_{i,t-1} + \beta_5 Return^{1q}_{i,t-1} \\ &+ \beta_6 Return^{1y}_{i,t-1} + \beta_7 Earnings\ News_{i,t-1} + \beta_8 Analyst\ News_{i,t-1} \\ &+ \gamma \boldsymbol{X}_{i,t} + \delta_t + \eta_j + \epsilon_{i,j,t}, \quad type = TA, NonTA\end{aligned}$$

See Table 3 for the definitions of the dependent and explanatory variables. All specifications include both calendar week and investor fixed effects. Standard errors are clustered by investor and calendar week, and associated t-statistics are reported in brackets, where *, **, and *** denote significance at levels 10%, 5%, and 1%.

	$\text{Sentiment}^{TA}_{i,j,t}$		$\text{Sentiment}^{NonTA}_{i,j,t}$	
	(1)	(2)	(3)	(4)
AI $\text{Signal}_{i,t-1}$	-0.503***	-0.520***	-0.214***	-0.206***
	[-10.25]	[-10.29]	[-3.74]	[-3.40]
$\text{Return}^{1d}_{i,t-1}$		0.015		0.009
		[0.65]		[0.33]
$\text{Return}^{1w}_{i,t-1}$		-0.027**		-0.016
		[-2.13]		[-0.90]
$\text{Return}^{1m}_{i,t-1}$		0.009		0.008
		[1.20]		[0.82]
$\text{Return}^{1q}_{i,t-1}$		0.017***		0.030***
		[3.32]		[6.28]
$\text{Return}^{1y}_{i,t-1}$		-0.024***		-0.020***
		[-9.71]		[-6.20]
Earnings $\text{News}_{i,t-1}$	-0.043***	-0.042***	-0.052***	-0.051***
	[-8.29]	[-8.38]	[-7.69]	[-7.64]
Analyst $\text{News}_{i,t-1}$	-0.067***	-0.065***	-0.079***	-0.078***
	[-13.03]	[-13.20]	[-11.10]	[-11.03]
Stock Characteristics	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes
Investor FEs	Yes	Yes	Yes	Yes
Observations	1,532,892	1,532,892	6,118,230	6,118,230
R-squared	0.193	0.195	0.165	0.166

Table A.4: **Robustness: Self-reported Sentiment Labeling**

This table focuses on investors' sentiments that are measured only from messages with StockTwits users' self-reported labeling in bullish and bearish.

$$\begin{aligned} Sentiment^{SelfLabel,type}_{i,j,t} = & \beta_1 AI\ Signal_{i,t} + \beta_2 Return^{1d}_{i,t-1} + \beta_3 Return^{1w}_{i,t-1} + \beta_4 Return^{1m}_{i,t-1} + \beta_5 Return^{1q}_{i,t-1} \\ & + \beta_6 Return^{1y}_{i,t-1} + \beta_7 Earnings\ News_{i,t-1} + \beta_8 Analyst\ News_{i,t-1} \\ & + \gamma X_{i,t} + \delta_t + \eta_j + \epsilon_{i,j,t}, \quad type = TA, NonTA \end{aligned}$$

where the dependent variable is the sentiment of investor j on stock i in week t, measured by the number of investor j's messages with self-reported bullish and bearish labeling, $Sentiment_{i,j,t} = \frac{N^{Bullish}_{i,j,t} - N^{Bearish}_{i,j,t}}{N^{Bullish}_{i,j,t} + N^{Bearish}_{i,j,t}}$ (see, Cookson et al., 2024b). In columns (1) and (2), the sentiment is measured based on technical analysis (TA)-related messages, while columns (3) and (4) focus on the sentiment of non-TA messages. See Table A.3 for definitions of explanatory and control variables. All specifications include both calendar week and investor fixed effects. Standard errors are clustered by investor and calendar week, and associated t-statistics are reported in brackets, where *, **, and *** denote significance at levels 10%, 5%, and 1%.

	$\text{Sentiment}^{TA}_{i,j,t}$		$\text{Sentiment}^{NonTA}_{i,j,t}$	
	(1)	(2)	(3)	(4)
AI Signal$_{i,t}$	-1.198***	-1.151***	-0.643***	-0.611***
	[-9.45]	[-9.69]	[-5.75]	[-5.63]
Return$^{1d}_{i,t-1}$		0.017		0.037
		[0.46]		[1.12]
Return$^{1w}_{i,t-1}$		0.006		-0.018
		[0.40]		[-1.06]
Return$^{1m}_{i,t-1}$		0.023**		0.021***
		[2.28]		[2.72]
Return$^{1q}_{i,t-1}$		0.015*		0.000
		[1.77]		[0.04]
Return$^{1y}_{i,t-1}$		-0.003		0.000
		[-1.06]		[0.09]
Earnings News$_{i,t-1}$	-0.027***	-0.026***	-0.017***	-0.016***
	[-4.09]	[-4.01]	[-3.21]	[-3.18]
Analyst News$_{i,t-1}$	-0.031***	-0.031***	-0.037***	-0.036***
	[-4.89]	[-4.83]	[-6.16]	[-6.11]
Stock Characteristics	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes
Investor FEs	Yes	Yes	Yes	Yes
Observations	137,501	137,501	291,528	291,528
R-squared	0.366	0.367	0.454	0.454

Table A.5: **Value-weighted Portfolios Sorted on Retail TA and AI Signal**

This table presents the average DGTW-adjusted returns of the value-weighted portfolios two-way sorted on retail TA sentiment and AI signal. Retail TA Bull (Bear) denotes the portfolio of stocks with retail TA sentiment greater (less) than its cross-sectional median. Retail Bull–Bear indicates the return difference of the Retail Bull and Bear portfolios. We also form portfolios based on retail non-TA sentiment. AI Buy (Sell) is the portfolio of stocks with AI Signal greater (less) than 0.5. AI Buy–Sell indicates the return difference between the AI Buy and Sell portfolios. Columns (5) and (6) focus on the two long-short strategies, one in which retail investors agree with AI in technical analysis (Buy/Bull–Sell/Bear), and the other in which retail investors disagree with AI (Buy/Bear–Sell/Bull). The average DGTW-adjusted returns are annualized in percentage and associated t-statistics are reported in brackets, where *, **, and *** denote significance at levels 10%, 5%, and 1%.

	(1)	(2)	(3)	(4)	(5)	(6)
	Retail TA					
	Bull		Bear		Buy/Bull	Buy/Bear
AI Signal	Buy	Sell	Buy	Sell	– Sell/Bear	– Sell/Bull
$r^{DGTW-adj}$	1.46	-4.36	0.68	1.19	0.26	5.05
	[1.28]	[-3.30]	[0.51]	[0.71]	[0.12]	[2.52]

	Retail Non-TA					
	Bull		Bear		Buy/Bull	Buy/Bear
AI Signal	Buy	Sell	Buy	Sell	– Sell/Bear	– Sell/Bull
$r^{DGTW-adj}$	1.42	-2.97	1.92	-2.18	3.60	4.89
	[1.27]	[-1.92]	[1.81]	[-1.79]	[1.99]	[2.59]

Table A.6: **Retail Usage of Fundamental Analysis and Other Strategies**

We repeat Table 2 to examine the usage of fundamental analysis (FA) and other strategies (OS) revealed by their posts on StockTwits. We estimate the following panel regression at the stock-investor-week level:

$$
\begin{aligned}
\textit{Strategy Usage}_{i,j,t+1} = & \beta_1 \textit{Technical Investor}_{j,t} + \beta_2 \textit{Short-term Investor}_{j,t} + \beta_3 \textit{Professional Investor}_{j,t} \\
& + \beta_4 \textit{Earnings News}_{i,t} + \beta_5 \textit{Analyst News}_{i,t} + \gamma \mathbf{X}_{i,t} + FEs + \epsilon_{i,j,t+1}.
\end{aligned}
$$

In columns (1)–(2), the dependent variable is the FA usage by investor j on stock i in week $t+1$, measured as the percentage of investor j's messages classified as FA-related by our classification LLM (TechBERT). Similarly, the dependent variable is investor j's use of other strategies in columns (3)–(4). See Table 2 for details on explanatory and control variables in the regressions. Columns (1) and (3) include calendar week fixed effects and columns (2) and (4) include both calendar week and investor fixed effects. Standard errors are clustered by investor and calendar week, and associated t-statistics are reported in brackets, where *, **, and *** denote significance at levels 10%, 5%, and 1%.

	Retail FA Usage$_{i,j,t+1}$		Retail OS Usage$_{i,j,t+1}$	
	(1)	(2)	(3)	(4)
Technical Investor$_{j,t}$	-0.046***		0.004	
	[-7.91]		[0.83]	
Short-term Investor$_{j,t}$	-0.019***		0.022***	
	[-3.23]		[5.86]	
Professional Investor$_{j,t}$	0.081***		0.021***	
	[5.76]		[3.82]	
Earnings News$_{i,t}$	0.016***	0.014***	-0.003***	-0.003***
	[9.25]	[13.15]	[-5.55]	[-6.20]
Analyst News$_{i,t}$	-0.014***	-0.004***	0.001	0.001
	[-7.43]	[-4.77]	[1.12]	[1.49]
Stock Characteristics	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes
Investor FEs	No	Yes	No	Yes
Observations	16,994,396	16,818,446	16,994,396	16,818,446
R-squared	0.026	0.276	0.006	0.114

Table A.7: **Longer Horizon Return Predictability**

This table examines the longer horizon return predictability of sentiments of StockTwits messages corresponding to various retail investment strategies. We estimate the panel regressions at the stock-week level,

$$\begin{aligned} Return_{i,t+k} =& \beta_1 Sentiment^{type}_{i,t} + \beta_2 Attention_{i,t} + \beta_3 Earnings\ News_{i,t} \\ &+ \beta_4 Analyst\ News_{i,t} + \gamma X_{i,t} + \delta_t + \eta_j + \epsilon_{i,t+1}, \quad k = 2, 3, 4 \end{aligned}$$

where $Sentiment^{type}, type = TA, FA, OS, NS$, represent the sentiments of the four message types: technical analysis (TA), fundamental analysis (FA), other strategies (OS), and non-strategy (NS). See Table 5 for definitions of other explanatory and control variables. All specifications include calendar week fixed effects. Standard errors are clustered by calendar week and associated t-statistics are reported in brackets, where *, **, and *** denote significance at levels 10%, 5%, and 1%.

	Return$_{i,t+2}$ (%)	Return$_{i,t+3}$ (%)	Return$_{i,t+4}$ (%)
	(1)	(2)	(3)
Sentiment$^{TA}_{i,t}$	-0.045	-0.030	-0.070*
	[-1.48]	[-0.94]	[-1.88]
Sentiment$^{FA}_{i,t}$	0.064**	-0.032	0.050
	[2.15]	[-0.87]	[1.30]
Sentiment$^{OS}_{i,t}$	-0.043	-0.022	0.006
	[-1.54]	[-0.78]	[0.18]
Sentiment$^{NS}_{i,t}$	0.023	0.019	0.060*
	[0.90]	[0.60]	[1.66]
Attention$_{i,t}$	-0.040	-0.016	-0.015
	[-1.26]	[-0.41]	[-0.43]
Earnings News$_{i,t}$	-0.071	-0.049	-0.062
	[-1.08]	[-0.74]	[-0.85]
Analyst News$_{i,t}$	0.036	-0.005	0.027
	[0.78]	[-0.10]	[0.52]
Stock Characteristics	Yes	Yes	Yes
Lagged Returns	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes
Observations	476,506	409,121	369,086
R-squared	0.125	0.126	0.127