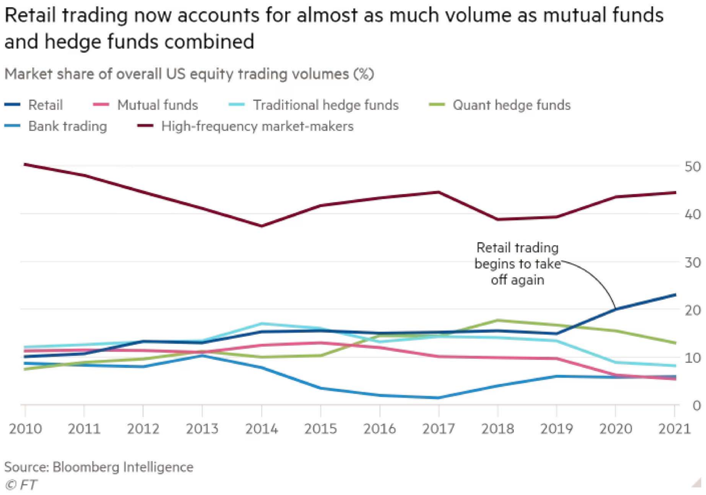
Wisdom or Whims?
Decoding the Language of Retail Trading with Social Media and AI

Shuaiyu Chen[a] Lin Peng[b] Dexin Zhou[b]

ABFER, May 2025

[a]Purdue University, [b]Baruch College, CUNY

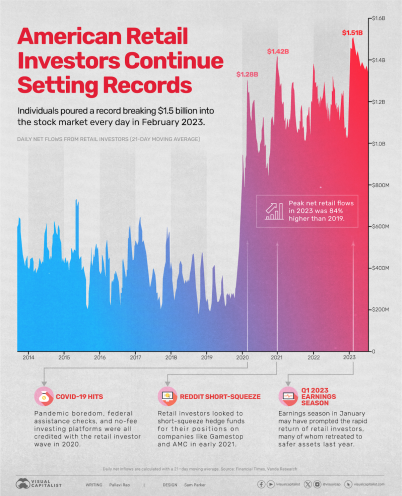
Tremendous Growth in Retail Trading



– Retail trading now matches the combined volume of mutual and hedge funds.

– Significant surge beginning around 2019, driven by fintech platforms and commission-free trading.

Record-Breaking Retail Investor Inflows



- In February 2023, retail investors poured a record $1.5 billion daily into the stock market.
- A dramatic 500% increase compared to pre-pandemic investment levels.
- Influenced by commission-free trading and events like the GameStop short squeeze.

Decoding Retail Investor Behavior: Beliefs and Strategies

- **Previous studies** typically explore retail investor behavior using:
  - **Trading account data** (*Barber and Odean, 2000, 2008*)
    - Insights on behavioral biases (e.g., overconfidence), but becoming outdated
  - **Aggregate retail order flows and Robinhood users' holdings** (*Kelley & Tetlock, 2013; Boehmer et al., 2021; Welch, 2022; Barber et al., 2024*)
    - Capturing collective retail impact, but missing investor-level profiles
  - **Survey-based methods** (*Choi & Robertson, 2020; Giglio et al., 2021; Chinco et al., 2022*)
    - Detailed beliefs, rich heterogeneity, but limited by small samples and static nature

Decoding Retail Investor Behavior: Beliefs and Strategies

- **Previous studies** typically explore retail investor behavior using:
  - **Trading account data** (*Barber and Odean, 2000, 2008*)
    - Insights on behavioral biases (e.g., overconfidence), but becoming outdated
  - **Aggregate retail order flows and Robinhood users' holdings** (*Kelley & Tetlock, 2013; Boehmer et al., 2021; Welch, 2022; Barber et al., 2024*)
    - Capturing collective retail impact, but missing investor-level profiles
  - **Survey-based methods** (*Choi & Robertson, 2020; Giglio et al., 2021; Chinco et al., 2022*)
    - Detailed beliefs, rich heterogeneity, but limited by small samples and static nature

- **Open Question:** Why do retail investors behave as they do?
  - Known behaviors, but unclear decision-making processes

Decoding Retail Investor Behavior: Beliefs and Strategies

- **Previous studies** typically explore retail investor behavior using:
  - **Trading account data** (*Barber and Odean, 2000, 2008*)
    - Insights on behavioral biases (e.g., overconfidence), but becoming outdated
  - **Aggregate retail order flows and Robinhood users' holdings** (*Kelley & Tetlock, 2013; Boehmer et al., 2021; Welch, 2022; Barber et al., 2024*)
    - Capturing collective retail impact, but missing investor-level profiles
  - **Survey-based methods** (*Choi & Robertson, 2020; Giglio et al., 2021; Chinco et al., 2022*)
    - Detailed beliefs, rich heterogeneity, but limited by small samples and static nature

- **Open Question:** Why do retail investors behave as they do?
  - Known behaviors, but unclear decision-making processes

- **Our Paper:** Leveraging LLMs and social media data
  - StockTwits: 100M messages, 800K users, 2010–2023
  - Capturing dynamic evolutions in investor strategies and sentiment

Results in a Nutshell

- **Classified investor strategies** using GPT-4 Turbo + BERT
  - 30% of posts related to specific investment strategies

Results in a Nutshell

- **Classified investor strategies** using GPT-4 Turbo + BERT
  - 30% of posts related to specific investment strategies
- **Retail strategies vary widely in type and performance**
  - **Technical Analysis (TA)**
    - 28% of strategy-related posts; highly diverse among users
    - TA sentiment negatively predicts future returns
    - A strategy betting against retail TA earns an annual return of 10% and a SR of 0.86
  - **Fundamental Analysis (FA)**
    - 44% of strategy-related posts
    - FA sentiment positively predicts future returns
    - A strategy trading with retail FA earns an annual return of 7.5% and a SR of 0.56
  - **Other Strategies (OS)**
    - 28% of strategy-related posts
    - Many related to options trading

Results in a Nutshell

- **Classified investor strategies** using GPT-4 Turbo + BERT
  - 30% of posts related to specific investment strategies

- **Retail strategies vary widely in type and performance**
  - **Technical Analysis (TA)**
    - 28% of strategy-related posts; highly diverse among users
    - TA sentiment negatively predicts future returns
    - A strategy betting against retail TA earns an annual return of 10% and a SR of 0.86
  - **Fundamental Analysis (FA)**
    - 44% of strategy-related posts
    - FA sentiment positively predicts future returns
    - A strategy trading with retail FA earns an annual return of 7.5% and a SR of 0.56
  - **Other Strategies (OS)**
    - 28% of strategy-related posts
    - Many related to options trading

- **Other findings**
  - Short-term sentiment generally leads to poorer return predictions for all strategy types
  - Experienced investors perform better using TA, but have no significant improvement in FA
  - TA sentiment's negative return prediction worsens after GameStop event
  - TA sentiment strongly linked to retail investor herding (e.g., Robinhood frenzies)

Contribution

- Contributes to the literature on social media and finance
  - *Cookson and Niessner (2021); Cookson et al. (2024); Cookson, Niessner & Schiller (2024); Cookson, Fos & Niessner (2025)*
    Note: Some tests in these papers classify technical/fundamental based on the topic modeling
  - Our LLM-based classifier outperforms traditional textual analysis (e.g., Bag-of-Words), yielding high agreement with human labeling based on validations

- Deepens understanding of retail investor trading strategies
  - Reveals significant heterogeneity even within self-reported investment styles
  - Highlights strategy-specific differences in the informativeness of retail sentiment
  - Helps explain other phenomena, like why AI-based technical trading strategies are profitable (*Jiang, Kelly & Xu, 2023; Murray, Xia & Xiao, 2024*)

- Expands understanding of retail trading informativeness
  - *Barber et al. (2022); Welch (2022); Eaton et al. (2022); Bradley et al. (2024)*
  - Informativeness strongly depends on specific strategies used by investors

- LLMs in economics and finance
  - *Korinek (2023); Fedyk et al. (2024)*
  - LLMs enable powerful analyses of extensive social media datasets
  - Overcomes traditional limitations of surveys and individual trading data

Data & Methods

Data

- US stocks, Jan 2010–Jun 2023
- **StockTwits:** Largest investor-focused social media platform
  - *Cookson and Niessner (2020), Cookson et al. (2024)*
  - Extensive dataset: 96 million messages, 840K users, 7,800 stocks
  - Messages: timestamps, ticker symbols (”cashtags”, e.g., $TSLA)
  - Sentiments: User-labeled (bullish/bearish), imputed if missing (using BERT or method introduced by *Cookson & Niessner (2020)*)
  - User Profiles: self-reported investment approach, holding horizon, and experience level
    - Available for 19% of users ▸ By Users
    - Available for 35% of messages ▸ By Messages
- Additional data sources: CRSP, Compustat, Ravenpack, TAQ, RobinTrack

Classifying Investment Strategies with LLMs

- Classify investor social media posts into major strategy types
  - Technical Analysis (TA)
  - Fundamental Analysis (FA)
  - Other Strategies (OS)
  - Non-Strategy (NS)
- Two-step classification via Knowledge Distillation (*Hinton (2015); Gu et al. (2023))*
  - "Teacher Model": Classify a random small sample (**20,000** messages) using GPT-4 Turbo
  - "Student Model": Fine-tune a BERT model on GPT-4 responses for large-scale classification
- Validation
  - GPT-4 Turbo outperforms traditional textual methods
  - High agreement with human labeling
  - Procedure similarly applied to FA and other strategy types

Step 1: *Teacher Model* - GPT4

Prompt provided to GPT-4 Turbo:

> *"You have a deep understanding of the language of social media and financial markets. Please analyze the message from an investor social media platform. 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?"*

- Cost-effective classification: total cost ≈ $70
- Validation
  - GPT-4 outperforms traditional Bag-of-Words (BoW) approach
  - GPT-human agreement: 90% correlation based on a random sample of 450 messages with GPT-BoW disagreement ▸ Validation
- GPT-4 performance also validated in recent studies:
  - Investment preferences: *Fedyk, Kakhbod, Li, Malmendier (2024)*
  - FOMC interpretations: *Hansen & Kazinnik (2024)*

Example: GPT-4 Responses

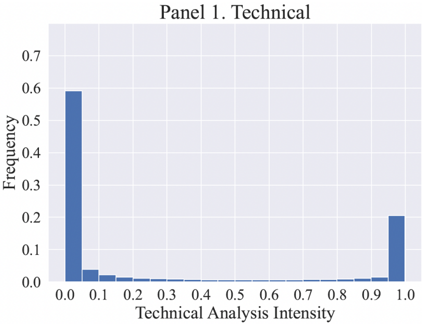
No.	Message	Ticker	Score	Indicators
1	$IOVA Biotechnology Company, Phase 2, Hammer, Support Line, Oversold, JMP Securities $38, Entry: Above $24	IOVA	2	Hammer, Support Line, Oversold
2	$CVS if it can hold firmly above $106 will signal entry at close. Stops tight at $104	CVS	2	Support Level, Stop Loss
3	$RETA 10 wk SMA has caught up. $300 stock btw, Livermore's finest	RETA	2	10 wk SMA
4	$ACOR this week's top loser (-10%). Expect Downtrend reversal	ACOR	1	Downtrend Reversal
5	$META About to break $100 level then breakdown further	META	1	Breakdown
6	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
7	$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	-
8	$MSFT Lmaooo you bears are dumb as shit. I sold all my Bitcoin to buy shares at $275 hand over fist.	MSFT	0	-
9	$MU I picked up some of the $25s for a punt...Company is undervalued massively...if they deliver, this soars > 15%.	MU	0	-

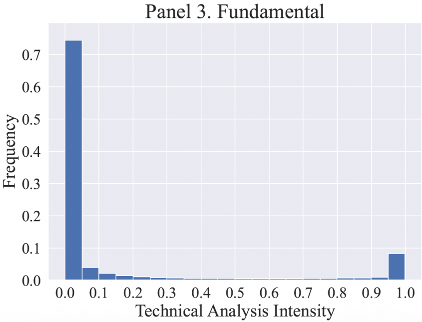
Step 2: *Student Model* — Fine-tuned BERT

- Fine-tune a smaller model (TechBERT) with GPT-4 labels
  - BERT: Robust NLP model, widely used for text classification (*Devlin et al., 2018; Gonzalez-Carvajal & Garrido-Merchan, 2020*)
  - Significant reduction in model complexity:
    - Parameters: from 1.7 trillion (GPT-4) down to 110 million (BERT)
    - Processing speed: Classifies entire dataset locally in 40 hours
- Evaluation and classification criteria:
  - Strong predictive accuracy (F1 Score: **0.83**)
  - Clearly separates TA from non-TA messages (distinct bimodal probability distribution)
  - TA classification threshold: probability $\geq 95\%$
- Why not use GPT-4 directly on all messages?
  - Estimated cost would be prohibitively high: around $500,000

Classification Shows Clear Distinctions

- TA usage by self-reported approach
- Clear bimodal distribution: low ambiguity in classification
- Self-identified "Technical" investors use more TA
- But TA still appears in messages from "Fundamental" investors
- TA classification threshold: 95%
  Our results robust across thresholds
- Distribution of technical words
  TF-IDF measure ▸ TF-IDF



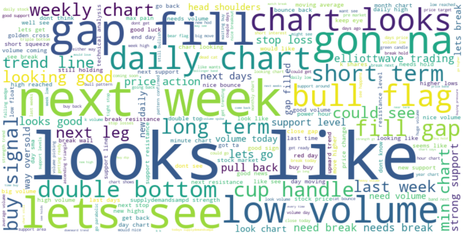


Salient Words in Technical Messages

- Top Unigram Words:
  - *chart, volume, break, support, today, day, next, back, resistance, gap, week, buy, close, short, breakout, long, low, time, bullish, hold, bounce, trend, level, move, high, strong, tomorrow, last, new*
  - Large overlap with Cookson and Niessner (2020, Table 2)
- Bigram Words Indicate Short Horizon:
  - *next week, short term*

Panel A: Unigram Word Cloud

Panel B: Bigram Word Cloud



Salient Words in Fundamental Messages

- Common terms highlight clear fundamental analysis themes:
  - *earnings, earnings call, price target, market cap, long term*
  - Emphasis on corporate valuation and financial performance

- Aligns closely with fundamental analysis topics investors usually discuss

Panel A: Unigram Word Cloud

Panel B: Bigram Word Cloud

Results

Classification of Investment Strategies

- Overall Message Classification:
  - Non-strategy related messages: 69%
  - Strategy-related messages: 31%
    - Fundamental Analysis (FA): 44%
    - Technical Analysis (TA): 28%
    - Other Strategies (OS): 28%
- Next Steps in Our Analysis:
  - What factors shape investors' strategy choices?
  - How do different investment strategies perform?
  - How does investor sentiment relate to actual trading?

Determinants of TA Usage (Message Level)

- Dep. Var: = 1 if a TA message
  - TA significantly higher (**+7%**) for self-identified technical investors
  - TA more common among short-term and professionals
- Within-user Variation:
  - Same user relies on TA more when firm-specific news are limited
  - Interesting to examine what trigger an investor to switch his strategies
- Much variation unexplained by traditional textual measures
  - Message-level TF-IDF frequency of TA/FA words (*Cookson & Niessner, 2020*)

	(1)	(2)	(3)	(4)
Technical Investor$_j$	0.076***	0.068***		
	[17.60]	[17.42]		
Swing/Day Trader$_j$	0.022***	0.022***		
	[3.89]	[4.22]		
Long-Term Investor$_j$	-0.028***	-0.023***		
	[-6.44]	[-5.65]		
Professional$_j$	0.044***	0.033***		
	[6.23]	[5.30]		
Novice$_j$	-0.018***	-0.013***		
	[-6.21]	[-5.31]		
Abnormal Turnover$_{i,t}$	-0.003***	-0.005***	-0.003***	-0.003***
	[-2.99]	[-6.37]	[-7.45]	[-8.10]
Abnormal News$_{i,t}$	-0.006***	-0.005***	-0.003***	-0.002***
	[-9.73]	[-11.86]	[-12.15]	[-10.15]
Technical$^{TF-IDF}_{i,j,t,n}$	0.825***	0.798***	0.683***	0.653***
	[40.95]	[40.26]	[42.09]	[39.56]
Fundamental$^{TF-IDF}_{i,j,t,n}$	-0.664***	-0.632***	-0.533***	-0.498***
	[-27.39]	[-26.73]	[-28.29]	[-28.52]
Log(Words$_{i,j,t,n}$)	0.047***	0.049***	0.044***	0.043***
	[16.62]	[18.92]	[28.62]	[26.76]
Date FE	No	Yes	Yes	Yes
Stock FE	No	Yes	Yes	No
Investor FE	No	No	Yes	No
Stock $\times$ Investor FE	No	No	No	Yes
Observations	21,641,362	21,641,218	21,623,813	20,630,883
R^2	0.063	0.084	0.213	0.287

Determinants of FA and OS Usage

- Fundamental Analysis (FA)
  - Less used by self-identified technical and short-term traders
  - More frequently used when firm-specific news is abundant
- Other Strategies (OS)
  - Profile similar to TA; positively associated with turnover
  - Professional investors and longer messages tend to use OS

	FA Usage		OS Usage	
	(1)	(2)	(1)	(2)
Technical Investor$_j$	-0.042***		0.014***	
	[-11.09]		[7.71]	
Swing/Day Trader$_j$	-0.015***		0.012***	
	[-3.54]		[5.23]	
Long-Term Investor$_j$	0.032***		-0.013***	
	[4.77]		[-5.90]	
Professional$_j$	0.018***		0.015***	
	[3.06]		[5.70]	
Novice$_j$	-0.016***		-0.011***	
	[-5.90]		[-7.43]	
Abnormal Turnover$_{i,t}$	-0.009***	-0.006***	0.004***	0.002***
	[-13.45]	[-14.35]	[11.89]	[9.32]
Abnormal News$_{i,t}$	0.006***	0.005***	-0.001***	-0.001***
	[11.17]	[13.76]	[-6.44]	[-7.38]
Technical$^{TF-IDF}_{i,j,t,n}$	-0.320***	-0.234***	0.220***	0.185***
	[-22.38]	[-23.63]	[15.55]	[15.10]
Fundamental$^{TF-IDF}_{i,j,t,n}$	0.760***	0.567***	-0.275***	-0.211***
	[16.87]	[17.12]	[-23.03]	[-20.94]
Log(Words$_{i,j,t,n}$)	0.137***	0.125***	0.029***	0.032***
	[47.24]	[58.25]	[23.94]	[47.78]
Date FE	Yes	Yes	Yes	Yes
Stock FE	Yes	No	Yes	No
Investor FE	No	No	No	No
Stock × Investor FE	No	Yes	No	Yes
Observations	21,641,218	20,630,883	21,641,218	20,630,883
R^2	0.161	0.325	0.021	0.145

Correlations between Sentiments Across Strategy Types

- Sentiments on messages by strategy types

$$Sentiment_{i,t} = \frac{N_{i,t}^{Bullish} - N_{i,t}^{Bearish}}{N_{i,t}^{Bullish} + N_{i,t}^{Bearish}}$$

- Low correlations among sentiments from different strategies
- Sentiment signals from each investment type provide distinct information

	TA	FA	OS	NS
TA	1.000			
FA	0.128	1.000		
OS	0.127	0.097	1.000	
NS	0.090	0.084	0.102	1.000

Predicting Next-Day Returns

– Heterogeneous Predictive Power of Investor Sentiments

  – TA and OS sentiment negatively predict next-day returns
  – FA sentiment positively predicts next-day returns
  – Non-strategy (NS) sentiment shows no predictive power

	Return$_{i,t+1}$ (%)				
	(1)	(2)	(3)	(4)	(5)
Sentiment$^{TA}_{i,t}$	-0.016**				-0.015**
	[-2.19]				[-2.25]
Sentiment$^{FA}_{i,t}$		0.014**			0.017***
		[2.17]			[2.88]
Sentiment$^{OS}_{i,t}$			-0.027***		-0.026***
			[-3.57]		[-3.73]
Sentiment$^{NS}_{i,t}$				-0.003	-0.002
				[-0.56]	[-0.30]
Attention$_{i,t}$	-0.056***	-0.056***	-0.056***	-0.056***	-0.056***
	[-5.47]	[-5.50]	[-5.46]	[-5.48]	[-5.47]
Stock Characteristics	Yes	Yes	Yes	Yes	Yes
Lagged Returns	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Observations	2,974,304	2,974,304	2,974,304	2,974,304	2,974,304
R^2	0.089	0.089	0.089	0.089	0.089

Economic Magnitudes: Trading Strategies

– Sentiment-weighted Long-Short Strategies (TA and OS sentiments multiplied by –1):

$$r^{i,L/S} = \frac{\sum_{j=1}^{N}(S_j^i - S^i)r_j}{\frac{1}{2}\sum_{j=1}^{N}\left|S_j^i - S^i\right|}, \quad \text{where } S^i = \frac{1}{N}\sum_{j=1}^{N} S_j^i, \quad i \in \{TA, FA, OS\}$$

– Annualized return and Sharpe Ratio:

  – Technical Analysis (TA): **10%, 0.86**
  – Fundamental Analysis (FA): **7.5%, 0.58**
  – Other Strategies (OS): **10%, 0.83**

Strategy	Avg. Daily Return (%)	t-statistic	Sharpe Ratio (Annual)
TA	0.04	2.91	0.86
FA	0.03	2.04	0.58
OS	0.04	2.92	0.83

Other Return Predictability Results

- Weekly Horizon Predictability:
  - No reversals observed in subsequent weeks ▸ Result
- Short vs. Long-Term Sentiment:
  - Short-term sentiment consistently underperforms across all strategy types ▸ Result
- Role of Investor Sophistication:
  - Investor sophistication reduces negative performance linked to TA
  - No effect observed for FA sentiment ▸ Result
- Impact of Market Events (GME Short-Squeeze):
  - Post-GME event, sentiment informativeness declines significantly
  - Particularly strong negative effect for TA sentiment ▸ Result

Links to Aggregate Retail Order Imbalance

- Do sentiments expressed on StockTwits reflect actual retail trading activity?
- Yes! Retail market order imbalance (OIB) strongly tied to intraday StockTwits sentiment
  - TA, FA, OS, and NS sentiments all positively correlate with OIB
  - Placebo tests: Overnight sentiments show weak links with retail OIB

	$\text{OIB}_{i,t}^{BJZZ}$ (%)	$\text{OIB}_{i,t}^{BHJOS}$ (%)
	(1)	(2)
Intraday TA Sentiment	0.732***	0.944***
	[17.64]	[24.52]
Intraday FA Sentiment	0.571***	0.586***
	[15.31]	[15.96]
Intraday OS Sentiment	0.723***	0.758***
	[19.35]	[22.67]
Intraday NS Sentiment	0.493***	0.619***
	[13.93]	[18.00]
Attention	0.384***	0.474***
	[2.74]	[3.47]
Controls	Yes	Yes
Time FE	Yes	Yes
Observations	2,974,934	2,974,934
R^2	0.009	0.012

Links to Attention-Induced Herding

- RH herding defined as stock i ranking among top 10 stocks on day t based on daily percentage increase in Robinhood users holding the stock (following *Barber et al., 2022*)

	RH Herding$_{i,t}$ (%)				
	(1)	(2)	(3)	(4)	(5)
Sentiment$_{i,t}^{TA}$	0.137***				0.117***
	[6.40]				[6.18]
Sentiment$_{i,t}^{FA}$		0.090***			0.066***
		[4.88]			[4.09]
Sentiment$_{i,t}^{OS}$			0.090***		0.066***
			[3.86]		[3.13]
Sentiment$_{i,t}^{NS}$				0.035***	0.019
				[2.76]	[1.62]
Attention$_{i,t}$	1.183***	1.185***	1.185***	1.188***	1.180***
	[3.70]	[3.70]	[3.70]	[3.70]	[3.71]
Controls	Yes	Yes	Yes	Yes	Yes
Lagged Returns	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Observations	554,877	554,877	554,877	554,877	554,877
R^2	0.013	0.013	0.013	0.013	0.013

Conclusion and Future Research

Conclusion

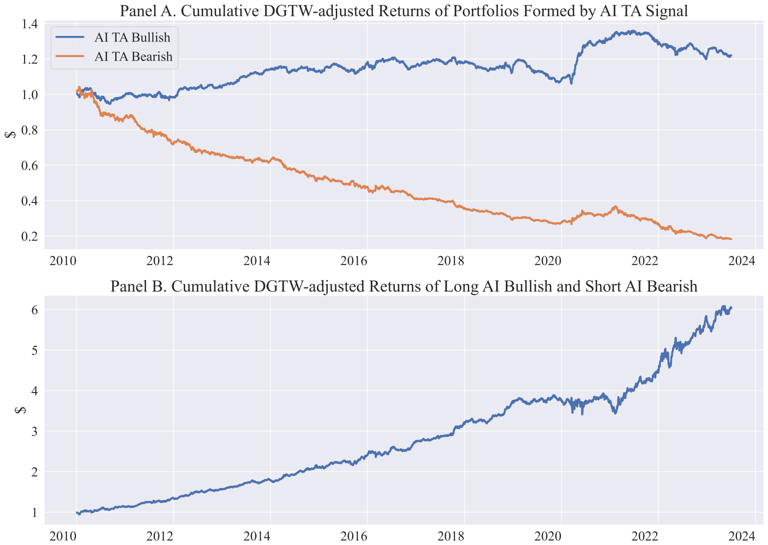
- LLMs effectively classify trading strategies from social media texts
  - Strategies exhibit substantial heterogeneity and temporal variation
- Retail trading strategies differ dramatically in predictive performance
  - Technical Analysis (TA) and Other Strategies (OS) negatively predict future returns
  - Fundamental Analysis (FA) positively predicts future returns
- Strategy classification aids in understanding investor trading
  - TA strongly associated with attention-driven trading
- Social media provides novel insights into retail investor behavior

Research Agenda

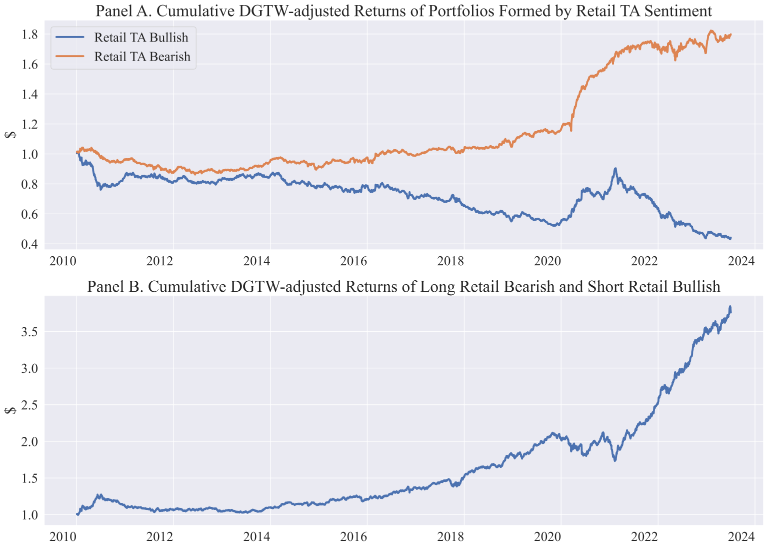
- Man vs. Machine: Technical Analysis and Trading Performance (with Lin Peng and Dexin Zhou)
  - Interaction between retail TA sentiment and AI-generated TA signals (Jiang, Kelly, & Xiu (2023); Murray, Xia, & Xiao (2024))
  - Key findings:
    - Retail TA sentiment **negatively correlated** with AI-generated signals (neural networks trained on historical price patterns)
    - AI-generated TA strategies profitable mainly by betting against retail sentiment
  - Helps explain when and why AI-powered trading is profitable
- Retail Sentiment in Option Markets (with Lin Peng, Yanbin Wu, and Dexin Zhou)
  - Currently no direct sentiment metrics in options markets – a good research opportunity!
  - Option strategy-specific sentiment metrics could help us understand informational transmission between options and underlying equities/ETFs

Performance of AI TA Strategies

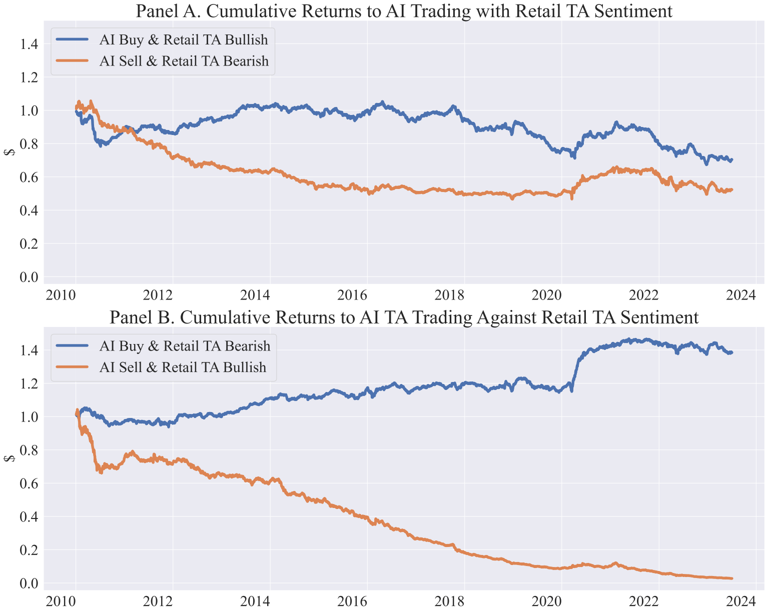
– AI TA signals: Training convolutional neural network (CNN) on historical returns (Murray, Xia, & Xiao (2024)) or directly on price charts (Jiang, Kelly, & Xiu (2023))



Performance of Retail TA Strategies



Interactions between AI and Retail TA Trading



Validation of GPT Classification

- We use the $20,000$ messages that GPT has classified
  - 2 means surely technical, 1 means possibly technical, 0 is not technical
- Bag-of-Words approach:
  - Use the dictionary of technical words in Cookson and Niessner (2020)
  - Assign technical score using TF-IDF weighting
  - Based on the technical score, we assign 2, 1, and 0 that match the distribution of the score generated by GPT
  - The classification has a 28% correlation with GPT-based classification
- Manual validation:
  - We select 450 messages where GPT and the BoW approaches do not agree
  - One RA conducted independent classification
  - Corr(GPT,Human)$= 0.9$; Corr(BoW,Human)$= -0.52$, indicating that GPT's classification is much more aligned with human.

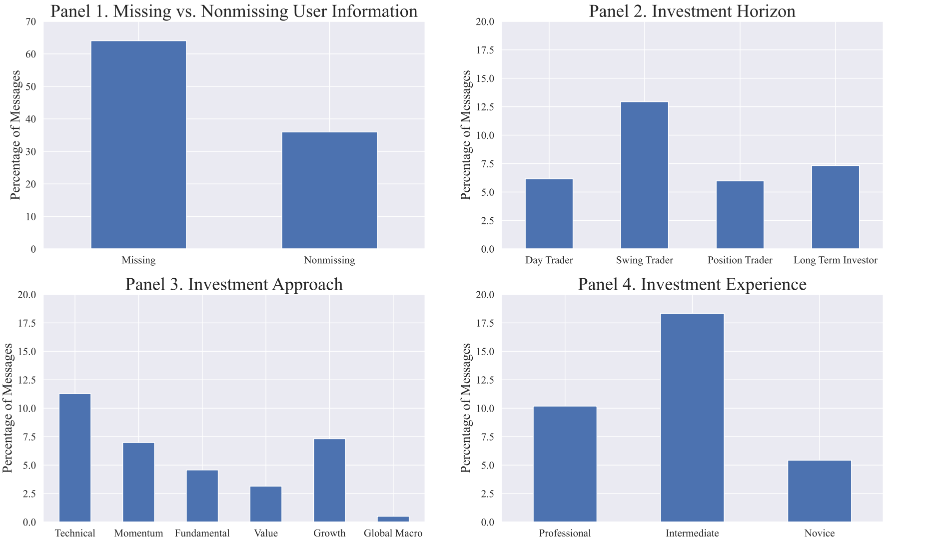
Back

Validation Examples

Message	GPT	BoW	Human
IOVA Biotecnology Company, Phase 2, Hammer, Support Line, Oversold, JMP Securities 38,*Q*4: Institutional Bought 77*M*,*Sold*13M, Speculation Trade, Entry: Above 24	2	0	2
CVS if it can hold firmly above 106 will signal entry at the close as well. Stops tight at 104	2	1	2
RETA 10 wk SMA has caught up. 300 stock btw, Livermore's finest	2	0	2
AMZN PT still 3226 today. Some nice upgrades could take it to 3250. POSSIBILITIES up in the 3300 area. Would have to see a nice pump early imo	0	1	0
ATVI Spyro the dragon news....should boost this to 53 today :)	0	2	0
LYFT are they starting to dump early? Didn't even hit 33..	0	2	0

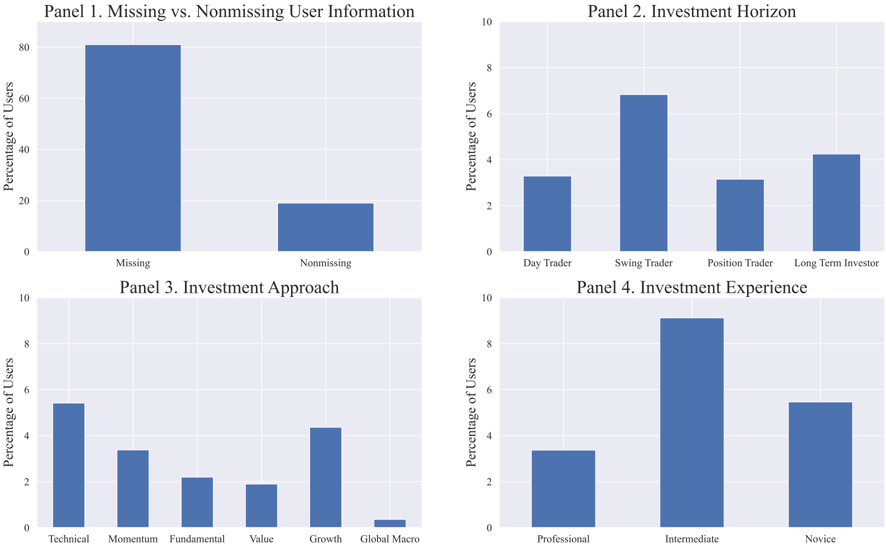
▸ Back

Message-Level Statistics



Back

User-Level Statistics



Back

Return Predictability at Longer Horizons

Panel A. Return Predictability at Longer Horizons			
	$\text{Return}_{i,t+1\rightarrow t+5}$ (%)	$\text{Return}_{i,t+6\rightarrow t+10}$ (%)	$\text{Return}_{i,t+11\rightarrow t+15}$ (%)
	(1)	(2)	(3)
$\text{Sentiment}^{TA}_{i,t}$	-0.060***	-0.056***	-0.016
	[-3.65]	[-3.65]	[-1.09]
$\text{Sentiment}^{FA}_{i,t}$	0.052***	-0.002	-0.001
	[3.88]	[-0.12]	[-0.08]
$\text{Sentiment}^{OS}_{i,t}$	-0.091***	-0.046***	-0.036**
	[-5.30]	[-2.73]	[-2.28]
$\text{Sentiment}^{NS}_{i,t}$	-0.027**	-0.001	0.007
	[-2.37]	[-0.12]	[0.64]
$\text{Attention}_{i,t}$	-0.164***	-0.083***	-0.065***
	[-8.91]	[-6.05]	[-5.34]
Stock Characteristics	Yes	Yes	Yes
Lagged Returns	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
N	2,972,286	2,970,213	2,968,174
R^2	0.105	0.108	0.112

Back

Return Predictability with Different Message Horizons

	$\text{Return}_{i,t+1\rightarrow t+5}$ (%)	$\text{Return}_{i,t+6\rightarrow t+10}$ (%)	$\text{Return}_{i,t+11\rightarrow t+15}$ (%)
	(1)	(2)	(3)
$\text{Sentiment}_{i,t}^{TA,Longterm}$	-0.045*	-0.049*	-0.016
	[-1.71]	[-1.95]	[-0.67]
$\text{Sentiment}_{i,t}^{FA,Longterm}$	0.051***	0.007	-0.030
	[2.66]	[0.37]	[-1.62]
$\text{Sentiment}_{i,t}^{OS,Longterm}$	-0.028	0.011	-0.066***
	[-1.00]	[0.38]	[-2.70]
$\text{Sentiment}_{i,t}^{TA,Daily}$	-0.149***	-0.075***	-0.033*
	[-6.50]	[-3.54]	[-1.65]
$\text{Sentiment}_{i,t}^{FA,Daily}$	-0.226***	-0.092***	-0.068**
	[-6.45]	[-2.93]	[-2.17]
$\text{Sentiment}_{i,t}^{OS,Daily}$	-0.237***	-0.086***	-0.062**
	[-8.53]	[-3.37]	[-2.50]
$\text{Attention}_{i,t}$	-0.153***	-0.078***	-0.058***
	[-8.54]	[-5.87]	[-4.96]
Stock Characteristics	Yes	Yes	Yes
Lagged Returns	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
N	2,972,286	2,970,213	2,968,174
R^2	0.105	0.108	0.112

Back

Investor Experience

	$\text{Return}_{i,t+1\to t+5}$ (%)	$\text{Return}_{i,t+6\to t+10}$ (%)	$\text{Return}_{i,t+11\to t+15}$ (%)
	(1)	(2)	(3)
$\text{Sentiment}_{i,t}^{TA}$	-0.087***	-0.089***	-0.032
	[-4.00]	[-4.37]	[-1.60]
$\text{Sentiment}_{i,t}^{FA}$	0.047***	0.007	0.001
	[2.77]	[0.40]	[0.04]
$\text{Sentiment}_{i,t}^{OS}$	-0.116***	-0.063***	-0.055***
	[-5.13]	[-2.89]	[-2.64]
$\text{Sentiment}_{i,t}^{NS}$	-0.031**	-0.007	0.016
	[-2.31]	[-0.47]	[1.20]
$\text{Sentiment}_{i,t}^{TA}$ × Fraction of Messages by $\text{Professional}_{i,t}$	0.107***	0.128***	0.063*
	[3.03]	[3.78]	[1.95]
$\text{Sentiment}_{i,t}^{FA}$ × Fraction of Messages by $\text{Professional}_{i,t}$	0.027	-0.026	-0.004
	[0.75]	[-0.77]	[-0.11]
$\text{Sentiment}_{i,t}^{OS}$ × Fraction of Messages by $\text{Professional}_{i,t}$	0.118***	0.084**	0.085**
	[2.90]	[2.15]	[2.24]
$\text{Sentiment}_{i,t}^{NS}$ × Fraction of Messages by $\text{Professional}_{i,t}$	0.035	0.032	-0.039
	[1.26]	[1.15]	[-1.45]
Fraction of Messages by $\text{Professional}_{i,t}$	0.054**	0.017	0.027
	[2.27]	[0.71]	[1.21]
$\text{Attention}_{i,t}$	-0.161***	-0.081***	-0.064***
	[-8.78]	[-5.94]	[-5.25]
Stock Characteristics	Yes	Yes	Yes
Lagged Returns	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
N	2,972,286	2,970,213	2,968,174
R^2	0.105	0.108	0.112

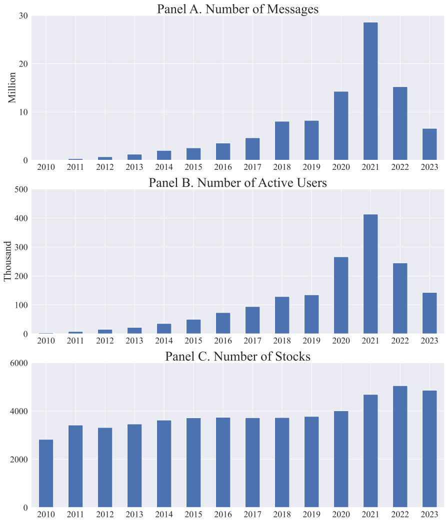
Back

Post-GME Short Squeeze

	Return$_{i,t+1\to t+5}$ (%)		Return$_{i,t+6\to t+10}$ (%)		Return$_{i,t+11\to t+15}$ (%)	
	(1)	(2)	(3)	(4)	(5)	(6)
Sentiment$_{i,t}^{TA}$	-0.060***	0.284***	-0.056***	0.252***	-0.016	0.144
	[-3.65]	[3.03]	[-3.65]	[2.64]	[-1.09]	[1.62]
Sentiment$_{i,t}^{FA}$	0.052***	0.132**	-0.002	0.024	-0.001	-0.061
	[3.88]	[2.15]	[-0.12]	[0.40]	[-0.08]	[-0.89]
Sentiment$_{i,t}^{OS}$	-0.091***	-0.019	-0.046***	0.238***	-0.036**	0.281***
	[-5.30]	[-0.23]	[-2.73]	[2.64]	[-2.28]	[3.37]
Sentiment$_{i,t}^{NS}$	-0.027**	-0.045	-0.001	0.138**	0.007	0.089
	[-2.37]	[-0.75]	[-0.12]	[2.04]	[0.64]	[1.39]
Sentiment$_{i,t}^{TA}$ × Post-GameStop Episode		-0.455***		-0.468***		-0.295**
		[-3.20]		[-3.39]		[-2.12]
Sentiment$_{i,t}^{FA}$ × Post-GameStop Episode		-0.176*		-0.173*		-0.043
		[-1.83]		[-1.84]		[-0.44]
Sentiment$_{i,t}^{OS}$ × Post-GameStop Episode		-0.219*		-0.455***		-0.491***
		[-1.67]		[-3.28]		[-3.85]
Sentiment$_{i,t}^{NS}$ × Post-GameStop Episode		-0.113		-0.345***		-0.220**
		[-1.25]		[-3.65]		[-2.30]
Attention$_{i,t}$	-0.164***	-0.026	-0.083***	0.046	-0.065***	0.119
	[-8.91]	[-0.13]	[-6.05]	[0.32]	[-5.34]	[1.00]
Stock Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Lagged Returns	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
N	2,972,286	530,098	2,970,213	529,747	2,968,174	529,368
R^2	0.105	0.082	0.108	0.091	0.112	0.087

Back

StockTwits Coverage



▸ Back

Classification by Technical Words TF-IDF Intensity