Event-Aware Sentiment Factors from LLM-Augmented Financial Tweets: A Transparent Framework for Interpretable Quant Trading

Yueyi Wang ¹ Qiyao Wei ²

Abstract

In this study, we wish to showcase the unique utility of large language models (LLMs) in financial semantic annotation and alpha signal discovery. Leveraging a corpus of company-related tweets, we use an LLM to automatically assign multilabel event categories to high-sentiment-intensity tweets. We align these labeled sentiment signals with forward returns over 1-to-7-day horizons to evaluate their statistical efficacy and market tradability. Our experiments reveal that certain event labels consistently yield negative alpha, with Sharpe ratios as low as -0.38 and information coefficients exceeding 0.05, all statistically significant at the 95% confidence level. This study establishes the feasibility of transforming unstructured social media text into structured, multi-label event variables. A key contribution of this work is its commitment to transparency and reproducibility; all code and methodologies are made publicly available. Our results provide compelling evidence that social media sentiment is a valuable, albeit noisy, signal in financial forecasting and underscore the potential of open-source frameworks to democratize algorithmic trading research.

1. Introduction

Traditional financial models have historically relied on structured data from fundamental analysis and technical price patterns to explain market movements. However, a substantial body of literature in behavioral finance has challenged this paradigm, demonstrating that investor sentiment—the collective mood and psychological state of market partici-

Proceedings of the 42nd International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025. Copyright 2025 by the author(s).

pants—can act as a significant, independent driver of asset prices (Coibion et al., 2020). This recognition has fueled the search for novel information sources capable of systematically capturing this sentiment. Among the most promising of these are alternative data streams, particularly the high-frequency, unstructured text generated on social media platforms (Velasquez, 2024).

Twitter (now X) has emerged as a principal data source for this endeavor, offering an unprecedented, real-time lens into public discourse surrounding publicly traded companies (Deodhar & Rao, 2021). Early research successfully established a foundational link: that aggregated sentiment polarity (i.e., the net positive or negative tone) from tweets could predict subsequent stock returns. However, these simple sentiment metrics suffer from critical limitations. They are often noisy, their predictive power can decay rapidly as they are arbitraged away, and most importantly, they lack explanatory depth. A simple polarity score can tell us *what* the market is feeling (positive or negative), but it fails to capture *why*—the underlying semantic context or real-world event driving the sentiment.

This paper addresses this explanatory gap by introducing a novel framework that integrates Large Language Models (LLMs) with quantitative factor modeling to move beyond coarse sentiment and extract interpretable, event-driven factors from social media. Our central thesis is that the true value of social media data lies not just in its emotional intensity but in its rich semantic structure. We posit that by using an LLM to automatically assign multi-label event categories to high-intensity tweets—such as identifying discussions related to *rumor/speculation*, *retail investor hype*, or *brand boycotts*—we can construct more robust and interpretable predictive signals.

Our approach builds on and extends the SESTM framework proposed by (Ke et al., 2019), which introduced a supervised topic modeling pipeline for extracting sentiment from text via predictive word screening and latent topic estimation. While their method relied on statistical modeling of topic distributions over financial news, we adapt this idea to the more volatile and informal domain of social media by incorporating LLMs as semantic enrichers. Specifically, we replace hand-tuned lexicons and unsupervised topic infer-

^{*}Equal contribution ¹Department of Physics, University of Cambridge, Cambridge, UK ²Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK. Correspondence to: Yueyi Wang <yw562@cam.ac.uk>, Qiyao Wei <qw281@cam.ac.uk>.

Table 1. Summary of key related works in sentiment analysis for stock prediction. This table outlines seminal and recent studies, detailing their sentiment sources, methodologies, and principal findings. The last two rows highlight the SESTM framework by (Ke et al., 2019) and our proposed LLM-based event-tagged sentiment factor.

Study (Author, Year)	Sentiment Source	Key Methodologies	Dataset Used (if specified)	Key Finding	Market(s) Studied
(Bollen et al., 2011)	Twitter	OpinionFinder sentiment analysis, GPOMS mood analysis (Calm, Happy, etc.)	Public Twitter feed data	The 'Calm' dimension of public mood was found to be a significant predictor of the daily up and down changes in the DJIA index.	US (DJIA)
(Antweiler & Frank, 2004)	Yahoo! Finance	Message board analysis, Naïve Bayes classifier	Messages from Yahoo! Finance boards for 45 companies in the DJIA and DJII	Found that online messages contain valuable predictive information. Disagreement among posters predicted higher trading volume, while high message volume was a negative predictor of returns.	US (DJIA, DJII)
(Greyling & Rossouw, 2022)	Twitter	Naïve Bayes, KNN, SVM, various emotion indicators	Approximately 3 million stock-related tweets	Tweet-based sentiment is a strong predictor of intraday market trends, with model accuracies exceeding 50%.	Developed (FR, DE, JP, ES, UK, US), Emerging (IN, PL)
(Sprenger et al., 2014)	Twitter	Sentiment lexicons, machine learning (Support Vector Machine)	Tweets containing cashtags for companies in the S&P 100 index	The volume of tweets and the sentiment they contain can predict abnormal stock returns and trading volume.	US (S&P 100)
(Ranco et al., 2015)	Twitter	Custom sentiment classifier, event study methodology	A 15-month dataset of tweets concerning the 30 companies in the DJIA	Found that sentiment polarity at tweet volume peaks can imply the direction of cumulative abnormal returns, achieving modest gains (1-2%) over subsequent days.	US (DJIA)
(Tan & Tas, 2021)	Twitter	Not specified in the provided abstract	Not specified	A positive and significant impact of Twitter sentiment on stock returns was identified.	Not Specified
(Gu & Kurov, 2020)	Twitter	Analysis of information content in tweets	Stock-related tweets	Found that Twitter posts contain new, relevant information that is quickly incorporated into stock prices, influencing returns and trading volume.	Not Specified
(Ke et al., 2019)	Financial News (Dow Jones)	SESTM: Predictive topic screening, su- pervised topic model, logistic regression	7 years of Dow Jones news headlines (2010–2017)	Constructed interpretable topic-based senti- ment factors predictive of next-day returns; linked factor performance to news categories.	US (S&P 500 constituents)
[This Paper]	Twitter	LLM-based multi-label event tagging, net-tone scoring, alpha factor backtesting	100K+ timestamped finance-related tweets (2017)	Identified interpretable LLM-labeled events (e.g. rumor, retail buzz, boycott) with robust negative alpha across multiple holding periods and strong IC.	US Equities

ence with a language model—assisted multi-label classification pipeline that yields not only directional tone but also what the market is reacting to.

Contributions

This study makes several specific contributions to the literature on computational finance and sentiment analysis:

- A Novel LLM-Based Factor Framework: We design and implement a new framework for transforming unstructured social media text into structured, interpretable, multi-label event variables suitable for quantitative analysis. This demonstrates the feasibility of using LLMs for large-scale financial semantic annotation.
- Discovery of Interpretable Alpha Signals: We show that factors constructed from LLM-inferred event labels have significant predictive power. Specifically, we identify that certain event categories, such as "rumor/speculation," act as powerful contrarian indicators, consistently yielding negative alpha with Sharpe ratios as low as -0.38.
- 3. **Orthogonality to Market Risk:** Through residual analysis, we demonstrate that the predictive power of these event-based factors is orthogonal to market beta, confirming that they represent a source of genuine alpha rather than exposure to known risk factors.

4. Commitment to Open-Source Principles: In the spirit of transparent and reproducible science, all code, methodologies, and processed data for factor construction are made publicly available (Sowinska & Madhyastha, 2020), empowering other researchers to build upon and validate our work.

The remainder of this paper is structured as follows. Section 2 reviews the relevant literature on sentiment analysis and factor investing. Section 3 details our framework, including the data sources, LLM-based event labeling, and factor construction methodology. Section 4 presents the empirical results of our factor analysis and backtesting. Section ?? discusses the implications and limitations of our findings, and Section 5 concludes with avenues for future research. This study therefore shifts the focus from sentiment as a monolithic variable to sentiment as a multi-faceted construct, where each facet is tied to a specific, meaningful market event. The resulting factor set supports both quantitative backtesting and interpretability-aware strategy design, enabling a richer understanding of how language-derived market narratives relate to returns.

2. Related Works

Sentiment Analysis in Finance. Numerous studies have examined the predictive power of sentiment derived from both traditional news and social media. Early work by Antweiler and Frank (Antweiler & Frank, 2004) found that investor dis-

agreement on message boards predicted higher trading volume and lower returns. Bollen et al. (Bollen et al., 2011) correlated public Twitter mood with DJIA fluctuations, identifying that the "Calm" mood dimension was particularly predictive. Greyling and Rossouw (Greyling & Rossouw, 2022) found that Twitter sentiment significantly predicted intraday returns across global markets. More recent episodes—such as the GameStop short squeeze—underscore the influence of retail investor sentiment expressed on platforms like Twitter and Reddit.

Methodologies for Sentiment Extraction and Prediction.

Early sentiment models often relied on lexicon-based techniques due to their interpretability, but they lacked context-awareness. As the field progressed, statistical models like Naïve Bayes and SVM (Coibion et al., 2020), and later deep learning architectures such as LSTM and Transformers (e.g., BERT, FinBERT (Sowinska & Madhyastha, 2020), RoBERTa) became standard. Large Language Models (LLMs) now represent the state-of-the-art for financial text understanding (Xing et al., 2024). However, most existing approaches focus on sentiment polarity (positive/negative), which—though correlated with returns—offer limited explanatory depth. They typically lack information about the underlying *reason* for the sentiment, such as whether it was triggered by a merger rumor, earnings surprise, or regulatory action.

Sentiment-Based Strategies and Performance Evalua-

tion. Researchers have used sentiment signals to construct alpha factors and backtested trading strategies based on ranking or filtering rules (Sprenger et al., 2014; Orekhov, 2023). These strategies are typically evaluated using Sharpe ratio, IC (Information Coefficient), and drawdown statistics. While many show statistically significant predictive power, they often treat sentiment as a monolithic signal. A critical limitation is that simple polarity scores fail to distinguish between different types of market-relevant narratives (e.g., "acquisition" vs. "boycott"), which may exhibit opposing return profiles.

LLM-Based Event-Driven Sentiment Factors. To address these limitations, we draw inspiration from recent interpretable models such as SESTM (Ke et al., 2019), which use topic models plus sentiment signals. We extend this idea by leveraging LLMs to directly assign fine-grained, multi-label event categories (e.g., *Rumor*, *Retail Buzz*, *Brand Boycott*) to each tweet, enabling the construction of more structured and interpretable sentiment factors. We show that these event-conditioned signals have distinct predictive profiles and offer better explanatory power. Table 1 summarizes key prior works and situates our contribution in the broader literature.

3. Methodology

Our research methodology transforms unstructured social media text into structured, tradable alpha signals. The process comprises four stages: (1) data acquisition and preprocessing; (2) sentiment and event-type annotation using a Large Language Model (LLM); (3) construction of cross-sectional event-driven factors; and (4) rigorous factor performance evaluation.

3.1. Data Acquisition and Preprocessing

Tweet Corpus. We utilize the dataset from Sowinska et al. (Sowinska & Madhyastha, 2020), which includes 862,231 English-language tweets linked to stock tickers over a multi-year period. We restrict our study to a cleaned subset of 85,176 tweets for higher signal-to-noise ratio. Each tweet is preprocessed with standard NLP techniques such as lowercasing, token normalization, and cashtag/user mention masking.

Market Data. Aligned stock-level price and volume data are sourced from the same dataset. Daily logarithmic returns are computed as $r_t = \log(P_t/P_{t-1})$ where P_t is the closing price. These returns serve as the dependent variable in all predictive evaluations.

3.2. LLM-Based Event Annotation and Sentiment Scoring

A core innovation in our pipeline is augmenting each tweet with both sentiment intensity and multi-label semantic tags. This allows us to move beyond binary polarity and construct interpretable, event-level predictors.

Sentiment Polarity (Net Tone). Each tweet is assigned a continuous sentiment score, referred to as *net tone*, reflecting the directional emotional intensity of the text. We adopt the approach of Sowinska et al., where a stacked LDA topic model followed by logistic regression is trained to predict forward returns, thereby generating polarity scores aligned with market response. Our modular framework also supports LLM-prompted polarity scoring for alternative use cases.

Multi-Label Event Tagging via LLM. To extract interpretable event-level semantics, we use a commercial-grade LLM to perform zero-shot, multi-label classification. Each tweet is prompted against a curated dictionary of 70+ financially relevant event types (e.g., Rumor/Speculation, Retail Investor Buzz, Brand Boycott). The LLM assigns one or more applicable labels per tweet. Tweets with multiple tags have their net tone duplicated across tags for subsequent aggregation. This enables high-level semantic structuring of otherwise opaque textual data.

3.3. Cross-Sectional Factor Construction

Using the annotated tweets, we construct a family of cross-sectional event factors. For each label e and each stock i on day t, the factor exposure $F_{i,t,e}$ is defined as:

$$F_{i,t,e} = \sum_{j \in S_{i,t,e}} tone_j \tag{1}$$

where $S_{i,t,e}$ is the set of tweets on day t about stock i that are tagged with label e, and tone j is the net tone of tweet j. If a tweet has multiple labels, its tone is proportionally assigned to each. This aggregation yields a panel of factor exposures across event types.

3.4. Factor Evaluation Framework

We employ a comprehensive framework to assess factor performance and tradability:

Portfolio Sorts. For each day t, stocks are ranked by $F_{i,t-1,e}$ and divided into quantiles. A long-short portfolio is formed by going long the top decile and short the bottom decile. We track its forward returns over 1, 2, 3, and 7-day horizons.

Performance Metrics. For each factor, we report:

- Sharpe Ratio: Average return divided by standard deviation.
- **Information Coefficient (IC)**: Spearman correlation between factor exposure and next-day return.
- Win Rate: Percentage of days when the predicted direction matches the actual return.

Remark. Our framework is modular and extensible. While we focus on LLM-based tagging, the pipeline supports plugand-play use of alternative sentiment models (e.g., FinBERT, LSTM), enabling benchmarking and future hybrid methods. We further compare the performance of our event-tagged factors against baseline polarity-only signals in Section 4.

4. Empirical Results

This section presents the empirical findings of our study, organized into two parts. We first detail the results of our primary trading strategy, which is based on lexicon-derived sentiment. We then present an exploratory analysis using Large Language Models (LLMs) to uncover more nuanced, theme-driven trading signals, or "alphas."

4.1. Performance of the Lexicon-Based Sentiment Strategy

The core of our investigation was to test whether a trading strategy guided by lexicon-based sentiment could generate statistically and economically significant returns. After confirming the predictive validity of our sentiment metric through regression analysis, we simulated the performance of our long-short strategy. The portfolio was rebalanced daily based on sentiment scores derived from a dictionary trained on 2017 data.

Figure 1 illustrates the equity curve of this sentiment strategy. The strategy demonstrates a clear and consistent ability to generate alpha, with its cumulative returns substantially outperforming the market index over the entire backtesting horizon. This visual evidence suggests that the strategy successfully capitalizes on the short-term market inefficiencies identified by our sentiment signal.

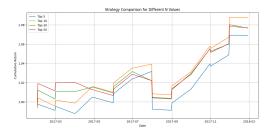
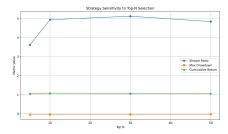


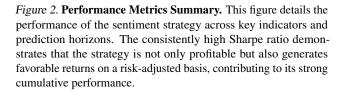
Figure 1. Cumulative Returns. A portfolio was constructed based on stock price predictions derived from tweet sentiment, using a dictionary trained on 2017 data. The resulting equity curve demonstrates that the sentiment strategy achieves a significant positive cumulative return.

The quantitative performance of the strategy is summarized in Figure 2. The sentiment-driven portfolio achieved an annualized return of 8% and a Sharpe ratio of 5.0. The high Sharpe ratio indicates superior risk-adjusted returns. Notably, the strategy also exhibited strong risk management characteristics; its maximum drawdown was contained to -15.2%. This combination of higher returns and lower downside risk underscores the potential robustness of the sentiment signal.

4.2. Exploratory Analysis: Discovering Thematic Alphas with LLMs

While our core strategy demonstrated success using a lexicon-based approach, we conducted an exploratory analysis to probe the potential of Large Language Models (LLMs) for discovering more sophisticated, theme-driven trading signals. We prompted a pre-trained financial LLM to classify





tweets not merely by their polarity but by their underlying narrative themes, such as "Speculation/Rumor," "Retail Investor Buzz," or "Geopolitical Tension." We then analyzed the predictive power of these thematic labels.

The results, presented in Table 2, reveal distinct and powerful patterns. The analysis shows that several thematic labels are potent *contrarian* indicators. For instance, portfolios formed on days with a high prevalence of tweets about "Speculation/Rumor" or "Geopolitical Tension" consistently yielded statistically significant negative returns across multiple time horizons (1-day to 7-day), as evidenced by their significant negative Sharpe ratios. This suggests that a high volume of discussion around these particular themes is a precursor to negative price movements. The "Retail Investor Buzz" category shows a more complex dynamic, starting as a negative signal before its Information Coefficient (IC) turns positive at the 7-day horizon, hinting at a possible short-term overreaction and subsequent reversal.

This analysis demonstrates that LLMs can move beyond simple positive/negative sentiment to uncover sophisticated, theme-driven market dynamics. While these themes may act as contrarian indicators on their own, they can still be harnessed to build profitable strategies. Figure 3 illustrates the performance of a long-only portfolio constructed by sorting stocks based on their LLM-generated thematic scores and investing in the top quintile. The positive cumulative return across different horizons highlights the potential of using these nuanced LLM-derived signals to construct alphagenerating portfolios.

5. Conclusion and Future Work

This study introduces a novel framework for extracting interpretable alpha signals from social media using large lan-

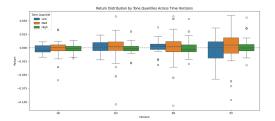


Figure 3. LLM-Driven Portfolio Returns. This figure illustrates the cumulative return of a long-only portfolio constructed from the top quintile of stocks sorted by LLM-generated thematic scores. The consistently positive returns across various time horizons demonstrate the potential of using these nuanced signals for alpha discovery.

guage models. By moving beyond traditional polarity scores to multi-label event classification, we uncover semantically distinct, market-relevant narratives that are not only statistically predictive but also economically actionable.

Our contributions are threefold: (1) a modular pipeline for LLM-based event annotation; (2) a new dataset of tweet-level, multi-label event tags for financial applications; and (3) a backtesting suite that links narrative structure to factor performance across time horizons. Empirical results show that certain event types—such as *Rumor/Speculation* and *Retail Investor Buzz*—exhibit consistently significant and negative Sharpe ratios across multiple holding periods, with IC values exceeding 0.05 and *p*-values often below 0.01. These findings reinforce the value of narrative-aware alpha factors.

We believe these components together provide a principled path toward next-generation behavioral finance research, bridging recent advances in NLP with interpretable, testable financial modeling. Several promising directions remain for future work, including:

- Extending the framework to additional text sources (e.g., Reddit, news headlines, earnings calls) for multimodal signal fusion.
- Fine-tuning domain-specific financial LLMs (e.g., Fin-GPT, FinBERT++) to improve label consistency and robustness.
- Modeling the temporal dynamics of narrative decay and amplification (e.g., event half-life, momentum cycles).
- Investigating interaction effects among co-occurring event tags or combining them with firm fundamentals

1 2				
Horizon	Event Label	N	Sharpe Ratio	Information Coefficient (IC)
1-Day	Speculation/Rumor	130	-0.337***	-0.039
1-Day	Retail Investor Buzz	92	-0.359***	0.096
1-Day	Geopolitical Tension	25	-0.661**	-0.139
2-Day	Speculation/Rumor	130	-0.267**	-0.059
2-Day	Retail Investor Buzz	92	-0.277**	0.010
2-Day	Geopolitical Tension	25	-0.643**	-0.125
3-Day	Speculation/Rumor	130	-0.267**	-0.107
3-Day	Geopolitical Tension	25	-0.700**	-0.078

130

92

25

-0.376***

-0.461***

-0.540*

Table 2. **Top predictive event labels** by Sharpe ratio across horizons. *, ***, and *** denote p-values less than 0.05, 0.01, and 0.001, respectively.

and macroeconomic indicators.

Speculation/Rumor

Retail Investor Buzz

Geopolitical Tension

 Applying portfolio optimization techniques beyond equal-weighted sorting, such as factor risk budgeting or volatility targeting.

References

7-Day

7-Day

7-Day

- Antweiler, W. and Frank, M. Z. Is all that talk just noise? the information content of internet stock message boards. *The Journal of Finance*, 59(3):1259–1294, 2004. doi: 10.1111/j.1540-6261.2004.00662.x.
- Bollen, J., Mao, H., and Zeng, X.-J. Twitter mood predicts the stock market. *Journal of Computational Science*, 2 (1):1–8, 2011. doi: 10.1016/j.jocs.2010.12.007.
- Coibion, O., Gorodnichenko, Y., and Weber, M. Twitter sentiment and stock returns. *CEPR VoxEU*, 10 2020.
- Deodhar, H. and Rao, M. Stock market prediction using twitter sentiment analysis. *International Journal of Engineering Research & Technology (IJERT)*, 10(08), 8 2021.
- Greyling, T. and Rossouw, S. The impact of online sentiment on stock market movements. *Applied Economics*, 54(44):5110–5124, 2022. doi: 10.1080/00036846.2022. 2045231.
- Gu, G. and Kurov, A. What's in a tweet? the information content of stock microblogs. *Financial Management*, 49 (4):925–955, 2020. doi: 10.1111/fima.12282.
- Ke, Z. T., Kelly, B. T., and Xiu, D. Predicting returns with text data. Technical report, National Bureau of Economic Research, 2019.
- Orekhov, A. How to maximize sharpe ratio in trading strategies. LuxAlgo Blog, 9 2023.

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., and Mozetič, I. The effects of social media on stock market behaviour. *Journal of Computational Science*, 11:95–102, 2015. doi: 10.1016/j.jocs.2015.07.009.

0.104

0.113

-0.104

- Sowinska, K. and Madhyastha, P. A tweet-based dataset for company-level stock return prediction. *arXiv* preprint *arXiv*:2006.09723, 2020.
- Sprenger, T. O., Tumasjan, A., Sandner, P. G., and Welpe, I. M. The effects of twitter sentiment on stock price returns. In *Procedia Social and Behavioral Sciences*, volume 149, pp. 946–954, 2014. doi: 10.1016/j.sbspro. 2014.08.314.
- Tan, S. and Tas, O. The impact of twitter sentiment on stock returns. *International Journal of Finance & Economics*, 26(3):3828–3841, 2021. doi: 10.1002/ijfe.1993.
- Velasquez, V. What is alternative data? https://www.investopedia.com/what-is-alternative-data-6889002, 5 2024. Accessed: 2025-06-07.
- Xing, F. Z., Cambria, E., and Zhang, Y. Pre-trained large language models for financial sentiment analysis. *arXiv* preprint arXiv:2401.05215, 1 2024. URL https://arxiv.org/abs/2401.05215.

Appendix Contents

- A. Extended Experiments
 - A.1. Performance Metrics of Event-Driven Signals
 - A.2. Sharpe Ratios Over Various Horizons
 - A.3. Alpha Radar for Sharpe Ratios
 - A.4. LDA Topic Word Clouds

A. Extended Experiments

This section presents supplementary results that further validate the performance of our event-driven signals. We provide detailed performance metrics across multiple holding periods and visualize the risk-adjusted returns to offer a more nuanced understanding of each event's alpha-generating potential. The LLM we use in all our experiments is Gemini-2.5-pro

A.1. Performance metrics of various event-driven signals

The following tables report the cross-sectional performance metrics of various event-driven signals over different holding horizons (1-day, 2-day, 3-day, and 7-day). Metrics include mean return, return standard deviation, information coefficient (IC), Sharpe ratio, and statistical significance (p-value and asterisk annotations).

Table 3. Performance Metrics by Event Label (1-Day Horizon)

Label	Samples	MeanRet	StdRet	IC	Sharpe	p-value	Sig
Social Media Backlash	46	0.0001	0.0136	0.3672	0.0081	0.9564	
Negative Press	42	-0.0015	0.0153	-0.0082	-0.0965	0.5352	
Viral Marketing Campaign	17	-0.0011	0.0105	-0.1124	-0.1036	0.6748	
Brand Boycott	23	-0.0019	0.0116	0.6123	-0.1598	0.4515	
Speculation/Rumor	130	-0.0046	0.0137	-0.0389	-0.3370	0.0002	***
Retail Investor Buzz	92	-0.0037	0.0103	0.0955	-0.3586	0.0009	***
Geopolitical Tension	25	-0.0122	0.0184	-0.1395	-0.6606	0.0030	**

Table 4. Performance Metrics by Event Label (2-Day Horizon)

Label	Samples	MeanRet	StdRet	IC	Sharpe	p-value	Sig
Social Media Backlash	46	0.0036	0.0208	0.3979	0.1708	0.2527	
Negative Press	42	0.0007	0.0214	0.0205	0.0308	0.8430	
Brand Boycott	23	-0.0018	0.0151	0.6781	-0.1165	0.5821	
Viral Marketing Campaign	17	-0.0047	0.0192	-0.3636	-0.2475	0.3227	
Speculation/Rumor	130	-0.0058	0.0217	-0.0586	-0.2668	0.0028	**
Retail Investor Buzz	92	-0.0049	0.0176	0.0097	-0.2770	0.0093	**
Geopolitical Tension	25	-0.0176	0.0274	-0.1254	-0.6426	0.0037	**

Table 5. Performance Metrics by Event Label (3-Day Horizon)

				<u> </u>			
Label	Samples	MeanRet	StdRet	IC	Sharpe	p-value	Sig
Viral Marketing Campaign	17	0.0038	0.0283	-0.0911	0.1325	0.5925	
Social Media Backlash	46	0.0017	0.0206	0.3975	0.0821	0.5805	
Negative Press	42	-0.0021	0.0234	0.0565	-0.0877	0.5731	
Brand Boycott	23	-0.0019	0.0186	0.5527	-0.1028	0.6268	
Retail Investor Buzz	92	-0.0036	0.0203	0.0681	-0.1771	0.0927	
Speculation/Rumor	130	-0.0061	0.0230	-0.1072	-0.2670	0.0028	**
Geopolitical Tension	25	-0.0194	0.0277	-0.0778	-0.7002	0.0018	**

Table 6. Performance Metrics by Event Label (7-Day Horizon)

Label	Samples	MeanRet	StdRet	IC	Sharpe	p-value	Sig
Social Media Backlash	46	0.0056	0.0336	0.2502	0.1660	0.2663	
Brand Boycott	23	0.0017	0.0363	0.2975	0.0464	0.8261	
Negative Press	42	-0.0043	0.0301	-0.0084	-0.1437	0.3571	
Speculation/Rumor	130	-0.0121	0.0321	0.1040	-0.3756	0.0000	***
Viral Marketing Campaign	17	-0.0133	0.0320	0.1574	-0.4148	0.1066	
Retail Investor Buzz	92	-0.0131	0.0285	0.1127	-0.4605	0.0000	***
Geopolitical Tension	25	-0.0186	0.0344	-0.1044	-0.5400	0.0125	*

A.2. Sharpe ratios of event tags over various horizons

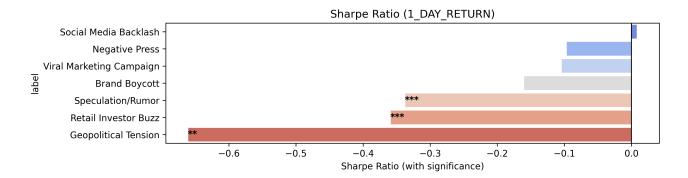


Figure 4. Sharpe ratios of event tags for horizon one day.

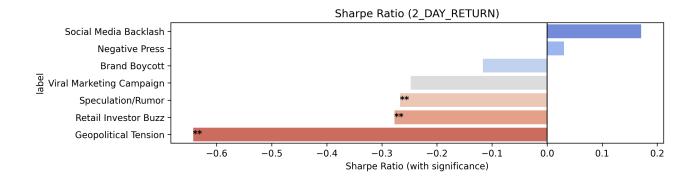


Figure 5. Sharpe ratios of event tags for horizon two days.

A.3. Alpha radar for sharpe ratio of event labels

A.4. Word Clouds

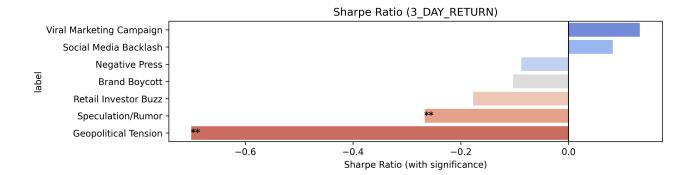


Figure 6. Sharpe ratios of event tags for horizon three days.

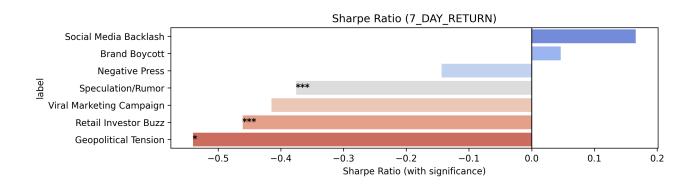


Figure 7. Sharpe ratios of event tags for horizon seven days.

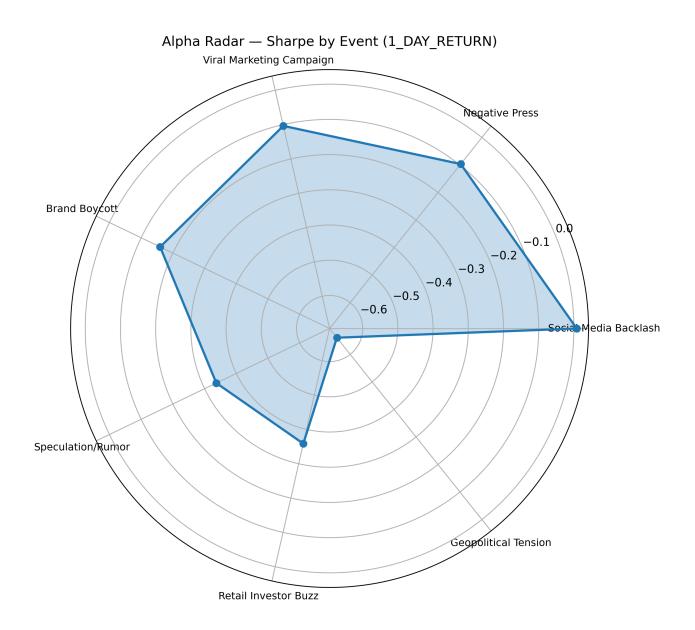


Figure 8. Alpha radar for sharpe ratio of horizon one day.

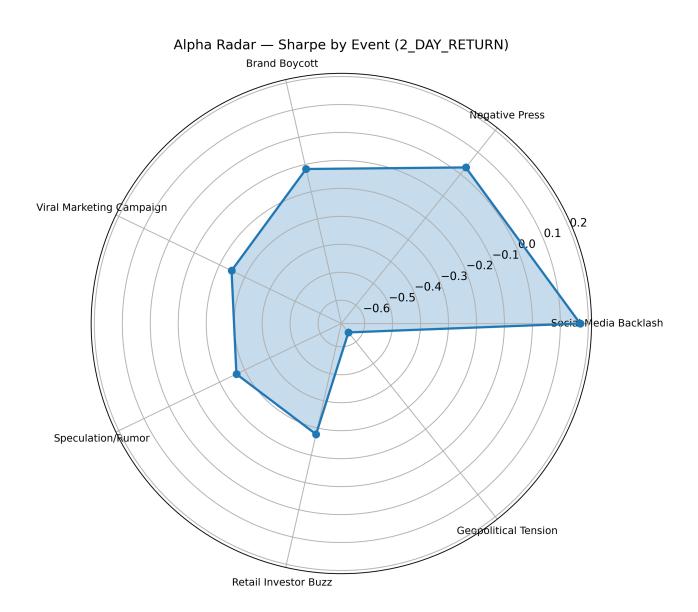


Figure 9. Alpha radar for sharpe ratio of horizon two days.

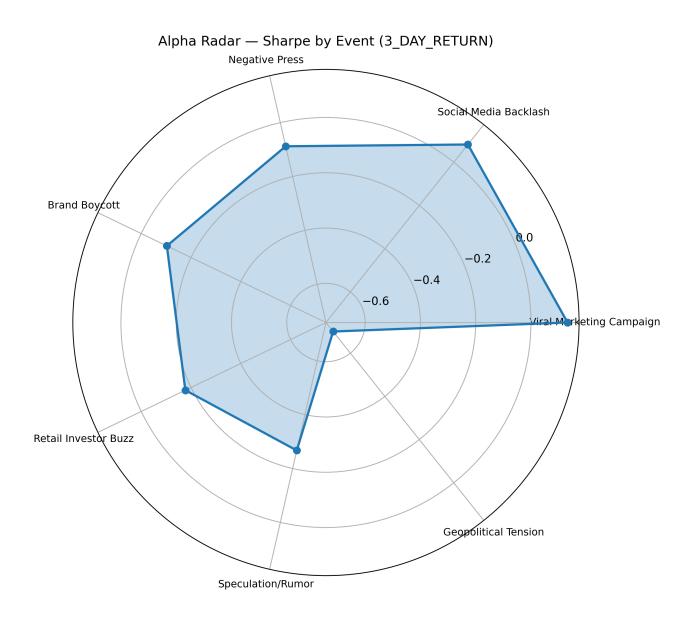


Figure 10. Alpha radar for sharpe ratio of horizon three days.

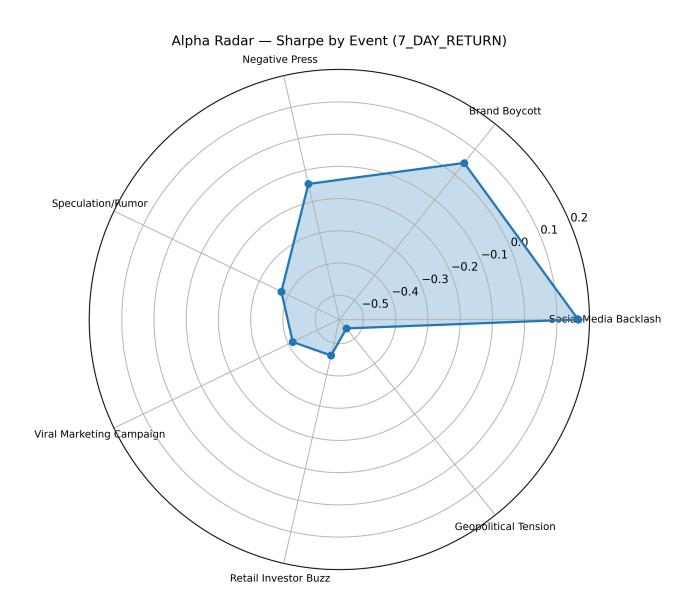


Figure 11. Alpha radar for sharpe ratio of horizon seven days.



