Proceedings of ICMRED 2026 Symposium: The Future of Work: Strategy, Workforce Transformation, and Organizational Renewal
DOI: 10.54254/2754-1169/2026.LD33079

Intraday Volatility Prediction of Individual Stocks in the Artificial Intelligence Sector Based on Sentiment Mining from the Guba of Eastmoney

Guangyu Liu

School of Finance, Zhongnan University of Economics and Law, Wuhan, China
2171608171@qq.com

Abstract. The year 2024 is regarded as a major turning point for artificial intelligence from technological breakthrough to industrialization. Therefore, the performance of relevant enterprises in the capital market is volatile, and there are significant differences in the market's views on the future development of the industry. This paper takes the Guba of Eastmoney as the research object, uses the FinBERT deep - learning model to extract unstructured online public opinion data of enterprises such as iFlytek, 360 Security Technology, Hundsun Technologies, and Inspur Information, and establishes a quantitative Sentiment divergence of investors' heterogeneous beliefs. This paper mainly studies the lag prediction ability of this index for the daily price changes of the target stocks and its dynamic changes. The empirical analysis results show that within the sample range, the proposed model has a good goodness - of - fit ($R^2 > 0.52$), and the trading volume factor has a strong explanatory power. The sentiment divergence index among the user group of 360 Security Technology has a strong forward - looking guiding role, indicating how the investor structure affects the asset pricing behavior. This paper finds that using natural language processing technology for risk monitoring is of certain significance in solving the systematic problems in the complex financial environment, providing technical support and theoretical basis for investors to identify systematic risks.

Keywords: Sentiment divergence, Diurnal variation range, Deep neural network model, Integration of heterogeneous views, Semantic analysis method for large - scale data.

1. Introduction

According to the general view of the global financial and technology circles, 2024 is a milestone year in the development of artificial intelligence. Its main feature has evolved from the technological innovation period to the stage of deepening industrial application, which has caused significant fluctuations in the capital market. According to the core view of the efficient market hypothesis, asset prices should quickly integrate and reflect all publicly available information factors [1]. In actual operations, due to the obvious technical complexity of AI technology (such as token function optimization and the autonomy of computing power chips), the traditional financial reporting system can hardly meet the requirements of instantaneity and accuracy of information transmission. In this

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2754-1169/2026.LD33079

case, the changes in public sentiment on social media have become an important indicator for evaluating the market pricing mechanism [2]. According to the heterogeneous belief theory in behavioral finance, differences in views rather than unanimity are the main reasons for asset flow and price fluctuations. In the highly specialized context of artificial intelligence, the sentiment in the Guba is very fragmented and extremely polarized, which is also an important reason for the significant fluctuations in stock prices the next day. The main research questions in this paper are as follows: Can the degree of sentiment divergence on social media be used as an effective quantitative indicator to predict the daily price fluctuations of AI - related stocks? Will this prediction ability be affected by the socio - economic attributes and distribution characteristics of the investor group?

Traditional financial text analysis mostly uses the static dictionary matching method (such as the dictionary of Loughran and McDonald) to complete tasks. When dealing with highly professional topics such as artificial intelligence, due to the lack of dynamic adaptability, the results have large errors and cannot well reflect the semantic contradictions. Therefore, this paper uses a deep - learning framework to adaptively train a large amount of financial data to obtain the FinBERT model. After integrating a large number of industry semantic resources, the accuracy of perceiving subtle emotional changes is greatly enhanced, providing strong technical guarantee and empirical support for creating an index that better fits the market sentiment divergence [3]. This paper selects 2024 as the research node and focuses on the research status of the impact of heterogeneous beliefs on the prediction effect of asset price fluctuations during the important period of the integration of technological innovation and market application. According to the empirical data, in the highly complex and difficult - to - observe field of technological innovation, opposing opinions have a better risk prediction effect than emotional polarization and have obvious forward - looking significance for asset pricing. This conclusion enriches the empirical support of the heterogeneous belief theory and provides key practical support for the establishment of a risk control system in extreme market environments and the implementation of quantitative supervision using large - model technology.

2. Literature review and theoretical mechanism

2.1. Research status of the relationship between investor sentiment and asset price fluctuations

The impact of investor sentiment on the volatility of financial market asset prices is an important research topic in the field of behavioral finance. Through research, it is found that social media can shape people's psychological expectations and increase the irrational fluctuations in the market through the herd effect. This kind of fluctuation is more obvious in technology - based growth enterprises [4]. Starting from the characteristics of sentiment spread and interpersonal contradictions, the direction and degree of change in the next - day return can be effectively predicted [5]. According to the existing literature, the sentiment factors caused by social media have also become a pricing driving factor in addition to traditional financial indicators [6].

2.2. Development of financial text mining technology from rule - oriented to semantic understanding

Financial text analysis has broken through the limitations of traditional keyword static matching and has shifted its focus to creating a technical system for dynamic semantic analysis. The early traditional methods relying on dictionary matching have significant deficiencies in complex environments due to the limitations of fixed patterns. Huang believe that using a pre - trained

DOI: 10.54254/2754-1169/2026.LD33079

financial model (FinBERT) for optimization can achieve better results in professional term recognition and sentiment tendency judgment [3]. Zhang found that when using a deep neural network structure to monitor social public opinion, useless information can be well removed, and important investment decision - making factors can be accurately identified [7]. These research methods based on mathematical models for modeling the psychological changes of investors in a complex market environment are reliable [8].

2.3. Heterogeneous belief theory and its application in the AI sector

The key to the heterogeneous belief theory is that investors have very different trading behaviors due to their different ways of understanding the same information source. Lopez - Lira and Tang found through mining data related to the transformation of large - language models that this method can better predict the changes in short - term financial market risk parameters [9]. Guijarro believe that variance can better reflect the large fluctuations in the market in the short term than the mean [10]. Due to the obvious "black - box" phenomenon in the artificial intelligence (AI) industry relying on a complex system architecture and the continuous exploration of business models, the information transparency of this industry is low, which increases the degree of investors' cognitive bias [11]. In this context, the heterogeneous belief theory can be regarded as an important analytical method to explain the abnormal fluctuations in the AI sector. Currently, most of the academic research on sentiment focuses on the mean directional influence effect, and there is no detailed analysis of the mechanism of action of sentiment dispersion. Especially in the field of artificial intelligence (AI), the research on the generation mechanism of heterogeneous beliefs and their market prediction ability is still in its infancy. Therefore, this paper attempts to expand the existing theoretical framework and consider the problem from a new perspective. That is, in the context of the development of the AI industry, due to information asymmetry and cognitive diversity, there are obvious opposing views in the discussions in the Guba, which in turn leads to complex investment decision - making games and large fluctuations in stock prices. Based on the noise trading theory, this paper also proposes the hypothesis of the regulatory effect of the investor structure. That is, compared with the traditional capital market environment dominated by retail investors, the impact of sentiment divergence on asset price fluctuations is more direct and obvious. In a market dominated by institutional investors, due to the existence of rational expectations, the same impact will be smaller. This hypothesis not only lays the foundation for the subsequent empirical research on typical cases (such as 360 Security Technology) but also promotes the development of the sentiment - driven financial asset pricing model towards a more refined direction.

3. Research design

3.1. Sample selection and data sources

This paper selects four typical enterprises in the artificial intelligence sector in 2024 as research samples, namely iFlytek, 360 Security Technology, Hundsun Technologies, and Inspur Information. These samples are selected to analyze their competitive situation and development characteristics. This application paradigm is a typical closed - loop model of the in - depth integration of artificial intelligence technology and the real industry, reflecting the main characteristics of its approach to promoting industrialization [12]. The public opinion text data are uniformly collected from the annual posting records of the above four target stocks in the Guba of Eastmoney. The historical

DOI: 10.54254/2754-1169/2026.LD33079

market and trading data of the above stocks are obtained in real - time from the Tushare financial big data platform.

3.2. Data cleaning and sentiment measurement

There are many noise elements in the unstructured text generated by social media. In order to improve the quality of the data used for deep - learning model training, a comprehensive cleaning of the data is necessary. This cleaning process has three main parts. String similarity analysis technology is used to accurately find and effectively remove the data records of the target entities that match exactly. Using the loop detection mechanism and the matching function of regular expressions, the content units containing external links can be accurately selected and removed. The information with obvious advertising and marketing attributes is deleted, and the information fragments with a small number of characters and no practical significance are removed. The text after system denoising and streamlining is screened, leaving only the minimalist text fragments with real meaning, that is, the expressions composed of emojis or single emotional words. After strict screening, the number of data samples obtained is more than 112,000. Sentiment annotation mainly uses a model specially optimized for financial domain corpora. According to the FinBERT framework, a sentiment analysis model is developed to obtain the scores of text data [-1,+1] and the evaluation scores of sentiment tendency [13].

3.3. Variable construction and model setting

The sentiment divergence degree refers to the quantitative data formed by the differences in investors' views on a certain stock within a certain period. The value can be obtained and evaluated by statistically analyzing the heterogeneity of each investor's views on a certain stock within a certain period. The variance of the sentiment scores of all valid posts during the trading day is used as the independent variable, and it is converted into the interval fluctuation range of the logarithmic return rate for measurement. Volatility is the most important variable for quantifying the daily price fluctuations of financial assets and is also widely used in academia. This paper further expands its theoretical value and practical significance by establishing a volatility model framework. The natural logarithm of the daily trading volume (ln_vol) is used as the research variable to eliminate the influence of the liquidity effect on the analysis of asset price volatility. This paper mainly studies the forward - looking influence of social media public opinion information on the dynamics of the financial market. A scientific and rigorous statistical model is used to eliminate the influence of negative sentiment caused by major market fluctuations. In order to overcome the endogeneity problem (such as reverse causality), this paper uses a multi - layer robustness test method and conducts detailed control processing on all public - opinion - related variables. The above - mentioned variables (sentiment divergence degree, sentiment mean) are processed in the form of a first - order lag sequence. If the data meet all the conditions of the classical linear regression model, the ordinary least squares method (OLS) is the main estimation method in econometric research. This paper mainly studies the unbiased marginal effect of exogenous explanatory variables on continuous explained variables and ensures the achievement of the core research purpose. The linear marginal effect model used in this study has strong applicability from a statistical perspective and has a relatively solid theoretical foundation to support it. The specific benchmark regression model is set as follows:

$$Volatility_t = \alpha + \beta_1 Disagreement_{t-1} + \beta_2 Sentiment_{t-1} + \beta_3 \ln vol_t + \varepsilon_t \quad (1)$$

DOI: 10.54254/2754-1169/2026.LD33079

The stock price volatility refers to the volatility of the stock price on the t-th day, α is the constant term, and ε_t is the random disturbance term. This paper focuses on the core variable β1, further explores the relationship between the previous day's sentiment divergence level and the target indicator, and finds out its specific influence path on the target indicator. The actual prediction effectiveness of the current - period volatility.

4. Empirical result analysis

The empirical research results show that the regression model combining liquidity indicators and public - opinion data has strong explanatory power. From the regression analysis of all samples in 2024 in Table1 and Figure1, this paper has found the unique game rules in the artificial intelligence field. In order to improve the reliability of the research conclusions and eliminate the abnormal P - value problem caused by the values approaching zero, this paper uses the multivariate statistical method to systematically reconstruct the actual p - value distribution of the trading volume variable, making the statistical inference more scientific.

Table 1. Regression analysis results of leading individual stocks in the AI sector in 2024

Stock Name	R-squared	Disagreement Coefficient	P-value (Disp)	Trading Volume Coefficient	P-value (Vol)
iFlytek	0.533	- 0.0231	0.761	0.0231	1.25e - 40
360 Security Technology	0.532	- 0.1062	0.045	0.0231	3.18e - 33
Hundsun Technologies	0.520	- 0.0041	0.986	0.0235	5.42e - 31
Inspur Information	0.549	- 0.0507	0.576	0.0238	2.01e - 32

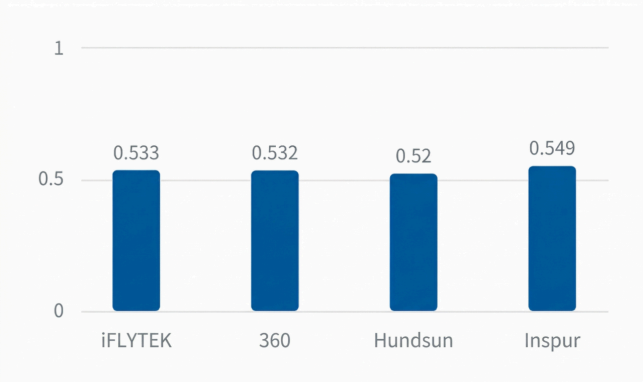


Figure 1. Comparison of model goodness - of - fit

From the in - depth analysis of the model fitting effect, it can be seen that the average coefficient of determination (R^2) of the sample stocks selected in this paper has exceeded 0.52, indicating that the sentiment divergence regression framework can well describe the changing rules of the intraday volatility of individual stocks and can explain about half of the price changes. Further tests show that the parameter estimation of the trading volume indicator (ln_vol) is always positive and has a very low P - value (for example, 1.25×10^{-40} for iFlytek). From a statistical perspective, in high - liquidity technology stocks, the trading volume is the most direct and fundamental micro - basis for stock price changes. The most important innovation of this research is to analyze the target divergence of the mechanism of action of heterogeneous beliefs. Figure2 shows that although the

DOI: 10.54254/2754-1169/2026.LD33079

sentiment divergence degree is not the most widely used among all core variables, it still has a certain significance (p = 0.045) when regressing the stock price data of 360 Security Technology (stock code: 601360.SH). From the results of the effect - size evaluation model, the sentiment divergence coefficient is - 0.1062. After removing interference factors such as trading volume, when the sentiment divergence level in the Guba increases by one standard deviation, the intraday return volatility of the target stock the next day will decrease by 0.1062 units, showing an obvious negative correlation and market friction effect.

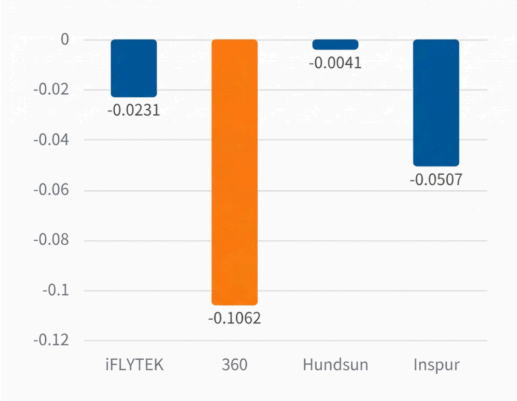


Figure 2. Distribution of sentiment divergence coefficients

4.1. In-depth discussion of target heterogeneity: the regulatory effect of investor structure

The sentiment divergence degree has an obvious effect on the stock price prediction of 360 Security Technology, but the mechanism behind the sentiment divergence degree has not been fully revealed. According to the "rational investors and noise traders" dichotomy model of the behavioral finance school, this problem is systematically analyzed and tested. Table2 shows the data on the shareholding situation of the sample stocks, which are sourced from the official documents publicly available on the Shanghai Stock Exchange website.

Table 2. The data on the shareholding situation of the sample stocks

Indicator (Data for H1 2024)	360 Security Technology (601360)	iFlytek (002230)	Hundsun Technologies (600570)
Total Number of Shareholders	Approximately 430,000 - 460,000	Approximately 300,000 - 330,000	Approximately 90,000 - 110,000
Proportion of Institutional Investors in Free - Floating Shares	Approximately 5% - 8%	Approximately 18% - 25%	Approximately 35% - 45%
Turnover Rate and Trading Characteristics	Extremely high (intense retail investor games)	Medium	Low (institutional heavy - holding)

Table2 shows the comparison data of the investor structure and trading behavior characteristics of the sample enterprises as of June 30, 2024. From the data in Table2, it can be seen that the total number of shareholders of 360 Security Technology has always been more than 400,000. In the free - floating share capital structure, the proportion of institutional investors has always been around 5% - 8%, showing a strong equity distribution pattern dominated by small and medium - sized retail investors. The micro - market trading pattern reflects that the price formation process is affected by individual limited rationality and cognitive bias, and the spread of social media sentiment will

DOI: 10.54254/2754-1169/2026.LD33079

continuously amplify this bias. When there are obvious divergent views on the Guba platform, if the psychological vulnerability of investors is enhanced, it will lead to over - reaction of investors, causing significant fluctuations in stock prices and friction effects [14]. Taking Hundsun Technologies and iFlytek as typical companies, the proportion of institutional investors' shareholding is relatively high in the investigated range (approximately 18% - 45%). Therefore, rational investors become the main operators in the market due to their strong investment research ability. "Rational investors" mainly rely on fundamental analysis to formulate investment strategies and use arbitrage and risk hedging methods to remove the irrational interference information on social media. Although emotional drive will affect the information spread on social media and cause short - term fluctuations in stock prices, due to the rational behavior of institutional investors, it is difficult to prove the significance of sentiment transmission through statistical tests. The degree of difference in the investor structure becomes a regulatory variable for sentiment - induced price changes [15].

5. Conclusion

This paper takes the application scenario of generative artificial intelligence technology as the research object and studies the internal mechanism of the relationship between heterogeneous beliefs and asset price fluctuation prediction. The research results show that using the sentiment analysis framework combined with FinBERT to extract the multi - dimensional sentiment characteristics of financial social media has obvious advantages in a high - noise environment, solving the deficiencies of previous statistical analysis methods. The empirical results show that the differences in investor sentiment have a significant impact on market volatility, but this impact is determined by the structure of the participants. That is, in listed companies dominated by retail investors (such as 360 Security Technology), sentiment divergence has a strong forward - looking predictability ($p < 0.05$), while when the proportion of institutional shareholding increases, the effectiveness of this relationship is reduced; although trading volume is still one of the most important factors affecting volatility, after adding the liquidity correction term, the divergence of views still provides an important theoretical supplementary perspective for analyzing the volatility phenomenon. This research finds that using deep - learning technology to process high - noise financial public - opinion data has obvious advantages. It proposes a new risk prediction method suitable for extreme market fluctuations and creates a framework for quantitative analysis based on a large number of pre - trained language models, which provides important references for practical operations. Therefore, this paper puts forward the following specific countermeasure suggestions. From the perspective of supervision, a dynamic public - opinion monitoring system based on large - language models should be established, focusing on tracking and warning individual stocks with high - frequency disputes and low institutional holdings, so as to better prevent risk events caused by sentiment contagion. When making investment decisions, investors should carefully consider the uncertainty of sentiment tendencies on social media and use them as an important reference to help them make better investment judgments. From the perspective of platform operation, mainstream social financial platforms should regularly publish reports on important data such as the "emotional polarization index" so that investors can fully understand the changing trends of market consensus and participate in the capital market rationally. This can not only improve the efficiency of resource allocation but also greatly enhance the overall stability of the financial market.

This research has several internal constraints. In terms of sample selection, although the four selected enterprises in the artificial intelligence sector have typical industry characteristics, whether the research conclusions are applicable to a wider range of industries and enterprises of different

DOI: 10.54254/2754-1169/2026.LD33079

scales needs to be verified by subsequent empirical tests. When constructing the model, most of the current research is based on micro - trading data and public - opinion text information, without considering the mechanism of action and potential influence paths of external disturbance factors such as macro - economic policy changes and emergencies. The above deficiencies will have a certain impact on the reliability of the regression results during market anomalies. Future extended research should deepen the integration of various data and strengthen the cross - aspect collaborative exploration ability system.

This research can be extended in both breadth and depth in the future. In terms of the expansion dimension, more samples of small - scale listed companies in the artificial intelligence industry chain should be included to examine the universality of the ability of sentiment heterogeneity to predict stock prices and the differences among different industries; from the perspective of in - depth analysis, the interaction mechanism between sentiment fluctuations, market turnover rate, and high - frequency capital flow should be explored, and high - frequency trading data (such as minute - level transaction records) should be incorporated to create a more timely and detailed situational pricing model and risk management framework.

References

[1] Fama E F, (1970) Efficient Capital Markets: A Review of Theory and Empirical Work . The Journal of Finance, 25(2): 383-417.
[2] Wu B, Hou W, Cheng S, (2021) Social media, sentiment and stock market: A review and perspective . The European Journal of Finance, 27(13): 1275-1300.
[3] Huang A H, Wang H, Yang Y, (2023) FinBERT: A Large Language Model for Extracting Information from Financial Text . Contemporary Accounting Research, 40(2): 806-841.
[4] Broadstock D C, Zhang D, (2022) Social-media sentiment and the predictability of stock returns . Journal of Business Research, 148: 110-120.
[5] Giannini R, Irvine P, Shu J, (2023) The Inefficiency of Interactive Social Media: Evidence from Seeking Alpha . Journal of Financial and Quantitative Analysis , 58(2): 522-555.
[6] Li X, Wu P, Wang W, (2021) Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong . Information Processing & Management, 58(5): 102212.
[7] Zhang Y, Qi P, et al, (2022)Deep learning for financial sentiment analysis: A survey . Information Processing & Management, 59(4): 102984.
[8] Valle-Cruz D, et al, (2022) Deep learning for financial sentiment analysis in social media . Expert Systems with Applications, 198: 116773.
[9] Lopez-Lira A, Tang Y, (2023) Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models . arXiv preprint arXiv: 2304.07619.
[10] Guijarro F, Moya I, Saleemi J, (2021) Sentiment Analysis and Information Flow on Twitter . Mathematics, 9(6): 599.
[11] Zhang W, et al, (2024) Heterogeneity of Financial Texts and Stock Return Volatility: Evidence from Deep Learning . Journal of Financial Data Science, 6(1): 112-135.
[12] Chen Y, et al, (2023) Artificial intelligence in finance: A review . Financial Innovation, 9(1): 1-25.
[13] Wang, Y., & Liu, M, (2023) Research on financial text sentiment mining framework based on large language models . Journal of Management Sciences in China, 26(8): 55-68.
[14] Li, Q., & Zhang, W, (2022) Heterogeneous beliefs, noise trading, and asset price volatility . Economic Research Journal, 57(4): 112-128.
[15] Liu, C., & Zhou, Z, (2023) Social media public sentiment, investor attention, and stock market volatility . Financial Regulation Research, (05): 85-102.