4th International Conference on Media, Economy, Communication and Intelligence Management (MECI 2026)

Article

A Multi-Agent Collaborative Investment Research Framework Integrating Large Language Models and Quantitative Factors for Financial Market Decision-Making

Zihan Zhou[1,*]

[1] HD Ningbo School, Ningbo, China
* Correspondence: Zihan Zhou, HD Ningbo School, Ningbo, China

Abstract: Financial investment decisions require comprehensive consideration of the enterprise's basic situation, market changes, and investment risks. Traditional prediction models based on Transformers mainly rely on structured data such as prices, while financial large language models, although capable of understanding and analyzing financial reports, still have deficiencies in quantitative analysis and risk control. This study designed a method consisting of four intelligent agents to conduct investment analysis by simultaneously utilizing enterprise financial information and market transaction data. Among them, the financial report analysis agent is responsible for extracting relevant information from the enterprise annual reports, the quantitative factor agent is used to analyze the characteristics of stock price changes, the risk control agent and the investment portfolio decision agent adjust the investment strategy according to market conditions. The study used the enterprise annual reports from 2021 to 2023, as well as the daily transaction data of Apple, Microsoft and NVIDIA from 2020 to 2023, and further extracted indicators such as stock price trends, price fluctuations and trading activity for analysis. Under the condition of using SPY as the market benchmark, the study conducted an out-of-sample test of the model for 2023. The experimental results show that compared with traditional Transformer models, this method has improved performance in terms of return and risk control, but still has certain gaps compared with the equal-weight strategy and the single-agent strategy of the FinGPT style. The findings from subsequent ablation experiments indicate that integrating enterprise information, market data, and risk management methods can help intelligent investment models develop a more comprehensive analysis process.

Keywords: multi-agent systems; financial large language models; quantitative factors; portfolio management; risk-adjusted performance

Received: 02 July 2026
Revised: 12 August 2026
Accepted: 26 August 2026
Published: 30 August 2026

1. Introduction

Financial investment decisions usually require the simultaneous reference to market transaction data and enterprise public information [1]. Stock prices, trading volumes, etc. can reflect short-term market changes, while annual reports provide longer-term information such as the enterprise's operating conditions, development plans, and risk factors [2]. Past quantitative strategies mainly built models around prices and trading behaviors and achieved good results in handling such structured data. However, their utilization of non-structured text such as financial reports is still limited. Some studies only extracted the information from them as artificially designed indicators [3]. In contrast, large language models have strong text understanding capabilities and can help analyze the key information in a large number of financial documents and identify potential risks.

However, when applied to actual investment decisions, issues such as insufficient stability of the analysis results and how to combine text information with risk control and portfolio management still need to be addressed [4].

In recent years, Transformer models have been widely applied to financial prediction tasks. By learning the patterns of changes in historical prices and trading volumes, these models can identify the complex relationships within market data. However, relying solely on historical transaction information may not fully account for the impact of changes in a company's fundamental situation [5]. At the same time, financial large language models such as FinGPT have begun to attempt to use financial reports for enterprise analysis and generate investment-related judgments [6]. However, in practical applications, if a model is tasked with multiple tasks such as information extraction, market judgment, and asset allocation, its decision-making process may be difficult to explain and it may be challenging to distinguish the impact of different information sources on the final outcome.

The multi-agent approach offers an alternative perspective for financial analysis. By assigning different tasks to different agents, the system can handle financial texts, market indicators, and risk factors separately, and then combine these results to make investment decisions. Currently, there are studies that attempt to combine data from news, social media, etc., or utilize interactions among agents to enhance analytical capabilities. However, these methods usually require a large amount of external data support and increase system operation costs. The actual impact of these methods on investment portfolio performance still needs further evaluation. Therefore, this study focuses on a more concise approach, that is, under limited data conditions, integrating financial report analysis, quantitative indicators, and risk management, and exploring the application value of the multi-agent framework in investment research.

This study designed a multi-agent collaboration system, integrating enterprise financial information, market transaction data, and risk factors into the same analysis process. The financial report analysis agent is responsible for reviewing the annual reports of enterprises, extracting key clues about the business situation and potential risks from them; the quantitative factor agent, by leveraging historical transaction data, continuously tracks the dynamic characteristics of market changes. On this basis, the risk control agent will comprehensively consider the enterprise's risk exposure, historical price fluctuations, and the concentration degree of the investment portfolio, making prudent revisions to the existing investment signals. Finally, the investment portfolio decision-making agent, based on the comprehensive judgment of the above aspects, determines the asset allocation plan for each week.

To test the practical performance of this method, this paper selects Apple, Microsoft and NVIDIA as the research objects, collects daily OHLCV trading data from 2020 to 2023, and analyzes it in combination with nine annual reports from 2021 to 2023. The experiment takes SPY as the market benchmark and compares the proposed method with the equal-weight strategy, the traditional Transformer model, and the FinGPT-style financial text intelligent agent. In the performance evaluation process, the performance of each strategy is uniformly measured using indicators such as annualized return rate, Sharpe ratio, maximum drawdown, volatility, and turnover rate.

This study mainly explores from the following aspects. Firstly, an attempt is made to combine the enterprise information in the annual report with quantitative market indicators, so that different types of information can jointly participate in the investment analysis process. Then, a signal fusion method combining confidence and risk factors is designed. Based on the results of financial text analysis and the performance of quantitative factors, the investment allocation is adjusted together. Finally, the study evaluates from the actual investment portfolio. It should be noted that this paper focuses on examining the feasibility of applying this framework in investment research, and does not attempt to prove that it can continuously achieve market excess returns.

2. Related Work

2.1. Quantitative Factors and Transformer-Based Investment

Traditional quantitative investment strategies typically seek patterns that can be used for investment decisions from historical trading data, and apply this information to asset screening and portfolio construction. Momentum indicators focus on whether the price trend is likely to persist, while moving averages are often used to determine whether current price changes deviate from the long-term trend [7]. Additionally, volatility can reflect the price fluctuations of assets and is an important reference indicator for assessing market risks and adjusting investment positions [8]. Meanwhile, fluctuations in trading volume can provide additional evidence from two dimensions: trading activity and the intensity of price changes. These market variables are often incorporated into statistical models or machine learning frameworks to predict future returns or price trends.

In recent years, the Transformer architecture has been applied to financial time series prediction research. Compared with traditional sequence models, this method can simultaneously consider the connections between multiple time points when analyzing historical data, thus having certain advantages in identifying the long-term influencing factors in price changes [9]. Based on this advantage, existing studies have adopted the Transformer model and combined multiple market data to attempt to predict future price trends [10]. However, it is worth noting that the final performance of such models is, to a large extent, dependent on the quality and completeness of the input data. When trained only using numerical variables such as price, yield, volatility, and trading volume, the Transformer has difficulty utilizing information contained in annual reports, such as the company's business background, risk factors, and management judgments.

2.2. Large Language Models in Financial Analysis

Large language models have provided new approaches for financial text analysis. Due to their ability to understand and process unstructured text information, these models have been applied to tasks such as financial sentiment analysis, financial report summaries, risk identification, question-answering systems, and investment auxiliary analysis [11]. In financial scenarios, large language models can extract information from text content such as management discussions and risk disclosures, helping to identify changes in business operations, cash flow pressures, regulatory impacts, and uncertainties facing future development [12].

FinGPT is an open-source financial large language model framework, mainly focusing on financial data organization, model adaptation, and related application development. This framework is used in financial tasks such as robot investment management, financial text analysis, and algorithmic trading. However, FinGPT itself is not a complete investment portfolio management system; its core objective is to provide basic resources and methodological support for the construction and application of language models in the financial field [13].

Therefore, in this study, FinGPT is used as a benchmark for financial text analysis. Specifically, this model reads the same annual report texts as the intelligent agent for financial report analysis and generates basic face scores at the stock level. Then, based on preset rules, these scores are mapped to portfolio weights. This approach is mainly used to examine the processing ability of a single financial large language model for financial text information.

2.3. Multi-Agent Financial Decision-Making

Multi-agent systems handle complex decision-making problems through task division, entrusting different analysis stages to intelligent agents with specific functions [14]. In the field of financial investment, existing research has attempted to construct different types of agent frameworks, such as those responsible for fundamental analysis, sentiment judgment, technical analysis, transaction execution, and risk management, respectively. TradingAgents further introduces bullish and bearish researchers, forming investment judgments through a structured discussion mechanism; FinMem focuses on the organization of investment memory and the integration ability of multiple sources of information [15].

Based on this, this study focuses on a more concise multi-agent investment framework. This paper mainly conducts research on the analysis of annual reports, the calculation of quantitative factors, risk control, and portfolio allocation under rule constraints. By clarifying the responsibilities of each module, this paper aims to analyze the role of different types of information in the investment decision-making process and further explore the following research questions:

RQ1: Compared with the traditional Transformer model and the single-agent strategy of FinGPT, does the proposed multi-agent framework improve the risk-adjusted investment performance?

RQ2: Can financial report information and quantitative factors provide complementary investment signals?

RQ3: Can the risk control agent reduce the maximum drawdown and decrease the unnecessary trading frequency?

3. Methodology

3.1. Overall Framework

This study has designed a multi-agent investment research framework, which is used to conduct investment analysis by integrating financial text information with quantitative market data. This framework consists of four functional modules, corresponding to financial report analysis, quantitative factor calculation, risk control, and investment portfolio decision-making.

The financial report analysis agent extracts the business information and potential risk factors of the enterprise from the annual report. The quantitative factor agent uses historical price and trading volume data to calculate market indicators such as momentum and volatility, and generates quantitative signals. Then, the risk control agent comprehensively considers the enterprise risks, market fluctuations, asset correlations, and investment portfolio concentration to adjust the initial investment signals.Finally, the investment portfolio decision-making agent determines the asset allocation weights based on the adjusted signals.

Assigning tasks such as text analysis, quantitative analysis, risk assessment and portfolio configuration to different intelligent agents respectively, this framework forms a clearly divided investment decision-making process. The overall architecture is shown in Figure 1.

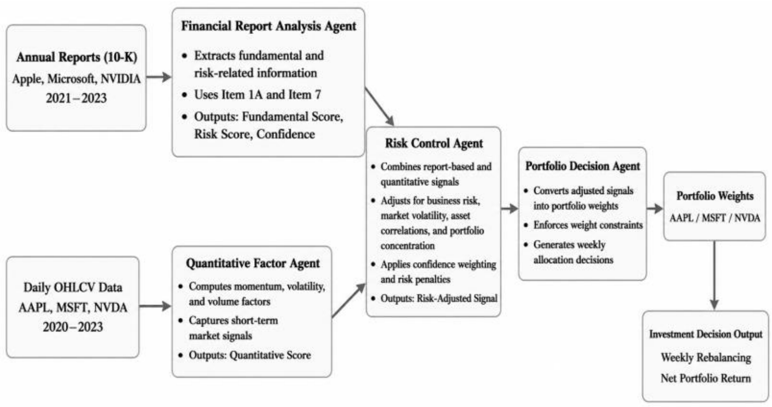

Figure 1. Architecture of the Proposed Multi-Agent Investment Research Framework.

3.2. Financial Report Analysis Agent

This study focuses on the "Risk Factors (Item 1A)" and "Management Discussion and Analysis (Item 7)". These two sections usually contain information about business changes, potential risks, and future development expectations, and are also one of the most important sources of information for investors. The research uses nine annual reports disclosed by Apple, Microsoft, and NVIDIA from 2021 to 2023. All the data comes from the company's official website and investor relations pages. These texts can provide more information about the company's fundamentals, such as operating conditions, cash flow performance, competitive environment, supply chain risks, and regulatory changes.

For each annual report issued by company i during reporting period t, the agent conducts an analysis from six aspects: revenue changes, net profit changes, operating cash flow, profitability, business risks, and management outlook, and based on this, generates a standardized fundamental score:

$$F_{i,t} \in [-1,1] \tag{1}$$

Among them, if $F_{i,t}$ is greater than 0, it indicates that the enterprise's fundamentals are positive. If it is equal to 0, it means the overall performance is neutral; and if it is less than 0, it indicates that the fundamentals are weak.

To assess the reliability of the analysis results, the model will also output the corresponding confidence score:

$$C_{i,t}^{F} \in [0,1] \tag{2}$$

When the financial data are consistent with the information disclosed by the management, the confidence level is relatively high; if there is a significant inconsistency between the two, or if the information provided in the report is insufficient, the confidence level should be reduced accordingly.

All annual reports are processed using a unified template. The model returns the stock code, report year, fundamental score, risk score, confidence level, and corresponding analysis basis. During the experiment, the annual report data was strictly used according to the information disclosure time. That is, only after the annual report is publicly released will the corresponding signals be used for subsequent investment decisions, and they will be retained until the next annual report is released.

3.3. Quantitative Factor Agent

The quantitative factor proxy uses the daily OHLCV data of AAPL, MSFT and NVDA from 2020 to 2023. All the data are from publicly available historical market data and have been formatted into a unified format.

This article selects four quantitative factors. The five-day momentum factor is defined as:

$$MOM_{i,t}^{5} = \frac{P_{i,t}}{P_{i,t-5}} - 1 \tag{3}$$

Among them, $P_{i,t}$ represents the closing price of stock i on the t-th trading day.

The twenty-day momentum factor is:

$$MOM_{i,t}^{20} = \frac{P_{i,t}}{P_{i,t-20}} - 1 \tag{4}$$

The annualized volatility factor on the 20th day is expressed as:

$$VOL_{i,t}^{20} = \sqrt{252}\,\mathrm{Std}(r_{i,t-19}, \dots, r_{i,t}) \tag{5}$$

Among them, $r_{i,t}$ represents the daily return rate of the stock.

The volume change factor is defined as:

$$VC_{i,t} = \frac{V_{i,t}}{\bar{V}_{i,t-20:t-1}} - 1 \tag{6}$$

where $V_{i,t}$ represents the trading volume on the current day, and $\bar{V}_{i,t-20:t-1}$ represents the average trading volume over the previous 20 to 1 trading days.

All factors are subjected to cross-sectional standardization on each decision day. The quantitative factor proxies calculate the comprehensive score based on the weighted results of each factor:

$$Q_{i,t} = \beta_1 Z(MOM_{i,t}^{5}) + \beta_2 Z(MOM_{i,t}^{20}) - \beta_3 Z(VOL_{i,t}^{20}) + \beta_4 Z(VC_{i,t}) \tag{7}$$

Among them, β_1 to β_4 represent the weights of each factor, which are determined by the validation set and remain unchanged during the out-of-sample testing stage. The volatility factor adopts a negative weight to reflect the preference for low-volatility stocks.

3.4. Risk Control Agent

The risk control agent comprehensively utilizes fundamental score, fundamental confidence, quantitative score, annual report risk score, historical volatility, current position status, and correlations among stocks to further adjust the investment signals.

The definition of the risk-adjusted signal is as follows:

$$S_{i,t} = \alpha C_{i,t}^{F} F_{i,t} + (1-\alpha) Q_{i,t} - \lambda R_{i,t} \quad (8)$$

Among them, $\alpha \in [0,1]$ is used to balance the influence of fundamental information and quantitative information; $R_{i,t}$ represents the normalized risk score; $\lambda \geq 0$ is the risk penalty coefficient.

During the signal adjustment process, the fundamental confidence level is used to measure the reliability of the financial report analysis results, to avoid the influence of highly uncertain analysis results on the final decision. The risk score takes into account the business operation risks and the recent market fluctuations, and the two types of information jointly participate in the correction of the investment signals. In addition, the model also combines the correlation between stocks to appropriately reduce the allocation ratio of stocks with similar earnings characteristics, in order to reduce the risk of excessive concentration in the investment portfolio.

The investment portfolio must meet the following constraints:

$$0 \leq w_{i,t} \leq 0.50 \quad (9)$$

$$\sum_{i=1}^{N} w_{i,t} = 1 \quad (10)$$

The holding ratio of a single stock shall not exceed 50%. When the 20-day volatility of a stock exceeds the threshold set during the verification stage, the corresponding investment signal will be reduced; if the annual report indicates that the enterprise has a high risk, the allocation ratio of this stock will also be reduced accordingly. Since this article does not consider the allocation of cash, the investment portfolio always maintains a full investment in the three stocks.

3.5. Portfolio Decision Agent

The portfolio decision agent uses the softmax function to calculate the risk-adjusted asset weights:

$$\widetilde{w}_{i,t} = \frac{\exp(S_{i,t})}{\sum_{j=1}^{N} \exp(S_{j,t})} \quad (11)$$

The higher the risk-adjusted score, the greater the weight assigned to the corresponding stock. Meanwhile, the sum of all initial weights remains at 1. Subsequently, a 50% limit is set for the single stock holding, and the excess portion is proportionally allocated to the remaining stocks.

The investment signals are generated on the last trading day of each week and are executed at the opening of the next trading day. The investment portfolio is held until the next weekly rebalancing. The transaction cost is calculated at 0.1% based on the change in the portfolio weights:

$$TC_t = 0.001 \sum_{i=1}^{N} | w_{i,t} - w_{i,t-1} | \quad (12)$$

Net portfolio returns are calculated after deducting transaction costs.

3.6. Agent Collaboration Process

This framework aims to integrate the advantages of different intelligent agents to achieve decision-making collaboration.

The financial report analysis agent first extracts the basic information and disclosed risks of the enterprise from the annual report, forming the result of text analysis. Thereafter, the quantitative factor agent generates quantitative signals based on market data such as momentum, volatility, and trading volume. Subsequently, the risk control agent integrates the two types of information and adjusts the signals in combination with confidence level, risk penalty, and position concentration constraints. Finally, the investment portfolio decision agent determines the weekly asset allocation weights based on the adjusted results.

Through this design, the large language model is mainly responsible for understanding financial texts, the quantitative module is responsible for market data analysis, the risk control module is used to constrain investment risks, and the investment portfolio module completes the final configuration decision. This framework combines text information with market signals while restricting the model from directly generating un-constrained trading decisions, ensuring that the investment process remains within clear risk management rules.

4. Data and Experimental Design

4.1. Data Sources

The empirical analysis selected three major American technology companies as the research samples, and conducted the analysis by combining their market trading data and annual report texts. Apple (AAPL), Microsoft (MSFT), and NVIDIA (NVDA) were the actual investment targets, while SPY was only used to measure the overall market performance.

The market data covers the daily trading records from January 2, 2020 to December 29, 2023. Each asset contains 1,006 observation data, mainly including fields such as date, opening price, highest price, lowest price, closing price and trading volume. The final data panel consists of 4,024 asset-day records. Among them, the data of MSFT, NVDA and SPY comes from the Stooq platform, while the data of AAPL comes from the public GitHub mirror. The original data source is Yahoo Finance. Due to possible differences in historical adjustment processing methods among different data providers, the different data sources of AAPL may bring certain limitations. All data have been uniformly formatted and maintained within the same time range.

The text data section includes nine annual reports, corresponding to the 10-K filings and annual reports of Apple, Microsoft and NVIDIA from 2021 to 2023 (As shown in Table 1).

Table 1. Data Sources and Experimental Configuration

Panel a. Data Sources					
Data	**Asset or company**	**Period**	**Records**	**Source**	**Use**
Daily OHLCV	AAPL	2020–2023	1,006	GitHub mirror; original source identified as Yahoo Finance	Factors and backtesting
Daily OHLCV	MSFT	2020–2023	1,006	Stooq	Factors and backtesting
Daily OHLCV	NVDA	2020–2023	1,006	Stooq	Factors and backtesting
Daily OHLCV	SPY	2020–2023	1,006	Stooq	Benchmark
Annual-report text	Apple	2021–2023	3	Official source	Fundamental analysis
Annual-report text	Microsoft	2021–2023	3	Official source	Fundamental analysis
Annual-report text	NVIDIA	2021–2023	3	Official source	Fundamental analysis

Panel B. Experimental Configuration	
Item	**Setting**
Investable assets	AAPL, MSFT, and NVDA
Market benchmark	SPY
Rebalancing frequency	Weekly
Signal execution	Next available trading day
Transaction cost	0.1% of portfolio turnover
Training period	2021
Validation period	2022

Out-of-sample test period	2023
Prediction horizon	Five trading days
Main baselines	Equal-Weight, Traditional Transformer, and FinGPT Single Agent
Main metrics	Annualized Return, Sharpe Ratio, and Maximum Drawdown

4.2. Data Processing and Time Alignment

The market data is first sorted by the transaction date, and missing values and duplicate records are checked. The daily return rate is calculated based on the closing prices of adjacent trading days. On this basis, the five-day momentum, twenty-day momentum, twenty-day volatility, and volume change indicators are calculated using the rolling window. Due to the lack of sufficient historical data for some samples, the corresponding initial observation records are deleted after the factor calculation is completed.

AAPL, MSFT and NVDA are considered as potential assets for the investment portfolio. SPY does not participate in the asset allocation. All quantitative signals are calculated based on the information available before the close of the decision day and the corresponding portfolio weights are executed on the next trading day.

For the annual report text, this article mainly analyzes the two sections: Item 1A, Risk Factors, and Item 7, Management's Discussion and Analysis. During the text preprocessing process, the table of contents, duplicate titles, page numbers, and repetitive content were removed, and a unified prompt template was used to analyze all companies and years. The fundamental score generated for each annual report is only used for investment decisions after the report is officially disclosed and continues until the release of the next annual report.

4.3. Training and Testing Design

The samples are divided in chronological order to conform to the characteristics of decision-making based on historical information in financial prediction tasks. The data from 2020 is used for the initialization of the rolling factors and to provide the necessary historical windows. The data from 2021 is used for model training, the data from 2022 is used to determine the factor weights, risk thresholds, and model hyperparameters. The data from 2023 serves as the out-of-sample test set to evaluate the model's performance on unknown data. Since the annual report information covers the period from 2021 to 2023, the main comparisons among the investment portfolio strategies are concentrated on the test period of 2023.

The traditional Transformer model constructs input features based on the OHLCV data of the previous 20 trading days and predicts the cumulative returns for the next 5 trading days. The model employs a Transformer encoder layer, two attention heads, and a 32-dimensional hidden layer. The learning rate is set at 10^{-3}, and early stopping is performed based on the validation set loss. Given that the research sample only includes three stocks, this model adopts a relatively simplified structure.

All model parameters, factor weights, and risk thresholds were determined solely based on the training set and validation set. The test data for 2023 was not involved in any parameter adjustments.

4.4. Baselines

The equal-weight strategy evenly distributes the investments among the three stocks. Each stock is initially assigned an equal weight of 1/3, and these weights are adjusted to equal weights during the weekly rebalancing process.

The traditional Transformer approach utilizes historical market variables to predict future five-day returns, and then converts the predicted returns into non-negative portfolio weights through the softmax function.

The FinGPT single-agent strategy and the financial report analysis agent use the same annual report text and generate standardized fundamental scores for each stock based on the text information. Then, the portfolio weights are determined according to the fixed allocation rules. This benchmark strategy does not include an independent quantitative factor module and a risk control module.

The proposed framework combines financial report analysis agent, quantitative factor agent, risk control agent and portfolio decision agent to complete the investment decision. To ensure fairness in the comparison, all strategies adopt the same asset range, testing period, rebalancing frequency and transaction cost settings.

4.5. Evaluation Metrics

The annualized return rate is used to measure the performance of an investment portfolio during the testing period:

$$\mathrm{AR} = \left[\prod_{t=1}^{T}(1 + r_{p,t})\right]^{252/T} - 1 \quad (13)$$

Among them, $r_{p,t}$ represents the daily net portfolio return rate, and T represents the number of observations within the testing period.

The Sharpe ratio is used to assess the risk-adjusted returns of an investment portfolio:

$$\mathrm{SR} = \frac{\bar{r}_p - r_f}{\sigma_p}\sqrt{252} \quad (14)$$

Maximum drawdown is defined as:Among them, $\bar{r}_p$ and σ_p represent the mean and standard deviation of the daily portfolio return rate respectively. Since no risk-free interest rate data was provided in the dataset, in this paper, r_f was set to 0 for all strategies.

The maximum drawdown is calculated as follows:

$$\mathrm{MDD} = \max_{t}\left(\frac{\max_{\tau\le t} V_\tau - V_t}{\max_{\tau\le t} V_\tau}\right) \quad (15)$$

where V_t represents the cumulative value of the investment portfolio. Additionally, this paper also reports the cumulative return rate, annualized volatility, and turnover rate to further analyze whether the performance of the returns is accompanied by an increase in risk or an increase in trading frequency. All investment portfolio results have deducted a 0.1% transaction cost.

5. Results and Discussion

5.1. Overall Investment Performance

Table 2 presents the performance of the five strategies during the out-of-sample testing period in 2023. During the same period, SPY achieved an annualized return of 26.41% and a Sharpe ratio of 1.86, and maintained the lowest maximum drawdown and volatility levels among all compared strategies. Due to the approximately 49.01%, 58.19%, and 239.02% increases in AAPL, MSFT, and NVDA respectively during the testing period, the strategy based on individual stock allocation achieved higher overall returns.

Table 2. Out-of-sample Investment Performance in 2023

Method	Annualized Return (%)	Sharpe Ratio	Maximum Drawdown (%)	Annualized Volatility (%)	Turnover (×)
SPY	26.41	1.86	9.97	13.08	-
Equal-Weight	104.38	2.87	12.17	26.08	0.96
Traditional Transformer	80.18	2.26	11.04	27.79	45.96
FinGPT-style Single Agent	90.70	2.73	13.17	24.75	2.66
Proposed Framework	90.12	2.71	13.71	24.87	7.80

The equal-weighting strategy achieved the highest annualized return and Sharpe ratio. As this strategy maintains the same allocation ratio for the three stocks at all times, its performance benefited from Nvidia's significant increase during the testing period. The annualized return of the proposed framework was 90.12%, and the Sharpe ratio was 2.71.

Both indicators were higher than those of SPY and the traditional Transformer model. However, compared to the equal-weighting strategy and the FinGPT style benchmark model, its performance still had some gaps.

From the perspective of risk indicators, the proposed framework did not achieve the lowest maximum drawdown. Its maximum drawdown was 13.71%, which was higher than 11.04% of the Transformer model and 9.97% of SPY. Meanwhile, the annualized volatility of this framework was 2.92 percentage points lower than that of the Transformer model, and the turnover rate was 7.80, significantly lower than 45.96 of the Transformer model. This indicates that this strategy does not rely on frequent trading to generate returns. However, its turnover rate is still higher than the equal-weight strategy and the single-factor strategy that only uses report information.

5.2. Cumulative Return and Drawdown Analysis

Table 2(a) presents the performance of the five strategies during the out-of-sample testing period in 2023. During the same period, SPY achieved an annualized return of 26.41% and a Sharpe ratio of 1.86, and maintained the lowest maximum drawdown and volatility levels among all compared strategies. Due to the approximately 49.01%, 58.19%, and 239.02% increases in AAPL, MSFT, and NVDA respectively during the testing period, the strategy based on individual stock allocation achieved relatively high returns overall.

The equal-weight strategy achieved the highest annualized return and Sharpe ratio. As this strategy maintains the same allocation ratio for the three stocks at all times, its performance benefited from the significant increase of Nvidia during the testing period. The annualized return of the proposed framework was 90.12%, and the Sharpe ratio was 2.71. Both indicators were higher than those of SPY and the traditional Transformer model. However, compared to the equal-weight strategy and the FinGPT style benchmark model, its performance still had some gaps.

From the perspective of risk indicators, the proposed framework did not achieve the lowest maximum drawdown. Its maximum drawdown was 13.71%, which was higher than 11.04% of the Transformer model and 9.97% of SPY. Meanwhile, the annualized volatility of this framework was 2.92 percentage points lower than that of the Transformer model, and the turnover rate was 7.80, significantly lower than 45.96 of the Transformer model. This indicates that this strategy does not rely on frequent trading to generate returns. However, its turnover rate is still higher than the equal-weight strategy and the single-factor strategy that only uses report information (As shown in Figure 2).

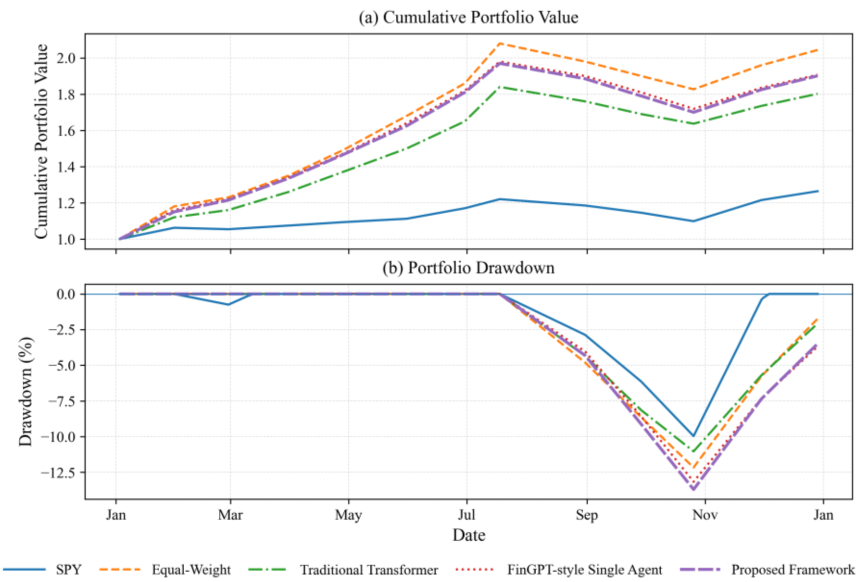

Figure 2. Out-of-sample Portfolio Performance in 2023: (A) Cumulative Portfolio Value and (B) Portfolio Drawdown.

5.3. Ablation Study

Table 3 further analyzes the impact of different framework components on the overall performance.

Table 3. Ablation Results in 2023

Variant	Annualized Return (%)	Sharpe Ratio	Maximum Drawdown (%)	Annualized Volatility (%)	Turnover (×)
Full Multi-Agent Framework	90.12	2.71	13.71	24.87	7.80
w/o Financial Report Agent	90.72	2.45	15.71	27.91	19.32
w/o Risk Control Agent	92.15	2.71	13.80	25.25	8.10
Equal Fusion without Confidence Weighting	88.90	2.64	14.23	25.34	11.50

After removing the financial report analysis agent, the annualized return rate slightly increased, but the Sharpe ratio dropped from 2.71 to 2.45. At the same time, the maximum drawdown increased by 2.00 percentage points, and the turnover rate rose from 7.80 to 19.32. Although this strategy achieved higher absolute returns compared to the complete framework, its risk-adjusted performance declined, and the trading frequency significantly increased, indicating that the annual report information mainly affected the risk control and signal stability of the strategy.

After removing the risk control agent, the annualized return rate increased to 92.15%, but the maximum drawdown, volatility and turnover rate all rose. The Sharpe ratio remained largely unchanged. In the context of the overall market increase in 2023, the risk control module had a relatively minor impact on the return rate, but played a certain role in reducing risk exposure and transaction adjustments. After further adopting an equal-weighted fusion method without including confidence weights, the strategy performance was still lower than that of the complete framework, indicating that the adjustment of confidence levels would affect the degree to which different reporting signals play a role in investment decisions.

5.4. Discussion

The experimental results show that financial report information and quantitative factors have a certain degree of complementarity in investment decisions. Annual report analysis agents mainly provide fundamental information such as the company's profitability, cash flow status, and disclosed risks. Quantitative factor agents, on the other hand, capture short-term market changes based on price, volatility, and trading volume changes. However, the significant increase in Nvidia's stock price in 2023 had a strong impact on some simple allocation strategies. Even without relying on complex analytical methods, maintaining a high holding ratio may still lead to higher returns.

Compared with the version without constraints, the risk control agent reduces the concentration, volatility and turnover rate of the investment portfolio, but the improvement in the maximum drawdown is not significant. This might be related to the overall market increase during the testing period. In an environment of market decline or increased volatility, the role of the risk control mechanism may become more obvious.

The FinGPT-style single-agent strategy verified the application value of annual report texts in investment analysis. However, the signals generated solely based on the report information are difficult to timely reflect changes in market prices. The framework of this paper incorporates quantitative factors to update investment signals, but these additional mechanisms did not bring higher returns during the 2023 testing period.

This study still has certain flaws. The experimental sample only includes three large American technology stocks and nine annual reports, and the out-of-sample test only covers 2023. Moreover, the AAPL price data and other market data come from different sources, and there may be differences in data processing. The study did not include news, macroeconomic indicators, quarterly reports, and restrictions in actual trading. Therefore, the backtesting results cannot directly represent the achievable returns in the real investment environment. Overall, this framework has verified the feasibility of the multi-agent investment analysis process under limited data conditions, but its applicability in a broader stock market still needs further examination.

6. Conclusion

6.1. Main Findings

This study proposes a four-agent collaborative investment research framework consisting of financial report analysis agent, quantitative factor agent, risk control agent and investment portfolio decision agent. This framework combines the textual information in the 10-K report with market factors such as momentum, volatility, and trading volume, and adjusts the investment signals through confidence-weighted, risk-penalized, and position concentration-limited methods.

The empirical analysis is based on the daily OHLCV data of AAPL, MSFT and NVDA from 2020 to 2023, as well as the nine annual reports published from 2021 to 2023, with SPY serving as the market benchmark. The performance of the strategy is evaluated through indicators such as annualized return rate, Sharpe ratio, maximum drawdown, annualized volatility and turnover rate.

During the out-of-sample testing in 2023, this framework achieved an annualized return rate of 90.12% and a Sharpe ratio of 2.71. Compared to the traditional Transformer model, this framework has improved in terms of return performance and risk-adjusted indicators, while maintaining a lower trading frequency. Its overall return is still lower than the single-factor strategy and the benchmark strategy of FinGPT style, and the maximum drawdown is also higher than that of SPY and the Transformer model. Further ablation experiments revealed that after removing the financial report analysis module, the Sharpe ratio decreased, and the maximum drawdown and turnover rate both increased, indicating that the annual report information mainly affects the signal stability and risk-adjusted performance of the strategy.

6.2. Practical Implications

The research results indicate that the multi-agent framework can help the investment analysis process achieve a clearer division of labor. The language model agent is mainly used to understand the textual information in financial reports, the quantitative agent focuses on the market changes in price and transaction data, the risk module is responsible for controlling the portfolio risk, and the portfolio module determines the stock allocation ratio based on the previous analysis results. This approach enables different types of information to be processed more specifically.

6.3. Future Work

Further research can expand the scope of the stock sample and incorporate additional information sources such as 10-Q filings, 8-K disclosures, financial news, and macroeconomic data. At the same time, longer out-of-sample data should be used to further observe the performance of this framework in different market environments, such as periods of market rise, decline, or increased volatility. Additionally, more effective confidence adjustment methods can be explored, including introducing dynamic transaction costs, cash allocation mechanisms, and testing a wider range of industry portfolios. Since the current experiment only covers three technology stocks and the performance of technology stocks in 2023 was particularly outstanding, the results of this paper mainly demonstrate the feasibility of applying this framework, but they cannot prove its stable and effective performance in a broader market.

References

1. Y. Yu, Z. Yao, H. Li, Z. Deng, Y. Jiang, Y. Cao, and Q. Xie, "Fincon: A synthesized llm multi-agent system with conceptual verbal reinforcement for enhanced financial decision making," *Advances in Neural Information Processing Systems*, vol. 37, pp. 137010–137045, 2024.
2. S. Fatemi and Y. Hu, "FinVision: A multi-agent framework for stock market prediction," in *Proc. 5th ACM Int. Conf. AI in Finance*, Nov. 2024, pp. 582–590.
3. X. Han, N. Wang, S. Che, H. Yang, K. Zhang, and S. X. Xu, "Enhancing investment analysis: Optimizing AI-Agent collaboration in financial research," in *Proc. 5th ACM Int. Conf. AI in Finance*, Nov. 2024, pp. 538–546.
4. P. Saha, J. Lyu, A. Saxena, T. Zhao, and D. Mehta, "Large language model agents for investment management: Foundations, benchmarks, and research frontiers," in *Proc. 6th ACM Int. Conf. AI in Finance*, Nov. 2025, pp. 736–744.
5. H. Chen, "Intelligent Agent-Based Market Research: Cloud-Orchestrated Large Language Models as Financial Analysts," *Cloud Computing and Data Science*, pp. 161–168, 2026.
6. Y. Dong, F. Wu, K. Zhang, Y. Dai, S. Zhang, W. Ye, and Z. Q. Cheng, "Large language model agents in finance: A survey bridging research, practice, and real-world deployment," in *Findings of the Association for Computational Linguistics: EMNLP 2025*, 2025, pp. 17889–17907.
7. Y. Wu, Q. Wang, Z. Long, R. Ye, Z. Lu, X. Zhang, and Z. Wei, "FinTeam: A Multi-agent Collaborative Intelligence System for Comprehensive Financial Scenarios," in *CCF Int. Conf. Natural Language Processing and Chinese Computing*, Singapore: Springer Nature Singapore, Aug. 2025, pp. 443–455.
8. H. Wang and Z. Yang, "A Multi-Agent Approach to Investor Profiling Using Large Language Models," in *2025 Int. Conf. Control, Automation and Diagnosis (ICCAD)*, Jul. 2025, pp. 1–6.
9. M. Wawer, J. A. Chudziak, and E. Niewiadomska-Szynkiewicz, "Large language models and the elliott wave principle: A multi-agent deep learning approach to big data analysis in financial markets," *Appl. Sci.*, vol. 14, no. 24, p. 11897, 2024.
10. S. Kundu, D. Sahoo, V. Li, J. Rabowsky, and A. Varshney, "A Multi-Agent Framework for Quantitative Finance: An Application to Portfolio Management Analytics," in *Proc. 2025 Conf. Empirical Methods Natural Language Process.: Industry Track*, Nov. 2025, pp. 812–824.
11. Q. Liu, L. Yao, Z. Wang, Y. Yang, Y. Tang, D. Cheng, and C. Jiang, "Large language model-based multi-agent systems for financial markets simulation: a survey," *Sci. China Inf. Sci.*, vol. 69, no. 7, p. 171201, 2026.
12. M. Rizinski and D. Trajanov, "AI Agents in Finance and Fintech: A Scientific Review of Agent-Based Systems, Applications, and Future Horizons," *Comput. Mater. Contin.*, vol. 86, no. 1, p. 1, 2026.
13. Y. Wu and Y. Li, "Agentic FinTech: A Comprehensive Survey on AI Agents in Finance in the Era of LLMs," SSRN, 2026. [Online]. Available: https://ssrn.com/abstract=6136529
14. A. Zhu and Y. Du, "A Role-Aware Multi-Agent Framework for Financial Education QA," in *Proc. 6th ACM Int. Conf. AI in Finance*, Nov. 2025, pp. 483–491.
15. X. Y. Liu, Y. Cao, and L. Deng, "Multimodal financial foundation models (mffms): Progress, prospects, and challenges," 2025.