Automate Strategy Finding with LLM in Quant Investment

Zhizhuo Kou[1], Holam Yu[2], Junyu Luo[3], Jingshu Peng[1], Xujia Li[1], Chengzhong Liu[1], Juntao Dai[3], Lei Chen[1, 2], Sirui Han[1*†] Yike Guo[1*],

[1] The Hong Kong University of Science and Technology

[2] The Hong Kong University of Science and Technology (Guangzhou)

[3] Peking University

zkouaa@connect.ust.hk {siruihan, yikeguo}@ust.hk

Abstract

We present a novel three-stage framework leveraging Large Language Models (LLMs) within a risk-aware multi-agent system for automate strategy finding in quantitative finance. Our approach addresses the brittleness of traditional deep learning models in financial applications by: employing prompt-engineered LLMs to generate executable alpha factor candidates across diverse financial data, implementing multimodal agent-based evaluation that filters factors based on market status, predictive quality while maintaining category balance, and deploying dynamic weight optimization that adapts to market conditions. Experimental results demonstrate the robust performance of the strategy in Chinese & US market regimes compared to established benchmarks. Our work extends LLMs capabilities to quantitative trading, providing a scalable architecture for financial signal extraction and portfolio construction. The overall framework significantly outperforms all benchmarks with 53.17% cumulative return on SSE50 [1] (Jan 2023 to Jan 2024), demonstrating superior risk-adjusted performance and downside protection on the market.

1 Introduction

Recent advances in LLMs and multi-agent systems are converging to transform quantitative finance. This synergistic relationship leverages LLMs' text comprehension and generation capabilities alongside multi-agent frameworks that simulate market dynamics, creating sophisticated approaches to portfolio management (Lee et al., 2020).

LLMs have evolved from supporting analytical tools to active participants in financial decision-making (Luo et al., 2025c). For example, BloombergGPT demonstrates superior performance in parsing market sentiment and answering domain-specific questions (Wu et al., 2023). Research shows that LLMs effectively generate trading actions by contextualizing price trends with news and earnings reports (Ding et al., 2023). Concurrently, multi-agent systems offer powerful approaches to portfolio optimization through decentralized interaction. The Multi-Agent Portfolio System demonstrates the portfolio managed by agents have achieved well-diversified returns with improved risk-adjusted performance (Lee et al., 2025). These frameworks capture complex market dynamics such as information sharing and strategic arbitrage that single-agent models cannot address (Spooner and Savani, 2020). The integration of these technologies creates sophisticated financial environments where LLM-enhanced agents demonstrate adaptive behavior. StockAgent exemplifies this approach with LLM-powered agents mimic diverse investor personas responding to market events (Zhang et al., 2024a). Hierarchical structures such as FinCon organize agents in manager-analyst relationships, facilitating collaboration through natural language communication (Yu et al., 2024). This convergence heralds a future of intelligent, distributed financial decision making that combines data-driven learning with human-like reasoning capabilities for more robust investment strategies (Yu et al., 2023b).

Alpha mining is the process of discovering trading signals that generate excess returns in financial markets, we identify three critical challenges in alpha mining: Rigidity of traditional methods that lack adaptability to dynamic markets (Tang et al., 2025); Data diversity and integration challenges despite machine learning advances (Cui et al., 2021; Yu et al., 2023a); and Adaptation to market variability despite progress in prediction (Xu and Cohen, 2018) and strategy formulation (Chen et al., 2019).

Our LLM-driven framework addresses these lim-

*Corresponding author.

†Project leader

[1]The SSE 50 Index tracks the performance of the 50 most influential large-cap blue-chip stocks on the Shanghai Stock Exchange.

Findings of the Association for Computational Linguistics: EMNLP 2025, pages 18517–18533
November 4-9, 2025 ©2025 Association for Computational Linguistics

itations through three components: Flexible Alpha Mining employs LLMs to extract, categorize, and filter alpha factors from financial literature, organizing them as momentum, fundamental, or liquidity factors with established independence (Xu and Cohen, 2018). Multi-agent Multimodal Market Evaluation conducts rigorous backtesting across diverse market conditions, with specialized agents evaluating factor effectiveness from multiple perspectives. Dynamic Strategy Optimization implements a weight gating layer that assigns optimal alpha factor weights based on current market conditions, ensuring adaptive strategy development.

Our methodology synthesizes cutting-edge machine learning techniques with established financial domain knowledge to create a robust interdisciplinary framework for alpha identification and optimization across diverse asset classes. This research is grounded in empirical quantitative investment practices, bridging theoretical advancements with practical applications in portfolio management.

Our main contributions are three-fold: ❶ A novel framework for identifying formulaic alpha factors using LLMs, leveraging their exploratory capabilities to establish an Alpha factory from multimodal information with incremental update functionality; ❷ Introduction of a multi-agent approach to portfolio management for evaluating relationships between market conditions and alpha factors, enabling specialized evaluation under different scenarios; ❸ Integration of advanced techniques from machine learning and finance, representing a significant advancement in developing robust, adaptive investment strategies without human intervention.

The proposed framework demonstrates versatility across various asset classes, enhancing its utility and practical effectiveness. To support future research and ensure reproducibility, we make source code publicly available at https://github.com/kouzhizhuo/Automate-Strategy-Finding-with-LLM-in-Quant-investment.

2 Problem Formulation

This section establishes the theoretical foundation for our research on alpha factor strategies in quantitative finance. We formulate a framework addressing three interconnected challenges: mathematically formalizing these alphas, developing dynamic methodologies to generate seed alphas, and defining and optimizing alpha factor strategies. Our approach integrates LLMs and multi-agent systems to overcome limitations in traditional quantitative trading methods.

We employ consistent notation throughout: n stocks observed over trading periods $t \in 1, 2, \ldots, T$, each characterized by m financial features. Alpha factors are denoted as $\alpha_{ij}^{(t)}$ for stock i in category j at time t, with corresponding weights w_i. Market conditions at time t are represented by $\mathcal{M}^{(t)}$, and alpha factor predictive power is measured using the Information Coefficient (IC) 1.

$$IC = \sigma(u, v) \tag{1}$$

A higher IC indicates stronger predictive relationships between alpha values and future returns. where $\sigma(u, v)$ is correlation coefficient between predicted alphas u and actual future returns v.

2.1 Alpha Factor Representation

The first challenge involves effectively representing alpha factors using mathematical expressions that capture meaningful financial signals (Lo, 2007). Given raw financial features $\mathbf{X}i^{(t)} = \{Xi1^{(t)}, X_{i2}^{(t)}, \ldots, X_{im}^{(t)}\}$ for each stock i at time t (such as price, volume, and volatility), we define alpha factors through two classes of operators:

Cross-Section Operators f_{cs}: These operators process data from a single time period, capturing instantaneous relationships between financial variables:

$$\alpha_{cs}^{(t)} = f_{\text{cs}}(\mathbf{X}i^{(t)}) \tag{2}$$

Time-Series Operators f_{ts}: These operators analyze data spanning multiple periods, identifying trends and temporal patterns:

$$\alpha_{ts}^{(t)} = f_{\text{ts}}(\mathbf{X}i^{(t)}, \mathbf{X}_i^{(t-1)}, \ldots, \mathbf{X}_i^{(t-n)}) \tag{3}$$

The seed alpha for stock i in category j at time t is formulated as:

$$\alpha_{ij}^{(t)} = w_{cs} \cdot f_{\text{cs}}(\mathbf{X}i^{(t)}) + wts \cdot f_{\text{ts}}(\mathbf{X}i^{(t)}, \ldots, \mathbf{X}_i^{(t-n)}) \tag{4}$$

where w_{cs} and w_{ts} are weights assigned to the cross-section and time-series components respectively. This representation allows for the flexible combination of different financial signals, enabling the creation of complex and nuanced alpha factors.

2.2 Alpha Mining and Selection

The second challenge addresses the limitations of traditional alpha mining methods, which often fail to adapt to changing market conditions. We formulate a dynamic selection approach that identifies the most relevant alphas for current market conditions using two complementary evaluation mechanisms:

Confidence Score Evaluation: Assesses the statistical reliability of each alpha factor:

$$\theta_{ij} = \mathbb{E}\left[IC(\alpha_{ij}^{(t)}|\mathcal{M}^{(t)})\right] \tag{5}$$

where θ_{ij} represents the confidence score for alpha α_{ij}, and $\mathbb{E}[\cdot]$ denotes the expected value. Higher confidence scores indicate more consistent performance across various market environments.

Risk Preference Evaluation: Examines the risk characteristics of each alpha factor:

$$\rho_{ij} = f_{\text{risk}}(\alpha_{ij}^{(t)}, \mathcal{M}^{(t)}) \tag{6}$$

where ρ_{ij} represents the risk score and f_{risk} is a function that evaluates how well the alpha performs under different risk scenarios.

The optimal set of seed alphas is selected by considering both confidence and risk evaluations:

$$\alpha_{ij} = \arg\max_{\alpha ij} \left[w_c \cdot \theta_{ij} + w_r \cdot \rho_{ij}\right] \tag{7}$$

where w_c and w_r are weights assigned to confidence and risk scores respectively, reflecting their relative importance in the selection process.

2.3 Strategy Optimization

The third challenge involves combining the selected alphas into an effective investment strategy. The final alpha strategy is defined as a weighted combination of the optimal alphas from each category:

$$\alpha^{(t)} = \sum_{j=1}^{k} w_j \cdot \alpha_{ij} \tag{8}$$

where w_j represents the weight assigned to category j, and k is the total number of alpha categories. These weights are dynamically adjusted based on current market status to maximize strategy performance while managing overall portfolio risk. This optimization completes our framework, transforming raw financial data into a robust trading strategy that can adapt to changing market environments.

3 Methodology

3.1 Framework Overview

Our framework comprises three interconnected components (Figure 1): the SAF, multi-agent decision-making, and weight optimization. The initial phase employs LLMs to analyze and categorize multimodal financial research documents, constructing a comprehensive SAF. The LLM's capability to process diverse datasets enables the creation of a robust set of seed alphas categorized into independent groups, aligning with established finance alpha mining principles (OpenAI, 2023). The second phase implements a multimodal multi-agent evaluation process that incorporates varied risk perspectives, enhancing strategy adaptability across different market conditions. This phase produces an optimized alpha set tailored to current market states and risk preferences. The final phase employs deep learning methods to optimize the weights of selected alphas, constructing a cohesive overall strategy.

The framework's dynamic architecture enables continuous refinement through incremental updates to the SAF as new research emerges and market conditions evolve. This adaptability ensures the strategy maintains relevance and robustness over time. The methodology's versatility permits application to any structured market globally, effectively replicating and enhancing professional investment research approaches.

3.2 LLM-Based Seed Alpha Generation

The first stage implements an LLM-based filtering and categorization process for alpha-related research. We utilize GPT-4o model to perform these tasks on a diverse corpus of financial research. Our initial dataset comprises 11 documents spanning both theoretical and applied aspects of alpha mining research (see Appendix A.2). Through this process, the system generates 9 distinct categories containing 100 seed alphas. The approach enhances the model's ability to extract intricate details and relationships within financial research, resulting in a more robust and diverse SAF.

The output is a structured set of seed alphas categorized into distinct financial domains such as Momentum, Mean Reversion, and Fundamental analysis et al. Each category includes specific alpha designations and corresponding executable formulations derived from the LLM's analysis (see Appendix A.3). This structured output forms the foundation for subsequent processing stages.

3.3 Multimodal Multi-Agent Evaluation

The second stage implements a comprehensive evaluation and selection of alpha factors through a multimodal and multi-agent system. Our system

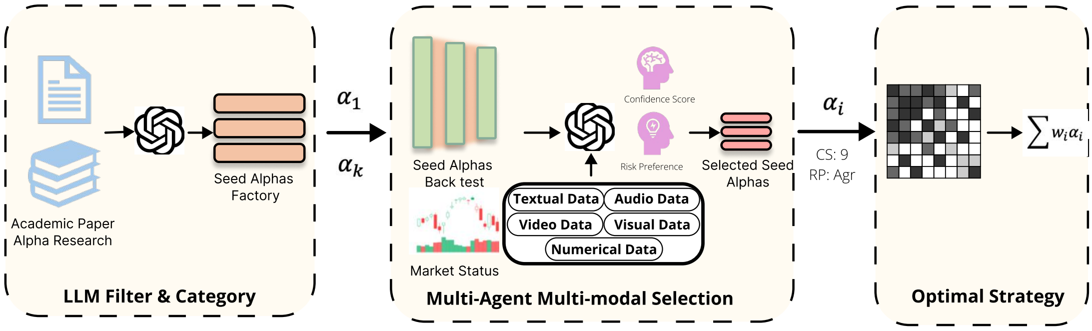

Figure 1: Overview of the strategy generate process in three components, a Seed Alpha Factory built using LLMs, a multi-agent decision-making system, and a weight optimization approach for overall strategy (CS stands for confidence score; RP stands for risk preference)

incorporates five types of multimodal data (detailed in Appendix A.4) encompassing textual, numerical, visual, audio, and video inputs. The multi-agent architecture comprises two agents Confidence Score Agent (CSA) and Risk Preference Agent (RPA). Selected alpha factors undergo rigorous backtesting using historical market data, with key evaluation metrics including the IC (Goodwin, 1998) 1 and Sharpe Ratio (Sharpe, 1994) 9:

$$\text{Sharpe Ratio} = \frac{\mathbb{E}[R(\alpha^{(t)}) - R_f]}{\sqrt{\text{Var}[R(\alpha^{(t)})]}} \tag{9}$$

where $R(\alpha^{(t)})$ represents the return of the strategy and R_f represents the risk-free rate. We developed a Category-Based Alpha Selection algorithm (detailed in Appendix A.5) to automate the selection. This algorithm systematically identifies and selects alphas from different categories based on their confidence scores, ensuring rigorous selection of factors that meet confidence thresholds across all categories while maintaining category diversity.

3.4 Weight Optimization

The final stage employs a 3-layer MLP to optimize the weights of selected seed alphas by mapping historical alpha calculations to future yields (Chen and et al., 2020). The architecture consists of an input layer processing daily alpha values, a hidden layer with ten ReLU-activated nodes, and a single-node output layer for yield prediction.

During training, the network employs backpropagation and gradient descent to minimize the loss function, quantifying the discrepancy between predicted and actual yields. We utilize a separate validation set to ensure model generalizability and prevent overfitting. The DNN processes input data through the hidden layer, transforming it with learned weights and biases, and generates the final output through the output layer's weights, biases, and activation function.

This methodology establishes a robust framework for predicting future yields based on historical alpha values, demonstrating the efficacy of deep learning techniques in optimizing alpha weights and enhancing investment strategy performance. Algorithm **??** provides a formal specification of our complete framework, illustrating the logical flow from multimodal data processing through multi-agent evaluation to final weight optimization.

4 Experiment

Our research aims to develop a comprehensive LLM-driven alpha mining framework that operates without human intervention. This framework is uniquely capable of processing multimodal information and adapting to varying market conditions. To validate the effectiveness of our framework, we have conducted a series of experiments.

4.1 Datasets

Our study focuses on financial data from the Chinese market and US market, specifically targeting the SSE 50 Index. Table 11 shows the experiment dataset, which encompasses six primary features as original inputs for our Alpha factors: open, high, low, close, volume (OHLCV), and volume-weighted average price (VWAP). To ensure rigorous evaluation and robust model performance. Our experiments integrate financial reports and factor performances of the 50 constituent companies of the SSE 50 Index, providing a comprehensive view of the market (Li and Mei, 2020). The evaluation considers various datasets, including financial reports from the specified periods and performance

Table 1: Summary of the experiment dataset

Aspect	Details
Primary Features	Open, High, Low, Close, Volume, VWAP
Alpha Factors	Custom factors based on price, volume, financial ratios, moving averages, sentiment analysis
Financial Reports	Quarterly and Annual reports from Index constituent companies
Time Periods	Jan 2019-Jun 2024
Market Coverage	SSE50, CSI300, SP500 Index
Evaluation Criteria	Causal relationships, Alpha factor performance, model robustness

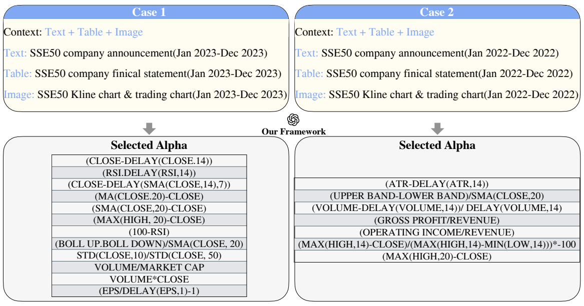

Figure 2: Sample Experiment on Different Market Status Input and Alpha Selection, Selected Alpha Depends on Different Context

metrics of different Alpha factors.

4.2 Multimodal Knowledge Extraction and Adaptive Alpha Discovery

We implement a prompt architecture (Figure 2) that incorporates multimodal market information into LLMs for comprehensive knowledge extraction and seed alpha selection. This architecture integrates textual financial sentiment data, numerical company financial statements, and visual trading charts to provide holistic market analysis (Luo et al., 2025b). Our contextual analysis mechanism dynamically adjusts parameters based on prevailing market trends and sector performance. Experimental validation demonstrates the framework's adaptability across varying market conditions. In Case 1, analyzing SSE50 data from Jan 2023 to Dec 2023, the model selected momentum and volume-based indicators such as price momentum, RSI, and MACD. Conversely, in Case 2, when processing the time window from Jan 2022 to Dec 2022, the model prioritized volatility and economic factors, including ATR, Bollinger Bands, and gross profit indicators.

These results confirm the framework's capacity to dynamically adapt to changing market conditions through effective multimodal data integration. This adaptability enables the identification of market-appropriate alphas, enhancing strategy robustness across diverse market environments.

Table 2: IC comparison of mean & selected

	Momentum	Mean Reversion	Volatility	Fundamental	Growth
Mean IC of SAF	0.0092	0.0135	0.0177	0.0118	0.0146
Mean IC of Selected SAF	**0.0208**	**0.0187**	**0.0258**	**0.0192**	**0.0217**

4.3 Comparative Performance Against Traditional Alpha Factories

When evaluating the performance of selected seed alpha signals, the primary metric is the IC. We evaluated five most common alpha categories: Momentum, Mean Reversion, Volatility, Fundamental and Growth. The results in Table 2 demonstrate that our LLM-driven framework consistently achieves higher average IC values across all categories, particularly in Volatility and Fundamental, indicating superior trading effectiveness compared to original.

4.4 Performance Evaluation of the Integrated Alpha Framework in Benchmark Comparison

Table 3: SSE50 2023 test combination of 12 Alphas

#	Alpha	Weight	IC(SSE50)
1	(CLOSE - DELAY(CLOSE, 14))	-0.1459	0.0209
2	(RSI - DELAY(RSI, 14))	-1.0265	-0.0225
3	(CLOSE - DELAY(SMA(CLOSE, 14), 7))	-0.1978	0.0193
4	(MA(CLOSE, 20) - CLOSE)	0.0556	-0.0186
5	(SMA(CLOSE, 20) - CLOSE)	-0.945	-0.0186
6	(MAX(HIGH, 20) - CLOSE)	-0.4053	-0.0185
7	(100-RSI)	-0.3199	0.0194
8	(BOLL_UP - BOLL_DOWN) / SMA(CLOSE, 20)	3.6186	0.0278
9	STD(CLOSE, 10) / STD(CLOSE, 50)	-0.183	0.0236
10	VOLUME / MARKET_CAP	-3.2145	-0.0194
11	VOLUME * CLOSE	-0.0058	0.0187
12	(EPS / DELAY(EPS, 1) - 1)	-1.8351	-0.0215
	Weighted Combination		-0.0587

Table 3 presents an example combination of 12 alphas generated by our framework, evaluated on the SSE50 constituent stock set. The table details the seed alphas selected by the LLM from each category, along with their respective weights and IC 1 values. The weight combination IC 1 value is quite high as -0.0587. Although some of the seed alphas exhibit relatively low IC 1 values individually, their removal results in a significant drop in the retrained combination weight, indicating their critical role in the overall performance. For example, if we remove alpha #6 the weight combination will drop to -0.055; once we remove the alpha #11, the weight combination will drop to merely 0.0491. This suggests that the seed alpha set selected by the LLM synergizes effectively, providing robust predictive power (Zhang et al., 2020). To address the ques-

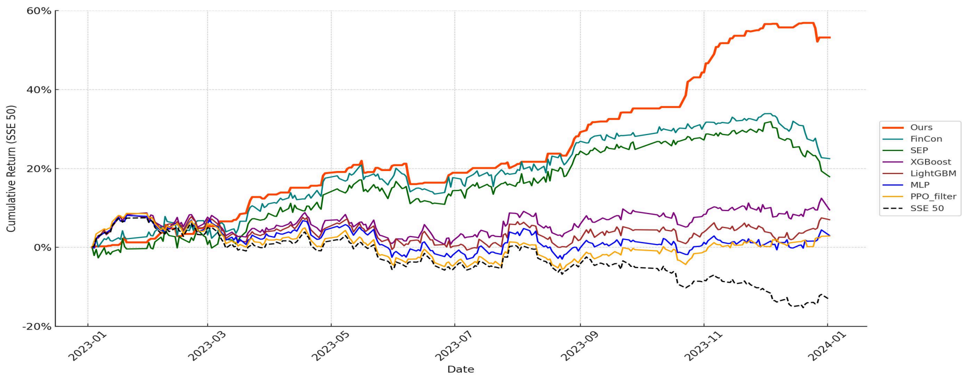

Figure 3: Cumulative return backtest result on SSE50. The line track the net worth of different methods

tion of whether our overall strategy, incorporating the LLM-driven framework and confidence scoring, can consistently beat the market, we conducted a backtest using a straightforward investment approach during the period from Jan 2023 to Jan 2024 (training: Jan 2021-Jun 2022; validation: Jul 2022-Dec 2022), on the SSE50 dataset we conduct the framework to construct the portfolio (A.8). The net worth progression of the respective strategies over the testing period is shown in Figure 3, and the comparison matrix shows in Table 4. Despite not explicitly optimizing for absolute returns, our framework demonstrates impressive performance in the backtest, achieving the highest profit compared to other methods. Our cumulative return for 2023 backtest comes to 53.17% positive and meanwhile the index performance is -11.73%. We also do the compression with 4 advanced statistical methods and 2 LLM based methods. The framework strategy substantially outperforms all competitors with the highest return (53.17%), best risk-adjusted metrics (Sharpe: 0.287, Sortino [2]: 0.208, Calmar [3]: 1.052), and lowest volatility (0.762%). LLM based mathods FinCon and SEP show moderate success (22.47%, 17.89%), while conventional machine learning methodologies yield marginal returns (XGBoost 9.53%). Notably, all evaluated strategies surpass the SSE50 benchmark, which exhibits negative performance (-13.22%). This approach demonstrates the efficacy of our LLM-driven framework consistently outperforming market benchmarks through dynamic adaptation to changing conditions.

Table 4: Performance comparison of trading strategies

Strategy	Final Return (%)	Sharpe Ratio	Volatility (%)	Sortino Ratio	Calmar Ratio
Ours	**53.173**	**0.287**	**0.762**	**0.208**	**1.052**
XGBoost (Chen and Guestrin, 2016)	9.532	0.038	1.019	0.067	0.103
LightGBM (Ke et al., 2017)	7.125	0.030	0.993	0.053	0.066
MLP	3.110	0.013	0.960	0.023	0.043
PPO_filter (Schulman et al., 2017)	2.865	0.013	0.886	0.024	0.017
FinCon (Yu et al., 2024)	22.474	0.077	1.196	0.126	0.232
SEP (Koa et al., 2024)	17.891	0.060	1.217	0.103	0.157
SSE 50	-13.22	-0.063	0.859	-0.111	-0.043

4.5 Framework Robustness Across Time Periods and Markets

We conducted rigorous cross-temporal and cross-market validation using two comprehensive datasets: CSI300 [4] constituents (accessed via Tushare API) and SP500 [5] constituents (accessed via CRSP platform) spanning January 2019 to December 2023. Our dataset incorporated daily OHLCV price metrics, quarterly financial statements, and relevant macroeconomic indicators to enable robust alpha factor construction. To ensure methodological integrity, we systematically partitioned the dataset into training, validation, and testing subsets with precise chronological segmentation as detailed in Table 5. This experimental design facilitates comprehensive evaluation of our framework's generalizability across different market environments and temporal contexts.

We employed consistent temporal partitioning across Chinese A-share and US markets to enable

[2]Sortino ratio computing excess return per unit of downside deviation, evaluating only negative return volatility.

[3]Calmar ratio divides annualized return by maximum drawdown, measuring reward relative to risk.

[4]CSI300 Index tracking the 300 largest companies listed on China's Shanghai and Shenzhen exchanges.

[5]SP500 Index measuring the performance of 500 large U.S. companies traded on American stock exchanges.

Table 5: Training, validation, and test Periods for CSI300 and SP500

Assets	Training Period	Validation Period	Test Period
CSI300	Jan 2019-Jun 2020	Jun-Dec 2020	Jan-Jun 2021
	Jan 2020-Jun 2021	Jun-Dec 2021	Jan-Jun 2022
	Jan 2021-Jun 2022	Jun-Dec 2022	Jan-Jun 2023
SP500	Jan 2019-Jun 2020	Jun-Dec 2020	Jan-Jun 2021
	Jan 2020-Jun 2021	Jun-Dec 2021	Jan-Jun 2022
	Jan 2021-Jun 2022	Jun-Dec 2022	Jan-Jun 2023

Table 6: Backtest performance results across different time windows

Time Window	Annual Return (%)		Cum. Return (%)		Max DD (%)	
	Strategy	Baseline	Strategy	Baseline	Strategy	Baseline
Our Strategy vs CSI300						
Jan-Jun 2021	**29.70**	7.31	**11.91**	3.10	-19.40	**-10.86**
Jan-Jun 2022	**12.78**	-30.37	**5.29**	-14.37	**-24.01**	-28.59
Jan-Jun 2023	**192.27**	9.13	**59.03**	3.85	-17.03	**-8.44**
Our Strategy vs SP500						
Jan-Jun 2021	**93.61**	29.67	**35.19**	12.59	-7.89	**-4.23**
Jan-Jun 2022	**2.77**	-44.22	**1.25**	-23.39	**-20.55**	-23.51
Jan-Jun 2023	**118.24**	35.22	**42.78**	14.76	-11.52	**-7.75**

Note:Max DD = Maximum Drawdown.

direct comparative analysis. Our findings demonstrate the framework's robust performance across diverse market environments. From Table 6 shows in the Chinese A-Share market, the strategy generated substantial alpha with CSI300 constituents, particularly in H1 2023 (192.27% annual return vs. benchmark 9.13%). Similarly, in the US market, the framework achieved strong returns with SP500 constituents (93.61% in H1 2021, 118.24% in H1 2023). Notably, during the H1 2022 market downturn, the strategy maintained positive returns in both markets (12.78% in CSI300, 2.77% in SP500) while their respective benchmarks declined significantly (-30.37% and -44.22%).

The empirical evidence substantiates both the temporal persistence and cross-market robustness of our findings, with notably beneficial countercyclical characteristics emerging during periods of market distress. The framework demonstrates consistent alpha generation across diverse market architectures, regulatory frameworks, and investor behavioral patterns, establishing its broad applicability. Specifically, the model maintains stable outperformance trajectories through bullish, bearish, and range-bound market conditions, thereby validating its structural integrity and adaptability to varying macroeconomic environments.

4.6 Ablation Study

Our ablation study systematically evaluates the contribution of CSA and RPA within a multi-agent

Table 7: Ablation study results: impact of agent components on performance metrics

Model Configuration	IC (In-Sample)	IC (Out-of-Sample)	Sharpe Ratio
Full Model	**0.059**	**0.047**	**1.94**
Without Confidence Score	0.054	0.032	1.51
Without Risk Preference	0.056	0.039	1.34

Table 8: Performance across market regimes (out-of-sample IC)

Model Configuration	Bull Market	Bear Market	Sideways Market
Full Model	**0.051**	**0.042**	**0.045**
Without Confidence Score	0.046	0.021	0.029
Without Risk Preference	0.049	0.028	0.037

framework. We evaluate three configurations of each, using the SSE50 index data (2010-2022) and measure performance by IC 1 and Sharpe Ratio 9.

Table 7 demonstrates that the complete model achieved superior performance, with an out-of-sample IC of 0.047 and Sharpe Ratio of 1.73. Removing the CSA caused substantial degradation, reducing out-of-sample IC by 31.9% and Sharpe Ratio by 22.5%. RPA removal also decreased performance metrics, with particularly significant impact on the Sharpe Ratio. These results indicate that while both components contribute meaningfully, the CSA plays a more critical role in maintaining predictive stability. To assess performance consistency across market conditions, we analyzed out-of-sample IC values during different market regimes (Table 8). The complete model maintained consistent performance across bull, bear, and sideways markets. In contrast, the model without CSA performed particularly poorly during bear markets(IC: 0.021 vs. 0.042). The model without RPA showed moderate degradation during non-bull markets, less severe than observed without CSA.

This analysis confirms that both components enhance performance stability, with the CSA providing especially critical functionality during adverse market conditions.

4.7 Sensitive Study

We analyzed the framework's sensitivity to agent weights and neural network hyperparameters. The CSA & RPA weight ratio of 0.6/0.4 provided superior performance across market regimes, particularly during bear markets (Sharpe: 10.37), indicating enhanced robustness to regime shifts (Table 9). Optimal neural network configuration (10 hidden nodes, learning rate: 0.001, batch size: 32,

Table 9: Sensitivity to agent weight configurations

Confidence Weight	Risk Weight	Sharpe Ratio Bull Market	Bear Market	Overall
1.0	0.0	4.32	6.60	8.10
0.8	0.2	-3.41	-7.23	-2.68
0.6	0.4	**8.70**	**10.37**	**11.39**
0.5	0.5	-0.49	-1.62	-5.10
0.4	0.6	-4.27	-4.41	-8.04
0.2	0.8	5.68	8.51	9.00
0.0	1.0	-0.61	1.78	-4.35

Table 10: Sensitivity to neural network hyperparameters

Hidden Nodes	Learning Rate	Batch Size	Regularization Parameter	Sharpe Ratio
5	0.001	32	0.001	9.69
10	0.001	32	0.001	**13.33**
20	0.001	32	0.001	6.93
10	0.0005	32	0.001	7.03
10	0.002	32	0.001	5.13
10	0.001	16	0.001	6.26
10	0.001	64	0.001	4.25
10	0.001	32	0.0005	-1.88
10	0.001	32	0.002	-6.91

regularization: 0.001) achieved a Sharpe ratio of 13.33, with performance particularly sensitive to regularization strength (Table 10).

The analysis revealed regime-dependent behavior across market conditions, and critical sensitivity to regularization strength. The optimal configuration achieved a robust overall Sharpe ratio of 11.39.

5 Related Work

Formulaic alphas in quantitative investment represent systematic, rule-based strategies that generate excess returns by exploiting specific market patterns and inefficiencies (Kakushadze, 2016). These strategies employ various methodologies, including genetic programming that involves structural and numerical mutations to generate novel alphas (Cong et al., 2021), enhanced time-series operators with mutual information as fitness measures (Lin et al., 2019), and algorithmic graphs for more complex predictions (Cui et al., 2021). Machine learning approaches utilize neural network architectures such as LSTM (Hochreiter and Schmidhuber, 1997) and Transformer models (Vaswani et al., 2017), while decision tree models like XGBoost (Chen and Guestrin, 2016) and LightGBM (Ke et al., 2017) offer interpretability advantages. Recent research focuses on integrating non-standard data sources, as demonstrated by REST (Xu et al., 2021b) and HIST (Xu et al., 2021a).

The development of **general-domain LLMs** has catalyzed interest in Finance LLMs (Fin-LLMs), although this specialized domain remains nascent (Novy-Marx, 2015; Yang et al., 2023; Zhao et al., 2023). Open-source LLMs such as LLaMA (Touvron et al., 2023), BLOOM (W. and et al., 2023), and Flan-T5 (W. and et al., 2022) provide flexibility but may underperform proprietary alternatives. Fine-tuned financial LLMs demonstrate enhanced domain-specific comprehension, yet their generative performance indicates the need for improved domain-specific datasets (Lewis and et al., 2020; Koa et al., 2024; Wen et al., 2025).

Multimodal LLMs have shown significant potential in investment contexts by processing diverse data types (L et al., 2023; Luo et al., 2025a), developing strategies that mitigate market volatility (K et al., 2024), and analyzing textual data to gauge investor sentiment (Zhao et al., 2024). Multi-agent LLM systems enhance market analysis capabilities by leveraging vast datasets to interpret financial reports and market sentiment (Zhang et al., 2024b), simulating various market scenarios (Talebirad and Nadiri, 2023), and facilitating parallel testing of diverse strategies (Wang et al., 2024). Implementation raises important considerations regarding transparency, accountability, and bias mitigation (Yu and et al., 2024; Mundhenk et al., 2021).

6 Conclusion

In this paper, we proposed a novel quantitative investment framework integrating LLMs and multi-Agent architectures to address instability in traditional approaches. Our system generates diversified alpha factors from multimodal financial data, constructs risk-calibrated trading agents, and employs a deep learning mechanism for dynamic agent weighting based on market conditions.

Experimental results confirm the framework's effectiveness across Chinese and US markets, our framework demonstrates significant outperformance versus SOTA alpha generation methods and benchmark indices across key financial metrics. This work successfully extends LLM capabilities to quantitative trading, creating a scalable, adaptive architecture for financial signal extraction that functions effectively without human intervention across diverse market regimes.

7 Limitations

Our framework presents several significant limitations. First, system efficacy is contingent upon input document quality, potentially perpetuating inherent biases (Deb et al., 2017; Ashok et al., 2018). Second, LLM-generated alphas occasionally lack the financial intuition characteristic of human analysts, resulting in theoretically sound but practically infeasible factors (Tuarob et al., 2017). Third, our multi-agent evaluation methodology presupposes persistent historical relationships between market conditions and alpha performance, an assumption that may prove tenuous during market regime shifts. Finally, our validation efforts have primarily targeted equity markets, with cross-asset applicability requiring additional empirical investigation. Future research directions should address these constraints through exploring Mixture of Experts (MoE) architectures to improve learning efficiency (Masoudnia and Ebrahimpour, 2014), adaptive agent architectures, transfer learning methodologies (Wang et al., 2023), related regulation requirements (Han and Xi, 2020), and computationally efficient implementations.

8 Acknowledgments

This paper is partially supported by grants from the HKUST Start-up Fund (R9911), Theme-based Research Scheme grant (T45-205/21-N), the InnoHK initiative of the Innovation and Technology Commission of the Hong Kong Special Administrative Region Government, and the research funding under HKUST-DXM AI for Finance Joint Laboratory (DXM25EG01). Also supported by National Key Research and Development Program of China Grant No. 2023YFF0725100, National Science Foundation of China (NSFC) under Grant No. U22B2060, Guangdong-Hong Kong Technology Innovation Joint Funding Scheme Project No. 2024A0505040012, the Hong Kong RGC GRF Project 16213620, RIF Project R6020-19, AOE Project AoE/E-603/18, Theme-based project TRS T41-603/20R, CRF Project C2004-21G, Key Areas Special Project of Guangdong Provincial Universities 2024ZDZX1006, Guangdong Province Science and Technology Plan Project 2023A0505030011, Guangzhou municipality big data intelligence key lab, 2023A03J0012, Hong Kong ITC ITF grants MHX/078/21 and PRP/004/22FX, Zhujiang scholar program 2021JC02X170, Microsoft Research Asia Collaborative Research Grant, HKUST-Webank joint research lab and 2023 HKUST Shenzhen-Hong Kong Collaborative Innovation Institute Green Sustainability Special Fund, from Shui On Xintiandi and the InnoSpace GBA. The authors are grateful to anonymous reviewers for their efforts and insightful suggestions to improve this article.

References

Aditya Ashok, Manimaran Govindarasu, and Venkataramana Ajjarapu. 2018. Online detection of stealthy false data injection attacks in power system state estimation. *IEEE Transactions on Smart Grid*, 9(3):1636–1646.

Chi Chen, Li Zhao, Jiang Bian, Chunxiao Xing, and Tie-Yan Liu. 2019. Investment behaviors can tell what inside: Exploring stock intrinsic properties for stock trend prediction. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2376–2384.

Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, San Francisco, CA, USA. ACM.

Yinpeng Chen and et al. 2020. Dynamic relu. In *European Conference on Computer Vision*. Springer International Publishing.

L. W. Cong, K. Tang, J. Wang, and Y. Zhang. 2021. Alphaportfolio: Direct construction through deep reinforcement learning and interpretable ai.

Can Cui, Wei Wang, Meihui Zhang, Gang Chen, Zhaojing Luo, and Beng Chin Ooi. 2021. Alphaevolve: A learning framework to discover novel alphas in quantitative investment. In *Proceedings of the 2021 International Conference on Management of Data (SIGMOD '21)*, pages 2208–2216, Virtual Event, China. ACM.

Chirag Deb, Fan Zhang, Junjing Yang, Siew Eang Lee, and Kwok Wei Shah. 2017. A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*, 74:902–924.

Yujie Ding, Shuai Jia, Tianyi Ma, Bingcheng Mao, Xiuze Zhou, Liuliu Li, and Dongming Han. 2023. Integrating stock features and global information via large language models for enhanced stock return prediction. *Preprint*, arXiv:2310.05627.

Thomas H Goodwin. 1998. The information ratio. *Financial Analysts Journal*, 54(4):34–43.

Sirui Han and Chao Xi. 2020. Financial regulation as interagency competition? The saga of venture capital rule-making in China. In Douglas W. Arner, Wai Yee Wan, Andrew Godwin, Wei Shen, and Evan Gibson, editors, *Research Handbook on Asian Financial Law*. Edward Elgar Publishing.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation*, 9(8):1735–1780.

Ouyang K, Liu Y, Li S, and et al. 2024. Modal-adaptive knowledge-enhanced graph-based financial prediction from monetary policy conference calls with llm.

Zura Kakushadze. 2016. 101 formulaic alphas. *Wilmott*, 2016(84):72–81.

Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017*, pages 3146–3154, Long Beach, CA, USA. Neural Information Processing Systems Foundation.

K. J. Koa, Y. Ma, R. Ng, and T.-S. Chua. 2024. Learning to generate explainable stock predictions using self-reflective large language models. In *Proceedings of the ACM on Web Conference 2024*, pages 4304–4315.

Li L, Chang T Y, and Wang H. 2023. Multimodal gen-ai for fundamental investment research.

Jean Lee, Nicholas Stevens, and Soyeon Caren Han. 2025. Large language models in finance (finllms). *Neural Computing and Applications*.

Jinho Lee, Raehyun Kim, Seok-Won Yi, and Jaewoo Kang. 2020. Maps: Multi-agent reinforcement learning-based portfolio management system. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, IJCAI-PRICAI-2020, page 4520–4526. International Joint Conferences on Artificial Intelligence Organization.

P. Lewis and et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. In *NeurIPS*, volume 33, pages 9459–9474.

Weiping Li and Feng Mei. 2020. Asset returns in deep learning methods: An empirical analysis on sse 50 and csi 300. *Research in International Business and Finance*, 54:101291.

Xiaoming Lin, Ye Chen, Ziyu Li, and Kang He. 2019. Stock alpha mining based on genetic algorithm. Technical report, Huatai Securities Research Center.

Andrew W Lo. 2007. Where do alphas come from?: A new measure of the value of active investment management. *A new measure of the value of active investment management (May 8, 2007)*.

Junyu Luo, Zhizhuo Kou, Liming Yang, Xiao Luo, Jinsheng Huang, Zhiping Xiao, Jingshu Peng, Chengzhong Liu, Jiaming Ji, Xuanzhe Liu, Sirui Han, Ming Zhang, and Yike Guo. 2025a. Finmme: Benchmark dataset for financial multi-modal reasoning evaluation. *Preprint*, arXiv:2505.24714.

Junyu Luo, Yuhao Tang, Yiwei Fu, Xiao Luo, Zhizhuo Kou, Zhiping Xiao, Wei Ju, Wentao Zhang, and Ming Zhang. 2025b. Sparse causal discovery with generative intervention for unsupervised graph domain adaptation. *Preprint*, arXiv:2507.07621.

Junyu Luo, Weizhi Zhang, Ye Yuan, Yusheng Zhao, Junwei Yang, Yiyang Gu, Bohan Wu, Binqi Chen, Ziyue Qiao, Qingqing Long, Rongcheng Tu, Xiao Luo, Wei Ju, Zhiping Xiao, Yifan Wang, Meng Xiao, Chenwu Liu, Jingyang Yuan, Shichang Zhang, and 7 others. 2025c. Large language model agent: A survey on methodology, applications and challenges. *Preprint*, arXiv:2503.21460.

Saeed Masoudnia and Reza Ebrahimpour. 2014. Mixture of experts: a literature survey. *Artificial Intelligence Review*, 42:275–293.

T. Nathan Mundhenk, Mikel Landajuela, Ruben Glatt, Daniel M. Faissol, and Brenden K. Petersen. 2021. Symbolic regression via neural-guided genetic programming population seeding. In *9th International Conference on Learning Representations (ICLR)*, Virtual Event, Austria.

Robert Novy-Marx. 2015. Fundamentally, momentum is fundamental momentum. Technical Report w20984, NBER.

OpenAI. 2023. Gpt-4 technical report.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *Preprint*, arXiv:1707.06347.

William F Sharpe. 1994. The sharpe ratio. *Journal of portfolio management*, 21(1):49–58.

Thomas Spooner and Rahul Savani. 2020. Robust market making via adversarial reinforcement learning. *Preprint*, arXiv:2003.01820.

Y. Talebirad and A. Nadiri. 2023. Multi-agent collaboration: Harnessing the power of intelligent llm agents.

Ziyi Tang, Zechuan Chen, Jiarui Yang, Jiayao Mai, Yongsen Zheng, Keze Wang, Jinrui Chen, and Liang Lin. 2025. Alphaagent: Llm-driven alpha mining with regularized exploration to counteract alpha decay. *Preprint*, arXiv:2502.16789.

H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, A. Rodriguez, A. Joulin, E. Grave, and G. Lample. 2023. Llama: Open and efficient foundation language models.

Suppawong Tuarob, Conrad S. Tucker, Soundar R. T. Kumara, C. Lee Giles, Aaron L. Pincus, David E. Conroy, and Nilam Ram. 2017. How are you feeling: A personalized methodology for predicting mental states from temporally observable physical and behavioral information. *Journal of Biomedical Informatics*, 68:1–19.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017*, pages 5998–6008, Long Beach, CA, USA. Neural Information Processing Systems Foundation.

B. W. and et al. 2023. Bloom: A 176b-parameter open-access multilingual language model.

H. W. and et al. 2022. Scaling instruction-finetuned language models.

Liping Wang, Jiawei Li, Lifan Zhao, Zhizhuo Kou, Xiaohan Wang, Xinyi Zhu, Hao Wang, Yanyan Shen, and Lei Chen. 2023. Methods for acquiring and incorporating knowledge into stock price prediction: A survey. *arXiv preprint arXiv:2308.04947*.

Q. Wang, Z. Wang, Y. Su, and et al. 2024. Rethinking the bounds of llm reasoning: Are multi-agent discussions the key?

Pengcheng Wen, Jiaming Ji, Chi-Min Chan, Juntao Dai, Donghai Hong, Yaodong Yang, Sirui Han, and Yike Guo. 2025. Thinkpatterns-21k: A systematic study on the impact of thinking patterns in llms. *Preprint*, arXiv:2503.12918.

Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David Rosenberg, and Gideon Mann. 2023. Bloomberggpt: A large language model for finance. *Preprint*, arXiv:2303.17564.

Wentao Xu, Weiqing Liu, Lewen Wang, Yingce Xia, Jiang Bian, Jian Yin, and Tie-Yan Liu. 2021a. Hist: A graph-based framework for stock trend forecasting via mining concept-oriented shared information.

Wentao Xu, Weiqing Liu, Chang Xu, Jiang Bian, Jian Yin, and Tie-Yan Liu. 2021b. Rest: Relational event-driven stock trend forecasting. In *WWW '21: The Web Conference 2021*, pages 1–10, Virtual Event / Ljubljana, Slovenia. ACM / IW3C2.

Yumo Xu and Shay B Cohen. 2018. Stock movement prediction from tweets and historical prices. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1970–1979.

Y. Yang, Y. Tang, and K. Y. Tam. 2023. Investlm: A large language model for investment using financial domain instruction tuning.

Shuo Yu, Hongyan Xue, Xiang Ao, Feiyang Pan, Jia He, Dandan Tu, and Qing He. 2023a. Generating synergistic formulaic alpha collections via reinforcement learning. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '23)*, pages 5476–5486, New York, NY, USA. Association for Computing Machinery.

Yangyang Yu and et al. 2024. Finmem: A performance-enhanced llm trading agent with layered memory and character design. In *Proceedings of the AAAI Symposium Series*, volume 3.

Yangyang Yu, Haohang Li, Zhi Chen, Yuechen Jiang, Yang Li, Denghui Zhang, Rong Liu, Jordan W. Suchow, and Khaldoun Khashanah. 2023b. Finmem: A performance-enhanced llm trading agent with layered memory and character design. *Preprint*, arXiv:2311.13743.

Yangyang Yu, Zhiyuan Yao, Haohang Li, Zhiyang Deng, Yupeng Cao, Zhi Chen, Jordan W. Suchow, Rong Liu, Zhenyu Cui, Zhaozhuo Xu, Denghui Zhang, Koduvayur Subbalakshmi, Guojun Xiong, Yueru He, Jimin Huang, Dong Li, and Qianqian Xie. 2024. Fincon: A synthesized llm multi-agent system with conceptual verbal reinforcement for enhanced financial decision making. *Preprint*, arXiv:2407.06567.

Chong Zhang, Xinyi Liu, Zhongmou Zhang, Mingyu Jin, Lingyao Li, Zhenting Wang, Wenyue Hua, Dong Shu, Suiyuan Zhu, Xiaobo Jin, Sujian Li, Mengnan Du, and Yongfeng Zhang. 2024a. When ai meets finance (stockagent): Large language model-based stock trading in simulated real-world environments. *Preprint*, arXiv:2407.18957.

Feng Zhang, Ruite Guo, and Honggao Cao. 2020. Information coefficient as a performance measure of stock selection models.

W. Zhang, L. Zhao, H. Xia, and et al. 2024b. A multimodal foundation agent for financial trading: Tool-augmented, diversified, and generalist. Technical report.

H. Zhao, Z. Liu, Z. Wu, and et al. 2024. Revolutionizing finance with llms: An overview of applications and insights.

W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, and et al. 2023. A survey of large language models.

A Appendix

A.1 Sample of Seed Alpha

The seed alpha representation depicted in Figure 4 illustrates a fundamental construct in quantitative finance for generating trading signals. This figure presents a comprehensive visualization of the Detrended Price Oscillator (DPO) formula, which compares the current closing price with a delayed simple moving average (SMA) of closing prices. Panel A shows the mathematical formulation of the seed alpha, expressed as `CLOSE - DELAY(SMA(CLOSE, 14), 7)`, which captures price momentum by measuring deviations from historical trends. Panel B transforms this formula into an equivalent expression tree, demonstrating the hierarchical relationship between operators and operands, which facilitates algorithmic implementation and analysis. Panel C provides a practical illustration through a tabulated step-by-step computation of this alpha formula on a sample time series, showing how the signal evolves over multiple trading days. This representation exemplifies how complex financial indicators can be systematically decomposed, formalized, and applied to market data for quantitative trading strategies.

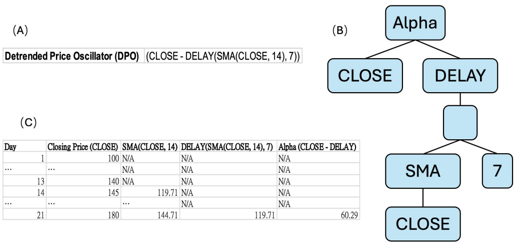

Figure 4: Seed Alpha Representation: (A) An example of the seed alpha formula. (B) Its equivalent expression tree. (C) Step-by-step computation of this seed alpha on an example time series.

A.2 Seed Alphas document lists for GPTs

The presented table offers a comprehensive overview of cutting-edge research pertinent to the development of a SAF in quantitative finance. These documents collectively represent the confluence of traditional financial methodologies with advanced computational techniques. Of particular significance is Kakushadze's "101 Formulaic Alphas," which provides a foundational repository of trading signals that can be categorized according to traditional financial factors. Lopez de Prado's work on "Causal Factor Investing" introduces scientific rigor to factor classification, ensuring independence between alpha categories. The integration of artificial intelligence is evident in multiple studies, including "Alpha-GPT" and "FinGPT," which leverage LLMs for alpha generation. Portfolio construction methodologies are addressed through reinforcement learning approaches in "AlphaPortfolio" and "Dynamic Graph-based Deep Reinforcement Learning." Performance enhancement strategies are explored in "Mastering Stock Markets with Efficient Mixture of Diversified Trading Experts," while market dependency analysis is covered in "Model-Free Implied Dependence and the Cross-Section of Returns." This literature collection provides quantitative researchers with the theoretical frameworks and methodological tools necessary to construct a Seed Alpha Factory that systematically generates, categorizes, and implements trading signals across independent financial categories, balancing traditional financial theory with contemporary computational advancements.

Sample prompt "Summarize the document information to help quantitative researchers build the Seed Alpha Factory according to traditional financial categories, ensuring that each category of seed alphas is independent."

A.3 Generate Seed Alpha factory

This taxonomy presents quantitative trading signals organized into eight categories: Momentum, Mean Reversion, Volatility, Fundamental, Liquidity, Quality, Growth, Technical, and Macroeconomic indicators. Each category targets distinct market phenomena with specific mathematical implementations. Momentum factors identify persistent trends, Mean Reversion signals detect market overreactions, Volatility metrics quantify price dispersion, and Fundamental factors evaluate company valuations. The framework also includes Liquidity measures of trading activity, Quality indicators of operational efficiency, Growth metrics of financial expansion, Technical indicators derived from price-volume patterns, and Macroeconomic signals reflecting broader economic conditions. The mathematical formulations provided enable researchers to implement diverse, uncorrelated alpha factors for robust quantitative trading strategies.

Category	Name	Short Code
Momentum	Price Momentum	(CLOSE - DELAY(CLOSE, 14))
	Volume Momentum	(VOLUME - DELAY(VOLUME, 14))
	RSI Momentum	(RSI - DELAY(RSI, 14))
	Rate of Change (ROC)	((CLOSE / DELAY(CLOSE, 14)) - 1)
	MACD Momentum	(MACD - DELAY(MACD, 14))
	Momentum Oscillator	((CLOSE - DELAY(CLOSE, 14)) / DELAY(CLOSE, 14))
	Chande Momentum Oscillator (CMO)	(SUM(IF(CLOSE - DELAY(CLOSE, 1) > 0, CLOSE - DELAY(CLOSE, 1), 0), 14) - SUM(IF(CLOSE - DELAY(CLOSE, 1) < 0, DELAY(CLOSE, 1) - CLOSE, 0), 14)) / (SUM(IF(CLOSE - DELAY(CLOSE, 1) > 0, CLOSE - DELAY(CLOSE, 1), 0), 14) + SUM(IF(CLOSE - DELAY(CLOSE, 1) < 0, DELAY(CLOSE, 1) - CLOSE, 0), 14)) * 100
	Stochastic Momentum Index (SMI)	((CLOSE - MIN(LOW, 14)) - (MAX(HIGH, 14) - CLOSE)) / (MAX(HIGH, 14) - MIN(LOW, 14))
	ATR Momentum	(ATR - DELAY(ATR, 14))
	Detrended Price Oscillator (DPO)	(CLOSE - DELAY(SMA(CLOSE, 14), 7))
	Average Directional Index (ADX) Momentum	(ADX - DELAY(ADX, 14))
Mean Reversion	Mean Reversion	(MEAN(CLOSE, 20) - CLOSE)
	Z-Score Mean Reversion	(CLOSE - MEAN(CLOSE, 20)) / STD(CLOSE, 20)
	Bollinger Bands	(CLOSE - LOWER_BAND) / (UPPER_BAND - LOWER_BAND)
	Keltner Channel	(CLOSE - LOWER_CHANNEL) / (UPPER_CHANNEL - LOWER_CHANNEL)
	Moving Average Reversion	(SMA(CLOSE, 20) - CLOSE)
	Exponential Moving Average (EMA) Reversion	(EMA(CLOSE, 20) - CLOSE)
	Distance from High	(MAX(HIGH, 20) - CLOSE)
	Distance from Low	(CLOSE - MIN(LOW, 20))
	Relative Strength Index (RSI) Reversion	(100 - RSI)
	Percent B	((CLOSE - LOWER_BAND) / (UPPER_BAND - LOWER_BAND)) * 100

Volatility	Standard Deviation	STD(CLOSE, 20)
	Average True Range (ATR)	ATR(14)
	Bollinger Band Width	(UPPER_BAND - LOWER_BAND) / SMA(CLOSE, 20)
	Historical Volatility	STD(RETURNS, 20) * SQRT(252)
	Volatility Ratio	STD(CLOSE, 10) / STD(CLOSE, 50)
	Chaikin Volatility	(EMA(HIGH - LOW, 10) / DELAY(EMA(HIGH - LOW, 10), 10)) - 1
	Garman-Klass Volatility	SQRT(0.5 * LOG(HIGH / LOW)^2 - (2 * LOG(2) - 1) * LOG(CLOSE / OPEN)^2)
	Parkinson Volatility	SQRT((1 / (4 * N * LOG(2))) * SUM(LOG(HIGH / LOW)^2, 20))
	Yang-Zhang Volatility	SQRT(VAR(LOG(CLOSE / OPEN)) + 0.5 * VAR(LOG(HIGH / OPEN) - LOG(LOW / OPEN)) + 0.25 * VAR(LOG(CLOSE / DELAY(OPEN, 1))))
	Ulcer Index	SQRT(MEAN(DRAWDOWN^2, 14))
Fundamental	Price-to-Earnings Ratio (P/E)	(CLOSE / EPS)
	Price-to-Book Ratio (P/B)	(CLOSE / BOOK_VALUE)
	Dividend Yield	(DIVIDENDS / CLOSE)
	Earnings Yield	(EPS / CLOSE)
	Sales-to-Price Ratio	(SALES / CLOSE)
	Cash Flow Yield	(OPERATING_CASH_FLOW / CLOSE)
Liquidity	Trading Volume	VOLUME
	Average Trading Volume	MEAN(VOLUME, 20)
	Volume Rate of Change (VROC)	(VOLUME - DELAY(VOLUME, 14)) / DELAY(VOLUME, 14)
	On-Balance Volume (OBV)	SUM(VOLUME * SIGN(CLOSE - DELAY(CLOSE, 1)))
	Liquidity Ratio	VOLUME / MARKET_CAP
	Turnover Rate	VOLUME / SHARES_OUTSTANDING
	Amihud Illiquidity Ratio	ABS(RETURN) / VOLUME
	High-Low Spread	(HIGH - LOW) / CLOSE
	Dollar Volume	VOLUME * CLOSE
	Debt-to-Equity Ratio	(TOTAL_DEBT / TOTAL_EQUITY)
	Return on Equity (ROE)	(NET_INCOME / EQUITY)
	Return on Assets (ROA)	(NET_INCOME / TOTAL_ASSETS)
	Gross Profit Margin	(GROSS_PROFIT / REVENUE)
	Price-to-Sales Ratio (P/S)	(CLOSE / SALES)
	Price-to-Cash Flow Ratio (P/CF)	(CLOSE / OPERATING_CASH_FLOW)
	Book-to-Market Ratio (B/M)	(BOOK_VALUE / CLOSE)
	Enterprise Value to EBITDA (EV/EBITDA)	(ENTERPRISE_VALUE / EBITDA)
	Bid-Ask Spread	(ASK_PRICE - BID_PRICE) / MID_PRICE
	High-Low Spread	(HIGH - LOW) / CLOSE
	Dollar Volume	VOLUME * CLOSE
Quality	Gross Profit Margin	(GROSS_PROFIT / REVENUE)
	Operating Profit Margin	(OPERATING_INCOME / REVENUE)
	Net Profit Margin	(NET_INCOME / REVENUE)
	Earnings Stability	STD(EPS, 5) / MEAN(EPS, 5)
	Debt to Equity Ratio	(TOTAL_DEBT / TOTAL_EQUITY)
	Interest Coverage Ratio	(EBIT / INTEREST_EXPENSE)
	Cash Conversion Cycle	(DIO + DSO - DPO)
	Asset Turnover Ratio	(REVENUE / TOTAL_ASSETS)
Growth	Earnings Growth Rate	(EPS / DELAY(EPS, 1) - 1)
	Revenue Growth Rate	(REVENUE / DELAY(REVENUE, 1) - 1)
	EBITDA Growth Rate	(EBITDA / DELAY(EBITDA, 1) - 1)

	Cash Flow Growth Rate	(CASH_FLOW / DELAY(CASH_FLOW, 1) - 1)
	Dividends Growth Rate	(DIVIDENDS / DELAY(DIVIDENDS, 1) - 1)
	Book Value Growth Rate	(BOOK_VALUE / DELAY(BOOK_VALUE, 1) - 1)
	Sales Growth Rate	(SALES / DELAY(SALES, 1) - 1)
	Asset Growth Rate	(ASSETS / DELAY(ASSETS, 1) - 1)
	Equity Growth Rate	(EQUITY / DELAY(EQUITY, 1) - 1)
	Retained Earnings Growth Rate	(RETAINED_EARNINGS / DELAY(RETAINED_EARNINGS, 1) - 1)
Technical	Moving Average (MA)	SMA(CLOSE, 20)
	Exponential Moving Average (EMA)	EMA(CLOSE, 20)
	Relative Strength Index (RSI)	RSI(14)
	Moving Average Convergence Divergence (MACD)	(EMA(CLOSE, 12) - EMA(CLOSE, 26))
	Bollinger Bands	UPPER_BAND - LOWER_BAND
	Stochastic Oscillator	((CLOSE - MIN(LOW, 14)) / (MAX(HIGH, 14) - MIN(LOW, 14))) * 100
	Average True Range (ATR)	ATR(14)
	Commodity Channel Index (CCI)	(TYPICAL_PRICE - SMA(TYPICAL_PRICE, 20)) / (0.015 * MEAN_DEV(TYPICAL_PRICE, 20))
	Williams %R	((MAX(HIGH, 14) - CLOSE) / (MAX(HIGH, 14) - MIN(LOW, 14))) * -100
Macro Economics	GDP Growth Rate	GDP - DELAY(GDP, n)
	Inflation Rate	CPI - DELAY(CPI, n)
	Unemployment Rate	UNEMPLOYMENT_RATE - DELAY(UNEMPLOYMENT_RATE, n)
	Interest Rate	INTEREST_RATE - DELAY(INTEREST_RATE, n)
	Industrial Production Index	IPI - DELAY(IPI, n)
	Retail Sales Growth	RETAIL_SALES - DELAY(RETAIL_SALES, n)
	Housing Starts Growth	HOUSING_STARTS - DELAY(HOUSING_STARTS, n)
	Consumer Confidence Index (CCI)	CCI - DELAY(CCI, n)
	Trade Balance	EXPORTS - IMPORTS
	Foreign Exchange Reserves	FX_RESERVES - DELAY(FX_RESERVES, n)

A.4 Multimodal Data types

Table 11 presents a taxonomy of multimodal data types essential for comprehensive quantitative finance research. The classification encompasses five categories: textual data (financial reports, news articles, social media discourse), numerical data (historical price series, returns, volatility metrics), visual data (charts and technical analysis patterns), audio data (financial broadcasts and market commentary), and video data (specialized financial news channels). This multimodal framework enables researchers to develop more robust predictive models by integrating complementary information channels, potentially identifying market inefficiencies that remain undetectable when analyzing isolated data types. By synthesizing diverse information formats, quantitative analysts can gain multidimensional insights into market dynamics, enhancing both analytical depth and predictive capability in financial modeling applications.

Table 11: Multimodal data types

Data Type	Description	Examples
Textual Data	Financial reports, academic papers, news articles, and other textual documents.	Trading forums' sentiment analysis and stock predictions, company disclosures, financial statements, Sina Finance
Numerical Data	Historical stock market data, financial metrics, and performance indicators.	Returns, log returns, annualized returns, volatility
Visual Data	Charts, graphs, and other visual representations of financial data.	Kline charts, trading charts
Audio Data	Financial news broadcasts.	Financial morning news radio, stock review radio, market discussion radio
Video Data	Financial news channels.	CCTV Securities Information Channel, CCTV News Broadcast (news affecting China's stock market)

A.5 Category-Based Alpha Selection

Algorithm 1 delineates a systematic methodology for alpha selection in quantitative investment strategies, employing a category-based approach to ensure diversification across multiple financial factors. The procedure operates on a structured set of alpha categories $\mathcal{C}$, each representing distinct market phenomena such as momentum, mean reversion, or volatility. For each category, the algorithm identifies superior alpha factors through the SelectBestAlphas function, which presumably evaluates historical performance metrics. The innovation lies in the subsequent dual-agent evaluation system: the RiskPreferenceAgent assesses each alpha's risk characteristics, while the ConfidenceScoreAgent evaluates the statistical robustness of its historical performance. These complementary evaluations are synthesized using weight parameters w_r and w_c to compute a comprehensive Final Score. Only alphas exceeding a predefined confidence threshold X are incorporated into the final selection set $\mathcal{A}$. This methodical approach ensures that the resulting alpha portfolio exhibits both category diversification and individual signal quality, potentially enhancing risk-adjusted returns while mitigating exposure to specific market regimes or factor deterioration.

Algorithm 1 Category-Based Alpha Selection

Input: Categories $\mathcal{C} = \{C_1, \ldots, C_m\}$, each C_i containing a set of alphas; confidence threshold X

Initialize selected alphas $\mathcal{A} \leftarrow \emptyset$
for each category $C_i \in \mathcal{C}$ **do**
    $\mathcal{A}_i \leftarrow$ SelectBestAlphas(C_i)
    for each $\alpha \in \mathcal{A}_i$ **do**
        risk_score $\leftarrow$ RiskPreferenceAgent(α)
        confidence_score $\leftarrow$ ConfidenceScoreAgent(α)
        final_score $\leftarrow$ w_r · risk_score + w_c · confidence_score
        if final_score $> X$ **then**
            $\mathcal{A} \leftarrow \mathcal{A} \cup \{\alpha\}$
        end if
    end for
end for
return $\mathcal{A}$

A.6 Dynamic Alpha Strategy Construction

Algorithm 2 Overall Framework Algorithm

Input: Multimodal financial data $\mathcal{D}$, market conditions $\mathcal{M}^{(t)}$, stocks $\mathcal{S}$, confidence threshold τ
Output: Optimized alpha strategy $\alpha^{(t)}$

/* Phase 1: SAF */
Initialize empty Seed Alpha Factory $\mathcal{F} \leftarrow \emptyset$
for each document $d \in \mathcal{D}$ **do**
    filtered_content $\leftarrow$ LLM.Filter(d)
    categories $\leftarrow$ LLM.Categorize(filtered_content)
    for each category $c \in$ categories **do**
        seed_alphas $\leftarrow$ LLM.GenerateAlphas(c)
        $\mathcal{F}_c \leftarrow \mathcal{F}_c \cup$ seed_alphas
    end for
end for
/* Phase 2: Multi-Agent Evaluation */
Initialize selected alphas $\mathcal{A} \leftarrow \emptyset$
for each category $c \in \mathcal{F}$ **do**
    for each alpha $\alpha_{ij} \in \mathcal{F}_c$ **do**
        $\theta_{ij} \leftarrow$ ConfidenceScoreAgent($\alpha_{ij}, \mathcal{M}^{(t)}$) ▷ Statistical evaluation
        $\rho_{ij} \leftarrow$ RiskPreferenceAgent($\alpha_{ij}, \mathcal{M}^{(t)}$) ▷ Risk assessment
        score$_{ij} \leftarrow w_c \cdot \theta_{ij} + w_r \cdot \rho_{ij}$ ▷ Combined score
        if score$_{ij} > \tau$ **then**
            $\mathcal{A}_c \leftarrow \mathcal{A}_c \cup \{\alpha_{ij}\}$
        end if
    end for
    if $\mathcal{A}_c \neq \emptyset$ **then**
        $\alpha_c^* \leftarrow \arg\max_{\alpha_{ij} \in \mathcal{A}_c}$ score$_{ij}$ ▷ Select best alpha in category
        $\mathcal{A} \leftarrow \mathcal{A} \cup \{\alpha_c^*\}$
    end if
end for
/* Phase 3: Weight Optimization */
Initialize MLP with architecture $\{|\mathcal{A}|, 10, 1\}$ ▷ Input, hidden, output layers
$\mathbf{X}_{\text{train}} \leftarrow$ ComputeHistoricalAlphas($\mathcal{A}$, $\mathcal{S}$, t_{start}, t_{end})
$\mathbf{y}_{\text{train}} \leftarrow$ FutureReturns($\mathcal{S}$, $t_{\text{start}} + 1$, $t_{\text{end}} + 1$)
$\mathbf{w} \leftarrow$ TrainMLP($\mathbf{X}_{\text{train}}$, $\mathbf{y}_{\text{train}}$) ▷ Learn optimal weights
for each stock $S_i \in \mathcal{S}$ **do**
    $\alpha_i^{(t)} \leftarrow \sum_{j=1}^{|\mathcal{A}|} w_j \cdot \alpha_{ij}^{(t)}$ ▷ Compute composite alpha
end for
return $\alpha^{(t)} = \{\alpha_1^{(t)}, \alpha_2^{(t)}, \ldots, \alpha_n^{(t)}\}$ ▷ Final alpha strategy

A.7 Sample prompt for Seed Alpha selection

1. The training set includes:

- **Financial Reports:** 6 quarters (from Jan 2021 to Jun 2022) for 50 companies listed in the SSE 50.
- **Factor Analysis Data:** 37 factors (from Jan 2021, to Jun 2022) divided into groups: Momentum, Mean Reversion, Volatility, Fundamental, Quality, Growth, Technical, Macro Economics. The metric used is the IC.

2. **Objective:** Learn the relationship between the performance of financial reports for the first four quarters from 2022 to 2023 and the factor analysis data (IC) for each of the 37 factors in the last quarter of 2022.

3. **When provided with the test set (performance of the first 4 quarters of 2023):**

- Select the factors that will perform best in the last quarter of the SSE 50.
- Provide a confidence score & the risk preference for your selection for each selected Alpha factor.

4. **Selection Criteria:**

- If no relationship between financial reports and IC can be found, select the Alpha factor with the highest IC value in each group.
- For verification of market information differences, if no relationship between financial reports and IR can be found, select the Alpha factor with the highest IR value.

A.8 Portfolio Construction Methods

Our investment methodology implements a daily portfolio reconstruction using a top-k/drop-n selection framework. At each trading session, we rank all securities by their alpha values—quantitative indicators of expected excess returns—and select the top k stocks for portfolio inclusion. This approach targets securities with the strongest signals, potentially exploiting short-term market inefficiencies. We employ an equal-weighting scheme across selected securities, distributing capital homogeneously among the top candidates, which offers diversification benefits and aligns with our view that alpha signals primarily provide directional indications rather than precise return forecasts.

To optimize transaction costs and operational efficiency, we limit portfolio turnover to a maximum of n securities per trading day. This constraint balances maintaining alignment with current alpha signals while minimizing trading friction costs. For our experiment in Section 4.4, we set $k = 13$ and $n = 5$ based on extensive backtesting. This configuration establishes a portfolio concentration that balances diversification against signal dilution, while limiting daily turnover to approximately 38%—a level that demonstrates favorable characteristics in our transaction cost modeling and represents an efficient trade-off between signal utilization and implementation costs.