MEME: MODELING THE EVOLUTIONARY MODES OF FINANCIAL MARKETS

Anonymous authors
Paper under double-blind review

ABSTRACT

LLMs have demonstrated significant potential in quantitative finance by processing vast unstructured data to emulate human-like analytical workflows. However, current LLM-based methods primarily follow either an *Asset-Centric* paradigm focused on individual stock prediction or a *Market-Centric* approach for portfolio allocation, often remaining agnostic to the underlying reasoning that drives market movements. In this paper, we propose a *Logic-Oriented* perspective grounded in the theory of Narrative Economics, modeling the financial market as a dynamic ecosystem of competing investment narratives, termed **Modes of Thought**. To operationalize this view, we introduce MEME (Modeling the Evolutionary Modes of Financial Markets). MEME extracts structured Investment Arguments from noisy multimodal data via multi-agent collaboration, uncovers latent consensus through probabilistic clustering, and tracks the lifecycle of these modes via temporal alignment and profitability evaluation. Extensive experiments on four heterogeneous stock pools from 2023 to 2025 demonstrate that MEME consistently outperforms seven SOTA baselines. Further ablation studies, sensitivity analysis, and a lifecycle case study validate MEME's capacity to identify evolving market consensus, with the recovered modes showing correspondence to established investment philosophies. Our implementation can be found at `https://anonymous.4open.science/r/MEME-FA3C`.

1 INTRODUCTION

LLMs (Team et al., 2025a; Liu et al., 2025; Yang et al., 2025; Team et al., 2025b) are increasingly being applied in quantitative finance (Blyth, 2014; McNeil et al., 2015). By processing vast amounts of unstructured data such as financial reports and social media discussions, multi-agent systems powered by LLMs can emulate the complex workflows of human analysts to distill insights and inform investment decisions. Current LLM-based methods can be classified into two categories. The Asset-Centric category focuses on predicting individual stock performance through forecasting (Koa et al., 2024; Zhang et al., 2024a) or Alpha Factor Mining (Tang et al., 2025; Shi et al., 2025b). The Market-Centric category (Tong et al., 2024; Guo et al., 2025; Li et al., 2025b) shifts focus to portfolio-level allocation. While effective, both paradigms are fundamentally object-oriented: they model "what" to buy or "how much" to allocate, while remaining agnostic to the underlying "why" that drives market movements.

We approach the problem from a different angle. The theory of Narrative Economics (Shiller, 2020) establishes that contagious narratives spread among market participants and drive economic fluctuations. The investor attention literature (Barber & Odean, 2008) further shows that the salience of information shapes trading behavior. Building on these insights, we adopt a *Logic-Oriented* perspective: we model the market as a dynamic ecosystem of competing investment logics, which we term **Modes of Thought**. In this view, price movements are manifestations of trading narratives gaining or losing consensus. Rather than predicting returns from features or optimizing portfolio weights, we aim to identify which investment logics are currently effective and let them guide portfolio construction.

Operationalizing this perspective requires addressing three challenges. First, raw multimodal data (fundamentals, news, technical indicators) is noisy and unstructured. Extracting coherent reasoning from such data demands a systematic filtering pipeline. Second, individual arguments are sparse and expressed in varied terminology. Uncovering the shared logic they represent requires aggregating

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

them into latent consensus patterns while allowing each argument to participate in multiple themes. Third, financial markets are non-stationary. Investment narratives emerge, gain consensus, and decay over time, and many sound plausible but fail to generate returns. A framework must track this evolution and distinguish historically validated logics from ineffective ones.

To address these challenges, we introduce MEME (Modeling the Evolutionary Modes of Financial Markets). MEME first employs a multi-agent system to distill raw data into structured Investment Arguments. It then identifies latent Modes of Thought through probabilistic clustering in a semantic space. Finally, it tracks these modes across time via temporal alignment and evaluates their historical profitability, ensuring that portfolio construction is guided by robust, historically validated reasoning rather than transient signals.

We evaluate MEME on four stock pools across the Chinese and U.S. markets from 2023 to 2025, comparing against 7 baselines spanning traditional ML, deep learning, and LLM-based methods. MEME achieves consistent improvements in both predictive accuracy and portfolio performance. We further provide ablation studies verifying each component, a sensitivity analysis, a signal orthogonality analysis confirming that MEME provides complementary information to price-volume models, and an out-of-sample test against data leakage. A lifecycle case study shows that the unsupervised modes recovered by MEME resemble established investment philosophies such as trend-following (Lefevre & Markman, 2010) and value investing (Graham & Zweig, 2003), suggesting that the identified Modes of Thought reflect reasoning patterns with real-world relevance.

In summary, our main contributions are as follows:

- We introduce MEME, a framework that models market dynamics as an evolutionary process of competing Modes of Thought, shifting the focus from assets to the lifecycle of investment logics.
- We design an end-to-end system that extracts structured arguments, identifies latent consensus via probabilistic modeling, and tracks the lifecycle of these modes through temporal alignment.
- We conduct extensive experiments on four stock pools with distinct styles across the Chinese and U.S. markets, achieving consistent improvements over seven SOTA baselines and providing an in-depth analysis of the identified market modes.

2 Related Work

This section reviews literature on financial trading. We categorize existing studies into asset-centric and market-centric approaches and discuss the theoretical foundation of our logic-oriented perspective.

2.1 Asset-Centric Approaches

Asset-centric research focuses on the micro-level forecasting of individual assets and the mining of predictive signals. This field includes traditional quantitative studies such as formulaic alpha factor mining and stock trend prediction (Blyth, 2014; McNeil et al., 2015). Researchers use genetic programming (Chen et al., 2021), reinforcement learning (Yu et al., 2023; Zhu & Zhu, 2025), and flow networks (Chen et al., 2025) to generate predictive expressions. Other studies employ factor modeling (Duan et al., 2022; Xiang et al., 2024; Duan et al., 2025; Shi et al., 2025a) and temporal dependency analysis (Yoo et al., 2021; Li et al., 2024b; Zhang et al., 2025a). Advanced models also integrate entity dependencies (Zhao et al., 2022; Xu et al., 2021; Qian et al., 2024; Song et al., 2025), causality modeling (Luo et al., 2023; Li et al., 2024a), and sentiment features (Araci, 2019; Du et al., 2024) or financial foundation models (Shi et al., 2025c). With the development of LLMs, agent-based methods have emerged. These systems perform multi-modal analysis (Zhang et al., 2024a), engage in expert debates (Xiao et al., 2025a), and use in-context self-reflection (Koa et al., 2024). Some frameworks also apply task-specific fine-tuning (Tian et al., 2025; Deng & Wang, 2025). In alpha factor mining, agents utilize iterative refinement (Li et al., 2024c; Tang et al., 2025; Cao et al., 2025) and navigated tree searching (Shi et al., 2025b) to find effective formulas.

Our method diverges from these asset-centric works by shifting the analytical target from micro-level mapping to macro-level reasoning. While traditional approaches treat assets as independent units and attempt to map noisy features directly to price movements, we posit that individual asset fluctuations are often idiosyncratic. Instead of predicting “what" an asset will do based on its internal signals, MEME extracts the underlying Investment Arguments that explain “why" a group of assets moves in tandem, thereby achieving higher reasoning purity by filtering out asset-specific noise.

2.2 MARKET-CENTRIC APPROACHES

Market-centric approaches examine the financial problem at the stock pool or portfolio level. Traditional research focuses on portfolio construction through inter-stock covariance matrix prediction (Zhu et al., 2022; Zheng et al., 2023) and multi-factor dimension reduction (Lin et al., 2021). Some studies apply reinforcement learning for weight optimization across a collection of assets (Zhang et al., 2024b). Recently, LLM-based systems have been used to synthesize information across the stock pool. For example, some works use diverse expert discussions to analyze constituent stocks for portfolio construction (Tong et al., 2024). Other methods employ multi-agent collaboration (Guo et al., 2025), investor behavior simulation (YANG et al., 2025), order-level simulation (Li et al., 2025a) or model-factor collaboration (Li et al., 2025b) to manage market interactions.

Our framework departs from these market-centric approaches in both perspective and methodology. First, while existing methods are primarily object-oriented as they treat the stock pool as a collection of assets to be weighted, we adopt a logic-oriented perspective that views the market as a dynamic competition of underlying narratives. Second, we shift the focus from spatial allocation to temporal evolution. Instead of merely optimizing asset weights for a static market snapshot, MEME models the complete lifecycle of Modes of Thought. By tracking how investment logics emerge, gain consensus, and decay through temporal alignment, we ensure that portfolio construction is guided by the historical robustness of a narrative rather than transient correlations.

Our logic-oriented perspective has deep roots in behavioral finance. Shiller's theory of Narrative Economics (Shiller, 2020) argues that contagious narratives spread among market participants and drive economic fluctuations. The investor attention literature (Barber & Odean, 2008) further shows that the salience and framing of information shape trading behavior. MEME operationalizes these insights within a quantitative pipeline: rather than modeling narratives as exogenous qualitative phenomena, we extract structured arguments from multimodal data, cluster them into latent Modes of Thought, and track their lifecycle to evaluate which narratives translate into realized returns.

3 METHODOLOGY

MEME is designed to construct daily portfolios from multi-modal financial data. Unlike traditional asset-centric or market-centric methods, MEME is logic-oriented. We focus on the "why" behind price movements by modeling the market as an evolving ecosystem of investment narratives. The process begins with raw daily data for a universe of N stocks over a period of T trading days. The input for a given stock s on day t is a data point $\mathcal{D}_{t,s}$, comprising information from a set of modalities $M = \{\text{Fundamental}, \text{News}, \text{Technical}\}$:

$$\mathcal{D}_{t,s} = \{\mathcal{D}_{t,s,m} \mid m \in M\}$$

where $\mathcal{D}_{t,s,m}$ is for stock s on day t from modality m. The final output is a daily portfolio weight vector $\mathbf{W}_t \in \mathbb{R}^N$.

MEME operationalizes this process through three sequential stages. First, we extract structured investment arguments from noisy multimodal data. Second, we identify latent modes of reasoning by clustering these arguments and tracking their evolution over time. Third, we evaluate the historical profitability of each mode to construct a final portfolio. The overall architecture of MEME is provided in Figure 1, and the pseudocode is available in Appendix 1.

3.1 ARGUMENT EXTRACTION: FROM NOISE TO STRUCTURE

The initial stage of MEME transforms unstructured, multi-modal financial data into structured records, which we term Investment Arguments. This stage is designed to overcome the structural chaos and high noise density in raw data. To manage the low signal-to-noise ratio inherent in raw data, we employ a multi-agent system that filters relevant information from each data modality and then generates structured arguments, each containing a directional stance, a rationale, and supporting evidence. This collaborative design ensures the purity of the reasoning by separating informative signals from irrelevant noise. This process is implemented through two steps:

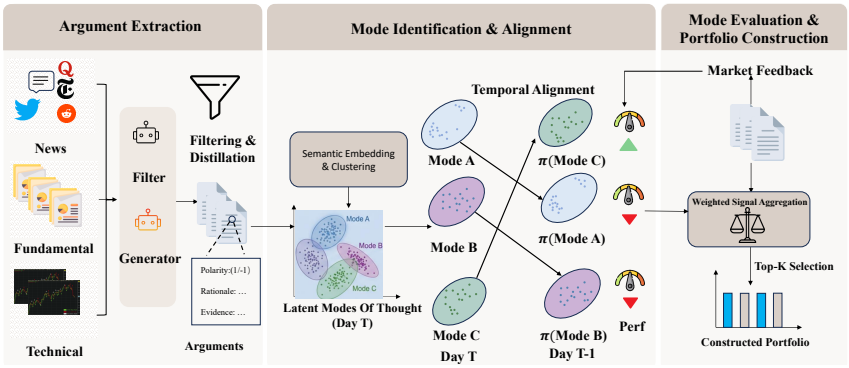


Figure 1: The overall architecture. MEME first mitigates structural noise by distilling raw multimodal data into structured Investment Arguments. To uncover latent consensus, it identifies Modes of Thought through probabilistic clustering and models market evolution via temporal alignment of these modes across consecutive days. Finally, MEME filters logical noise and construct the portfolio by evaluating the historical profitability of each mode through an alignment-based scoring mechanism.

Information Filtering. For each modality $m \in M$, a filter agent, $\mathcal{F}_{\text{filter},m}$, processes the raw data $\mathcal{D}_{t,s,m}$ to extract a concise set of relevant information, $\mathcal{I}_{t,s,m}$.

$$\mathcal{I}_{t,s,m} = \mathcal{F}_{\text{filter},m}(\mathcal{D}_{t,s,m}) \tag{1}$$

Argument Generation. The filtered information from all modalities is consolidated. A generator agent, $\mathcal{F}_{\text{gen}}$, then synthesizes this unified information set to a series of structured investment arguments, $\{\mathcal{A}_{t,s,i}\}_i$.

$$\{\mathcal{A}_{t,s,i}\}_i = \mathcal{F}_{\text{gen}}\left(\bigcup_{m \in M} \mathcal{I}_{t,s,m}\right) \tag{2}$$

Each argument $\mathcal{A}_{t,s,i}$ is a tuple containing a polarity p, an analytical rationale a, and its supporting evidence e:

$$\mathcal{A}_{t,s,i} = (p_{t,s,i}, a_{t,s,i}, e_{t,s,i}) \tag{3}$$

where $p_{t,s,i} \in \{+1, -1\}$ indicates the argument's directional stance.The design details are in the Appendix A.10.

3.2 Mode Identification and Alignment

After extracting individual arguments, we identify and track the underlying consensus they represent. This step is motivated by the observation from Narrative Economics that market dynamics are driven by the competition among investment narratives rather than isolated signals. Isolated arguments are often sparse and may use varied terminology for similar concepts. Besides, the focus of investors may shift over time. Therefore, this stage first identifies latent modes of reasoning by clustering arguments in a semantic space on a given day. Subsequently, it aligns these modes across consecutive days to model their temporal evolution.

3.2.1 Mode Identification: From Arguments to Latent Consensus

We model the shared logic among arguments by identifying latent modes within a semantic space. We use a probabilistic approach because market themes often overlap. A single argument might reflect multiple themes simultaneously. To implement this, we apply a Gaussian Mixture Model (GMM) (Duda & Hart, 1973; Dempster et al., 1977) to the collection of daily argument embeddings. This process yields a set of modes, each defined by a Gaussian distribution, and assigns each argument a membership probability to every mode. This captures the nuanced reality of market consensus better than rigid classification.

Formally, we generate a semantic representation for each argument on a given day t. We concatenate the argument's rationale $a_{t,s,i}$ and evidence $e_{t,s,i}$, then encode the result into a D-dimensional vector $\mathbf{x}_{t,s,i}$ using an encoder $\mathcal{E}$:

$$\mathbf{x}_{t,s,i} = \mathcal{E}(\text{concat}(a_{t,s,i}, e_{t,s,i})) \in \mathbb{R}^D \tag{4}$$

We then apply a GMM to the collection of all daily argument embeddings $\{\mathbf{x}_{t,s,i}\}_{s,i}$. This defines K modes for day t, where each mode $\mathcal{M}_{t,k}$ is a Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}_{t,k}, \boldsymbol{\Sigma}_{t,k})$. The GMM also computes the posterior probability $\omega_{t,s,i,k}$ that argument $\mathcal{A}_{t,s,i}$ belongs to mode $\mathcal{M}_{t,k}$. The relationship is captured by a vector $\boldsymbol{\omega}_{t,s,i}$:

$$\boldsymbol{\omega}_{t,s,i} = [\omega_{t,s,i,1}, \ldots, \omega_{t,s,i,K}], \quad \text{where} \sum_{k=1}^{K} \omega_{t,s,i,k} = 1 \tag{5}$$

3.2.2 Temporal Mode Alignment: From Snapshot to Evolution

To understand the persistence of themes, we connect the identified modes across time. A static daily analysis is insufficient as the reasoning focus of a mode can shift. We address this by linking related modes across consecutive days, allowing us to track the complete lifecycle of investment logics as they emerge and decay.

Let $\{\boldsymbol{\mu}_{t-1,k}\}_{k=1}^{K}$ and $\{\boldsymbol{\mu}_{t,k}\}_{k=1}^{K}$ be the sets of mode centroids for days $t-1$ and t. We seek a permutation π that maximizes the total similarity between matched modes by applying a linear-assignment algorithm (Jonker & Volgenant, 1987):

$$\max_{\pi \in S_K} \sum_{k=1}^{K} s(\boldsymbol{\mu}_{t-1,k}, \boldsymbol{\mu}_{t,\pi(k)}) \tag{6}$$

where $s(\cdot, \cdot)$ is the cosine similarity and S_K is the set of all permutations of $\{1, \ldots, K\}$. The resulting permutation π establishes the temporal correspondence of each mode.

3.3 Mode Evaluation and Portfolio Construction

The final stage of MEME evaluates the historical effectiveness of each mode by leveraging the mode evolution process and the performance of each mode at every historical cross section. It helps us distinguish between enduring market wisdom and narratives that sound plausible but fail in practice. We develop a mechanism to quantify the historical profitability of each mode. This performance score is then used to weight new signals and generate final stock rankings for the future portfolio.

Formally, at the close of trading on day t, we evaluate the performance of all modes from day $t-1$. The one-day forward excess return of stock s is:

$$r_{t,s} = \frac{c_{t,s} - c_{t-1,s}}{c_{t-1,s}} - \frac{1}{N} \sum_{i=1}^{N} \frac{c_{t,i} - c_{t-1,i}}{c_{t-1,i}} \tag{7}$$

where $c_{t,s}$ is the closing price at day t. The realized score of an argument $\mathcal{A}_{t-1,s,i}$ is:

$$\text{Score}(\mathcal{A}_{t-1,s,i}) = p_{t-1,s,i} \cdot r_{t,s} \tag{8}$$

The daily aggregated score for each mode k on day $t-1$, $\text{AggScore}_{t-1,k}$, is the responsibility-weighted average score of all arguments:

$$\text{AggScore}_{t-1,k} = \frac{\sum_{s,i} \omega_{t-1,s,i,k} \cdot \text{Score}(\mathcal{A}_{t-1,s,i})}{\sum_{s,i} \omega_{t-1,s,i,k}} \tag{9}$$

To filter out fleeting, high-frequency anomalies, we update a long-term performance score, $\text{Perf}_{t-1,k}$, using an exponential moving average. This update uses the temporal mapping π' from the prior step, which links day $t-2$ to $t-1$, to ensure performance continuity, where $\lambda \in [0, 1]$ is a smoothing factor:

$$\text{Perf}_{t-1,k} = \lambda \cdot \text{Perf}_{t-2,\pi'^{-1}(k)} + (1-\lambda) \cdot \text{AggScore}_{t-1,k} \tag{10}$$

Finally, we construct our daily portfolio $\mathbf{W}_t$. By using historically proven narratives to weight new signals, we ensure that our decisions are guided by robust reasoning. The process begins by quantifying the alignment of each new argument $\mathcal{A}_{t,s,i}$ with the established modes from day $t-1$. This alignment is computed as a posterior probability, $\hat{\omega}_{t,s,i,k}$, using the GMM parameters $\{\pi_{t-1,k}, \boldsymbol{\mu}_{t-1,k}, \boldsymbol{\Sigma}_{t-1,k}\}_{k=1}^{K}$ from the previous day:

$$\hat{\omega}_{t,s,i,k} = \frac{\pi_{t-1,k} \mathcal{N}(\mathbf{x}_{t,s,i} | \boldsymbol{\mu}_{t-1,k}, \boldsymbol{\Sigma}_{t-1,k})}{\sum_{j=1}^{K} \pi_{t-1,j} \mathcal{N}(\mathbf{x}_{t,s,i} | \boldsymbol{\mu}_{t-1,j}, \boldsymbol{\Sigma}_{t-1,j})} \tag{11}$$

This probability distribution then weights the long-term performance of the modes, $\text{Perf}_{t-1,k}$, to produce a predicted score for the new argument:

$$\widehat{\text{Score}}(\mathcal{A}_{t,s,i}) = \sum_{k=1}^{K} \hat{\omega}_{t,s,i,k} \cdot \text{Perf}_{t-1,k} \tag{12}$$

We then aggregate these argument-level scores to compute a composite signal for each stock s. Let $\mathcal{I}^+_{t,s} = \{i \mid p_{t,s,i} = +1\}$ and $\mathcal{I}^-_{t,s} = \{i \mid p_{t,s,i} = -1\}$ be the index sets for positive and negative arguments. The signal is the difference between the average predicted scores of its positive and negative arguments:

$$\text{Signal}_{t,s} = \frac{1}{|\mathcal{I}^+_{t,s}| + \epsilon} \sum_{i \in \mathcal{I}^+_{t,s}} \widehat{\text{Score}}(\mathcal{A}_{t,s,i}) - \frac{1}{|\mathcal{I}^-_{t,s}| + \epsilon} \sum_{i \in \mathcal{I}^-_{t,s}} \widehat{\text{Score}}(\mathcal{A}_{t,s,i}) \tag{13}$$

where ϵ is a small constant for numerical stability. The resulting signal vector, $\text{Signal}_t \in \mathbb{R}^N$, provides a ranking for all stocks. Based on this ranking, we employ a top-K strategy with equal weight to construct the portfolio $\mathbf{W}_t$ at the open of the next trading day.

4 Experiments

4.1 Setup

Datasets: We evaluate MEME on four stock universes from both the Chinese and U.S. markets: the SSE 50, CSI 300, and CSI 500 from the A-share market, and the Nasdaq 100 from the U.S. market. Our experiments run from 2023 Q4 to 2025 Q4, a period that encompasses multiple market style transitions, including the bear market of late 2023 to early 2024, a comprehensive bull market in September 2024, a structured market phase in mid-2025, and an out-of-LLM-knowledge period in 2025 Q4. The details of the dataset description and composition are provided in Appendix A.5.

Performance metrics: Following previous work (Yoo et al., 2021; Li et al., 2024b; Guo et al., 2025; Li et al., 2025b), we utilize two types of metrics to evaluate the performance of MEME: (1) Correlation-based metrics: Information Coefficient (IC), IC Information Ratio (ICIR), Rank Information Coefficient (RIC), and RIC Information Ratio (RICIR). For these correlation metrics, the prediction target is daily forward return. (2) Portfolio-based metrics: Annualized Return (AR), Maximum Drawdown (MDD), and Sharpe Ratio (SR). The details of these metrics are in the Appendix A.6.

Baselines: We compare MEME with various competitive baselines across different types: (1) Traditional machine learning methods: LightGBM (Ke et al., 2017). (2) Deep learning techniques: DTML (Yoo et al., 2021), FactorVAE (Duan et al., 2022), MASTER (Li et al., 2024b). (3) Agent-based approaches centered on a single asset: SEP (Koa et al., 2024), TradingAgents (Xiao et al., 2025b). (4) Agent-based approaches centered on the market: R&D-Agent(Q) (Li et al., 2025b). The details of all baselines are provided in the Appendix A.7.

Implementation Details: DeepSeek V3.1 (DeepSeek-AI, 2024) serves as the backbone LLM in MEME and all agent-based baselines for fair comparisons. We use Qwen3 Embedding-8B (Zhang et al., 2025b) as the embedding model in mode identification. The number of modes K is set to 15 for the SSE 50 and 20 for the rest. The smoothing factor λ is set to 0.5. The small constant ϵ is set to 1e-5.

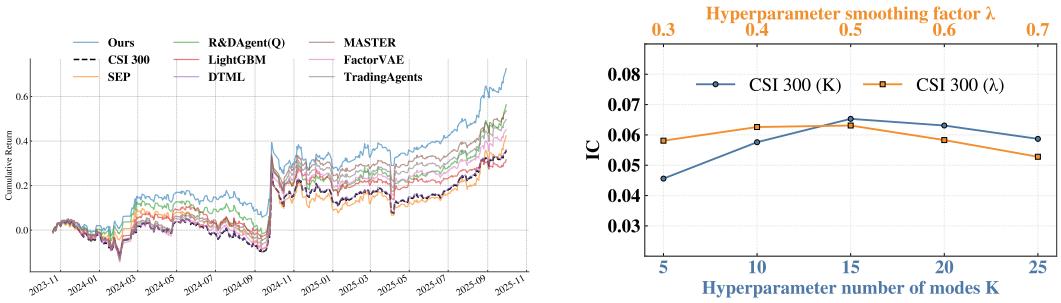
4.2 Main Experiments

Table 1 exhibits the main experiments on four datasets from 2023 Q4 to 2025 Q3. Our analysis yields two key findings. First, MEME demonstrates strong predictive capabilities, consistently achieving higher IC, RIC, and AR across all datasets. Second, MEME shows improved stability across different market conditions, obtaining higher ICIR, RICIR, and SR while also reducing the MDD.

We also draw the backtest curve on CSI 300. As shown in Figure 2a, MEME takes the performance lead in most backtest periods. Furthermore, MEME demonstrates enhanced robustness by sustaining low drawdown during the early 2024 liquidity crisis and the April 2025 trade frictions. To verify that the performance is not concentrated in high-information-event periods, we analyze excess returns

Table 1: Performance comparison on four stock pools. **Bold** and underlined numbers represent the best and second-best performance. The symbols ↑ and ↓ indicate that higher and lower values are better. The *Benchmark Index* represents the buy-and-hold strategy of the corresponding market index.

Dataset	Method	Correlation Metrics (%)				Portfolio Metrics		
		$IC^{\uparrow}$	$ICIR^{\uparrow}$	$RIC^{\uparrow}$	$RICIR^{\uparrow}$	$AR^{\uparrow}$	$MDD^{\downarrow}$	$SR^{\uparrow}$
SSE 50	LightGBM	2.63	13.56	2.98	17.14	21.19%	15.26%	0.9779
	DTML	2.97	13.89	3.19	14.53	18.47%	16.17%	0.9767
	FactorVAE	3.05	13.25	3.56	13.98	19.06%	16.55%	0.9263
	Master	4.30	18.42	5.42	24.84	24.64%	**10.71%**	1.3761
	SEP	3.65	17.89	3.16	15.30	22.58%	16.62%	0.9812
	TradingAgents	4.32	21.50	5.31	25.27	24.78%	11.05%	1.3543
	R&D-Agent(Q)	4.27	20.89	5.03	25.80	22.70%	11.21%	1.2988
	MEME	**6.69**	**33.16**	**5.71**	**31.01**	**29.18%**	10.93%	**1.4105**
	Benchmark Index	–	–	–	–	16.18%	13.95%	0.8601
CSI 300	LightGBM	1.18	9.86	2.15	17.50	12.89%	19.08%	0.5864
	DTML	3.51	23.25	4.16	25.58	19.13%	15.61%	0.8788
	FactorVAE	3.47	20.81	3.98	26.82	18.79%	16.20%	0.7892
	Master	4.37	22.60	5.26	28.30	24.46%	19.71%	1.0891
	SEP	2.98	19.76	3.42	24.75	18.22%	16.98%	0.9730
	TradingAgents	4.18	23.67	5.09	25.78	23.69%	18.40%	1.1310
	R&D-Agent(Q)	4.85	28.72	5.45	27.81	25.02%	16.50%	1.3216
	MEME	**6.78**	**42.17**	**6.21**	**40.65**	**31.91%**	**10.79%**	**1.4863**
	Benchmark Index	–	–	–	–	15.72%	17.74%	0.7265
CSI 500	LightGBM	2.18	14.01	4.81	33.11	17.41%	24.52%	0.6872
	DTML	3.64	30.47	4.18	34.72	21.04%	23.27%	0.7971
	FactorVAE	2.98	26.25	4.97	34.63	20.72%	24.99%	0.7104
	Master	2.74	15.63	4.60	23.21	16.36%	26.64%	0.6570
	SEP	2.88	18.76	3.95	21.86	15.78%	25.32%	0.5865
	TradingAgents	4.06	31.56	5.01	33.06	23.29%	18.64%	1.0517
	R&D-Agent(Q)	4.15	29.09	4.88	32.19	23.50%	19.89%	1.0343
	MEME	**6.91**	**42.15**	**6.73**	**35.06**	**32.16%**	**16.43%**	**1.2269**
	Benchmark Index	–	–	–	–	17.90%	24.55%	0.6824
Nasdaq 100	LightGBM	3.31	23.82	3.54	22.46	30.88%	22.16%	1.2109
	DTML	3.91	29.22	4.12	27.96	34.76%	19.68%	1.5647
	FactorVAE	3.28	22.74	3.49	21.88	30.25%	22.47%	1.1635
	Master	4.03	29.62	4.10	26.58	36.28%	18.34%	1.7200
	SEP	3.52	24.41	3.73	23.19	31.46%	21.84%	1.2860
	TradingAgents	3.85	28.67	4.07	27.04	35.91%	18.72%	1.6706
	R&D-Agent(Q)	3.74	27.91	3.92	**28.35**	34.12%	20.11%	1.4986
	MEME	**4.01**	**29.85**	**4.16**	28.12	**37.64%**	**17.86%**	**1.8173**
	Benchmark Index	–	–	–	–	29.02%	22.93%	1.3314



(a) Backtest on the CSI 300.

(b) Sensitivity of K and λ on IC.

Figure 2: Backtest experiment and parameter sensitivity analysis on the CSI 300.

separately during earnings seasons and non-earnings periods in Appendix A.8. MEME achieves notable annualized excess returns during both periods, indicating that MEME captures value from the ongoing evolution of investment narratives rather than relying on information bursts.

Table 2: Ablation study of different components. We incrementally add components SAG (Structured Argument Generation), MOT (Mode Of Thoughts), PM (Probabilistic Modeling), and TA (Temporal Alignment) to evaluate their contribution. More ablation study can be found in Appendix A.2.

Components				SSE 50				CSI 300				CSI 500			
SAG	MOT	PM	TA	IC	ICIR	RIC	RICIR	IC	ICIR	RIC	RICIR	IC	ICIR	RIC	RICIR
✗	✗	✗	✗	1.87	10.11	1.06	7.89	0.79	6.92	0.94	5.88	0.18	1.68	0.24	3.63
✓	✗	✗	✗	2.51	12.73	1.92	9.33	3.16	14.40	1.89	8.65	0.21	1.65	0.38	4.70
✓	✓	✗	✗	2.47	14.66	1.91	10.85	5.30	30.43	5.24	29.36	4.90	38.09	4.87	36.20
✓	✓	✓	✗	3.18	16.95	3.08	17.38	6.44	38.59	5.61	32.67	5.22	37.14	5.10	34.12
✓	✓	✓	✓	**6.69**	**33.16**	**5.71**	**31.01**	**6.78**	**42.17**	**6.21**	**40.65**	**6.91**	**42.15**	**6.73**	**35.06**

4.3 Experiments on Data Leakage Concern

To validate MEME against potential data leakage, we conduct a temporal out-of-sample test on four datasets using new data from Q4 2025, which falls after the LLM knowledge cutoff. As shown in Table 7 in the Appendix, MEME maintains its performance advantage over all baselines across all four datasets. This strongly indicates that MEME's effectiveness stems from our methodological design and not from contamination within the LLM's training corpus.

4.4 Ablation Study

We conduct ablation studies to evaluate the contribution of each core component of MEME. Table 2 presents the results of incrementally adding each component: Structured Argument Generation (SAG), Modes of Thought (MOT), Probabilistic Modeling (PM), and Temporal Alignment (TA).

The first row establishes a baseline by using the LLM directly on raw data, which yields the lowest performance across all datasets and demonstrates the necessity of structured analysis. Introducing SAG alone provides a performance lift on smaller stock pools like SSE 50. However, its effectiveness diminishes on the larger CSI 500, where metrics show minimal improvement, suggesting that simply extracting arguments is insufficient for complex scenarios.

The integration of MOT, even with a deterministic clustering approach, leads to a performance leap, particularly on CSI 300 and CSI 500. This highlights the critical role of organizing arguments into coherent reasoning patterns. Subsequently, replacing the k-means clustering with our Probabilistic Modeling component further enhances performance across all datasets. This confirms the benefit of modeling the ambiguous relationship between arguments and their underlying modes of thought.

Finally, adding the Temporal Alignment to form the complete model also brings significant gains. This result validates that modeling the lifecycle of thought patterns is essential for generating robust predictive signals. We provide extended ablation studies in Appendix A.2 to further validate the design choice of MEME, including text-information utilization, argument extraction pipeline, chosen of clustering techniques and the contribution of narrative structure from temporal scoring.

4.5 Parameter Sensitivity Analysis

We investigate the sensitivity of MEME to its key hyperparameters: the number of modes K and the smoothing factor λ. The analysis is conducted on the CSI 300, and the results are presented in Figure 2b. More results are provided in Figure 5a

Overall, MEME demonstrates robustness to variations in both parameters, maintaining stable performance across a considerable range of values.

For the number of modes K, performance degrades when the value is too small. An insufficient K forces conceptually distinct arguments into the same mode, which negatively impacts the quality of the resulting signals. Regarding the smoothing factor λ, excessively large values also harm performance. As defined in Equation 10, a large λ places greater weight on the long-term historical performance of modes relative to their most recent effectiveness. This reduces the model's responsiveness to recent shifts in the predictive power of the thought patterns, thereby degrading the final results.

4.6 Cost Time Analysis

We compare the daily average running time of MEME with SOTA LLM-based baselines. As shown in Figure 5b in the Appendix, MEME achieves lower computational cost by filtering noise via

structured argument extraction and by aligning information across assets and trading days through mode identification and alignment, thereby reducing the overhead of explicit LLM reflection.

4.7 Signal Orthogonality Analysis

Beyond absolute performance, a practical value of any new signal source lies in its ability to provide complementary information to existing models (DeMiguel et al., 2009). We compute the daily Pearson correlation between the stock-level signals generated by MEME and those from three DL baselines that rely on price-volume features. As shown in Table 3, the correlations are near-zero or negative across all datasets. This confirms that MEME produces a distinct, low-correlation signal source. This orthogonality suggests that MEME can serve as a complementary alpha source in multi-model ensemble strategies.

Table 3: Correlation between MEME and DL baseline signals.

Model	SSE 50	CSI 300	CSI 500	Nasdaq 100
LightGBM	−0.07	−0.11	−0.09	0.05
DTML	−0.06	−0.04	0.02	-0.02
MASTER	−0.02	−0.07	−0.11	-0.06

4.8 Case Study

To investigate what Modes of Thought look like in practice, we conduct a case study on CSI 300. Each mode is a cluster of Investment Arguments that share similar reasoning logic. We track the long-term performance score Perf for each mode and classify them by sustained effectiveness across market regimes (bear market: Oct 2023 to Sep 2024; bull market: Sep 2024 to present) into four categories: Long-term effective, Bull-market effective, Bear-market effective, and Ineffective. The detailed classification criteria and concrete argument examples are provided in Appendix A.9.

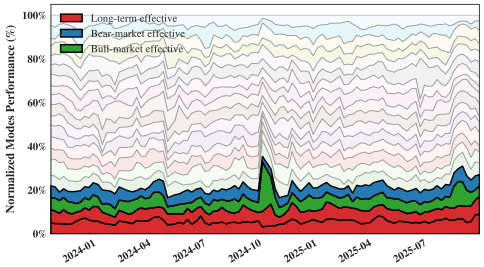


Figure 3: Normalized performance evolution of Modes of Thought categories on CSI 300.

Figure 3 shows how the relative performance contribution of each category evolves over time (normalized via Softmax across all modes per day). Several patterns emerge. First, a large portion of modes remain ineffective, consistent with the rarity of sustained profitability in financial markets. Second, the composition of effective modes shifts with market conditions: bear-market modes, which center on dividend yield and balance-sheet strength, dominated during the early 2024 liquidity crisis, resembling the value investing philosophy of Graham (Graham & Zweig, 2003); bull-market modes, which rely on technical momentum and capital flows, gained dominance during the late 2024 rally, resembling the trend-following approach of Livermore (Lefevre & Markman, 2010). The fact that purely unsupervised clustering recovers patterns resembling well-established investment philosophies suggests that the identified Modes of Thought reflect meaningful reasoning patterns with real-world correspondence.

5 Conclusion

In this paper, we introduce MEME, which models financial markets from a Logic-Oriented perspective grounded in Narrative Economics. By extracting structured Investment Arguments, identifying latent Modes of Thought through probabilistic clustering, and tracking their lifecycle via temporal alignment, MEME shifts the analytical focus from individual assets to the evolution of investment logics. Experiments on four stock pools demonstrate consistent improvements over SOTA baselines. The lifecycle case study further reveals that the unsupervised modes recovered by MEME exhibit correspondence to established investment philosophies, suggesting that the framework identifies reasoning patterns with real-world relevance. We discuss the scope of our framework, including its reliance on text coverage and behavior on small-cap universes in Appendix A.3. In the future, We hope we can extend MEME to text-sparse environments and explore richer temporal modeling.

AI USE STATEMENT

Generative AI is a component of the method studied in this paper: MEME uses large language models to filter raw multimodal data and to extract structured Investment Arguments, and all such uses are described in Section 3 and the Appendix. Beyond this methodological use, we used generative AI tools only as general-purpose writing and coding assistants, for polishing the language of the manuscript and for routine coding support during implementation and plotting. We did not use generative AI tools to generate research ideas, to derive the formulations presented in this paper, to produce experimental results, or to write the related work survey. All AI-assisted text was reviewed and edited by the authors, and all AI-assisted code was verified and tested by the authors. We take responsibility for the final content of this work, including all text, claims and artifacts.

REFERENCES

Dogu Araci. Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*, 2019.

Brad M Barber and Terrance Odean. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The review of financial studies*, 21(2):785–818, 2008.

Stephen Blyth. *An introduction to quantitative finance*. Oxford University Press, 2014.

Lang Cao, Zekun Xi, Long Liao, Ziwei Yang, and Zheng Cao. Chain-of-alpha: Unleashing the power of large language models for alpha mining in quantitative trading. *arXiv preprint arXiv:2508.06312*, 2025.

Binqi Chen, Hongjun Ding, Ning Shen, Jinsheng Huang, Taian Guo, Luchen Liu, and Ming Zhang. Alphasage: Structure-aware alpha mining via gflownets for robust exploration. *arXiv preprint arXiv:2509.25055*, 2025.

Tianxiang Chen, Wei Chen, and Luyao Du. An empirical study of financial factor mining based on gene expression programming. In *2021 4th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE)*, pp. 1113–1117. IEEE, 2021.

DeepSeek-AI. Deepseek-v3 technical report, 2024. URL https://arxiv.org/abs/2412.19437.

Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The review of Financial studies*, 22(5):1915–1953, 2009.

Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society: series B (methodological)*, 39(1): 1–22, 1977.

Zheye Deng and Jiashu Wang. Alphaquanter: An end-to-end tool-orchestrated agentic reinforcement learning framework for stock trading. *arXiv preprint arXiv:2510.14264*, 2025.

Kelvin Du, Frank Xing, Rui Mao, and Erik Cambria. Financial sentiment analysis: Techniques and applications. *ACM Computing Surveys*, 56(9):1–42, 2024.

Yitong Duan, Lei Wang, Qizhong Zhang, and Jian Li. Factorvae: A probabilistic dynamic factor model based on variational autoencoder for predicting cross-sectional stock returns. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pp. 4468–4476, 2022.

Yitong Duan, Weiran Wang, and Jian Li. Factorgcl: A hypergraph-based factor model with temporal residual contrastive learning for stock returns prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 173–181, 2025.

Richard O Duda and Peter E Hart. Pattern classification and scene analysis. *A Wiley-interscience publication*, 1973.

Benjamin Graham and Jason Zweig. *The intelligent investor*. HarperBusiness Essentials New York, 2003.

Taian Guo, Haiyang Shen, Jinsheng Huang, Zhengyang Mao, Junyu Luo, Zhuoru Chen, Xuhui Liu, Bingyu Xia, Luchen Liu, Yun Ma, et al. Mass: Multi-agent simulation scaling for portfolio construction. *arXiv preprint arXiv:2505.10278*, 2025.

Roy Jonker and Anton Volgenant. A shortest augmenting path algorithm for dense and sparse linear assignment problems. *Computing*, 38(4):325–340, 1987.

Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.

Kelvin JL Koa, Yunshan Ma, Ritchie Ng, and Tat-Seng Chua. Learning to generate explainable stock predictions using self-reflective large language models. In *Proceedings of the ACM Web Conference 2024*, pp. 4304–4315, 2024.

Edwin Lefevre and Jon D Markman. *Reminiscences of a stock operator: With new commentary and insights on the life and times of Jesse Livermore*. John Wiley & Sons, 2010.

Junjie Li, Yang Liu, Weiqing Liu, Shikai Fang, Lewen Wang, Chang Xu, and Jiang Bian. Mars: a financial market simulation engine powered by generative foundation model. In *The Thirteenth International Conference on Learning Representations*, 2025a. URL https://openreview.net/forum?id=Yqk7EyT52H.

Shuqi Li, Yuebo Sun, Yuxin Lin, Xin Gao, Shuo Shang, and Rui Yan. Causalstock: Deep end-to-end causal discovery for news-driven multi-stock movement prediction. *Advances in Neural Information Processing Systems*, 37:47432–47454, 2024a.

Tong Li, Zhaoyang Liu, Yanyan Shen, Xue Wang, Haokun Chen, and Sen Huang. Master: Market-guided stock transformer for stock price forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 162–170, 2024b.

Yuante Li, Xu Yang, Xiao Yang, Minrui Xu, Xisen Wang, Weiqing Liu, and Jiang Bian. R&d-agent-quant: A multi-agent framework for data-centric factors and model joint optimization. In *The Thirty-Ninth Annual Conference on Neural Information Processing Systems*, 2025b.

Zhiwei Li, Ran Song, Caihong Sun, Wei Xu, Zhengtao Yu, and Ji-Rong Wen. Can large language models mine interpretable financial factors more effectively? a neural-symbolic factor mining agent model. In *Findings of the Association for Computational Linguistics ACL 2024*, pp. 3891–3902, 2024c.

Hengxu Lin, Dong Zhou, Weiqing Liu, and Jiang Bian. Deep risk model: a deep learning solution for mining latent risk factors to improve covariance matrix estimation. In *Proceedings of the Second ACM International Conference on AI in Finance*, pp. 1–8, 2021.

Aixin Liu, Aoxue Mei, Bangcai Lin, Bing Xue, Bingxuan Wang, Bingzheng Xu, Bochao Wu, Bowei Zhang, Chaofan Lin, Chen Dong, et al. Deepseek-v3. 2: Pushing the frontier of open large language models. *arXiv preprint arXiv:2512.02556*, 2025.

Di Luo, Weiheng Liao, Shuqi Li, Xin Cheng, and Rui Yan. Causality-guided multi-memory interaction network for multivariate stock price movement prediction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 12164–12176, 2023.

Alexander J McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton university press, 2015.

Hao Qian, Hongting Zhou, Qian Zhao, Hao Chen, Hongxiang Yao, Jingwei Wang, Ziqi Liu, Fei Yu, Zhiqiang Zhang, and Jun Zhou. Mdgnn: Multi-relational dynamic graph neural network for comprehensive and dynamic stock investment prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 14642–14650, 2024.

Hao Shi, Weili Song, Xinting Zhang, Jiahe Shi, Cuicui Luo, Xiang Ao, Hamid Arian, and Luis Angel Seco. Alphaforge: A framework to mine and dynamically combine formulaic alpha factors. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 12524–12532, 2025a.

Yu Shi, Yitong Duan, and Jian Li. Navigating the alpha jungle: An llm-powered mcts framework for formulaic factor mining. *arXiv preprint arXiv:2505.11122*, 2025b.

Yu Shi, Zongliang Fu, Shuo Chen, Bohan Zhao, Wei Xu, Changshui Zhang, and Jian Li. Kronos: A foundation model for the language of financial markets. *arXiv preprint arXiv:2508.02739*, 2025c.

Robert J Shiller. Narrative economics: How stories go viral and drive major economic events. 2020.

Lingyun Song, Haodong Li, Siyu Chen, Xinbiao Gan, Binze Shi, Jie Ma, Yudai Pan, Xiaoqi Wang, and Xuequn Shang. Multi-scale temporal neural network for stock trend prediction enhanced by temporal hyepredge learning. *Proceedings of the IJCAI, Montreal, QC, Canada*, pp. 16–22, 2025.

Ziyi Tang, Zechuan Chen, Jiarui Yang, Jiayao Mai, Yongsen Zheng, Keze Wang, Jinrui Chen, and Liang Lin. Alphaagent: Llm-driven alpha mining with regularized exploration to counteract alpha decay. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pp. 2813–2822, 2025.

Gemma Team, Aishwarya Kamath, Johan Ferret, Shreya Pathak, Nino Vieillard, Ramona Merhej, Sarah Perrin, Tatiana Matejovicova, Alexandre Ramé, Morgane Rivière, et al. Gemma 3 technical report. *arXiv preprint arXiv:2503.19786*, 2025a.

Kimi Team, Yifan Bai, Yiping Bao, Guanduo Chen, Jiahao Chen, Ningxin Chen, Ruijue Chen, Yanru Chen, Yuankun Chen, Yutian Chen, et al. Kimi k2: Open agentic intelligence. *arXiv preprint arXiv:2507.20534*, 2025b.

Feng Tian, Flora D Salim, and Hao Xue. Tradinggroup: A multi-agent trading system with self-reflection and data-synthesis. *arXiv preprint arXiv:2508.17565*, 2025.

Hanshuang Tong, Jun Li, Ning Wu, Ming Gong, Dongmei Zhang, and Qi Zhang. Ploutos: Towards interpretable stock movement prediction with financial large language model. *arXiv preprint arXiv:2403.00782*, 2024.

Quanzhou Xiang, Zhan Chen, Qi Sun, and Rujun Jiang. Rsap-dfm: Regime-shifting adaptive posterior dynamic factor model for stock returns prediction. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24. International Joint Conferences on Artificial Intelligence Organization*, 2024.

Yijia Xiao, Edward Sun, Tong Chen, Fang Wu, Di Luo, and Wei Wang. Trading-r1: Financial trading with llm reasoning via reinforcement learning. *arXiv preprint arXiv:2509.11420*, 2025a.

Yijia Xiao, Edward Sun, Di Luo, and Wei Wang. Tradingagents: Multi-agents llm financial trading framework. In *The First MARW: Multi-Agent AI in the Real World Workshop at AAAI 2025*, 2025b.

Wentao Xu, Weiqing Liu, Lewen Wang, Yingce Xia, Jiang Bian, Jian Yin, and Tie-Yan Liu. Hist: A graph-based framework for stock trend forecasting via mining concept-oriented shared information. *arXiv preprint arXiv:2110.13716*, 2021.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, et al. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025.

Yuzhe YANG, Yifei Zhang, Minghao Wu, Kaidi Zhang, Yunmiao Zhang, Honghai Yu, Yan Hu, and Benyou Wang. Twinmarket: A scalable behavioral and social simulation for financial markets. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=h60y6zlPyl.

Jaemin Yoo, Yejun Soun, Yong-chan Park, and U Kang. Accurate multivariate stock movement prediction via data-axis transformer with multi-level contexts. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 2037–2045, 2021.

Shuo Yu, Hongyan Xue, Xiang Ao, Feiyang Pan, Jia He, Dandan Tu, and Qing He. Generating synergistic formulaic alpha collections via reinforcement learning. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 5476–5486, 2023.

Wentao Zhang, Lingxuan Zhao, Haochong Xia, Shuo Sun, Jiaze Sun, Molei Qin, Xinyi Li, Yuqing Zhao, Yilei Zhao, Xinyu Cai, et al. A multimodal foundation agent for financial trading: Tool-augmented, diversified, and generalist. In *Proceedings of the 30th acm sigkdd conference on knowledge discovery and data mining*, pp. 4314–4325, 2024a.

Wentao Zhang, Yilei Zhao, Shuo Sun, Jie Ying, Yonggang Xie, Zitao Song, Xinrun Wang, and Bo An. Reinforcement learning with maskable stock representation for portfolio management in customizable stock pools. In *Proceedings of the ACM Web Conference 2024*, pp. 187–198, 2024b.

Xu Zhang, Zhengang Huang, Yunzhi Wu, Xun Lu, Erpeng Qi, Yunkai Chen, Zhongya Xue, Qitong Wang, Peng Wang, and Wei Wang. Multi-period learning for financial time series forecasting. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, pp. 2848–2859, 2025a.

Yanzhao Zhang, Mingxin Li, Dingkun Long, Xin Zhang, Huan Lin, Baosong Yang, Pengjun Xie, An Yang, Dayiheng Liu, Junyang Lin, Fei Huang, and Jingren Zhou. Qwen3 embedding: Advancing text embedding and reranking through foundation models. *arXiv preprint arXiv:2506.05176*, 2025b.

Yu Zhao, Huaming Du, Ying Liu, Shaopeng Wei, Xingyan Chen, Fuzhen Zhuang, Qing Li, and Gang Kou. Stock movement prediction based on bi-typed hybrid-relational market knowledge graph via dual attention networks. *IEEE transactions on knowledge and data engineering*, 35(8):8559–8571, 2022.

Xiaolin Zheng, Mengpu Liu, and Mengying Zhu. Deep hashing-based dynamic stock correlation estimation via normalizing flow. In *IJCAI*, pp. 4993–5001, 2023.

Haoren Zhu, Shih-Yang Liu, Pengfei Zhao, Yingying Chen, and Dik Lun Lee. Forecasting asset dependencies to reduce portfolio risk. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 4397–4404, 2022.

Zhoufan Zhu and Ke Zhu. Alphaqcm: Alpha discovery in finance with distributional reinforcement learning. In *Forty-second International Conference on Machine Learning*, 2025.

A APPENDIX

A.1 PSEUDOCODE OF MEME

Algorithm 1 MEME

Input: Raw multi-modal data $\{\mathcal{D}_t\}_{t\in T}$, historical prices $\{P_t\}_{t\in T}$, number of modes K, EMA factor λ

Output: Sequence of daily portfolios $\{\mathbf{W}_t\}_{t\in T}$

Initialize long-term performance scores $\text{Perf}_{0,k} \leftarrow 0$ for $k = 1, \ldots, K$ Initialize mode parameters for day 0, $\text{Modes}_0 \leftarrow \emptyset$

for *each trading day* $t \in T$ **do**

  `// -- Stage 1: Argument Extraction --`

  Extract structured arguments $\{\mathcal{A}_{t,s,i}\}_{s,i}$ from raw data $\mathcal{D}_t$ via Eqs. (1-3)

  `// -- Stage 2: Mode Identification and Alignment --`

  Encode arguments to get embeddings $\{\mathbf{x}_{t,s,i}\}$ via Eq. (4)

  Fit GMM on $\{\mathbf{x}_{t,s,i}\}$ to identify daily modes $\text{Modes}_t = \{\mathcal{M}_{t,k}\}_{k=1}^{K}$ and responsibilities $\{\boldsymbol{\omega}_{t,s,i}\}$

  if $t > 0$ **then**

    `// Align today's modes with yesterday's`

    Find permutation π_t by aligning centroids of Modes_t and Modes_{t-1} via Eq. (6)

    `// -- Stage 3.1: Mode Evaluation (Backward-looking) --`

    `// Update performance scores using t-1's predictions and t's returns`

    Calculate aggregated score $\text{AggScore}_{t-1,k}$ for each mode from day $t-1$ via Eqs. (7-9) Update long-term performance $\text{Perf}_{t-1,k}$ using Perf_{t-2} and π_{t-1} via EMA in Eq. (10)

    `// -- Stage 3.2: Portfolio Construction (Forward-looking) --`

    `// Generate trading signals for day t using updated scores`

    Calculate predicted score $\widehat{\text{Score}}(\mathcal{A}_{t,s,i})$ for each new argument via Eqs. (11-12) using Perf_{t-1} Aggregate argument scores into a final stock signal $\text{Signal}_{t,s}$ via Eq. (13) Construct portfolio $\mathbf{W}_t$ based on a Top-K ranking of Signal_t

  end

  else

    $\mathbf{W}_t \leftarrow$ EqualWeighted Portfolio

  end

end

return *Sequence of daily portfolios* $\{\mathbf{W}_t\}_{t\in T}$

A.2 EXTENDED ABLATION STUDIES

We conduct four groups of additional ablation experiments to further validate the design choices in MEME. All experiments use the same test period (2023 Q4 to 2025 Q3) and report correlation metrics. Results are summarized in Table 4.

Text Information Utilization. To verify that MEME's advantage stems from its structured reasoning pipeline rather than merely having access to text data, we augment two DL baselines (LightGBM and MASTER) with text features by appending average daily Qwen3-Embedding vectors of news headlines to their original feature sets. Both text-augmented baselines produce positive IC, confirming that news embeddings carry useful signal. However, simple feature concatenation treats the embedding as a static numeric vector, discarding the structured argument semantics that MEME explicitly models. The substantial gap between MASTER+Text and MEME demonstrates that our framework's value lies in how it processes text, not merely in having access to it.

Argument Extraction Pipeline. We test three degraded configurations: (1) W/O Information Filtering removes the filter agent and feeds raw data directly to the argument generator; (2) W/O Argument Structure retains the filter but replaces the structured argument format (a, e) in Equation 3 with the filtered summary directly, using the LLM only to assign a polarity label; (3) W/O Both applies both modifications. Removing filtering alone reduces IC moderately. Removing the argument structure causes even larger degradation, showing that argument generation is the primary signal source. Combining both yields near-zero IC, confirming that the two stages are individually necessary and jointly sufficient.

Table 4: Extended ablation studies across three datasets. All values are in percent. Each group isolates a specific design dimension of MEME.

Method	SSE 50				CSI 300				CSI 500			
	IC	ICIR	RIC	RICIR	IC	ICIR	RIC	RICIR	IC	ICIR	RIC	RICIR
Text Information Utilization												
LightGBM+Text	2.62	16.59	2.87	20.40	3.19	23.77	3.65	27.45	2.85	23.51	3.02	20.38
MASTER+Text	3.26	17.13	3.38	19.95	4.03	23.69	4.25	26.60	3.61	25.19	3.88	27.52
Argument Extraction Pipeline												
W/O Information Filtering	4.92	25.84	4.18	22.67	4.59	26.54	4.41	25.18	4.05	24.06	3.99	20.46
W/O Argument Structure	3.17	18.48	3.31	19.57	3.95	23.06	3.64	17.60	2.92	16.97	2.41	16.86
W/O Both	0.76	5.42	0.58	4.73	0.61	4.95	0.52	4.18	0.43	3.67	0.39	2.94
Clustering Method in Mode Identification												
DTM	5.42	30.21	3.89	26.18	5.46	36.84	4.99	34.32	5.88	37.27	5.94	31.37
Spherical K-Means	2.97	20.14	2.81	19.92	3.92	29.87	4.28	33.84	4.29	32.10	4.17	34.32
Narrative Structure vs. Temporal Scoring												
Momentum Baseline	2.62	13.72	2.15	13.10	3.41	16.39	3.08	15.46	2.23	11.70	2.53	13.89
SAG+MOT (no PM, no TA)	2.47	14.66	1.91	10.85	5.30	30.43	5.24	29.36	4.90	38.09	4.87	36.20
MEME (Full)	**6.69**	**33.16**	**5.71**	**31.01**	**6.78**	**42.17**	**6.21**	**40.65**	**6.91**	**42.15**	**6.73**	**35.06**

Clustering Method in Mode Identification. We compare our GMM-based approach with two alternatives: Dynamic Topic Models (DTM), which relies on word co-occurrence, and Spherical K-Means, a continuous-space hard-assignment method. GMM consistently outperforms both. DTM operates on discrete word distributions and degrades on the short texts of Investment Arguments. Spherical K-Means performs hard assignment, losing the multi-narrative participation captured by GMM's soft probabilities, and projecting all embeddings onto the unit sphere discards the distance and density structure of the original space. GMM models per-component covariance, distinguishing compact high-consensus modes from diffuse low-consensus ones.

Narrative Structure vs. Temporal Scoring. To isolate the contribution of the mode-based narrative structure from the temporal scoring mechanism, we implement a momentum baseline that scores each stock by the average realized excess return of its arguments from the preceding 5 trading days, bypassing the mode structure entirely. On SSE 50, momentum and SAG+MOT perform comparably, as a concentrated 50-stock universe has limited narrative diversity. However, as the universe expands, the advantage of mode clustering becomes pronounced: on CSI 500, SAG+MOT substantially outperforms momentum. The full MEME demonstrates that combining narrative structure with temporal evaluation yields the strongest results.

A.3 Further Discussions and Limitations

To examine the generalizability of MEME under varying text coverage conditions, we conduct experiments on the CSI 1000, a small-cap index where textual information (news and analyst reports) is inherently sparser than for large- and mid-cap stocks.

Table 5: Performance on the CSI 1000 (2023 Q4 to 2025 Q3). All correlation values are in percent.

Method	$IC^{\uparrow}$	$ICIR^{\uparrow}$	$RIC^{\uparrow}$	$RICIR^{\uparrow}$	$AR^{\uparrow}$	$MDD^{\downarrow}$	$SR^{\uparrow}$
LightGBM	3.38	21.77	4.64	20.18	20.17%	29.86%	0.78
DTML	4.20	**34.20**	4.15	30.34	27.18%	30.74%	1.01
FactorVAE	3.89	25.40	4.23	27.84	21.30%	31.93%	0.85
MASTER	**4.71**	28.27	**6.27**	**35.41**	**29.46**%	30.61%	**0.99**
SEP	1.65	11.92	1.91	12.13	11.04%	31.28%	0.34
TradingAgents	2.06	15.79	2.23	14.57	15.28%	30.76%	0.43
R&D-Agent(Q)	3.85	23.21	4.08	22.60	22.84%	29.48%	0.77
MEME	4.23	24.70	4.69	22.19	25.62%	**27.69**%	0.88
Benchmark	–	–	–	–	12.01%	33.34%	0.49

As shown in Tables 5 and 6, MEME does not outperform all DL baselines on the CSI 1000. MASTER, which relies exclusively on dense price-volume features, achieves a higher IC (4.71 vs. 4.23) over the full period. This gap reflects a fundamental trade-off: price-volume models benefit from uniformly dense quantitative data regardless of text coverage, while MEME relies on textual information that is

Table 6: Performance on the CSI 1000 (2025 Q4). All correlation values are in percent.

Method	$IC^{\uparrow}$	$ICIR^{\uparrow}$	$RIC^{\uparrow}$	$RICIR^{\uparrow}$	$AR^{\uparrow}$	$MDD^{\downarrow}$	$SR^{\uparrow}$
LightGBM	2.14	19.80	4.07	24.32	3.64%	8.26%	0.34
DTML	2.67	18.39	3.54	21.36	3.21%	9.18%	0.29
FactorVAE	2.51	18.05	3.82	23.66	3.77%	8.64%	0.36
MASTER	**3.63**	**25.38**	**5.06**	**38.17**	**9.86**%	**6.42**%	**1.22**
SEP	1.96	16.45	2.87	21.40	1.92%	9.85%	0.18
TradingAgents	2.73	24.06	3.16	22.54	4.94%	8.01%	0.51
R&D-Agent(Q)	2.91	23.13	3.10	24.67	4.17%	7.44%	0.58
MEME	3.48	25.14	3.65	30.08	6.31%	6.98%	0.90
Benchmark	–	–	–	–	1.08%	6.91%	0.21

inherently sparser for small-cap constituents. Analyst coverage, media attention, and financial report detail are all systematically lower for CSI 1000 stocks compared to the SSE 50 or CSI 300, limiting the quality of the extracted Investment Arguments.

However, two observations merit attention. First, MEME consistently outperforms all LLM-based baselines (SEP, TradingAgents, and R&D-Agent(Q)) on the CSI 1000 in both periods. This suggests that the structured argument extraction and mode-based reasoning pipeline extracts signal from text more effectively than alternative LLM-based approaches, even under text-sparse conditions. Second, MEME achieves the lowest MDD (27.69%) over the full period, indicating that the temporal mode evaluation mechanism provides a degree of risk control even when the raw text signal is weaker.

These results highlight a clear boundary condition for MEME: its effectiveness scales with the availability and quality of textual information. We view extending MEME to text-sparse environments as an important direction for future work.

A.4 Temporal Out-of-Sample Test (2025 Q4)

To validate that MEME's performance does not stem from contamination within the LLM's training corpus, we conduct a temporal out-of-sample test using data from 2025 Q4, which falls after the knowledge cutoff of the backbone LLM. As shown in Table 7, MEME maintains its advantage over all baselines across all four datasets, confirming that its effectiveness stems from the methodological design rather than data leakage.

A.5 Dataset composition details

A.5.1 Stock pool details

- **SSE 50**: This index represents the 50 largest and most liquid A-share stocks on the Shanghai Stock Exchange. It captures the performance of leading blue-chip companies.
- **CSI 300**: This index is a broad market benchmark for China's A-share market. It consists of the 300 largest and most liquid stocks from both the Shanghai and Shenzhen exchanges, offering a comprehensive representation of the large-cap segment.
- **CSI 500**: This index tracks the performance of the next 500 largest and most liquid A-share stocks after the constituents of the CSI 300. It primarily reflects the behavior of mid and small-cap companies, providing a distinct universe from the large-cap focused indices.
- **Nasdaq 100**: This index comprises the 100 largest non-financial companies listed on the Nasdaq exchange in the U.S. market. It is heavily weighted towards the technology sector and serves as a representative benchmark for large-cap growth stocks.

A.5.2 Dataset construction details

Fundamental Modality

- Financial Statements, including Income Statement, Balance Statement, CashFlow Statement
- Valuation Metrics, including Price-to-Earnings Ratio, Price-to-Book Ratio, Price-to-Cash Flow Ratio, Price-to-Sales Ratio, and dividend Ratio.

News Modality

Table 7: Experiments on 2025 Q4. Performance comparison on four datasets after the LLM knowledge cutoff.

Dataset	Method	Correlation Metrics (%)				Portfolio Metrics		
		$IC^{\uparrow}$	$ICIR^{\uparrow}$	$RIC^{\uparrow}$	$RICIR^{\uparrow}$	$AR^{\uparrow}$	$MDD^{\downarrow}$	$SR^{\uparrow}$
SSE 50	LightGBM	2.26	20.33	2.37	22.24	8.10%	4.89%	0.8523
	DTML	2.45	15.81	3.24	22.91	5.80%	5.53%	0.5334
	FactorVAE	1.79	10.60	**2.88**	19.85	2.84%	5.28%	0.4998
	Master	2.37	**27.96**	2.48	22.08	6.49%	4.81%	0.6820
	SEP	2.30	18.55	1.91	17.30	3.03%	6.06%	0.4120
	TradingAgents	2.18	20.06	2.05	18.87	4.18%	5.64%	0.5226
	R&D-Agent(Q)	2.58	23.88	2.65	21.80	7.88%	6.10%	0.6851
	MEME	**2.67**	24.35	2.83	**26.42**	**8.25%**	**4.79%**	**0.8894**
	Benchmark Index	–	–	–	–	5.66%	4.85%	0.7653
CSI 300	LightGBM	1.16	9.84	3.02	20.54	-2.56%	7.42%	-0.2579
	DTML	3.44	21.08	4.27	26.47	3.20%	5.01%	0.4789
	FactorVAE	2.28	16.62	3.70	21.22	0.99%	5.89%	0.1335
	Master	3.47	17.18	3.77	18.13	0.79%	6.16%	0.0924
	SEP	1.93	11.72	2.34	14.29	2.06%	5.43%	0.1910
	TradingAgents	2.61	15.34	3.07	19.89	4.87%	5.12%	0.4120
	R&D-Agent(Q)	4.09	23.01	3.89	26.20	5.42%	5.31%	0.5838
	MEME	**4.12**	**32.11**	**4.53**	**29.39**	**9.77%**	**4.87%**	**0.7353**
	Benchmark Index	–	–	–	–	-0.92%	7.46%	-0.0837
CSI 500	LightGBM	2.39	12.24	4.71	28.40	2.39%	8.73%	0.0618
	DTML	3.24	20.49	1.89	16.54	14.06%	9.02%	0.6495
	FactorVAE	2.23	18.80	3.66	17.75	12.51%	8.75%	0.6122
	Master	2.88	20.42	4.65	33.28	19.84%	6.21%	1.1254
	SEP	3.01	16.35	3.24	19.49	4.71%	8.96%	0.2427
	TradingAgents	3.87	21.64	3.93	28.32	8.22%	8.40%	0.5680
	R&D-Agent(Q)	4.38	27.70	4.73	33.59	13.98%	7.22%	0.9606
	MEME	**5.15**	**32.61**	**5.31**	**34.97**	**20.56%**	**5.95%**	**1.1433**
	Benchmark Index	–	–	–	–	2.88%	10.08%	0.2281
Nasdaq 100	LightGBM	2.94	22.18	3.28	25.44	9.88%	7.26%	0.7195
	DTML	3.49	28.31	3.88	29.96	14.06%	6.42%	1.2038
	FactorVAE	2.87	22.54	3.19	24.36	9.42%	7.38%	0.6650
	Master	3.72	27.03	3.82	31.40	15.21%	6.18%	1.3649
	SEP	3.08	23.46	3.35	24.95	10.61%	7.04%	0.7807
	TradingAgents	3.55	27.86	3.96	30.47	14.74%	**6.03%**	1.2943
	R&D-Agent(Q)	3.31	25.62	3.63	27.88	12.32%	6.83%	1.0459
	MEME	**3.67**	**28.74**	**4.08**	**31.22**	**15.68%**	6.12%	**1.4385**
	Benchmark Index	–	–	–	–	7.46%	7.91%	0.4871

- News collected from Wind [1] (for A-share markets) and AlphaVantage [2] (for the U.S. market).
- Financial Report Collected from EastMoney [3] (for A-share markets) and AlphaVantage (for the U.S. market).

Technical Modality

- **Price and Volume Series**: Daily, weekly, and monthly adjusted close prices, trading volumes, and market cap.
- **Technical Indicators**:
  - Trend-following: Daily, weekly, and monthly moving averages.
  - Momentum: Daily, weekly, and monthly Moving Average Convergence Divergence (MACD) and Relative Strength Index (RSI).
  - Volatility: Daily, weekly, and monthly Bollinger Bands (BOLL).
  - Oscillators: Daily, weekly, and monthly Stochastic Oscillator (KDJ).
- **Contextual Metadata**: The primary industry classification for each stock.
- **Benchmark Data**: Daily price series for the corresponding industry indices.

A.6 Performance Metrics Details

We assess the performance of a predictive model denoted by f. At each time step t, for a universe of N_t assets, the model generates a vector of predictive scores $\mathbf{s}_t \in \mathbb{R}^{N_t}$. The efficacy of these scores is measured against the vector of subsequent asset returns $\mathbf{r}_t \in \mathbb{R}^{N_t}$. Unless specified otherwise, all statistical moments such as expectation $\mathbb{E}[\cdot]$, variance $\mathrm{Var}[\cdot]$, and covariance $\mathrm{Cov}[\cdot,\cdot]$ are computed cross-sectionally over the asset universe at time t.

The fundamental measure of predictive accuracy is the daily cross-sectional Pearson correlation between the scores and returns.

$$\rho_t = \frac{\mathrm{Cov}(\mathbf{s}_t, \mathbf{r}_t)}{\sqrt{\mathrm{Var}(\mathbf{s}_t)\,\mathrm{Var}(\mathbf{r}_t)}} \tag{14}$$

For portfolio-based metrics, we simulate a strategy where the daily portfolio return, R_t^p, is constructed from the factor scores $\mathbf{s}_t$. The cumulative wealth over a period of T time steps is given by $W_T = \prod_{t=1}^{T}(1 + R_t^p)$. Let A be the number of trading periods in a year, e.g., $A = 252$ for daily data, and let $r_{t,f}$ be the risk-free rate.

Information Coefficient The Information Coefficient (IC) measures the time-series average of the daily cross-sectional correlations, quantifying the mean predictive power of the factor.

$$\mathrm{IC} = \mathbb{E}_t[\rho_t] \tag{15}$$

Information Ratio The Information Ratio (IR) evaluates the consistency of the factor's predictive power. It is the time-series mean of the daily correlations scaled by their standard deviation, thus rewarding stable performance.

$$\mathrm{IR} = \frac{\mathbb{E}_t[\rho_t]}{\mathrm{Std}_t(\rho_t)} \tag{16}$$

Rank Information Coefficient The Rank Information Coefficient (Rank IC) is the time-series average of the daily Spearman rank correlations. It assesses the model's ability to correctly order assets by predicted performance and is robust to outliers and non-linear monotonic relationships.

$$\text{Rank IC} = \mathbb{E}_t[\mathrm{Corr}(\mathrm{rank}(\mathbf{s}_t), \mathrm{rank}(\mathbf{r}_t))] \tag{17}$$

Rank Information Ratio The Rank Information Ratio (Rank IR) measures the stability of the factor's ordinal predictive power over time.

$$\text{Rank IR} = \frac{\mathbb{E}_t[\mathrm{Corr}(\mathrm{rank}(\mathbf{s}_t), \mathrm{rank}(\mathbf{r}_t))]}{\mathrm{Std}_t(\mathrm{Corr}(\mathrm{rank}(\mathbf{s}_t), \mathrm{rank}(\mathbf{r}_t)))} \tag{18}$$

[1] https://www.wind.com.cn/

[2] https://www.alphavantage.co/

[3] https://data.eastmoney.com/report/

Annualized Return The Annualized Return (AR) represents the total return of the factor-induced portfolio scaled to a one-year period.

$$\text{AR} = A \cdot \mathbb{E}_t[R_t^p] \tag{19}$$

Maximum Drawdown The Maximum Drawdown (MDD) quantifies the largest peak-to-trough decline in the portfolio's cumulative wealth, serving as a critical measure of tail risk.

$$\text{MDD} = \max_{t \le T} \left(1 - \frac{W_t}{\max_{u \le t} W_u}\right) \tag{20}$$

Sharpe Ratio The annualized Sharpe Ratio (SR) measures the risk-adjusted excess return of the portfolio. It is defined as the annualized ratio of the average excess return to the standard deviation of excess returns.

$$\text{SR} = \frac{\mathbb{E}_t[R_t^p - r_{t,f}]}{\text{Std}_t(R_t^p - r_{t,f})} \cdot \sqrt{A} \tag{21}$$

A.7 Baseline Details

We conduct experiments on various SOTA baselines:

Machine Learning Approaches

- LightGBM (Ke et al., 2017): A high-efficiency, leaf-wise gradient boosting decision tree framework by Microsoft research. We utilize the official python package [4] to implement it.

Deep Learning Approaches

- **DTML**: DTML (Yoo et al., 2021) presents an attention-based architecture designed to capture inter-stock correlations for investment decision-making. Our implementation is based on the official code[5].
- **FactorVAE**: FactorVAE (Duan et al., 2022) is a probabilistic dynamic factor model that leverages a variational autoencoder framework. The implementation is adapted from the open-source repository[6].
- **MASTER**: MASTER (Li et al., 2024b) is a Transformer-based model for stock price forecasting. It models both contemporaneous and temporal stock correlations while integrating market-wide information. We utilize the authors' provided implementation[7].

Agent-Based Approaches

- **SEP**: SEP (Koa et al., 2024) is an agent-based approach where a self-reflective agent, trained with PPO, learns to generate explainable predictions for individual stocks. We build our tests upon the official code[8].
- **TradingAgents**: TradingAgents (Xiao et al., 2025b) is a multi-agent framework that constructs portfolios by simulating collaborative dynamics within trading firms. Our experiments leverage the publicly accessible implementation[9].
- **R&D-Agent(Q)**: R&D-Agent(Q) (Li et al., 2025b) is an LLM-driven framework for collaborative factor model development in quantitative finance. We employ the official implementation[10].

A.8 Backtest Experiment Details

We use an index-enhanced strategy with transaction cost of 1.5e-4 and daily rebalancing, holding the top 20% stocks with equal weights. For the A-share market, we exclude stocks that hit the daily price

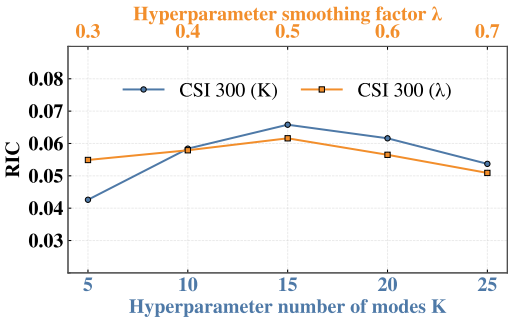
[4] https://github.com/microsoft/LightGBM
[5] https://github.com/ceteris11/DTML
[6] https://github.com/x7jeon8gi/FactorVAE
[7] https://github.com/SJTU-DMTai/MASTER
[8] https://github.com/koa-fin/sep
[9] https://github.com/TauricResearch/TradingAgents
[10] https://github.com/microsoft/RD-Agent

limit or are suspended from trading. For the U.S. market, we exclude suspended stocks. To examine whether MEME's performance depends on high-information-event periods, we partition the CSI 300 backtest into earnings seasons and non-earnings periods, and plot the cumulative excess return over time.

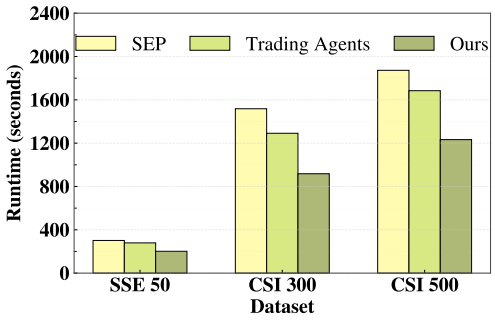


Figure 4: Cumulative excess return on the CSI 300 with earnings seasons highlighted. The upward trend is distributed across both earnings (shaded) and non-earnings periods.

As shown in Figure 4, the cumulative excess return curve trends upward consistently across both earnings and non-earnings windows. Quantitatively, MEME achieves an annualized excess return of 13.89% during earnings seasons and 21.20% during non-earnings periods on the CSI 300. The higher return during non-earnings periods indicates that the framework is not driven by information bursts from earnings releases. This robustness stems from two design choices. First, MEME processes multimodal data (fundamental, technical, and news) simultaneously, generating Investment Arguments that capture both event-triggered and ongoing reasoning. Second, the temporal mode tracking mechanism continuously updates the lifecycle of investment narratives, enabling MEME to adapt to evolving market conditions regardless of event intensity.



(a) Sensitivity of K and λ on RIC.



(b) Average daily running time.

Figure 5: (a) Parameter sensitivity analysis on RIC. (b) Running time comparison of MEME with LLM-based baselines.

A.9 Detailed examples of effective modes

A.9.1 Classification of modes

- Long-term effective: Across both bull and bear markets, the selected mode exhibits positive values in Equation 10 for over 65% of the observed periods.
- Bull-market effective: the selected mode is not long-term effective, and exhibits positive values in Equation 10 for over 70% of the bull-market periods. As shown in the examples below, this mode centers on technical momentum and capital flows, resembling the trend-following philosophy of Livermore (Lefevre & Markman, 2010).
- Bear-market effective: the selected mode is not long-term effective and bull-market effective, and exhibits positive values in Equation 10 for over 70% of the bear-market periods. As shown in the examples below, this mode centers on dividend yield, cash flow, and balance-sheet strength, resembling the value investing principles of Graham (Graham & Zweig, 2003).
- Non-effective: None of the above.

A.9.2 LONG-TERM EFFECTIVE MODES

Modes example & summary

Modes examples


```
{"converted_arguments": [
{"code": "000001.SZ",
"name": "Ping An Bank",
"date": "2024-01-30",
"argument": {
"p": 1, "a": "The stock is significantly undervalued with
substantial potential for mean reversion.", "e": "Based on a relative
valuation and historical quantile analysis framework, the current P/E
ratio of 3.80 is only above the 2.8th percentile of its historical
data, far below the median of 8.84. Similarly, the P/B ratio of
0.40 is at the 2.8th percentile, below the median of 0.82, indicating
significant room for valuation recovery."
}},

{"code": "000895.SZ",
"name": "Henan Shuanghui Investment & Development", "date":
"2024-06-25",
"argument": {
"p": 1,
"a": "The inventory management strategy is poised to enhance
profitability.",
"e": "Based on the company's statement on optimizing frozen goods
inventory to maximize returns, combined with industry data showing
a month-over-month contraction in hog production, this cost control
optimization is expected to directly improve the gross profit margin."
}},

{ "code": "603259.SH",
"name": "WuXi AppTec",
"date": "2024-07-08",
"argument": { "p": 1,
"a": "Utilizing a model that links buyback activities with
shareholder confidence, the cancellation of stock options sends a
positive signal.",
"e": "WuXi AppTec's announcement on July 1st to cancel 31,500 stock
options, viewed in the context of over 1,600 A-share companies
conducting nearly 100 billion RMB in buybacks in the first half of
the year, reflects the management's proactive stance on supporting the
company's valuation."
}},

{ "code": "000725.SZ",
"name": "BOE Technology Group A",
"date": "2025-05-27",
"argument": {
"p": -1,
"a": "Intensifying industry competition and capital expenditure
pressures are likely to weigh on short-term performance.",
"e": "A competitive landscape analysis reveals that competitors like
HKC and Tianma have signed new display projects totaling 72 billion
RMB for mass production. This acceleration in industry capacity
expansion means BOE Technology will face heightened challenges from
price competition and market integration."
}},

{"code": "601138.SH",
"name": "Foxconn Industrial Internet",
"date": "2025-04-12",
"argument":
```

```
"p": 1,
"a": "Earnings growth is the primary driver for stock price
appreciation.",
"e": "According to a performance-driven analytical framework,
the company's net profit for the first quarter grew by 24%-27%
year-over-year, with revenue increasing by 34%-35.6%, setting a
new record for the period. This high-growth data substantiates
expectations for an upward trajectory in the stock price."
}}
],}
```


Modes summaries

The overarching mode of thought is a multi-faceted, evidence-based reasoning process. It synthesizes diverse data from quantitative metrics and corporate actions to industry trends--through specific analytical frameworks to construct a structured investment thesis on a stock's future performance.

A.9.3 BULL-MARKET EFFECTIVE MODES

Modes example & summary

Modes examples


```
{"converted_arguments": [
{"code": "000063.SZ",
"name": "ZTE Corporation",
"date": "2024-01-30",
"argument": {
"p": -1, "a": "The trend has deteriorated on a weekly basis,
confirming the continuation of a mid-term downtrend.",
"e": "The weekly closing price of 20.78 has fallen below the
Bollinger Band middle rail of 25.73, the RSI has dropped sharply to
30.9, and the negative MACD value has widened to -0.62."
}},

{"code": "688256.SH",
"name": "Cambricon Technologies", "date": "2025-07-30",
"argument": {
"p": 1,
"a": "A monthly-level breakout above the upper Bollinger Band,
coupled with an overbought RSI, indicates strong upward momentum.",
"e": "Based on trend continuation logic, the recent monthly closing
price of 1492.49 is significantly above the upper Bollinger Band of
1127.98, and the monthly RSI has reached 84.26, confirming a clear
breakout pattern." }},

{ "code": "600519.SH",
"name": "Kweichow Moutai",
"date": "2024-05-06",
"argument": { "p": 1,
"a": "The stock is exhibiting a strong short-term technical breakout,
suggesting the price is likely to continue rising.",
"e": "Based on a bullish moving average alignment framework, the
daily closing price has stayed above all short-term moving averages
(5-day, 10-day, 20-day) for three consecutive days, and the 5-day
MA has crossed above the 10-day and 20-day MAs, indicating strong
short-term upward momentum."
}},

{ "code": "688981.SH",
"name": "SMIC",
"date": "2025-07-30",
"argument": {
```

```
"p": 1,
"a": "The short-term stock price is poised to benefit from
institutional capital inflows and positive market sentiment.",
"e": "Margin financing net purchases on July 29th reached a two-month
high of 164 million RMB, and the margin financing balance ratio rose
to 4.01% on July 30th. Concurrently, recommendations from Minsheng
Securities and Everbright Securities as a top pick for August signal
strong institutional optimism."
}},

{"code": "603986.SH",
"name": "GigaDevice Semiconductor",
"date": "2025-04-12",
"argument":
"p": 1,
"a": "The stock price is likely to increase, as inferred from
financing activities and capital flow dimensions.",
"e": "A net margin financing purchase of 65.38 million RMB on June
30th marked the fourth consecutive day of net purchases, indicating
active capital inflow from the market and reflecting strengthening
investor confidence."
}}
],}
```


Modes summaries

This thinking mode prioritizes technical analysis and market sentiment. It interprets chart patterns, momentum indicators (RSI, MACD), and capital flows (margin financing) as direct predictors of short-term price movements, inferring investor conviction and forecasting trend continuation or reversal based on these signals.

A.9.4 BEAR-MARKET EFFECTIVE MODES

Modes example & summary

Modes examples


```
{"converted_arguments": [
{"code": "600941.SH",
"name": "China Mobile",
"date": "2024-01-30",
"argument": {
"p": 1, "a": "An attractive dividend yield provides downside
protection and enhances investment appeal.",
"e": "Utilizing a yield analysis framework, the dividend yield
stands at 4.15%, surpassing the industry average of 2.44%. The
company's abundant cash flow, with cash and equivalents consistently
exceeding 3.3 trillion RMB and no interest-bearing debt, supports the
sustainability of its dividend policy."
}},

{"code": "601816.SH",
"name": "Beijing-Shanghai High-Speed Railway", "date": "2024-06-27",
"argument": {
"p": 1,
"a": "The implementation of a cash dividend may attract stable,
yield-seeking capital inflows.",
"e": "Based on a dividend event-driven analysis, the company's
execution of a cash dividend of 1.12 RMB per 10 shares, with the
record date completed on June 27th, highlights its prominent
high-dividend characteristic." }},

{ "code": "688981.SH",
"name": "SMIC",
"date": "2024-05-12",
```

```
"argument": { "p": -1,
"a": "Deteriorating cash flow combined with a lack of dividends
diminishes investment attractiveness and may lead to valuation
pressure.",
"e": "Through P/CF and dividend yield analysis, the company's cash
flow return is insufficient. The P/CF ratio of 14.55 is above its
historical 36th percentile while cash and equivalents have declined.
The 0% dividend yield presents a clear disadvantage compared to the
industry's average yield."
}},

{ "code": "601012.SH",
"name": "LONGi Green Energy",
"date": "2025-04-24",
"argument": {
"p": -1,
"a": "The balance sheet reveals growing financial pressure, which
constrains potential stock price appreciation.",
"e": "Cash and equivalents have decreased while total debt remains
high, causing the debt-to-asset ratio to rise to 0.592. Persistently
high levels of accounts receivable and inventory signal pressure on
operational efficiency."
}},

{"code": "600900.SH",
"name": "China Yangtze Power",
"date": "2025-08-27",
"argument":
"p": 1,
"a": "The stock price is likely to increase, as inferred from
financing activities and capital flow dimensions.",
"e": "Applying dividend and cash flow analysis, the current dividend
yield of 3.41% is above the industry's 2.67%, and a P/CF ratio of
11.38 below its historical median indicates a reasonable valuation.
An improving financial structure, seen in the reduced debt-to-asset
ratio, enhances dividend sustainability."
}}
],}
```


Modes summaries
This thinking mode prioritizes financial health and shareholder returns. It assesses a company's intrinsic value and risk profile by analyzing balance sheet strength, cash flow generation, and dividend policies, viewing a strong dividend and stable financials as key indicators of investment quality and downside protection.

A.10 Implementation details of Argument Extraction

A.10.1 Implementation details of Filter Agent

System & User Prompts

System prompts
You are an expert on financial market analysis, portfolio management and quantitative trading. Strictly follow the user's requirements to generate sound, logical and detailed responses. Your output should only contain a JSON object. Do not include any explanations, additional text, notes, code block markers like "` or "`json.
User prompts for Information Filtering.
Context: You are part of an automated investment analysis framework. Your analysis will serve as a primary input for a downstream synthesis agent. The goal is to reduce noise and provide a clear, factual summary of the current situation for **[Asset Name] ([Asset Ticker])** as of **[Analysis Date]**.

Input Data:
You will be provided with raw [Modality Name] data. [Raw Data for the specified modality, e.g., a table of fundamental ratios, a list of recent news headlines, or a time-series of technical indicators]
Task: 1. **Analyze:** Thoroughly review the provided raw data.
2. **Identify Salience:** Identify the most salient points, patterns, and anomalies that could materially impact the asset's future price performance. Disregard trivial or redundant information.
3. **Summarize:** Distill your findings into a concise, structured summary. The summary must be objective and based strictly on the provided data.
Constraints:
1. Focus exclusively on the provided data. Do not introduce external information or make speculative predictions.
2. The output must be a factual distillation, not a final investment recommendation.
3. Present the information in a clear, structured format using the specified headings.
Output Format:
{
"Modality_name": [Modality Name],
"Analysis_summary": "...",
"Asset_code": [Asset Ticker]
}

A.10.2 Implementation details of Generator Agent

System & User Prompts

User prompts for Argument Generation.
Context: You have received concise, pre-filtered intelligence reports for [Asset Name] ([Asset Ticker]) as of [Analysis Date]. Your task is to look for confluence or divergence between these reports to construct distinct lines of reasoning for either a bullish or bearish stance.
Input Data:
You will be provided with a collection of structured summaries from different analytical perspectives.
1. **Fundamental Analysis Summary:**[fundamental_output]
2. **Technical Analysis Summary:**[technical_output]
3. **Narrative Analysis Summary:**[news_output]
Task:
1. **Synthesize:** Review all provided analysis summaries. Identify distinct lines of reasoning by connecting facts across different reports. A single fact is not an argument; a synthesis of facts is.
2. **Formulate Arguments:** For each distinct line of reasoning, construct a structured investment argument. An argument can be supported by evidence from one or more of the analytical summaries.
3. **Assign Stance:** Clearly define each argument as either bullish or bearish.
Constraints:
1. The output must be a list of structured JSON objects.
2. Each argument must be a self-contained, logical thesis.
3. The 'rationale' should be your own synthesized conclusion, not a direct copy of the input.
4. The 'evidence' must be a direct quote or a close, factual paraphrase of the key observations from the input summaries that support your rationale.
Output Format:
Produce a valid JSON array of argument objects. Each object must conform to the following schema:
{
"p": 1 for bullish. -1 for bearish,

```
"a":  "argument analytical rationale",
"e":  "argument evidence"
}
```