FinKario: Event-Enhanced Automated Construction of Financial Knowledge Graph

Xiang Li[1]*, **Penglei Sun[1]***, **Wanyun Zhou[1]**,
Zikai Wei[2], **Yongqi Zhang[1]†**, **Xiaowen Chu[1]†**
[1]The Hong Kong University of Science and Technology (Guangzhou)
[2]International Digital Economy Academy
{xli906, psun012, yongqizhang, xwchu}@hkust-gz.edu.cn

Abstract

Individual investors are disadvantaged in financial markets, overwhelmed by abundant information and lacking professional analysis. Equity research reports are crucial resources, offering valuable insights. By leveraging these reports, large language models (LLMs) can enhance investors' decision-making and strengthen financial analysis. However, two key challenges limit their effectiveness: (1) the rapid evolution of market events outpaces the slow update cycles of existing knowledge bases, and (2) the long-form, unstructured nature of financial reports hinders timely, context-aware integration by LLMs. To address these challenges, we tackle both data and methodology aspects. We introduce the Event-Enhanced Automated Construction of Financial Knowledge Graph **(FinKario)**, a dataset with over 305,360 entities, 210,328 relational triples, and 19 relation types. FinKario integrates real-time company fundamentals and events through prompt-driven extraction guided by institutional templates, providing structured, accessible financial insights for LLMs. We further propose a Two-stage, Graph-based retrieval strategy **(FinKario-RAG)** to optimize retrieval over evolving, large-scale financial knowledge. Experiments show that FinKario with FinKario-RAG achieves superior trend prediction accuracy, outperforming financial LLMs by **18.81%** and institutional strategies by **17.85%** on average in backtesting. [1]

1 Introduction

The financial market is dominated by institutional investors, leaving individual investors at a significant disadvantage (Davis, 1996; Schlachter, 2013). Individual investors often struggle to make informed decisions due to limited access to professional-grade analysis (Hilton, 2001). Equity research reports, which provide expert insights into market trends and company performance, serve as a critical resource to bridge this gap (Zhou et al., 2024; Kapellas and Siougle, 2017; Bian et al., 2026). LLMs can reason over complex financial text and interpret context (Yang et al., 2023; Wu et al., 2023; Araci, 2019; Tang et al., 2026a), making them well-suited for analyzing research reports and enabling timely, scalable insights. Recent work at the intersection of data mining and financial trading (Wang et al., 2025; Zhang et al., 2024; Tang et al., 2026c) further underscores their potential in this domain.

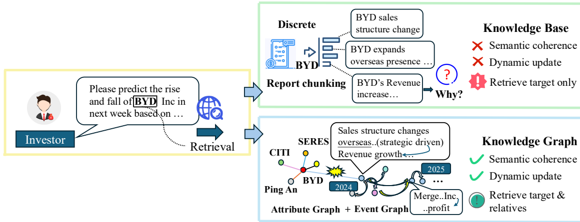


Figure 1: Comparison of Traditional Single-research Report Retrieval vs. FinKario Retrieval.

However, LLM-based research report analysis faces two primary challenges: **Ch1**, the rapid evolution of financial events (Cheng et al., 2020; Xu et al., 2025; Bian et al., 2026), which outpaces the slow update cycles of existing knowledge bases, as shown in Table 1; and **Ch2**, the inherent complexity of processing long-form, unstructured data (Xia et al., 2024; Sarmah et al., 2024; Hu et al., 2026a; Yu et al., 2026; Hu et al., 2026b). Events such as earnings releases, product launches, and regulatory changes are key drivers of market behavior and are essential for understanding temporal shifts in asset performance (MacKinlay, 1997; Thompson, 1995). Financial markets generate a continuous stream of such events, often reflected in research reports, yet both traditional knowledge bases and

*Equal contribution
†Corresponding author
[1]Our code is available at https://github.com/Jackson906E/FinKario.

Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 9826–9845
July 2-7, 2026

modern LLMs struggle to keep pace. As illustrated in Figure 1 and Table 1, two key limitations persist. **Lim1**: Current knowledge-base approaches rely on static, target-only retrieval and report chunking, which lack semantic coherence, provide limited contextual explanations (i.e., "why"), and cannot incorporate evolving market signals in real time. **Lim2**: Existing event-centric Financial Knowledge Graphs (FKGs) still depend on manual or semi-automated construction pipelines (Zhu et al., 2023; Li and Sanna Passino, 2024; Kertkeidkachorn et al., 2023; Colombo et al., 2025; Dong et al., 2025; Zhu et al., 2025), leading to outdated representations in dynamic environments. Although LLMs excel at processing natural language, their internal knowledge is static and infrequently updated, resulting in persistent knowledge lag (Singh et al., 2024; Adewale et al., 2023; Tang et al., 2026b).

To address these issues, we propose **FinKario**, a dual-structured financial knowledge graph built from equity research reports that automatically and dynamically captures financial attributes, indicators, and events (**Ch1**). It comprises two subgraphs: an **attribute subgraph** for stable fundamentals and an **event subgraph** for time-sensitive events such as quarterly performance and key profitability drivers (**Ch2**). The construction pipeline begins with automated schema generation for both subgraphs, using prompt-driven extraction guided by professional institutional frameworks such as the CFA handbook (**Lim2**). The event schema follows a top-down structure derived from academic reports from the University of Wisconsin and is refined into a detailed event ontology based on the Financial Industry Business Ontology (FIBO). Based on these schemas, LLMs extract structured knowledge from research reports using domain-specific templates. A quality-control module then improves reliability by correcting erroneous or outdated information, normalizing entities, and completing missing attributes with support from the Tushare platform. We collect a corpus of research reports from August 2024 to March 2025. The resulting FinKario instance contains over 305,360 entities, 210,328 relational triples, and 19 relation types. To further tackle retrieval over dynamically evolving, large-scale financial knowledge, we propose **FinKario-RAG**, which first retrieves information directly related to the queried entity and then expands to related entities and relationships, providing holistic context essential for accurate predictions (**Lim1**).

To evaluate FinKario, we conduct backtests comparing predictive performance against powerful LLMs and institutional strategies. On average, FinKario combined with FinKario-RAG outperforms financial LLMs by 18.81% and institutional strategies by 17.85% in predictive accuracy. Ablation studies further show that FinKario-RAG surpasses mainstream retrieval methods by an average of 12.70% in predictive accuracy. The principal contributions of this paper are as follows:

- We introduce **FinKario**, a dynamic, event-driven financial knowledge graph with over 305,360 entities, 210,328 relational triples, and 19 relation types, supporting automated updates without manual intervention or predefined domain knowledge, and enabling professional template–driven schema construction.
- We propose **FinKario-RAG**, a retrieval strategy that integrates industry-level and index-level perspectives to overcome the limitations of single-target retrieval, enabling holistic, realistic analysis aligned with practical investment scenarios and supporting retrieval over large-scale, dynamically evolving financial knowledge.
- We empirically validate FinKario-RAG through extensive experiments, showing that our method surpasses the runner-up by **58.14%** in Sharpe ratio, **30.86%** in accumulative rate of return, and **1.04%** in predictive accuracy, demonstrating its effectiveness for financial analysis and stock trend forecasting.

2 Related Work

2.1 Financial Knowledge Graph

In recent years, Financial Knowledge Graphs (FKGs) have attracted growing attention for improving financial data analysis and decision-making. Early work mainly relied on standard natural language processing techniques, including semantic recognition, classification, and Named Entity Recognition (NER). For instance, Wang et al. (Wang et al., 2021) proposed datasets and evaluations for constructing financial knowledge graphs, enabling comprehensive assessment. Chen et al. (Chen et al., 2024b) introduced a retrieval-augmented framework that incorporates structured FKGs into language models to improve reliability and accuracy in market analysis and report generation. Sun et al. (Sun et al., 2025) proposed a knowledge-enhanced prompt learning framework for financial news recommendation, using FKGs to provide context-aware, accurate recommendations. More recently, Arun et al. (Arun et al., 2025) proposed an agentic and reflection-driven framework for large-scale FKGs construction. However, most existing methods still depend on predefined schemas and manual input, underscoring the need for more automated and schema-independent FKGs construction.

2.2 Automatic Knowledge Graph Construction

Recent work on Automatic Knowledge Graph Construction (AKGC) with LLMs increasingly aims to reduce manual schema engineering. Zhang et al. (Zhang and Soh, 2024) proposed EDC, which decouples extraction, schema definition, and canonicalization, allowing schemas to be pre-defined or self-generated. Ding et al. (Ding et al., 2024) introduced TKGCon, leveraging Wikipedia-derived ontologies, existing theme-specific KGs, and LLM-generated relation sets to build fine-grained, timely KGs. SAC-KG (Chen et al., 2024a) further integrated generation, verification, and pruning to iteratively construct high-precision domain KGs. In contrast, Su et al. (Su et al., 2020) designed a rule-based framework over relational databases for power systems, relying on static schemas and device metadata. Despite these advances, most methods either expand pre-existing graphs or use prompt-based extraction without grounding in authoritative domain templates. This gap is particularly evident in finance, where consistency and interpretability are crucial, yet AKGC approaches rarely incorporate professional, domain-specific schemas.

2.3 LLMs in Finance

The use of Large Language Models (LLMs) in finance has progressed through several stages. Early work emphasized domain-specific pretraining. Yang et al. (Yang et al., 2023) introduced FinGPT, an open-source alternative emphasizing real-time market adaptability. Subsequent research expanded the functional scope of financial LLMs. Li et al. (Li et al., 2024) proposed Alphafin, a retrieval-augmented stock-chain framework that dynamically integrates financial data from multiple sources. Recent innovations target multimodal understanding and advanced training. Gan et al. (Gan et al., 2024) introduced MME-Finance, a benchmark for cross-modal comprehension of financial texts, tables, and charts. Fin-R1 (Liu et al., 2025) further improves financial reasoning via high-quality CoT datasets, supervised fine-tuning, and reinforcement learning. Overall, the field is moving toward frameworks that combine retrieval, multi-agent collaboration, and multimodal reasoning, while challenges remain in temporal reasoning and explainability under dynamic market conditions.

3 Methodology

3.1 FinKario Construction

To support structured interpretation of financial narratives, we adopt a dual–schema design with an **Attribute Graph** and an **Event Graph**, generated via a schema-guided extraction function:

$$\mathcal{F} : (\mathcal{D} \times \mathcal{S}) \rightarrow \mathcal{G},$$

where $\mathcal{F}$ maps a document corpus $\mathcal{D}$ and schema $\mathcal{S}$ into a structured graph $\mathcal{G}$.

Table 1 presents our knowledge graph alongside existing financial knowledge graphs. Compared to the existing ones, which rely on manually defined schemas and lack comprehensive knowledge updates throughout the process. To construct these graphs in a fully automated manner, we introduce **FinKario**, a dataset built upon four primary modules.

3.1.1 Domain Corpus Acquisition

We collect raw equity research reports from East Money[2] and convert each report into standardized Markdown using MinerU[3] (Wang et al., 2024). A refinement step removes non-informative content such as disclaimers, images, and repetitive legal statements, yielding a clean corpus $\mathcal{D}'$ for downstream processing.

3.1.2 Schema Construction

We construct two complementary schemas: the Attribute Graph Schema $\mathcal{S}_A$ and the Event Graph Schema $\mathcal{S}_E$.

• **Schema for Attribute Graph.** We use standardized equity research templates from CFA Institute[4] and J.P. Morgan[5] as structural priors, denoted

[2] https://www.eastmoney.com

[3] https://github.com/opendatalab/MinerU

[4] https://www.cfainstitute.org/sites/default/files/-/media/documents/support/research-challenge/challenge/rc-equity-research-report-essentials.pdf

[5] https://www.wallstreetprep.com/knowledge/sample-equity-research-report/

Table 1: Comparison of key characteristics of financial knowledge graph construction methods, including events, dynamics, automation, major sources, and the number of entities, relations, and triples.

Knowledge	Event	Dynamic Updated	Automation	Entities	Relations	Triples	Major Source
FR2KG (Wang et al., 2021)	✗	✗	✗	17,799	13	1,328	Research Report
KGEEF (Cheng et al., 2020)	✓	✗	✗	5,262,423	/	325,786	News
T-FinKB (Zhu et al., 2023)	✗	✓	✗	3,974	16	/	News
FinKG (Kertkeidkachorn et al., 2023)	✗	✗	✗	37,382,905	12	30M+	Market data
FEEKG (Liu et al., 2024)	✓	✗	✗	112,000	12	/	News
FinDKG (Li and Sanna Passino, 2024)	✗	✓	✗	13,645	15	/	News
FMAG (Chen et al., 2024b)	✗	✓	✗	8,052	5	5,664	Research Report
HiDy (Zhu et al., 2023)	✗	✓	✗	51,000	34	505,800	Market data
FNRKPL (Sun et al., 2025)	✗	✗	✗	11,432	6	52,384	News
EKG (Colombo et al., 2025)	✗	✗	✗	/	9	/	Market data
FinRipple (Xu et al., 2025)	✓	✗	✗	/	4	/	News
FinKario (Ours)	✓	✓	✓	305,360	19	210,328	Research Report

θ_{CFA} and θ_{JPM}. Prompting the LLM with these templates produces a set of attribute-level relation types: $\mathcal{S}_A = \text{LLM}(\text{Prompt}_{\text{attr}}; \theta_{\text{CFA}}, \theta_{\text{JPM}})$ covering relations such as *Industry*, *Risk Factors*, and *Exchange*. Figure 2(b) illustrates $\mathcal{S}_A$.

• **Schema for Event Graph.** We construct a hierarchical event schema $\mathcal{S}_E$ using a top-down approach. At the first level, we generate high-level driven categories by prompting the LLM based on the institutional template θ_{WIS} from the University of Wisconsin: $\mathcal{C} = \text{LLM}(\text{Prompt}_{\text{cat}}; \theta_{\text{WIS}})$, where $\mathcal{C} = \{c_1, c_2, \ldots, c_m\}$ represents the set of high-level event categories. For each category $c_i \in \mathcal{C}$, we construct prompts grounded in the Financial Industry Business Ontology (FIBO) [6], denoted as $\mathcal{O}_{\text{FIBO}}$, to generate corresponding low-level event ontology: $\mathcal{O}_{c_i} = \text{LLM}(\text{Prompt}_{\text{event}}; c_i, \mathcal{O}_{\text{FIBO}})$. The resulting schema is defined as: $\mathcal{S}_E = \bigcup_{i=1}^{m} \{(c_i, o) \mid o \in \mathcal{O}_{c_i}\}$. Figure 3 visualizes the tree-structured event schema.

3.1.3 Knowledge Population

For each refined Markdown document $\mathcal{D}'$, entities are extracted at each timestamp $\tau \in T$ via a dedicated prompt guided by the previously formed schema $\mathcal{S}_A$ and $\mathcal{S}_E$:

$$\mathcal{E}_{\mathcal{A}\tau}; \mathcal{E}_{\mathcal{E}\tau}, \mathcal{R}_{\mathcal{E}\tau} = \text{LLM}(\text{Prompt}, D', \mathcal{S}_A, \mathcal{S}_E, \tau),$$

where $\mathcal{E}_{\mathcal{A}\tau}$ and $\mathcal{E}_{\mathcal{E}\tau}$ denote the set of extracted entities at timestamp τ. These timestamped entity sets $\mathcal{E}_{\mathcal{A}\tau}$ and relation types $\mathcal{R}_A$ are combined to form the attribute knowledge graph for each stock s:

$$\mathcal{G}_A^{(s)} = \bigcup_{\tau \in T} \{(e_h, r, e_t, \tau) \mid e_h, e_t \in \mathcal{E}_{A,\tau},\ r \in \mathcal{R}_A\},$$

where e_h and e_t denote head and tail entities, respectively.

The Event Knowledge Graph is constructed as

$$\mathcal{G}_E^{(s)} = \bigcup_{\tau \in T} \{(e_s, r', e_o, \tau) \mid e_s, e_o \in \mathcal{E}_{E,\tau},\ r' \in \mathcal{R}_{E,\tau}\},$$

where e_s and e_o are subject and object entities and r' denotes event-trigger relations.

The final financial knowledge graph for stock s integrates both components:

$$\mathcal{G}_{\text{FinKario}}^{(s)} = \mathcal{G}_A^{(s)} \cup \mathcal{G}_E^{(s)},$$

capturing both structured financial information along with inferred event interactions. Figure 4 illustrates the daily temporal structure of FinKario.

3.1.4 Quality Control Refinement

To ensure the reliability of the constructed knowledge graph, we implement a refinement module that addresses common issues in financial text extraction, including entity ambiguity, missing numeric values, and extraction errors. Specifically, the module performs entity normalization, attribute completion via the Tushare platform[7], and error or placeholder correction via the LLM. The complete refinement pipeline is detailed in Algorithm 1 in the appendix, and yields the refined knowledge graph $\mathcal{G}'^{(s)}_{\text{FinKario}}$.

3.2 FinKario-RAG

The Two-stage Graph-based retrieval augmented generation pipeline (FinKario-RAG) converts $\mathcal{G}'^{(s)}_{\text{FinKario}} = (\mathcal{E}, \mathcal{R})$ into actionable investment advice through three interlocking modules, as illustrated in Fig. 2.

3.2.1 Knowledge Graph Vectorization & Ingestion

To support semantic retrieval, the event-augmented financial knowledge graph $\mathcal{G}'^{(s)}_{\text{FinKario}}$ is vectorized

[6] https://spec.edmcouncil.org/fibo/ontology

[7] https://tushare.pro

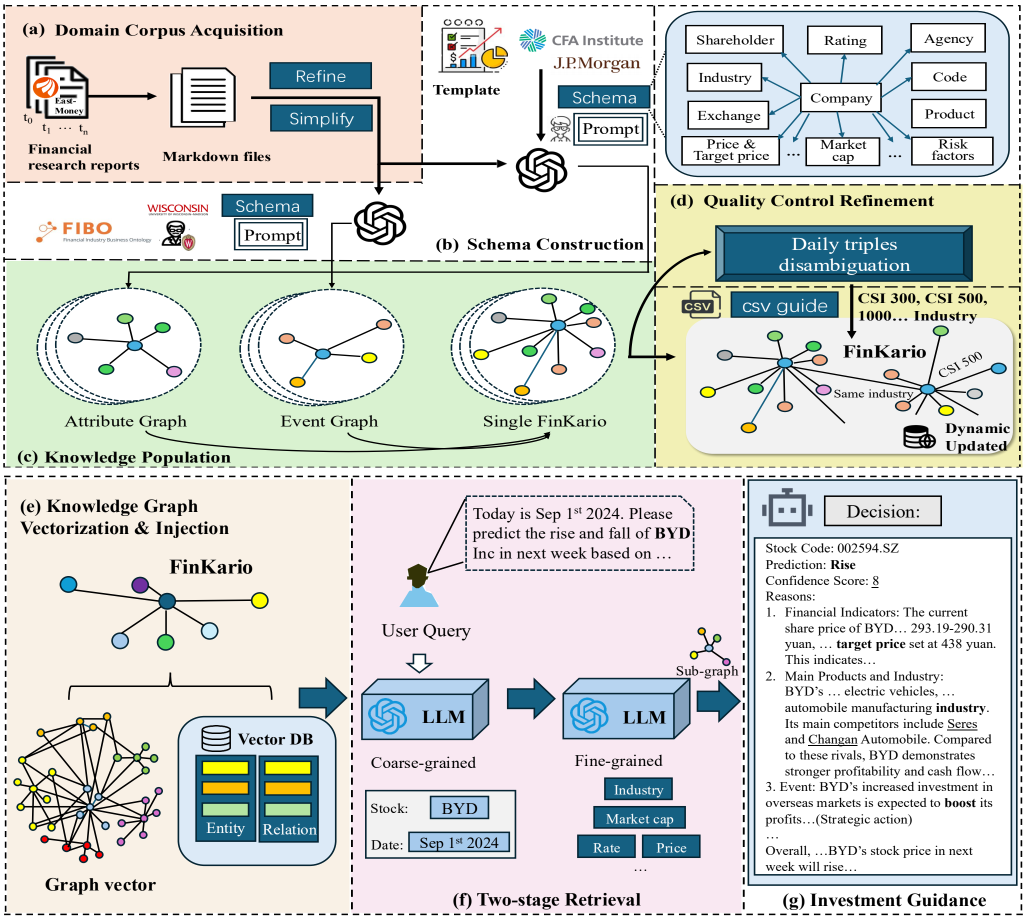


Figure 2: The overall framework of FinKario and FinKario-RAG: (a)–(d) The construction process of FinKario; (e)–(g) Details of the FinKario-RAG pipeline.

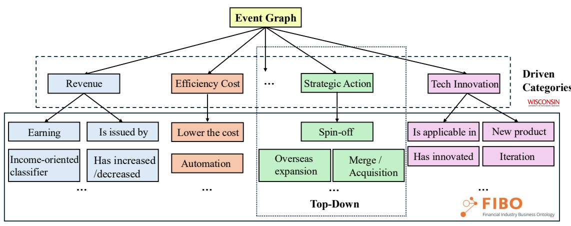


Figure 3: Tree-Structured Schema for Event Graph.

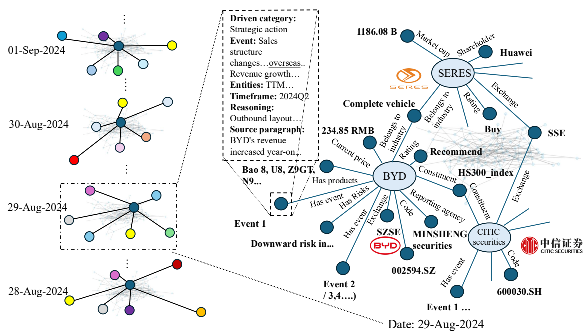


Figure 4: Temporal Visualization of FinKario.

into three components: entity-level, relation-level, and graph-level representations.

• **Entity and Relation Embedding.** We encode all entities and relations using a graph encoder Φ to obtain the structural embedding set:

$$\mathbf{Z}_{\text{local}} = \Phi(\mathcal{G}'^{(s)}_{\text{FinKario}}) = \{\mathbf{e}_i\}_{i=1}^{|\mathcal{E}|} \cup \{\mathbf{r}_j\}_{j=1}^{|\mathcal{R}|},$$

where $\mathbf{e}_i$ and $\mathbf{r}_j$ denote the latent vectors of the i-th entity and j-th relation, respectively.

• **Graph-level Embedding.** To capture global context and topological semantics, we also compute a graph-level vector using a readout function ρ, where $\mathbf{g}_{\text{global}} = \rho(\mathcal{G}'^{(s)}_{\text{FinKario}})$.

• **Vector Store Indexing.** The unified representation $\mathbf{Z}_{\text{FinKario}} = \mathbf{Z}_{\text{local}} \cup \{\mathbf{g}_{\text{global}}\}$ is normalized and stored in the vector database $\mathbb{V}$, which supports efficient maximum inner product search during downstream retrieval.

3.2.2 Two-stage Retrieval

Given a user query q, the system first encodes it into a dense vector $\mathbf{h}_q = \Psi(q)$ via a language model encoder Ψ. Retrieval is performed in two stages:

• **Coarse-grained Retrieval.** The first stage aims to identify rough semantic anchors such as relevant stocks and dates. This is achieved by matching $\mathbf{h}_q$ against indexed stock and date representations in the vector store:

$$\mathbb{V}_{\text{coarse}} = \mathcal{R}_{\text{coarse}}(\mathbf{h}_q, \mathbb{V}, k_c),$$

returning the top-k_c coarse candidates (e.g., `BYD`, Sep 1st 2024).

• **Fine-grained Retrieval.** Building on the coarse results, a finer-grained retrieval is conducted to collect surrounding financial entities—such as `industry`, `market cap`, and `price` by searching over relevant portions of the vector set:

$$\mathbb{V}_{\text{fine}} = \mathcal{R}_{\text{fine}}(\mathbf{h}_q, \mathbb{V}_{\text{coarse}}, k_f),$$

where k_f represents the number of related entities from fine-grained process. To facilitate structured reasoning, the retrieved fine-level vectors are mapped back to their original graph context to reconstruct a semantically aligned subgraph:

$$\mathcal{G}_{\text{sub}} = \texttt{Mapping}(\mathbb{V}_{\text{fine}}), \quad \mathcal{G}_{\text{sub}} \subseteq \mathcal{G}'^{(s)}_{\text{FinKario}}.$$

where $\texttt{Mapping}(\cdot)$ aligns vector-retrieved entities with their corresponding nodes and edges in $\mathcal{G}'^{(s)}_{\text{FinKario}}$, yielding an adaptive subgraph $\mathcal{G}_{\text{sub}}$ centered on the queried stock and its contextual relations. This Two-stage retrieval enables FinKario-RAG to construct a compact, semantically aligned subgraph that bridges user intent with local financial context for downstream reasoning.

3.2.3 Investment Guidance

The subgraph $\mathcal{G}_{\text{sub}}$ and the user query q are jointly fed into the final reasoning model:

$$y = \text{LLM}_{\text{Analyst}}(q, \mathcal{G}_{\text{sub}}),$$

where y includes a predicted movement label (e.g., `Rise` or `Fall`), an associated confidence level, and a textual rationale grounded in the retrieved knowledge. This completes the FinKario-RAG pipeline by converting graph-derived evidence into interpretable and actionable investment guidance.

4 Experiment Results

4.1 Experiment Setup

Given a universe of stocks $\mathcal{S}$, for any stock $s \in \mathcal{S}$ on a given trading day t, we evaluate a long-only trading strategy driven by FinKario-RAG signals. The strategy operates as follows: (1) **Signal Generation.** On trading day t, FinKario-RAG generates a signal $\gamma_{s,t}$ for stock s; (2) **Entry Rule.** If $\gamma_{s,t}$ indicates a buy signal, a position is initiated on the next trading day $t+1$ at the closing price $c_{s,t+1}$, introducing a one-day execution lag between publication and trade execution to mitigate information leakage; (3) **Exit Rule.** The position is held until the last trading day of the following week, denoted as $\tau(t)$, at which point the stock is sold at the closing price $c_{s,\tau(t)}$.

4.2 Dataset & Metrics

We evaluate our model using a multi-source dataset that combines textual and financial data. The raw research reports are collected from the East Money website, covering the period from August 28, 2024 to February 28, 2025. Corresponding stock price data for backtesting is obtained from Tushare, spanning from August 28, 2024 to March 07, 2025. In addition, we incorporate index components and industry classification information provided by Wind platform to support graph construction and semantic enrichment.

To evaluate model performance, we adopt six widely used metrics: **Annualized Rate of Return (ARR)**, **Volatility (VOL)**, **Sharpe Ratio (SR)**, **Maximum Drawdown (MDD)**, **Calmar Ratio (CR)**, **Accuracy (ACC)**, and **Sortino Ratio (STO)**. These jointly assess risk-return characteristics and directional predictive accuracy, offering a comprehensive view of the investment performance of FinKario-RAG.

4.3 Baseline

Our proposed approach is evaluated against four categories of baselines:

• **Market Indices.** Market indices are standard passive benchmarks. We report results on several representative indices, including the **CSI 300**, **CSI 500**, **SSE Composite Index**, and **SSE Dividend Index**, which cover major segments of the Chinese market.

• **Vanilla LLMs.** We include general-purpose language models such as Qwen3-8B (Yang et al.,

2025) and GPT-4o-mini (Hurst et al., 2024), which are not specifically tuned for financial tasks.

• **Financial Domain LLMs.** We evaluate several open-source financial language models, including FinMA (Xie et al., 2023), FinGPT (Yang et al., 2023), DISC-FinLLM (Chen et al., 2023), XuanYuan-6B [8] and Stock-Chain (Li et al., 2024). These models are tailored for financial forecasting, making them suitable for downstream backtesting and comparison.

• **Financial Institutions.** The selected institutions were chosen as leading brokerages that frequently publish research reports.

4.4 Experimental Results

Figure 5 shows that FinKario-RAG consistently outperforms all benchmark models in cumulative returns, demonstrating the effectiveness of the knowledge graph–enhanced retrieval framework. In late September 2024, most strategies surged following favorable Chinese fiscal policies, before entering a period of pullback and consolidation, during which SOOCHOW remained relatively stable. A major turning point occurred in early February 2025, when many strategies rebounded: Stock-Chain experienced a sharp spike, whereas FinKario-RAG exhibited steady and accelerating growth, ultimately surpassing nearly all competitors by early March. Overall, these results highlight FinKario-RAG's strong adaptability to market shifts and robust performance across volatile periods.

In addition to the visual insights from cumulative NAV trends, Table 2 presents a quantitative comparison that further validates FinKario-RAG's superiority. FinKario-RAG achieves the highest scores in ARR (2.633), SR (4.926), CR (15.315), ACC (0.581), and STO (18.717), while maintaining a moderate Maximum Drawdown (MDD) of 0.172. These results reflect a balanced and effective investment strategy.

In terms of ARR, FinKario-RAG outperforms Guolian-Minsheng (2.012) by 30.8%, SOOCHOW (1.625) by 62.0%, and exceeds Stock-Chain by a significant 123.7%. For risk-adjusted returns, FinKario-RAG's SR, CR, and STO represent improvements of 58.1%, 24.4%, and 113.2%, respectively, over the runner-up performers. Although its VOL (0.534) is not the lowest, it remains well-controlled relative to its high returns.

Regarding predictive accuracy, FinKario-RAG achieves a leading ACC of 0.581, outperforming Guolian-Minsheng (0.575), China Fortune (0.573), and RAG (0.559). While some institutional strategies also demonstrate strong accuracy, performance varies considerably across institutions, with the maximum accuracy gap reaching 0.164.

Overall, FinKario-RAG strikes a robust risk-reward balance. Its knowledge graph-enhanced design delivers both superior profitability and effective risk control, positioning it as a state-of-the-art approach for LLM-based quantitative investment.

4.5 Ablation Study

We conduct two ablation studies. All variants are built on the same backbone, GPT-4o-mini, ensuring consistency across comparisons.

4.5.1 Knowledge Ablation Study

Table 3 presents an ablation study over different knowledge sources, comparing FinKario with existing financial knowledge. Removing the Event Graph (w/o Event Graph) results in a dramatic decrease in ARR, showing an 87.2% drop, from 2.633 to 0.336, and a reduction in SR by 81.1%, from 4.926 to 0.932. In contrast, removing the Attribute Graph (w/o Attribute Graph) also leads to performance degradation, but the decline is less pronounced across all indicators, underscoring the complementary role of the Event Graph in enhancing model performance. Additionally, we examine the impact of raw markdown content, which leads to a noticeable decline in performance. The model struggles to filter out useful information, particularly when processing long-text, highlighting its inefficiency in handling unstructured data. Furthermore, we compare the open-source knowledge base **HiDy** [9], which integrates multiple sources of financial data. The results suggest that HiDy provides valuable supplementary knowledge but does not outperform the full knowledge graph structure in improving performance, with a reduction in SR by 72.5%.

4.5.2 Retrieval Ablation Study

Table 4 highlights the ablation study on varied retrieval approaches. We integrate FinKario with traditional retrieval methods. The Vanilla RAG approach shows the lowest performance across all metrics, with ACC and SR both dropping by 28.9% and 84.6%, respectively. In contrast, the GraphRAG shows moderate gains over the Vanilla

[8] https://huggingface.co/Duxiaoman-DI/XuanYuan-6B

[9] https://zenodo.org/records/12630355

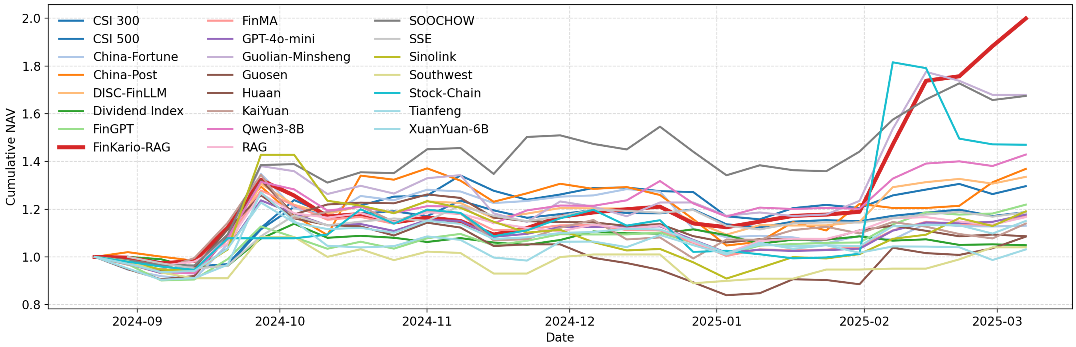


Figure 5: Accumulated returns (AR) of each baseline strategy on the financial-report dataset from August 28, 2024 to March 7, 2025. The figure shows the net asset value (NAV) curves over the weekly backtesting period.

Table 2: Performance comparison across market indices, vanilla LLMs, financial domain LLMs, and institutional strategies.

Model	ARR↑	VOL↓	SR↑	MDD↓	CR↑	ACC↑	STO↑
CSI 300	0.392	0.295	1.330	0.091	4.332	-	3.861
CSI 500	0.648	0.342	1.894	0.137	4.729	-	3.990
SSE	0.390	0.265	1.471	0.082	4.748	-	3.400
Dividend Index	0.096	0.233	0.411	0.079	1.208	-	0.929
Qwen3-8b	0.941	0.459	2.051	0.132	7.130	0.475	5.234
GPT-4o-mini	0.351	0.372	0.944	0.178	1.977	0.471	2.221
RAG (4o-mini)	0.336	0.360	0.932	0.197	1.703	0.559	2.408
FinMA	0.348	0.389	0.895	0.214	1.623	0.479	2.300
FinGPT	0.443	0.327	1.355	0.103	4.294	0.475	3.075
DISC-FinLLM	0.729	0.468	1.559	0.163	4.462	0.474	4.335
XuanYuan-6B	0.318	0.373	0.852	0.170	1.868	0.471	1.930
Stock-Chain	1.177	1.211	0.971	0.190	6.182	0.546	3.103
Tianfeng	0.054	0.542	0.100	0.225	0.242	0.411	0.282
Southwest	0.121	0.485	0.249	0.173	0.701	0.492	0.281
Sinolink	0.391	0.648	0.604	0.365	1.070	0.438	1.694
SOOCHOW	1.625	0.522	3.115	0.132	12.311	0.557	8.522
Guolian-Minsheng	2.012	0.647	3.108	0.169	11.880	0.575	7.915
Guosen	0.167	0.456	0.366	0.197	0.845	0.460	0.810
Huaan	0.170	0.471	0.361	0.333	0.509	0.435	1.020
KaiYuan	0.181	0.473	0.383	0.279	0.650	0.552	0.954
China-Fortune	0.263	0.537	0.489	0.216	1.218	0.573	1.028
China-Post	0.830	0.559	1.485	0.236	3.519	0.440	3.606
FinKario-RAG	2.633	0.534	4.926	0.172	15.315	0.581	18.171

Table 3: Ablation study on the impact of different knowledge sources. $'w/'$ uses other knowledge sources instead of FinKario; $'w/o'$ removes parts of FinKario for ablation.

	Knowledge	ARR↑	SR↑	MDD↓	ACC↑
FinKario-RAG	w/ Research report	0.336	0.932	0.197	0.559
	w/ HiDy	0.462	1.353	0.174	0.455
	w/o Event Graph	0.386	0.903	0.177	0.474
	w/o Attribute Graph	2.230	4.691	0.181	0.433
	FinKario (Ours)	**2.633**	**4.926**	**0.172**	**0.581**

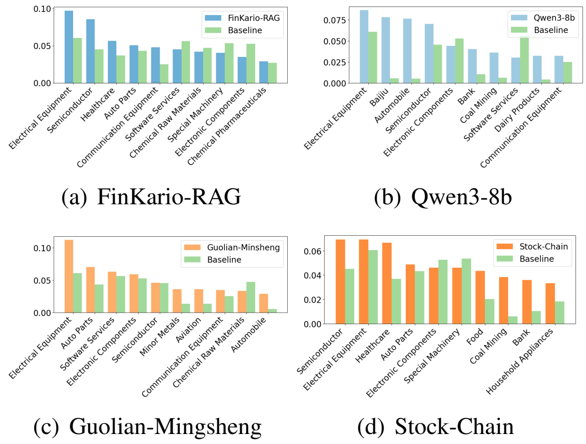
Table 4: Ablation study of varied retrieval approach.

	Method	ARR↑	SR↑	MDD↓	ACC↑
FinKario	Vanilla RAG	0.377	0.758	**0.120**	0.413
	GraphRAG	0.344	0.915	0.152	0.471
	LightRAG	0.821	1.313	0.140	0.495
	FinKario-RAG (Ours)	**2.633**	**4.926**	0.172	**0.581**

RAG, with ACC and SR improvements of 14% and 20.7%, respectively, demonstrating the potential of incorporating graph-based retrieval. The LightRAG method, which builds on GraphRAG while balancing computational cost and performance, shows further improvements, with ACC and SR gains of 5.1% and 43.5%, respectively. However, both methods still underperform compared to FinKario-RAG. These results underscore the advantage of the FinKario-RAG framework, which delivers superior performance across all metrics.

4.5.3 Case Study

To illustrate model behavior, we compare FinKario-RAG with the best-performing models in each category: Qwen3-8B (vanilla LLM), Guolian-Mingsheng (institution), and Stock-Chain (financial LLM). As depicted in Figure 6, FinKario-RAG exhibits a strong concentration in high-growth sec-



(a) FinKario-RAG

(b) Qwen3-8b

(c) Guolian-Mingsheng

(d) Stock-Chain

Figure 6: Visualization of Model Industry Preferences vs. Baseline Preferences. The baseline reflects the original industry distribution derived from raw research reports.

tors such as Electrical Equipment, Semiconductor, and Healthcare—an allocation strategy that closely aligns with the technology-led market rally observed around February 2025.

In contrast, Qwen3-8B and Stock-Chain present broader and less focused industry allocations across their top 3–4 sectors. Guolian-Mingsheng narrows its focus mainly to Auto Parts and Semiconductor, which aligns with FinKario-RAG to some extent and helps sustain performance in the later market stages. Although Stock-Chain also overweights Semiconductor, its simultaneous heavy allocation to sectors like Food, Coal Mining and Bank results in an implicit internal hedging effect, partially diluting return potential during sector rallies. FinKario-RAG's industry targeting strategy, by contrast, reflects higher consistency and adaptability to sector momentum.

5 Conclusion

This work presents **FinKario**, a fully automated dataset constructed with event-enhanced knowledge grounded in professional institutional templates, ensuring both domain alignment and scalable dynamic updates. By integrating **FinKario-RAG**, a Two-stage Graph-based RAG mechanism, our approach overcomes the limitations of single-target retrieval and reduces hallucination risk when handling large, dynamically evolving graphs. Extensive experiments demonstrate that FinKario-RAG significantly outperforms both vanilla LLMs and state-of-the-art FinLLMs in stock trend forecasting. To the best of our knowledge, this is the first benchmark against real-world institutional strategies, providing a more grounded and practical performance comparison.

Limitations

While FinKario-RAG demonstrates promising retrieval and reasoning performance, it remains limited in several aspects. The current framework focuses solely on textual financial data, lacking integration of multi-modal inputs such as numerical tables, figures, and time-series patterns that often convey critical quantitative insights. In addition, our experiments are primarily conducted on Chinese financial corpora. Future work could address these limitations by incorporating multimodal financial inputs, such as tables, charts, and time-series data, to enrich the retrieval context and enhance predictive robustness. Moreover, extending the framework to multilingual financial corpora would be valuable, as our professional templates based on CFA and JP Morgan standards are naturally suitable for cross-lingual adaptation.

Ethical Considerations

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

This work is intended strictly for research purposes and does not constitute financial, investment, or trading advice. Although the proposed framework produces structured predictions related to market reactions, it is not designed, evaluated, or validated for direct deployment in real-world trading systems. Inappropriate or unregulated use of such predictions may lead to financial losses, particularly under distribution shifts, market regime changes, or unforeseen economic conditions. Any potential real-world application of such systems would require appropriate safeguards, including domain-expert oversight and careful risk management, to mitigate misuse and unintended consequences.

Our goal is to evaluate and enhance the ability of large language models to understand, structure, and reason over complex financial information, particularly for the automated construction of financial knowledge graph schemas and event representations. This research aims to provide information and processed graph-based resources that may facilitate future academic research on the understanding and reasoning of financial information.

6 Acknowledgments

This work was partially supported by the Hong Kong Collaborative Research Fund (CRF) (Grants No. C7004-22G and C6015-23G), the Guangdong Provincial Natural Science Foundation (Grant No. 2025A1515010304), the Guangdong Province Project (Grant No. 2024QN11X088), and the Guangzhou Science and Technology Planning Project (Grant No. 2025A03J4491).

References

Titilope Tosin Adewale, Titilayo Deborah Olorunyomi, and Theodore Narku Odonkor. 2023. Big data-driven financial analysis: A new paradigm for strategic insights and decision-making. *Journal of Financial Innovation and Analytics*, 1(1):1–15.

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

Abhinav Arun, Fabrizio Dimino, Tejas Prakash Agarwal, Bhaskarjit Sarmah, and Stefano Pasquali. 2025. Finreflectkg: Agentic construction and evaluation of financial knowledge graphs. In *Proceedings of the 6th ACM International Conference on AI in Finance*, pages 283–290.

Haonan Bian, Zhiyuan Yao, Sen Hu, Zishan Xu, Shaolei Zhang, Yifu Guo, Ziliang Yang, Xueran Han, Huacan Wang, and Ronghao Chen. 2026. Realmem: Benchmarking llms in real-world memory-driven interaction. *Preprint*, arXiv:2601.06966.

Hanzhu Chen, Xu Shen, Qitan Lv, Jie Wang, Xiaoqi Ni, and Jieping Ye. 2024a. SAC-KG: Exploiting large language models as skilled automatic constructors for domain knowledge graph. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4345–4360. Association for Computational Linguistics.

Wei Chen, Qiushi Wang, Zefei Long, Xianyin Zhang, Zhongtian Lu, Bingxuan Li, Siyuan Wang, Jiarong Xu, Xiang Bai, Xuanjing Huang, and 1 others. 2023. Disc-finllm: A chinese financial large language model based on multiple experts fine-tuning. *arXiv preprint arXiv:2310.15205*.

Yuemin Chen, Feifan Wu, Jingwei Wang, Hao Qian, Ziqi Liu, Zhiqiang Zhang, Jun Zhou, and Meng Wang. 2024b. Knowledge-augmented financial market analysis and report generation. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track*, pages 1207–1217.

Dawei Cheng, Fangzhou Yang, Xiaoyang Wang, Ying Zhang, and Liqing Zhang. 2020. Knowledge graph-based event embedding framework for financial quantitative investments. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and*

Development in Information Retrieval, pages 2221–2230.

Andrea Colombo, Teodoro Baldazzi, Luigi Bellomarini, Emanuel Sallinger, and Stefano Ceri. 2025. Template-based explainable inference over high-stakes financial knowledge graphs. *computational complexity*, 7(24):47.

E Philip Davis. 1996. The role of institutional investors in the evolution of financial structure and behaviour. *The Future of the Financial System*, 33:49–99.

Linyi Ding, Sizhe Zhou, Jinfeng Xiao, and Jiawei Han. 2024. Automated construction of theme-specific knowledge graphs. *arXiv preprint arXiv:2404.19146*.

Peijie Dong, Zhenheng Tang, Xiang Liu, Lujun Li, Xiaowen Chu, and Bo Li. 2025. Can compressed llms truly act? an empirical evaluation of agentic capabilities in llm compression. In *International Conference on Machine Learning*, pages 14169–14202. PMLR.

Ziliang Gan, Yu Lu, Dong Zhang, Haohan Li, Che Liu, Jian Liu, Ji Liu, Haipang Wu, Chaoyou Fu, Zenglin Xu, Rongjunchen Zhang, and Yong Dai. 2024. Mme-finance: A multimodal finance benchmark for expert-level understanding and reasoning. *Preprint*, arXiv:2411.03314.

Denis J Hilton. 2001. The psychology of financial decision-making: Applications to trading, dealing, and investment analysis. *The Journal of Psychology and Financial Markets*, 2(1):37–53.

Sen Hu, Yuxiang Wei, Jiaxin Ran, Zhiyuan Yao, Xueran Han, Huacan Wang, Ronghao Chen, and Lei Zou. 2026a. Does memory need graphs? a unified framework and empirical analysis for long-term dialog memory. *Preprint*, arXiv:2601.01280.

Sen Hu, Zhiyu Zhang, Yuxiang Wei, Xueran Han, Zhenheng Tang, Huacan Wang, and Ronghao Chen. 2026b. Clonemem: Benchmarking long-term memory for ai clones. *Preprint*, arXiv:2601.07023.

Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, and 1 others. 2024. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*.

K Kapellas and G Siougle. 2017. Financial reporting practices and investment decisions: A review of the literature. *Industrial Engineering & Management*, 6(04):1–9.

Natthawut Kertkeidkachorn, Rungsiman Nararatwong, Ziwei Xu, and Ryutaro Ichise. 2023. Finkg: A core financial knowledge graph for financial analysis. In *2023 IEEE 17th International Conference on Semantic Computing (ICSC)*, pages 90–93. IEEE.

Xiang Li, Zhenyu Li, Chen Shi, Yong Xu, Qing Du, Mingkui Tan, Jun Huang, and Wei Lin. 2024. Alphafin: Benchmarking financial analysis with retrieval-augmented stock-chain framework. *arXiv preprint arXiv:2403.12582*.

Xiaohui Victor Li and Francesco Sanna Passino. 2024. Findkg: Dynamic knowledge graphs with large language models for detecting global trends in financial markets. In *Proceedings of the 5th ACM International Conference on AI in Finance*, pages 573–581.

Zhaowei Liu, Xin Guo, Fangqi Lou, Lingfeng Zeng, Jinyi Niu, Zixuan Wang, Jiajie Xu, Weige Cai, Ziwei Yang, Xueqian Zhao, and 1 others. 2025. Fin-r1: A large language model for financial reasoning through reinforcement learning. *arXiv preprint arXiv:2503.16252*.

Zhenghao Liu, Zhijian Zhang, and Xi Zeng. 2024. Risk identification and management through knowledge association: A financial event evolution knowledge graph approach. *Expert Systems with Applications*, 252:123999.

A Craig MacKinlay. 1997. Event studies in economics and finance. *Journal of economic literature*, 35(1):13–39.

Bhaskarjit Sarmah, Dhagash Mehta, Benika Hall, Rohan Rao, Sunil Patel, and Stefano Pasquali. 2024. Hybridrag: Integrating knowledge graphs and vector retrieval augmented generation for efficient information extraction. In *Proceedings of the 5th ACM International Conference on AI in Finance*, pages 608–616.

Michael C Schlachter. 2013. How you compare to the big funds: Advantages and disadvantages of individual investors. In *INVEST LIKE AN INSTITUTION: PROFESSIONAL STRATEGIES FOR FUNDING A SUCCESSFUL RETIREMENT*, pages 1–14. Springer.

Kuldeep Singh, Simerjot Kaur, and Charese Smiley. 2024. Finqapt: Empowering financial decisions with end-to-end llm-driven question answering pipeline. In *Proceedings of the 5th ACM International Conference on AI in Finance*, pages 266–273.

Zheng Su, Mukai Hao, Qiang Zhang, Bo Chai, and Ting Zhao. 2020. Automatic knowledge graph construction based on relational data of power terminal equipment. In *2020 5th International Conference on Computer and Communication Systems (ICCCS)*, pages 761–765. IEEE.

ShaoBo Sun, Xiaoming Pan, Shuang Qi, and Jun Gao. 2025. Knowledge enhanced prompt learning framework for financial news recommendation. *Pattern Recognition*, page 111461.

Zhenheng Tang, Xin He, Tiancheng Zhao, Fanjunduo Wei, Xiang Liu, Peijie Dong, Qian Wang, Qi Li, Huacan Wang, Ronghao Chen, and 1 others. 2026a. Llm agent memory: A survey from a unified representation–management perspective.

Zhenheng Tang, Xiang Liu, Qian Wang, Eunsol Choi, Bo Li, and Xiaowen Chu. 2026b. Are dilemmas and conflicts in llm alignment solvable? a view from priority graph. *arXiv preprint arXiv:2603.15527*.

Zichen Tang, Zirui Zhang, Qian Wang, Zhenheng Tang, Bo Li, and Xiaowen Chu. 2026c. Is your llm-as-a-recommender agent trustable? llms' recommendation is easily hacked by biases (preferences). *arXiv preprint arXiv:2603.17417*.

Rex Thompson. 1995. Empirical methods of event studies in corporate finance. *Handbooks in Operations Research and Management Science*, 9:963–992.

Bin Wang, Chao Xu, Xiaomeng Zhao, Linke Ouyang, Fan Wu, Zhiyuan Zhao, Rui Xu, Kaiwen Liu, Yuan Qu, Fukai Shang, and 1 others. 2024. Mineru: An open-source solution for precise document content extraction. *arXiv preprint arXiv:2409.18839*.

Mengyu Wang, Tiejun Ma, and Shay B Cohen. 2025. Pre-training time series models with stock data customization. *arXiv preprint arXiv:2506.16746*.

Wenguang Wang, Yonglin Xu, Chunhui Du, Yunwen Chen, Yijie Wang, and Hui Wen. 2021. Data set and evaluation of automated construction of financial knowledge graph. *Data Intelligence*, 3(3):418–443.

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.

Bolun Xia, Vipula Rawte, Aparna Gupta, and Mohammed Zaki. 2024. Fetilda: evaluation framework for effective representations of long financial documents. *ACM Transactions on Knowledge Discovery from Data*, 18(7):1–27.

Qianqian Xie, Weiguang Han, Xiao Zhang, Yanzhao Lai, Min Peng, Alejandro Lopez-Lira, and Jimin Huang. 2023. Pixiu: A large language model, instruction data and evaluation benchmark for finance. *arXiv preprint arXiv:2306.05443*.

Yuanjian Xu, Jianing Hao, Kunsheng Tang, Jingnan Chen, Anxian Liu, Peng Liu, and Guang Zhang. 2025. Finripple: Aligning large language models with financial market for event ripple effect awareness. *arXiv preprint arXiv:2505.23826*.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, and 1 others. 2025. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*.

Hongyang Yang, Xiao-Yang Liu, and Christina Dan Wang. 2023. Fingpt: Open-source financial large language models. *Preprint*, arXiv:2306.06031.

Zixuan Yu, Zhenheng Tang, Tongliang Liu, Chengqi Zhang, Xiaowen Chu, and Bo Han. 2026. Rethinking deep research from the perspective of web content distribution matching. *arXiv preprint arXiv:2603.07241*.

Bowen Zhang and Harold Soh. 2024. Extract, define, canonicalize: An llm-based framework for knowledge graph construction. *arXiv preprint arXiv:2404.03868*.

Wentao Zhang, Lingxuan Zhao, Haochong Xia, Shuo Sun, Jiaze Sun, Molei Qin, Xinyi Li, Yuqing Zhao, Yilei Zhao, Xinyu Cai, and 1 others. 2024. 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*, pages 4314–4325.

Tianyu Zhou, Pinqiao Wang, Yilin Wu, and Hongyang Yang. 2024. Finrobot: Ai agent for equity research and valuation with large language models. *arXiv preprint arXiv:2411.08804*.

Xinyi Zhu, Liping Wang, Hao Xin, Xiaohan Wang, Zhifeng Jia, Jiyao Wang, Chunming Ma, and Yuxiang Zengt. 2023. T-finkb: A platform of temporal financial knowledge base construction. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, pages 3671–3674. IEEE.

Yuanbing Zhu, Zhenheng Tang, Xiang Liu, Ang Li, Bo Li, Xiaowen Chu, and Bo Han. 2025. OracleKV: Oracle guidance for question-independent KV cache compression. In *ICML 2025 Workshop on Long-Context Foundation Models*.

A Appendix

A.1 Quality Control Refinement Algorithm

The complete refinement pipeline is detailed in Algorithm 1.

A.2 Prompts in FinKario

A.2.1 The Prompt for Acquiring Schema of Attribute Graph

We leverage standardized equity research templates from authoritative sources (e.g., the CFA Institute and J.P. Morgan) as reference guides. We design a prompt that guides the model to identify core company attributes for schema construction, including name, ticker, rating, market capitalization, and more. Specifically, the attribute schema comprises 11 relation types: Stock Ticker, Primary Exchange, Primary Industry, Investment Rating, Current Stock Price, Market Capitalization, Target Price, Major Shareholders, Risk Assessment, Key Products, Research Institution. Figure 7 illustrates an example of the automatically acquired schema for our Attribute Graph.

Figure 7: Prompt for acquiring schema of Attribute Graph.

A.2.2 The Prompt for Attribute Graph construction

After acquiring the schema of the Attribute Graph, this structure serves as the foundation for attribute-level knowledge population from financial research reports. Figure 8 provides an example of the designed prompt used to guide the model in extracting these attributes during the knowledge population process.

A.2.3 The Prompt for acquiring schema of Event Graph

We automatically construct the top-down event-driven schema by prompting LLMs to extract high-level driven categories from the Wisconsin handbook and further match low-level event ontologies from FIBO to support interpretable Event Graph construction. Figure 9 illustrates the resulting tree-structured schema. Each event category corresponds to a high-level driver and is associated with its fine-grained instances or relations. Below is the list of categories and their typical subtypes:

- **Supply**: is provided by, Capacity Adjustment, Market Action, Holds
- **Demand**: Sales, Consumption, Performance, Is needed by
- **Revenue**: Earning, Profit, Income-oriented classifier, Is issued by, Has increased / decreased
- **Efficiency Cost**: Lower the cost, Automation
- **Strategic Action**: Merger / Acquisition, Overseas expansion, Spin-off
- **Technology Innovation**: Is applicable in, New product, Has innovated, Iteration
- **Policy Regulation**: Regulatory action, Governs, License
- **Macro**: Interest rate, GDP, Disaster

A.2.4 The Prompt for Event Graph construction

After establishing the event schema, we designed a tailored prompt to guide large language models in extracting structured event-level information from equity research reports. As illustrated in Figure 10, the prompt instructs the model to (1) identify the subject and object of the event, (2) extract relevant entities such as company names, products, and indicators, (3) annotate the timeframe, and (4) determine a driven category from options like "Supply", "Demand", or "Strategic Action".

To ensure relevance and accuracy, the prompt restricts extraction to events directly tied to company activities. The resulting JSON output includes not only the core event tuple but also a reasoning statement that explains the event linkage. The figure presents both an illustrative example and a real-case

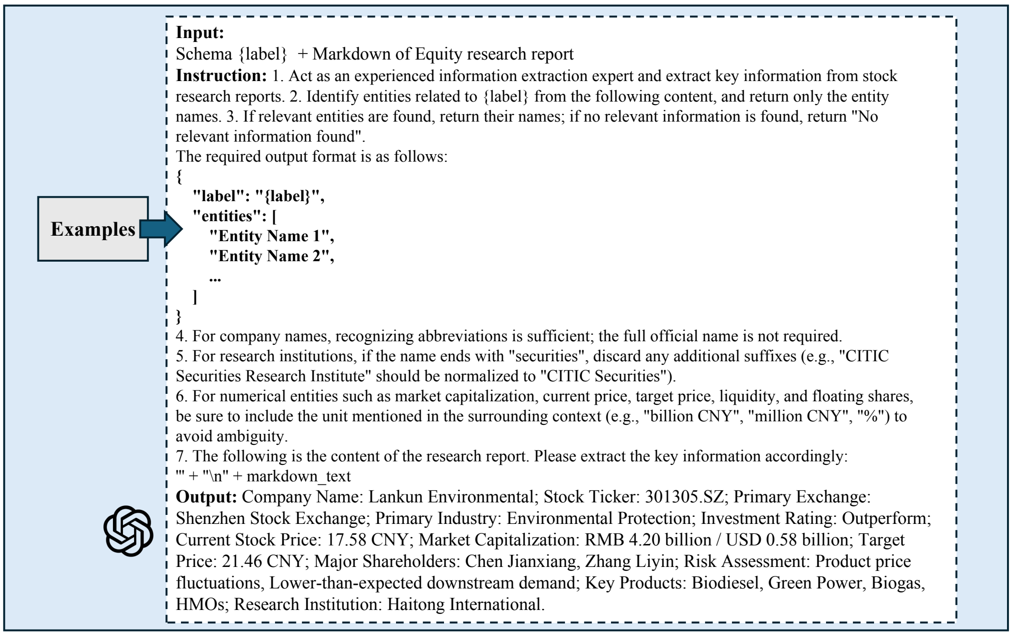


Figure 8: Prompt for Attribute Graph construction.

output to demonstrate the clarity and consistency achieved through our prompt design.

A.3 Statistics of trading signals

Table 5 shows the number of trading signals generated by different strategies. The frequency varies substantially, ranging from 300 (China-Post) to 8,211 (FinMA). FinKario-RAG generates 689 signals, which is relatively conservative compared to several baselines. This variation in trading frequency is crucial because more signals typically result in higher transaction costs through increased brokerage fees and market impact. To evaluate this trade-off between signal frequency and profitability, we analyze trading costs in the next section.

Table 5: Number of Trading Signals Generated by Different Strategies

Strategy	Trading Signals
DISC-FinLLM	5835
FinGPT	1545
FinMA	8211
GPT-4o-mini	6737
Qwen3-8B	499
RAG	1752
Stock-Chain	390
XuanYuan-6B	3410
SOOCHOW	832
Guolian-Mingsheng	697
China-Post	300
Tianfeng	788
FinKario-RAG	**689**

A.4 The Case Study of Investment Query Response

This subsection presents a case study comparing how various models—including FinLLMs such as FinGPT, XuanYuan-6B, Stock-Chain, and FinKario-RAG, as well as an advanced vanilla LLM (GPT-4o-mini), respond to a user query about investment analysis for Haier Biomedical (Haier Bio). The query asks the model to analyze the stock's investment potential and predict its future price trend.

As shown in Figure 11, FinGPT, XuanYuan-6B, and GPT-4o-mini consistently emphasize the limitations of LLMs in delivering definitive investment predictions. These models cite the inherent uncertainty of market dynamics and the lack of access to real-time data as major barriers. Their responses generally avoid direct suggestions, instead encouraging users to consider macroeconomic conditions, company fundamentals, and to consult financial professionals. In contrast, Stock-Chain attempts a more analytical response by summarizing company fundamentals and macro-level trends. However, its output includes factual inaccuracies such as misidentifying the stock code, and it fails to synthesize comparative insights across the industry, falling short of the user's request for a cross-company investment evaluation. Moreover, it does not offer actionable investment guidance, only suggesting that decisions require consideration of multiple external factors.

FinKario-RAG addresses these shortcomings by accurately grounding its analysis in correct entity identifiers, comparing the target stock with other industry players, and offering nuanced conclusions. This demonstrates the effectiveness of our financial knowledge graph construction and retrieval approach, which goes beyond traditional single-document retrieval methods to support more context-aware and investor-aligned responses.

A.5 Supplementary Experiments

A.5.1 Metric & Baseline Settings

To evaluate the performance of our model, we adopt six widely used metrics: **Annualized Rate of Return (ARR)**, which measures the compound annual growth rate of the portfolio value over the evaluation period; **Volatility (VOL)**, which quantifies the annualized standard deviation of weekly returns and indicates the portfolio's risk level; **Sharpe Ratio (SR)**, which quantifies risk-adjusted performance by dividing the annualized return by the annualized volatility; **Maximum Drawdown (MDD)**, which measures the largest peak-to-trough decline in portfolio value, representing the worst-case loss scenario; **Calmar Ratio (CR)**, which assesses risk-adjusted returns by dividing the annualized return by the absolute maximum drawdown; **Accuracy (ACC)**, which measures the percentage of correct directional predictions generated by FinKario-RAG trading signals; **Sortino Ratio (STO)**, which measures downside-risk-adjusted performance by normalizing returns with respect to downside volatility only, thereby complementing SR by focusing on negative return deviations. Together, these metrics provide a comprehensive view of both the predictive quality and practical investment performance of our model.

For the baselines, the markdown reports are not passed to the model in full. Instead, each report is cleaned and then segmented into 1000 token with overlap. These chunks are embedded and stored in a vector index, and only the top-k retrieved chunks are provided to the LLM at inference time.

A.5.2 The Supplementary Experiments of Institutional Strategy

In the main manuscript, we compared our method against institutional agencies that published at least 300 equity research reports. To enable a more comprehensive evaluation of institutional strategy effectiveness, we additionally included agencies that have rated at least 100 reports in this supplementary analysis. The cumulative performance of these strategies is illustrated in the NAV curve shown in Figure 12, while detailed performance metrics are summarized in Table 7.

For example, Guolian-Minsheng, SOOCHOW, and Caixin demonstrate strong annualized returns, with Guolian-Minsheng reaching 2.012, SOOCHOW at 1.625, and Caixin also at 1.625. Compared to low-return agencies such as Tianfeng and Cinda, whose annualized returns are below 0.06, these top performers yield more than 25 times higher returns. However, high returns do not always equate to stability. Caixin, despite having the highest Sharpe Ratio of 3.816, shows relatively low accuracy at 0.455, indicating that effective market timing can sometimes matter more than raw predictive accuracy. In contrast, Guolian-Minsheng achieves both a high return and an accuracy of 0.575, the best among all institutions. Zhongtai and Ping-An exhibit the lowest volatility values at 0.312 and 0.324, respectively, which are about 40% lower than the average across institutions. Nonetheless, their returns remain moderate, indicating a trade-off between risk and reward. Meanwhile, Shanxi and Dongxing record negative returns, highlighting potential weaknesses in their investment strategies during weekly backtesting. Some institutions, such as China-Post and Pacific, deliver reasonably high returns above 0.8 and 1.5, respectively. However, their performance metrics such as drawdown and accuracy remain suboptimal, indicating inconsistent prediction quality. Agencies like Guosen and KaiYuan exhibit more balanced profiles with moderate returns and volatility, yet lack standout performance in any single dimension.

Overall, these results illustrate the fragmented quality and strategic effectiveness among institutional players. Against this backdrop, FinKario-RAG delivers the strongest performance across all major return-oriented metrics, including the highest annualized return, Sharpe Ratio, and accuracy, while keeping risk measures such as maximum drawdown and volatility within acceptable bounds,

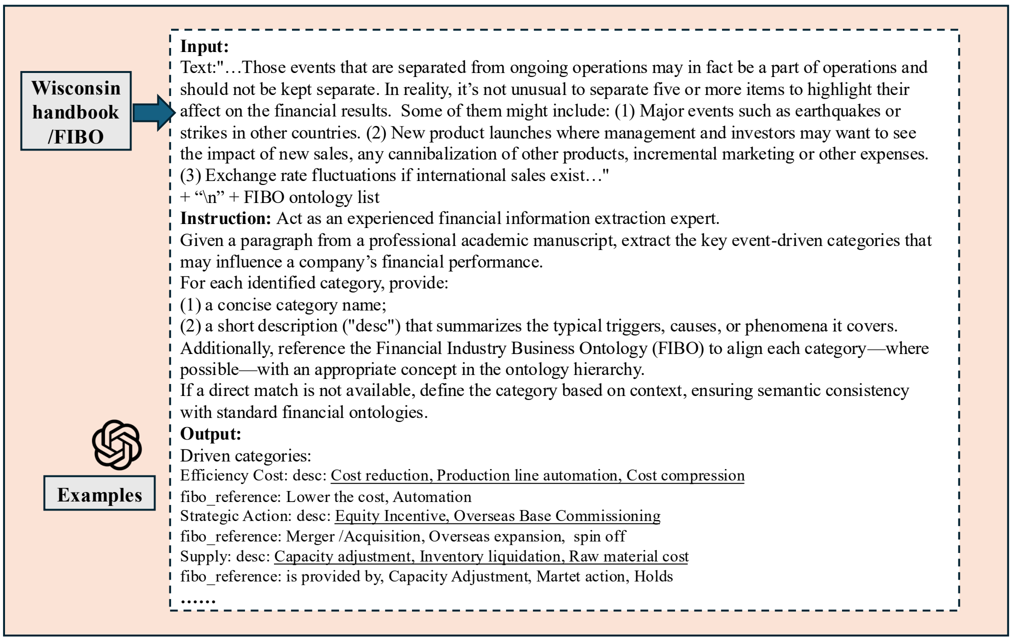


Figure 9: Prompt for acquiring schema of Event Graph.

thereby reinforcing the advantage of our structured retrieval and knowledge-grounded approach.

A.5.3 The Supplementary Experiments with Trading Cost

To further assess the robustness of FinKario-RAG under realistic market conditions, we incorporate a transaction cost of 0.005 (i.e., 0.5% per trade for both entry and exist direction) into the backtesting process. This adjustment accounts for commission fees commonly encountered in real-world trading. As shown in Figure 13, all strategies experience a moderate reduction in cumulative returns after considering trading costs, reflecting the expected impact of transaction frictions. Market indices such as CSI 300 and CSI 500 show smaller declines due to their lower trading frequency and more stable holdings, whereas model-driven strategies exhibit relatively larger drops owing to higher portfolio turnover. Nevertheless, Table 6 demonstrates that FinKario-RAG continues to outperform both baseline LLMs and institutional benchmarks across key performance metrics (ARR, SR, and CR), indicating its strong generalization ability and stability even when trading expenses are included. These results highlight that FinKario-RAG maintains superior risk-adjusted profitability while remaining resilient to realistic transaction environments.

A.5.4 The Supplementary Experiments under Different Market Regimes

To further validate robustness, we extend the backtesting period to June 6, 2025, covering more diverse and evolving market conditions. As shown in Figure 14, FinKario-RAG consistently maintains stable outperformance over baselines throughout the extended horizon. The results demonstrate that the proposed method generalizes well across different market regimes and remains resilient under longer-term evaluation.

Examples

Input:
Schema {choice} + Markdown of Equity research report
Instruction: Please extract events from the following paragraph and return the results in JSON format. (1) Identify the subject and object components of each event and output logically consistent relations. (2) Extract involved entities, including companies, products, and indicators. (3) Annotate the time frame of the event (if available). (4) Only consider events centered around companies; ignore entities that are not directly related to company activities. (5) Based on the type of event, assign a driven category, which must be selected from the following options: {choices}
Example Input:
"Due to production cuts by original manufacturers driving up storage prices, Demingli's net profit in Q4 2023 increased by 500%."
Example Output:
{"subject": "Production cuts by original manufacturers", "object": "Increase in net profit ", "entities": {"Company": ["Demingli"], "Product": ["Storage"], "Indicator": ["Net Profit"]},"timeframe": "2023Q4", "reasoning": "The production cuts directly led to an increase in storage prices. Demingli's 500% increase in net profit is linked to this. The event is primarily driven by factors related to Supply.", "driven_category": "Supply"
}
Output:
{"subject": "Jinbo Biotech's active expansion into overseas markets", "object": "Improvement in brand awareness and market competitiveness", "entities": {"Product": ["ProtYouth Collagen Peptide Essence", "ProtYouth Triple Collagen Cream", "Single-ingredient Luxurious Collagen Essence"],"Indicator": ["Brand Awareness"]}, "timeframe": "By the end of January 2025", "confidence": 0.8, "reasoning": "Jinbo Biotech's active expansion into overseas markets is the reason behind the improvement in brand awareness and market competitiveness. The time frame is explicitly indicated as the end of January 2025.", "source_paragraph": "…", "driven_category": "Strategic Action"}

Figure 10: Prompt for Event Graph construction.

Table 6: Performance comparison across vanilla LLMs, financial domain LLMs, and institutional strategies. (Considering trading cost)

Model	ARR↑	VOL↓	SR↑	MDD↓	CR↑
Qwen3-8b	0.160	0.460	0.349	0.236	0.679
GPT-4o-mini	-0.203	0.372	-0.547	0.302	-0.673
RAG	-0.213	0.361	-0.589	0.302	-0.704
FinMA	-0.205	0.389	-0.527	0.320	-0.641
FinGPT	-0.147	0.325	-0.452	0.220	-0.668
DISC-FinLLM	0.029	0.468	0.062	0.273	0.107
XuanYuan-6B	-0.224	0.373	-0.600	0.287	-0.780
Stock-Chain	0.282	1.175	0.240	0.285	0.987
Tianfeng	-0.384	0.542	-0.709	0.378	-1.015
Southwest	-0.248	0.399	-0.623	0.244	-1.018
Sinolink	-0.173	0.640	-0.271	0.443	-0.391
SOOCHOW	0.587	0.524	1.120	0.173	3.400
Guolian-Minsheng	0.752	0.627	1.201	0.273	2.756
Guosen	-0.316	0.453	-0.697	0.347	-0.911
Huaan	-0.305	0.464	-0.657	0.414	-0.735
KaiYuan	-0.307	0.474	-0.648	0.369	-0.831
China-Fortune	-0.257	0.535	-0.480	0.342	-0.751
China-Post	0.088	0.553	0.159	0.302	0.292
FinKario-RAG	1.222	0.537	2.274	0.260	4.692

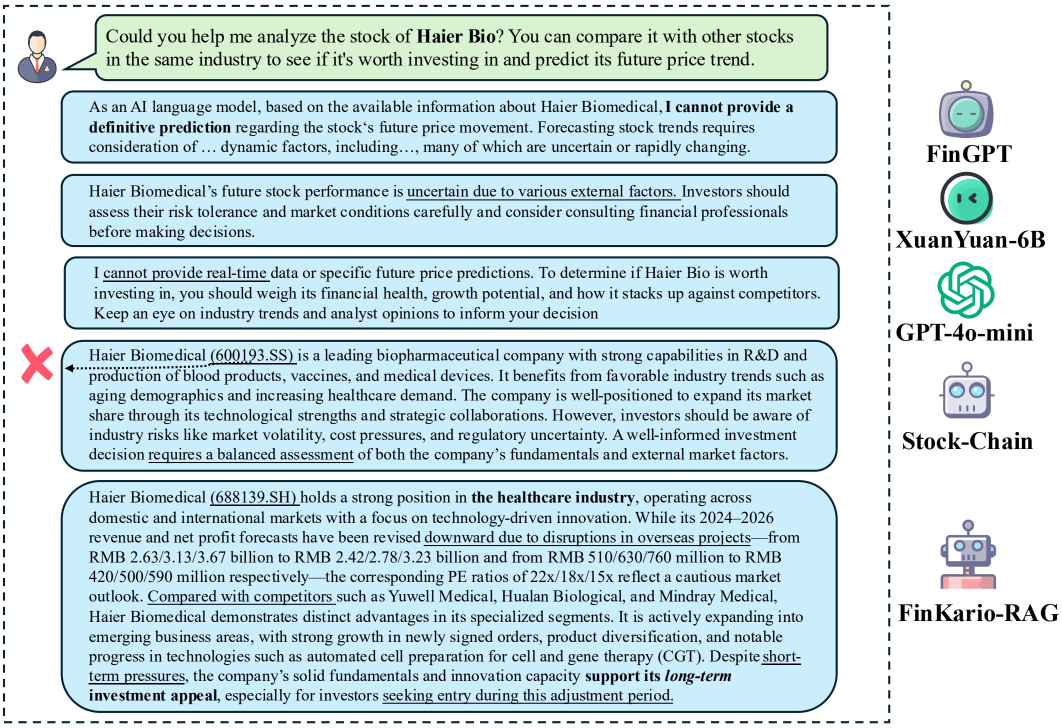


Figure 11: Evaluation of investment suggestions from FinGPT, XuanYuan-6B, GPT-4o-mini, Stock-Chain, and FinKario

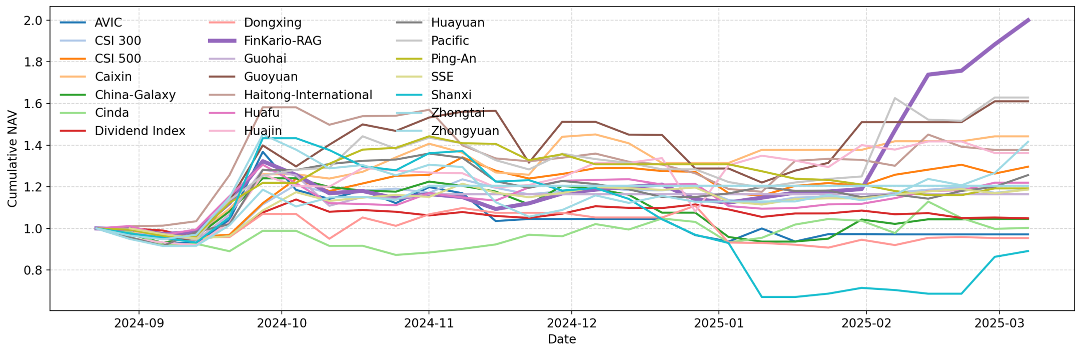


Figure 12: Accumulated returns (AR) of each institutional strategy on the financial-report dataset from August 28, 2024 to March 7, 2025. The figure shows the net asset value (NAV) curves over the weekly backtesting period.

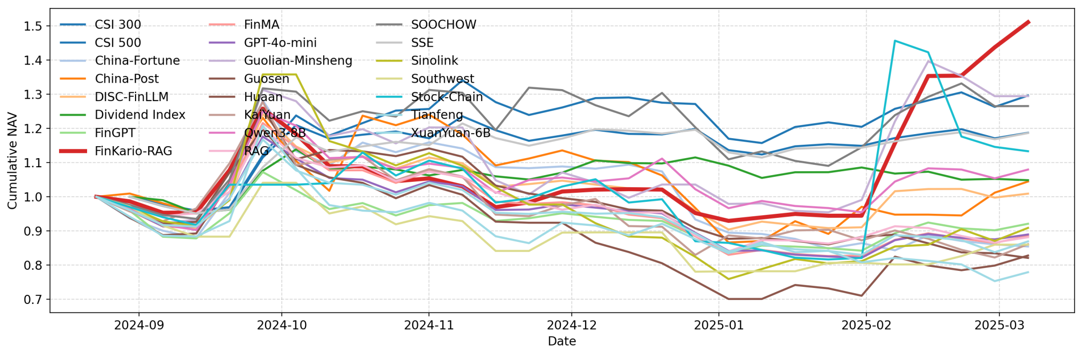


Figure 13: Accumulated returns (AR) of each baseline strategy with transaction cost on the financial-report dataset from August 28, 2024 to March 7, 2025. The figure shows the net asset value (NAV) curves over the weekly backtesting period.

Table 7: Performance of institutional agencies that published at least 100 equity research reports.

Agency	ARR↑	VOL↓	SR↑	MDD↓	CR↑	ACC↑
Tianfeng	0.054	0.542	0.100	0.225	0.242	0.411
Southwest	0.121	0.485	0.249	0.173	0.701	0.492
Sinolink	0.391	0.648	0.604	0.365	1.070	0.438
SOOCHOW	1.625	0.522	3.115	0.132	12.311	0.557
Guolian-Minsheng	2.012	0.647	3.108	0.169	11.880	0.575
Guosen	0.167	0.456	0.366	0.197	0.845	0.460
Huaan	0.170	0.471	0.361	0.333	0.509	0.435
KaiYuan	0.181	0.473	0.383	0.279	0.650	0.552
China-Fortune	0.263	0.537	0.489	0.216	1.218	0.573
China-Post	0.830	0.559	1.485	0.236	3.519	0.440
Pacific	1.557	0.582	2.674	0.170	9.138	0.458
China-Galaxy	0.533	0.452	1.179	0.289	1.847	0.482
Huafu	0.465	0.381	1.222	0.162	2.881	0.409
Zhongtai	0.432	0.312	1.388	0.068	6.358	0.328
Cinda	0.005	0.413	0.011	0.117	0.041	0.496
Shanxi	-0.198	0.817	-0.243	0.532	-0.373	0.440
Guohai	0.340	0.387	0.879	0.136	2.506	0.402
Huajin	0.813	0.569	1.429	0.192	4.233	0.358
Ping-An	0.403	0.324	1.243	0.196	2.056	0.538
Guoyuan	1.488	0.623	2.391	0.220	6.768	0.491
Huayuan	0.553	0.421	1.314	0.168	3.292	0.409
Haitong-International	0.852	0.571	1.490	0.256	3.322	0.395
Zhongyuan	0.964	0.561	1.718	0.270	3.565	0.483
Caixin	1.625	0.426	3.816	0.126	12.897	0.455
AVIC	-0.091	0.589	-0.154	0.329	-0.276	0.437
Dongxing	-0.087	0.391	-0.223	0.181	-0.481	0.415
FinKario-RAG	2.633	0.534	4.926	0.172	15.315	0.581

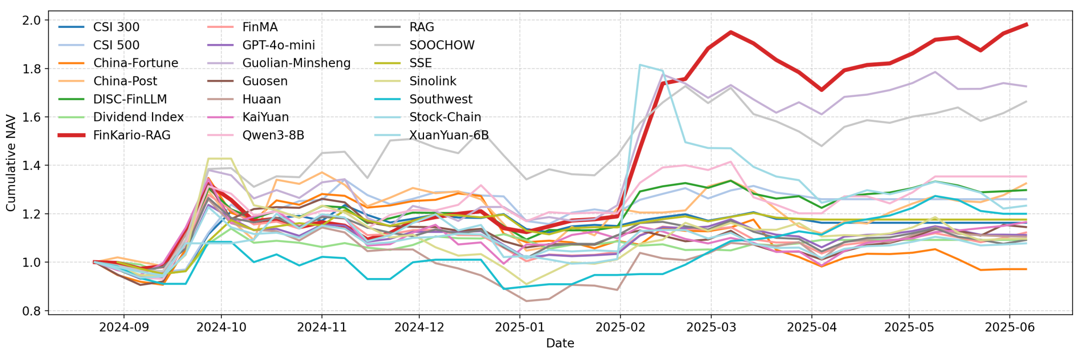


Figure 14: Accumulated returns (AR) of each institutional strategy on the financial-report dataset from August 28, 2024 to June 6, 2025. The figure shows the net asset value (NAV) curves over the weekly backtesting period.

Algorithm 1 Quality Control Refinement

Require: Raw knowledge graph $\mathcal{G}^{(s)}_{\text{FinKario}}$, reference dictionary $\mathcal{T}_{\text{ref}}$, and LLM($\cdot$) instantiated as GPT-4o-mini

Ensure: Refined knowledge graph $\mathcal{G}'^{(s)}_{\text{FinKario}}$

// Step 1: Entity Normalization

for all entity $e \in \mathcal{G}^{(s)}_{\text{FinKario}}$ **do**
  if e is a name variant (e.g., "BYD Inc.", "BYD Auto") **then**
    Replace e with canonical form (e.g., "BYD")
  end if
end for

// Step 2: Attribute Completion

for all triple $(e_h, r, e_t) \in \mathcal{G}^{(s)}_{\text{FinKario}}$ **do**
  if r is a numeric attribute (e.g., "Price", "Cap") and (e_t is missing or lacks unit) **then**
    Query $\mathcal{T}_{\text{ref}}$ for value and unit (e.g., CNY, USD, billions)
    Replace e_t with correct value and unit
  end if
end for

// Step 3: Error Correction via LLM

for all triple $(e_h, r, e_t) \in \mathcal{G}^{(s)}_{\text{FinKario}}$ **do**
  if e_t contains a placeholder (e.g., "No relevant information was found", "Extraction error") **then**
    Re-feed the source Markdown passage to `LLM(·)`
    Replace e_t with the corrected output from the LLM
  end if
end for
return $\mathcal{G}'^{(s)}_{\text{FinKario}}$