Junjie ZHANG, Shuoling LIU, Tongzhe ZHANG, Yuchen SHI, 2025. A survey on large language model-based alpha mining. *Frontiers of Information Technology & Electronic Engineering*, 26(10):1809-1821. https://doi.org/10.1631/FITEE.2500386

A survey on large language model-based alpha mining

Front Inform Technol Electron Eng

Key words: Alpha mining; Quantitative investment; Large language models (LLMs); LLM agents; Fintech

Corresponding author: Yuchen SHI E-mail: shiyuchen@efunds.com.cn
ORCID: https://orcid.org/0000-0002-1885-8043

Motivation

- **Alpha mining** aims to discover data-driven signals predictive of future cross-sectional returns.
- Traditional methods:
  - Human-designed: interpretable but slow and narrow.
  - Algorithmic mining: scalable but opaque and less interpretable.
- **Large language models (LLMs) bridge both worlds**, enabling natural-language-driven alpha discovery with speed, structure, and semantic reasoning.
- This paper provides the **first systematic review** of LLM-based alpha mining frameworks from an **agentic and engineering** perspective.

Front Inform Technol Electron Eng

Main idea and research framework

- **Objective:** Explore how LLMs enhance the entire alpha mining pipeline by acting as **miners**, **evaluators**, and **interactive assistants**.
- **Approach:**
  - Conducted a structured literature review (2023–2025) using keywords such as LLM alpha mining, financial signal generation, and LLM quantitative trading.
  - Analyzed **representative frameworks** (Alpha-GPT, QuantAgent, AlphaAgent, R&D-Agent, FAMA, etc.).
  - Compared design paradigms, roles, architectures, and performance.
- **Core concept:**

  LLMs transform natural-language hypotheses into executable factors and enhance feedback loops between **domain intuition** and **algorithmic rigor**, forming a hybrid intelligence paradigm.

Front Inform Technol Electron Eng

LLM-enhanced alpha mining pipeline

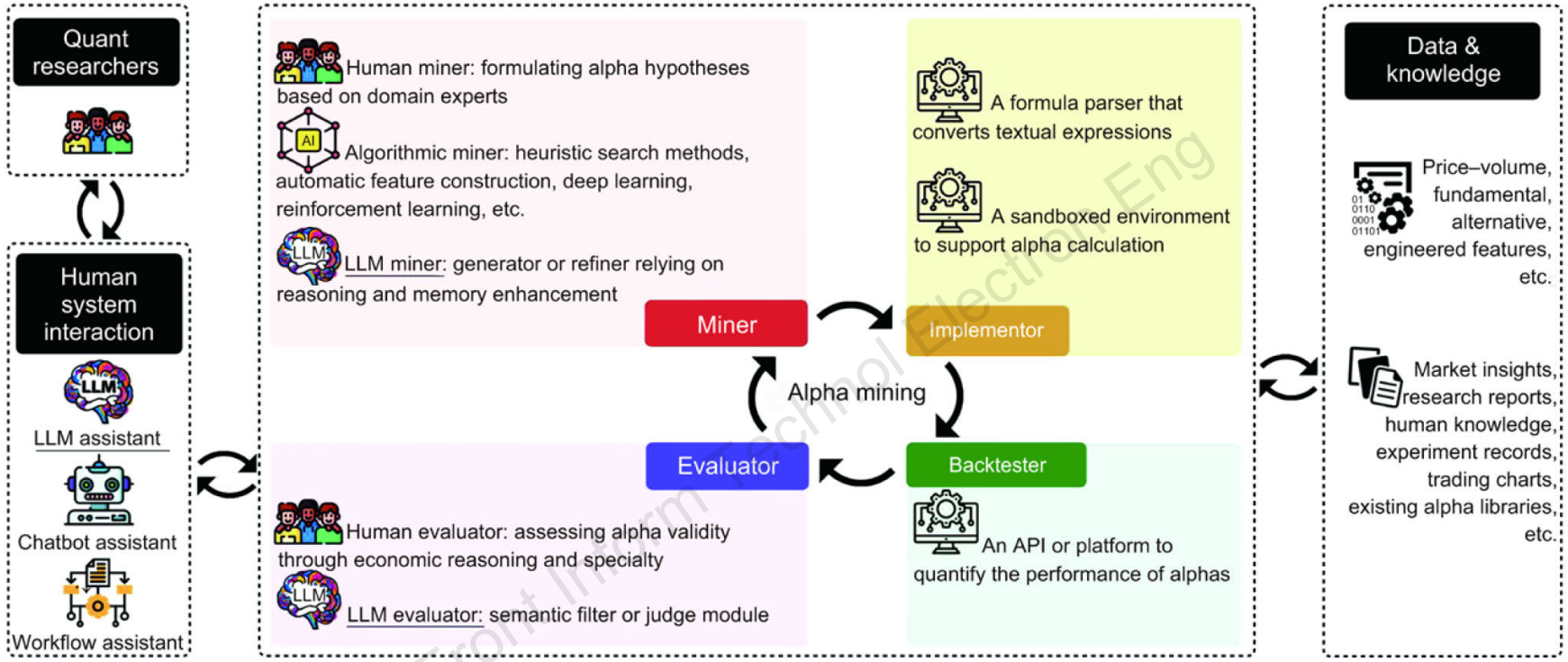


Fig. 1 A general alpha mining pipeline integrating human experts, algorithmic tools, and LLMs

Core roles of LLMs:

1. **Miner:** synthesizes new factor expressions from text, data, or prior alphas.
2. **Evaluator:** judges plausibility, originality, and interpretability.
3. **Interactive assistant:** supports conversational and iterative refinement.

Representative LLM-based alpha mining frameworks

Table 1 Summary of representative LLM-based alpha mining frameworks

Framework	Open source	LLM used	Data type	Knowledge base	LLM role
Alpha-GPT (Wang SZ et al., 2023)	No	GPT-3.5-turbo-16k	Price–volume	Existing alpha library and historical experiment records	Miner and interactive assistant
Cheng-Tang-2023 (Cheng and Tang, 2024)	No	ChatGPT	Price–volume	None	Miner and evaluator
GPT-signal (Wang YN et al., 2024)	Planned	GPT-4	Fundamental	None	Miner
Kou-2024 (Kou et al., 2024)	Yes	ChatGPT	Price–volume	Multimodal data: market dynamics, financial reports, trading charts, etc.	Miner
QuantAgent (Wang SZ et al., 2024)	No	GPT-4	Price–volume	Historical experiment records	Miner and evaluator
FAMA (Li ZW et al., 2024)	No	GPT-3.5 (text-davinci-002)	Price–volume	Existing alpha library	Miner
R&D-Agent (Li YT et al., 2025)	Yes	GPT-4o-(mini), o3-(mini), GPT-4.1, GPT-4-turbo	Price–volume	Historical experiment records	Miner, evaluator, and interactive assistant
AlphaAgent (Tang et al., 2025)	No	GPT-3.5-turbo	Price–volume	Human knowledge, research reports, and market insights	Miner and evaluator
LLM-Powered MCTS (Shi Y et al., 2025)	No	GPT-4.1	Price–volume	Existing alpha library and historical experiment records	Miner and evaluator
Chain-of-Alpha (Cao L et al., 2025)	No	GPT-4o, DeepSeek-V3, Qwen3-32B	Price–volume	Existing alpha library and historical experiment records	Miner

MCTS: Monte Carlo tree search; FAMA: foundation for the advancement of monetary affairs

Front Inform Technol Electron Eng

Reported experimental results

Table 2 LLM-based factor mining experiments and baseline comparisons

Framework	Data scope	Disclosed backtest metric	Comparison with baseline
Alpha-GPT (Wang SZ et al., 2023)	US S&P500 (2012–2021)	Single factor IC=0.020–0.025; annualized return=5%–10%	IC doubled compared with LLM-only generation (0.010→0.020); AR improved from marginally positive to 5%–10%
Cheng-Tang-2023 (Cheng and Tang, 2024)	CRSP US Stock (2021–2022)	Single factor AR=66.16%, Sharpe=4.49; equal-weighted GPT factor portfolio AR=88%, Sharpe=2.46	Excess returns unexplained by Fama–French five-factor model; superior to heuristic factor design in novelty and efficiency
GPT-signal (Wang YN et al., 2024)	US S&P500 sectoral data (healthcare, IT, energy) (2016–2020)	IC improved by 3%–5%; annualized return improved by 5%–8%; portfolio Sharpe from 1.2 to >1.5; strongest gains in IT and healthcare; development time reduced by 30%	IC and Sharpe improved by 30% compared with traditional financial-signal models
QuantAgent (Wang SZ et al., 2024)	China A-share, 500 stocks (2023)	IC from 0.009 to ∼0.018 after five self-improvement cycles; Sharpe increased from 0.85 to >1.25; AR improved by ∼5%–7% compared to a static baseline	Dual-loop (inner+outer) mechanism outperforming ablated single-loop settings: without the outer loop IC stagnated at ∼0.012 and without the inner loop Sharpe dropped to <1; overall performance nearly doubled relative to static factors
FAMA (Li ZW et al., 2024)	US S&P500 (2015–2022)	RankIC=0.054, RankICIR=0.485, annualized return ≈38.4%, Sharpe ≈6.67	Dual-loop (inner+outer) mechanism outperforming single-loop variants; IC stagnation (∼0.012) without outer loop; Sharpe <1 without inner loop; performance nearly doubled vs. static factors
Kou-2024 (Kou et al., 2024)	China SSE 50 (2021–2023)	cum. return 53.17%, Sharpe >1.5, MaxDD −7%	AR improved by +30%–40%; Sharpe improved by 0.5 compared to classical and deep (ALSTM, Transformer) models; superior to AlphaEvolve and RL-based alpha factories in return and drawdown control

Front Inform Technol Electron Eng

Reported experimental results

Table 2 (Cont'd)

AlphaAgent (Tang et al., 2025)	China CSI500, US S&P500 (2015–2024)	CSI500: AR=11.00%, IR=1.488, IC=0.0212, MaxDD=−9.36%; S&P500: AR=8.74%, IR=1.055, IC=0.0056, MaxDD=−9.10%	Superior AR, IR, and IC vs. LSTM, Transformer, LightGBM, TRA, StockMixer, AlphaForge, R&D-Agent, and LLM baselines (OpenAI-o1, DeepSeek-R1)
R&D-Agent-Quant (Li YT et al., 2025)	China CSI300 (2016–2024)	RD-Factor: IC=0.0497, IR=1.36, ARR=11.84%, MaxDD=−9.10%; RD-Model: IC=0.0469, IR=1.70, ARR=10.09%, MaxDD=−6.94%; RD-Agent (Q): IC=0.0532, IR=1.74, ARR=14.21%, MaxDD=−7.42%	Superior to Alpha101 factor library, GP-based methods, and earlier RD-Agent; also superior to deep models (LSTM/Transformer) on CSI300 in both IC and risk-adjusted return
LLM-Powered MCTS (Shi Y et al., 2025)	China CSI300, CSI1000 (2011–2024); US S&P500 (2015–2024)	CSI300: AR=8.20%, IR=0.94, IC=0.0420, RankIC=0.0395; CSI1000: AR=13.90%, IR=1.36, IC=0.0800, RankIC=0.0730; S&P500: IC=0.0132, RankIC=0.0130	Superior IC/RankIC/AR/IR vs. GP, DSO, AlphaGen, AlphaForge, CoT/ToT, FAMA, and AlphaAgent; consistent advantage on CSI300/CSI1000 and S&P500
Chain-of-Alpha (Cao L et al., 2025)	China CSI500 & CSI1000 (2010–2025)	CSI500: IC=0.0485, RankIC=0.0771, ICIR=0.3047, RankICIR=0.5013, AR=0.1324, IR=1.4178; CSI1000: IC=0.0672, RankIC=0.0902, ICIR=0.4630, RankICIR=0.6228, AR=0.1471, IR=1.4043	Superior to Alpha101/158/360, GP, DSO, AlphaGen, AlphaForge, and LLM baselines (LLM+CoT/ToT/MCTS)

ALSTM: attention-based long short-term memory; AR: autoregressive; CoT: chain-of-thought; CRSP: center for research in security prices; CSI: China securities index; DSO: direct search optimization; GP: genetic programming; IR: information ratio; LSTM: long short-term memory; SSE: Shanghai stock exchange; ToT: tree-of-thought; TRA: trading return analysis; cum.: cumulative

Front. Inform. Technol. Electron. Eng.

Challenges and future directions

Challenges

- Simplified and non-standard evaluation protocols
- Limited numerical reasoning and financial logic understanding
- Low diversity and originality in factor generation
- Weak iterative improvement dynamics
- Temporal data leakage risks
- Compliance and explainability constraints in production use

Front Inform Technol Electron Eng

Future directions

- Domain-aligned reasoning: fine-tuning with financial corpora and numeric reasoning
- Multimodal extension: structured, textual, and environmental, social, and governance (ESG)/satellite data integration
- Revised evaluation frameworks: unified benchmarks and regime-aware testing
- General-purpose quant platforms: integration of LLMs with human and algorithmic agents

Conclusions

LLMs:

- amplify, not replace, human and algorithmic intelligence;
- enable faster hypothesis testing and interpretable factor discovery;
- represent a hybrid paradigm merging intuition, automation, and language reasoning; and
- are expected to reshape quantitative research into scalable, explainable, and interactive systems.

Front Inform Technol Electron Eng



Junjie ZHANG is a PhD candidate at Nanyang Technological University. Before this, he earned his MS degree in Artificial Intelligence from Tsinghua University. He has published several papers as the first author in leading international conferences including ICML and ACL. His current research focuses on reinforcement learning and large language models, exploring their applications across diverse domains.



Yuchen SHI is an Associate Director of Frontier Technology at E Fund Management Co., Ltd. She received her PhD degree from the Department of Industrial Systems Engineering and Management at the National University of Singapore, and conducted research with the Singapore-ETH Centre on statistical modeling and monitoring of complex systems. Before joining E Fund, she worked at Huawei Digital Power AI Lab as a Senior Algorithm Engineer, leading projects on industrial visual inspection and anomaly detection. Her current work focuses on the development of multi-agent architectures for financial applications and large-model-based quantitative investment systems.

Front Inform Technol Electron Eng