AUTOMATE STRATEGY FINDING WITH LLM IN QUANT INVESTMENT

Morning

25 March, 2025

TEAM MEMBERS

Hoang Van Duong

Nguyen The Cuong

Nguyen Gia Huy

Overview

1. Terminology
2. Abstract
3. Introduction
4. Problem Formulation
5. Methodology
6. Experiment
7. Related Work
8. Conclusion

TERMINOLOGY

Quantitative investment
(đầu tư định lượng)

Quantitative investment = Đầu tư dựa trên dữ liệu + Mô hình toán học + Tự động hóa

Dựa trên mô hình toán học

Được kiểm định kỹ

Tự động hóa & tốc độ cao

Không bị chi phối bởi cảm xúc

TERMINOLOGY

ALPHA

Alpha (Quantitative Alpha) là một chỉ số

- Alpha dương cho thấy chiến lược có thể tạo ra lợi nhuận vượt trội.
- Alpha âm cho thấy chiến lược không hiệu quả.

ALPHA

Có nhiều cách tính alpha (momentum, volatility, mean reversion...) nhưng không phải lúc nào cũng hiệu quả.

- Nếu thị trường sôi động → Dùng biến động giá (volatility-based alpha).
- Nếu thị trường ổn định → Có thể dùng động lượng (momentum-based alpha).

Trước khi chọn: Gọi chung là **Seed Alpha** (các alpha tiềm năng).
Sau khi chọn: Gọi là **Selected Alpha** (alpha chính thức đưa vào chiến lược).

ATR
AVERAGE TRUE RANGE

Chỉ báo phân tích kỹ thuật do J. Welles Wilder phát triển.

- ATR dùng để đo mức độ biến động giá của một tài sản trong một khoảng thời gian nhất định.
- Đo độ mạnh/yếu của biến động.

Close - Giá đóng cửa

Open - Giá mở cửa

Delay - Lấy loại dữ liệu quá khứ

VD: CLOSE - DELAY(CLOSE,5)

SMA - Trung bình giá đơn giản

EMA - Trung bình lũy thừa

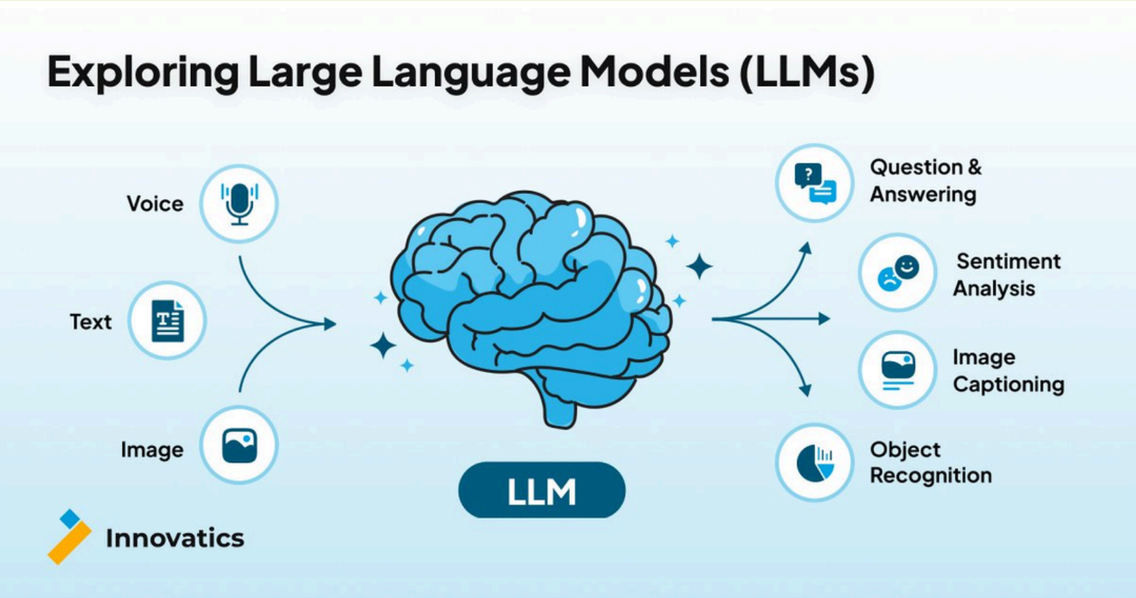
(Thêm trọng số cho giá gần đây)

TERMINOLOGY

LLMs là gì?

Large Language Models
là các mô hình ngôn ngữ lớn,
một dạng trí tuệ nhân tạo (AI)

Ví dụ: ChatGPT, ClaudeAI,

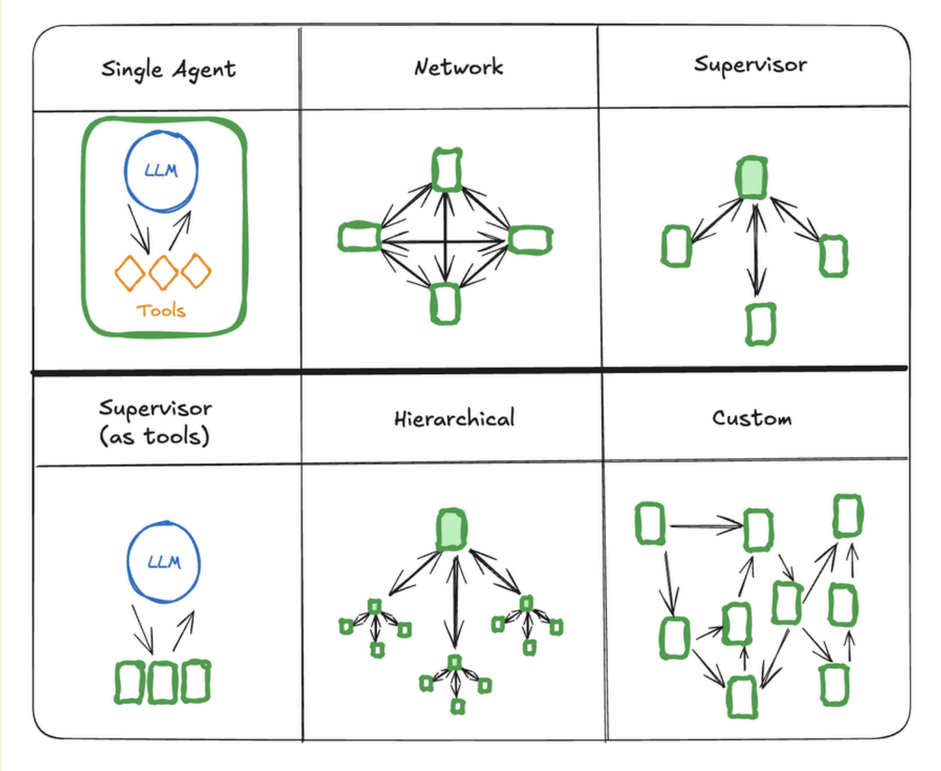


TERMINOLOGY

multi-Agent là gì?

Multi-Agent System

- là hệ thống bao gồm nhiều tác nhân (agents)
- có thể tương tác, hợp tác hoặc cạnh tranh để thực hiện các nhiệm vụ phức tạp.



LLMs có thể làm "não" cho từng agent, giúp chúng suy luận, giao tiếp.

1. ABSTRACT

Dự án đề xuất một framework đầu tư định lượng hiện đại, kết hợp sức mạnh của LLM và multi-agent để khai thác các yếu tố alpha và tối ưu hóa chiến lược giao dịch. Hệ thống cho phép phân tích dữ liệu tài chính đa phương thức, đánh giá yếu tố alpha theo điều kiện thị trường thực tế và xây dựng chiến lược đầu tư linh hoạt, thích ứng. Kết quả cho thấy mô hình mang lại hiệu suất vượt trội và có tiềm năng ứng dụng rộng rãi trong đầu tư tài chính AI.

2. INTRODUCTION

Thị trường dữ liệu thay thế (Alternative Data):

- Năm 2023: Đạt 7.2 tỷ USD
- Tốc độ tăng trưởng kép hàng năm từ 2023–2030: 52.1%

Đầu tư định lượng đã tăng cường đáng kể khả năng phân tích và xử lý khối lượng lớn dữ liệu tài chính

Thử thách:

- Phương pháp truyền thống thiếu linh hoạt
- Dữ liệu đa dạng, khó tích hợp
- Thị trường thay đổi liên tục

2. INTRODUCTION

Giải pháp:

Sử dụng LLMS

- Khai thác Alpha linh hoạt
- Multi-Agent multimodal
- Tối ưu chiến lược linh hoạt

2. INTRODUCTION

Mô hình đầu tư

1. Nghiên cứu – tìm kiếm tín hiệu
2. Giao dịch – chọn alpha theo thị trường
3. Quản trị rủi ro – đánh giá theo khẩu vị rủi ro
4. Quản lý danh mục – tổng hợp chiến lược đầu tư

Figure 1: Pipeline of Quant Trading Company

2. INTRODUCTION

Đóng góp chính:

- Đề xuất framework xây dựng Alpha factory linh hoạt dựa trên LLM, có khả năng cập nhật liên tục
- Ứng dụng kiến trúc Multi-Agent vào tài chính để đánh giá alpha theo thị trường
- Tối ưu hóa chiến lược giao dịch mà không cần can thiệp thủ công

3. PROBLEM FORMULATION

Framework giúp hỗ trợ tạo ra các yếu tố alpha và chiến lược trong giao dịch định lượng (quant invesment).

- Yếu tố Seed Alpha
- Tận dụng LLMs và multi-agent system

3.1 ALPHA FACTOR & STRATEGY

Tính toán Alpha:

1. Với n cổ phiếu trong T phiên giao dịch tới, mỗi cổ phiếu có những đặc trưng (giá, khối lượng...).
2. Từ các đặc trưng này, sinh ra các alpha khác nhau (α_{11}, α_{12}, ..., α_{kk}).
3. Chọn alpha tốt nhất từ từng nhóm dựa trên kết quả back-test & điều kiện thị trường hiện tại.
4. Từ các alpha được chọn, xây dựng chiến lược tổng hợp:

$$\alpha = \sum w_i \alpha_i$$

trong đó, w_i là trọng số của mỗi alpha.

3.2 SEED ALPHA MINING & SELECTION

- **Hạn chế** của phương pháp Alpha truyền thống:
  - thiếu linh hoạt
  - không tối ưu trong môi trường biến động

- **Giải pháp** đề xuất:
  - Áp dụng mô hình ngôn ngữ lớn (LLM) và hệ thống tự động để lựa chọn Seed Alpha hiệu quả.
  - Multi-agent: Khai thác thêm dữ liệu mạng xã hội, tin tức... để tăng độ chính xác.

>> Framework mới: Kết hợp học máy + tài chính định lượng → hệ thống linh hoạt theo biến động thị trường.

3.3 SEED ALPHA FORMULAIC

Framework Seed Alpha:

- Áp dụng **quy tắc toán học** do LLM tạo ra, kết hợp nhiều biểu thức tài chính trên dữ liệu hàng ngày.
- Áp dụng toán tử thời gian

Ví dụ: CLOSE - DELAY(SMA(CLOSE, 14), 7)

→ So sánh giá đóng cửa hiện tại với trung bình động 14 phiên, lùi 7 ngày.

- Cập nhật và cải tiến liên tục:
  - Mô hình không cố định mà thích ứng với dữ liệu mới và kiến thức tài chính mới nhất.

3.2 SELECTED ALPHA FORMULA

Ví dụ:

- CLOSE - DELAY(SMA(CLOSE, 14), 7)

→ So sánh giá đóng cửa hiện tại với trung bình động 14 phiên, lùi 7 ngày.

Nếu có nhiều đầu vào thì sao?

$$\text{Alpha} = \sum w_i \alpha_i$$

với w_i là các trọng số của mỗi alpha.

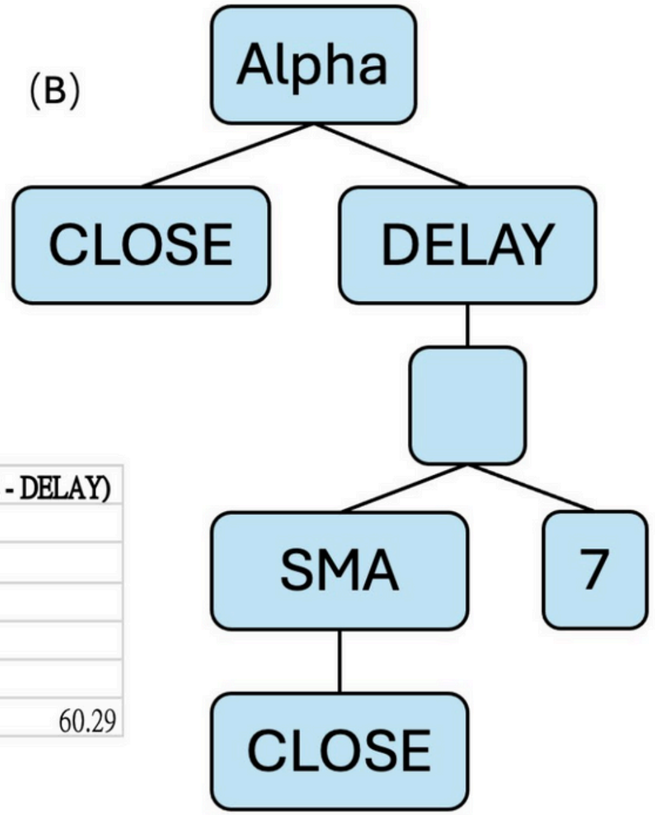
#	Alpha	Weight
1	(CLOSE - DELAY(CLOSE, 14))	-0.1459
2	(RSI - DELAY(RSI, 14))	-1.0265
3	(CLOSE - DELAY(SMA(CLOSE, 14), 7))	-0.1978
4	(MA(CLOSE, 20) - CLOSE)	0.0556
5	(SMA(CLOSE, 20) - CLOSE)	-0.945
6	(MAX(HIGH, 20) - CLOSE)	-0.4053
7	(100-RSI)	-0.3199
8	(BOLL_UP - BOLL_DOWN) / SMA(CLOSE, 20)	3.6186
9	STD(CLOSE, 10) / STD(CLOSE, 50)	-0.183
10	VOLUME / MARKET_CAP	-3.2145
11	VOLUME * CLOSE	-0.0058
12	(EPS / DELAY(EPS, 1) - 1)	-1.8351

3.2 SEED ALPHA MINING & SELECTION

(A)

Detrended Price Oscillator (DPO)	(CLOSE - DELAY(SMA(CLOSE, 14), 7))

(B)



(C)

Day	Closing Price (CLOSE)	SMA(CLOSE, 14)	DELAY(SMA(CLOSE, 14), 7)	Alpha (CLOSE - DELAY)
1	100	N/A	N/A	N/A
...	...	N/A	N/A	N/A
13	140	N/A	N/A	N/A
14	145	119.71	N/A	N/A
...	...	...	N/A	N/A
21	180	144.71	119.71	60.29

Figure 2: (A) An example of the seed alpha. (B) Its equivalent expression tree. (C) Step-by-step computation of this seed alpha on an example time series.

4. METHODOLOGY

1. **Seed Alpha Factory:**
   - Sử dụng LLMs để lọc và tạo tập hợp mạnh mẽ các Seed Alpha từ dữ liệu tài chính.
2. **Multi-Agent:**
   - Hệ thống lựa chọn Alpha dựa trên rủi ro đa dạng và điều kiện thị trường, đảm bảo độ tin cậy thông qua cơ chế chấm điểm.
3. **Tối ưu hóa trọng số:**
   - Dùng Mạng Nơ-ron Sâu (DNN) để dự đoán lợi suất, hỗ trợ xây dựng chiến lược đầu tư hiệu quả.

4.1 Overview of the framework

- Cấu trúc 3 phần:
  - Seed Alphas Factory
  - Quy trình ra quyết định multi-agent
  - Phương pháp tối ưu hóa trọng số

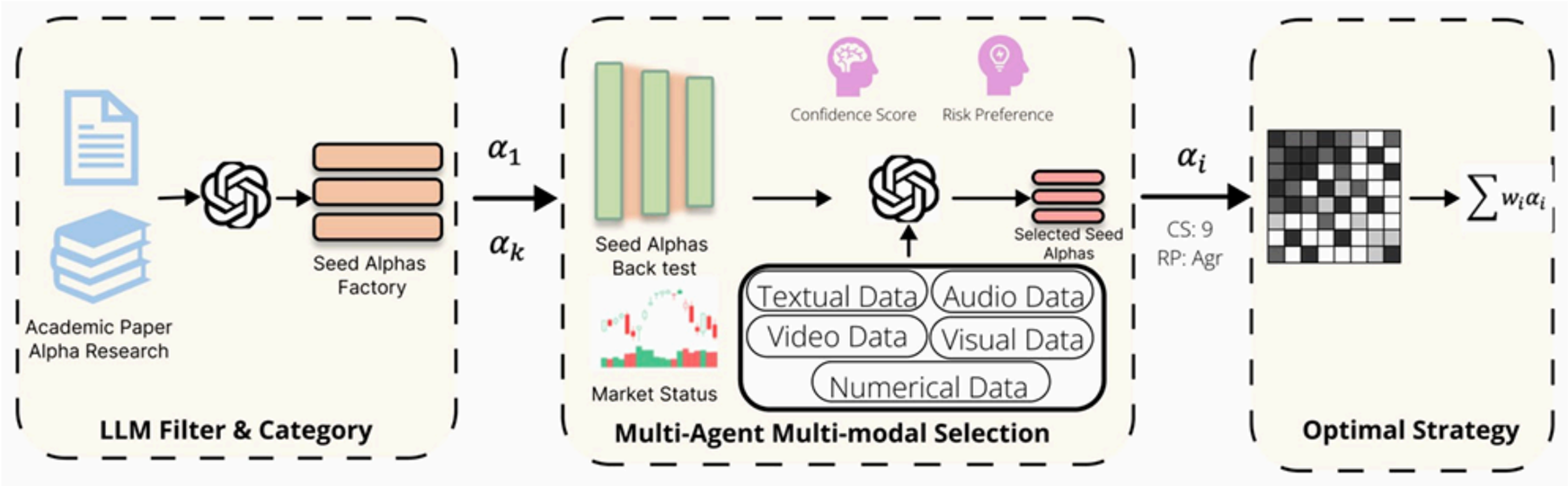


Figure 3: Overview of the strategy generate process (CS stands for confidence score; RP stands for risk preference).

4.1 Overview of the Framework

- **Giai đoạn 1 – Tạo Seed Alpha:**
  - Sử dụng mô hình ngôn ngữ lớn (LLM) để lọc & phân loại dữ liệu, tạo tập hợp Seed Alpha mạnh mẽ.
- **Giai đoạn 2 – Ra quyết định multi-agent:**
  - Lựa chọn Alpha phù hợp theo điều kiện thị trường & rủi ro.
- **Giai đoạn 3 – Tối ưu hóa trọng số:**
  - Dùng kỹ thuật học sâu để đánh giá hiệu suất Seed Alpha và xây dựng chiến lược đầu tư.

4.2 LLM FILTER & CATEGORY

- Mục tiêu:
  - Đẩy nhanh việc hiểu biết và khai thác Alpha bằng cách dùng LLM (sử dụng bản ChatGPT gọi là **Alpha Grail**).
- Chức năng của Alpha Grail:
  - Hỗ trợ xây dựng Seed Alpha Factory thông qua xử lý dữ liệu nghiên cứu định lượng.

- Kết quả giai đoạn này:
  - Tạo tập hợp Alpha có cấu trúc, gồm nhiều danh mục theo từng lĩnh vực tài chính, là nền tảng cho các bước tiếp theo như ra quyết định và tối ưu hóa trọng số.

4.2 LLM FILTER & CATEGORY

DETAILED

Alpha Grail – Xây dựng Seed Alpha Factory

🔍 **Input:**

- 11 bài nghiên cứu về tài chính & khoa học máy tính.
- Văn bản, hình ảnh, bảng biểu, đồ thị tài chính.
- Thông tin về cách phân loại alpha.
- The specific instruction given to Alpha Grail was: "Summarize the document information to help quantitative researchers build the Seed Alpha Factory according to traditional financial categories, ensuring that each category of seed alphas is independent."

📊 **Output:**

- Seed Alphas được phân loại theo Động lượng, Hồi quy trung bình, Biến động, và Phân tích cơ bản.

4.3 MULTIMODAL & MULTI-AGENT EVALUATE ALPHA SET

- Tổng quan:
  - Đánh giá toàn diện các yếu tố Alpha bằng cách tích hợp nhiều loại dữ liệu (văn bản, hình ảnh, bảng biểu,...) và phân tích từ nhiều góc nhìn.
  - Kết hợp dữ liệu đa phương tiện
  - Phong phú quá trình đánh giá, giúp hiểu rõ động thái thị trường.

- **Multi-agent system:**
  - Nhiều tác nhân với tiêu chí riêng biệt cùng **chấm điểm Alpha theo chỉ số rủi ro**
  - → Đảm bảo độ tin cậy do nhiều agent có nhiều khẩu vị rủi ro khác nhau.

4.3 Multimodal & Multi-Agent Evaluate Alpha Set

Table 1: Multimodal data types

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

4.3 MULTIMODAL & MULTI-AGENT EVALUATE ALPHA SET

- Tự động hóa tìm selected alphas:
  - Sử dụng thuật toán Category-Based Alpha Selection và hàm SelectBestAlphas để xác định bộ Alpha vượt ngưỡng điểm tin cậy.

- Kết quả:
  - Bộ Alpha có tiềm năng nhất, làm nền tảng cho chiến lược đầu tư thích ứng với thị trường thay đổi.

4.3 Multimodal & Multi-Agent Evaluate Alpha Set

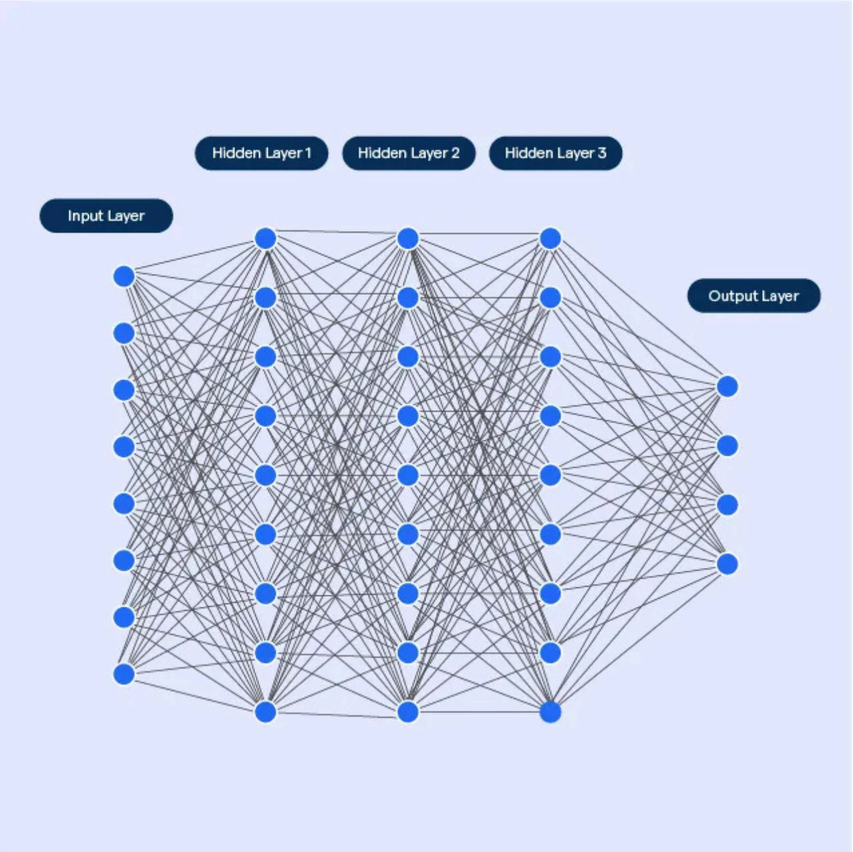
Algorithm 1 Category-Based Alpha Selection

Input: categories $\mathcal{C} = \{C_1, ..., C_m\}$, each C_i containing a set of Alphas, confidence threshold X selected Alphas $\mathcal{A} = \{\}$

```
for each category C_i ∈ C do
    A_i = SelectBestAlphas(C_i);
    for each α ∈ A_i do
        if ConfidenceScore(α) > X then
            A = A ∪ {α};
        end if
    end for
end for
return A
```

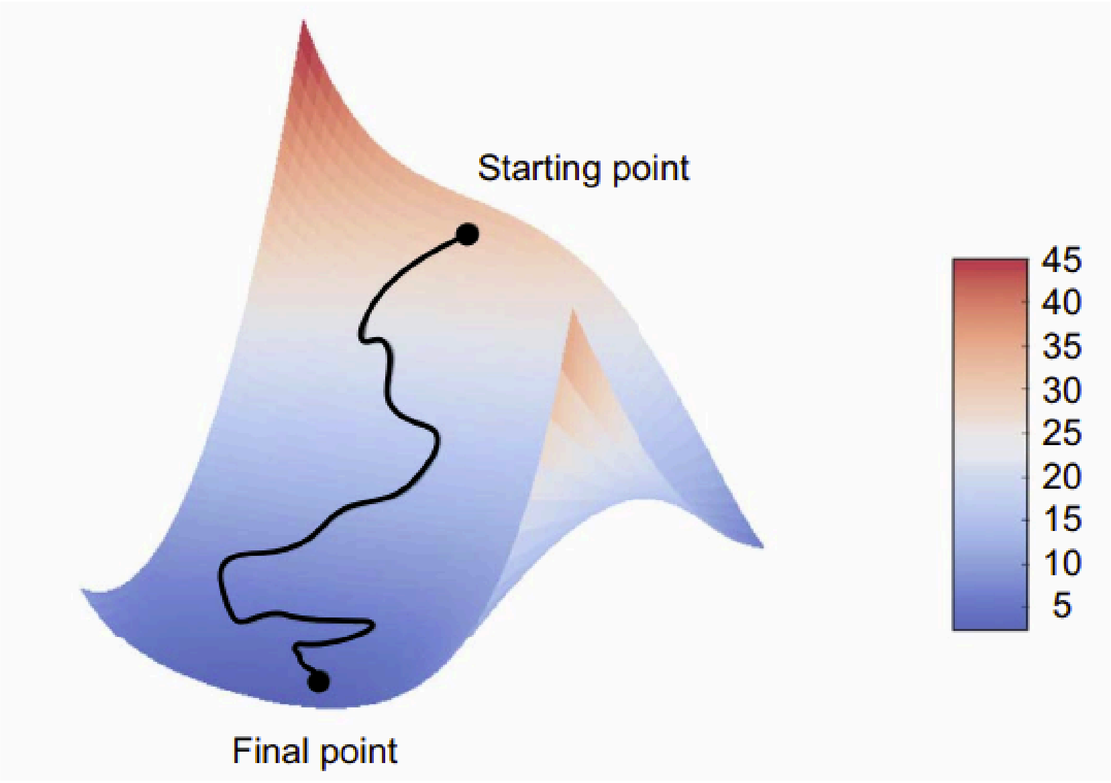
4.4 OPTIMAL WEIGHT ALPHA STRATEGY

- Công cụ sử dụng:
  - Mạng Nơ-ron Sâu (DNN) được dùng để dự đoán lợi suất tương lai từ giá trị alpha.
- Cấu trúc mạng:
  - Gồm 3 lớp:
    - đầu vào (nhận alpha)
    - lớp ẩn (10 nút, dùng ReLU)
    - lớp đầu ra (dự đoán lợi suất)



4.4 OPTIMAL WEIGHT ALPHA STRATEGY

- Quá trình huấn luyện:
  - Sử dụng backpropagation và gradient descent để tối ưu hoá trọng số
  - Tránh overfitting bằng cách dùng tập xác thực riêng biệt.
- Hiệu quả:
  - Thiết lập nền tảng vững chắc cho chiến lược đầu tư chính xác và linh hoạt hơn nhờ tối ưu trọng số alpha.



Thử nghiệm Framework

Đánh Giá Hiệu Quả của Framework

- Xác định khả năng nhận diện alpha phù hợp với điều kiện thị trường

Đánh Giá Tri Thức Đa Nguồn

- Tích hợp thông tin từ các nguồn khác nhau
- Nâng cao khả năng hiểu thị trường và phân tích cảm tính

Cơ Chế Tính Điểm Tin Cậy

- Nghiên cứu hiệu quả giảm thiểu nguy cơ hallucination

Đánh Giá Hiệu Suất Chiến Lược Tổng Thể

- Đánh giá hiệu quả giao dịch

DATASET

Dữ liệu tài chính từ thị trường Trung Quốc, đặc biệt là Chỉ số SSE 50.
bao gồm 50 công ty lớn, tiêu biểu nhất của thị trường A-shares trên sàn Thượng Hải

Dataset

- **Dữ liệu chính:** Giá mở, giá đóng , giá cao nhất, giá thấp nhất, khối lượng giao dịch, và Volume Weighted Average Price - Giá trung bình có trọng số theo khối lượng.
- **Yếu tố Alpha:** giá, khối lượng, chỉ số tài chính, trung bình động, phân tích tâm lý.
- **Báo cáo tài chính:** báo cáo quý và năm của các công ty SSE 50.
- **Giai đoạn:** Huấn luyện (2021-2022), kiểm tra (2023).
- **Phạm vi thị trường:** SSE 50.
- **Đánh giá:** mối quan hệ nhân quả, hiệu suất Alpha và độ ổn định mô hình.

Table 2: Summary of the experiment dataset

Aspect	Details
Primary Features	Open, Close, High, Low, Volume, VWAP
Alpha Factors	Custom factors based on price, volume, financial ratios, moving averages, sentiment analysis
Financial Reports	Quarterly and Annual reports from SSE 50 constituent companies
Time Periods	- Training: Jan 1, 2021 - Dec 31, 2022 - Test: Jan 1, 2023 - Dec 31, 2023
Market Coverage	SSE 50 Index, covering major sectors in the Chinese A-shares market
Evaluation Criteria	Causal relationships, Alpha factor performance, model robustness

Câu hỏi 1

Khung mô hình có tìm alpha mới theo điều kiện thị trường không?

Đầu vào

- tin tức tài chính
- báo cáo tài chính công ty
- biểu đồ giao dịch

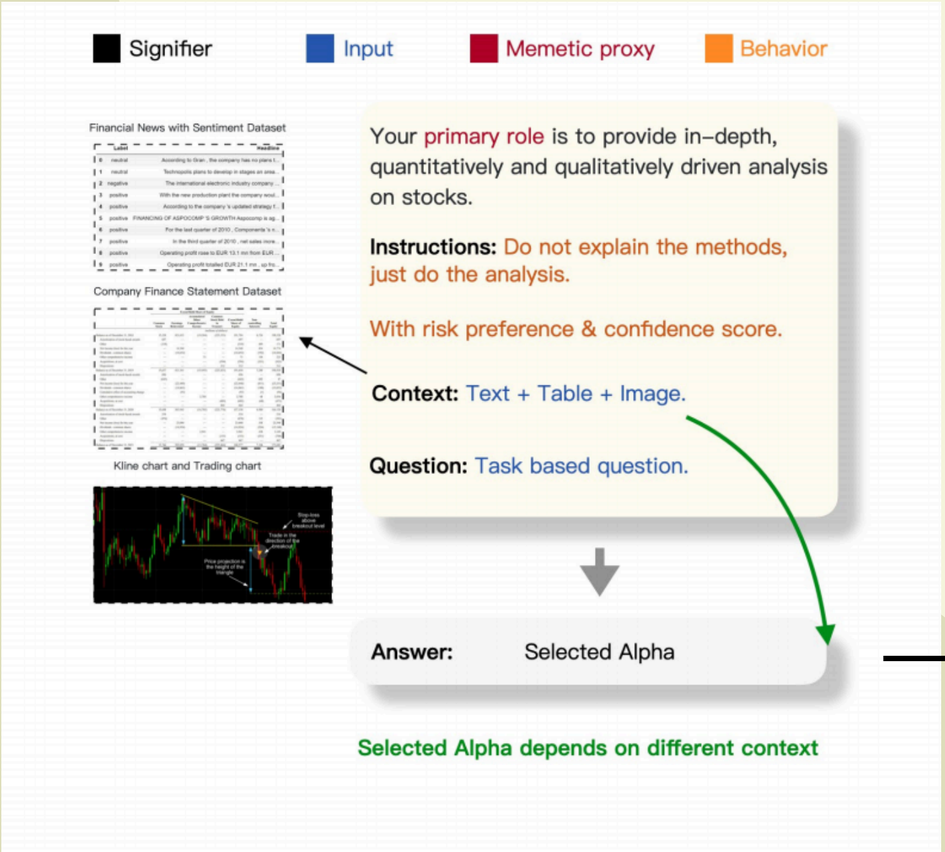


Figure 4: Multimodal LLM Analysis architecture

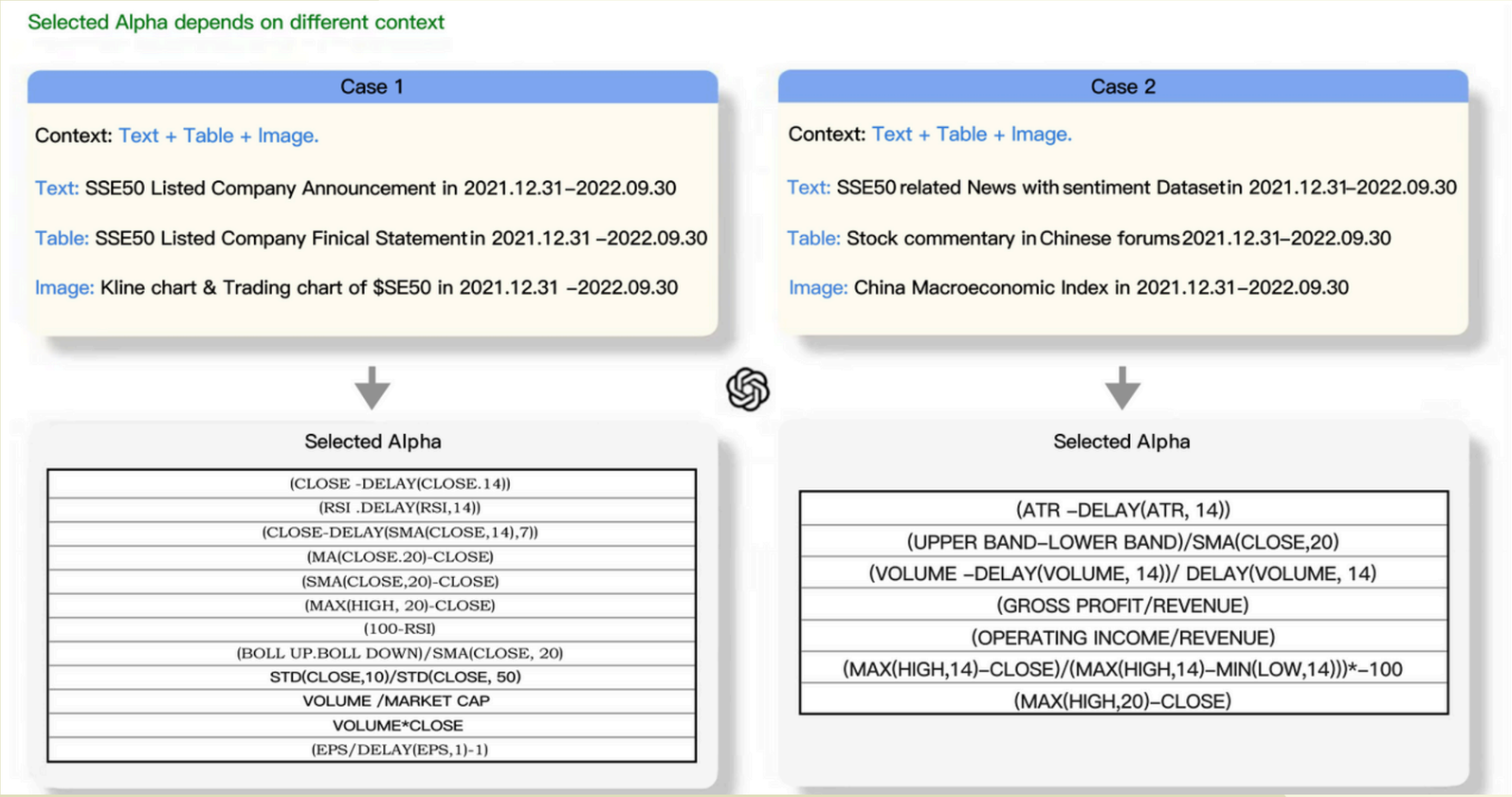
Đầu ra

Câu hỏi 1

Khung mô hình có tìm alpha mới theo điều kiện thị trường không?

Case 1: Dữ liệu tập trung vào thông báo, báo cáo

=> Alpha nghiêng về động lượng giá và khối lượng.



Case 2: Dữ liệu tập trung vào tin tức thị trường, chỉ số vĩ mô

=> Alpha thiên về biến động (volatility) và yếu tố kinh tế.

Câu hỏi 2

RQ2: LLM có khai thác alpha hiệu quả hơn alpha factory hiện tại không?

IC (Information Coefficient)

đo lường độ tương quan giữa giá trị alpha dự đoán và lợi nhuận thực tế.

=> IC càng cao, alpha càng dự báo tốt lợi nhuận tương lai.

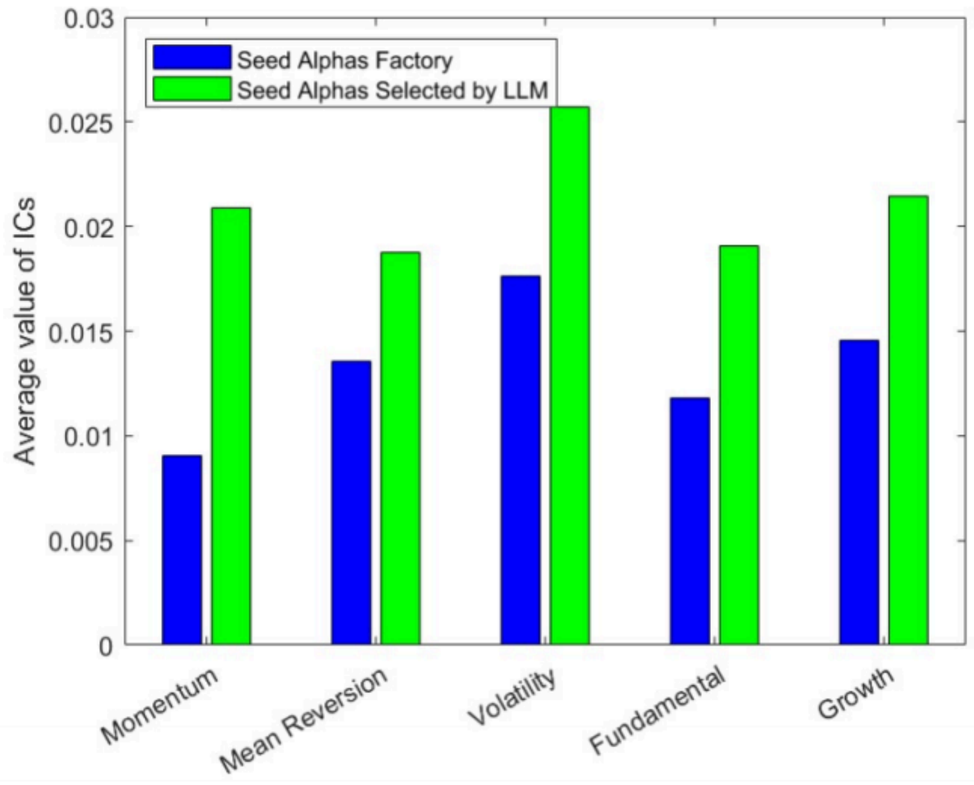


Figure 6: Average IC Comparation after LLM Selection

Câu hỏi 3

Chiến lược dựa trên LLM có thể duy trì lợi nhuận cao hơn thị trường không?

Backtest:

- Danh mục đầu tư có 13 cổ phiếu, nhưng nếu một số cổ phiếu không còn nằm trong top 13 theo alpha, hệ thống sẽ bán những mã đó và mua những mã mới.
- Để kiểm soát chi phí giao dịch, nó sẽ không mua/bán hơn 5 mã cổ phiếu trong một ngày.

=> Mục tiêu là tối ưu hóa danh mục mà không phát sinh quá nhiều chi phí giao dịch.

Câu hỏi 3

Chiến lược dựa trên LLM có thể duy trì lợi nhuận cao hơn thị trường không?

Lợi nhuận tích luỹ

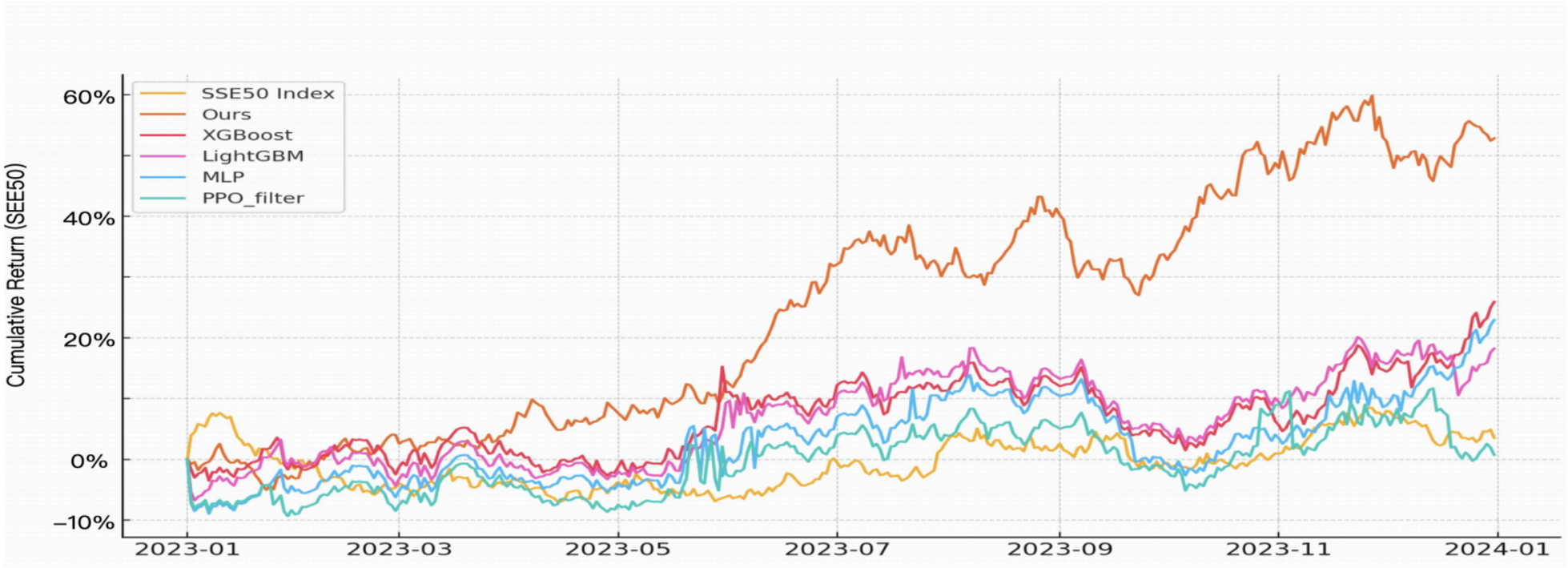


Figure 7: Backtest Result on SSE50 Comparison Between Index and Our performance

Conclusion

Giới thiệu Framework đầu tư chứng khoán định lượng bằng cách kết hợp LLMs và multi-agent.

1. **LLMs** được sử dụng để khai thác và tạo ra các yếu tố alpha từ nhiều nguồn thông tin tài chính khác nhau.
2. **Multi-agent** đánh giá điều kiện thị trường và xây dựng các chiến lược giao dịch cá nhân hóa theo mức độ rủi ro khác nhau.
3. **Học sâu** về trọng số sẽ chọn lọc và điều chỉnh trọng số của các tác nhân giao dịch theo thời gian thực.

THANK YOU