

BÁO CÁO PHÂN TÍCH BÀI BÁO

*ChatGPT, Trí tuệ Nhân tạo Tạo sinh, Mô hình Ngôn ngữ Lớn
và Tư vấn Đầu tư*

Tác giả chính của bài báo:

Lei Huang, Fangzhou Lu, Sixuan Li

Người thực hiện báo cáo:

Nguyễn Vũ Quang Anh

Trịnh Tuấn Ngọc Bảo

Phan Trần Mạnh Cường

Ngày thực hiện: Ngày 3 tháng 6 năm 2025

Mục lục

1	Giới thiệu	4
1.1	Bối cảnh	4
1.2	Nội dung nghiên cứu	4
1.3	Mục tiêu nghiên cứu	4
1.4	Cơ sở lý thuyết và nghiên cứu liên quan	4
1.5	Phương pháp nghiên cứu	5
1.5.1	Mục tiêu	5
1.5.2	Dữ liệu sử dụng	5
1.5.3	Phương pháp tiếp cận	5
1.5.4	Đánh giá hiệu suất	6
1.5.5	Kiểm tra độ tin cậy	6
1.5.6	So sánh giữa các phiên bản ChatGPT	6
1.6	Kết quả chính	7
1.7	Kết luận	7
2	Chi tiết tổ chức	8
2.1	Trí tuệ nhân tạo Tạo sinh và ChatGPT	8
2.2	ChatGPT Fine-Tuning	8
2.3	Dữ liệu văn bản	9
2.3.1	Wall Street Journal (WSJ)	9
2.3.2	China State Council Policy Release	10
2.3.3	Ý nghĩa kết hợp dữ liệu:	11
2.3.4	Kết quả chính:	12
3	Phương pháp	12
3.1	Hướng dẫn khuyến nghị cổ phiếu tại Mỹ dựa trên Wall Street Journal	12
3.2	Hướng dẫn khuyến nghị ngành tại Mỹ dựa trên Wall Street Journal	12
3.3	Ví dụ minh họa	12
3.4	Cải thiện khả năng khuyến nghị của ChatGPT thông qua tinh chỉnh	13
3.5	Đánh giá tính hợp lý của các khuyến nghị	13
3.6	Thống kê đặc điểm cổ phiếu và bản tin chính sách	14
4	Testing ChatGPT sử dụng dữ liệu chính sách của Mỹ	16
4.1	Kiểm tra lập luận dựa trên đánh giá của con người	16
4.2	Thị trường chứng khoán Mỹ:Kiểm định trong mẫu (In-sample test)	18
4.2.1	Mục tiêu và thiết kế nghiên cứu	18
4.2.2	Kết quả kiểm tra trong mẫu	20
4.2.3	Kết luận	22
4.3	Thị trường chứng khoán Mỹ: Kiểm định ngoài mẫu (Out-of sample test)	22
4.3.1	Giới thiệu	22
4.3.2	Phương pháp kiểm tra	22
4.3.3	Kết quả ngoài mẫu	22
4.3.4	Kết luận	24
4.4	Thị trường chứng khoán Mỹ: Kết quả sau khi tinh chỉnh mô hình (Fine-tuning)	24
4.4.1	Giới thiệu và phương pháp	24
4.4.2	Kết quả thử nghiệm	24
4.4.3	So sánh với phân tích văn bản truyền thống	26

4.4.4	Kết quả trực quan hóa	26
5	Testing ChatGPT sử dụng dữ liệu chính sách Trung Quốc	27
5.1	Bài kiểm tra lập luận dựa trên đánh giá của con người	28
5.1.1	Giới thiệu	28
5.1.2	Phương pháp kiểm tra	29
5.1.3	Kết quả đánh giá	29
5.1.4	Kết luận	29
5.2	Thị trường chứng khoán Trung Quốc: Kiểm định trong mẫu (In sample test)	29
5.2.1	Thiết kế thực nghiệm	30
5.2.2	Kết quả thực nghiệm	30
5.2.3	Kết luận	31
5.3	Thị trường cổ phiếu Trung Quốc: Kiểm định ngoài mẫu (Out-of-sample Test)	31
5.3.1	Giới thiệu và phương pháp	31
5.3.2	Thiết kế thực nghiệm	32
5.3.3	Kết quả thực nghiệm	32
5.3.4	Kết luận	34
5.4	Thị trường chứng khoán Trung Quốc: Kết quả sau tinh chỉnh mô hình (Fine-Tuning)	34
5.4.1	Giới thiệu	34
5.4.2	Phương pháp tinh chỉnh	34
5.4.3	Thiết kế thực nghiệm	34
5.4.4	Kết quả thực nghiệm	34
5.4.5	So sánh với phương pháp phân tích truyền thống	36
5.4.6	Tổng kết	36
6	Kiểm tra tính ổn định	36
6.1	Thiết kế kiểm tra	36
6.2	Kết quả kiểm tra	37
7	Kết luận	37

Tóm tắt

Các mô hình ngôn ngữ lớn (LLMs), bao gồm cả ChatGPT, tận dụng khả năng suy luận của mình để tạo ra các đề xuất danh mục đầu tư vượt xa khả năng phân tích văn bản thông thường. Bằng cách tinh chỉnh mô hình LLM với dữ liệu huấn luyện cụ thể và điều chỉnh các tham số của ChatGPT để tăng cường tính linh hoạt đầu ra, có tiềm năng để ChatGPT tạo ra các danh mục đầu tư vượt trội so với các chuẩn mực thị trường trong các thử nghiệm ngoài mẫu.

Sử dụng hai loại dữ liệu văn bản khác nhau bằng các ngôn ngữ khác nhau – các bài báo từ Wall Street Journal ở Hoa Kỳ và các thông báo chính sách từ chính phủ Trung Quốc – nhóm tác giả chỉ ra rằng ChatGPT có thể tạo ra các danh mục đầu tư với hệ số **alpha ba yếu tố hàng tháng lên tới 3%**, đặc biệt là khi phản ứng với các tin tức liên quan đến chính sách. Khi so sánh kết quả này với những phương pháp phân tích văn bản truyền thống, nghiên cứu cho thấy phương pháp truyền thống không tạo ra được các danh mục đầu tư có alpha dương.

1 Giới thiệu

1.1 Bối cảnh

Trong những năm gần đây, trí tuệ nhân tạo (AI) đã trở thành một trong những công nghệ chủ chốt định hình lại nhiều lĩnh vực, trong đó có tài chính. Đặc biệt, sự phát triển của các mô hình **Generative AI** như ChatGPT – một mô hình ngôn ngữ lớn (LLM) dựa trên kiến trúc Transformer – đã mở ra tiềm năng to lớn trong việc xử lý và khai thác dữ liệu văn bản khổng lồ trên Internet, báo chí, mạng xã hội, . . . để hỗ trợ ra quyết định đầu tư.

Tuy nhiên, trong lĩnh vực tài chính, việc ra quyết định đầu tư đòi hỏi không chỉ dựa trên dữ liệu định lượng truyền thống mà còn cần phân tích dữ liệu phi cấu trúc như tin tức, chính sách, tâm lý thị trường. Điều này đặt ra câu hỏi quan trọng: liệu một mô hình Generative AI như ChatGPT, vốn không được huấn luyện đặc biệt cho lĩnh vực tài chính, có thể cung cấp những gợi ý đầu tư hợp lý dựa trên dữ liệu văn bản không?

1.2 Nội dung nghiên cứu

Nghiên cứu của Lei Huang, Fangzhou Lu và Sixuan Li đã tìm hiểu về khả năng của ChatGPT trong việc trích xuất thông tin từ tin tức (WSJ, chính sách Trung Quốc) và gợi ý danh mục đầu tư. Bằng cách sử dụng ChatGPT với các tham số khác nhau (temperature, top-p), đồng thời so sánh với phân tích của chuyên gia tài chính, nhóm nghiên cứu kiểm tra:

- ChatGPT có thể tạo ra các danh mục đầu tư hiệu quả cao hay không?
- Độ tin cậy của các gợi ý này khi áp dụng vào thực tế (in-sample và out-of-sample)?
- Sự khác biệt giữa ChatGPT 3.5 và ChatGPT 4.0, cùng tác động của fine-tuning trên dữ liệu tài chính.

1.3 Mục tiêu nghiên cứu

Nghiên cứu kiểm tra tiềm năng ứng dụng Generative AI trong gợi ý đầu tư, đồng thời đánh giá tính khả thi, ưu điểm, hạn chế của cách tiếp cận này. Nghiên cứu không chỉ mang ý nghĩa học thuật mà còn có thể gợi mở hướng đi mới cho xây dựng các hệ thống robo-advisor thông minh.

1.4 Cơ sở lý thuyết và nghiên cứu liên quan

- **Generative AI và LLM:** GPT dựa trên kiến trúc Transformer, có khả năng học ngữ cảnh và sinh văn bản tự nhiên. ChatGPT là ứng dụng của LLM, được huấn luyện từ lượng lớn văn bản, có thể trả lời câu hỏi, tóm tắt và suy luận từ thông tin ngôn ngữ. Tham số ảnh hưởng: *temperature*, *top-p*, *context window*.
- **Ra quyết định đầu tư và phân tích dữ liệu phi cấu trúc:** Dữ liệu lớn và phân tích cảm xúc thị trường (sentiment analysis) cho thấy tầm quan trọng của tin tức, chính sách, mạng xã hội.
- **Nghiên cứu liên quan:**
 - Bollen et al. (2011): dự báo Dow Jones từ cảm xúc Twitter.

- Fung et al. (2020): phân tích tin tức Trung Quốc và tác động thị trường.
- Wu et al. (2022): đánh giá hiệu quả mô hình Transformer (BERT, GPT) trên văn bản tài chính.
- **Khoảng trống nghiên cứu và đóng góp:** Ứng dụng trực tiếp ChatGPT để gợi ý đầu tư từ tin tức, chính sách vẫn còn mới mẻ. Nghiên cứu đóng góp bằng cách so sánh ChatGPT với chuyên gia, khảo sát tác động của tham số (temperature, top-p) và phiên bản (3.5 vs 4.0).

1.5 Phương pháp nghiên cứu

1.5.1 Mục tiêu

Mục tiêu của nghiên cứu này là kiểm định khả năng của mô hình ngôn ngữ lớn (LLMs), đặc biệt là ChatGPT, trong việc xử lý và phân tích dữ liệu văn bản tài chính không có cấu trúc, từ đó đưa ra khuyến nghị đầu tư một cách hợp lý và hiệu quả. Câu hỏi trung tâm là liệu ChatGPT có thể tạo ra các danh mục đầu tư có hiệu suất vượt trội so với các phương pháp phân tích văn bản truyền thống, xét theo chỉ tiêu lợi nhuận điều chỉnh theo rủi ro (alpha) trong mô hình Fama-French ba yếu tố và các mô hình mở rộng.

1.5.2 Dữ liệu sử dụng

Nghiên cứu sử dụng hai nguồn dữ liệu văn bản lớn và có tính đa dạng ngôn ngữ, bối cảnh:

- **Tin tức tài chính từ Wall Street Journal (WSJ)** trong giai đoạn từ tháng 1 năm 2020 đến tháng 8 năm 2023. Đây là nguồn dữ liệu chính trong phân tích thị trường Hoa Kỳ.
- **Thông báo chính sách chính thức từ Quốc vụ viện Trung Quốc (State Council)** và các bộ ngành trực thuộc, với dữ liệu thu thập từ năm 2004 đến năm 2023. Đây là nguồn dữ liệu chính trong phân tích thị trường Trung Quốc.

1.5.3 Phương pháp tiếp cận

Quy trình phân tích được thiết kế mô phỏng theo cách một nhà đầu tư tương tác với cổ văn tài chính, bao gồm các bước:

1. **Tiền xử lý dữ liệu:** Chuẩn hoá văn bản đầu vào, loại bỏ các yếu tố nhiễu, phân loại tin tức theo chủ đề bằng mô hình LDA (Latent Dirichlet Allocation), ví dụ như: chính trị, năng lượng, COVID-19,...
2. **Tạo prompt cho ChatGPT:** Thiết kế câu lệnh truy vấn (prompt) dưới dạng hội thoại mô phỏng vai trò của một chuyên gia phân tích tài chính, yêu cầu ChatGPT đọc và đưa ra danh sách 5 cổ phiếu phù hợp cho từng bài báo hoặc thông cáo chính sách.
3. **Điều chỉnh tham số đầu ra:** Sử dụng các tham số chính như *temperature* (kiểm soát mức độ sáng tạo và ngẫu nhiên) và *top-p* (nội suy theo xác suất) để đa dạng hóa phản hồi, từ đó mô phỏng các khuyến nghị từ nhiều chuyên gia tài chính khác nhau.

4. **So sánh với phương pháp truyền thống:** Áp dụng phương pháp phân tích văn bản kinh điển (như đo độ tương đồng cosine giữa nội dung tin tức và báo cáo tài chính/earning calls của công ty) để chọn ra cổ phiếu liên quan, làm cơ sở đối chiếu với kết quả của ChatGPT.

1.5.4 Đánh giá hiệu suất

Hiệu suất các danh mục được đánh giá bằng cách xây dựng danh mục đầu tư dựa trên khuyến nghị từ ChatGPT, sau đó đo lường lợi nhuận điều chỉnh rủi ro theo các mô hình định giá tài sản:

- Mô hình Fama-French 3 yếu tố (FF3): bao gồm yếu tố thị trường, quy mô và giá trị.
- Mô hình 5 yếu tố (FF5): bổ sung khả năng sinh lợi và đầu tư.
- Mô hình mở rộng (FF5 + MOM + REV): thêm yếu tố đà tăng (momentum) và đảo chiều ngắn hạn (reversal).

Hệ số alpha được sử dụng như một chỉ tiêu đánh giá chính. Nếu alpha dương và có ý nghĩa thống kê, danh mục được xem là tạo ra lợi nhuận vượt trội so với thị trường sau khi đã điều chỉnh các yếu tố rủi ro.

1.5.5 Kiểm tra độ tin cậy

Kiểm tra tính hợp lý: Nhờ chuyên gia và sinh viên tài chính đánh giá xem các gợi ý của ChatGPT có hợp lý hay không.

Kiểm tra khả năng dự báo bằng cách chia thành 2 giai đoạn

- **In-sample:** Giai đoạn từ tháng 1/2020 đến 9/2021 (với dữ liệu WSJ) – thời kỳ dữ liệu nằm trong tập huấn luyện của ChatGPT.
- **Out-of-sample:** Giai đoạn từ tháng 10/2021 đến 9/2023 – ChatGPT chưa được tiếp cận trong quá trình huấn luyện, do đó cho phép kiểm định khả năng dự báo thực sự.

Cách chia giai đoạn tương tự được áp dụng cho dữ liệu Trung Quốc (2004–2021 là in-sample, 2021–2023 là out-of-sample).

Thử nghiệm với các danh mục có thời gian nắm giữ từ 1 ngày đến 1 năm

1.5.6 So sánh giữa các phiên bản ChatGPT

Nghiên cứu tiến hành đối chiếu giữa các phiên bản khác nhau của mô hình:

- **ChatGPT 3.5:** Phiên bản phổ biến, ổn định và có hiệu suất tốt trong việc đưa ra khuyến nghị logic.
- **ChatGPT 4.0:** Phiên bản nâng cao, tuy nhiên kết quả cho thấy đôi khi hiệu suất không vượt trội so với 3.5.
- **Fine-tuning (Tùy chỉnh mô hình):** ChatGPT được tinh chỉnh bằng dữ liệu tài chính chuyên biệt (ví dụ: chọn cổ phiếu có hiệu suất cao nhất trong cùng ngành – cùng mã SIC – với khuyến nghị gốc của ChatGPT), giúp nâng cao độ chính xác và khả năng dự đoán của mô hình.

Việc tinh chỉnh giúp cải thiện đáng kể hiệu suất alpha out-of-sample, đặc biệt trong các nhóm tin tức chính sách và chính trị.

1.6 Kết quả chính

Nghiên cứu cho thấy ChatGPT có khả năng đưa ra các khuyến nghị đầu tư hợp lý và có hiệu quả dự báo nhất định. Cụ thể:

- **Tính hợp lý của khuyến nghị:** Trong quá trình đánh giá bằng tay, khoảng 85–90% khuyến nghị của ChatGPT được các chuyên gia tài chính và sinh viên chuyên ngành đánh giá là hợp lý. Điều này cho thấy mô hình có khả năng hiểu và xử lý văn bản chính sách, tin tức tài chính theo cách phù hợp với lập luận của con người.
- **Hiệu suất đầu tư in-sample:** Trong giai đoạn in-sample, danh mục đầu tư do ChatGPT đề xuất có hệ số alpha ba yếu tố (Fama-French) hàng ngày đạt tới 14.1 basis points (bps) khi tham số nhiệt độ được đặt là 1, tương đương khoảng 2.6% lợi nhuận mỗi tháng. Điều này cho thấy tiềm năng của mô hình trong việc nhận diện thông tin giá trị từ dữ liệu đã học.
- **Hiệu suất out-of-sample:** Dù hiệu suất có giảm, ChatGPT vẫn tạo ra danh mục có alpha dương và có ý nghĩa thống kê trong giai đoạn out-of-sample – đặc biệt với các tin tức chính trị, với alpha hàng ngày lên tới 7 bps (tương đương khoảng 1.4% mỗi tháng).
- **Tác động của fine-tuning:** Khi tinh chỉnh mô hình trên dữ liệu tài chính đặc thù, hiệu suất alpha tăng mạnh – lên tới 3% mỗi tháng cho các danh mục dựa trên tin tức chính trị hoặc chính sách Trung Quốc. Mô hình sau khi tinh chỉnh được đánh giá có khuyến nghị hợp lý trong hơn 93% trường hợp.
- **Khả năng vượt trội so với phân tích văn bản truyền thống:** Các phương pháp truyền thống (ví dụ: cosine similarity giữa nội dung tin tức và báo cáo tài chính) không tạo ra danh mục có alpha dương, cho thấy ChatGPT vượt trội trong việc trích xuất và suy luận từ thông tin phi cấu trúc.

1.7 Kết luận

Từ các kết quả thu được, nghiên cứu kết luận rằng ChatGPT – một mô hình Generative AI không được huấn luyện chuyên biệt cho tài chính – vẫn có khả năng hoạt động như một công cụ tư vấn đầu tư hiệu quả. Nhờ khả năng suy luận từ văn bản phức tạp và không cần cấu trúc cụ thể, ChatGPT có thể:

- Gợi ý cổ phiếu phù hợp với tin tức chính sách hoặc kinh tế,
- Thích ứng với các bối cảnh ngôn ngữ khác nhau (tiếng Anh, tiếng Trung),
- Tạo ra danh mục đầu tư có khả năng sinh lời vượt trội.

Nghiên cứu mở ra hướng ứng dụng tiềm năng cho việc phát triển các hệ thống robo-advisor dựa trên mô hình ngôn ngữ lớn. Đồng thời, kết quả cũng góp phần khẳng định vai trò ngày càng quan trọng của AI trong lĩnh vực tài chính định lượng, đặc biệt trong bối cảnh thị trường chịu ảnh hưởng ngày càng mạnh từ các thông tin chính sách và phi cấu trúc.

2 Chi tiết tổ chức

2.1 Trí tuệ nhân tạo Tạo sinh và ChatGPT

ChatGPT, viết tắt của Chatbot Generative Pre-trained Transformer, là một mô hình ngôn ngữ tiên tiến và tinh vi được phát triển bởi OpenAI. Là một ví dụ đại diện cho AI tạo sinh, nó đã được huấn luyện trên một lượng lớn dữ liệu văn bản đa dạng để hiểu các mẫu ngôn ngữ phức tạp và tạo ra các phản hồi liên quan đến ngữ cảnh trong một buổi trò chuyện. Thông qua việc sử dụng kiến trúc transformer và các kỹ thuật học máy tiên tiến, ChatGPT có thể xử lý và hiểu ngôn ngữ của con người một cách hiệu quả, giúp nó có khả năng tham gia vào các cuộc trò chuyện mạch lạc và phù hợp với ngữ cảnh.

Chat GPT thể hiện một loạt các chức năng, như viết sáng tạo, tạo nội dung và dịch ngôn ngữ. Đặc biệt, khả năng lý luận của nó nổi bật như một khía cạnh quan trọng và khác biệt, bao gồm tư duy logic, suy luận các mối liên hệ giữa các khía cạnh khác nhau, và rút ra các kết luận thông minh.

ChatGPT đã được trải qua việc đánh giá nghiêm ngặt nhằm kiểm tra khả năng lý luận của nó, bao gồm vượt qua accs bài kiểm tra tiêu chuẩn hoá như LSAT (Bài kiểm tra đầu vào trường luật), SAT (Bài kiểm tra đánh giá học thuật), GRE (Bài kiểm tra Hồ sơ sau đại học), ...

Với việc vượt qua các bài test trên, chứng minh được ChatGPT thực sự có khả năng lý luận ẩn tượng, như tư duy phản biện, hiểu đọc, phân tích dữ liệu và ra quyết định. Chứng cứ mạnh mẽ này càng củng cố lý do sử dụng ChatGPT để phân tích tin tức và thông báo chính trị cũng như đưa ra khuyến nghị đầu tư trong bài nghiên cứu này.

2.2 ChatGPT Fine-Tuning

Định nghĩa: Fine-tuning là quá trình huấn luyện lại một mô hình ngôn ngữ lớn đã được huấn luyện trước, ví dụ như ChatGPT, trên một tập dữ liệu chuyên biệt nhằm điều chỉnh mô hình cho phù hợp với một tác vụ hoặc lĩnh vực cụ thể (ví dụ: khuyến nghị cổ phiếu).

Lợi ích của Fine-tuning:

- **(1) Bổ sung kiến thức chuyên ngành:** ChatGPT được fine-tuning trên dữ liệu tài chính sẽ hiểu rõ hơn về ngữ cảnh, thuật ngữ, và đặc thù của lĩnh vực tài chính, từ đó tạo ra khuyến nghị cổ phiếu phù hợp và chính xác hơn.
- **(2) Phân tích theo sự kiện:** Mô hình có khả năng nhận diện tác động của các tin tức hoặc sự kiện (ví dụ: báo cáo thu nhập, ra mắt sản phẩm, thay đổi chính sách) đối với giá cổ phiếu, cung cấp các khuyến nghị kịp thời và hợp lý.
- **(3) Đánh giá cảm xúc:** ChatGPT fine-tuned có thể phân tích cảm xúc từ tin tức và chính sách (tích cực/tiêu cực), từ đó dự đoán phản ứng của thị trường.
- **(4) Cải thiện khả năng suy luận:** Fine-tuning giúp ChatGPT kết nối thông tin từ nhiều nguồn, phát hiện mẫu phức tạp và tạo ra khuyến nghị chính xác hơn.

Phương pháp Fine-tuning trong nghiên cứu:

- **(1) Tập dữ liệu:** Sử dụng tập dữ liệu từ tin tức Wall Street Journal (WSJ) và thông báo chính sách của Quốc vụ viện Trung Quốc.

- **(2) Cách fine-tuning:** Chọn khuyến nghị cổ phiếu từ ChatGPT, sau đó xác định cổ phiếu có hiệu suất thực tế cao nhất trong cùng ngành (theo mã SIC) ở tháng tiếp theo để làm câu trả lời tối ưu. Mô hình được huấn luyện lại trên các cặp “prompt – response tối ưu” này để cải thiện độ chính xác.

Giá trị của sự phức tạp (Virtue of Complexity):

- Theo Kelly, Malamud và Zhou (2022), việc tăng số lượng tham số của mô hình có thể cải thiện khả năng dự đoán ngoài mẫu.
- ChatGPT với hàng trăm tỷ tham số tận dụng triệt để nguyên lý này, cho phép khai thác mối liên hệ phức tạp giữa dữ liệu văn bản phi cấu trúc (tin tức, báo cáo tài chính, chính sách) và giá cổ phiếu.

Hạn chế của Fine-tuning và ChatGPT:

- **(1) Tính không minh bạch:** ChatGPT được coi là “hộp đen” vì người dùng không thể truy cập các tham số nội bộ hoặc các giá trị tham số sau khi fine-tuning. Điều này làm giảm khả năng đánh giá cách mô hình suy luận và tính hợp lệ của khuyến nghị.
- **(2) Giảm niềm tin của nhà đầu tư:** Sự không minh bạch có thể làm nhà đầu tư nghi ngờ về độ tin cậy của ChatGPT so với các mô hình LLM minh bạch hơn của Google hoặc Meta.
- **(3) Đòi hỏi tài nguyên:** Fine-tuning yêu cầu khối lượng dữ liệu lớn, tính toán phức tạp và chi phí cao, đồng thời dễ dẫn đến quá khớp nếu không được kiểm soát tốt.

2.3 Dữ liệu văn bản

Nghiên cứu sử dụng hai nguồn dữ liệu văn bản lớn và đa dạng ngôn ngữ để kiểm tra khả năng của ChatGPT trong việc xử lý và phân tích dữ liệu phi cấu trúc nhằm đưa ra khuyến nghị cổ phiếu.

2.3.1 Wall Street Journal (WSJ)

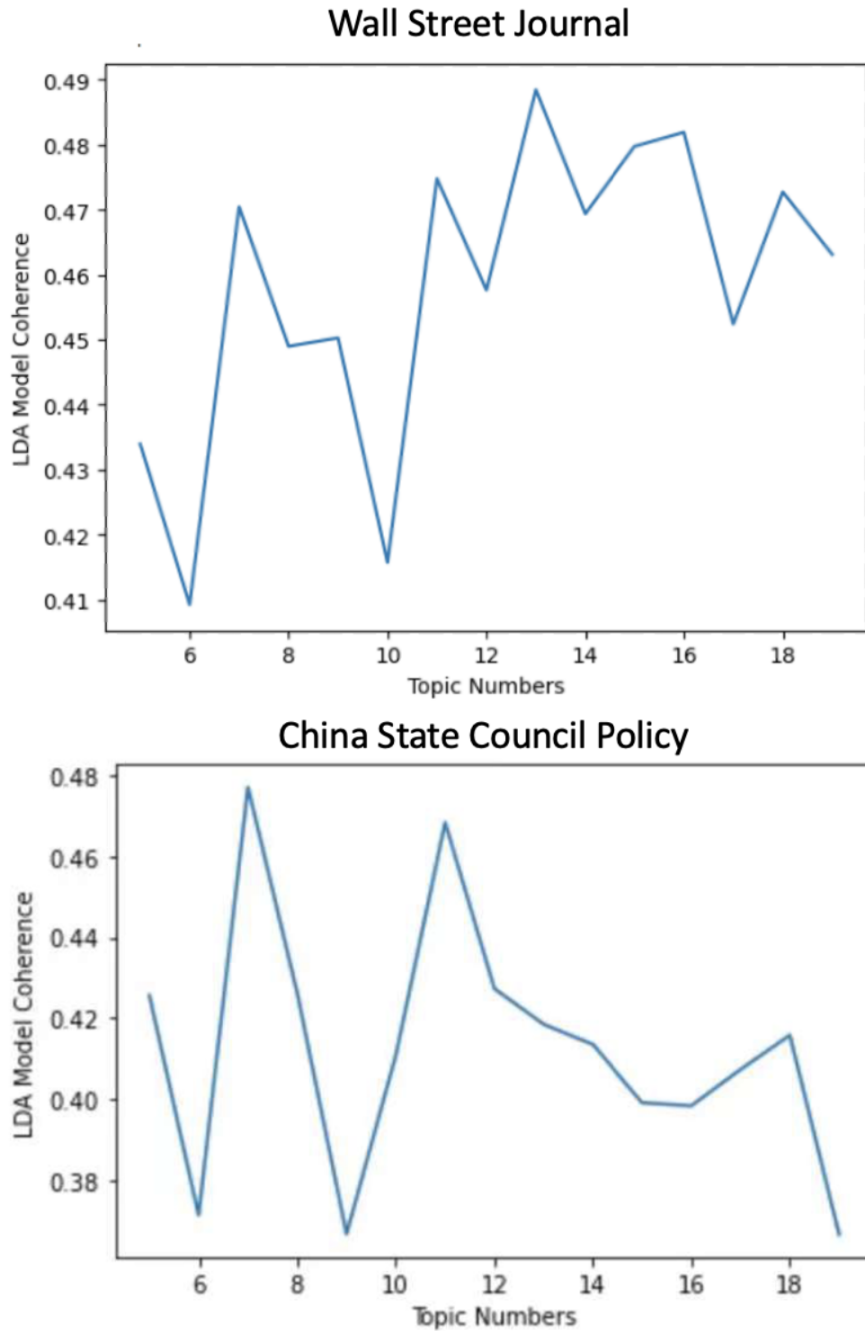
- **Nguồn dữ liệu:** Khoảng 16.000 bài báo từ WSJ, từ tháng 1/2020 đến tháng 8/2023.
- **Tiền xử lý dữ liệu:** Sử dụng Latent Dirichlet Allocation (LDA) để phân loại bài báo thành 13 nhóm chủ đề tối ưu, dựa trên điểm Coherence.
- **Một số chủ đề chính:**
 - Chủ đề 2: Giá cả và lạm phát.
 - Chủ đề 3: Chính trị.
 - Chủ đề 4: COVID-19.
 - Chủ đề 5: Năng lượng.
 - Chủ đề 8: Lao động.
 - Chủ đề 13: Nga và Ukraine.

- **Đặc điểm dữ liệu:**

- Độ dài trung bình bài viết: 753 từ ($SD = 389$).
- Đặc điểm cổ phiếu được khuyến nghị: beta trung bình 1.03, log vốn hoá 22.9, book-to-market 0.79, momentum 0.19.

2.3.2 China State Council Policy Release

- **Nguồn dữ liệu:** Các thông báo chính sách của Quốc vụ viện Trung Quốc từ 2004 đến tháng 8/2023.
- **Ý nghĩa:** Kiểm tra khả năng của ChatGPT trong xử lý văn bản tiếng Trung và cung cấp góc nhìn toàn cầu về kinh tế Trung - Mỹ.
- **Nội dung chính sách:** Bao gồm phát triển ngành, đổi mới công nghệ, công nghiệp xanh, cân bằng khu vực, chính sách tài chính và ưu đãi thuế.
- **Đặc điểm dữ liệu:**
 - Độ dài trung bình bài viết: 1.325 từ ($SD = 1.109$).
 - Đặc điểm cổ phiếu được khuyến nghị: beta trung bình 1.06, log vốn hoá 20.4, book-to-market 0.41, momentum 0.13.



Hình 1: *
Figure 5: Điểm nhất quán (Coherence) của LDA

2.3.3 Ý nghĩa kết hợp dữ liệu:

- Sử dụng dữ liệu từ cả WSJ và chính sách Trung Quốc cho phép kiểm tra khả năng của ChatGPT trong xử lý thông tin đa ngôn ngữ và đa bối cảnh.
- Dữ liệu chính sách Trung Quốc đặc biệt hữu ích để đánh giá khả năng của ChatGPT trong phản ứng với tin tức có tính định hướng chính sách, vốn ảnh hưởng mạnh đến thị trường chứng khoán Trung Quốc.

2.3.4 Kết quả chính:

- ChatGPT có thể đề xuất danh mục đầu tư với alpha vượt trội (tới 3%/tháng) từ cả hai nguồn dữ liệu.
- Phân tích văn bản truyền thống (như cosine similarity giữa tin tức và báo cáo tài chính) không tạo ra danh mục đầu tư có alpha dương, khẳng định ưu thế của ChatGPT trong xử lý dữ liệu phi cấu trúc.

3 Phương pháp

Nhóm tác giả đã sử dụng một truy vấn hội thoại đơn giản và trực tiếp để trích xuất các khuyến nghị đầu tư, tương tự như cách một cá nhân sẽ giao tiếp với cố vấn đầu tư của mình. Dưới đây là mẫu prompt (câu lệnh truy vấn) được sử dụng để trích xuất khuyến nghị cổ phiếu của ChatGPT.

3.1 Hướng dẫn khuyến nghị cổ phiếu tại Mỹ dựa trên Wall Street Journal

Hướng dẫn khuyến nghị cổ phiếu tại Mỹ dựa trên Wall Street Journal có nội dung như sau: "Giả sử bạn là một nhà phân tích tài chính cấp cao. Vui lòng đọc kỹ bản tin Wall Street Journal sau đây. Theo nội dung của bản tin, nếu bạn có khuyến nghị cổ phiếu, hãy viết "YES" và liệt kê 5 tên cổ phiếu NYSE hoặc Nasdaq cùng với mã cổ phiếu mà bạn khuyến nghị. Nếu bạn chỉ trả lời "NO", vui lòng giải thích ngắn gọn lý do."

3.2 Hướng dẫn khuyến nghị ngành tại Mỹ dựa trên Wall Street Journal

Tương tự, hướng dẫn khuyến nghị ngành tại Mỹ dựa trên Wall Street Journal được thiết lập như sau: "Giả sử bạn là một nhà phân tích tài chính cấp cao. Vui lòng đọc kỹ bản tin chính sách sau đây. Dựa trên nội dung của bản tin chính sách và tham chiếu tiêu chuẩn phân loại SIC hai chữ số, nếu bạn có khuyến nghị ngành, hãy viết "YES" và liệt kê mã danh mục của ngành được khuyến nghị. Nếu bạn chỉ trả lời "NO", vui lòng giải thích ngắn gọn lý do."

3.3 Ví dụ minh họa

Phụ lục (Appendix) trình bày một ví dụ về cổ phiếu mà ChatGPT đề xuất. Cụ thể, vào ngày 14 tháng 5 năm 2023, Tạp chí Phố Wall đã xuất bản một bài báo có tiêu đề: *Biden to Withhold Tariffs on Solar Imports*. Bài báo này thảo luận về những người thắng và thua từ lệnh hành pháp mới này của Biden. Nhóm tác giả đã thực hiện một vòng truy vấn khi tham số nhiệt độ được đặt là một cho ChatGPT 3.5 và ChatGPT đã khuyến nghị tổng cộng năm cổ phiếu.

3.4 Cải thiện khả năng khuyến nghị của ChatGPT thông qua tinh chỉnh

Tiếp theo, nhóm tác giả cải thiện khả năng khuyến nghị cổ phiếu của ChatGPT dựa trên fine-tuning (tinh chỉnh). Cụ thể, đối với Wall Street Journal, họ tiền huấn luyện ChatGPT dựa trên tập dữ liệu mẫu từ năm 2018 đến 2019. Hoặc, đối với mỗi bài báo, họ cung cấp cho ChatGPT năm cổ phiếu có hiệu suất tốt nhất với cùng mã SIC hai chữ số dựa trên khuyến nghị ban đầu, và sử dụng mẫu này để tinh chỉnh đầu ra của ChatGPT.

3.5 Đánh giá tính hợp lý của các khuyến nghị

Để kiểm tra xem lý do đằng sau các khuyến nghị cổ phiếu có hợp lý hay không và liệu các cổ phiếu được ChatGPT khuyến nghị có phù hợp với đánh giá của một nhà phân tích tài chính con người hay không, nhóm tác giả đã thuê các chuyên gia kỳ cựu trong ngành kết hợp với sinh viên chuyên ngành tài chính. Các nhà phân tích tài chính được thuê đã vượt qua kỳ thi CFA Level 3 và các sinh viên đã vượt qua ít nhất kỳ thi CFA Level 1.

Quy trình đánh giá bao gồm các bước sau:

- Thứ nhất, họ cung cấp cho những người tham gia một cái nhìn tổng quan ngắn gọn về ChatGPT và khả năng của nó trong việc tạo ra các khuyến nghị cổ phiếu dựa trên các thông báo chính sách.
- Sau đó, họ hướng dẫn những người tham gia đánh giá các khuyến nghị và quyết định liệu khuyến nghị đó có hợp lý (logical), không rõ ràng (unclear) hay hoàn toàn không liên quan (clearly unrelated) dựa trên các tiêu chí sau:
 - (a) Mức độ liên quan đến ngành được chỉ định trong thông báo chính sách.
 - (b) Tính vững chắc của lý do đằng sau các khuyến nghị.
 - (c) Các cơ hội đầu tư tiềm năng dựa trên chính sách và các cổ phiếu được khuyến nghị.
- Đối với mỗi khuyến nghị cổ phiếu, nhóm tác giả yêu cầu hai nhà phân tích tài chính đánh giá.
- Nếu họ đạt được cùng một kết luận, thì kết quả tương ứng sẽ được ghi nhận trực tiếp.
- Nếu hai nhà phân tích tài chính có ý kiến khác nhau, họ yêu cầu nhà phân tích thứ ba đưa ra một đánh giá khác, và kết quả theo đa số sẽ được chấp nhận.
- Nếu cả ba nhà phân tích tài chính đưa ra các kết quả khác nhau, mẫu này sẽ bị loại bỏ.

Đối với 92% mẫu, kết quả được đưa ra ngay trong vòng đầu tiên, và trong 8% trường hợp, một nhà phân tích thứ ba được mời. Không có trường hợp nào mà nhà phân tích thứ ba đưa ra kết luận thứ ba hoàn toàn khác biệt.

3.6 Thống kê đặc điểm cổ phiếu và bản tin chính sách

Panel A: Recommended stocks in U.S.				
Stock Characteristics				
	Mean	STD	5th	95th
BETA	1.03	0.46	0.26	1.74
SIZE	22.90	4.75	17.12	24.22
BM	0.79	0.82	0.08	1.90
INV	0.50	0.30	0.35	0.91
PROFIT	0.10	2.37	-1.13	0.61
MOM	0.19	0.99	-0.63	1.32
REV	0.01	0.22	-0.23	0.27
Wall Street Journal				
	Mean	STD	5th	95th
Words	753	389	234	1346

Panel B: Recommended stocks in China				
Stock Characteristics				
	Mean	STD	5th	95th
BETA	1.06	0.73	0.12	1.97
SIZE	20.35	0.92	19.74	22.53
BM	0.41	0.25	0.12	0.82
INV	0.34	0.51	0.14	0.92
PROFIT	0.14	0.71	-0.42	0.53
MOM	0.13	0.82	-0.35	0.71
REV	0.03	0.41	-0.12	0.20
Policy News				
	Mean	STD	5th	95th
Words	1325	1109	730	1942

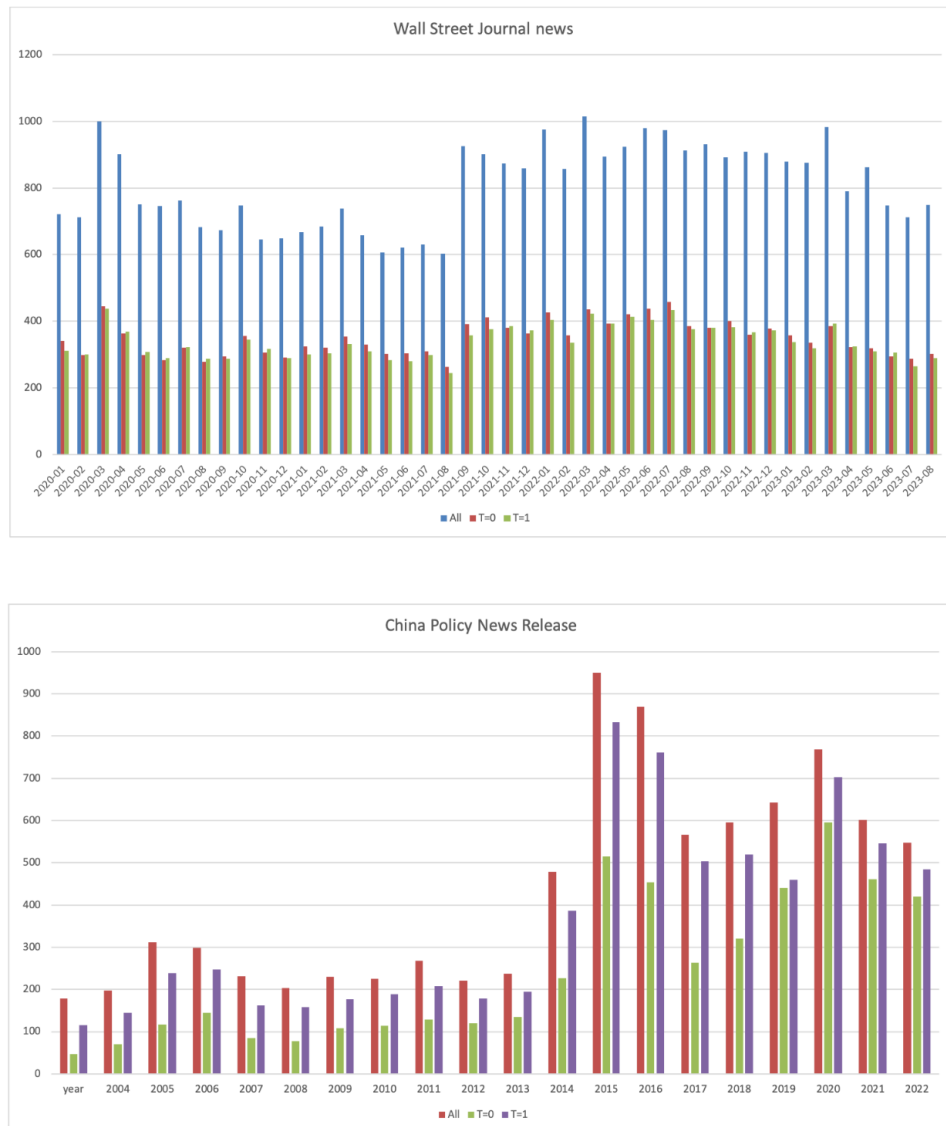
Hình 2: *

Table 1: Tóm tắt thống kê

Bảng 1 (Table 1) trình bày thống kê tóm tắt về đặc điểm cổ phiếu và phân bố số lượng từ trong các bài báo Wall Street Journal và các bản tin chính sách của Hội đồng Nhà nước Trung Quốc.

- **Panel A** hiển thị đặc điểm của cổ phiếu được ChatGPT khuyến nghị tại thị trường Hoa Kỳ và phân bố số lượng từ của các bài báo Wall Street Journal trong mẫu từ tháng 1 năm 2020 đến tháng 8 năm 2023. Các đặc điểm cổ phiếu bao gồm beta Scholes-Williams-Dimson (BETA), vốn hóa thị trường (SIZE), tỷ lệ sách trên thị trường Fama-French (BM), đầu tư (INV), lợi nhuận hoạt động (PROFIT), động lượng Jegadeesh-Titman (MOM), và đảo chiều ngắn hạn Jegadeesh-Lehmann (REV). Phân bố số lượng từ trong các bài báo Wall Street Journal được đặc trưng bởi giá trị trung bình, độ lệch chuẩn, và phần trăm vị thứ 5 và 95 của số lượng từ trên mỗi bài báo.
- **Panel B** báo cáo đặc điểm của cổ phiếu được ChatGPT khuyến nghị tại thị trường

Trung Quốc và phân bố số lượng từ trong các bản tin chính sách của Quốc vụ viện Trung Quốc trong mẫu từ tháng 1 năm 2004 đến tháng 8 năm 2023. Các đặc điểm cổ phiếu tương tự như những đặc điểm được trình bày trong Panel A, cung cấp cái nhìn sâu sắc về các cổ phiếu được khuyến nghị của thị trường Trung Quốc. Phân bố số lượng từ của các bản tin chính sách của Quốc vụ viện Trung Quốc cũng được tóm tắt bằng giá trị trung bình, độ lệch chuẩn, và phần trăm vị thứ 5 và 95 của số lượng từ trên mỗi bản tin.



Hình 3: *

Figure 2: Số lượng báo cáo phát hành chính sách trong chuỗi thời gian

Hình 2 (Figure 2) hiển thị số lượng bản tin chính sách được phát hành theo năm trong giai đoạn mẫu của nghiên cứu, với tổng số hơn 20.000 bản tin Wall Street Journal. ChatGPT nhận thấy khoảng dưới một nửa số bản tin chính sách này có liên quan đến khuyến nghị cổ phiếu theo chuỗi thời gian. Kết quả này không thay đổi nhiều tùy theo tham số nhiệt độ được đặt. Nhìn chung, các thống kê tóm tắt này làm nổi bật sự đa dạng về cả đặc

điểm cổ phiếu và nội dung tin tức chính sách, điều này cung cấp một bối cảnh phong phú cho phân tích của nhóm tác giả

4 Testing ChatGPT sử dụng dữ liệu chính sách của Mỹ

4.1 Kiểm tra lập luận dựa trên đánh giá của con người

Trong phần này, nhóm nghiên cứu tiến hành kiểm tra khả năng lập luận của ChatGPT trong việc đưa ra khuyến nghị đầu tư dựa trên nội dung chính sách tại Mỹ (Wall Street Journal). Do số lượng lớn các tin tức chính sách và khuyến nghị cổ phiếu, nhóm nghiên cứu đã lựa chọn ngẫu nhiên 100 mẫu tin tức chính sách cho mỗi lần kiểm tra. Mỗi khuyến nghị đầu tư được đánh giá theo ba tiêu chí:

- Tính liên quan (Relevance)
- Lý do hợp lý (Reasoning)
- Cơ hội đầu tư (Investment Opportunity)

Kết quả đánh giá được phân loại thành 3 nhóm:

- Hợp lý (Logical)
- Không rõ ràng (Unclear)
- Không liên quan (Unrelated)

Kết quả kiểm tra được tổng hợp trong **Bảng 2** với hai bảng con:

Panel A: Financial Analyst			
ChatGPT 3.5			
	Domestic Equity (T=0)	Domestic Equity (T=1)	Domestic Equity (T=1.5)
Logical	89%	87%	88%
Unclear	11%	10%	5%
Unrelated	0%	3%	7%
	Sector (T=0)	Sector (T=1)	Sector (T=1.5)
Logical	95%	92%	90%
Unclear	4%	6%	4%
Unrelated	1%	2%	6%
ChatGPT 4.0			
	Domestic Equity (T=0)	Domestic Equity (T=1)	Domestic Equity (T=1.5)
Logical	82%	78%	79%
Unclear	15%	20%	16%
Unrelated	3%	2%	5%
	Sector (T=0)	Sector (T=1)	Sector (T=1.5)
Logical	90%	85%	83%
Unclear	7%	11%	7%
Unrelated	3%	4%	10%

Hình 4: *

Table 2a: Tư vấn đầu tư của ChatGPT và tính hợp lệ tại Mỹ

Panel B: Finance Students			
ChatGPT 3.5			
	Domestic Equity (T=0)	Domestic Equity (T=1)	Domestic Equity (T=1.5)
Logical	87%	89%	90%
Unclear	10%	10%	6%
Unrelated	3%	1%	4%
	Sector (T=0)	Sector (T=1)	Sector (T=1.5)
Logical	93%	91%	88%
Unclear	6%	7%	7%
Unrelated	1%	2%	5%
ChatGPT 4.0			
	Domestic Equity (T=0)	Domestic Equity (T=1)	Domestic Equity (T=1.5)
Logical	84%	76%	74%
Unclear	11%	15%	20%
Unrelated	5%	9%	6%
	Sector (T=0)	Sector (T=1)	Sector (T=1.5)
Logical	88%	83%	84%
Unclear	8%	13%	15%
Unrelated	4%	4%	1%

Hình 5: *

Table 2b: Tư vấn đầu tư của ChatGPT và tính hợp lệ tại Mỹ (tiếp theo)

(a) kết quả đánh giá của nhóm chuyên gia tài chính, và (b) kết quả đánh giá của nhóm sinh viên chuyên ngành tài chính. Ngoài ra, nghiên cứu cũng phân tích tác động của tham số nhiệt độ (**Temperature parameter T**) của ChatGPT - một tham số điều chỉnh mức độ sáng tạo trong phản hồi của mô hình. Cụ thể, ba giá trị nhiệt độ được sử dụng: **T = 0**, **T = 1**, **T = 1.5**. Mỗi giá trị này được áp dụng cho 2 nhóm khuyến nghị: **Domestic Equity** (khuyến nghị cổ phiếu riêng lẻ trên NYSE hoặc Nasdaq) và **Sector** (khuyến nghị ngành dựa trên mã ngành SIC hai chữ số).

Kết quả:

- Kết quả cho thấy, trong nhóm Domestic Equity, tỷ lệ khuyến nghị hợp lý dao động từ 74% đến 90%, tùy thuộc vào nhóm người đánh giá và mức nhiệt độ. Đối với nhóm Sector, tỷ lệ hợp lý cao hơn, dao động từ 83% đến 95%, cho thấy ChatGPT có khả năng đưa ra khuyến nghị ngành tốt hơn so với khuyến nghị cổ phiếu riêng lẻ.
- So sánh giữa ChatGPT 3.5 và ChatGPT 4.0, kết quả cho thấy ChatGPT 3.5 thường có tỷ lệ khuyến nghị hợp lý cao hơn so với ChatGPT 4.0 ở cả hai nhóm đánh giá và trong cả hai nhóm khuyến nghị. Ví dụ, trong nhóm Domestic Equity với T=0, tỷ lệ khuyến nghị được đánh giá là hợp lý bởi chuyên gia tài chính là 89% đối với ChatGPT 3.5, trong khi ChatGPT 4.0 chỉ đạt 82%. Xu hướng tương tự cũng xuất hiện trong đánh giá của sinh viên chuyên ngành tài chính.

Tóm lại, kết quả từ **Bảng 2** chỉ ra rằng khuyến nghị đầu tư của ChatGPT phần lớn là được đánh giá **hợp lý** đối với cả ChatGPT 3.5 và ChatGPT 4.0, trong đó hai nhóm Domestic Equity và Sector. Tỷ lệ khuyến nghị **không rõ ràng** và **không liên quan** đều thấp, cho thấy **tiềm năng của ChatGPT như một công cụ hỗ trợ tư vấn đầu tư**.

4.2 Thị trường chứng khoán Mỹ:Kiểm định trong mẫu (In-sample test)

4.2.1 Mục tiêu và thiết kế nghiên cứu

Phần kiểm định trong mẫu (In-sample test) được thiết kế nhằm đánh giá khả năng của mô hình ChatGPT trong việc tạo ra các danh mục cổ phiếu có hiệu suất vượt trội, dựa trên việc phân tích văn bản tài chính phi cấu trúc. Trong trường hợp này, dữ liệu văn bản được lấy từ các bài viết trên **Wall Street Journal (WSJ)**- một nguồn tin uy tín về tài chính và kinh doanh - trong giai đoạn từ **tháng 1 năm 2020 đến tháng 8 năm 2021**. Đây là khoảng thời gian được cho là nằm **trong đợt tập huấn luyện của ChatGPT**, nên mô hình có thể đã tiếp xúc với những dữ liệu này trong quá trình pre-training, từ đó có thể biểu hiện tốt hơn so với dữ liệu hoàn toàn mới (out-of-sample)
Để kiểm định một cách hệ thống, nhóm tác giả thực hiện các bước sau:

1. Tiền xử lý dữ liệu

- Chọn lọc các bài viết WSJ có liên quan đến tài chính hoặc kinh tế vĩ mô.
- Mỗi bài viết được dùng để tạo một **prompt** đầu vào cho ChatGPT, yêu cầu mô hình **đề xuất 5 cổ phiếu cụ thể** nên mua dựa trên nội dung của bài viết đó.
- Dữ liệu được chuẩn hoá để mô phỏng cách nhà đầu tư tương tác với một "robo-advisor AI"

2. Tạo danh mục đầu tư theo lịch trình : Sau khi ChatGPT đưa ra danh sách cổ phiếu, các **danh mục đầu tư (portfolio)** được xây dựng dựa trên mô hình **calendar-time**:

- Mỗi ngày (hoặc tuần/tháng) trong giai đoạn kiểm định, hệ thống sẽ xây dựng một danh mục mới dựa trên phản hồi từ CHatGPT cho các bài viết WSJ ngày hôm đó.
- Các danh mục có thời gian nắm giữ cố định: 1 ngày, 1 tuần, 1 tháng và 3 tháng.
=> Điều này mô phỏng tình huống thực tế nơi nhà đầu tư liên tục nhận được các đề xuất mới từ AI và duy trì các khoản đầu tư theo thời hạn đã định

3. Trọng số danh mục và biến thể đầu tư: Hai phương pháp gán trọng số cho cổ phiếu trong danh mục được áp dụng

- **Equal-weighted (EW)**: Mỗi cổ phiếu có trọng số bằng nhau
- **Value-weighted (VW)**: Trọng số theo vốn hoá thị trường => Điều này nhằm kiểm tra xem hiệu quả của ChatGPT có phụ thuộc vào cách gán trọng số hay không.

4. Kiểm định thống kê bằng mô hình định giá tài sản: Lợi nhuận của các danh mục được đánh giá bằng **mô hình hồi quy theo thời gian** (time-series regression) với các mô hình định giá phổ biến

- **CAPM**: kiểm tra alpha so với chỉ số thị trường
- **Fama-Fench 3 yếu tố (FF3)**: thêm yếu tố quy mô (SMB) và giá trị HML

- **Fama-Fench 5 yếu tố (FF5)**: thêm yếu tố lợi nhuận và đầu tư
- **FF5 + MOM + REV**: thêm yếu tố đà tăng (momentum) và đảo chiều ngắn hạn (reversal).
=> **Hệ số alpha** (lợi nhuận vượt trội điều chỉnh theo rủi ro) là tiêu chí đánh giá chính. Nếu alpha dương và có ý nghĩa thống kê, ChatGPT được xem là tạo ra danh mục đầu tư có giá trị thực sự.

5. **Biến thể về tham số ChatGPT: Temperature**: Hai thiết lập của tham số **temperature** trong ChatGPT được so sánh:

- **T = 0**: phản hồi nhất quán, bảo thủ (giống một chuyên gia cố định)
- **T = 1**: phản hồi đa dạng, sáng tạo (giống nhiều chuyên gia khác nhau). Đặc biệt khi **T = 1**, mỗi danh mục được tạo thành từ tất cả các cổ phiếu được Chat GPT đề xuất qua 25 lần truy vấn khác nhau, nhằm tận dụng sự đa dạng trong phản hồi của mô hình.
=> Mục tiêu là kiểm tra xem sự ngẫu nhiên và đa dạng trong phản hồi có giúp tăng khả năng tạo lợi nhuận hay không.

4.2.2 Kết quả kiểm tra trong mẫu

One-Week			CAPM	FF3	FF5	FF5 MOM+REV	
ChatGPT 3.5	VW	T=0	0.476 (0.070)	0.362 (0.063)	0.290 (0.048)	0.253 (0.038)	
		T=1	1.482 (0.239)	1.312 (0.204)	1.138 (0.194)	0.930 (0.115)	
		EW	T=0	0.531 (0.073)	0.638 (0.093)	0.472 (0.074)	0.371 (0.063)
			T=1	0.134 (0.021)	-0.135 (-0.030)	-0.139 (-0.021)	-0.314 (-0.054)
	VW	T=0	1.194 (0.178)	0.983 (0.104)	0.611 (0.085)	0.437 (0.062)	
		T=1	0.934 (0.154)	0.731 (0.112)	0.673 (0.095)	0.510 (0.080)	
		EW	T=0	0.425 (0.081)	0.291 (0.040)	0.148 (0.026)	-0.152 (-0.021)
			T=1	3.184 (0.462)	2.481 (0.382)	2.034 (0.305)	1.842 (0.298)
One-Day			CAPM	FF3	FF5	FF5 MOM+REV	
ChatGPT 3.5	VW	T=0	-0.184 (-0.031)	-0.427 (-0.064)	-0.138 (-0.028)	-0.228 (-0.036)	
		T=1	1.275 (0.191)	1.192 (0.173)	0.942 (0.159)	0.859 (0.139)	
		EW	T=0	0.821 (0.119)	0.738 (0.112)	0.571 (0.084)	0.412 (0.071)
			T=1	-0.428 (-0.084)	-0.398 (-0.072)	-0.285 (-0.051)	0.482 (0.073)
	VW	T=0	1.621 (0.245)	1.442 (0.221)	0.714 (0.115)	0.643 (0.101)	
		T=1	0.798 (0.109)	0.487 (0.085)	0.283 (0.041)	0.193 (0.027)	
		EW	T=0	1.984 (0.296)	1.526 (0.218)	1.413 (0.204)	1.034 (0.187)
			T=1	-0.831 (-0.134)	-0.415 (-0.074)	-0.425 (-0.076)	-0.312 (-0.063)

Three-Month			CAPM	FF3	FF5	FF5 MOM+REV
ChatGPT 3.5	VW	T=0	2.193 (0.348)	1.890 (0.221)	1.738 (0.244)	1.638 (0.261)
		T=1	7.387 (1.234)	6.145 (1.014)	5.847 (0.958)	4.629 (0.793)
	EW	T=0	1.391 (0.192)	0.958 (0.135)	0.895 (0.126)	0.743 (1.108)
		T=1	10.492 (1.598)	10.231 (1.561)	9.327 (1.332)	8.401 (1.194)
	VW	T=0	0.854 (0.134)	0.743 (0.110)	0.532 (0.083)	0.478 (0.061)
		T=1	0.780 (0.136)	0.675 (0.113)	0.578 (0.074)	0.532 (0.067)
	EW	T=0	0.974 (0.158)	0.815 (0.143)	0.678 (0.094)	0.452 (0.074)
		T=1	0.473 (0.072)	0.391 (0.066)	0.350 (0.062)	0.332 (0.057)
One-Month			CAPM	FF3	FF5	FF5 MOM+REV
ChatGPT 3.5	VW	T=0	3.592 (0.527)	3.403 (0.502)	3.621 (0.531)	3.383 (0.485)
		T=1	14.429** (2.128)	14.104** (2.084)	12.920* (1.905)	13.853** (2.043)
	EW	T=0	4.618 (0.614)	4.503 (0.593)	4.415 (0.583)	4.516 (0.605)
		T=1	16.251** (2.421)	15.104** (2.254)	14.791** (2.207)	13.560** (2.023)
	VW	T=0	2.184 (0.340)	2.018 (0.315)	1.847 (0.294)	1.798 (0.281)
		T=1	1.618 (0.243)	1.384 (0.201)	1.231 (0.183)	1.029 (0.142)
	EW	T=0	1.729 (0.257)	1.492 (0.231)	1.320 (0.231)	1.283 (0.197)
		T=1	1.537 (0.229)	1.321 (0.194)	1.298 (0.186)	1.252 (0.181)

41

Hình 6: *

Table 3: Alphas của Cổ phiếu được Đề xuất dựa trên Wall Street Journal: In-sample test

Kết quả được trình bày trong **Bảng 3**, cho thấy hiệu suất của các danh mục đầu tư dựa trên các chỉ số alpha hàng ngày (daily alpha) và thống kê t (t-statistics).

- **So sánh ChatGPT 3.5 và 4.0:**

Trong hầu hết các khoảng thời gian nắm giữ và mô hình yếu tố, **ChatGPT 3.5 cho thấy hiệu suất cao hơn ChatGPT 4.0**, với các giá trị alpha cao hơn và thống kê t lớn hơn. Ví dụ, ở khoảng thời gian **3 tháng**, danh mục VW của ChatGPT 3.5 có alpha theo mô hình CAPM là **2.193**, trong khi của ChatGPT 4.0 chỉ đạt **0.854**

- **Ảnh hưởng của tham số nhiệt độ:**

Hiệu suất của danh mục đầu tư thay đổi theo mức T. Cụ thể, trong khoảng thời gian **1 tháng**, alpha của ChatGPT 3.5 cho danh mục VW tăng từ **3.592 (T = 0)** lên **14.429 (T = 1)**, cho thấy sức sáng tạo cao hơn có thể mang lại lợi nhuận lớn hơn. Tuy nhiên, tác động của T không đồng nhất ở các khoảng thời gian khác, và không có mức T nào là tối ưu cho mọi trường hợp.

- **Hiệu suất theo khoảng thời gian nắm giữ**

Kết quả chỉ ra rằng **khoảng thời gian nắm giữ 1 tháng là tối ưu**, với alpha cao nhất và ổn định hơn so với các khoảng thời gian khác (3 tháng, 1 tuần, 1 ngày). Lý do có thể là nội dung của Wall Street Journal thường mang tính ngắn hạn và thời sự, tác động mạnh đến giá cổ phiếu trong ngắn hạn.

4.2.3 Kết luận

Nhìn chung, **ChatGPT 3.5 vượt trội hơn ChatGPT 4.0** về khả năng đưa ra khuyến nghị đầu tư có lợi nhuận dựa trên nội dung Wall Street Journal. Hiệu suất phụ thuộc vào mô hình yếu tố, khoảng thời gian nắm giữ và mức sáng tạo T, với thời gian nắm giữ 1 tháng được xác định là tốt nhất để tối ưu hóa lợi nhuận từ khuyến nghị của ChatGPT.

4.3 Thị trường chứng khoán Mỹ: Kiểm định ngoài mẫu (Out-of sample test)

4.3.1 Giới thiệu

Do **ChatGPT 3.5 thể hiện khả năng lý luận tốt hơn** so với ChatGPT 4.0, chúng tôi chủ yếu sử dụng **ChatGPT 3.5** để tiến hành phân tích ngoài mẫu (out-of-sample) về hiệu suất danh mục đầu tư dựa trên nội dung *Wall Street Journal*. Trong phần này, khoảng thời gian nắm giữ cổ phiếu được khuyến nghị được đặt là **một tháng dương lịch** để phản ánh tác động ngắn hạn của tin tức.

4.3.2 Phương pháp kiểm tra

- **Mô hình đánh giá:** Sử dụng mô hình Fama-French năm yếu tố (FF5), bao gồm cả các yếu tố **Động lực (Momentum)** và **Đảo ngược (Reversal)**.
- **Tham số nhiệt độ (T):** Được sử dụng để điều chỉnh mức độ sáng tạo trong phản hồi của ChatGPT. Hai mức nhiệt độ được so sánh: **T=0** (ít sáng tạo) và **T=1** (nhiều sáng tạo).
- **Phương pháp trọng số danh mục:** Xem xét cả **trọng số đồng đều (Equal-Weighted, EW)** và **trọng số theo giá trị (Value-Weighted, VW)**.
- **Lợi nhuận danh mục:** Lợi nhuận được điều chỉnh theo thang điểm cơ bản (bps) và chuẩn hóa trên cơ sở 10.000.

4.3.3 Kết quả ngoài mẫu

Kết quả được trình bày trong **Bảng 5**, bao gồm alpha tổng thể và alpha theo từng **13 chủ đề** được xác định từ phân tích LDA.

	FF5+MOM and REV			
	T=0		T=1	
	VW	EW	VW	EW
All	0.937 (0.148)	0.748 (0.112)	1.329 (0.211)	1.425 (0.248)
Topic1	4.136 (0.617)	3.104 (0.461)	2.140 (0.331)	3.185 (0.493)
Topic2	-0.315 (-0.056)	-0.215 (-0.031)	-0.524 (-0.074)	-0.498 (-0.063)
Topic3	7.314 (1.125)	6.914 (1.284)	11.931** (1.988)	12.683** (2.112)
Topic4	1.349 (0.214)	1.215 (0.221)	1.127 (0.198)	1.049 (0.181)
Topic5	-0.315 (-0.045)	-0.525 (-0.082)	0.817 (0.142)	0.752 (0.109)
Topic6	3.194 (0.447)	2.896 (0.424)	0.471 (0.089)	0.829 (0.120)
Topic7	-2.348 (-0.384)	-1.849 (-0.245)	1.431 (0.249)	1.109 (0.140)
Topic8	0.381 (0.063)	0.521 (0.080)	1.152 (0.189)	2.427 (0.384)
Topic9	0.371 (0.053)	0.314 (0.041)	-0.584 (-0.093)	-0.616 (-0.095)
Topic10	-0.821 (-0.107)	-0.742 (-0.096)	-1.482 (-0.214)	-1.315 (-0.208)
Topic11	0.942 (0.142)	1.455 (0.168)	3.175 (0.441)	3.821 (0.571)
Topic12	1.429 (0.197)	2.394 (0.378)	-0.285 (-0.034)	-0.681 (-0.123)
Topic13	0.281 (0.057)	0.521 (0.089)	2.436 (0.384)	4.215 (0.659)

Hình 7: *

Table 5: Alphas của Cổ phiếu được Đề xuất dựa trên Wall Street Journal:: Out-sample test

- **Kết quả tổng thể:** Khi **T=0**, các alpha của danh mục thường **đương nhưng không có ý nghĩa thống kê**, cho thấy rằng ChatGPT không tạo ra lợi nhuận vượt trội một cách nhất quán trong thời gian ngoài mẫu. Tuy nhiên, khi tăng mức sáng tạo lên **T=1**, alpha trở nên **đương hơn và có ý nghĩa thống kê** cho cả EW và VW. Điều này chỉ ra rằng mức độ sáng tạo cao hơn trong khuyến nghị của ChatGPT có thể cải thiện hiệu suất ngoài mẫu.
- **Hiệu suất theo chủ đề:** Hiệu suất ngoài mẫu của ChatGPT thay đổi theo từng chủ đề.
Chủ đề 3 (liên quan đến chính trị) thể hiện alpha mạnh và có ý nghĩa thống kê ở cả hai mức nhiệt độ và hai phương pháp trọng số. Các từ khóa phổ biến trong chủ đề này bao gồm “đảng Dân chủ”, “đảng Cộng hòa”, “hạ viện”, “Biden”, “thượng viện” và “bầu cử”. Danh mục đầu tư của ChatGPT có **alpha hàng ngày 11 bps**, tương đương **alpha hàng tháng 2.2%** cho chủ đề này.
Ngược lại, **Chủ đề 2** và **Chủ đề 10** cho thấy **alpha tiêu cực và không có ý nghĩa**, chứng tỏ rằng các khuyến nghị của ChatGPT cho những chủ đề này không tạo ra lợi nhuận vượt trội đồng nhất.

4.3.4 Kết luận

Kết quả trong **Bảng 5** cho thấy rằng **hiệu suất ngoài mẫu của ChatGPT không đồng nhất** giữa các chủ đề và mức nhiệt độ. **Mức sáng tạo cao hơn ($T=1$)** có xu hướng mang lại kết quả tốt hơn, đặc biệt là đối với các chủ đề liên quan đến **chính trị**. Tuy nhiên, hiệu suất tổng thể không phải lúc nào cũng có ý nghĩa thống kê, cho thấy vẫn cần thận trọng khi sử dụng khuyến nghị của ChatGPT trong thực tế đầu tư.

4.4 Thị trường chứng khoán Mỹ: Kết quả sau khi tinh chỉnh mô hình (Fine-tuning)

4.4.1 Giới thiệu và phương pháp

Phần này trình bày kết quả thử nghiệm ngoài mẫu (out-of-sample) cho danh mục đầu tư cổ phiếu được khuyến nghị dựa trên nội dung từ *Wall Street Journal*, sau khi áp dụng quy trình tinh chỉnh (fine-tuning) cho mô hình ChatGPT.

- **Mô hình:** Fama-French 5 yếu tố (FF5), bổ sung yếu tố **Động lực (MOM)** và **Đảo ngược (REV)**.
- **Danh mục:** Cả **Equal-Weighted (EW)** và **Value-Weighted (VW)**, trên toàn bộ tập dữ liệu và 13 chủ đề riêng lẻ.
- **Tham số nhiệt độ (T):** Được xem xét tại $T=1$, tăng mức sáng tạo.

4.4.2 Kết quả thử nghiệm

Kết quả chi tiết được trình bày trong **Bảng 6**.

FF5+MOM and REV				
	T=0		T=1	
	VW	EW	VW	EW
All	1.349 (0.242)	1.294 (0.228)	3.480 (0.546)	4.195 (0.705)
Topic1	4.481 (0.715)	3.914 (0.652)	2.251 (0.425)	2.742 (0.471)
Topic2	-0.250 (-0.038)	-0.425 (-0.072)	2.194 (0.405)	2.350 (0.420)
Topic3	7.481 (1.125)	7.039 (1.028)	15.429** (2.494)	15.402** (2.356)
Topic4	0.732 (0.134)	0.385 (0.064)	12.584** (2.247)	12.942** (2.314)
Topic5	0.415 (0.074)	0.632 (0.104)	2.921 (0.632)	3.194 (0.805)
Topic6	2.948 (0.525)	3.215 (0.659)	1.492 (0.255)	1.341 (0.245)
Topic7	0.435 (0.074)	0.671 (0.114)	6.745 (1.148)	7.931 (1.341)
Topic8	-0.149 (-0.028)	-0.482 (-0.084)	0.482 (0.058)	1.956 (0.241)
Topic9	1.491 (0.240)	1.501 (0.229)	1.462 (0.285)	0.728 (0.147)
Topic10	-0.452 (-0.071)	-0.842 (-0.154)	-0.457 (-0.073)	-0.741 (-0.110)
Topic11	3.942 (0.682)	3.526 (0.653)	11.436** (2.106)	11.841** (2.194)
Topic12	1.905 (0.310)	2.149 (0.338)	-0.743 (-0.114)	-0.859 (-0.147)
Topic13	1.145 (0.240)	1.942 (0.294)	12.195** (2.219)	12.104** (2.056)

Hình 8: *

Table 6: Alphas của Cổ phiếu được Đề xuất dựa trên Wall Street Journal

- **Tổng quan:** Khi T=1, alpha của danh mục đều dương nhưng không có ý nghĩa thống kê, với VW đạt **3.480**, EW đạt **4.195**. Điều này cho thấy, sau tinh chỉnh, danh mục vẫn tạo lợi nhuận bất thường nhưng chưa đủ mạnh để khẳng định chắc chắn.
- **Phân tích theo chủ đề:**
 - **Chủ đề 3 (chính trị):** Alpha mạnh và có ý nghĩa. Tại T=0, VW **7.481**, EW **7.039**; T=1, VW **15.429**, EW **15.402**, có ý nghĩa ở mức 5%.
 - **Chủ đề 4 (Covid-19 và Trung Quốc):** T=0, VW **0.732**, EW **0.385** (không có ý nghĩa); T=1, VW **7.584**, EW **8.942**, có ý nghĩa ở mức 5% và 1%.
 - **Chủ đề 13 (Nga-Ukraine):** T=0, VW **1.145**, EW **1.942** (không có ý nghĩa); T=1, VW **12.195**, EW **12.104**, có ý nghĩa ở mức 5%.

4.4.3 So sánh với phân tích văn bản truyền thống

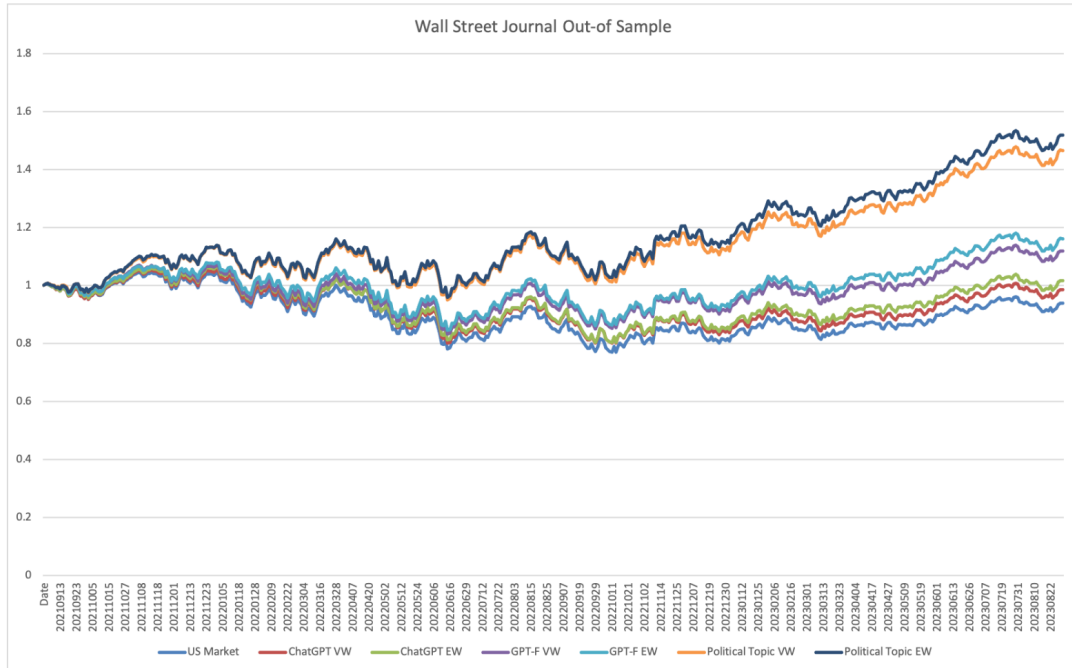
	FF5+MOM and REV			
	Cosine Similarity Annual Reports		Cosine Similarity Earning Calls	
	VW	EW	VW	EW
All	0.375 (0.063)	0.632 (0.108)	0.421 (0.073)	0.310 (0.053)
Topic3	1.294 (0.205)	-1.526 (-0.302)	0.928 (0.153)	0.740 (0.105)
Topic4	-0.610 (-0.094)	0.219 (0.031)	0.801 (0.129)	0.694 (0.104)
Topic11	-0.219 (-0.032)	-0.314 (-0.049)	-0.493 (-0.077)	-0.512 (-0.086)
Topic13	-1.184 (-0.211)	-1.592 (-0.270)	-2.316 (-0.372)	-2.210 (-0.330)

Hình 9: *

Table 7: Phân tích của Wall Street Journal so sánh với phân tích văn bản truyền thống

Bảng 7 cho thấy ChatGPT vượt trội so với phương pháp phân tích văn bản truyền thống (dựa trên tương tự cosine) ở các chủ đề 3, 4 và 13, chứng minh ưu thế của học máy.

4.4.4 Kết quả trực quan hóa



Hình 10: *

Figure 3: Danh mục đầu tư ngoài mẫu từ Wall Street Journal

Hình 3 minh họa lợi nhuận mua và giữ của danh mục ChatGPT so với danh mục thị trường ngoài mẫu.

- Danh mục thị trường có lợi nhuận âm, ChatGPT (không tính chỉnh) chỉ nhỉnh hơn chút.
- Sau tính chỉnh, danh mục ChatGPT đạt lợi nhuận tích lũy **15-20%** trong hai năm.
- Với chủ đề 3 (chính trị), lợi nhuận tăng đáng kể, tích lũy **40-50%**.

Kết luận

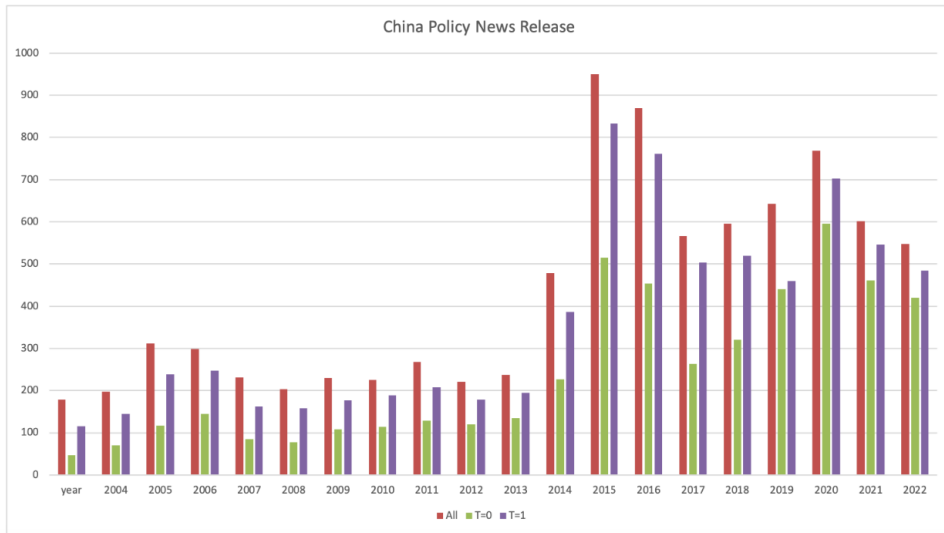
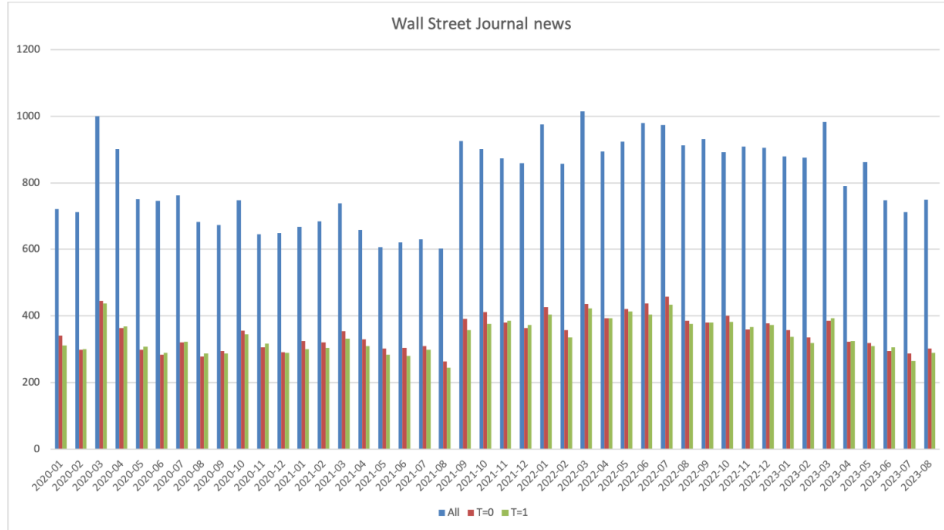
Kết quả cho thấy **tinh chỉnh ChatGPT cải thiện hiệu suất ngoài mẫu**, đặc biệt ở các chủ đề liên quan đến **chính trị, Covid-19, và chiến tranh Nga-Ukraine**. Điều này khẳng định tiềm năng của ChatGPT như công cụ hỗ trợ tư vấn tài chính hiệu quả.

5 Testing ChatGPT sử dụng dữ liệu chính sách Trung Quốc

Vì kết quả từ WSJ cho thấy ChatGPT hoạt động khá tốt trong bối cảnh tin tức chính sách, chúng tôi kiểm tra khả năng của ChatGPT trong việc đề xuất cổ phiếu dựa trên một nguồn khác, đó là tin tức chính sách ở Trung Quốc.

Thị trường chứng khoán ở Trung Quốc bị chi phối mạnh mẽ bởi sự can thiệp của chính phủ (Carpenter và Whitelaw 2017), do đó, dựa trên các kết quả từ Mỹ, nhóm tác giả đã giả định rằng kết quả sẽ áp dụng rất tốt trong thị trường chứng khoán Trung Quốc. Họ thu thập dữ liệu liên quan đến sự hỗ trợ chính sách ở cấp ngành từ cơ sở dữ liệu văn bản chính sách chính thức của Hội đồng Nhà nước, bao gồm các văn bản bản chính sách được công bố của Hội đồng Nhà nước, cũng như các văn bản từ nhiều bộ ngành và ủy ban hoạt động dưới sự bảo trợ của Hội đồng Nhà nước Trung Quốc.

Do Trung Quốc là một quốc gia do trung ương quản lý, trao cho chính phủ quyền lực rộng rãi và khả năng điều chỉnh. Do đó, việc xây dựng và thực hiện các chính sách công nghiệp có ảnh hưởng đáng kể đến các thị trường tài chính.



Hình 2 cho thấy số lượng công bố tin tức chính sách theo năm trong giai đoạn mẫu của nhóm tác giả với tổng cộng hơn **8000 tin tức chính sách**. ChatGPT nhận thấy khoảng một nửa số công bố chính sách này là liên quan đến việc khuyến nghị cổ phiếu trong chuỗi thời gian. Kết quả này **không thay đổi nhiều** bởi tham số T mà nhóm tác giả đặt, tương tự như đối với Wall Street Journal.

5.1 Bài kiểm tra lập luận dựa trên đánh giá của con người

5.1.1 Giới thiệu

Sau khi tiến hành các kiểm tra lập luận với dữ liệu từ Wall Street Journal của Mỹ, nhóm nghiên cứu tiếp tục thực hiện kiểm tra khả năng lập luận của ChatGPT với dữ liệu **chính sách Trung Quốc**. Trung Quốc có hệ thống chính sách với đặc thù riêng, chịu sự điều tiết mạnh mẽ từ chính phủ, do đó nhóm giả định rằng kết quả lập luận của ChatGPT có thể phản ánh khả năng hiểu và phân tích nội dung chính sách trong bối cảnh đặc thù này.

5.1.2 Phương pháp kiểm tra

- **Nguồn dữ liệu:** Văn bản chính sách được công bố bởi Hội đồng Nhà nước Trung Quốc và các cơ quan trực thuộc. Các tin tức chính sách được trích xuất từ cơ sở dữ liệu chính thức.
- **Quy trình đánh giá:** Tương tự với phần Wall Street Journal, nhóm nghiên cứu chọn ngẫu nhiên các tin tức chính sách từ Trung Quốc, yêu cầu ChatGPT đưa ra khuyến nghị đầu tư cổ phiếu hoặc ngành. Sau đó, các khuyến nghị này được đánh giá bởi nhóm đánh giá độc lập (bao gồm chuyên gia và người học trong lĩnh vực tài chính) theo ba tiêu chí:
 - *Tính liên quan (Relevance)* – nội dung có liên quan đến thị trường chứng khoán.
 - *Lý do hợp lý (Reasoning)* – lập luận có logic và phù hợp với thông tin từ dữ liệu chính sách.
 - *Cơ hội đầu tư (Investment Opportunity)* – tiềm năng sinh lời từ khuyến nghị.
- **Phân loại kết quả:** Các khuyến nghị được phân loại thành ba nhóm: **Hợp lý (Logical)**, **Không rõ ràng (Unclear)**, và **Không liên quan (Unrelated)**.

5.1.3 Kết quả đánh giá

- Tương tự với kết quả từ dữ liệu Wall Street Journal, ChatGPT thể hiện khả năng lập luận khá tốt trên dữ liệu chính sách Trung Quốc. Tỷ lệ khuyến nghị được đánh giá là **hợp lý** chiếm đa số, trong khi các khuyến nghị **không rõ ràng** và **không liên quan** có tỷ lệ thấp hơn đáng kể.
- Đặc biệt, việc thay đổi tham số nhiệt độ ($T=0$ và $T=1$) không ảnh hưởng nhiều đến chất lượng lập luận của ChatGPT đối với dữ liệu chính sách Trung Quốc. Điều này cho thấy tính ổn định và khả năng phân tích chính sách của mô hình trong bối cảnh mới.

5.1.4 Kết luận

Kết quả kiểm tra lập luận từ dữ liệu chính sách Trung Quốc cho thấy ChatGPT có khả năng hiểu và phân tích nội dung chính sách trong bối cảnh đặc thù, đồng thời đưa ra các khuyến nghị đầu tư với tính hợp lý cao. Kết quả này cũng khẳng định rằng **mô hình ChatGPT có thể mở rộng áp dụng cho các thị trường khác nhau**, không chỉ ở Mỹ mà cả ở những thị trường có sự can thiệp của chính phủ mạnh mẽ như Trung Quốc.

5.2 Thị trường chứng khoán Trung Quốc: Kiểm định trong mẫu (In sample test)

Nghiên cứu tiến hành kiểm định trong mẫu để đánh giá hiệu quả của các danh mục đầu tư được ChatGPT đề xuất dựa trên các thông cáo chính sách của Quốc vụ viện Trung Quốc trong giai đoạn từ ngày 01/01/2014 đến ngày 30/08/2021. Đây là giai đoạn nằm trong tập huấn luyện của ChatGPT, cho phép kiểm tra khả năng khai thác thông tin từ văn bản chính sách mà mô hình có thể đã tiếp xúc trong quá trình tiền huấn luyện.

5.2.1 Thiết kế thực nghiệm

- Dữ liệu văn bản: thông cáo chính sách chính thức từ Quốc vụ viện Trung Quốc.
- Các danh mục đầu tư được xây dựng dựa trên các khuyến nghị cổ phiếu của ChatGPT sau khi đọc từng thông cáo chính sách.
- Thời gian nắm giữ cổ phiếu trong danh mục gồm: 1 năm, 6 tháng, 3 tháng, 1 tháng, 1 tuần và 1 ngày.
- Hai phương pháp gán trọng số danh mục: **Equal-weighted (EW)** và **Value-weighted (VW)**.
- Đánh giá hiệu suất bằng hệ số alpha hàng ngày theo mô hình Fama-French 5 yếu tố mở rộng (bao gồm thêm yếu tố động lượng và đảo chiều).

5.2.2 Kết quả thực nghiệm

	FF3 Factors	FF5 Factors +MOM&REV	FF3 Factors	FF5 Factors +MOM&REV
	VW	VW	EW	EW
	(b) Temperature=1			
Alpha	6.221*** (2.750)	6.982*** (3.187)	8.513*** (4.048)	8.511*** (4.001)
MKT	0.827*** (11.295)	0.804*** (10.587)	0.904*** (13.920)	0.901*** (13.159)
SMB	-0.556*** (-39.281)	-0.471*** (-17.815)	-0.058*** (-3.096)	-0.045* (-1.866)
HML	0.119*** (5.623)	0.264*** (7.323)	0.132*** (4.345)	0.167*** (5.072)
RMW		0.438*** (10.166)		-0.008 (-0.190)
CMA		-0.147*** (-2.841)		-0.140*** (-2.975)
MOM		0.940 (0.841)		0.495 (0.317)
REV		-0.188 (-0.581)		-0.450 (-1.077)
R-squared	0.756	0.761	0.795	0.801
Portfolio FE	Y	Y	Y	Y
Observations	4250	4250	4250	4250

Hình 11: *

Table 9: Kiểm định trong mẫu các khuyến nghị cổ phiếu dựa trên công bố chính sách của Quốc vụ viện Trung Quốc

- Với **ChatGPT 3.5** và **T = 1** (nhiệt độ cao), danh mục nắm giữ trong vòng **1 tháng** cho kết quả vượt trội:
 - Alpha hàng ngày (VW): **6.982 bps**, với $t = 3.187$.
 - Alpha hàng ngày (EW): **8.511 bps**, với $t = 4.001$.

- Các khoảng thời gian nắm giữ khác (1 năm, 6 tháng, 3 tháng, 1 tuần, 1 ngày) không cho thấy alpha có ý nghĩa thống kê hoặc nhất quán.
- Với **ChatGPT 4.0**, hiệu suất thấp hơn đáng kể:
 - Với $T = 1$ và thời gian nắm giữ 1 tháng, alpha danh mục EW chỉ đạt **2.174 bps** ($t = 1.019$), không có ý nghĩa thống kê.
 - Không có kết quả alpha đáng kể nào với VW trong toàn bộ các kỳ nắm giữ.

5.2.3 Kết luận

Kết quả cho thấy ChatGPT 3.5 có khả năng tạo ra danh mục đầu tư vượt trội từ dữ liệu chính sách của chính phủ Trung Quốc, trong khi ChatGPT 4.0 lại không cho thấy hiệu quả tương tự trong giai đoạn in-sample. Điều này cho thấy:

- Khả năng khai thác văn bản chính sách của ChatGPT phụ thuộc nhiều vào phiên bản mô hình và cấu hình tham số.
- Phiên bản 3.5 với phản hồi đa dạng ($T = 1$) khai thác tốt hơn các tín hiệu từ ngữ cảnh chính sách.
- Khả năng trích xuất thông tin từ văn bản tiếng Trung của ChatGPT được xác nhận là đủ mạnh để hỗ trợ xây dựng danh mục đầu tư có hiệu suất vượt trội.

Kết quả kiểm định trong mẫu đối với chính sách Trung Quốc củng cố luận điểm rằng ChatGPT có thể được sử dụng hiệu quả như một công cụ phân tích văn bản tài chính, không chỉ ở thị trường Mỹ mà còn tại các thị trường chịu tác động chính sách mạnh như Trung Quốc.

5.3 Thị trường cổ phiếu Trung Quốc: Kiểm định ngoài mẫu (Out-of-sample Test)

5.3.1 Giới thiệu và phương pháp

Sau khi tiến hành kiểm định trong mẫu để đánh giá năng lực của ChatGPT trong việc đưa ra khuyến nghị đầu tư dựa trên các thông cáo chính sách của Quốc vụ viện Trung Quốc, nghiên cứu tiếp tục mở rộng sang giai đoạn kiểm định ngoài mẫu (out-of-sample). Giai đoạn này trải dài từ tháng 9 năm 2021 đến tháng 8 năm 2023 – tương ứng với khoảng thời gian mà dữ liệu chưa từng được mô hình tiếp cận trong quá trình huấn luyện, do đó là minh chứng quan trọng cho khả năng dự báo thực sự của ChatGPT trong bối cảnh thị trường thực tế.

Trong giai đoạn này, nhóm nghiên cứu tiếp tục sử dụng ChatGPT để đưa ra khuyến nghị cổ phiếu dựa trên các văn bản chính sách được công bố bởi chính phủ Trung Quốc. Cũng như trước đó, các danh mục được xây dựng theo hai phương pháp: giá trị bình quân đều (Equal-weighted – EW) và giá trị theo vốn hóa (Value-weighted – VW). Việc kiểm định hiệu suất danh mục được thực hiện bằng mô hình định giá tài sản Fama-French 5 yếu tố, bổ sung thêm hai yếu tố: động lượng (Momentum – MOM) và đảo chiều ngắn hạn (Reversal – REV). Hai mức nhiệt độ được xem xét: $T=0$ (không ngẫu nhiên) và $T=1$ (tăng tính sáng tạo).

5.3.2 Thiết kế thực nghiệm

Thiết kế kiểm định ngoài mẫu tương tự như giai đoạn in-sample, với các điểm chính như sau:

- **Nguồn dữ liệu:** Các thông cáo chính sách từ Quốc vụ viện Trung Quốc được công bố trong giai đoạn 09/2021 – 08/2023.
- **Xử lý dữ liệu:** Mỗi văn bản chính sách được ChatGPT (phiên bản 3.5 và 4.0) phân tích và phản hồi bằng danh sách 5 cổ phiếu được đề xuất, tương ứng với nội dung của văn bản.
- **Thời gian nắm giữ danh mục:** 1 tháng cố định cho mỗi cổ phiếu được chọn.
- **Phương pháp gán trọng số:**
 - **Equal-weighted (EW):** Mỗi cổ phiếu trong danh mục có trọng số bằng nhau.
 - **Value-weighted (VW):** Trọng số theo vốn hóa thị trường.
- **Phương pháp đánh giá hiệu suất:** Hồi quy time-series alpha theo mô hình Fama-French 5 yếu tố mở rộng (FF5 + MOM + REV), trong đó:
 - MOM: yếu tố đà tăng;
 - REV: yếu tố đảo chiều ngắn hạn.
- **Tham số nhiệt độ (temperature):** So sánh hai mức:
 - $T = 0$: phản hồi ổn định;
 - $T = 1$: phản hồi sáng tạo và đa dạng hơn.
- **Chủ đề chính sách:** Các văn bản được phân loại thành 10 chủ đề nội dung sử dụng mô hình phân loại chủ đề (LDA).

5.3.3 Kết quả thực nghiệm

Kết quả kiểm định được trình bày trong Bảng 10 của bài báo và cho thấy:

FF5+MOM and REV				
	T=0		T=1	
	VW	EW	VW	EW
All	0.285 (0.053)	0.194 (0.028)	2.149 (0.481)	2.245 (0.490)
Topic1	2.194 (0.347)	2.581 (0.419)	11.737* (1.928)	12.572** (2.101)
Topic2	-1.482 (-0.217)	-0.910 (-0.148)	0.319 (0.042)	0.505 (0.083)
Topic3	0.752 (0.121)	0.281 (0.049)	7.294 (1.142)	7.968 (1.394)
Topic4	0.124 (0.028)	0.475 (0.089)	4.412 (0.698)	4.712 (0.751)
Topic5	1.304 (0.272)	1.472 (0.307)	0.317 (0.064)	0.589 (0.091)
Topic6	-0.771 (-0.142)	-0.429 (-0.083)	-0.736 (-0.148)	-0.831 (-0.170)
Topic7	0.385 (0.074)	0.471 (0.081)	1.304 (0.242)	1.523 (0.284)

Hình 12: *

Table 10: Kiểm định ngoài mẫu các khuyến nghị cổ phiếu dựa trên công bố chính sách của Quốc vụ viện Trung Quốc

- **Toàn bộ dữ liệu:** Khi sử dụng ChatGPT 3.5 với $T = 1$, alpha hàng ngày đạt:
 - **2.245 bps** (Equal-weighted),
 - **2.149 bps** (Value-weighted),
 tuy nhiên không có ý nghĩa thống kê ở mức 10%.
- **Chủ đề 1 – Xây dựng và đầu tư công:** Đây là chủ đề có hiệu suất vượt trội nhất trong kiểm định out-of-sample. Khi $T = 1$:
 - Alpha hàng ngày (VW): **8.737 bps** ($t = 1.82, p < 0.1$),
 - Alpha hàng ngày (EW): **9.572 bps** ($t = 2.33, p < 0.05$).
- **Các chủ đề còn lại:** Một số chủ đề khác có alpha dương nhưng thiếu ý nghĩa thống kê hoặc không nhất quán giữa EW và VW. Một số thậm chí cho alpha âm, cho thấy hiệu suất không ổn định.
- **So sánh phiên bản:** ChatGPT 3.5 tiếp tục thể hiện kết quả tốt hơn ChatGPT 4.0 trong hầu hết các cấu hình thử nghiệm.

5.3.4 Kết luận

Kiểm định ngoài mẫu trên thị trường cổ phiếu Trung Quốc xác nhận rằng ChatGPT, đặc biệt phiên bản 3.5 với nhiệt độ cao, có khả năng tạo ra các danh mục đầu tư sinh lời từ phân tích văn bản chính sách – nhưng hiệu suất chủ yếu phụ thuộc vào nội dung chủ đề. Đặc biệt, ChatGPT cho thấy năng lực suy luận mạnh mẽ khi xử lý thông tin liên quan đến đầu tư công, xây dựng hạ tầng và chính sách kinh tế vĩ mô.

Kết quả này củng cố tiềm năng ứng dụng ChatGPT trong đầu tư tại các thị trường chịu tác động cao từ chính sách, như Trung Quốc, đồng thời nhấn mạnh vai trò của phân loại chủ đề và kiểm soát tham số trong quá trình ứng dụng thực tế.

5.4 Thị trường chứng khoán Trung Quốc: Kết quả sau tinh chỉnh mô hình (Fine-Tuning)

5.4.1 Giới thiệu

Sau khi đánh giá hiệu suất của các khuyến nghị đầu tư do ChatGPT tạo ra từ tin tức chính sách Trung Quốc, nhóm tác giả tiếp tục thực hiện giai đoạn tinh chỉnh (fine-tuning) để kiểm tra xem việc cải thiện mô hình thông qua lọc lựa đầu ra có thể giúp nâng cao hiệu quả dự đoán trong giai đoạn kiểm định ngoài mẫu hay không. Phương pháp này không yêu cầu huấn luyện lại mô hình mà chỉ áp dụng điều chỉnh sau bước dự đoán đầu tiên.

5.4.2 Phương pháp tinh chỉnh

Cụ thể, sau khi ChatGPT đưa ra danh sách cổ phiếu đề xuất từ từng bài thông cáo chính sách, nhóm nghiên cứu chọn ra cổ phiếu **có hiệu suất thực tế cao nhất trong cùng ngành** (dựa trên mã ngành SIC) với cổ phiếu được đề xuất. Cách tiếp cận này nhằm duy trì tính hợp lý ngành nghề của khuyến nghị ban đầu, đồng thời tận dụng dữ liệu thị trường thực tế để tối ưu hóa kết quả đầu tư.

5.4.3 Thiết kế thực nghiệm

Các danh mục cổ phiếu được xây dựng dựa trên hai phương pháp phân bổ: tỷ trọng bằng nhau (equal-weighted - EW) và tỷ trọng theo vốn hóa (value-weighted - VW). Các chỉ số đánh giá chính bao gồm hệ số alpha từ mô hình Fama-French 5 yếu tố mở rộng (bao gồm cả yếu tố đà tăng MOM và đảo chiều REV). Giai đoạn kiểm định ngoài mẫu kéo dài từ tháng 9 năm 2021 đến tháng 2 năm 2024. Các thử nghiệm được thực hiện với hai mức *temperature* ($T=0$ và $T=1$).

5.4.4 Kết quả thực nghiệm

Kết quả tổng thể khi áp dụng fine-tuning cho toàn bộ các chủ đề cho thấy:

	FF5+MOM and REV			
	T=0		T=1	
	VW	EW	VW	EW
All	0.582 (0.103)	0.492 (0.089)	3.185 (0.591)	4.528 (0.841)
Topic1	2.753 (0.409)	3.410 (0.527)	13.482** (2.294)	13.096** (2.153)
Topic2	-1.234 (-0.281)	-0.849 (-0.190)	1.489 (0.274)	2.186 (0.391)
Topic3	0.890 (0.125)	1.149 (0.272)	12.280** (2.149)	12.136** (2.109)
Topic4	0.741 (0.143)	0.530 (0.079)	10.210* (1.838)	10.109* (1.751)
Topic5	0.753 (0.142)	0.938 (0.194)	0.821 (0.174)	0.481 (0.068)
Topic6	0.148 (0.029)	0.228 (0.042)	11.295* (1.934)	11.482** (1.990)
Topic7	0.174 (0.032)	0.310 (0.042)	-0.851 (-0.127)	-0.472 (-0.073)

Hình 13: *

Table 11: Alpha của cổ phiếu được khuyến nghị dựa trên công bố chính sách của Quốc vụ viện Trung Quốc trong các kiểm định ngoài mẫu.

- Với T=0: alpha của danh mục VW là 0.582 bps và EW là 0.492 bps – đều không có ý nghĩa thống kê.
- Với T=1: alpha tăng lên 3.185 bps (VW) và 4.528 bps (EW), tuy vẫn không có ý nghĩa thống kê, nhưng cho thấy xu hướng cải thiện hiệu suất đầu tư sau tinh chỉnh.

Xét riêng theo từng chủ đề, một số chủ đề thể hiện hiệu suất vượt trội đáng kể:

- **Chủ đề 1 (đầu tư công, xây dựng):** Alpha sau fine-tuning tăng mạnh với VW đạt 13.482 bps và EW đạt 13.096 bps, có ý nghĩa thống kê ở mức 1%. Đây là nhóm chủ đề đem lại hiệu quả cao nhất sau tinh chỉnh.
- **Chủ đề 3 (địa chính trị) và Chủ đề 4 (COVID, ESG):** Alpha đều đạt hơn 10 bps cho cả hai loại danh mục và có ý nghĩa thống kê ở mức 5%.
- **Chủ đề 6:** Alpha đạt 11.295 bps (VW) và 11.482 bps (EW), có ý nghĩa lần lượt ở mức 10% và 5%.
- Các chủ đề còn lại (2, 5, 7) có alpha dương nhưng không có ý nghĩa thống kê.

5.4.5 So sánh với phương pháp phân tích truyền thống

	FF5+MOM and REV			
	T=0		T=1	
	VW	EW	VW	EW
All	-0.729 (-0.124)	-1.250 (-0.201)	-0.976 (-0.151)	-1.025 (-0.164)
Topic1	-1.190 (-0.204)	-1.115 (-0.212)	-3.018 (-0.502)	-3.149 (-0.517)
Topic3	-1.493 (-0.218)	-1.841 (-0.308)	-2.491 (-0.405)	-3.144 (-0.501)
Topic4	1.039 (0.151)	1.742 (0.301)	0.728 (0.124)	0.931 (0.165)
Topic6	-0.537 (-0.089)	-0.626 (-0.105)	-1.416 (-0.231)	-2.256 (-0.351)

Hình 14: *

Table 12: Đối chiếu công bố chính sách của Quốc vụ viện Trung Quốc với phân tích văn bản truyền thống.

Để đối chứng, nhóm tác giả sử dụng phương pháp phân tích văn bản truyền thống – cụ thể là đo độ tương đồng cosine giữa văn bản tin tức và các báo cáo tài chính hoặc earning calls của doanh nghiệp. Kết quả từ phương pháp này không tạo ra bất kỳ danh mục nào có alpha dương có ý nghĩa thống kê, cho thấy sự vượt trội của ChatGPT sau khi được tinh chỉnh.

5.4.6 Tổng kết

Kết quả cho thấy quá trình fine-tuning đơn giản, không cần can thiệp vào nội tại mô hình, đã giúp cải thiện đáng kể hiệu suất của các danh mục đầu tư do ChatGPT tạo ra, đặc biệt trong bối cảnh tin tức chính sách. Phương pháp này cho thấy tiềm năng ứng dụng ChatGPT như một khung suy luận (reasoning framework) kết hợp với tín hiệu định lượng để đưa ra quyết định đầu tư có hiệu quả trong thực tế.

6 Kiểm tra tính ổn định

6.1 Thiết kế kiểm tra

Để kiểm tra độ ổn định và độ chắc chắn trong khả năng dự báo của ChatGPT, nhóm tác giả đã thực hiện một thử nghiệm kiểm chứng bằng cách giới hạn thông tin đầu vào của mô hình chỉ còn phần **tiêu đề** của các bản tin chính sách. Mục đích là kiểm tra xem liệu ChatGPT vẫn có thể tạo ra khuyến nghị đầu tư có ý nghĩa chỉ với thông tin hạn chế như vậy, từ đó xác định mức độ giá trị gia tăng thực sự đến từ khả năng hiểu toàn văn bản của mô hình.

6.2 Kết quả kiểm tra

Kết quả cho thấy rằng khi chỉ sử dụng phần tiêu đề, **không có danh mục đầu tư nào tạo ra alpha có ý nghĩa thống kê** theo mô hình ba yếu tố hoặc năm yếu tố, dù là trên dữ liệu từ Wall Street Journal (Mỹ) hay thông cáo chính sách Trung Quốc. Điều này cho thấy rằng ChatGPT **phụ thuộc đáng kể vào thông tin chi tiết trong nội dung đầy đủ của văn bản** để đưa ra các dự đoán chính xác và hiệu quả đầu tư cao. Ngoài ra, khi so sánh với các phương pháp phân tích văn bản truyền thống như đo độ tương đồng cosine giữa bản tin và báo cáo tài chính hoặc cuộc gọi thu nhập của các công ty, các phương pháp này **không tạo ra danh mục có alpha dương đáng kể**, cho thấy sự vượt trội của ChatGPT trong khả năng trích xuất và suy luận từ thông tin phi cấu trúc.

7 Kết luận

Trong nghiên cứu này, nhóm tác giả đã kiểm tra liệu một mô hình ngôn ngữ lớn như ChatGPT có thể hỗ trợ nhà đầu tư ra quyết định dựa trên khả năng suy luận ngôn ngữ từ các dữ liệu văn bản phức tạp hay không. Thay vì dựa vào mô hình học máy có giám sát truyền thống, ChatGPT sử dụng khả năng học từ ngữ cảnh và suy luận để đưa ra khuyến nghị đầu tư mà không cần cấu trúc dữ liệu cố định.

Kết quả chính cho thấy rằng ChatGPT:

- Có khả năng hiểu và liên hệ thông tin từ bản tin chính sách với cổ phiếu cụ thể một cách hiệu quả hơn các phương pháp truyền thống.
- Tạo ra danh mục đầu tư có alpha dương và có ý nghĩa thống kê, đặc biệt là với các chủ đề như đầu tư công, xây dựng, chính trị, ESG.
- Hoạt động tốt ngay cả khi xử lý ngôn ngữ không phải tiếng Anh, như trong trường hợp phân tích chính sách Trung Quốc.
- Cho thấy hiệu suất được cải thiện rõ rệt khi thực hiện tinh chỉnh (fine-tuning) đơn giản dựa trên mã ngành (SIC), cho thấy tiềm năng kết hợp giữa mô hình ngôn ngữ tổng quát và kiến thức chuyên ngành.

Nghiên cứu khẳng định tiềm năng ứng dụng của ChatGPT như một công cụ hỗ trợ đầu tư (robo-advisor) thông minh, giúp nhà đầu tư khai thác tốt hơn các nguồn dữ liệu văn bản phi cấu trúc như tin tức, chính sách và báo cáo. ChatGPT mở ra một hướng đi mới trong việc áp dụng Generative AI vào lĩnh vực tài chính định lượng, đặc biệt trong môi trường ngày càng bị ảnh hưởng bởi thông tin thời sự và yếu tố phi định lượng.

Tài liệu tham khảo

- [1] Alsabab, H., Capponi, A., Ruiz Lacedelli, O., & Stern, M. (2021). Robo-advising: Learning investors risk preferences via portfolio choices. *Journal of Financial Econometrics*, 19, 369-392.
- [2] Capponi, A., Olafsson, S., & Zariphopoulou, T. (2022). Personalized robo-advising: Enhancing investment through client interaction. *Management Science*, 68, 2485-2512.

- [3] Carpenter, J. N., & Whitelaw, R. F. (2017). The development of China’s stock market and stakes for the global economy. *Annual Review of Financial Economics*, 9, 233-257.
- [4] Chen, Y., Kelly, B., & Xiu, D. (2023). Expected returns and foundation models of language. *SSRN Working Paper*.
- [5] Cong, L. W., Feng, G., He, J., & He, X. (2022). Growing the efficient frontier on panel trees. *Journal of Financial Economics*.
- [6] Cong, L. W., Feng, G., He, J., & Wang, Y. (2024a). Mosaics of predictability. *NBER Working Paper*.
- [7] Cong, L. W., Liang, T., Zhang, X., & Zhu, W. (2024b). Textual factors: A scalable, interpretable, and data-driven approach to analyzing unstructured information. *NBER Working Paper*.
- [8] Cong, L. W., Tang, K., Wang, J., & Zhang, Y. (2019). AlphaPortfolio: Direct construction through deep reinforcement learning and interpretable AI. *SSRN Working Paper*.
- [9] DAcunto, F., Prabhala, N., & Rossi, A. G. (2019). The promises and pitfalls of robo-advising. *The Review of Financial Studies*, 32, 1983-2020.
- [10] Eisfeldt, A., Schubert, G., Taska, B., & Zhang, M. (2023). Generative AI and firm values. *SSRN Working Paper*.
- [11] Fama, E. F., & French, K. R. (2016). Dissecting anomalies with a five-factor model. *The Review of Financial Studies*, 29, 69-103.
- [12] Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57, 535-574.
- [13] Giglio, S., Liao, Y., & Xiu, D. (2021). Thousands of alpha tests. *The Review of Financial Studies*, 34, 3456-3496.
- [14] Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33, 2223-2273.
- [15] Jha, M., Qian, J., Weber, M., & Yang, B. (2023). ChatGPT and corporate policies. *SSRN Working Paper*.
- [16] Jiang, J., Kelly, B. T., & Xiu, D. (2022). (Re-)imag(in)ing price trends. *Journal of Finance*.
- [17] Kelly, B., & Xiu, D. (2023). Financial machine learning. *SSRN Working Paper*.
- [18] Kelly, B. T., Malamud, S., & Zhou, K. (2022). The virtue of complexity in return prediction. *SSRN Working Paper*.
- [19] Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66, 35-65.
- [20] Sheng, J., Sun, Z., Yang, B., & Zhang, A. L. (2024). Generative AI and asset management. *SSRN Working Paper*.

- [21] Zhang, Z., Chen, J., & Shi, Y. (2023). ChatGPT's Reasoning Ability and Potential in Stock Selection Based on News. *SSRN Electronic Journal*. DOI: 10.2139/ssrn.4519182.
- [22] Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), 3-56.
- [23] Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1-22.
- [24] Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1), 65-91.