

# ChatGPT, Generative AI, LLMs, and Investment Advisory\*

Lei Huang<sup>a</sup>

Fangzhou Lu<sup>a</sup>

Sixuan Li<sup>a</sup>

December 17, 2024

## Abstract

Large language models (LLMs), including ChatGPT, leverage their reasoning capabilities to create portfolio recommendations beyond the reach of conventional textual analysis. By refining the LLM model with specific training data and adjusting ChatGPT's parameters for enhanced output flexibility, there is potential for ChatGPT to craft portfolios that outperform market benchmarks in out-of-sample tests. Utilizing two distinct types of textual data in different languages, articles from the Wall Street Journal in the U.S. and policy announcements from the Chinese government, we show that ChatGPT can produce portfolios with a monthly three-factor alpha of up to 3%, particularly in response to policy-related news. When comparing these outcomes to those generated by traditional textual analysis methods, we observe that the conventional methods fail to produce portfolios with positive alpha.

JEL Codes: G12, G14.

Keywords: Generative AI, ChatGPT, Textual analysis, Machine Learning

---

<sup>a</sup>HKU Business School, The University of Hong Kong. Please send correspondence to: lufz@hku.hk. We also thank colleagues at HKU, especially Alan Kwan and Dunhong Jin, for their helpful comments.

# ChatGPT, Generative AI, LLMs, and Investment Advisory

## Abstract

Large language models (LLMs), including ChatGPT, leverage their reasoning capabilities to create portfolio recommendations beyond the reach of conventional textual analysis. By refining the LLM model with specific training data and adjusting ChatGPT's parameters for enhanced output flexibility, there is potential for ChatGPT to craft portfolios that outperform market benchmarks in out-of-sample tests. Utilizing two distinct types of textual data in different languages, articles from the Wall Street Journal in the U.S. and policy announcements from the Chinese government, we show that ChatGPT can produce portfolios with a monthly three-factor alpha of up to 3%, particularly in response to policy-related news. When comparing these outcomes to those generated by traditional textual analysis methods, we observe that the conventional methods fail to produce portfolios with positive alpha.

JEL Codes: G12, G14.

Keywords: Generative AI, ChatGPT, Textual analysis, Machine Learning

# 1 Introduction

The recent emergence of generative Artificial Intelligence (AI) has changed the way people think about the limits of what AI can do. While textual analysis and machine learning have greatly improved our understanding of finance in both asset pricing and corporate finance, both methodologies are still to some extent limited by pre-specific structures or human supervision (Cong, Liang, Zhang, and Zhu, 2024b; Kelly and Xiu, 2023; Chen, Kelly, and Xiu, 2023). Generative AI combined with machine learning provides a new perspective to help potential portfolio managers determine the optimal portfolio and potentially serve as retail investors' financial advisors (Cong, Tang, Wang, and Zhang, 2019; Capponi, Olafsson, and Zariphopoulou, 2022; Sheng, Sun, Yang, and Zhang, 2024).

In this paper, we test whether generative AI such as ChatGPT can help investors make investment decisions. ChatGPT is a large language model with reasoning ability trained on a massive amount of diverse text data from the Internet developed by OpenAI. Its reasoning ability is based on the patterns and associations learned from a vast amount of text data during the training process. It relies on the transformer architecture to understand and generate human-like text responses. Its reasoning ability can be broken down into a few key components: pattern recognition, inference, knowledge recall, analogical reasoning, probabilistic reasoning, and error analysis. Therefore, we hypothesize that, unlike traditional textual analysis, ChatGPT has the ability to recommend stocks or assets based on complicated textual data with only limited pre-determined structures or instructions.

Our methodology relies on ChatGPT and is straightforward and intuitive. It involves querying ChatGPT using textual data to determine suitable investments, such as equity, ETFs, mutual funds, or investments in specific industries, for an investor. ChatGPT exhibits a notable advantage over other textual analyses by virtue of its unique capacity to establish direct associations with specific firms. This capability allows ChatGPT to effectively comprehend news data that presents challenges for conventional forms of analysis,

including traditional textual analysis methods.<sup>1</sup>

In this paper, we construct equity portfolios by following a similar approach to that of an investor consulting with a professional financial analyst.<sup>2</sup> ChatGPT can give equity recommendations based on textual data such as news, and offer an explanation for the rationale of such recommendation. ChatGPT can also offer varying recommendations on the same news due to the influence of the model parameter such as ChatGPT “temperature”.<sup>3</sup> This parameter determine the level of randomness in the generated responses. For example, when the temperature is set to zero, ChatGPT tends to offer consistent recommendations, albeit with somewhat diminished levels of creativity and practicality. Conversely, higher temperature settings empower ChatGPT to generate a wide array of imaginative suggestions, mirroring the dynamic in the real world where different investment managers might provide distinct investment counsel based on the same news, consequently enriching the overall diversity of responses.

Our methodology can be applied to a wide range of news-based data and policy-based textual data. In this paper, we first apply ChatGPT on Wall Street Journal news from 2020 January to 2023 August. Figure 1 shows the word cloud of the kind of news we obtain from Wall Street Journal. We require ChatGPT to recommend five stocks to buy based on each news we collect. This resulted in a total of around 8,400 stock recommendations in our sample period per one round of queries.

To test whether these stocks actually make sense from a reasoning point of view, we manually check whether the reasoning behind those stocks’ recommendations makes sense and whether the stocks recommended by ChatGPT is consistent with a human financial analyst’s judgement. For example, Appendix Table 1 shows that on Oct 10th 2022, the Wall Street Journal publish the news titled: Biden to Withhold Tariffs on Solar Imports.

---

<sup>1</sup>Traditional textual analysis of news requires the presence of firm names for effective processing and analysis.

<sup>2</sup>We demonstrate an example of ChatGPT response in detail in the appendix.

<sup>3</sup>There are other parameters such as “Top P” that can also affects the randomness of response from ChatGPT. We find that varying these other parameters work similarly as varying ChatGPT temperature.

The main text suggests that “The decision would be a win for U.S. solar developers and utilities”. Therefore, it would be reasonable to recommend stocks that are U.S. solar developers and utilities. We hire financial analysts with at least three years of working experience in equity research firms as well as finance-major students to help us check the validity of those recommendations.<sup>4</sup> Based on finance professionals’ evaluation, we find that 87% of the recommendation is logical, 10% is unclear, and only 3% are clearly unrelated from random sampling using ChatGPT 3.5. ChatGPT 4.0’s result is also similar albeit slightly worse in most cases. The evaluations from finance-major students are around the same.

Next, we test whether ChatGPT exhibits predictive ability in the financial markets. In order for our results to have any return predictability, limits of arbitrage must be presented and the news must contain useful information regarding the future cash flow or discount rate. We present our results separately for the in-sample period: January 2020- September 2021, and the out-of-sample period: October 2021 to September 2023. We divide our sample into these two sample periods because ChatGPT’s training data involves all potential Internet data and news up till September 2021. Therefore, trading days after September 2021 naturally becomes the out-of-sample test sample for our purpose. We construct portfolio with varying holding period. The holding period of the stocks ranges from one day to one year. The CAPM, three-factor and five-factor alphas are highest when the holding period is one calendar month. In the earlier sample period between January 2020-September 2021, specifically, the ChatGPT calendar-time long-short portfolio is capable of generating a daily three-factor alpha of 3.4 basis points (bps), and a five-factor alpha of 3.6 bps when the ChatGPT temperature parameter is set to 0. These results are nevertheless statistically insignificant. When we set the temperature parameter to 1 to allow for more creativity from the ChatGPT side, the three-factor alpha is 14.1 bps, and the five-factor alphas is 12.9 bps per day accordingly. These results suggest a monthly in-sample alpha of close to 2.6%.

In terms of recommended portfolio characteristics, ChatGPT recommends a wide range

---

<sup>4</sup>The financial analysts we hire have also passed the CFA Level 3 exam and the students we hire have at least passed CFA Level 1 exam.

of stocks that is hard to pin down with either a Fama-French three-factor or a five-factor model (Fama and French, 2016). These stocks are neither stocks with high momentum nor reversal. This suggests that ChatGPT can adapt its response depending on different news and offer a diversified portfolio as recommendations. When we take a closer check at the specific kind of news that potentially has the strongest return predictability, it is news related to politics. Specifically, high frequency keywords such as “democrat”, “republican”, “house”, “biden”, “senate”, and “vote” are in this category of news. This suggests that either limits of arbitrage is especially concentrated in this category of news, or ChatGPT has a superior understanding of political document, or a combination of both.

In out-of-sample period from September 2021 to February 2024, and when the parameter is set to zero, ChatGPT portfolios exhibit a 7 bps daily five-factor alpha for news related to politics but not other news category. This result suggests a monthly alpha of 1.4% for ChatGPT portfolios based on political news. When the temperature parameter is set to 1, we repetitively query the ChatGPT 25 times to model the distribution of different financial advisors’ recommendations. When we set the temperature parameter to one to allow for more creativity from the ChatGPT side, the three-factor alpha is 1.3 bps, and the five-factor alpha is 1.4 bps per day accordingly. These results suggest a monthly alpha of close to 28 bps. This result is weaker than the results of in-sample tests, but nevertheless shows that the in-sample predictability is not spurious. The results for political related news based ChatGPT portfolio remain economically and statistically significant when the temperature parameter is set to one.

We also query ChatGPT to recommend sectors based on Wall Street Journal news. We use the two-digit SIC code to categorize sectors. Based on the previous analysis, given that ChatGPT can recommend stocks reasonably consistent with financial profession’s judgement, we expect that ChatGPT can also recommend sectors reasonably well. We find that 95% of the recommendation is logical, 4% is unclear, and only 1% are clearly unrelated.

Nevertheless, these industry portfolios do not generate alpha.<sup>5</sup>

Next, given that GPT is a general-purpose large language model without specialized training in finance data, we fine-tune the model using our proprietary training dataset to enhance return predictability. Specifically, during the in-sample period, we refine the model based on the original ChatGPT recommendations. For each news article, we identify the best-performing stock within the next month that shares the same two-digit SIC code as the originally recommended stock by the untrained ChatGPT model, using this as our ideal optimal response. The fine-tuning process adjusts the parameters of the ChatGPT model, enabling us to effectively “train” our large language model to better accommodate the specific task of recommending the most suitable stocks.

After fine-tuning in out-of sample tests, when the temperature parameter is set to 1 in the US out of sample period, the five-factor (also includes Momentum and Reversal) alpha is 3.4 for value-weighted portfolio and 4.1 bps for equal weighted portfolio accordingly. The monthly alpha in the politics topics subsample is further improved by 15 bps per day. This implies a monthly alpha of 3%. Moreover, ChatGPT portfolios based on news topic related to Covid and China (topic 4) as well as Russo-Ukrainian War (topic 13) also generates alphas around 12 bps per day. After fine-tuning, financial analysts find 93% of the recommendation to be logical, 7% is unclear, and only 1% are clearly unrelated. This suggests that fine-tuning improves the quality of equity recommendations. Overall, our results indicate that ChatGPT has the ability to serve as a robo-financial advisor and exhibit certain financial forecasting abilities.

Since the Wall Street Journal results suggest that the ChatGPT apply particular well in the context of policy news, we test ChatGPT’s ability to recommend stocks based on an alternative source, policy news in China. China’s stock market is highly dictated by government intervention (Carpenter and Whitelaw, 2017), therefore, based on the US results, we would hypothesize the result would apply extremely well in the Chinese stock market.

---

<sup>5</sup>We obtain similar results if we use ChatGPT to recommend industry based on three-digit SIC code instead.

We acquire data pertaining to industry-level policy support from the official policy document database of the State Council, which encompasses policy documents published by the State Council, as well as those from various ministries and commissions operating under the auspices of the State Council. China, as a centrally governed country, bestows its government with extensive power and regulatory capabilities. Consequently, the formulation and implementation of industrial policies wield significant influence over the financial markets. If ChatGPT can also recommend quality stocks in the Chinese stock market, it will not only suggest that ChatGPT can recommend stocks in different textual language, different culture, and different market dynamics, but also verify that ChatGPT’s equity recommendation applies particularly well for politics related text.

Therefore, we test whether the individual stocks that ChatGPT recommends in the Chinese stock markets can actually be linked to the specific policy news of the State Council and whether these recommendations are reasonable, similarly as for the Wall Street Journal data in the US. We require ChatGPT to recommend five stocks based on each policy release. This resulted in a total of around 10,000 stock recommendations in our sample period per one round of queries. Similarly as for the US data, We find that 85% of the recommendation is logical, 13% is unclear, and only 2% are clearly unrelated given finance professionals’ evaluation. This result is similar and comparable to the U.S. result we document.

Next, we test whether ChatGPT exhibits predictive ability in the Chinese financial markets. We present our results separately for the in-sample period: January 2004- September 2021, and the out-of-sample period: September 2021 to August 2023. In the earlier sample period between January 2004-September 2021, for holding period of one calendar month, specifically, the ChatGPT calendar-time portfolio is capable of generating a daily three-factor alpha of 4.5 basis points (bps), and a five-factor (including MOM &REV) alpha of 4.6 bps when the ChatGPT temperature parameter is set to 0 using equal-weighted methods. When we set the temperature parameter to one to allow for more creativity from the ChatGPT



side, the three-factor alpha and five-factor alpha is around the same.<sup>6</sup> In terms of recommended portfolio characteristics, ChatGPT recommends a wide range of stocks that is hard to pin down with either a Fama-French three-factor or five-factor model (Fama and French, 2016). This is consistent with the results we observe for the U.S. main sample and suggests that ChatGPT can adapt its response depending on different news and offer a diversified portfolio as recommendations.

In out-of-sample period and when the parameter is set to zero, ChatGPT exhibits a daily alpha especially for topic related to government investment, construction, and ESG related issue (topic 1). The ChatGPT portfolio has an alpha of around 2.5 bps for both the equal and value weighted methods. These results are not statistically significant. When the temperature parameter is set to one, we repetitively query the ChatGPT to model the scenarios of the distribution of different financial advisors' recommendations, 25 times for each parameter. ChatGPT calendar-time portfolio is capable of generating a daily a five-factor alpha of 2.1 bps for all news and a significant daily alpha of 12 bps for topic 1. Those results become much more economically and statistically significant after we fine-tune our models as in the US sample. These results suggest that indeed ChatGPT works well in terms of policy related news and it has the the ability to work with non-English textual data.

As a robustness check, we want to test whether incremental info is helpful for ChatGPT for out-of-sample predictions. Therefore, We test ChatGPT's capacity when ChatGPT can only "read" the title of the announcement policy. We hypothesize that the predictive power should be smaller given only the title of the news announcement. Combined with the results when we use ChatGPT to analyze the full text, the results suggest that ChatGPT does generate incremental predictive results from additional textual information in out-of-sample analysis. Moreover, we find that ChatGPT has incremental predictability than when we use standard benchmark textual analysis model. For example, if we calculate the cosine similarity between 10-K or earning calls of the firms and the news and choose the five stocks

---

<sup>6</sup>We obtain those results when the Top P parameter is set to 0.95. When we adjust this parameter, this result is also similar.

with the top five cosine score, those portfolios do not generate alpha for either Wall Street Journal or China State Council policy news. Therefore, this suggests that ChatGPT are able to extract extra information beyond simply finding stocks with annual report or earning calls that have high similar content as the news content.

Moreover, we benchmark these results using traditional textual analysis methods. Specifically, we employ the cosine similarity method to identify the top five firms whose earnings call textual data and annual report textual data are most similar to the Wall Street Journal news articles and China state policy news data used in our ChatGPT model. Our findings indicate that, overall, and particularly for topics that generate significantly positive alpha portfolios, we find that traditional textual analysis methods yield no comparable results. This suggests that ChatGPT’s machine learning mechanism is superior and goes beyond the capabilities of commonly used textual analysis methods.

Our research is related to the emerging literature on LLM and AIGC. Eisfeldt, Schubert, Taska, and Zhang (2023) show that firms that have higher exposure to AI have higher expected returns. Jha, Qian, Weber, and Yang (2023) show that ChatGPT can extract investment policies from earning calls and this methods perform better than traditional machine learning. Chen et al. (2023) show that LLM models such as Google’s BERT, RoBERTa and OPT by Meta can generate sentiment based on stock specific news and under the assumption of limits of arbitrage, generate investment with high Sharpe ratios. Sheng, Sun, Yang, and Zhang (2024) show that ChatGPT can help hedge fund managers increase their returns and outperform the ones that don’t use ChatGPT. Cong, Liang, Zhang, and Zhu (2024b) develop a general approach to analyze large scale text based data using neural network language processing and generative statistical models. Different from Chen et al. (2023), we rely on ChatGPT to generate direct stock recommendations from news or policies that otherwise cannot be linked to specific stocks.

Our research is related to the literature on robo-advising as ChatGPT or a modified version of the LLM model can potentially serve as robot advisors. D’Acunto, Prabhala,

and Rossi (2019) show that robot-advising can increase investors’ portfolio diversification. Alsabah, Capponi, Ruiz Lacedelli, and Stern (2021) show that robo-advisors with exploration-exploitation algorithms can potentially help improve the performance of stand-alone investors. Capponi, Olafsson, and Zariphopoulou (2022) shows that robot advising helps the most during an economic contraction when investors tend to over-decrease their positions when the market risk-return tradeoff is more favorable. Our results suggest LLM model can potentially help retail investors understand complicated policy news and choose better portfolios.

Our research is related to the literature on the application of Machine Learning in finance. Gu, Kelly, and Xiu (2020) shows the importance of machine learning in the field of empirical asset pricing. Giglio, Liao, and Xiu (2021) use machine learning techniques to develop a test for detecting true alphas. Jiang, Kelly, and Xiu (2022) show that traditional technical analysis images can help to predict stock return. Cong, Feng, He, and Wang (2024a) use a tree-based machine learning model to study the grouped heterogeneity in return predictability across different assets. Cong, Feng, He, and He (2022) use a P-tree model to generalize security sorting for constructing effective test assets and latent pricing factors. We find that a specific machine learning technique, LLM Model, helps predict future stock returns.

Finally, our research is related to the literature on textual analysis. But unlike the traditional textual analysis relies on a pre-specific structure to generate sentiment analysis that leads to recommended portfolios, our approach relies on AI’s reasoning ability to analyze text that cannot be easily parsed with traditional textual analysis methods. Loughran and McDonald (2011) show that with a pre-determined business dictionary different from a traditional word dictionary, this dictionary helps analyze financial statements and reports. Gentzkow, Kelly, and Taddy (2019) show that most textual data can generate predictions for future asset prices, and the interpretation is not necessarily causal. Our result shows that without relying a specific dictionary, ChatGPT can generate stock recommendations based on textual data.

## 2 Institutional Details

### 2.1 Generative AI and ChatGPT

ChatGPT, an abbreviation for Chatbot Generative Pre-trained Transformer, is an advanced and sophisticated language model developed by OpenAI. As a representative example of generative AI, it has been trained on vast amounts of diverse textual data to understand complex language patterns and generate contextually relevant responses in a conversational setting. Through the use of transformer architecture and advanced machine learning techniques, ChatGPT can effectively process and comprehend human language, making it capable of engaging in coherent and contextually appropriate conversations with users.

ChatGPT demonstrates a broad spectrum of functionalities, such as creative writing, content generation, and language translation. Notably, its reasoning ability emerges as a paramount and distinguished aspect, encompassing a range of cognitive faculties, including logical thinking, inferring interconnections among diverse concepts, and drawing well-informed conclusions. So far, ChatGPT has underwent rigorous evaluation aimed at testing its reasoning ability, including passing standardized tests such as LSAT (Law School Admission Test), SAT (Scholastic Assessment Test), GRE (Graduate Record Examinations), and more.<sup>7</sup> These accomplishments showcase that ChatGPT indeed has impressive reasoning abilities, such as critical thinking, reading comprehension, data analysis, and decision-making. This strong evidence further reinforces the justification for using ChatGPT to analyze news and political announcements and offer investment recommendations in this paper.

---

<sup>7</sup>Specifically, ChatGPT-4 achieved a commendable score of 710 out of 800 on the SAT reading and writing section, placing it above 93% of human test takers. Such an impressive performance suggests that ChatGPT-4 possesses the potential to meet the academic standards required for admission to prestigious Ivy League institutions like Harvard, Columbia, and Princeton. On the LSAT exam, ChatGPT-4 achieved a score of 163 in the 88th percentile, putting it in a good position to be admitted into a top 20 law school. In the GRE Verbal test, ChatGPT-4 achieved an outstanding score of 169 out of 170, surpassing 99% of all test takers.

## 2.2 ChatGPT Fine-Tuning

Fine-tuning is the process of training a pre-trained language model, such as ChatGPT, on a specific dataset to adapt it to a particular task or domain. In the context of stock recommendations, the fine-tuning process can be used to make the model more sensitive and knowledgeable about market dynamics, news, and policy releases. Fine-tuning can help improve stock recommendations for the following four reasons.

First, fine-tuning can help with domain-specific knowledge: By fine-tuning ChatGPT on a dataset containing financial news, articles, and policy announcements, the model can learn to better understand the language and context of the financial world. This enables it to generate more accurate and relevant stock recommendations. Second, fine-tuning can help with Event-driven analysis: Fine-tuning can help ChatGPT to identify the impact of specific news and policy announcements on stock prices. This way, the model can make recommendations based on real-world events, such as earnings reports, product launches, or government policy changes. Third, fine-tuning can improve sentiment analysis: Through fine-tuning, ChatGPT can learn to gauge the sentiment of financial news and policy releases. Understanding the sentiment of these sources can help the model predict how the market will react to certain events, which in turn can inform its stock recommendations. Finally, fine-tuning can lead to improved reasoning: Fine-tuning can enhance ChatGPT's ability to reason and make connections between various pieces of information. This improved reasoning can help the model generate more informed and accurate stock recommendations based on the available data.

In light of the “virtue of complexity” concept, as expounded upon by Kelly, Malamud, and Zhou (2022), utilizing ChatGPT for stock recommendations offers numerous advantages. This concept posits that as the number of parameters used to estimate an econometric model increases, it can yield better out-of-sample estimations. ChatGPT, with its vast number of parameters, is ideally suited to capitalize on this notion. Its ability to process and comprehend large volumes of textual information enables it to generate informed recommendations

grounded in its extensive understanding of market dynamics, corporate news, and policy releases. Moreover, the fine-tuning capabilities of ChatGPT allow for enhanced domain-specific knowledge and improved reasoning, enabling the model to discern intricate patterns and make connections that may elude traditional analytical methods. Consequently, embracing the sophistication of ChatGPT as an auxiliary tool for stock recommendations, in accordance with the “virtue of complexity” principle, can provide investors with a more comprehensive and nuanced perspective, thus potentially leading to superior investment decisions.

In spite of the potential benefits associated with utilizing ChatGPT for stock recommendations, certain drawbacks must be considered, particularly when comparing it to Google and Meta’s large language models (LLMs). One notable disadvantage is the increased opacity of ChatGPT’s LLM, which can be perceived as more of a black box compared to its counterparts. This is due to the fact that users typically cannot access the parameter values associated with normal usage or the parameter values after fine-tuning. Such limited transparency may hinder users’ ability to evaluate the model’s reasoning process and the validity of its recommendations. Consequently, this opacity could raise concerns about the trustworthiness and reliability of ChatGPT-generated recommendations, potentially compromising investors’ confidence in the model’s capacity to provide valuable insights and informed advice. As a result, the lack of transparency in ChatGPT’s LLM constitutes a drawback when considering its application for stock recommendations, particularly when juxtaposed against more transparent alternatives.

## **2.3 Textual Data**

### **2.3.1 Wall Street Journal**

In our baseline analysis, we incorporate all Wall Street Journal articles, resulting in a comprehensive dataset of approximately 16,000 articles spanning from January 2020 to August 2023. Utilizing the Latent Dirichlet Allocation (LDA) method to categorize these news ar-

ticles into distinct groups, the LDA Coherence score identifies the optimal number of news clusters as 13.<sup>8</sup> A few categories emerge with clear thematic labels. For instance, Topic 2 focuses on price and inflation, Topic 3 pertains to politics, Topic 4 addresses COVID-19, Topic 5 encompasses energy-related matters, Topic 8 deals with labor issues, and Topic 13 covers Russia and Ukraine.

### **2.3.2 China’s State Council Policy release**

In assessing ChatGPT’s stock recommendation capabilities, we not only consider textual data that have been extensively used in the literature such as the Wall Street Journal, but also other seemingly orthogonal source, such as the State Council Policy releases, which could provide a useful robustness check for ChatGPT’s ability to process and interpret information in a different language and context. Furthermore, the inclusion of Chinese policy releases offers valuable insights into the complex interdependence between the world’s two largest economies, which can enhance the predictive capabilities of ChatGPT. By adopting a broader range of source materials, we can more comprehensively assess the performance of ChatGPT’s stock recommendations, thereby contributing to the growing trend in finance research towards a global perspective on market dynamics.

China, as a centrally governed state, attaches immense significance to the issuance of State Council policies. These policies span diverse domains, including support for strategic emerging industries, the promotion of upgrading traditional industries, the facilitation of technological innovation, the advancement of green and environmentally sustainable industries, and the fostering of balanced regional development. Playing a vital role in guiding the country’s economic development and industrial upgrading, these policies are pivotal in shaping the nation’s economic landscape.

The impact of State Council policies on the financial market is profound. Firstly, policy releases direct capital flows into various industrial sectors, thus influencing the profitability

---

<sup>8</sup>Figure 5 shows the result from the LDA Coherence score.

and development prospects of relevant industrial companies. Secondly, industrial policies often encompass an array of financial incentives, tax benefits, and supportive measures to encourage investment and foster growth in targeted sectors. These incentives may take the form of grants, subsidies, or preferential tax rates, providing direct financial support to enterprises operating within the specified industries. In addition, industrial policies can affect the competitiveness of specific firms. Companies that adhere to the policy goals and criteria may receive preferential treatment, enhancing their financial standing and credit ratings. These firms can access more favorable financing opportunities, which helps them enhance their competitive edge and pursue growth opportunities. Therefore, State Council policies can give some indication of how stocks and other assets might perform in the future, serving as important signals for investors. By scrutinizing the content and implications of these policies, investors can gain valuable insights into the government’s strategic priorities and potential market trends, and then make informed decisions about their investments.

### 3 Methodology

We use a simple and straightforward conversational query to extract investment advisory similar as how one individual would communicate with his or her investment advisor. Here are the sample prompt we use to extract ChatGPT’s equity recommendation.

U.S. Equity Recommendation Instruction based on Wall Street Journal: “Suppose you are a senior financial analyst. Please carefully read the following Wall Street Journal news. According to the content of the news, if you have recommended stocks, write ”YES” and list 5 NYSE or Nasdaq stock names and the stock codes that you recommend. If you just answered ”NO”, please briefly explain the reason.”

U.S. Sector Recommendation Instruction based on Wall Street Journal: “Suppose you are a senior financial analyst. Please carefully read the following policy news. Based on the content of the policy news and referring to SIC two-digit classification standard, if you have



recommended sectors, write "YES" and list the category code of the recommended sector. If you have just answered "NO", please briefly explain the reason."

The Appendix shows an example of the stock that ChatGPT offers. On May 14th 2023, The Wall Street Journal published an article titled: Biden to Withhold Tariffs on Solar Imports. This article talks about the winners and losers for this new executive order by Biden. We conduct one rounds of query when the temperature parameter is set to one for ChatGPT 3.5 and the ChatGPT recommends five stocks in total.

We later improve ChatGPT's stock recommendation ability based on fine-tuning. Specifically, for Wall Street Journal, we pre-train the ChatGPT based on the sample in 2018 to 2019, or for each news, we give ChatGPT the top five performing stocks with the same two SIC code based on the initial recommendation, and use this sample to fine-tune ChatGPT's output.

Next, we hired industry veteran coupled with finance-major students to help us check whether the reasoning behind those stocks' recommendations makes sense and whether the stocks recommended by ChatGPT belong to the specific industry that the industrial policy is about. The financial analysts we hire have passed the CFA Level 3 exam and the students we hire have at least passed CFA Level 1 exam. Firstly, we provided the participants with a brief overview of ChatGPT and its capabilities in generating stock recommendations based on policy announcements. Then instructed the participants to evaluate the recommendations and decide whether the recommendation is logical, unclear, and clearly unrelated based on the following criteria: a. Relevance to the industry specified in the policy announcement.<sup>9</sup> b. The soundness of the reasoning behind the recommendations. c. Potential investment opportunities based on the policy and the recommended stocks. For each stock recommendation, we ask two financial analysts to evaluate.<sup>10</sup> If they reach the same conclusion, then

---

<sup>9</sup>For example, if an industry policy is about promoting electric vehicles, then recommending electric vehicle firms or their suppliers with the potential to grow would seem logical while recommending traditional car manufacturers would be unclear if no obvious rationale is provided, and recommending real estate firms would clearly be unrelated. We hire master of finance students to help us check the validity of those recommendations.

<sup>10</sup>For a slightly larger sample, we ask finance-major students to help with the evaluation, the results look

we directly get our corresponding result, if the two financial analysts have different opinions, we ask a third analysts to give another evaluation, and the majority vote gets to our result. If all three financial analysts give different results, we discard this sample. For 92% of our sample, we get the result in the first round, in 8% of the cases, we invite a third analyst. Under no scenario that the third analysts reach a third conclusion.

Table 1 presents the summary statistics for the stock characteristics and word distribution in Wall Street Journal articles and China’s State Council policy news. Panel A displays the characteristics of stocks recommended by ChatGPT in the U.S. market and the word distribution of the Wall Street Journal articles in the sample from January 2020 to August 2023. The stock characteristics include the Scholes-Williams-Dimson beta (BETA), market capitalization (SIZE), Fama-French book-to-market ratio (BM), investment (INV), operating profitability (PROFIT), Jegadeesh-Titman momentum (MOM), and Jegadeesh-Lehmann short-term reversal (REV). The word distribution of the Wall Street Journal articles is characterized by the mean, standard deviation, and 5th and 95th percentiles of the number of words per article.

Panel B reports the characteristics of stocks recommended by ChatGPT in the Chinese market and the word distribution of China’s State Council policy news in the sample from January 2004 to August 2023. The stock characteristics are similar to those presented in Panel A, providing insights into the Chinese market’s recommended stocks. The word distribution of China’s State Council policy news is also summarized by the mean, standard deviation, and 5th and 95th percentiles of the number of words per news item. Figure 2 shows the number of policy news releases by year over our sample period for a total of over 20,000 Wall Street Journal news. ChatGPT finds roughly less than half of these policy releases are relevant for stock recommendation in the time series. This result does not vary much by the temperature parameter we set. Overall, the summary statistics highlight the diversity in both stock characteristics and policy news content, which provides a rich context

---

highly similar.

for our analysis.

## 4 Testing ChatGPT using U.S. policy data

### 4.1 Reasoning Test based on human rating

Due to the large quantity of policy news and stock recommendations, we randomly sample one hundred policy news per run to conduct the manual check. Table 2 presents an evaluation of ChatGPT’s investment advisory based on human ratings in terms of Relevance, Reasoning, and Investment Opportunity. The results are categorized into logical, unclear, and unrelated recommendations, and are further divided into two panels: (a) ChatGPT 3.5 and (b) ChatGPT 4.0. Panel A presents the results of financial analysts’ evaluations, while Panel B showcases the results of finance-major students’ evaluations. The temperature parameter ( $T$ ) is adjusted to control the creativity of ChatGPT’s recommendations, with values of 0, 1, and 1.5. The subsamples include Domestic Equity and Sector recommendation separately.

In the Domestic Equity subsample, where ChatGPT recommends NYSE or Nasdaq stocks based on the full text of the Wall Street Journal, the percentage of logical recommendations ranges from 74% to 90% across different evaluators and temperature settings. The Sector subsample, where ChatGPT recommends a sector (two-digit SIC code) to invest in based on the Wall Street Journal content, exhibits a slightly higher percentage of logical recommendations, ranging from 83% to 95%.

If we compare ChatGPT 3.5 and ChatGPT 4.0, we observe that ChatGPT 3.5 generally yields a higher percentage of logical recommendations across both evaluator groups and subsamples. For instance, in the Domestic Equity subsample with  $T=0$ , financial analysts rate 89% of ChatGPT 3.5 recommendations as logical, while only 82% of ChatGPT 4.0 recommendations receive the same rating. A similar trend is observed among finance students’ evaluations. This finding suggests that ChatGPT 3.5 outperforms ChatGPT 4.0 in terms of

providing logical investment advice.

In summary, the table demonstrates that ChatGPT’s investment advisory, in terms of Relevance, Reasoning, and Investment Opportunity, is predominantly logical across both versions 3.5 and 4.0, and across different subsamples and temperature settings. The unclear and unrelated recommendations are generally lower, indicating the potential effectiveness of ChatGPT as an investment advisory tool.

## 4.2 US Equity: In-sample test

Table 3 presents the in-sample tests of ChatGPT’s recommended equity based on the Wall Street Journal content, focusing on the daily alphas of the constructed ChatGPT calendar-time portfolios. The table reports results for various holding periods—three months, one month, one week, and one day—across different panels. The temperature parameter (T) is used to adjust the creativity of ChatGPT’s recommendations, with T=0 and T=1 considered. When the temperature is set to one, the portfolio includes all recommended stocks from 25 queries as different queries generate potentially different portfolios.<sup>11</sup> The daily alphas are reported for the CAPM model, Fama-French three-factor model (FF3), five-factor model (FF5), and five-factor model including momentum and reversal (FF5 MOM+REV) between the sample period of January 1st, 2020 and August 30th, 2021. The equal-weighted (EW) and value-weighted (VW) portfolio returns are both considered, and returns are scaled by 10,000 (reported in basis points). T-statistics are reported in parentheses.

For the three-month holding period, the performance of ChatGPT 3.5 recommendations is generally better than that of ChatGPT 4.0, with higher alphas across all factor models and both EW and VW portfolios. This pattern is observed across the majority of holding periods, suggesting that ChatGPT 3.5 generally outperforms ChatGPT 4.0 in providing profitable investment recommendations.

Regarding the temperature parameter’s influence, the results reveal that the performance

---

<sup>11</sup>We omit repetitive equity recommendations for the same news.

varies across different holding periods and factor models. For example, in the one-month holding period with ChatGPT 3.5, the CAPM alpha for the VW portfolio increases from 3.592 (T=0) to 14.429 (T=1), suggesting that higher creativity levels can improve the recommendations' profitability.

It is noteworthy that the one-month holding period consistently exhibits the strongest performance across both ChatGPT 3.5 and ChatGPT 4.0 recommendations. Specifically, the alphas in the one-month holding period are higher than those in other holding periods, such as three months, one week, or one day, across all factor models and both EW and VW portfolios. This finding suggests that ChatGPT's investment recommendations based on the Wall Street Journal content are particularly effective for short-term investment horizons. The superior performance of the one-month holding period may be attributed to the time-sensitive nature of the information contained in the Wall Street Journal, which could have a more immediate impact on stock prices.

Overall, the results in Table 3 indicate that ChatGPT 3.5 generally outperforms ChatGPT 4.0 in terms of providing profitable investment recommendations based on the Wall Street Journal content. The temperature parameter's influence on the performance is not consistent across all holding periods and factor models, suggesting that the optimal level of creativity may depend on the specific context. One-month holding period is the optimal holding period given ChatGPT's recommendations.

Next we investigate that whether stocks recommended by ChatGPT exhibit any specific characteristics. Table 4 presents the in-sample tests of ChatGPT's equity recommendations based on Wall Street Journal content, focusing on the factor exposure of the constructed ChatGPT calendar-time portfolios. The holding period for each ChatGPT recommended stock is one calendar month. The temperature parameter (T) is used to adjust the creativity of ChatGPT's recommendations, with T=0 and T=1 considered. In terms of factor exposure, the market factor (MKT) is consistently positive and statistically significant across all specifications, indicating that the ChatGPT recommended portfolios are positively exposed to

the market risk. Meanwhile, the size factor (SMB) has a mixed exposure, being negative and significant for VW portfolios, but positive and significant for EW portfolios. This suggests that the recommended portfolios have different tilts toward small-cap stocks depending on the weighting scheme. The value factor (HML) and the other factors (RMW, CMA, MOM, and REV) show varying levels of significance and coefficients across different specifications, indicating that ChatGPT’s recommendations do not have a consistent exposure to these factors.

### 4.3 US Equity: Out-of-sample test

Due to the fact that ChatGPT 3.5 exhibits better reasoning ability, we mainly use ChatGPT 3.5 for our out-of-sample portfolio analysis when using the holding period as one calendar month.<sup>12</sup> Table 5 presents the out-of-sample tests of ChatGPT’s equity recommendations based on Wall Street Journal content, focusing on the Fama-French five-factor model alphas (which also include the Momentum and Reversal factor). The holding period for each ChatGPT recommended stock is one calendar month. The temperature parameter (T) is used to adjust the creativity of ChatGPT’s recommendations, with T=0 and T=1 considered. The equal-weighted (EW) and value-weighted (VW) portfolio returns are both considered, and returns are scaled by 10,000 (reported in basis points).

The table shows the alphas for the whole out-of-sample tests and separately for 13 topics generated from the LDA analysis. For the overall out-of-sample test, the alphas are generally positive but not statistically significant when T=0, indicating that the model’s recommendations do not consistently generate excess returns in the out-of-sample period. However, when T=1, the alphas become more positive and statistically significant for both EW and VW portfolios, suggesting that increased creativity in ChatGPT’s recommendations can lead to

---

<sup>12</sup>We also conduct the baseline analysis using ChatGPT 4.0 and the results are not better than those when we generate using ChatGPT 3.5. Therefore, for brevity, we do not report those results. Moreover, we also conducted analysis using different holding period, similar to the in-sample results, the one-month holding period is also optimal for out-of-sample analysis. We also mainly vary the parameter temperature instead of Top P in our analysis. The Top P parameter is set as 0.95.

better out-of-sample performance.

When examining the results by topic, the performance varies across different topic categories. For instance, Topic 3 demonstrates strong and statistically significant alphas across both temperature settings and weighting schemes, indicating that ChatGPT’s recommendations are particularly effective for this specific topic. When we take a closer check at Topic 3, it is news related to politics. Specifically, high frequency keywords such as “democrat”, “republican”, “house”, “Biden”, “senate”, and “vote” are in this category of news. ChatGPT portfolios exhibit a 11 bps daily alpha for a five-factor model. This result suggests a monthly alpha of 2.2% for ChatGPT portfolios based on political news. In contrast, some topics, such as Topic 2 and Topic 10, show negative and insignificant alphas, suggesting that the model’s recommendations for these topics do not consistently generate excess returns in the out-of-sample period.

In summary, Table 5 reveals that ChatGPT’s equity recommendations based on the Wall Street Journal content exhibit varying out-of-sample performance depending on the topic and temperature setting. While the overall out-of-sample performance is not consistently significant, higher creativity levels ( $T=1$ ) tend to yield better results, particularly in certain topic categories.

#### 4.4 US Equity: Fine-Tuning

Table 6 presents the results of out-of-sample tests for the Fama-French five factors model, augmented with Momentum (MOM) and Reversal (REV) factors, on recommended equity portfolios based on Wall Street Journal articles. The table displays alphas for both equal-weighted (EW) and value-weighted (VW) portfolios, considering all topics and 13 individual topics.

The overall results for all topics show that, when time  $T=1$ , the alphas are positive and but statistically insignificant for both VW (3.480) and EW (4.195) portfolios. This indicates that the recommended equity portfolios exhibit abnormal returns after adjusting for the

Fama-French five factors, MOM, and REV factors.

When examining the alphas for individual topics, the results vary. For Topic 3 or political related topic, the alphas are even more positive and highly significant for both VW (7.481) and EW (7.039) when  $T=0$ , and even more so when  $T=1$  with VW (15.429) and EW (15.402) alphas significant at the 5% level. This suggests that the recommended equity portfolios for Topic 3 consistently outperform the market, and can potentially generate a monthly five-factor alpha up to 3% per month.

For Topic 4 (related to Covid=19 and China), the results display a noticeable change between  $T=0$  and  $T=1$ . When  $T=0$ , the alphas for both VW (0.732) and EW (0.385) portfolios are positive but not statistically significant, suggesting that the recommended equity portfolios do not exhibit abnormal returns after accounting for the Fama-French five factors, MOM, and REV factors at this time. However, when  $T=1$ , the alphas for VW (7.584) and EW (8.942) portfolios turn positive and statistically significant at the 5% and 1% levels, respectively.

Topic 13 (related to the Russo-Ukrainian War) also exhibits interesting results. When  $T=0$ , the alphas for both VW (1.145) and EW (1.942) portfolios are positive and statistically insignificant. However, when  $T=1$ , the alphas for both VW (12.195) and EW (12.104) portfolios not only remain positive but also become highly statistically significant at the 5% level. This demonstrates that the recommended equity portfolios for Topic 13 generate abnormal returns after adjusting for the risk factors at the initial time period.

We benchmark these results using traditional textual analysis methods. Specifically, we employ the cosine similarity method to identify the top five firms whose earnings call textual data and annual report textual data are most similar to the Wall Street Journal news articles used in our ChatGPT model. Our findings indicate that, overall, and particularly for topics such as Topic 3, Topic 4, and Topic 13, Table 7 shows that traditional textual analysis methods yield no comparable results. This suggests that ChatGPT’s machine learning mechanism is superior and goes beyond the capabilities of commonly used textual analysis



methods.

Figure 3 summarize the buy-and-hold return of various ChatGPT portfolios as well as the U.S. value-weighted market portfolio over time in the out-of-sample period. The market portfolio’s return is overall negative, and without fine-tuning, the ChatGPT portfolio does slightly better than the market portfolio. After fine-tuning, the overall ChatGPT equal-weighted and value weighted portfolio generates a cumulative return between 15% to 20% over a two year period. If we only consider news related to political topic (topic 3), the ChatGPT portfolio earns significantly higher return than the market portfolio and the cumulative return over the two year out-of-sample period is between 40% to 50%. Overall, these results suggest that the ChatGPT can generate portfolios that out-perform the market in out-of-sample periods.

## 5 Testing ChatGPT using Chinese policy data

Since the Wall Street Journal results suggest that the ChatGPT apply particular well in the context of policy news, we test ChatGPT’s ability to recommend stocks based on an alternative source, policy news in China. China’s stock market is highly dictated by government intervention (Carpenter and Whitelaw, 2017), therefore, based on the US results, we would hypothesize the result would apply extremely well in the Chinese stock market. We acquire data pertaining to industry-level policy support from the official policy document database of the State Council, which encompasses policy documents published by the State Council, as well as those from various ministries and commissions operating under the auspices of the State Council. China, as a centrally governed country, bestows its government with extensive power and regulatory capabilities. Consequently, the formulation and implementation of industrial policies wield significant influence over the financial markets.

Figure 2 shows the number of policy news releases by year over our sample period for a total of over 8,000 policy news. ChatGPT finds roughly half of these policy releases are

relevant for stock recommendation in the time series. This result does not vary much by the temperature parameter we set, similarly as for the Wall Street Journal.

## 5.1 Reasoning Test based on human rating

Table 7 presents the results of an evaluation of ChatGPT’s investment advisory for the Chinese market, specifically examining the recommendation’s Relevance, Reasoning, and Investment Opportunity. The evaluation is conducted by financial analysts (Panel A) and finance-major students (Panel B). The table showcases the results for both ChatGPT 3.5 and ChatGPT 4.0, with varying temperature parameters ( $T=0$ ,  $T=1$ , and  $T=1.5$ ) that adjust the creativity of ChatGPT’s recommendations.

For Domestic Equity recommendations, the results indicate that ChatGPT 3.5 generally receives higher logical ratings than ChatGPT 4.0 across all temperature parameters and both evaluator groups. Specifically, when  $T=1$ , 89% of financial analysts and 84% of finance students rate ChatGPT 3.5’s recommendations as logical, compared to 83% for ChatGPT 4.0 in both groups.

In the case of Sector recommendations, ChatGPT 3.5 consistently outperforms ChatGPT 4.0 in terms of logical ratings across all temperature parameters and both evaluator groups. For instance, when  $T=0$ , 97% of financial analysts and 94% of finance students rate ChatGPT 3.5’s recommendations as logical, compared to 89% for ChatGPT 4.0 in both groups. Similar to Domestic Equity recommendations, the percentage of logical ratings for Sector recommendations decreases as the temperature parameter increases for ChatGPT 4.0, while it remains relatively stable for ChatGPT 3.5.

Overall, the results suggest that ChatGPT 3.5 outperforms ChatGPT 4.0 in terms of providing logically rated investment advice for the Chinese market. Additionally, the results indicate that, for both versions of ChatGPT, a lower temperature parameter ( $T=0$  or  $T=1$ ) tends to yield more logical recommendations, as compared to a higher temperature parameter ( $T=1.5$ ).

## 5.2 Chinese Equity: In-sample test

Table 8 reports the alphas of equity recommendations based on China State Council policy releases using ChatGPT 3.5 and ChatGPT 4.0, with varying temperature parameters ( $T=0$  and  $T=1$ ). The alphas are calculated using the Chinese five-factor model, including momentum and reversal factors, for the in-sample period from January 1st, 2014 to August 30th, 2021.

The table presents results for different holding periods: one year, six months, three months, one month, one week, and one day, with both equal-weighted (EW) and value-weighted (VW) portfolio returns. The returns are scaled by 10,000 (reported in basis points), and the t-statistics are shown in parentheses.

For the GPT 3.5 model with a temperature setting of 1, the one-month holding period yields a statistically significant daily alpha of 6.982 basis points (t-statistic = 3.187) for the value-weighted portfolio. Similarly, the equal-weighted portfolio's one-month holding period produces a significant daily alpha of 8.511 basis points (t-statistic = 4.001). The other holding periods do not show consistently significant results.

In the GPT 4.0 model, the result indicates that the one-month holding period under the temperature setting of 1 yields a daily alpha of 2.174 basis points (t-statistic = 1.019) for the equal-weighted portfolio. Although the t-statistic is lower than in the previous table, the results still suggest a positive alpha for this holding period. The value-weighted portfolio does not exhibit any significant results across different holding periods.

Overall, the results suggest that ChatGPT 3.5 exhibits more positive and statistically significant alphas across different holding periods and temperature parameters, compared to ChatGPT 4.0. This indicates that ChatGPT 3.5's equity recommendations based on China State Council policy releases tend to perform better in the Chinese market. However, it is essential to interpret these results with caution, as the performance may vary across different holding periods and temperature parameters.

Next we investigate that whether stocks recommended by ChatGPT exhibit any spe-

cific characteristics. Table 9 shows that similar to our results for Wall Street Journal, the recommended stocks in the U.S. do not exhibit any particular characteristics.

### 5.3 Chinese Equity: Out-of-sample test

Table 10 presents the out-of-sample test results of the Fama-French five factors model, augmented with Momentum (MOM) and Reversal (REV) factors, on equity recommendations based on China State Council policy releases. The table displays alphas for both equal-weighted (EW) and value-weighted (VW) portfolios, considering all topics together and seven individual topics.

The overall results for all topics show that, when  $T=1$ , the alphas are positive but statistically insignificant at the 10% level for both VW (2.149) and EW (2.245) portfolios. This indicates that the recommended equity portfolios exhibit signs of positive abnormal returns after adjusting for the Fama-French five factors, MOM, and REV factors.

When examining the alphas for individual topics, the results vary. For Topic 1, the alphas are positive but insignificant for both VW (2.194) and EW (2.581) when  $T=0$ . However, they become highly statistically significant when  $T=1$  with VW (8.737) and EW (9.572) alphas, significant at the 10% and 5% level. This suggests that the recommended equity portfolios for Topic 1 consistently outperform the market, even after accounting for the risk factors. From keywords of this topic, this topic is related to construction and investment project of governments.

The results for other topics are mixed, with some topics demonstrating positive and significant alphas, while others do not. This suggests that the performance of recommended equity portfolios based on China State Council policy releases varies across different topics and weighting schemes.

## 5.4 Chinese Equity: Fine-Tuning

Table 11 presents the out-of-sample test results of the Fama-French five factors model, coupled with Momentum (MOM) and Reversal (REV) factors, on equity recommendations based on China State Council policy releases after fine-tuning. The table displays alphas for both equal-weighted (EW) and value-weighted (VW) portfolios, considering all topics together and seven individual topics for the period between September 2021 and February 2024.

For the overall results across all topics, when  $T=1$ , the alphas are positive for both VW (3.185) and EW (4.528) portfolios, but not statistically significant, with t-statistics of 0.591 and 0.841, respectively. This indicates that the recommended equity portfolios do not exhibit significant abnormal returns after adjusting for the Fama-French five factors, MOM, and REV factors following the fine-tuning process.

For Topic 1, the alphas are positive for both VW (2.753) and EW (3.410) when  $T=0$ , but not statistically significant. When  $T=1$ , the alphas for VW (13.482) and EW (13.096) portfolios are positive and significant at the 1% level, suggesting that the recommended equity portfolios for Topic 1 consistently outperform the market after fine-tuning. Topic 2 shows negative alphas for both VW (-1.234) and EW (-0.849) portfolios when  $T=0$ , but they are not statistically significant. When  $T=1$ , the alphas turn positive for both VW (1.489) and EW (2.186), though they are still not statistically significant.

For Topic 3, the alphas are positive but not statistically significant when  $T=0$ . However, when  $T=1$ , the alphas for both VW (12.280) and EW (12.136) are positive and significant at the 5% level. In the case of Topic 4, the alphas are positive but not statistically significant when  $T=0$ . When  $T=1$ , the alphas for both VW (10.210) and EW (10.109) are positive and significant at the 5% level. Topic 5 shows positive alphas when  $T=0$ , but they are not statistically significant. When  $T=1$ , the alphas for both VW (0.821) and EW (0.481) portfolios are positive, but not statistically significant.

For Topic 6, the alphas are positive but not statistically significant when  $T=0$ . When  $T=1$ , the alphas for both VW (11.295) and EW (11.482) are positive and significant at the

10% and 5% levels, respectively. In the case of Topic 7, the alphas are positive but not statistically significant when  $T=0$ . When  $T=1$ , the alphas for both VW (-0.851) and EW (-0.472) portfolios are negative, but not statistically significant.

Again, we benchmark these results using traditional textual analysis methods. Specifically, we employ the cosine similarity method to identify the top five firms whose earnings call textual data and annual report textual data are most similar to the China State policy news data used in our ChatGPT model. Our findings indicate that, overall, and particularly for topics such as Topic 1, Topic 3, Topic 4, and Topic 6, Table 12 shows that traditional textual analysis methods yield no comparable results. This again suggests that ChatGPT's machine learning mechanism is superior and goes beyond the capabilities of commonly used textual analysis methods.

Overall, the results after fine-tuning suggest that the performance of recommended equity portfolios are improved after fine-tuning similarly as in the Wall Street Journal analysis. The recommended equity portfolios exhibit significant abnormal returns after adjusting for the Fama-French five factors, MOM, and REV factors for Topics 1, 3, 4, and 6 when  $T=1$ . However, the performance is not consistently significant across all topics. Figure 4 summarizes the buy-and-hold return of various ChatGPT portfolios as well as the China value-weighted market portfolio over time in the out-of-sample period. The market portfolio's return is overall negative, and without fine-tuning, the ChatGPT portfolio does slightly better than the market portfolio. After fine-tuning, the overall ChatGPT equal-weighted and value weighted portfolio generates a cumulative return between 10% to 20% over a two year period. If we only consider news related to construction and investment project (topic 1), the ChatGPT portfolio earns significantly higher return than the market portfolio and the cumulative return over the two year out-of-sample period is around 40%. Overall, these results suggest that the ChatGPT can generate portfolios that out-perform the market in out-of-sample periods.

## 6 Robustness Checks

As a robustness check, we want to test whether incremental info is helpful for ChatGPT for out-of-sample predictions. Therefore, We test ChatGPT’s capacity when ChatGPT can only “read” the title of the announcement policy. We hypothesize that the predictive power should be smaller only given the title of the news announcement. In the out-of-sample period, we find none of the recommended portfolios have significantly positive three-factor and five-factor alpha regardless of the news topic in either U.S. or China. Combined with the results when ChatGPT was analyzing the full text, the results suggest that ChatGPT does generate incremental predictive results from full textual information in out-of-sample analysis.

Moreover, we find that when we use standard benchmark textual analysis models, these traditional textual analysis do not generate any return predictability. For example, if we calculate the cosine similarity between 10-K (annual report) or earning calls of the firms and the news and choose the top five with the highest cosine score, those portfolios do not generate alpha for either Wall Street Journal or China industry policy news. Therefore, this suggests that ChatGPT are able to extract extra information beyond simply finding stocks with annual report or earning calls that have high similar content as the news content.

## 7 Conclusion

In this paper, we test whether generative AI such as ChatGPT can help investors make investment decisions based on its reasoning ability. Therefore, we hypothesize that, unlike supervised machine learning, ChatGPT has the ability to recommend stocks or assets based on complicated textual data without pre-determined structures. ChatGPT exhibits a notable advantage over other textual analyses by virtue of its unique capacity to establish direct associations with specific firms. This capability allows ChatGPT to effectively comprehend news data that presents challenges for conventional forms of analysis, including traditional

machine learning methods.

We apply ChatGPT on Wall Street Journal and China’s policy release data set as an example, and show that ChatGPT can generate up to a monthly five-factor alpha of 3% per month. Overall, our results indicate that ChatGPT has the ability to serve as a robo-financial advisor and exhibit certain financial forecasting abilities. It is able to provide valuable insights for investors in these market segments.

Our methodology can be applied to a wide range of news-based data and policy-based textual data. Finally, we also find that there is no significant difference when we use ChatGPT 4.0 or ChatGPT 3.5 for our baseline results. We also test alternative LLM models such as Anthropic’s Claude 2.0, and Baidu’s Wenxinyiyan, but no other LLM models exhibit similar reasoning ability as ChatGPT. This again shows the uniqueness of ChatGPT and why we present our main results based on ChatGPT rather than other LLM models. Our research continues to shed light on the potential use of AI and suggests LLM model can potentially help retail investors understand complicated policy news and choose better portfolios. Our result shows that without relying on any specific dictionary, ChatGPT can generate stock recommendations based on textual data.



## References

- Alsabah, 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.
- Capponi, A., Olafsson, S., Zariphopoulou, T., 2022. Personalized robo-advising: Enhancing investment through client interaction. *Management Science* 68, 2485–2512.
- 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.
- Chen, Y., Kelly, B., Xiu, D., 2023. Expected returns and foundation models of language. SSRN Working Paper .
- Cong, L. W., Feng, G., He, J., He, X., 2022. Growing the efficient frontier on panel trees. *Journal of Financial Economics* .
- Cong, L. W., Feng, G., He, J., Wang, Y., 2024a. Mosaics of predictability. NBER Working Paper .
- 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 .
- Cong, L. W., Tang, K., Wang, J., Zhang, Y., 2019. AlphaPortfolio: Direct construction through deep reinforcement learning and interpretable AI. SSRN Working Paper .
- D'Acunto, F., Prabhala, N., Rossi, A. G., 2019. The promises and pitfalls of robo-advising. *The Review of Financial Studies* 32, 1983–2020.
- Eisfeldt, A., Schubert, G., Taska, B., Zhang, M., 2023. Generative AI and firm values. SSRN Working Paper .
- Fama, E. F., French, K. R., 2016. Dissecting anomalies with a five-factor model. *The Review of Financial Studies* 29, 69–103.
- Gentzkow, M., Kelly, B., Taddy, M., 2019. Text as data. *Journal of Economic Literature* 57, 535–74.
- Giglio, S., Liao, Y., Xiu, D., 2021. Thousands of alpha tests. *The Review of Financial Studies* 34, 3456–3496.
- Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. *The Review of Financial Studies* 33, 2223–2273.
- Jha, M., Qian, J., Weber, M., Yang, B., 2023. Chatgpt and corporate policies. SSRN Working Paper .
- Jiang, J., Kelly, B. T., Xiu, D., 2022. (re-) imag (in) ing price trends. *Journal of Finance* .

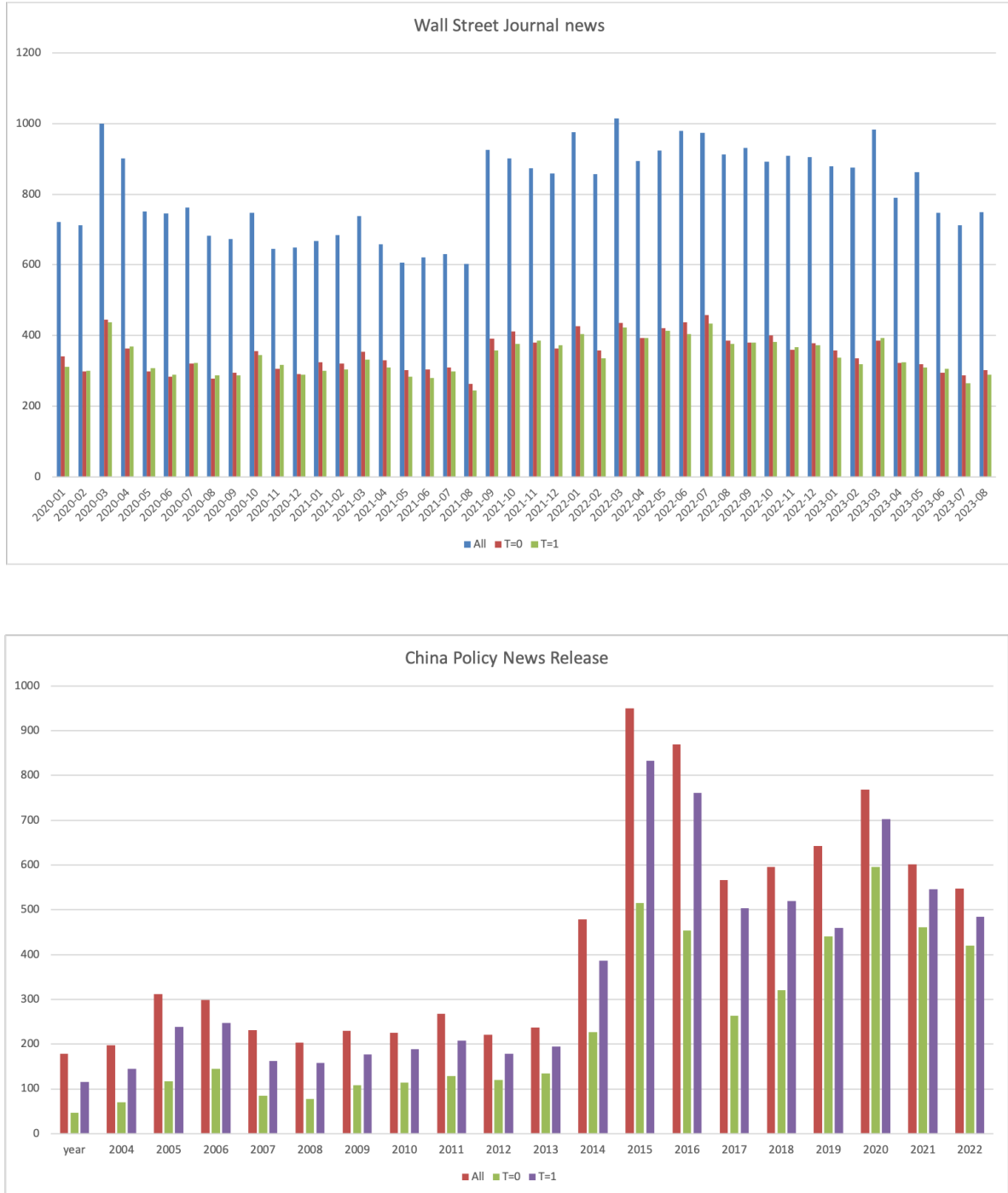
- Kelly, B., Xiu, D., 2023. Financial machine learning. SSRN Working Paper .
- Kelly, B. T., Malamud, S., Zhou, K., 2022. The virtue of complexity in return prediction .
- Loughran, T., McDonald, B., 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-ks. *The Journal of Finance* 66, 35–65.
- Sheng, J., Sun, Z., Yang, B., Zhang, A. L., 2024. Generative AI and asset management. SSRN Working Paper .

Figure 1: Words clouds of Wall Street Journal and China State-Council policy release



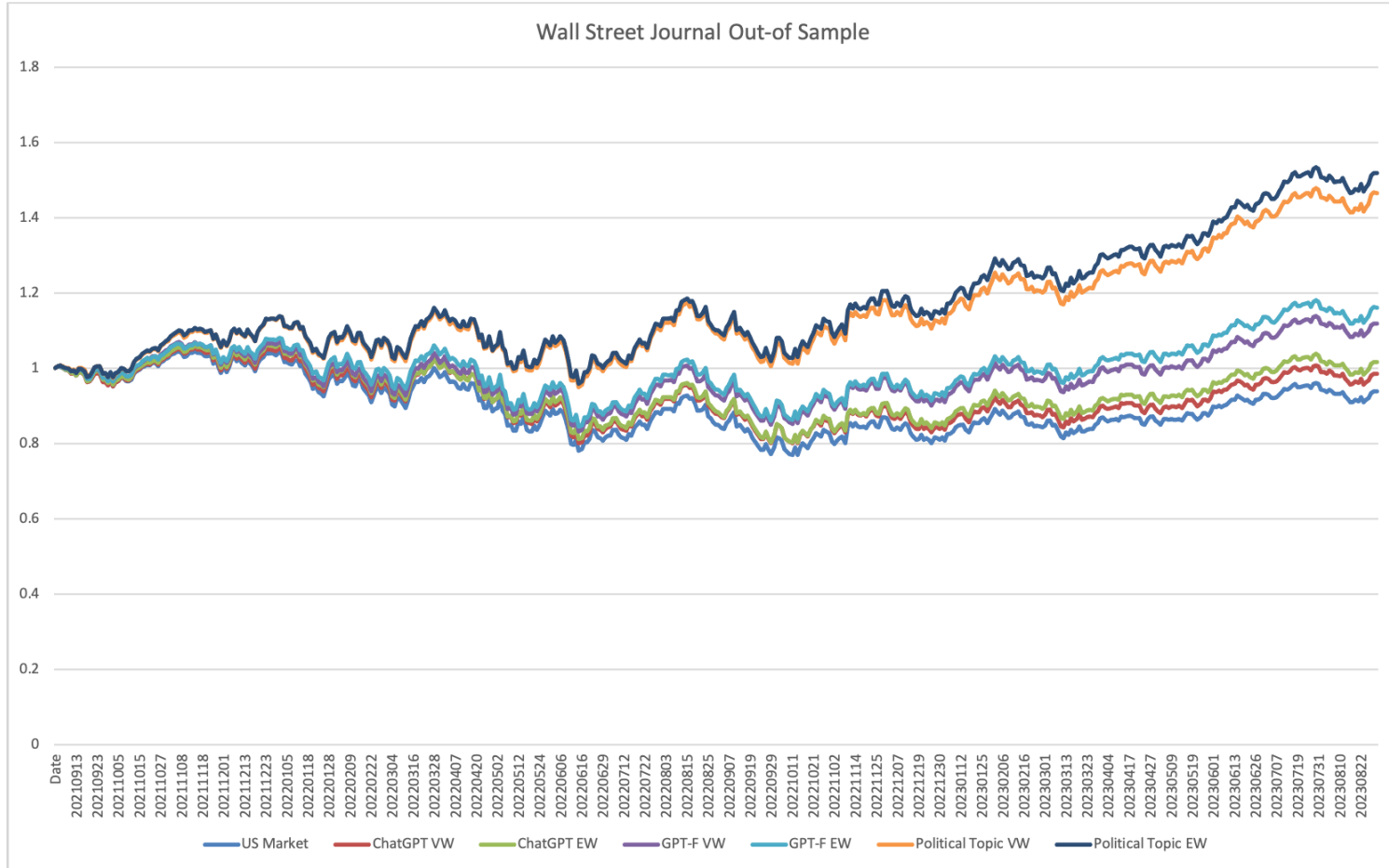
Note: Words clouds of Wall Street Journal and China State-Council policy release.

Figure 2: Number of policy release reports in the time series



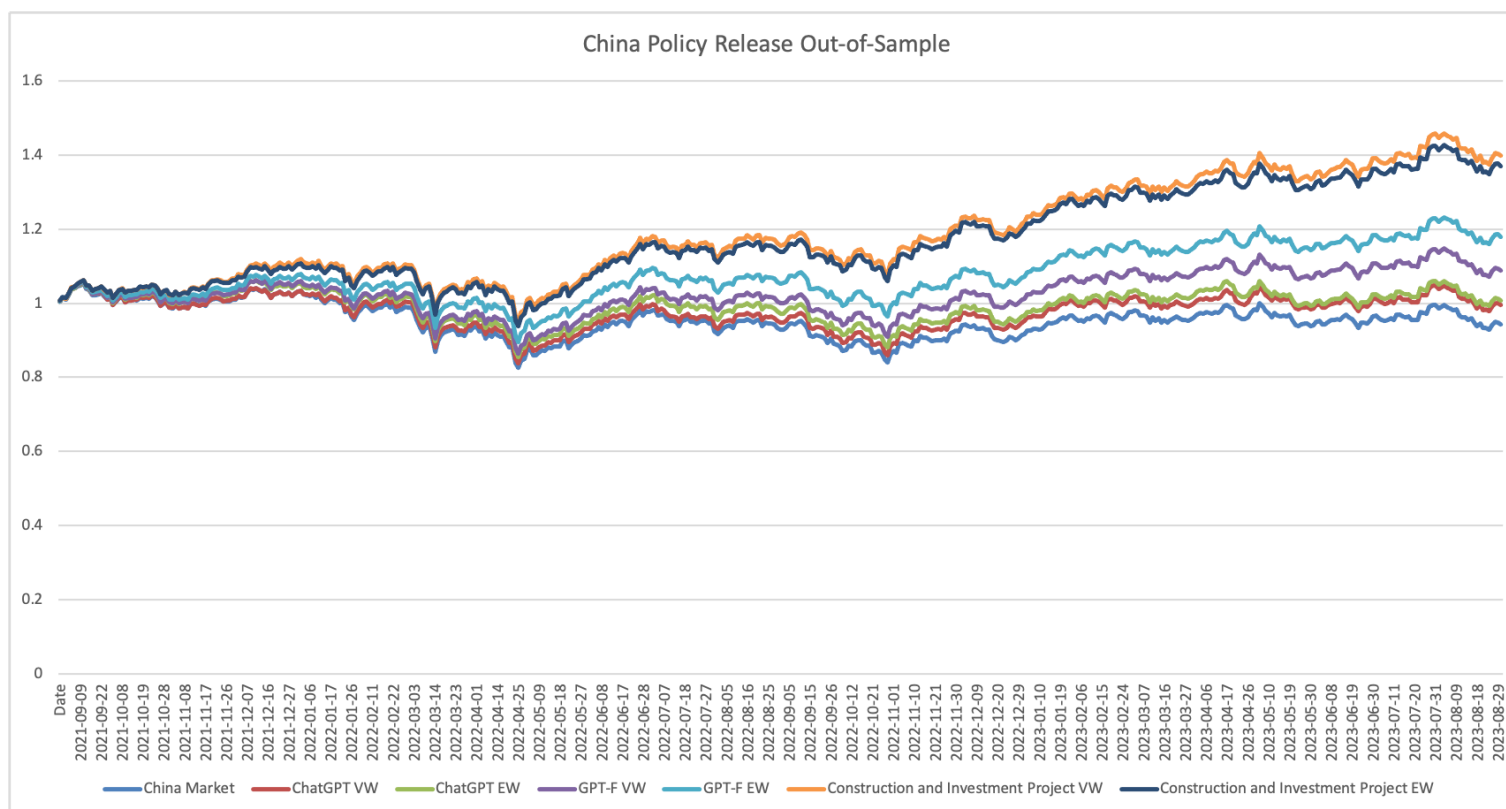
Note: This Figure shows the monthly number of Wall Street Journal news articles and annual number of Chinese policy release reports. The total number of news articles is represented by “All.” The variable  $T=0$  indicates the number of news articles identified by ChatGPT as relevant to stock recommendations when the temperature parameter is set to zero.  $T=1$  represents the average number of news articles identified as relevant to stock recommendations by ChatGPT when the temperature parameter is set to one.

Figure 3: Wall Street Journal out-of-sample portfolios



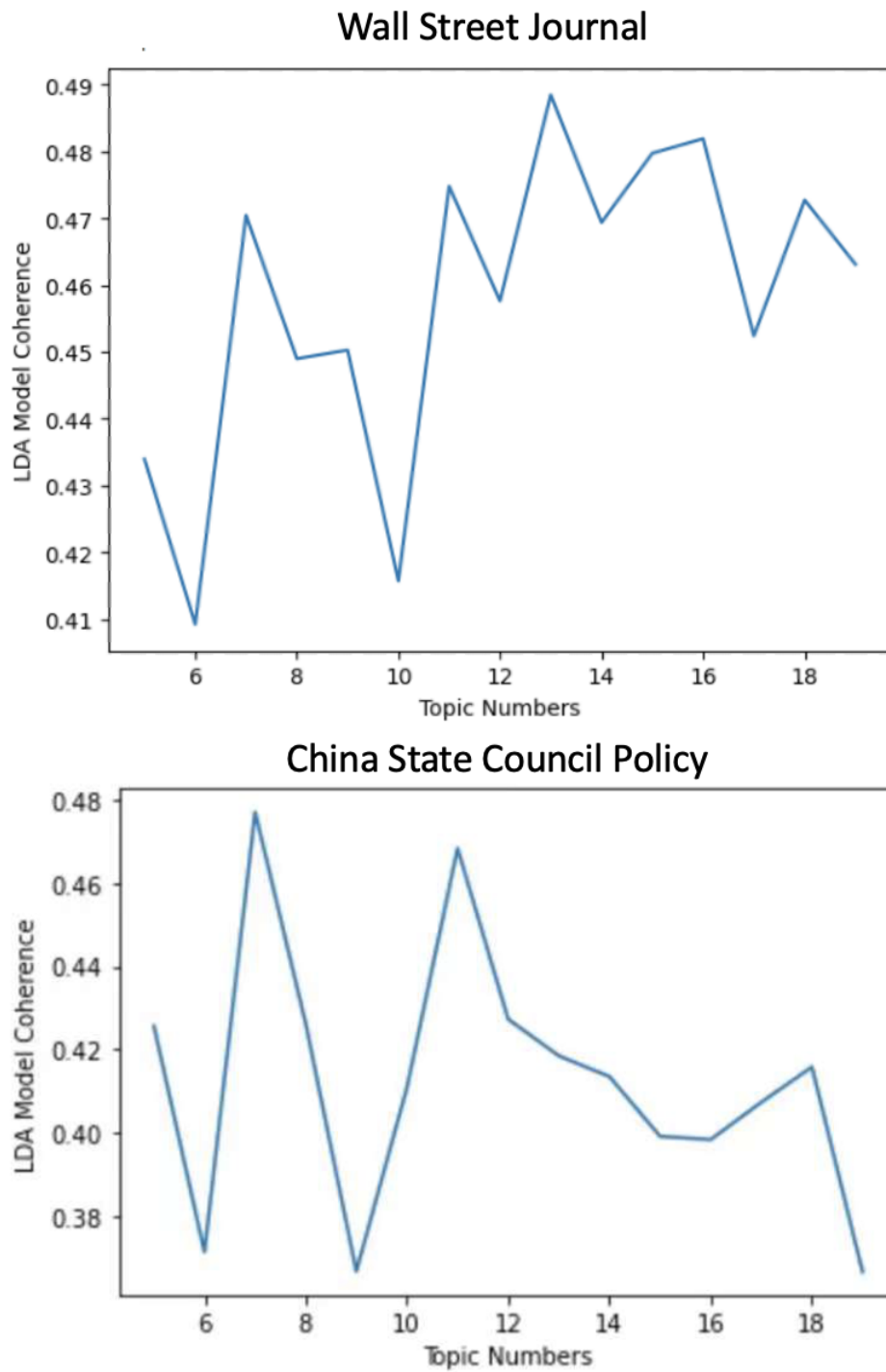
Note: Buy-and-hold returns of the U.S. market portfolio, as well as the ChatGPT value-weighted, equal weighted, fine-tuned value-weighted, fine-tuned equal-weighted, political topic only (topic 3) value-weighted, political topic only (topic 3) equal-weighted portfolios.

Figure 4: China Policy Release out-of-sample portfolios



Note: Note: Buy-and-hold returns of the China market portfolio, as well as the ChatGPT value-weighted, equal weighted, fine-tuned value-weighted, fine-tuned equal-weighted, construction and investment project (topic 1) value-weighted, construction and investment project (topic 1) equal-weighted portfolios.

Figure 5: LDA Coherence score



Note: LDA Coherence Score for Wall Street Journal and China policy release.

Table 1: Summary Statistics

**Notes.** Panel A shows the characteristics of stocks ChatGPT recommended in U.S. as well as the word distribution of the Wall Street Journal in the sample from Jan.2020 to Aug.2023. BETA is the Scholes-Williams-Dimson beta obtained from regressing daily firm return on daily current, lead, and lagged market returns over the previous month. SIZE is the log of total market value of equity at the end of the previous month in USD. BM is the Fama-French book-to-market ratio of book value of equity to market value of equity at the end of the previous calendar year. INV, is the change in total assets from the fiscal year ending in year t-2 to the fiscal year ending in t-1, divided by t-1 total assets. Operating profitability, PROFIT, is measured with accounting data for the fiscal year ending in year t-1 and is revenues minus cost of goods sold, minus selling, general, and administrative expenses, minus interest expense all divided by book equity. MOM is Jegadeesh-Titman momentum defined as the cumulative stock return over months t-12 to t-1. REV is Jegadeesh-Lehmann short-term reversal defined as the return on the stock over the previous month. Words are the number of words in each news. Panel B shows the characteristics of stocks ChatGPT recommended in China as well as the word distribution of China’s State Council policy news in the sample from Jan.2004 to Aug 2023.

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



Table 2: ChatGPT’s investment advisory and its validity in U.S.

**Notes.** Human rating of ChatGPT’s investment advisory based on the recommendation’s Relevance, Reasoning, and Investment Opportunity. The rating is based on three categories: logical, unclear, and unrelated. We present the result of financial analysts’ evaluation in Panel A and the result of finance-major students’ evaluation in Panel B. We also present results using both ChatGPT 3.5 and ChatGPT 4.0. T is the temperature parameter we use to adjust the creativity of ChatGPT’s recommendation. Domestic Equity is the subsample where the ChatGPT recommends NYSE or Nasdaq stocks based on the full text of the Wall Street Journal. Sector is the subsample where the ChatGPT recommends the sector (two-digit SIC code) to invest based on the content of the Wall Street Journal.

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%

Table 2: ChatGPT’s investment advisory and its validity in U.S. (continued)

**Notes.** Human rating of ChatGPT’s investment advisory based on the recommendation’s Relevance, Reasoning, and Investment Opportunity. The rating is based on three categories: logical, unclear, and unrelated. We present the result of financial analysts’ evaluation in Panel A and the result of finance-major students’ evaluation in Panel B. We also present results using both ChatGPT 3.5 and ChatGPT 4.0. T is the temperature parameter we use to adjust the creativity of ChatGPT’s recommendation. Domestic Equity is the subsample where the ChatGPT recommends NYSE or Nasdaq stocks based on the full text of the Wall Street Journal. Sector is the subsample where the ChatGPT recommends the sector (two-digit SIC code) to invest based on the content of the Wall Street Journal.

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%

Table 3: Alphas of Recommended Equity based on Wall Street Journal: In-sample tests

**Notes.** We construct ChatGPT calendar-time portfolios and the holding periods for each ChatGPT recommended stock are three months, one-month, one-week, and one day in different panels. When the temperature parameter equals zero ( $T=0$ ), we query ChatGPT 3.5 (4.0) and construct one unique ChatGPT calendar-time portfolio. When temperature equals one ( $T=1$ ), we query ChatGPT 3.5 (4.0) 25 times and construct 25 ChatGPT calendar-time portfolios. Daily alphas are presented for the CAPM model, Fama-French three-factor model (FF3), five-factor model (FF5), and five-factor model including momentum and reversal (FF5 MOM+REV) during the sample period January 1st, 2020-August 30th 2021. EW is the equal-weighted portfolio return. VW is the value-weighted portfolio return. Returns are scaled by 10,000 (report in basis points). T-statistics are reported in parentheses and the standard errors are clustered at the calendar day level.

Three-Month			CAPM	FF3	FF5	FF5 MOM+REV
ChatGPT 3.5	VW	T=0	2.193	1.890	1.738	1.638
			(0.348)	(0.221)	(0.244)	(0.261)
		T=1	7.387	6.145	5.847	4.629
			(1.234)	(1.014)	(0.958)	(0.793)
	EW	T=0	1.391	0.958	0.895	0.743
			(0.192)	(0.135)	(0.126)	(1.108)
		T=1	10.492	10.231	9.327	8.401
			(1.598)	(1.561)	(1.332)	(1.194)
ChatGPT 4.0	VW	T=0	0.854	0.743	0.532	0.478
			(0.134)	(0.110)	(0.083)	(0.061)
		T=1	0.780	0.675	0.578	0.532
			(0.136)	(0.113)	(0.074)	(0.067)
	EW	T=0	0.974	0.815	0.678	0.452
			(0.158)	(0.143)	(0.094)	(0.074)
		T=1	0.473	0.391	0.350	0.332
			(0.072)	(0.066)	(0.062)	(0.057)

One-Month			CAPM	FF3	FF5	FF5 MOM+REV
ChatGPT 3.5	VW	T=0	3.592	3.403	3.621	3.383
			(0.527)	(0.502)	(0.531)	(0.485)
		T=1	14.429**	14.104**	12.920*	13.853**
			(2.128)	(2.084)	(1.905)	(2.043)
	EW	T=0	4.618	4.503	4.415	4.516
			(0.614)	(0.593)	(0.583)	(0.605)
		T=1	16.251**	15.104**	14.791**	13.560**
			(2.421)	(2.254)	(2.207)	(2.023)
ChatGPT 4.0	VW	T=0	2.184	2.018	1.847	1.798
			(0.340)	(0.315)	(0.294)	(0.281)
		T=1	1.618	1.384	1.231	1.029
			(0.243)	(0.201)	(0.183)	(0.142)
	EW	T=0	1.729	1.492	1.320	1.283
			(0.257)	(0.231)	(0.231)	(0.197)
		T=1	1.537	1.321	1.298	1.252
			(0.229)	41(0.194)	(0.186)	(0.181)

Table 3: Alphas of Recommended Equity based on Wall Street Journal: In-sample tests (continued)

One-Week			CAPM	FF3	FF5	FF5 MOM+REV
ChatGPT 3.5	VW	T=0	0.476	0.362	0.290	0.253
			(0.070)	(0.063)	(0.048)	(0.038)
		T=1	1.482	1.312	1.138	0.930
			(0.239)	(0.204)	(0.194)	(0.115)
	EW	T=0	0.531	0.638	0.472	0.371
			(0.073)	(0.093)	(0.074)	(0.063)
		T=1	0.134	-0.135	-0.139	-0.314
			(0.021)	(-0.030)	(-0.021)	(-0.054)
ChatGPT 4.0	VW	T=0	1.194	0.983	0.611	0.437
			(0.178)	(0.104)	(0.085)	(0.062)
		T=1	0.934	0.731	0.673	0.510
			(0.154)	(0.112)	(0.095)	(0.080)
	EW	T=0	0.425	0.291	0.148	-0.152
			(0.081)	(0.040)	(0.026)	(-0.021)
		T=1	3.184	2.481	2.034	1.842
			(0.462)	(0.382)	(0.305)	(0.298)

One-Day			CAPM	FF3	FF5	FF5 MOM+REV
ChatGPT 3.5	VW	T=0	-0.184	-0.427	-0.138	-0.228
			(-0.031)	(-0.064)	(-0.028)	(-0.036)
		T=1	1.275	1.192	0.942	0.859
			(0.191)	(0.173)	(0.159)	(0.139)
	EW	T=0	0.821	0.738	0.571	0.412
			(0.119)	(0.112)	(0.084)	(0.071)
		T=1	-0.428	-0.398	-0.285	0.482
			(-0.084)	(-0.072)	(-0.051)	(0.073)
ChatGPT 4.0	VW	T=0	1.621	1.442	0.714	0.643
			(0.245)	(0.221)	(0.115)	(0.101)
		T=1	0.798	0.487	0.283	0.193
			(0.109)	(0.085)	(0.041)	(0.027)
	EW	T=0	1.984	1.526	1.413	1.034
			(0.296)	(0.218)	(0.204)	(0.187)
		T=1	-0.831	-0.415	-0.425	-0.312
			(-0.134)	(-0.074)	(-0.076)	(-0.063)

Table 4: In-Sample Tests of Equity Recommendations based on Wall Street Journal: Factor exposure

**Notes.** We construct ChatGPT calendar-time portfolios and the holding period for each ChatGPT recommended stocks is one calendar month. When the temperature parameter equals zero, we query ChatGPT 3.5 and construct one unique ChatGPT calendar-time portfolio. When temperature equals one, we query ChatGPT 3.5 25 times and construct 25 ChatGPT calendar-time portfolios. Daily alphas are presented for the Fama-French three-factor and five-factor models (also include the one-year momentum and one-month reversal factors) during the sample period January 1st, 2020-August 30th 2021. Weight denotes the weights method when calculating the daily return. EW is the equal-weighted portfolio return. VW is the value-weighted portfolio return. Returns are scaled by 10,000 (report in basis points). Portfolio FE is the portfolio fixed effect for each ChatGPT recommended portfolio. T-statistics are reported in parentheses and the standard errors are clustered at the calendar day level.

	FF3 Factors	FF5 Factors +MOM&REV	FF3 Factors	FF5 Factors +MOM&REV
	VW	VW	EW	EW
	(a) Temperature=0			
Alpha	3.403 (0.502)	3.383 (0.485)	4.503 (0.593)	4.516 (0.605)
MKT	0.840*** (4.292)	0.792*** (3.289)	1.280*** (3.422)	0.948*** (3.019)
SMB	-0.218*** (-9.291)	-0.428*** (-8.312)	0.194*** (-4.259)	0.293*** (-4.178)
HML	0.284** (2.901)	0.129 (1.210)	-0.179 (-1.411)	-0.275 (-1.265)
RMW		0.384 (1.209)		-0.284 (-0.112)
CMA		0.147 (1.213)		0.258** (2.101)
MOM		0.164 (1.073)		-0.095 (-0.156)
REV		0.251 (0.823)		-0.326 (-1.382)
R-squared	0.738	0.783	0.801	0.848

Table 4: In-Sample Tests of Equity Recommendations based on Wall Street Journal: Factor exposure (continued)

**Notes.** We construct ChatGPT calendar-time portfolios and the holding period for each ChatGPT recommended stocks is one calendar month. When the temperature parameter equals zero, we query ChatGPT 3.5 and construct one unique ChatGPT calendar-time portfolio. When temperature equals one, we query ChatGPT 3.5 25 times and construct 25 ChatGPT calendar-time portfolios. Daily alphas are presented for the Fama-French three-factor and five-factor models (also include the one-year momentum and one-month reversal factors) during the sample period January 1st, 2020-August 30th 2021. Weight denotes the weights method when calculating the daily return. EW is the equal-weighted portfolio return. VW is the value-weighted portfolio return. Returns are scaled by 10,000 (report in basis points). Portfolio FE is the portfolio fixed effect for each ChatGPT recommended portfolio. T-statistics are reported in parentheses and the standard errors are clustered at the calendar day level.

	FF3 Factors	FF5 Factors +MOM&REV	FF3 Factors	FF5 Factors +MOM&REV
	VW	VW	EW	EW
	(b) Temperature=1			
Alpha	14.104** (2.084)	13.853** (2.043)	15.104** (2.254)	13.560** (2.023)
MKT	0.859*** (4.276)	0.759*** (3.175)	1.095*** (5.085)	1.163*** (5.165)
SMB	-0.749*** (-3.391)	-0.683*** (-3.184)	0.326*** (4.193)	0.409*** (3.194)
HML	0.184 (1.482)	0.189 (1.501)	-0.296 (-1.184)	-0.291 (-1.325)
RMW		0.193 (0.284)		0.819 (1.039)
CMA		0.391*** (3.384)		0.104 (0.812)
MOM		0.165 (0.257)		-0.487 (-0.957)
REV		-0.376 (-0.847)		0.578* (1.859)
Portfolio FE	Y	Y	Y	Y
R-squared	0.742	0.751	0.831	0.823

Table 5: Alphas of Recommended Equity based on Wall Street Journal: Out-of-sample tests

**Notes.** This Table shows the Fama-French five factors (also includes Momentum and Reversal) alphas between September 2021 and February 2024. We separately show the alphas for the whole out-of-sample tests and separately for 13 topics (Topics are generated from the LDA analysis). Weight denotes the weights method when calculating the daily return. EW is the equal-weighted portfolio return. VW is the value-weighted portfolio return. Returns are scaled by 10,000 (report in basis points). T-statistics are reported in parentheses and the standard errors are clustered at the calendar day level.

	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)

Table 6: Fine-Tuning: Alphas of Recommended Equity based on Wall Street Journal in out-of-sample tests

**Notes.** This Table shows the Fama-French five factors (also includes Momentum and Reversal) alphas after fine-tuning between September 2021 and February 2024. We separately show the alphas for the whole out-of-sample tests and separately for 13 topics. Weight denotes the weights method when calculating the daily return. EW is the equal-weighted portfolio return. VW is the value-weighted portfolio return. Returns are scaled by 10,000 (report in basis points). T-statistics are reported in parentheses and the standard errors are clustered at the calendar day level.

	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)



Table 7: Benchmark Wall Street Journal analysis with traditional textual analysis

**Notes.** This Table shows the Fama-French five factors (also includes Momentum and Reversal) alphas after fine-tuning between September 2021 and February 2024. We separately show the alphas for the whole out-of-sample tests and separately for different topics using traditional textual analysis. Weight denotes the weights method when calculating the daily return. EW is the equal-weighted portfolio return. VW is the value-weighted portfolio return. Returns are scaled by 10,000 (report in basis points). T-statistics are reported in parentheses and the standard errors are clustered at the calendar day level.

	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)

Table 8: ChatGPT’s investment advisory and its validity in China

**Notes.** Human Rating of ChatGPT’s investment advisory based on the recommendation’s Relevance, Reasoning, and Investment Opportunity. The rating is based on three categories: logical, unclear, and unrelated. We present the result of financial analysts’ evaluation in Panel A and the result of finance-major students’ evaluation in Panel B. We also present results using both ChatGPT 3.5 and ChatGPT 4.0. T is the temperature parameter we use to adjust the creativity of ChatGPT’s recommendation. Domestic Equity is the subsample where the ChatGPT recommends Chinese A-share stocks based on the full text of the policy news release. Sector is the subsample where the ChatGPT recommends the sector (CSRC 20 industry classification) to invest based on the content of the policy news release.

Panel A: Financial Analyst			
ChatGPT 3.5			
	Domestic Equity (T=0)	Domestic Equity (T=1)	Domestic Equity (T=1.5)
Logical	85%	89%	85%
Unclear	13%	2%	11%
Unrelated	2%	9%	4%
	Sector (T=0)	Sector (T=1)	Sector (T=1.5)
Logical	97%	95%	94%
Unclear	1%	3%	5%
Unrelated	2%	2%	1%
ChatGPT 4.0			
	Domestic Equity (T=0)	Domestic Equity (T=1)	Domestic Equity (T=1.5)
Logical	79%	83%	75%
Unclear	17%	12%	17%
Unrelated	4%	5%	8%
	Sector (T=0)	Sector (T=1)	Sector (T=1.5)
Logical	89%	90%	85%
Unclear	3%	4%	2%
Unrelated	8%	6%	13%

Table 7: ChatGPT’s investment advisory and its validity in China (continued)

**Notes.** Human Rating of ChatGPT’s investment advisory based on the recommendation’s Relevance, Reasoning, and Investment Opportunity. The rating is based on three categories: logical, unclear, and unrelated. We present the result of financial analysts’ evaluation in Panel A and the result of finance-major students’ evaluation in Panel B. We also present results using both ChatGPT 3.5 and ChatGPT 4.0. T is the temperature parameter we use to adjust the creativity of ChatGPT’s recommendation. Domestic Equity is the subsample where the ChatGPT recommends Chinese A-share stocks based on the full text of the policy news release. Sector is the subsample where the ChatGPT recommends the sector (CSRC 20 industry classification) to invest based on the content of the policy news release.

Panel B: Finance Students			
ChatGPT 3.5			
	Domestic Equity (T=0)	Domestic Equity (T=1)	Domestic Equity (T=1.5)
Logical	82%	84%	80%
Unclear	16%	10%	13%
Unrelated	2%	6%	7%
	Sector (T=0)	Sector (T=1)	Sector (T=1.5)
Logical	94%	90%	85%
Unclear	1%	4%	2%
Unrelated	5%	6%	13%
ChatGPT 4.0			
	Domestic Equity (T=0)	Domestic Equity (T=1)	Domestic Equity (T=1.5)
Logical	79%	83%	75%
Unclear	17%	12%	17%
Unrelated	4%	5%	8%
	Sector (T=0)	Sector (T=1)	Sector (T=1.5)
Logical	89%	90%	85%
Unclear	3%	4%	2%
Unrelated	8%	6%	13%

Table 8: Alphas of Equity Recommendation based on China State Council Policy Release: In-sample tests

**Notes.** We construct ChatGPT calendar-time portfolios and the holding periods for each ChatGPT recommended stock are one year, six months, three months, one-month, one-week, and one day in different panels. When the temperature parameter equals zero ( $T=0$ ), we query ChatGPT 3.5 (4.0) and construct one unique ChatGPT calendar-time portfolio. When temperature equals one ( $T=1$ ), we query ChatGPT 3.5 (4.0) 25 times and construct 25 ChatGPT calendar-time portfolios. Daily alphas are presented for the Chinese five-factor model including momentum and reversal between the sample period January 1st, 2014-August 30th 2021. EW is the equal-weighted portfolio return. VW is the value-weighted portfolio return. Returns are scaled by 10,000 (report in basis points). T-statistics are reported in parentheses and the standard errors are clustered at the calendar day level.

China FF5 MOM+REV				One Year	Six Month	Three Month	One Month	One Week	One Day
GPT 3.5	VW	$T=0$		-0.284	0.482	1.379	2.731	1.249	-0.382
				(-0.128)	(0.246)	(0.682)	(1.365)	(0.582)	(-0.193)
		$T=1$		1.482	0.528	1.632	6.982***	2.481	-0.847
				(0.621)	(0.230)	(0.738)	(3.187)	(1.185)	(-0.409)
	EW	$T=0$		0.149	-0.174	0.810	2.980	2.485	1.392
				(0.029)	(-0.086)	(0.384)	(1.426)	(1.220)	(0.685)
		$T=1$		0.385	0.637	2.492	8.511***	1.374	0.738
				(0.164)	(0.336)	(1.290)	(4.001)	(0.688)	(0.281)
GPT 4.0	VW	$T=0$		-0.189	-0.371	0.359	1.385	0.849	-0.164
				(-0.083)	(-0.164)	(0.170)	(0.748)	(0.381)	(-0.081)
		$T=1$		1.394	0.642	0.291	0.819	-0.284	1.108
				(0.638)	(0.384)	(0.150)	(0.416)	(-0.149)	(0.547)
	EW	$T=0$		0.583	0.184	0.621	1.472	0.391	0.164
				(0.294)	(0.097)	(0.329)	(0.782)	(0.192)	(0.092)
		$T=1$		0.602	0.831	1.390	2.174	1.105	-1.483
				(0.282)	(0.365)	(0.501)	(1.019)	(0.661)	(-0.547)

Table 9: In-Sample Tests of Equity Recommendations based on China State Council Policy Release: Factor exposure

**Notes.** We construct ChatGPT calendar-time portfolios and the holding period for each ChatGPT recommended stocks is one calendar month. When the temperature parameter equals zero, we query ChatGPT 3.5 and construct one unique ChatGPT calendar-time portfolio. When temperature equals one, we query ChatGPT 3.5 25 times and construct 25 ChatGPT calendar-time portfolios. Daily alphas are presented for the Fama-French three-factor and five-factor models (also include the one-year momentum and one-month reversal factors) during the sample period January 1st, 2004-August 30th 2021. Weight denotes the weights method when calculating the daily return. EW is the equal-weighted portfolio return. VW is the value-weighted portfolio return. Returns are scaled by 10,000 (report in basis points). Portfolio FE is the portfolio fixed effect for each ChatGPT recommended portfolio. T-statistics are reported in parentheses and the standard errors are clustered at the calendar day level.

	FF3 Factors	FF5 Factors +MOM&REV	FF3 Factors	FF5 Factors +MOM&REV
	VW	VW	EW	EW
	(a) Temperature=0			
Alpha	2.501 (1.280)	2.731 (1.365)	3.092 (1.530)	2.980 (1.426)
MKT	0.829*** (8.391)	0.835*** (7.816)	0.843*** (8.281)	0.856*** (9.902)
SMB	-0.247*** (-13.014)	-0.275*** (-8.306)	0.239*** (13.519)	0.188*** (5.867)
HML	0.004 (0.149)	-0.141*** (-3.113)	-0.047* (-1.772)	-0.039*** (-0.889)
RMW		0.476*** (8.866)		-0.115** (-2.207)
CMA		0.074 (1.144)		-0.017 (-0.268)
MOM		-0.582 (-0.839)		-0.281 (-0.374)
REV		0.832 (1.194)		0.840 (0.953)
R-squared	0.621	0.641	0.686	0.691

Table 9: In-Sample Tests of Equity Recommendations based on China State Council Policy Release: Factor exposure(continued)

	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

Table 10: Alphas of Recommended Equity based on China State Council Policy Release in out-of-sample tests

**Notes.** This Table shows the Fama-French five factors (also includes Momentum and Reversal) alphas. We separately show the alphas for the whole out-of-sample tests and separately for 7 topics between September 2021 and February 2024. Weight denotes the weights method when calculating the daily return. EW is the equal-weighted portfolio return. VW is the value-weighted portfolio return. Returns are scaled by 10,000 (report in basis points). T-statistics are reported in parentheses and the standard errors are clustered at the calendar day level.

	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)

Table 11: Fine-Tuning: Alphas of Recommended Equity based on China State Council Policy Release in out-of-sample tests

**Notes.** This Table shows the Fama-French five factors (also includes Momentum and Reversal) alphas. We separately show the alphas for the whole out-of-sample tests and separately for 7 topics between September 2021 and February 2024. Weight denotes the weights method when calculating the daily return. EW is the equal-weighted portfolio return. VW is the value-weighted portfolio return. Returns are scaled by 10,000 (report in basis points). T-statistics are reported in parentheses and the standard errors are clustered at the calendar day level.

	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)



Table 12: Benchmark China State Council Policy Release with traditional textual analysis  
**Notes.** This Table shows the Fama-French five factors (also includes Momentum and Reversal) alphas after fine-tuning between September 2021 and February 2024. We separately show the alphas for the whole out-of-sample tests and separately for different topics using traditional textual analysis. Weight denotes the weights method when calculating the daily return. EW is the equal-weighted portfolio return. VW is the value-weighted portfolio return. Returns are scaled by 10,000 (report in basis points). T-statistics are reported in parentheses and the standard errors are clustered at the calendar day level.

	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)

## A Online Appendix

### A.1 Sample Wall Street Journal news and ChatGPT response

#### Biden to Withhold Tariffs on Solar Imports

The White House is preparing to announce on Monday that it won't impose any new tariffs on solar imports for two years, in a move that is aimed at getting stalled solar-power projects on track, according to people familiar with the decision. The decision would be a win for U.S. solar developers and utilities—which are highly dependent on imported solar panels—and a loss for manufacturers trying to build up a domestic solar supply chain. The decision would come in the wake of the Commerce Department's decision to investigate whether Chinese solar producers are illegally circumventing solar tariffs by routing operations through four countries in Southeast Asia—Cambodia, Thailand, Vietnam and Malaysia. If the investigation concludes that tariffs were being circumvented, it could lead to levies being imposed retroactively on imports from those countries—a scenario that created a chilling effect on new solar development, according to several companies and lawmakers. Companies including NextEra Energy Inc. and Xcel Energy Inc. have said the solar-panel probe could significantly delay solar projects, and California Gov. Gavin Newsom told the Commerce Department it would delay numerous new solar and battery storage projects expected to come online through 2024. By announcing the two-year pause in any new tariffs, the White House is trying to give more certainty to developers and utilities that have complained its trade policies are causing the slowdown in adding new solar capacity nationwide, according to a senior administration official. At the same time, the Commerce Department investigation would continue, the official said. The probe was launched after California-based Auxin Solar Inc., a small maker of solar panels, accused Chinese companies of circumventing tariffs by routing operations through Cambodia, Thailand, Vietnam and Malaysia that then export materials to the U.S.

The company said that Chinese solar producers have sought to undercut U.S. manufac-

turers by shifting production to countries not covered by tariffs. The International Energy Agency said last month new solar development in the U.S. is on pace to decline 6.8% this year—one of the few places in the world it is slowing down—because of uncertainty over tax and trade policy. In addition to the two-year pause on new solar tariffs, administration officials say the White House will unveil plans to increase domestic solar manufacturing capacity to 22.5 gigawatts by 2024, or three times the manufacturing capacity when President Biden took office. That would enable millions of homes to switch to solar energy, the officials said. Abigail Ross Hopper, president and chief executive of the Solar Energy Industries Association, cheered the expected action from Mr. Biden to pause new tariffs. These steps “protect existing solar jobs, will lead to increased employment in the solar industry and foster a robust solar manufacturing base here at home,” she said. The levies on imported solar cells and modules apply to products made in China. Known as antidumping and countervailing duties, they were first imposed by the Obama administration in 2012. The investigation could extend those tariffs to cells and modules completed in Cambodia, Malaysia or Vietnam, using parts and components from China. Those countries now produce roughly 80% of U.S. solar imports. The tariff issue has split the solar industry, pitting solar installers who buy imported panels against U.S. manufacturers who want protection from what they see as cut-rate Chinese competitors. Mr. Biden has been caught between these competing interests—on the one hand hoping to boost U.S. manufacturing and on the other hoping to accelerate the transition to clean renewable energy. Mr. Biden must also decide what to do with the tariffs on tens of billions of dollars of Chinese imports inherited from the Trump administration amid escalating inflation and the need to confront China. Many businesses, economists and even some in his own administration such as Treasury Secretary Janet Yellen, have pressed the president to eliminate some of the tariffs as a way to ease inflation pressure. Mr. Biden himself has said tackling inflation is his “top economic priority.” But Mr. Biden also faces growing pressure to stand tough against China’s economic practices as the war in Ukraine fuels worries about authoritarian governments.

### **ChatGPT Recommend stocks:**

Recommended stocks1: First Solar, FSLR, NASDAQ

Recommended stocks2: SunPower Corporation, SPWR, NASDAQ

Recommended stocks3: Enphase Energy Inc., ENPH, NASDAQ

Recommended stocks4: SolarEdge Technologies, SEDG, NASDAQ

Recommended stocks5: Sunrun Inc., RUN, NASDAQ

## **A.2 Prompt for ChatGPT in Chinese subsample**

Equity Recommendation Instruction: Suppose you are a senior financial analyst. Please carefully read the following policy announcement. According to the content of the announcement, if you have recommended stocks, write "YES" and list 5 A-share (China domestic listed) stock names and the stock codes that you recommend. If you just answered "NO", please briefly explain the reason.

Domestic Sector Recommendation Instruction based on Policy News: Suppose you are a senior financial analyst. Please carefully read the following policy news. Based on the content of the policy news and referring to the Chinese Security Regulatory Committee (CSRC) 2012 industry classification standard, if you have recommended sectors, write "YES" and list the first-level category code of the recommended sector. If you have just answered "NO", please briefly explain the reason.

The Appendix A.3 shows an example of the stock that ChatGPT offers. On May 14th 2023, The National Development and Reform Commission National Energy Administration released a policy titled: National Development and Reform Commission and National Energy Administration on Accelerating the Construction of Charging Infrastructure and Better Supporting New Energy Vehicles in Rural Areas and Rural Revitalization Implementation Opinions. This policy talks about the new energy reform in China. We conduct two rounds of queries when the temperature parameter is set to one for ChatGPT 3.5 and the ChatGPT recommends 10 stocks in total.

### A.3 Sample China State Council Policy and ChatGPT response

We execute our prompt programming at the following site:

[https://www.ndrc.gov.cn/xxgk/zcfb/tz/202305/t20230517\\_1355814.html](https://www.ndrc.gov.cn/xxgk/zcfb/tz/202305/t20230517_1355814.html)

Here is the translated text of the policy: National Development and Reform Commission and National Energy Administration on Accelerating the Construction of Charging Infrastructure and Better Supporting New Energy Vehicles in Rural Areas and Rural Revitalization Implementation Opinions

General Office [2023] No. 545

To the People's Governments of all provinces, autonomous regions, and municipalities directly under the Central Government, the Xinjiang Production and Construction Corps, State Grid Corporation of China, and China Southern Power Grid Co., Ltd.:

China has built the largest charging infrastructure system in the world, covering the largest area and serving the most comprehensive range of vehicles, providing strong support for the rapid development of new energy vehicles. However, there are still problems in rural areas, such as insufficient construction of public charging infrastructure, difficulty in installing and sharing charging facilities in residential communities, and prominent contradictions between supply and demand during peak periods, which restrict the release of consumption potential of new energy vehicles in rural areas. Moderately ahead-of-schedule construction of charging infrastructure and optimization of the purchase and use environment for new energy vehicles are of great significance for promoting new energy vehicles in rural areas, guiding green travel for rural residents, and promoting the comprehensive revitalization of villages. In order to do a good job in related work, with the approval of the State Council, the following implementation opinions are formulated.

I. Innovate the construction, operation, and maintenance mode of charging infrastructure in rural areas (1) Strengthen the layout and construction of public charging infrastructure. Support local governments to carry out county and township public charging network planning according to local conditions, and coordinate with land and space planning and distri-

bution network planning to accelerate the realization of "full coverage of charging stations in suitable areas for new energy vehicles" and "full coverage of charging piles in townships". Rationally promote the construction of centralized public charging stations, prioritize the allocation of public charging facilities in county and township enterprises and institutions, commercial buildings, transportation hubs (stations), highway service areas (stations), and extend to resettlement areas for relocated residents, key villages for rural tourism, and accelerate the construction of charging piles in places along highways and qualified gas stations.

(2) Promote the sharing of community charging infrastructure construction. Accelerate the construction of charging facilities in existing residential communities in rural areas, carry out charging facility construction condition transformation according to local conditions, and allocate a certain proportion of public charging parking spaces in residential communities with installation conditions. Implement the requirements for the construction of charging infrastructure in newly-built residential communities, promote the construction of charging facilities or reserve installation conditions for fixed parking spaces to meet the needs of direct metering and power connection. Implement the management responsibilities of grassroots management institutions such as street offices, strengthen guidance and supervision of residential community management units, and establish a "one-stop" coordination promotion and complaint resolution mechanism. Residential community management units should actively assist users in installing charging facilities and explore mechanisms for cooperation with charging facility operators. Guide communities to promote sharing models such as "nearby parking space sharing", "community time-sharing sharing", and "multi-vehicle one-pile".

(3) Increase support for the construction and operation of charging networks. Encourage qualified localities to introduce special support policies for the construction and operation of public charging infrastructure in rural areas. Use tools such as local government special bonds to support the construction of charging and battery-swapping infrastructure that meets the requirements for highway service areas (stations), public bus and electric vehicle

stations, and passenger vehicle stations on high-speed highways and ordinary national and provincial trunk roads. Coordinate the development of rural charging networks and power transmission and distribution networks, increase support for land guarantees and other efforts, carry out supporting power grid construction and transformation, and enhance the support and guarantee capabilities of rural power grids. By 2030, centralized charging and battery swapping facilities implementing two-part electricity prices will be exempted from demand (capacity) electricity charges, and restrictions on the investment efficiency of related distribution network construction by power grid enterprises will be relaxed, with full inclusion in the recovery of transmission and distribution prices.

(4) Promote new models such as smart and orderly charging. Improve the intelligence level of newly built charging infrastructure, incorporate smart and orderly charging into the scope of charging infrastructure and new energy vehicle product functions, encourage the charging piles equipped with new energy vehicles to have orderly charging functions, and accelerate the formation of unified industry standards. Encourage the research on key technologies such as bidirectional interaction between electric vehicles and power grids (V2G) and coordinated control of photovoltaic energy storage and charging, and explore the construction of integrated charging infrastructure that provides photovoltaic power generation, energy storage, and charging in rural areas with low charging pile utilization. Implement peak-valley time-of-use electricity pricing policies to encourage users to charge during off-peak periods.

(5) Enhance the operation and maintenance service experience of charging infrastructure. In conjunction with the charging facility environment, power grid infrastructure conditions, and operation and maintenance requirements in rural areas, carry out the revision and typical design of charging facility construction standards. Improve the operation and maintenance system of charging facilities, enhance facility availability and fault handling capabilities, and promote interconnection and interoperability among public charging and battery swapping network operators' platforms. Encourage parking lots and charging facility operators to

innovate technologies and management measures, guide the separate parking of fuel vehicles and new energy vehicles, and maintain a good charging order. Use technical means to predict the concentrated periods and areas of charging demand in advance, and ensure service guarantee.

2. Support the purchase and use of new energy vehicles in rural areas

(6) Enrich the supply of new energy vehicles. Encourage new energy vehicle enterprises to optimize their configurations through differentiated strategies according to the characteristics of rural consumers, and develop more economical and practical models, especially new energy light cargo vans, mini trucks, and light trucks. Improve the evaluation system for used new energy vehicles, strengthen inspection and refurbishment of used new energy vehicles, and encourage enterprises to provide high-quality used new energy vehicles for the rural market.

(7) Accelerate the promotion and application in public areas. Speed up the promotion and application of new energy vehicles in county and township party and government agencies, schools, hospitals, and other units, and increase the proportion of new energy vehicles used in official vehicles according to local conditions, playing a leading and demonstrative role. Encourage qualified localities to increase support for the application of new energy vehicles in public transportation, road passenger transport, taxis, law enforcement, sanitation, logistics distribution, and other fields.

(8) Provide diversified purchase support policies. Encourage qualified localities to provide support such as consumption coupons for rural household-registered residents who purchase new energy vehicles within their county of household registration. Encourage relevant automobile companies and qualified localities to provide trade-in rewards for the replacement of low-speed electric vehicles with new energy vehicles. Encourage local governments to strengthen collaboration with enterprises, carry out activities such as offering free charging coupons with car purchases. Increase auto consumption credit support in rural areas, encourage financial institutions to reasonably determine down payment ratios, loan interest rates, and repayment terms under the premise of legal compliance and controllable risks.



(9) Increase publicity and guidance efforts. Actively promote through news reports, expert comments, and internet new media, support local governments and industry institutions to organize new energy vehicle manufacturers to carry out joint brand exhibitions, test rides, and test drives, encourage new energy vehicle companies to cooperate with upstream battery companies in the industry chain to provide free “battery check-ups” for rural car purchases within three years, and enhance consumers’ acceptance of new energy vehicles.

3. Strengthen publicity, service management, and administration of new energy vehicles in rural areas

(10) Strengthen the sales and service network. Encourage new energy vehicle companies to sink their sales network, guide vehicle companies and third-party service companies to accelerate the construction of joint business outlets, establish supporting after-sales service systems, regularly carry out maintenance and after-sales service activities in rural areas, provide emergency rescue and other services, and alleviate purchase and usage concerns. Encourage vocational colleges to train new energy vehicle maintenance technicians for rural areas, provide relevant vocational education in automotive maintenance and charging pile maintenance, and effectively connect local employment promotion with supporting new energy vehicle consumption.

(11) Strengthen safety supervision. Improve the safety supervision system for new energy vehicles, use various means to adapt to local conditions, enhance the safety level of new energy vehicles and batteries, strictly manage charging facilities in rural areas, guide charging facility operators to access government charging facility supervision platforms, strictly comply with the safety conditions of supporting power supply and centralized charging venues, ensure compliance with relevant laws and regulations, national standards or industry standards, strengthen safety training for management personnel, and periodically conduct hazard investigations on existing charging piles. Guide rural residents to install and use independent charging piles and reasonably equip them with leakage protectors and grounding equipment to enhance the level of electricity safety.

All regions and relevant departments should effectively strengthen organizational leadership, clarify responsibilities and division of labor, take the initiative to act, and promote the implementation and effectiveness of related policies and measures as soon as possible. Improve purchase and use policies, further strengthen the charging infrastructure network, and ensure “construction, management, and sustainability.” Create a favorable environment for the promotion and use of new energy vehicles in rural areas, and better meet the production and living needs of the masses.

National Development and Reform Commission National Energy Administration May 14, 2023

ChatGPT Response 1st round:

1. Longyuan Technology (300105): The company focuses on new energy fields such as wind power generation and solar power generation.
2. East Sunrise (300118): The company is mainly engaged in the research and development, production, and sales of solar photovoltaic products.
3. Guodian NARI (600406): The company is involved in energy internet-related fields such as smart grid and power automation equipment.
4. TBEA (600089): The company is mainly engaged in the research and development, production, and sales of power transmission and distribution equipment.
5. Zhongcai Energy Conservation (000609): The company is mainly engaged in technology research and development and equipment manufacturing in the fields of energy conservation and environmental protection.

ChatGPT Response 2nd round:

6. Guodian Power (600795): The company is mainly engaged in power generation, supply, and related services.
7. Tongwei Group (600438): The company’s business includes photovoltaic manufactur-

ing, new energy power generation, and more.

8. Huaneng International (600011): The company is mainly engaged in various energy power generation businesses, including hydropower, thermal power, and wind power.

9. Guotou Power (600886): The company's business covers various energy power generation sectors, including thermal power, hydropower, and wind power.

10. Sichuang Electronics (600990): The company is mainly engaged in technology research and development and product manufacturing in the fields of smart grid and energy informatization.

These recommended stocks cover various sectors of the energy industry, including renewable energy, power grid, energy-saving technology, and power generation. Investing in these stocks could provide diversified exposure to the energy market in China, which is in line with the policy's focus on promoting clean and efficient energy utilization.

#### Sector Recommendation:

Based on the news, which focuses on accelerating the construction of charging infrastructure, I would recommend the following sectors in China's domestic A-share market:

1. (C36) - Automobile Manufacturing
2. (C38) - Electrical Machinery and Equipment Manufacturing
3. (C39) - Computer, Communication, and Other Electronic Equipment Manufacturing
4. (C40) - Instrument and Meter Manufacturing

These sectors are related to electric vehicles, charging infrastructure, communication, and electronic equipment, which are essential components of the charging infrastructure mentioned in the news.