Using Generative AI to predict the weather impact on future stock returns*

Yi Zhou†

June 12, 2024

Abstract

This study explores the use of Generative AI, specifically OpenAI's ChatGPT, for forecasting the impacts of severe weather events on stock returns. Employing prompts that assess textual weather descriptions, ChatGPT, a powerful generative AI large language model (LLM), provides predictions incorporated into econometric models. Results show that when ChatGPT forecasts negative stock impacts from storms, larger, more profitable firms with lower leverage and higher liquidity experience lower subsequent returns, suggesting investor underreaction to weather risk. ChatGPT's predictive abilities are stronger during favorable economic conditions like uptrends, low volatility, and robust employment growth, implying investor underreaction amid bullish sentiment.

JEL classification: G02, G11, G12, G13, G14.
Keywords: Weather risk, stock returns, Generative AI, ChatGPT, large language model (LLM)

*I deeply appreciate Kelvin Kay Droegemeier for his tremendous encouragement and support. I thank Vikash Rungta at Meta for numerous insightful discussions about Generative AI. I also thank Charlie Flanagan (ex-Googler) at Balyasny Asset Management for sharing his expertise on Large Language Modeling (LLM). All remaining errors are mine.

†Department of Finance, College of Business, San Francisco State University, 1600 Holloway Avenue, San Francisco, CA 94132, USA, Email: yizhou88@sfsu.edu, Office: 415-338-2661, Fax: 415-338-0596.

Using Generative AI to predict the weather impact on future stock returns

Abstract

This study explores the use of Generative AI, specifically OpenAI's ChatGPT, for forecasting the impacts of severe weather events on stock returns. Employing prompts that assess textual weather descriptions, ChatGPT, a powerful generative AI large language model (LLM), provides predictions incorporated into econometric models. Results show that when ChatGPT forecasts negative stock impacts from storms, larger, more profitable firms with lower leverage and higher liquidity experience lower subsequent returns, suggesting investor underreaction to weather risk. ChatGPT's predictive abilities are stronger during favorable economic conditions like uptrends, low volatility, and robust employment growth, implying investor underreaction amid bullish sentiment.

JEL classification: G02, G11, G12, G13, G14.

Keywords: Weather risk, stock returns, Generative AI, ChatGPT, large language model (LLM)

1 Introduction

A burgeoning body of literature has emerged on Generative AI, particularly focusing on ChatGPT and its applications in finance, as exemplified by Chen et al. (2023), Chen et al. (2024), Fedyk et al. (2024), Kim et al. (2024b), Lopez-Lira and Tang (2023), Tan et al. (2024), and Li et al. (2024a), among others. There has also been a long stream of literature on climate finance, exemplified by Sautner et al. (2023a), Sautner et al. (2023b), Ilhan et al. (2023), Pankratz and Schiller (2024), Li et al. (2024b), Addoum et al. (2023), and Engle et al. (2020), among others. This paper seeks to bridge the gap between Generative AI applications in finance and climate risk literature, leveraging Generative AI to predict how weather patterns influence future stock returns.

Generative Artificial Intelligence (Generative AI or Gen AI) is a broad category of artificial intelligence focused on generating new content–such as text, images, music, or code–based on the data it has been trained on. It encompasses various techniques and models used to create original, human-like content. Large Language Models (LLMs) are a specific type of generative AI designed to process and generate human-like text. LLMs are advanced machine learning models trained on vast amounts of diverse data to perform various natural language processing tasks, including text generation, translation, summarization, and question answering. These models can understand and generate text in a manner that mimics human communication. A prominent example of an LLM is ChatGPT, a conversational AI model developed by OpenAI. Built on the Transformer architecture, ChatGPT (Chat Generative Pre-trained Transformer) is designed to generate human-like text responses based on the input it receives, enabling it to engage in dialogue, answer questions, and assist with various tasks. The "GPT" in ChatGPT stands for "Generative Pre-trained Transformer," indicating that it uses generative AI techniques and is pre-trained on a vast amount of text data before being fine-tuned for specific tasks. In summary, generative AI is the overarching concept, LLMs are a specific type of generative AI focused on language, and ChatGPT is a practical implementation of an LLM tailored for interactive, conversational use, as well as a wide range of other tasks that involve generating human-like text. Specifically, ChatGPT typically works by taking a prompt or input from the user and generating a relevant response based on the patterns and information learned during its training process.

This study utilizes Generative AI, specifically ChatGPT models, to forecast the effect of severe weather events on cross-sectional stock returns. Prompt engineering is a key component of this research, which requires creating specialized textual inputs to aid AI models in generating appropriate and focused responses. In this study, a customized prompt was carefully constructed to direct ChatGPT to assess the probable influence of weather events on stock values. The prompt reads as follows:

> *Forget all previous instructions. Pretend you are a financial expert with stock recommendation experience. I'll provide a description of a weather episode for a county on a specific day, and you need to determine whether this weather episode will significantly impact the stock prices of firms located in the county the next day. Respond with YES if the impact is significant, NO for all the others..*

This strategy is known as a 'zero-shot' task, in which the model completes the work based purely on the presented instructions and without any prior particular training on similar tasks. In other words, the model leverages its broad training on diverse data to understand and respond to the new, specific query without needing additional, task-specific training data.

The study converts ChatGPT's 'YES' and 'NO' responses into numerical scores, where 'YES' is assigned a value of one, and 'NO' is assigned a value of zero. These scores are then utilized in econometric modeling, which involves two crucial steps: episode description recommendation and returns regression. In the first stage, the overall tone of the episode description is evaluated, while in the second stage, these descriptions are used to forecast cross-sectional stock returns over daily, weekly, and monthly time periods.

Our analysis reveals that the coefficients on the dummy variable are negative and statistically significant. This indicates that when ChatGPT predicts a significant impact on future stock prices due to a storm event, the firm's subsequent returns tend to be lower. This finding suggests that ChatGPT's 'YES' recommendations are associated with negative abnormal returns in the periods following the storm event. Moreover, the predictability of returns is robust across various future time horizons, including the next trading day, the next week, the next two weeks, the next three weeks, and the next month.

The cross-sectional heterogeneity analysis reveals several consistent patterns regarding ChatGPT's ability to predict future stock returns using weather storm data. The analysis shows that ChatGPT's storm-based signals exhibit particularly strong predictive power for large firms, firms with relatively low leverage levels, firms with higher profitability levels, firms with lower bid-ask spreads (higher liquidity), and firms that experienced a higher frequency of storm events. The negative coefficients suggest market participants may underreact to weather risk. The models demonstrate greater effectiveness in forecasting returns for these subsets of firms, even after accounting for other firm-level factors that could drive future stock performance. Notably, the negative coefficients consistently indicate that ChatGPT can effectively forecast lower future returns for companies affected by severe weather, regardless of other firm-level characteristics.

The time-series heterogeneity analysis reveals several consistent patterns regarding ChatGPT's ability to forecast stock returns using weather storm data. ChatGPT's storm-based signals exhibit particularly strong predictive power during market uptrends, periods of low volatility, times of high total non-farm payroll growth, and periods of low smoothed recession probability. This suggests that market participants may underreact to weather risk during favorable economic conditions. The models demonstrate greater effectiveness in predicting returns during these specific time-series conditions, even after accounting for other macroeconomic factors that could influence market performance. The consistent patterns indicate that ChatGPT can effectively forecast stock returns under these temporal subsamples characterized by positive economic environments, regardless of other macroeconomic variables.

In summary, this study demonstrates that when ChatGPT predicts negative impacts on stock prices based on weather storm data, the affected firms tend to experience subsequent declines in their share prices. This effect is particularly pronounced for larger corporations with higher profitability, lower leverage, and more liquid stocks. Moreover, ChatGPT's storm-based forecasting models exhibit heightened accuracy during favorable economic conditions, such as market uptrends, low volatility periods, times of robust employment growth, and lower recession probabilities. These findings suggest that investors may systematically underreact to the implications of severe weather events for certain firms, especially during bullish market phases.

The paper is structured as follows. Section 2 reviews the recent literature related to this paper. Section 3 describes the data sources for the key variables. Section 4 introduces the Generative AI prompt engineering methods. Section 5 presents the summary statistics, empirical evidence and robustness checks. Finally, Section 6 offers the conclusions.

2 Literature Review

This paper contributes to the growing body of research focusing on ChatGPT and finance. Chen et al. (2023) leverage LLMs to extract contextualized representations of news text for stock return prediction, capturing both syntax and semantics more comprehensively than traditional word-based methods. Their study across 16 international markets and 13 languages shows that news information is inefficiently incorporated into prices with delays, consistent with limits-to-arbitrage, and that trading strategies exploiting fresh news alerts generate higher Sharpe ratios. Chen et al. (2024) find that positive news extracted by ChatGPT from the front pages of the Wall Street Journal can predict stock market movements and is linked to macroeconomic conditions. Investors tend to underreact to positive news, especially during economic downturns, high information uncertainty, and high novelty of news, while negative news only impacts contemporaneous returns. Traditional textual analysis methods, including word lists and models like Bidirectional Encoder Representations from Transformers (BERT)[1], show minimal predictability. ChatGPT excels in identifying economic-related news that influences the stock market. Fedyk et al. (2024) find that OpenAI's ChatGPT (GPT4) accurately predicts investment preferences across demographics, capturing nuances like women rating stocks lower than men and older individuals preferring cash, by comparing over 1,200 survey participants' preferences to ChatGPT-generated ratings. Their study reveals that while ChatGPT's free-form responses closely align with human themes like risk, return, knowledge, and experience, and ChatGPT responses tend to be transitive, human responses are more prone to violating transitivity, suggesting AI's potential in augmenting human surveys for preference elicitation in finance, particularly for applications like robo-advising. Kim et al. (2024b) assess whether GPT-4 can perform financial statement analysis comparable to professional human analysts, finding that even without narrative or industry-specific data, the LLM outperforms human analysts in predicting future earnings direction, especially in challenging situations, and matches the accuracy of specialized machine learning models. Their study reveals that the LLM's predictions are not based on its training memory but rather on generating useful narrative insights, and trading strategies based on GPT-4's predictions yield higher Sharpe ratios and alphas than those based on other models, indicating LLMs' potential for a central role in financial decision-making. Lopez-Lira and Tang (2023) find that categorizing news headlines as positive, negative, or neutral for companies' stock prices using ChatGPT and other LLMs shows a significant correlation between LLM scores and subsequent daily stock returns, surpassing traditional methods, with more complex LLMs exhibiting higher accuracy and offering higher Sharpe ratios than basic models. They attribute these patterns to economic theories concerning information diffusion frictions, limits to arbitrage, and investor sophistication, with predictability strengthening among smaller stocks and

[1]BERT is a transformer-based machine learning model for natural language processing tasks created by Google. BERT is meant to comprehend the context of a word in search queries, making it particularly useful for tasks such as question answering and language comprehension. Its bidirectional nature means that when analyzing a sentence, it examines the context from both sides (left and right), which improves comprehension ability.

following negative news, propose an interpretability technique to evaluate LLMs' reasoning, and suggest that integrating advanced language models into investment decisions can enhance prediction accuracy and trading performance. Tan et al. (2024) employ LLMs (LLMs) like BERT, RoBERTa, FinBERT, Baichuan, and ChatGLM to extract contextualized representations of nearly 2.2 million Chinese news articles from 2008 to 2023 and predict stock returns in the Chinese equity market, observing significant return predictability from news tones across models that leads to high annualized returns. Their analysis shows LLMs demonstrate effectiveness in return forecasting beyond sentiment classification, particularly benefiting stocks with lower market caps, shorting activity, institutional ownership, and state ownership, while news assimilation into prices occurs rapidly within two days, and the extracted news tones influence future investor trading directions and reflect firms' fundamentals.

Furthermore, using GPT-4-Turbo to analyze unusual financial communication in S&P 500 earnings calls, Beckmann et al. (2024) find it often leads to negative market reactions and increased trading, highlighting LLMs' potential in financial analysis. Bond et al. (2023) find ChatGPT and BARD effectively forecast stock market returns by analyzing business news to derive a sentiment indicator negatively correlated with short-term S&P 500 returns, outperforming traditional sentiment classifiers. Bertomeu et al. (2023) find that Italy's ban on ChatGPT in March 2023 led to about 9% underperformance in stocks of highly exposed Italian firms, particularly smaller and newer companies, while disrupting the information environment with fewer analyst forecasts and wider bid-ask spreads. Dou et al. (2024) show that AI-powered trading strategies enable speculators to develop collusive behaviors, achieving supra-competitive profits through price-trigger strategies and homogenized learning biases, impacting price informativeness and market liquidity even in efficient or noisy markets. Glasserman et al. (2023) find an increase in news novelty, measured by entropy changes in news text distribution, is linked to negative stock returns and macroeconomic outcomes over the next year, with entropy serving as a superior market return predictor and a hedge against aggregate risk. Itoh and

Okada (2023) show that using Shikiho's comprehensive corporate data, ChatGPT classify Japanese firms into positive and negative outlook groups, informing portfolios that suggest the market may underappreciate this factual information, highlighting issues in market efficiency and price discovery. Jha et al. (2024) show a firm-level ChatGPT investment score from conference calls predicts future capital expenditures and investments up to nine quarters ahead, correlating with CFO survey responses and indicating high-score firms face significant negative abnormal returns, showcasing ChatGPT's versatility in measuring corporate policies like dividends and employment. Kim et al. (2024a) find that generative AI tools like ChatGPT create concise, information-rich summaries of corporate disclosures, improving stock market reaction explanations and reducing information asymmetry, highlighting their value in summarizing complex information for investors. Kim and Nikolaev (2023) investigate how considering the narrative context around financial statement numbers helps investors interpret earnings more effectively. They find that analysts incorporate this context into their forecasts, increasing the relevance of earnings over time and revealing diverse patterns in earnings persistence across sectors. Kim and Nikolaev (2024) investigate the impact of incorporating narrative context into profitability measurements using a large language model. They find that context-enhanced measurements outperform conventional measures and improve portfolio performance, addressing limitations in traditional asset pricing models and emphasizing the value of context in investment strategies. Li et al. (2024a) apply generative AI models to analyze 2.4 million analyst reports from 2000 to 2020, revealing insights into analysts' perceptions of corporate culture. They find that business strategy and management team are key factors, with innovation and adaptability significantly impacting outcomes. These insights influence stock recommendations and market reactions, highlighting the importance of understanding corporate culture in assessing firm value. LoGrasso (2024) evaluate OpenAI's ChatGPT's ability to provide high-quality investment recommendations for casual investors. They find that the GPT-4 model, using information available at the time of stock selection from 1985 to 2021, achieved average alphas of around 1% per month for two-year holding periods, with consistent performance across individual portfolios. Nakano and Yamaoka (2023) investigate text data as a new source of alpha returns, proposing a sentiment analysis scheme based on ChatGPT for non-English text data. They find that an investment strategy based on this analysis outperforms market benchmarks and traditional strategies, particularly noting mean reversion following negative news for Japanese large-cap companies. Pelster and Val (2023) investigate the capability of ChatGPT-4, with internet access, to offer valuable investment advice. Through a live experiment, they establish a positive correlation between ChatGPT-4 ratings and future earnings announcements and stock returns, leading to a successful investment strategy based on ChatGPT-4's attractiveness ratings. Shah et al. (2024) focus on the influence of Federal Open Market Committee (FOMC) pronouncements on financial market returns. They construct a comprehensive dataset of FOMC communications, introducing a hawkish-dovish classification task, and identify RoBERTa-large as the best model. The study evaluates the impact of the FOMC's policy stance on treasury and stock markets, providing publicly available resources for further exploration. Wang and Ling (2024) examine the impact of the ChatGPT launch on the Chinese capital market. They find that funds related to ChatGPT experience improved Sharpe ratios and excess returns post-launch, with variations based on fund company ownership, investment styles, and historical returns. Increased fund inflows and returns of ChatGPT concept stocks significantly contribute to this improved performance.

This paper is also related to a stream of literature on the impact of climate risk on the financial market. Sautner et al. (2023a) develop a method to identify firms' climate change exposures from earnings call transcripts using machine learning, capturing various shocks, and show that these measures predict outcomes related to the net-zero transition and are priced in financial markets. Sautner et al. (2023b) estimate the risk premium for S&P 500 stocks' climate change exposure from 2005 to 2020, finding an overall insignificant unconditional risk premium but noting positive trends pre-financial crisis and post-2014. Ilhan et al. (2023) demonstrate through a survey and empirical evidence that institutional investors highly value and actively seek climate risk disclosures. Ilhan et al. (2021) highlight the necessity for robust regulatory measures to address climate change, revealing how climate policy uncertainty is reflected in option market pricing, with higher costs for downside tail risk protection observed for firms with more carbon-intensive operations. Cuculiza et al. (2024) find that firms more sensitive to temperature changes have lower future profitability, riskier policies, and lower subsequent returns, suggesting mispricing that nonlocal investors and analysts contribute to, enabling a trading strategy exploiting this mispricing to generate over 4% annual risk-adjusted returns from 1968-2020. Ilhan (2022) show that U.S. households more exposed to future sea level rise risks hold less stock market participation compared to unexposed neighbors. Kölbel et al. (2024) use BERT to assess regulatory climate risk disclosures' effects on CDS, finding that disclosing transition risks tends to increase CDS spreads post-2015 Paris Climate Agreement, while disclosing physical risks decreases them. Pankratz and Schiller (2024) investigate how physical climate events like heat and floods at supplier locations negatively impact operating income and lead to customers terminating relationships with suppliers having higher realized climate exposure. Li et al. (2024b) develop text-based measures to quantify firms' exposure to physical and transition climate risks from earnings call transcripts, finding that firms with high transition risk exposure are discounted by investors and respond differently through investments, innovation, and employment policies. Addoum et al. (2023) examine the impact of extreme temperatures on corporate profitability across industries, finding significant bidirectional effects on earnings in over 40% of industries. Ginglinger and Moreau (2023) use firm-level data on forward-looking physical climate risk to examine its impact on capital structure, finding that greater physical climate risk leads to lower leverage in the post-2015 period due to increased disclosure standards and higher expected distress and operating costs. Bartrama et al. (2022) show that California's cap-and-trade program leads to regulatory arbitrage, with financially constrained firms shifting emissions and output to other states, undermining the policy's effectiveness. Engle et al. (2020) propose a procedure to dynamically hedge climate change risk by extracting innovations from constructed climate news series through textual analysis, using ESG scores to build hedge portfolios that effectively hedge innovations in climate news both in-sample and out-of-sample. Gounopoulos and Zhang (2024) find that companies increase cash reserves in response to rising climate risks, particularly financially constrained firms with low environmental awareness, who rely more on equity issuance and cost cuts than debt to bolster cash holdings. Lin et al. (2023) show that production inflexibility coupled with product price uncertainty creates price risk, significantly impacting firms' liquidity management, with higher electricity price volatility leading to increased cash holdings among firms using inflexible production technologies. Pankratz et al. (2022) link firm performance, analyst forecasts, and earnings announcement returns to firm-specific heat exposure measures, finding that increased exposure to extremely high temperatures reduces revenues and operating income, with analysts and investors failing to fully anticipate the economic repercussions of heat as a physical climate risk.

3 Data

We utilize two databases, "Storm Events Database" and Center for Research in Security Prices (CRSP). The CRSP daily and monthly returns dataset includes information on daily and monthly stock returns, stock prices, trading volumes, and shares outstanding for a wide array of companies listed on major U.S. stock exchanges, including those on the NYSE, AMEX, and Nasdaq.

The "Storm Events Database" is provided by the National Weather Service (NWS) of the National Oceanic and Atmospheric Administration (NOAA). It is available at https://www.ncdc.noaa.gov/stormevents/. The database is freely accessible to the public at no cost. The dataset encompasses several variables, which we have tailored for inclusion in this paper.

- $ID_{episode}$: A unique identifier assigned by NWS to a storm episode, which includes significant multiple weather events that may lead to fatalities, injuries, property damage, and commercial disruption. A weather episode represents a prolonged period of similar weather conditions that may encompass multiple weather events.

- ID_{event}: A unique identifier assigned by NWS to a single event within a storm episode. A weather event is a specific, short-term occurrence.
- $Narrative_{episode}$: Provides a detailed overview of the storm episode's nature and activity as reported by NWS.
- $Narrative_{event}$: Offers specific details of the individual event as reported by NWS.
- *Yearmonth*: Represents the year and month when the weather event began.
- *Year*: Indicates the specific year of the weather event's start.
- *Month*: Lists the month in which the weather event was recorded, such as January, April, July, or September.
- *State*: Specifies the state where the weather event occurred.
- *County_Name*: Identifies the county name where the weather event took place.
- *Type_of_Event*: Describes the type of weather event, including thunderstorm wind, winter storm, or strong wind.
- *Magnitude*: Indicates the severity of the weather event, primarily used for wind speeds (in knots). For each county and year-month, the highest value of *Magnitude* is selected.
- $Damage_{property}$: The damage to property in dollars caused by the weather event.
- $Damage_{crops}$: The damage to crops in dollars caused by the weather event.
- $Deaths_{direct}$: The number of deaths directly attributed to the weather event.
- $Deaths_{indirect}$: The number of deaths indirectly associated with the weather event.
- *Death*: The total number of deaths, combining both direct and indirect fatalities.
- $Injuries_{direct}$: The number of injuries directly caused by the weather event.

- $Injuries_{indirect}$: The number of injuries indirectly related to the weather event.
- $Injury$: The total number of injuries, combining both direct and indirect injuries.

For each county and each year-month, the values for $Damage_{property}$, $Damage_{crops}$, $Death$, and $Injury$ are aggregated by summing their values.

In our sample which spans from March 2022 through December 2023, there are a total of 52,166 storm events. The types of storm events included in this dataset are: thunderstorm wind (38,460 (73.73%)), high wind (6,787 (13.01%)), marine thunderstorm wind (5,097 (9.77%)), strong wind (1,710 (3.28%)), marine high wind (107 (0.21%)), and marine strong wind (5 (0.01%)). Figure 1 indicates a clear seasonal pattern with storm events peaking in mid-year (June and July) for both 2022 and 2023. The year 2023 shows a higher number of storm events compared to 2022, with July 2023 being particularly notable for its high count. Table 1 presents the geographical distribution of storm events. Texas has the highest number of storm events, with 2,803 events, making up 5.37% of the total. This is due to Texas's large geographic area and diverse climate. States in the Southeastern and Central United States, such as Georgia, Virginia, Illinois, and Kansas, also show high storm frequencies. This region is often referred to as "Tornado Alley" and is known for its high incidence of tornadoes and severe weather. The table also highlights significant storm activity in states bordering the Gulf of Mexico and the Atlantic Ocean, such as Alabama, Florida, and North Carolina. These areas are prone to hurricanes and tropical storms.

To give a concrete example, on August 6, 2023, in Illinois, there was a thunderstorm wind event with a magnitude of 52[2]. The *episode_id* is 184619. The *episode_narrative* is

> *A vigorous short-wave trough interacted with an approaching low pressure system and associated warm frontal boundary to trigger clusters of strong to severe thunderstorms across west-central Illinois during the late afternoon of August 6th. Due to enhanced low-level wind shear along and north of the warm front, many of the cells began rotating and spinning up occasional funnel clouds. One of the cells intensified*

[2]Please see https://www.weather.gov/bmx/event_03242023.

as it pushed into Sangamon County, producing a long-track tornado that was on the ground for over 25 miles from just north of Pawnee in far southeastern Sangamon County to southeast of Stonington in Christian County. The tornado produced EF-1 damage along much of its path, peaking with EF-2 damage just south of Willeys in Christian County. This particular thunderstorm cell dropped 2-inch diameter hail north of Taylorville and went on to produce a downburst with 80 mph wind gusts in far eastern Shelby County near Lake Mattoon. The storms eventually congealed into a line as they pushed into eastern Illinois later in the evening, with scattered wind damage, hail, and flash flooding..

Another example, on September 9, 2023, in Virginia, there was a thunderstorm wind event with a magnitude of 50. The *episode_id* is 185704. The *episode_narrative* is

Scattered severe thunderstorms in advance of a frontal boundary produced damaging winds across portions of central and southeast Virginia..

For the county of Southampton, the *event_id* is 1139867. The *event_narrative* is *Power line was downed on Seacock Chapel Road*. For the county of Suffolk, the *event_id* is 1139869. The *event_narrative* is *Trees were downed near the Holland Bypass on Route 58*.

For the county of Effingham, the *event_id* is 1130577. The *event_narrative* is *A trained spotter estimated a 60 mph wind gust*. For the county of Clay, the *event_id* is 1130578. The *event_narrative* is *An emergency manager estimated sustained winds of 50 mph with gusts to around 60 mph*.

Word Clouds

A word cloud, also known as a tag cloud, is a visual representation where the size of each word corresponds to its frequency or relevance within a given text. It is created by tokenizing the text into individual words, calculating the frequency of each word, and then rendering the words with varying font sizes based on their frequencies more frequent words appear larger and more prominent. Figures 2 depicts the textual information of *episode_narrative*, using word clouds. The font size of each word reflects its significance within the provided narratives, with less frequent words displayed smaller and more frequent words appearing larger.

Figure 2 highlights the overall weather conditions and main activities during the episode, including terms like "severe thunderstorm," "wind damage," "wind gust," "shower thunderstorm," and "heavy rain," among others. The distinctive characteristic of these descriptions is that they consist of scientific words, which differ in nature from those found in the sentiment analysis by Chen et al. (2023) and Tan et al. (2024). The news in Chen et al. (2023) and Tan et al. (2024) exhibit distinctive positive or negative tones. For example, Chen et al. (2023) observe positive words such as raise, gain, approval, strong, and outstanding, which are often positively correlated with realized returns, and negative words such as fall, low, loss, and charge, which are often negatively correlated with realized returns.

However, because *episode_narrative* and *event_narrative* depict natural processes, their tones differ from those seen in news. Instead of communicating thoughts or emotional responses, these narratives present factual and scientific accounts of weather phenomena. The language utilized in these accounts is focused on describing the meteorological conditions, physical impacts, and technical components of the storm events. This provides specific terms for weather patterns, measures, and impacts. As a result, the word clouds produced by these narratives represent a more neutral, objective viewpoint geared at accurately describing the events and their repercussions, rather than expressing subjective thoughts or sentiments. This is in stark contrast to the tone of financial news. In this study, for brevity reason, we focus on the *episode_narrative*.

4 Methods

4.1 Generative AI Prompt Engineering

This paper utilizes Generative AI to predict the impact of weather conditions on cross-sectional stock returns. Generative AI refers to a class of artificial intelligence algorithms capable of generating new data, such as text, images, or music, similar to the data they were trained on. These models learn patterns from large datasets and can produce creative and coherent outputs based on that learning. Generative AI models like GPT-3 and GPT-4 form the foundation of ChatGPT, a conversational agent designed to engage in dialogue with users by generating human-like text based on the input it receives.

ChatGPT is a specific application of Generative AI that employs a version of the GPT (Generative Pre-trained Transformer) model to understand and respond to user queries naturally and coherently. It can be used for various applications, such as customer support, content creation, tutoring, and more. Prompts play a critical role in guiding Generative AI models like ChatGPT to produce specific and relevant responses. They are the key to unlocking the potential of ChatGPT in particular contexts.

Prompts are vital in steering language models like ChatGPT to produce appropriate responses for specific tasks and queries. These textual inputs provide the necessary context and instructions, allowing the model to comprehend the assigned task and prepare for the desired interaction. For example, a prompt can direct ChatGPT to assume the role of a financial expert, language translator, or storyteller. The quality and clarity of the prompt significantly influence the nature and quality of the AI-generated output.

The prompt initiates the model's response generation process, varying in length from a concise sentence to an intricate multi-sentence paragraph, contingent upon the task's complexity. Upon receiving the prompt, the model meticulously examines its syntactic structure and semantic meaning, generates an array of potential responses, and ultimately selects the most fitting option based on coherence, relevance, and grammatical accuracy.

Prompts unlock the capability of language models to undertake a diverse array of language-related tasks, including translation, text summarization, question answering, and the generation of coherent, human-like text. These textual inputs enable the models to adapt to specific contextual settings and tailor their responses to align with user requirements.

For the purposes of our research, we employed the following prompt and applied it to a dataset comprised of narratives describing various weather episodes:

Forget all previous instructions. Pretend you are a financial expert with stock recommendation experience. I'll provide a description of a weather episode for a county on a specific day, and you need to determine whether this weather episode will significantly impact the stock prices of firms located in the county the next day. Respond with YES if the impact is significant, NO for all the others..

This prompt instructs the language model, ChatGPT, to assume the role of a financial expert with experience in stock recommendations. The prompt is specifically designed for financial analysis and asks ChatGPT to evaluate a given weather episode and its potential impact on a firm's stock price. We set the temperature parameter of GPT models to 0, the goal is to maximize the reproducibility and consistency of the generated responses. The prompt explicitly states that the narrative describing the weather episode is the sole source of information provided to ChatGPT. It is an implicit assumption that this narrative contains sufficient details for an expert in the financial industry to reasonably evaluate its potential impact on stock prices.

Here, we ask ChatGPT to perform what is known as a "zero-shot" task. In a zero-shot scenario, the ChatGPT model is given a task without any prior examples or specific training on that task. The model relies solely on its pre-existing knowledge and the instructions provided in the prompt. The ChatGPT model is not given any examples of weather descriptions and corresponding stock price impacts before making predictions. It has to understand and perform the task based purely on the instructions given in the prompt.

The prompt explicitly tells the ChatGPT to pretend to be a financial expert and make a prediction based on the given weather description. This clear instruction is what guides the ChatGPT in generating the desired response without any previous training on similar tasks. The prompt clearly defines the task and the expected output format, enabling the ChatGPT to generate a response even though it hasn't been specifically trained on this exact task with similar examples before.

For example, consider the following description of a storm episode in Madison County, Alabama, on June 7, 2022: *Clusters of thunderstorms developed during the early afternoon* *hours in north Alabama. The thunderstorms produced large hail and damaging winds in Madison County. Flash flooding was also reported in Madison and Marshall Counties.* ChatGPT's response is YES.

For the following description of a storm episode in Madison County, Alabama, on June 6, 2022: *Two clusters of strong to severe thunderstorms moved west out of Georgia into Alabama, including north Alabama, during the afternoon and evening hours. The thunderstorms were strong enough to knock down trees in Madison County during the afternoon and Marshall County during the evening hours,* ChatGPT's response is NO.

4.2 Econometric Modeling

We prompt ChatGPT to provide a recommendation for an episode description and transform it into a numerical ChatGPT score, where "YES" is 1 and "NO" is 0. Using the scores generated by ChatGPT as inputs, we conduct econometric modeling. Our study involves two key modeling steps: episode description recommendation and returns regression. The first step, episode description recommendation, allows us to assess the overall tone of the episode description. The second step, returns regression, involves predicting cross-sectional stock returns over daily, weekly, and monthly horizons based on the episode descriptions.

We estimate the following OLS regressions of the next day's, week's, and month's stock returns on the ChatGPT score. All of our results are out-of-sample by empirical design:

$$r_{i,t} = \beta^{'} ChatGPTScore_{i,t-1} + \delta_{t-1} + \delta_i + \epsilon_{i,t},$$

where the dependent variable, $r_{i,t}$, is stock i's return over a subsequent trading day, week, or month. δ_{t-1} and δ_i are date and firm fixed effects, respectively. Standard errors are clustered at the firm level.

By including firm fixed effects δ_i, we account for any time-invariant firm-specific factors, isolating the effect of the ChatGPT score from these unobserved influences. By including date fixed effects δ_{t-1}, we control for any common shocks, trends, or macroeconomic events that occur on specific dates and could influence stock returns across all firms. Clustering standard errors at the firm level accounts for within-firm correlation in the residuals, recognizing that observations from the same firm may be more similar to each other than observations from different firms. Clustering ensures that the standard errors are robust to within-firm correlations, leading to more reliable statistical inference.

5 Empirical Analysis

5.1 Descriptive Analysis

Table 2 presents the descriptive statistics of the sample: In Panel A, ”Episode Description Length” represents the length of the episode description measured in words. ”Episode ChatGPT Score” is a binary variable indicating whether ChatGPT’s response was ”YES” (coded as 1) or ”NO” (coded as 0). The episode description length ranges from 44 words to 5,794 words, with a mean of 336 words and a median of 244 words. The mean of the ’Episode ChatGPT Score’ variable is 0.258, indicating that out of the 644 episodes, 166 received a ”YES” response from ChatGPT, while 478 received a ”NO” response.

In Panel B, the *Magnitude* variable exhibits a wide range, spanning from 35 to 96, with an average around 54.81. The distribution of *Magnitude* is spread out, as indicated by a standard deviation of 8.43. The $Log(Damage_{property} + 1)$ variable is heavily skewed, with a median (7.60) higher than the 25th percentile (0.00) but lower than the mean (5.35), suggesting the presence of significant outliers. Most $Log(Damage_{crops} + 1)$ values are zero, as evidenced by a very low mean (0.04) and a relatively high standard deviation (0.60), implying occasional high crop damage values that contribute to the large spread.

The data reveals that both property and crop damage tend to exhibit extreme value outliers, with property damage having a higher occurrence of non-zero values compared to crop damage. $Log(Death + 1)$ occurrences are rare in the dataset, as reflected by the very low mean (0.02) and the 25th, median, and 75th percentiles all being 0, indicating that most observations reported no deaths. Similarly, $Log(Injury + 1)$ occurrences are also rare, with a very low mean (0.04) and the 25th, median, and 75th percentiles all being 0, suggesting that most observations reported no injuries.

Panel C shows that the correlations between most of the key variables are quite small, ranging from -0.14 to 0.18, except that the correlations between $Log(Death + 1)$ and $Log(Injury + 1)$ is moderate at 0.56.

5.2 Baseline Results

Table 3 reports the return prediction using OLS regression. The dependent variables are the stock returns over various future time horizons: the next trading day (Model 1), the next week (Model 2), the next two weeks (Model 3), the next three weeks (Model 4), and the next month (Model 5). The key independent variable of interest is a dummy variable that takes the value of 1 if ChatGPT recommends "YES", and 0 otherwise. The regression includes control variables such as the return on the event day, the magnitude of the storm event, property damage, crop damage, and the number of deaths and injuries caused by the storm event. Additionally, we control for date and firm fixed effects. Standard errors are clustered at the firm level to account for potential within-firm correlation.

The main results in Table 3 demonstrate robust return predictability over various future time horizons: the next trading day (Model 1), the next week (Model 2), the next two weeks (Model 3), the next three weeks (Model 4), and the next month (Model 5). The negative and statistically significant coefficients on the dummy variable, which takes the value of 1 if ChatGPT recommends "YES", and 0 otherwise, indicate that when ChatGPT predicts a significant impact on future stock prices due to a storm event, the firm's subsequent returns tend to be lower. This finding suggests that ChatGPT's "YES" recommendations are associated with negative abnormal returns in the periods following the storm event.

Specifically, the coefficient for the dummy variable is -0.002 (-0.2% daily return) with a t-stat of -2.09 for the next trading day (Model 1), the coefficient for the dummy variable is -0.014 (-1.4% weekly return) with a t-stat of -4.72 for the next week (Model 2), the coefficient for the dummy variable is -0.027 (-2.7% biweekly return) with a t-stat of -6.01 for the next two weeks (Model 3), the coefficient for the dummy variable is -0.027 (-2.7% return over three weeks) with a t-stat of -4.86 for the next three weeks (Model 4), and the coefficient for the dummy variable is -0.024 (-2.4% monthly return) with a t-stat of -3.18 for the next month (Model 5).

All models suggest that the significance of ChatGPT recommendations even after controlling for other characteristics that might influence future returns. Overall, ChatGPT effectively forecast future stock returns based on the weather storm data.

5.3 Cross-sectional Heterogeneity

Table 4 presents the return prediction results for firms with market capitalizations above the mean in our sample. The study highlights ChatGPT's superior ability to forecast future stock returns using weather storm data for these larger firms.

Specifically, the results indicate a significant negative impact of the dummy variable on returns across various time frames. For the next trading day (Model 1), the coefficient is -0.003, corresponding to a -0.3% daily return, with a t-statistic of -2.23. Over the following week (Model 2), the coefficient increases to -0.016, or -1.6% weekly return, with a t-statistic of -4.36. This trend continues with a biweekly return (Model 3) where the coefficient is -0.028 (-2.8%) and the t-statistic is -5.43. For the next three weeks (Model 4), the coefficient reaches -0.031, equating to a -3.1% return, with a t-statistic of -4.89. Finally, for the monthly return (Model 5), the coefficient is -0.030, or -3.0%, with a t-statistic of -3.34. The results from all models remain robust even after controlling for other characteristics that could potentially impact future returns.

As shown in Table 4, the negative coefficients on the large firm dummy variable indicate that large companies tend to experience lower returns following severe weather events. These findings suggest that market participants may underreact to weather risk for larger firms. The analysis reveals that ChatGPT's predictive power for future stock returns, based on weather storm data, is particularly strong for larger firms.

The return predictions for firms with a debt-to-equity ratio below the sample mean, reported in Table 5, suggest that ChatGPT's ability to anticipate future stock market performance by leveraging weather storm information is more pronounced for companies with lower debt levels. Specifically, for firms with below-average debt-to-equity ratios, the dummy variable coefficient indicates a negative relationship between storm exposure and subsequent returns. This effect is statistically significant across multiple time horizons, even after controlling for various firm characteristics that could influence future returns.

Over the next trading day, the coefficient of -0.002 (-0.2%) has a t-statistic of -1.81 (Model 1). For the following week, the coefficient is -0.014 (-1.4%) with a t-statistic of -4.59 (Model 2). Looking two weeks out, the coefficient is -0.026 (-2.6%) with a t-statistic of -5.71 (Model 3). At the three-week horizon, the coefficient is -0.023 (-2.3%) with a t-statistic of -4.02 (Model 4). Finally, for the next monthly period, the coefficient is -0.025 (-2.5%) with a t-statistic of -3.24 (Model 5).

The consistent negative coefficients across these models indicate that ChatGPT's storm-based signals have greater predictive power for forecasting returns among firms with relatively low leverage. This pattern holds even after accounting for other firm-level factors that could drive future stock performance.

Table 6 reports the return prediction for the subsample of firms with a return-to-assets ratio above the mean return-to-assets ratio of the full sample. The outcomes highlight that ChatGPT's competence in forecasting stock returns using weather storm data as an input is superior for firms with with high return-to-assets ratio.

Across multiple time horizons, the dummy variable coefficients exhibit a consistent negative relationship between storm exposure and subsequent returns for these high return-on-assets firms. Over the next trading day, the coefficient is -0.002 (-0.2%) with a t-statistic of -1.65 (Model 1). For the following week, the coefficient is -0.011 (-1.1%) with a t-statistic of -3.64 (Model 2). Looking two weeks out, the coefficient is -0.026

(-2.6%) with a t-statistic of -5.76 (Model 3). At the three-week horizon, the coefficient is -0.027 (-2.7%) with a t-statistic of -5.26 (Model 4). Finally, for the next monthly period, the coefficient is -0.023 (-2.3%) with a t-statistic of -3.24 (Model 5).

These findings demonstrate the predictive power of ChatGPT's storm-based signals for anticipating returns among firms with relatively high profitability levels. The results are robust to the inclusion of control variables capturing other firm-level factors that could influence future stock performance.

Table 7 presents the return predictions for firms with a bid-ask spread below the sample mean. The results suggest that ChatGPT's weather storm data-driven models exhibit enhanced effectiveness in forecasting stock returns for upcoming periods when applied to companies with narrower bid-ask spreads.

Across multiple future time horizons, the dummy variable coefficients consistently indicate a negative relationship between storm exposure and subsequent returns for firms with relatively low bid-ask spreads. For the next week, the coefficient is -0.011 (-1.1%) with a t-statistic of -3.62 (Model 2). Looking two weeks out, the coefficient is -0.027 (-2.7%) with a t-statistic of -5.91 (Model 3). At the three-week horizon, the coefficient is -0.026 (-2.6%) with a t-statistic of -4.79 (Model 4). Finally, for the next monthly period, the coefficient is -0.023 (-2.3%) with a t-statistic of -3.06 (Model 5).

These findings demonstrate the predictive power of ChatGPT's storm-based signals for anticipating returns among firms with lower bid-ask spreads, a measure of liquidity. The results remain robust even after accounting for other firm-level characteristics that could potentially influence future stock performance.

Table 8 presents the return predictions using OLS regression for firms in the top third of storm event frequency. The results indicate a consistent negative relationship between storm exposure and subsequent returns for these companies that experienced a higher incidence of storm events.

Across multiple future time horizons, the dummy variable coefficients exhibit statistically significant negative values. Over the next trading day, the coefficient is -0.006

(-0.6%) with a t-statistic of -2.86 (Model 1). For the following week, the coefficient is -0.019 (-1.9%) with a t-stat of -3.40 (Model 2). Looking two weeks out, the coefficient is -0.027 (-2.7%) with a t-stat of -3.40 (Model 3). At the three-week horizon, the coefficient is -0.044 (-4.4%) with a t-stat of -4.68 (Model 4). Finally, for the next monthly period, the coefficient is -0.040 (-4.0%) with a t-stat of -3.40 (Model 5).

These findings demonstrate the predictive power of ChatGPT's storm-based signals for forecasting returns among firms that experienced a higher frequency of storm events. The results remain robust even after accounting for other firm-level characteristics that could potentially influence future stock performance.

In summary, the cross-sectional heterogeneity analysis reveals several consistent patterns regarding ChatGPT's ability to predict future stock returns using weather storm data. For large companies, which tend to experience lower returns following severe weather events, the negative coefficients suggest market participants may underreact to weather risk. ChatGPT's storm-based signals exhibit particularly strong predictive power for these larger firms. The models also demonstrate greater effectiveness in forecasting returns for companies with relatively low leverage levels, even after accounting for other firm-level factors that could drive future stock performance.

Furthermore, the findings show ChatGPT's storm data-driven models anticipate returns more accurately among firms with higher profitability levels. These results remain robust to the inclusion of control variables capturing other characteristics that could influence future returns. Similarly, for companies with lower bid-ask spreads, a measure of liquidity, ChatGPT's storm-based signals demonstrate predictive power that holds after controlling for other potential performance drivers.

Notably, the models exhibit enhanced predictive capabilities for firms that experienced a higher frequency of storm events. Regardless of other firm-level characteristics, the negative coefficients consistently indicate ChatGPT can effectively forecast lower future returns for companies affected by severe weather.

5.4 Time-series Heterogeneity

Table 9 presents the return predictions for periods when the S&P 500 level[3] is above its mean during the sample period. The results indicate that ChatGPT's ability to utilize weather storm data to forecast upcoming stock returns is particularly effective during market uptrend periods.

Specifically, for the next trading day (Model 1), the coefficient for the dummy variable is -0.003, indicating a -0.3% daily return, with a t-statistic of -2.32. Over the following week (Model 2), the coefficient is -0.019, corresponding to a -1.9% weekly return, with a t-statistic of -5.47. This trend continues for the biweekly return (Model 3), where the coefficient is -0.035, equating to a -3.5% return, with a t-statistic of -6.66. For the next three weeks (Model 4), the coefficient is -0.038, reflecting a -3.8% return, supported by a t-statistic of -5.91. Finally, for the monthly return (Model 5), the coefficient is -0.023, indicating a -2.3% return, with a t-statistic of -2.84.

These findings demonstrate ChatGPT's ability to utilize weather storm data to forecast stock returns is particularly effective during market uptrends. The results remain robust even after accounting for other firm-level characteristics that could potentially influence future stock performance.

Table 10 presents the return predictions for periods when the VIX level is below its mean during the sample period. The VIX[4], also known as the CBOE Volatility Index or the "fear index," measures the expected volatility of the S&P 500 index over the next 30 days. A high VIX level generally indicates increased volatility and uncertainty in the stock market, while a low VIX level suggests lower volatility. The results indicate that ChatGPT's ability to utilize weather storm data to forecast upcoming stock returns is particularly effective during periods of low volatility.

Specifically, for the next trading day (Model 1), the coefficient for the dummy variable is -0.003, indicating a -0.3% daily return, with a t-statistic of -2.46. Over the following

[3] https://fred.stlouisfed.org/series/SP500
[4] https://fred.stlouisfed.org/series/vixcls

week (Model 2), the coefficient is -0.019, corresponding to a -1.9% weekly return, with a t-statistic of -5.79. This trend continues for the biweekly return (Model 3), where the coefficient is -0.034, equating to a -3.4% return, with a t-statistic of -6.47. For the next three weeks (Model 4), the coefficient is -0.036, reflecting a -3.6% return, supported by a t-statistic of -5.77. Finally, for the monthly return (Model 5), the coefficient is -0.018, indicating a -1.8% return, with a t-statistic of -2.10.

These findings demonstrate that ChatGPT's ability to utilize weather storm data to forecast stock returns is particularly effective during periods of low volatility. The results remain robust even after accounting for other firm-level characteristics that could potentially influence future stock performance.

Table 11 presents the return predictions for periods when the total non-farm payroll growth is above its mean during the sample period. Total Non-farm Payroll Growth[5] refers to the increase in the number of paid workers in the United States economy, excluding those in the farming sector and certain other categories such as government employees, private home employees, and nonprofit organization employees. High total non-farm payroll growth means that the economy is experiencing a significant increase in the number of jobs outside the farming sector. This is generally considered a positive sign for the economy.

The results in Table 11 indicate that ChatGPT's ability to utilize weather storm data to forecast upcoming stock returns is particularly strong during periods of high total non-farm payroll growth.

Specifically, for the next trading day (Model 1), the coefficient for the dummy variable is -0.003, indicating a -0.3% daily return, with a t-statistic of -2.42. Over the following week (Model 2), the coefficient is -0.019, corresponding to a -1.9% weekly return, with a t-statistic of -5.47. This trend continues for the biweekly return (Model 3), where the coefficient is -0.035, equating to a -3.5% return, with a t-statistic of -6.61. For the next three weeks (Model 4), the coefficient is -0.037, reflecting a -3.7% return, supported by

[5]https://fred.stlouisfed.org/series/PAYEMS

a t-statistic of -5.87. Finally, for the monthly return (Model 5), the coefficient is -0.023, indicating a -2.3% return, with a t-statistic of -2.80. The results remain robust even after accounting for other firm-level characteristics that could potentially influence future stock performance.

Table 12 presents the return predictions for periods when the Smoothed Recession Probability is below its mean during the sample period. "Smoothed Recession Probability"[6] is a statistical measure that estimates the likelihood of an economy being in a recession at a given time, based on historical data and economic indicators. This probability is often derived from models that use various economic variables, such as GDP growth, unemployment rates, industrial production, and other leading indicators, to forecast the state of the economy. Smoothed Recession Probability provides a refined and stable estimate of the chances that the economy is currently in, or is about to enter, a recession, based on comprehensive economic data and sophisticated modeling techniques. When the "Smoothed Recession Probability" is low, it indicates that the likelihood of the economy being in a recession, or entering one in the near future, is minimal.

The results of Table 12 show that ChatGPT's ability to utilize weather storm data to forecast upcoming stock returns is particularly effective during low Smoothed Recession Probability periods.

Specifically, for the next trading day (Model 1), the coefficient for the dummy variable is -0.002, indicating a -0.2% daily return, with a t-statistic of -2.22. Over the following week (Model 2), the coefficient is -0.014, corresponding to a -1.4% weekly return, with a t-statistic of -4.71. This trend continues for the biweekly return (Model 3), where the coefficient is -0.029, equating to a -2.9% return, with a t-statistic of -6.42. For the next three weeks (Model 4), the coefficient is -0.029, reflecting a -2.9% return, supported by a t-statistic of -5.14. Finally, for the monthly return (Model 5), the coefficient is -0.029, indicating a -2.9% return, with a t-statistic of -3.89. The results remain robust even after accounting for other firm-level characteristics that could potentially influence future stock

[6] https://fred.stlouisfed.org/series/RECPROUSM156N

performance.

In summary, the time-series heterogeneity analysis reveals several consistent patterns regarding ChatGPT's ability to forecast stock returns using weather storm data. During market uptrends, ChatGPT's storm-based signals exhibit particularly strong predictive power, suggesting that market participants may underreact to weather risk during these periods. The models also demonstrate greater effectiveness in predicting returns during times of low volatility, even after accounting for other macroeconomic factors that could influence market performance.

Furthermore, the findings show ChatGPT's storm data-driven models anticipate returns more accurately during periods of high total non-farm payroll growth. These results remain robust to the inclusion of control variables capturing other economic indicators that could impact stock returns. Similarly, during periods of low Smoothed Recession Probability, ChatGPT's storm-based signals demonstrate predictive power that holds after controlling for other potential market drivers.

Notably, the models exhibit enhanced predictive capabilities during times characterized by these favorable economic conditions. Regardless of other macroeconomic variables, the consistent patterns indicate ChatGPT can effectively forecast stock returns under these specific time-series conditions.

For brevity reasons, we don't report the results for firms with smaller sizes, high leverage ratios, low profitability, high bid-ask spreads, low storm event frequency, or during market downturns, high VIX period, low total non-farm payroll growth periods, high smoothed recession probability periods, as they are weakly or not significant.

6 Conclusion

This study bridges the burgeoning literature on generative AI, particularly LLMs like ChatGPT, with climate finance research, aiming to forecast the impact of severe weather events on stock returns. Generative AI, especially LLMs, has shown significant potential in financial analysis by generating insightful responses to tailored prompts. By carefully engineering prompts to assess weather storm episode descriptions, ChatGPT provides 'YES' or 'NO' predictions on whether a weather event will significantly affect a firm's future stock prices. These predictions are then incorporated into econometric models analyzing cross-sectional stock returns over different time horizons.

Our findings indicate that ChatGPT's predictions, when forecasting a significant impact on stock prices due to severe weather events, are associated with negative abnormal returns. This pattern holds across various future time horizons, including daily, weekly, and monthly returns. The robustness of these results suggests that ChatGPT's 'YES' recommendations effectively signal subsequent declines in stock prices, highlighting the model's predictive power in financial markets.

Further analysis reveals that ChatGPT's predictive accuracy is particularly strong for large firms with high profitability, low leverage, and high liquidity. These firms tend to experience more pronounced declines in stock prices following severe weather events. The models exhibit stronger predictive power for these types of firms, suggesting market participants may systematically underreact to weather risk, especially among bigger corporations.

Furthermore, the analysis uncovers that ChatGPT's forecasting abilities are heightened during favorable economic conditions such as market uptrends, low volatility periods, robust employment growth, and lower recession probabilities. The models demonstrate comparative advantages in predicting returns during these bullish phases, indicating investors may overlook weather implications when overall sentiment is positive.

Overall, this study underscores the value of integrating generative AI, like ChatGPT, in financial analysis and climate risk assessment. It highlights the emerging potential of integrating generative AI into investment decision-making processes. By accurately predicting the financial impact of weather events, ChatGPT can provide valuable insights for investors, enabling them to make more informed decisions and potentially exploit market inefficiencies related to weather-related risks.

References

Addoum, Jawad M., David T. Ng, and Ariel Ortiz-Bobea (2023), "Temperature Shocks and Industry Earnings News," *Journal of Financial Economics*, vol. 150, 1–45.

Bartrama, Söhnke M., Kewei Hou, and Sehoon Kim (2022), "Real Effects of Climate Policy: Financial Constraints and Spillovers," *Journal of Financial Economics*, vol. 143, 668–696.

Beckmann, Lars, Heiner Beckmeyer, Ilias Filippou, Stefan Menze, and Guofu Zhou (2024), "Unusual Financial Communication–Evidence from ChatGPT, Earnings Calls, and the Stock Market," Working Paper.

Bertomeu, Jeremy, Yupeng Lin, Yibin Liu, and Zhenghui Ni (2023), "Capital Market Consequences of Generative AI: Early Evidence from the Ban of ChatGPT in Italy," Working Paper.

Bond, Shaun A., Hayden Klok, and Min Zhu (2023), "Large Language Models and Financial Market Sentiment," Working Paper.

Chen, Jian, Guohao Tang, Guofu Zhou, and Wu Zhu (2024), "ChatGPT, Stock Market Predictability and Links to the Macroeconomy," Working Paper.

Chen, Yifei, Bryan Kelly, and Dacheng Xiu (2023), "Expected Returns and Large Language Models," Working Paper.

Cuculiza, Carina, Alok Kumar, Wei Xin, and Chendi Zhang (2024), "Temperature Sensitivity, Mispricing, and Predictable Returns," Working Paper.

Dou, Winston Wei, Itay Goldstein, and Yan Ji (2024), "AI-Powered Trading, Algorithmic Collusion, and Price Efficiency," Working Paper.

Engle, Robert F., Stefano Giglio, Bryan Kelly, Heebum Lee, and Johannes Stroebel (2020), "Hedging Climate Change News," *Review of Financial Studies*, vol. 33, 1184–1216.

Fedyk, Anastassia, Ali Kakhbod, Peiyao Li, and Ulrike Malmendier (2024), "ChatGPT and Perception Biases in Investments: An Experimental Study," Working Paper.

Ginglinger, Edith and Quentin Moreau (2023), "Climate Risk and Capital Structure," *Management Science*, vol. 69, 7492–7516.

Glasserman, Paul, Harry Mamaysky, and Jimmy Qin (2023), "New News is Bad News," Working Paper.

Gounopoulos, Dimitrios and Yu Zhang (2024), "Temperature Trend and Corporate Cash Holdings," *Finaical Management*, pages 1–24.

Ilhan, Emirhan (2022), "Sea Level Rise and Portfolio Choice," Working Paper.

Ilhan, Emirhan, Philipp Krueger, Zacharias Sautner, and Laura T. Starks (2023), "Climate Risk Disclosure and Institutional Investors," *Review of Financial Studies*, vol. 36, 2617–2650.

Ilhan, Emirhan, Zacharias Sautner, and Grigory Vilkov (2021), "Carbon Tail Risk," *Review of Financial Studies*, vol. 34, 15401571.

Itoh, Satoshi and Katsuhiko Okada (2023), "The Power of Large Language Models: A ChatGPT-driven Textual Analysis of Fundamental Data," Working Paper.

Jha, Manish, Jialin Qian, Michael Weber, and Baozhong Yang (2024), "ChatGPT and Corporate Policies," Working Paper.

Kim, Alex G., Maximilian Muhn, and Valeri V. Nikolaev (2024a), "Bloated Disclosures: Can ChatGPT Help Investors Process Information?" Working Paper.

Kim, Alex G., Maximilian Muhn, and Valeri V. Nikolaev (2024b), "Financial Statement Analysis with Large Language Models," Working Paper.

Kim, Alex G. and Valeri Nikolaev (2023), "Context-Based Interpretation of Financial Information," Working Paper.

Kim, Alex G. and Valeri Nikolaev (2024), "Profitability Context and the Cross-Section of Stock Returns," Working Paper.

Kölbel, Julian F., Markus Leippold, Jordy Rillaerts, and Qian Wang (2024), "Ask BERT: How Regulatory Disclosure of Transition and Physical Climate Risks Affects the CDS Term Structure," *Journal of Financial Econometrics*, vol. 22, 30–69.

Li, Kai, Feng Mai, Rui Shen, Chelsea Yang, and Tengfei Zhang (2024a), "Dissecting Corporate Culture Using Generative AI–Insights from Analyst Reports," Working Paper.

Li, Qing, Hongyu Shan, Yuehua Tang, and Vincent Yao (2024b), "Corporate Climate Risk: Measurements and Responses," vol. 00, 1–53.

Lin, Chen, Thomas Schmid, and Michael S. Weisbach (2023), "Price Risk, Production Flexibility, and Liquidity Management: Evidence from Electricity Generating Firms," Working Paper.

LoGrasso, Marc F. (2024), "Could ChatGPT Have Earned Abnormal Returns?" Working Paper.

Lopez-Lira, Alejandro and Yuehua Tang (2023), "Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models," Working Paper.

Nakano, Masafumi and Takuya Yamaoka (2023), "Enhancing Sentiment Analysis based Investment by Large Language Models in Japanese Stock Market," Working Paper.

Pankratz, Nora, Rob Bauer, and Jeroen Derwall (2022), "Climate Change, Firm Performance, and Investor Surprises," Working Paper.

Pankratz, Nora M. C. and Christoph M. Schiller (2024), "Climate Change and Adaptation in Global Supply-Chain Networks," *Review of Financial Studies*, vol. 00, 149.

Pelster, Matthias and Joel Val (2023), "Can ChatGPT assist in picking stocks?" Working Paper.

Sautner, Zacharias, Laurence Van Lent, Grigory Vilkov, and Ruishen Zhang (2023a), "Firm-Level Climate Change Exposure," *Journal of Finance*, vol. 78, 1449–1498.

Sautner, Zacharias, Laurence Van Lent, Grigory Vilkov, and Ruishen Zhang (2023b), "Pricing Climate Change Exposure," *Management Science*, vol. 69, 75407561.

Shah, Agam, Suvan Paturi, and Sudheer Chava (2024), "Trillion Dollar Words: A New Financial Dataset, Task & Market Analysis," Working Paper.

Tan, Lin, Huihang Wu, and Xiaoyan Zhang (2024), "Large Language Models and Return Prediction in China," Working Paper.

Wang, Lulu and Aifan Ling (2024), "Fund Performance Driven by ChatGPT: Evidence from Chinese Fund Market," Working Paper.

Figure 1: Storm Event Count

This figure plots the storm event count from March 2022 to December 2023.

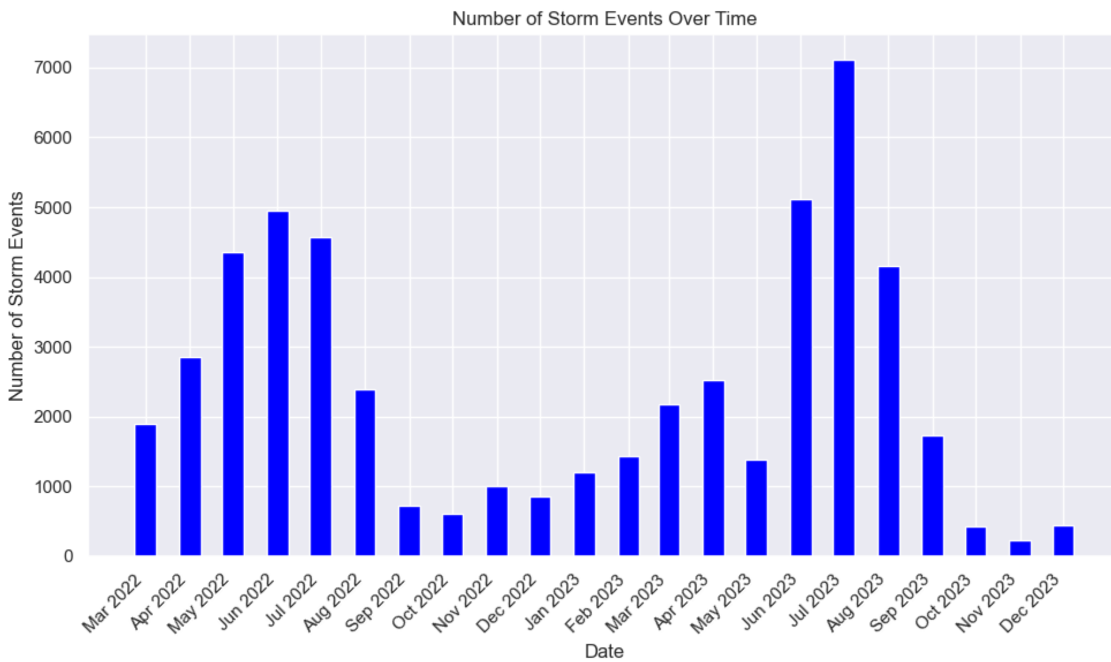

Figure 2: The Word Cloud Graph for Episode Narrative

The image depicts a visual portrayal of the words found within the text data from the *episode_narrative* field. The prominence or visibility of each word in the visualization corresponds to its frequency of occurrence throughout the specified text corpus.

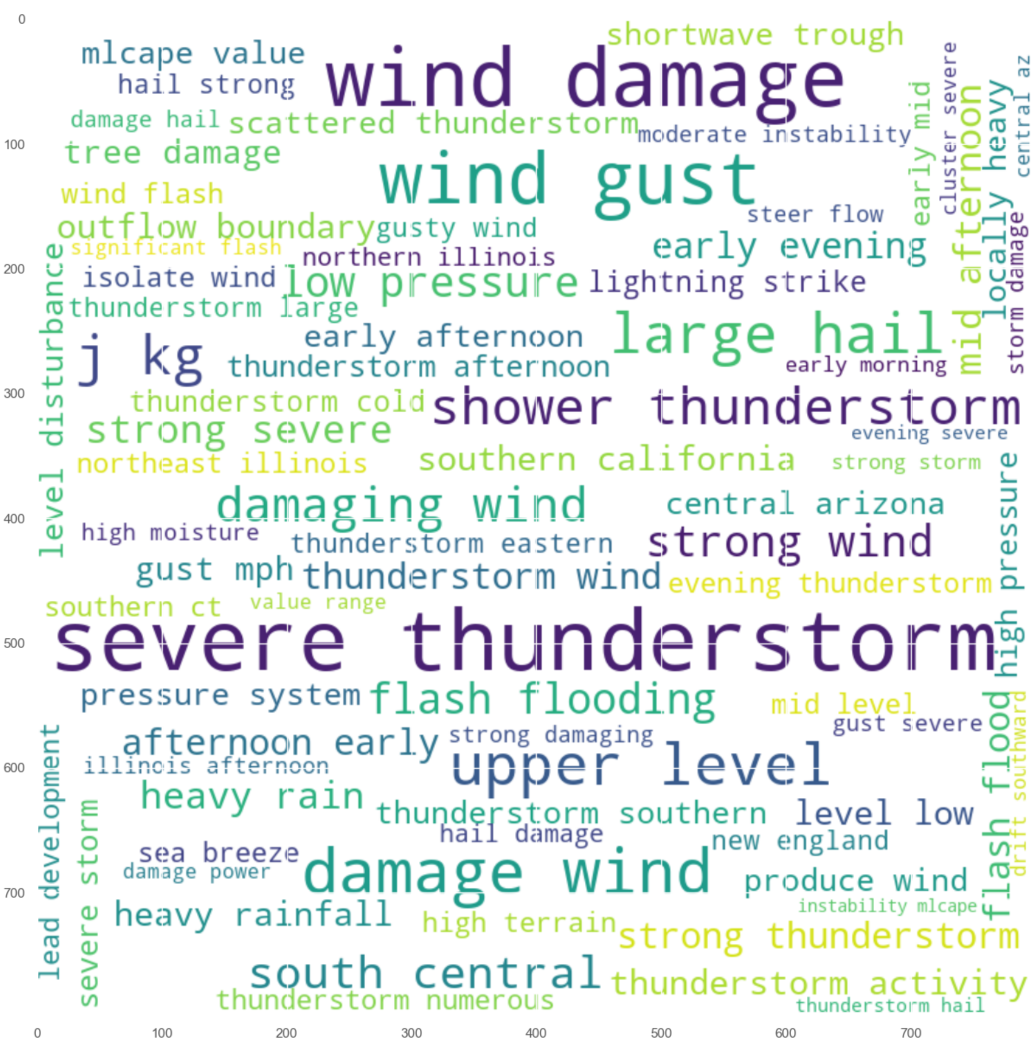

Table 1: **Storm Event Distribution by State**

This table presents the storm event distribution by state. The sample period is from March 2022 to December 2023.

State	Number of Storm Events	Percentage
Texas	2,803	5.37%
Georgia	2,662	5.10%
Virginia	1,914	3.67%
Illinois	1,897	3.64%
Kansas	1,845	3.54%
Kentucky	1,758	3.37%
North Carolina	1,737	3.33%
Pennsylvania	1,685	3.23%
Alabama	1,661	3.18%
New York	1,626	3.12%
South Dakota	1,444	2.77%
Tennessee	1,383	2.65%
Ohio	1,379	2.64%
South Dakota	1,329	2.55%
Nebraska	1,250	2.40%
Oklahoma	1,228	2.35%
Mississippi	1,194	2.29%
Missouri	1,172	2.25%
California	1,154	2.21%
Florida	1,151	2.21%
Minnesota	1,075	2.06%
Indiana	1,053	2.02%
Montana	1,023	1.96%
Others	16,743	32.10%
Total	52,166	100.00%

Table 2: **Descriptive Analysis**

This table presents the summary statistics. In Panel A, "Episode Description Length" is the length of the episode description. "Episode ChatGPT Score" is the ChatGPT score which is 1 if ChatGPT's response is YES; 0 if ChatGPT's response is NO. In Panels B and C, *Magnitude* is the weather event's severity scale for wind speeds which is measured in knots. $Damage_{property}$ and $Damage_{crops}$ are the damage to property and crops in dollars caused by the weather event, respectively. $Deaths_{direct}$ and $Deaths_{indirect}$ are the number of deaths directly and indirectly attributed to the weather event, respectively. *Death* is the total number of deaths, combining both direct and indirect fatalities. $Injuries_{direct}$ and $Injuries_{indirect}$ are the number of injuries directly and indirectly attributed to the weather event, respectively. *Injury* is the total number of injuries, combining both direct and indirect injuries. For each county and each year-month, the values for $Damage_{property}$, $Damage_{crops}$, *Death*, and *Injury* are aggregated by summing their values. The sample period is from March 2022 to December 2023.

Panel A

	N	Mean	Std	Min	25%	50%	75%	Max
Episode Description Length	644	336	349	44	159	244	400	5,794
Episode ChatGPT Score	644	0.258	0.438	0	0	0	1	1

Panel B

	N	Mean	Std	Min	25%	50%	75%	Max
$Magnitude$	3,682	54.81	8.43	35.00	50.00	52.00	59.00	96.00
$Log(Damage_{property}+1)$	3,682	5.35	4.88	0.00	0.00	7.60	9.62	18.20
$Log(Damage_{crops}+1)$	3,682	0.04	0.60	0.00	0.00	0.00	0.00	13.30
$Log(Death+1)$	3,682	0.02	0.12	0.00	0.00	0.00	0.00	1.79
$Log(Injury+1)$	3,682	0.04	0.19	0.00	0.00	0.00	0.00	1.61

Panel C

Variable	$Magnitude$	$Log(Damage_{crops}+1)$	$Log(Damage_{property}+1)$	$Log(Death+1)$	$Log(Injury+1)$
$Magnitude$	1.00				
$Log(Damage_{crops}+1)$	0.05	1.00			
$Log(Damage_{property}+1)$	0.02	0.04	1.00		
$Log(Death+1)$	-0.07	-0.01	-0.12	1.00	
$Log(Injury+1)$	0.18	0.02	-0.14	0.56	1.00

Table 3: **Return Prediction**

This table reports the return prediction using OLS regression. The dependent variables are the stock returns over various future time horizons: the next trading day (Model 1), the next week (Model 2), the next two weeks (Model 3), the next three weeks (Model 4), and the next month (Model 5). The key independent variable of interest is a dummy variable that takes the value of 1 if ChatGPT recommends "YES", and 0 otherwise. The regression includes control variables such as the return on the event day, the magnitude of the storm event, property damage, crop damage, and the number of deaths and injuries caused by the storm event. Additionally, we control for date and firm fixed effects. Standard errors are clustered at the firm level to account for potential within-firm correlation. The sample period is from March 2022 to December 2023. t-statistics are shown in parentheses. $*p < 0.1; **p < 0.05; ***p < 0.01$.

	Model 1	Model 2	Model 3	Model 4	Model 5
	r_{t+1}	r_{t+7}	r_{t+14}	r_{t+21}	r_{t+30}
$I_{ChatGPT-YES}$	−0.002**	−0.014***	−0.027***	−0.027***	−0.024***
	(−2.09)	(−4.72)	(−6.01)	(−4.86)	(−3.18)
r_t	0.019	0.002	−0.095	0.065	−0.147
	(0.91)	(0.03)	(−1.20)	(0.69)	(−0.89)
$Magnitude$	0.000	0.001	0.001	0.000	0.000
	(1.52)	(1.52)	(1.08)	(1.28)	(1.25)
$Log(Damage_{property} + 1)$	0.000	−0.001***	0.000	−0.001**	−0.001*
	(1.62)	(−2.78)	(0.24)	(−2.37)	(−1.78)
$Log(Damage_{crops} + 1)$	0.000	−0.002	−0.002	0.005	0.012
	(0.05)	(−0.81)	(−0.66)	(1.20)	(1.26)
$Log(Death + 1)$	−0.005	0.004	−0.018	0.029	−0.025
	(−1.17)	(0.31)	(−0.91)	(1.10)	(−0.64)
$Log(Injury + 1)$	0.001	0.002	0.018	0.001	−0.012
	(0.33)	(0.20)	(1.63)	(0.05)	(−0.65)
$Constant$	−0.013***	−0.022**	0.002	−0.042**	−0.067***
	(−3.63)	(−2.23)	(0.11)	(−2.33)	(−2.66)
Date FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
N	3,467	3,421	3,419	3,410	3,342
$Adjusted\ R^2$	5.60%	11.30%	15.50%	19.80%	18.10%

Table 4: **Return Prediction for Firms with Large Market Caps**

This table reports the return prediction using OLS regression for the subsample of firms whose market capitalization exceeds the mean market capitalization across the full sample. The market capitalization is calculated by multiplying the stock price by the number of publicly traded shares, recorded in thousands. The dependent variables are the stock returns over various future time horizons: the next trading day (Model 1), the next week (Model 2), the next two weeks (Model 3), the next three weeks (Model 4), and the next month (Model 5). The key independent variable of interest is a dummy variable that takes the value of 1 if ChatGPT recommends "YES", and 0 otherwise. The regression includes control variables such as the return on the event day, the magnitude of the storm event, property damage, crop damage, and the number of deaths and injuries caused by the storm event. Additionally, we control for date and firm fixed effects. Standard errors are clustered at the firm level to account for potential within-firm correlation. The sample period is from March 2022 to December 2023. t-statistics are shown in parentheses. $*p < 0.1; **p < 0.05; ***p < 0.01$.

	Model 1	Model 2	Model 3	Model 4	Model 5
	r_{t+1}	r_{t+7}	r_{t+14}	r_{t+21}	r_{t+30}
$I_{ChatGPT-YES}$	−0.003**	−0.016***	−0.028***	−0.031***	−0.030***
	(−2.23)	(−4.36)	(−5.43)	(−4.89)	(−3.34)
r_t	0.031	−0.057	−0.097	0.187	−0.209
	(1.11)	(−0.83)	(−1.05)	(1.54)	(−0.92)
$Magnitude$	0.000	0.000	0.000	0.000	0.001
	(1.35)	(1.59)	(1.17)	(1.33)	(1.22)
$Log(Damage_{property} + 1)$	0.000	0.000	0.000	−0.001	−0.001
	(1.19)	(1.49)	(0.60)	(−1.25)	(−0.91)
$Log(Damage_{crops} + 1)$	0.000	−0.003*	−0.006***	0.000	−0.001
	(1.12)	(−1.67)	(−3.02)	(0.41)	(−0.24)
$Log(Death + 1)$	0.001	0.001	−0.018	0.008	−0.038
	(0.28)	(0.07)	(−0.99)	(0.29)	(−1.01)
$Log(Injury + 1)$	0.000	−0.003	0.015	−0.004	−0.001
	(0.07)	(−0.25)	(1.28)	(−0.28)	(−0.04)
$Constant$	−0.007*	−0.014	0.009	−0.035*	−0.073**
	(−1.74)	(−1.22)	(0.53)	(−1.74)	(−2.32)
Date FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
N	2,204	2,154	2,150	2,139	2,106
$Adjusted\ R^2$	7.80%	16.60%	20.90%	25.70%	20.90%

Table 5: **Return Prediction for Firms with Low Debt-to-Equity Ratio**

This table reports the return prediction using OLS regression for the subsample of firms with a debt-to-equity ratio below the mean debt-to-equity ratio of the full sample. The dependent variables are the stock returns over various future time horizons: the next trading day (Model 1), the next week (Model 2), the next two weeks (Model 3), the next three weeks (Model 4), and the next month (Model 5). The key independent variable of interest is a dummy variable that takes the value of 1 if ChatGPT recommends "YES", and 0 otherwise. The regression includes control variables such as the return on the event day, the magnitude of the storm event, property damage, crop damage, and the number of deaths and injuries caused by the storm event. Additionally, we control for date and firm fixed effects. Standard errors are clustered at the firm level to account for potential within-firm correlation. The sample period is from March 2022 to December 2023. t-statistics are shown in parentheses. $*p < 0.1; **p < 0.05; ***p < 0.01$.

	Model 1	Model 2	Model 3	Model 4	Model 5
	r_{t+1}	r_{t+7}	r_{t+14}	r_{t+21}	r_{t+30}
$I_{ChatGPT-YES}$	−0.002*	−0.014***	−0.026***	−0.023***	−0.025***
	(−1.81)	(−4.59)	(−5.71)	(−4.02)	(−3.24)
r_t	0.03	−0.019	−0.097	0.085	−0.164
	(1.41)	(−0.34)	(−1.18)	(0.90)	(−0.96)
$Magnitude$	0.000	0.001	0.001	0.000	0.000
	(1.48)	(1.18)	(1.16)	(0.92)	(0.56)
$Log(Damage_{property} + 1)$	0.000	−0.001***	0.000	−0.001**	−0.002*
	(0.99)	(2.83)	(−1.14)	(−2.32)	(−1.84)
$Log(Damage_{crops} + 1)$	0.000	−0.002	−0.003	0.004	0.011
	(0.49)	(−0.88)	(0.98)	(0.75)	(1.19)
$Log(Death + 1)$	−0.006	0.016	−0.008	0.026	−0.010
	(−1.26)	(0.97)	(−0.40)	(0.96)	(−0.24)
$Log(Injury + 1)$	0.002	0.001	0.017	0.008	-0.007
	(0.69)	(0.12)	(1.53)	(0.53)	(0.42)
$Constant$	−0.013***	−0.071***	−0.087***	−0.065***	−0.100**
	(−2.97)	(−6.26)	(−5.21)	(−2.74)	(−2.56)
Date FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
N	2,969	2,924	2,927	2,916	2,858
$Adjusted\ R^2$	2.80%	9.50%	12.80%	17.70%	14.50%

Table 6: **Return Prediction for Firms with High ROA**

This table reports the return prediction using OLS regression for the subsample of firms with a return-to-assets ratio above the mean return-to-assets ratio of the full sample. The dependent variables are the stock returns over various future time horizons: the next trading day (Model 1), the next week (Model 2), the next two weeks (Model 3), the next three weeks (Model 4), and the next month (Model 5). The key independent variable of interest is a dummy variable that takes the value of 1 if ChatGPT recommends "YES", and 0 otherwise. The regression includes control variables such as the return on the event day, the magnitude of the storm event, property damage, crop damage, and the number of deaths and injuries caused by the storm event. Additionally, we control for date and firm fixed effects. Standard errors are clustered at the firm level to account for potential within-firm correlation. The sample period is from March 2022 to December 2023. t-statistics are shown in parentheses. $*p < 0.1; **p < 0.05; ***p < 0.01$.

	Model 1	Model 2	Model 3	Model 4	Model 5
	r_{t+1}	r_{t+7}	r_{t+14}	r_{t+21}	r_{t+30}
$I_{ChatGPT-YES}$	−0.002*	−0.011***	−0.026***	−0.027***	−0.023***
	(−1.65)	(−3.64)	(−5.76)	(−5.26)	(−3.24)
r_t	−0.016	0.002	−0.095	0.086	−0.335*
	(−0.61)	(0.02)	(−1.05)	(0.75)	(−1.66)
$Magnitude$	0.000	0.000	0.000	0.000	0.000
	(1.29)	(1.43)	(1.17)	(1.13)	(1.18)
$Log(Damage_{property} + 1)$	0.000	−0.001***	0.000	−0.001**	−0.002**
	(1.06)	(−3.71)	(0.37)	(−2.58)	(−2.09)
$Log(Damage_{crops} + 1)$	0.000	−0.001	−0.004	0.003	0.007
	(0.66)	(−0.26)	(−0.89)	(0.51)	(1.35)
$Log(Death + 1)$	−0.001	0.014	−0.014	0.036	−0.035
	(−0.19)	(0.99)	(−0.81)	(1.35)	(−1.14)
$Log(Injury + 1)$	0.001	−0.001	0.019*	−0.001	−0.004
	(0.23)	(−0.10)	(1.71)	(−0.07)	(−0.22)
$Constant$	−0.014***	−0.012	0.016	−0.035*	−0.062**
	(−4.04)	(−1.14)	(0.99)	(−1.95)	(−2.40)
Date FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
N	2,544	2,517	2,516	2,513	2,478
$Adjusted\ R^2$	3.80%	9.10%	13.10%	18.80%	16.70%

Table 7: **Return Prediction for Firms with Narrow Bid-Ask Spreads**

This table reports the return prediction using OLS regression for the subsample of firms with a bid-ask spread below the mean bid-ask spread of the full sample. The dependent variables are the stock returns over various future time horizons: the next trading day (Model 1), the next week (Model 2), the next two weeks (Model 3), the next three weeks (Model 4), and the next month (Model 5). The key independent variable of interest is a dummy variable that takes the value of 1 if ChatGPT recommends "YES", and 0 otherwise. The regression includes control variables such as the return on the event day, the magnitude of the storm event, property damage, crop damage, and the number of deaths and injuries caused by the storm event. Additionally, we control for date and firm fixed effects. Standard errors are clustered at the firm level to account for potential within-firm correlation. The sample period is from March 2022 to December 2023. t-statistics are shown in parentheses. $*p < 0.1; **p < 0.05; ***p < 0.01$.

	Model 1	Model 2	Model 3	Model 4	Model 5
	r_{t+1}	r_{t+7}	r_{t+14}	r_{t+21}	r_{t+30}
$I_{ChatGPT-YES}$	−0.001	−0.011***	−0.027***	−0.026***	−0.023***
	(−1.37)	(−3.62)	(−5.91)	(−4.79)	(−3.06)
r_t	−0.011	−0.043	−0.034	0.083	−0.099
	(−0.46)	(−0.66)	(−0.35)	(0.78)	(−0.50)
$Magnitude$	0.000	0.000	0.000	0.000	0.000
	(1.14)	(1.19)	(1.15)	(1.12)	(0.80)
$Log(Damage_{property} + 1)$	0.000	−0.001***	−0.001**	−0.001***	−0.002*
	(0.75)	(−4.00)	(−2.13)	(−2.75)	(−1.75)
$Log(Damage_{crops} + 1)$	0.000	−0.001	−0.001	0.007	0.004
	(0.69)	(−0.30)	(−0.27)	(1.64)	(0.65)
$Log(Death + 1)$	−0.007	0.009	−0.027	0.019	−0.051
	(−1.59)	(0.63)	(−1.43)	(0.72)	(−1.35)
$Log(Injury + 1)$	0.003	−0.002	0.023**	0.010	0.006
	(1.03)	(−0.17)	(2.12)	(0.76)	(0.49)
$Constant$	−0.010***	−0.010	0.019	−0.038**	−0.066***
	(−2.78)	(−0.92)	(1.17)	(−2.04)	(−2.60)
Date FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
N	2,591	2,561	2,562	2,555	2,516
$Adjusted\ R^2$	3.70%	10.70%	14.90%	21.20%	18.10%

Table 8: **Return Predictions for Firms with the Highest Third of Storm Event Frequency**

This table reports the return prediction using OLS regression for the subsample of firms the highest third of storm event frequency. The dependent variables are the stock returns over various future time horizons: the next trading day (Model 1), the next week (Model 2), the next two weeks (Model 3), the next three weeks (Model 4), and the next month (Model 5). The key independent variable of interest is a dummy variable that takes the value of 1 if ChatGPT recommends "YES", and 0 otherwise. The regression includes control variables such as the return on the event day, the magnitude of the storm event, property damage, crop damage, and the number of deaths and injuries caused by the storm event. Additionally, we control for date and firm fixed effects. Standard errors are clustered at the firm level to account for potential within-firm correlation. The sample period is from March 2022 to December 2023. t-statistics are shown in parentheses. $*p < 0.1; **p < 0.05; ***p < 0.01$.

	Model 1	Model 2	Model 3	Model 4	Model 5
	r_{t+1}	r_{t+7}	r_{t+14}	r_{t+21}	r_{t+30}
$I_{ChatGPT-YES}$	−0.006***	−0.019***	−0.027***	−0.044***	−0.040***
	(−2.86)	(−3.40)	(−3.40)	(−4.68)	(−3.40)
r_t	−0.065*	−0.055	−0.182	−0.158	−0.459
	(−1.82)	(−0.61)	(−1.47)	(−1.19)	(−1.64)
$Magnitude$	0.000	0.001	0.001	0.001	0.002
	(0.19)	(1.64)	(1.41)	(1.35)	(1.40)
$Log(Damage_{property} + 1)$	0.000**	−0.001**	0.002**	0.000	−0.002
	(2.10)	(−2.06)	(2.50)	(0.20)	(−1.49)
$Log(Damage_{crops} + 1)$	0.001	−0.003	−0.011***	−0.007**	0.010***
	(1.17)	(−1.31)	(−5.50)	(−2.54)	(3.91)
$Log(Death + 1)$	0.004	0.066**	0.034	0.134	0.165
	(0.27)	(2.45)	(0.68)	(1.34)	(1.32)
$Log(Injury + 1)$	0.016***	−0.005	0.002	−0.029	−0.041*
	(3.78)	(−0.36)	(0.14)	(−1.37)	(−1.66)
$Constant$	−0.014**	−0.021	−0.025	−0.100***	−0.145***
	(−2.33)	(−1.02)	(−0.88)	(−3.16)	(−3.36)
Date FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
N	1,468	1,479	1,479	1,476	1,441
$Adjusted\ R^2$	17.00%	19.90%	22.50%	26.30%	23.40%

Table 9: **Return Prediction During Market Uptrend Periods**

This table reports the return prediction using OLS regression for the subsample of periods when the S&P 500 level is above its mean during the sample period. The dependent variables are the stock returns over various future time horizons: the next trading day (Model 1), the next week (Model 2), the next two weeks (Model 3), the next three weeks (Model 4), and the next month (Model 5). The key independent variable of interest is a dummy variable that takes the value of 1 if ChatGPT recommends "YES", and 0 otherwise. The regression includes control variables such as the return on the event day, the magnitude of the storm event, property damage, crop damage, and the number of deaths and injuries caused by the storm event. Additionally, we control for date and firm fixed effects. Standard errors are clustered at the firm level to account for potential within-firm correlation. The sample period is from March 2022 to December 2023. t-statistics are shown in parentheses. $*p < 0.1; **p < 0.05; ***p < 0.01$.

	Model 1	Model 2	Model 3	Model 4	Model 5
	r_{t+1}	r_{t+7}	r_{t+14}	r_{t+21}	r_{t+30}
$I_{ChatGPT-YES}$	−0.003**	−0.019***	−0.035***	−0.038***	−0.023***
	(−2.32)	(−5.47)	(−6.66)	(−5.91)	(−2.84)
r_t	−0.038	−0.052	−0.232**	−0.004	−0.303
	(−1.24)	(−0.70)	(−2.09)	(−0.03)	(−1.19)
$Magnitude$	0.000	0.000	0.000	0.000	0.001
	(0.19)	(1.17)	(1.13)	(0.52)	(1.18)
$Log(Damage_{property} + 1)$	0.000	0.000	0.000	−0.001	−0.002*
	(1.37)	(1.30)	(0.96)	(−1.14)	(−1.92)
$Log(Damage_{crops} + 1)$	0.000	0.002	0.008	0.010	0.026
	(0.29)	(0.96)	(1.54)	(1.63)	(1.42)
$Log(Death + 1)$	0.001	0.004	0.019	0.051	−0.008
	(0.30)	(0.22)	(0.66)	(0.96)	(−0.13)
$Log(Injury + 1)$	0.000	0.010	−0.015	−0.022	−0.061
	(0.03)	(0.59)	(−0.55)	(−0.45)	(−0.98)
$Constant$	0.003	−0.032***	−0.032*	−0.027	−0.019
	(0.68)	(−2.68)	(−1.82)	(−1.18)	(−0.64)
Date FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
N	2,082	2,030	2,028	2,028	1,990
$Adjusted\ R^2$	6.90%	13.10%	15.10%	19.50%	20.20%

Table 10: **Return Prediction During Market Low VIX Periods**

This table reports the return prediction using OLS regression for the subsample of periods when the VIX level is below its mean during the sample period. The dependent variables are the stock returns over various future time horizons: the next trading day (Model 1), the next week (Model 2), the next two weeks (Model 3), the next three weeks (Model 4), and the next month (Model 5). The key independent variable of interest is a dummy variable that takes the value of 1 if ChatGPT recommends "YES", and 0 otherwise. The regression includes control variables such as the return on the event day, the magnitude of the storm event, property damage, crop damage, and the number of deaths and injuries caused by the storm event. Additionally, we control for date and firm fixed effects. Standard errors are clustered at the firm level to account for potential within-firm correlation. The sample period is from March 2022 to December 2023. t-statistics are shown in parentheses. $*p < 0.1; **p < 0.05; ***p < 0.01$.

	Model 1	Model 2	Model 3	Model 4	Model 5
	r_{t+1}	r_{t+7}	r_{t+14}	r_{t+21}	r_{t+30}
$I_{ChatGPT-YES}$	−0.003**	−0.019***	−0.034***	−0.036***	−0.018**
	(−2.46)	(−5.79)	(−6.47)	(−5.77)	(−2.10)
r_t	−0.029	−0.054	−0.233**	−0.019	−0.285
	(−0.97)	(−0.75)	(−2.15)	(−0.14)	(1.11)
$Magnitude$	0.000	0.000	0.000	0.000	0.000
	(0.40)	(1.25)	(1.55)	(0.48)	(0.88)
$Log(Damage_{property} + 1)$	0.000	0.000	0.000	−0.001	−0.002*
	(1.07)	(1.27)	(1.00)	(−0.96)	(−1.76)
$Log(Damage_{crops} + 1)$	0.000	0.002	0.008	0.010	0.026
	(0.28)	(0.96)	(1.62)	(1.61)	(1.39)
$Log(Death + 1)$	0.010*	0.011	0.022	0.055	−0.072
	(1.84)	(0.62)	(0.82)	(1.11)	(−1.22)
$Log(Injury + 1)$	−0.010**	0.002	−0.02	−0.029	−0.003
	(−2.06)	(0.10)	(−0.89)	(−0.65)	(−0.05)
$Constant$	−0.011**	−0.059***	−0.072***	−0.053*	−0.086**
	(−2.25)	(−4.52)	(−3.54)	(−1.90)	(−2.01)
Date FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
N	2,163	2,110	2,109	2,107	2,069
$Adjusted\ R^2$	7.50%	13.80%	16.30%	19.30%	20.00%

Table 11: **Return Prediction During High Total Non-farm Payroll Growth Periods**

This table reports the return prediction using OLS regression for the subsample of periods when the Total Non-farm Payroll Growth Rate is above its mean during the sample period. The dependent variables are the stock returns over various future time horizons: the next trading day (Model 1), the next week (Model 2), the next two weeks (Model 3), the next three weeks (Model 4), and the next month (Model 5). The key independent variable of interest is a dummy variable that takes the value of 1 if ChatGPT recommends "YES", and 0 otherwise. The regression includes control variables such as the return on the event day, the magnitude of the storm event, property damage, crop damage, and the number of deaths and injuries caused by the storm event. Additionally, we control for date and firm fixed effects. Standard errors are clustered at the firm level to account for potential within-firm correlation. The sample period is from March 2022 to December 2023. t-statistics are shown in parentheses. $*p < 0.1; **p < 0.05; ***p < 0.01$.

	Model 1	Model 2	Model 3	Model 4	Model 5
	r_{t+1}	r_{t+7}	r_{t+14}	r_{t+21}	r_{t+30}
$I_{ChatGPT-YES}$	−0.003**	−0.019***	−0.035***	−0.037***	−0.023***
	(−2.42)	(−5.47)	(−6.61)	(−5.87)	(−2.80)
r_t	−0.04	−0.05	−0.229**	−0.015	−0.295
	(−1.31)	(−0.69)	(−2.07)	(−0.11)	(−1.16)
$Magnitude$	0.000	0.000	0.000	0.000	0.001
	(0.02)	(1.21)	(1.59)	(0.52)	(1.19)
$Log(Damage_{property} + 1)$	0.000	0.000	0.000	−0.001	−0.002*
	(1.57)	(1.35)	(0.95)	(−1.12)	(−1.90)
$Log(Damage_{crops} + 1)$	0.000	0.002	0.008	0.010	0.026
	(0.28)	(0.96)	(1.54)	(1.63)	(1.42)
$Log(Death + 1)$	0.001	0.004	0.019	0.051	−0.009
	(0.32)	(0.21)	(0.65)	(0.95)	(−0.13)
$Log(Injury + 1)$	0.000	0.010	−0.015	−0.022	−0.061
	(0.06)	(0.59)	(−0.55)	(−0.44)	(−0.99)
$Constant$	0.004	0.01	0.057**	0.052*	0.011
	(0.45)	(0.50)	(2.00)	(1.79)	(0.32)
Date FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
N	2,094	2,043	2,041	2,041	2,002
$Adjusted\ R^2$	6.90%	13.00%	15.20%	19.50%	20.20%

Table 12: **Return Prediction During Low Smoothed Recession Probability Periods**

This table reports the return prediction using OLS regression for the subsample of periods when the Smoothed Recession Probability is below its mean during the sample period. The dependent variables are the stock returns over various future time horizons: the next trading day (Model 1), the next week (Model 2), the next two weeks (Model 3), the next three weeks (Model 4), and the next month (Model 5). The key independent variable of interest is a dummy variable that takes the value of 1 if ChatGPT recommends "YES", and 0 otherwise. The regression includes control variables such as the return on the event day, the magnitude of the storm event, property damage, crop damage, and the number of deaths and injuries caused by the storm event. Additionally, we control for date and firm fixed effects. Standard errors are clustered at the firm level to account for potential within-firm correlation. The sample period is from March 2022 to December 2023. t-statistics are shown in parentheses. $*p < 0.1; **p < 0.05; ***p < 0.01$.

	Model 1	Model 2	Model 3	Model 4	Model 5
	r_{t+1}	r_{t+7}	r_{t+14}	r_{t+21}	r_{t+30}
$I_{ChatGPT-YES}$	−0.002**	−0.014***	−0.029***	−0.029***	−0.029***
	(−2.22)	(−4.71)	(−6.42)	(−5.14)	(−3.89)
r_t	0.014	0.004	−0.106	0.048	−0.197
	(0.63)	(0.07)	(−1.29)	(0.51)	(−1.20)
$Magnitude$	0.000	0.000	0.001	0.000	0.001
	(1.44)	(1.23)	(1.01)	(1.54)	(1.64)
$Log(Damage_{property} + 1)$	0.000	−0.001***	0.000	−0.001**	−0.001*
	(1.26)	(−3.12)	(0.36)	(−2.36)	(−1.68)
$Log(Damage_{crops} + 1)$	0.000	−0.002	−0.002	0.005	0.012
	(0.12)	(−0.78)	(−0.63)	(1.20)	(1.27)
$Log(Death + 1)$	−0.007	0.003	−0.018	0.026	−0.01
	(−1.45)	(0.21)	(−0.90)	(0.98)	(−0.25)
$Log(Injury + 1)$	0.003	0.003	0.020*	0.005	−0.022
	(1.02)	(0.34)	(1.74)	(0.32)	(−1.26)
$Constant$	−0.013***	−0.020*	0.003	−0.041**	−0.080***
	(−3.64)	(−1.92)	(0.19)	(−2.23)	(−3.16)
Date FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
N	3,248	3,256	3,252	3,247	3,181
$Adjusted\ R^2$	4.50%	10.90%	14.70%	19.20%	17.60%