



**HAL**  
open science

# Sentiment Analysis of Bloomberg Markets Wrap Using ChatGPT: Application to the NASDAQ

Baptiste Lefort, Eric Benhamou, Jean-Jacques Ohana, David Saltiel, Beatrice Guez, Thomas Jacquot

► **To cite this version:**

Baptiste Lefort, Eric Benhamou, Jean-Jacques Ohana, David Saltiel, Beatrice Guez, et al.. Sentiment Analysis of Bloomberg Markets Wrap Using ChatGPT: Application to the NASDAQ. 2024. hal-04739924

**HAL Id: hal-04739924**

**<https://hal.science/hal-04739924v1>**

Preprint submitted on 16 Oct 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Sentiment Analysis of Bloomberg Markets Wrap Using ChatGPT: Application to the NASDAQ

B.Lefort, E.Benhamou, JJ.Ohana, D. Saltiel, B. Guez, T.Jacquot

Ai For Alpha

France

{baptiste.lefort, eric.benhamou,jean-jacques.ohana,  
david.saltiel,beatrice.guez,thomas.jacquot}@aiforalpha.com

## Abstract

In this paper, we use a comprehensive dataset of daily Bloomberg Financial Market Summaries spanning from 2010 to 2023, published by Yahoo Finance, CNews and multiple large financial medias, to determine how global news headlines may affect stock market movements. To make this analysis more effective, we employed ChatGPT. First, from the vast pool of daily financial updates, we identify the top global news headlines that could potentially have a significant influence on stock prices. Second, for each headline, we question ChatGPT to answer whether the news might lead to a rise, fall in stock prices or is indecisive. This two-stage method proves more effective than posing a direct question to the entire text. By gathering ChatGPT's predictions day by day, we formed an overall *market sentiment* score. We transform this score into a practical investment strategy in the NASDAQ index, demonstrating the significance of minimizing noise in sentiment scores by initially accumulating and then detrending them. This approach showcases that ChatGPT's analysis of news headlines can provide valuable insights into future stock market behaviors and be a valuable tool to develop intuitive NLP-driven investment strategies leveraging news predictive power.

**Keywords:** sentiment analysis, ChatGPT, stock exchange, financial news

## 1. Introduction

Finance has a longstanding tradition of employing Natural Language Processing (NLP) to extract valuable insights from textual data and news (Tetlock, 2007; Schumaker and Chen, 2009). The financial world has always been at the forefront of embracing technological innovation. From the inception of ATMs and electronic trading to the burgeoning realm of fintech, financial services have undergone significant evolution, especially with the infusion of AI and ML technologies (Arner et al., 2015; Fatouros et al., 2023).

Sentiment analysis stands out as a cornerstone in this transformation (Poria et al., 2016). It plays a crucial role in deciphering market sentiments, offering invaluable predictive insights. Historically, the financial sector leaned on handpicked word lists and basic ML techniques for sentiment analysis (Tetlock, 2007; Schumaker and Chen, 2009). Yet, with NLP's rapid advancements, a slew of advanced methods has come to the fore. Models like BERT and its finance-centric sibling, FinBERT, have elevated sentiment analysis's precision (Devlin et al., 2018; Liu et al., 2021).

However, the financial realm brings its set of challenges for sentiment analysis (Loughran and McDonald, 2011). Financial news is a complex mesh of domain-specific jargons and layered emotions. A singular piece of news might carry different sentiments for multiple financial entities, making general sentiment analysis tools potentially misleading. Furthermore, these tools often struggle

with context-specific outputs, making them less versatile in diverse scenarios (Poria et al., 2017). Indeed, undertaking natural language processing (NLP) in finance is notably challenging due to the specificity of the corpus, as evidenced by diverse studies on financial texts, sentiment lexicons, and financial reports across various languages and financial systems (Li et al., 2022; Moreno-Ortiz et al., 2020; Ghaddar and Langlais, 2020) and can require knowledge graph (Oksanen et al., 2022) or language-specific corpus (Masson and Paroubek, 2020; Jabbari et al., 2020; Zmandar et al., 2022). Converting also a sentiment score into an investment strategy is notably difficult (Yuan et al., 2020; lordache et al., 2022)

With the advent of Large Language Models (LLMs), an AI paradigm has emerged with transformative potential (George and George, 2023). GPT, particularly its conversational variant, ChatGPT, has shown promise in refining financial applications (OpenAI, 2023). By leveraging ChatGPT's prowess in language comprehension, financial entities can enhance their sentiment analysis depth. This proficiency translates to better-informed investment decisions, optimized risk management, and more effective portfolio strategies. Furthermore, ChatGPT's capability to convey intricate financial insights in understandable terms makes it a potential game-changer in democratizing financial knowledge (Yue et al., 2023).

In this study, we design a sentiment analysis of Bloomberg markets wrap news using ChatGPT. Besides, we develop a two-step prompt-based pro-

cess to extract information from text and convert this into a sentiment score. Finally, we convert this score into an investable strategy with two underlying strategies: a long and short strategy. Using the NASDAQ as a benchmark, we evaluate the improvement of our strategy and show improved risk statistics like Sharpe, Sortinor or Calmar ratio versus a passive investment. To sum up, the contributions of this paper are three folds:

1. We designed a two-step ChatGPT based sentiment analysis extraction from Bloomberg markets wrap news.
2. We proposed a method for converting the sentiment score into investable allocations.
3. We show that this approach overperforms the passive investment strategy.

The rest of this paper is organized as follows. Section 2 briefly reviews the related work. Section 4 presents the data collection including time stamp used and data filtering. Section 3 discusses the GPT Capabilities in zero-shot learning. Section 5 explains how proper prompts based on a two steps approach can effectively lead to a robust sentiment score. Section 6 provides details about the conversion into an investable strategy. The evaluations and improvements are presented in Section 7 and finally Section 8 concludes.

## 2. Related works

In the realm of finance and economics, several recent scholarly works have employed ChatGPT, such as Hansen and Kazinnik (2023), Cowen and Tabarrok (2023), Korinek (2023); Lopez-Lira and Tang (2023), and Noy and Zhang (2023). Hansen and Kazinnik (2023) elucidates how Large Language Models (LLMs) like ChatGPT can decipher FedSpeak, the nuanced language employed by the Federal Reserve to convey monetary policy decisions. Lopez-Lira and Tang (2023) explains proper prompting for forecasting stock returns. Both Cowen and Tabarrok (2023) and Korinek (2023) expound upon ChatGPT's utility in economics education and research. Meanwhile, Noy and Zhang (2023) underscores ChatGPT's capability to augment productivity in professional writing tasks. Furthermore, Yang and Menczer (2023) showcases ChatGPT's aptitude for distinguishing credible news outlets.

Simultaneously, research by Xie et al. (2023) posits that ChatGPT's performance is comparable to rudimentary methods like linear regression for numerical data-based prediction tasks. Additionally, Ko and Lee (2023) endeavored to employ ChatGPT in portfolio selection, albeit without discernible success. Our hypothesis attributes these

varied outcomes to their reliance on historical numerical data for prediction, whereas ChatGPT's forte lies in textual tasks.

This research also aligns with the burgeoning literature that amalgamates textual analysis and machine learning to address myriad finance-centric research queries, as evidenced by studies like Jegadeesh and Wu (2013), Rapach et al. (2013), and Campbell et al. (2014), among others.

Our paper offers a novel perspective to this body of literature. It pioneers the assessment of ChatGPT's proficiency in forecasting the NASDAQ trends, a pivotal task for which it has not been explicitly trained, traditionally referred to as zero shot learning. Instead of leveraging finance-specific data, we hinge on ChatGPT's intrinsic NLP capabilities. Moreover, we introduce an innovative prompting method to leverage ChatGPT's analytical processes by finding headlines, then converting these headlines into a sentiment, and finally aggregating carefully these score with both a cumulative sum and a detrending process to filter out noise. Such insights not only augment the nascent literature on deciphering intricate news with LLM models but also differentiate our study from contemporaneous works that use chatGPT in a more brute-force way.

Our paper further complements the literature that harnesses linguistic critiques of news pieces to discern sentiment and subsequently forecast stock returns. Certain studies, like Tetlock (2007) and Garcia (2013), focus on the correlation between media sentiment and overall stock returns. It is in spirit close to (Lopez-Lira and Tang, 2023) but addresses the issue of using news summaries for the NASDAQ index as opposed to company-specific news.

Finally, our paper intersects with literature examining the repercussions of AI-centric technologies on stock market news. Works like Agrawal et al. (2019) and Webb (2019), among others, explore job vulnerabilities and exposures to AI innovations and their implications for employment dynamics and productivity. Amid the AI ascension, our research seeks to evaluate the latent capabilities of AI, particularly LLMs, within the financial sphere.

## 3. Assessing GPT Capabilities

ChatGPT, developed by OpenAI, is rooted in the GPT (Generative Pre-trained Transformer) architecture and represents one of the state of the art achievements in natural language processing (NLP) to date (Vaswani et al., 2017). Leveraging the depth of the Transformer architecture, GPT has been instrumental in reshaping outcomes in numerous NLP tasks. Central to GPT's architecture is its use of the transformer block, characterized by self-

attention mechanisms that highlight salient parts of the input, thereby enriching the model's contextual understanding.

### 3.1. Benchmark with other models

Through rigorous training, ChatGPT has honed its skills across a spectrum of language tasks (Kocorí et al., 2023). Its proficiency in generating human-like text renders it invaluable for applications like chatbots. Notably, ChatGPT, despite its linguistic prowess, hasn't been explicitly trained for financial forecasting, prompting our exploration of its potential in stock return predictions.

Given ChatGPT's extensive training on vast text corpora, it possesses nuanced linguistic comprehension. We postulate that ChatGPT could discern valuable cues from news headlines regarding markets' trajectories, even without specific financial training. Our hypothesis suggests that ChatGPT may surpass traditional sentiment indicators in forecasting returns, benefiting from its refined linguistic capabilities and leveraging the latent potential of zero-shot learning.

Our evaluation aims to assess if ChatGPT can extract predictive sentiment scores from financial markets news summaries.

### 3.2. Zero-shot learning

In the context delineated by Wei et al. (2023), the zero-shot learning approach signifies a model's inherent capability to predict or make determinations about unfamiliar, previously unexperienced tasks without a requirement for task-specific training data. Instead of equipping ChatGPT through targeted training on stock market forecasts stemming from news headlines, the model utilizes its extensive foundational training over diverse textual corpora to discern insights regarding these novel tasks as in Törnberg (2023), Lopez-Lira and Tang (2023) or Hansen and Kazinnik (2023).

The research leverages ChatGPT's zero-shot learning in this manner: It refrains from directly instructing the model with historical stock market datasets or particularized financial narratives. Rather, it draws upon ChatGPT's pre-existing vast training.

## 4. Data collection

In the domain of Natural Language Processing (NLP), the integrity and quality of data are paramount. The principle that suboptimal input invariably results in suboptimal output emphasizes the imperative of rigorous data preprocessing. It is essential to meticulously curate the dataset by eliminating any erroneously aggregated news articles,

as these can introduce inaccuracies into subsequent model predictions. The task of sanitizing a dataset encompassing daily pertinent news spanning from 2010 to 2023 proved to be an intricate endeavor. The dataset under scrutiny comprises daily financial market summaries from Bloomberg for the specified period, with a primary focus on global news headlines that have the potential to sway stock market dynamics.

### 4.1. Key words

To feed our model, we explored a comprehensive exploration of diverse sources. Our primary data collection was initially sourced from the renowned website, 'Investing'. The acquired news articles were systematically categorized into five distinct segments: Most Popular, Stock Markets, Commodities and Futures, Economy and Economic Indicators, and lastly, Forex and Cryptocurrencies. The objective behind this categorization was to generate a word cloud to discern the most salient news keywords for each respective category, as illustrated in figure 1.

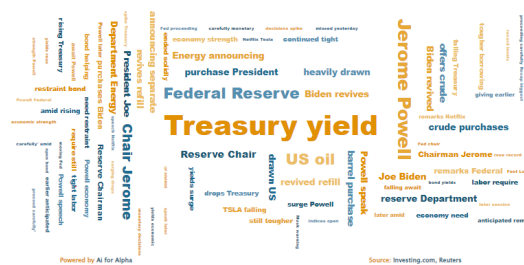


Figure 1: Example of daily Most Popular Word cloud

### 4.2. Headlines

Subsequently, we extended our data collection efforts by integrating multiple sources from Bloomberg (BBG), encompassing the Global Markets Wrap, Market Talks, and Morning Reports, as we remarked that these news were subsequently spread into larger diffusion through Yahoo News, CNews, CNBC, and many other mainstream news channels. The objective behind this comprehensive data acquisition was to encapsulate diverse news emanating from distinct global regions, namely Asia, Europe, and the United States, facilitating the generation of informative headlines.

## 5. Prompt engineering

### 5.1. Naive approach

We then iterated to construct an index termed the "Global Equities Sentiment Indicator." The initial methodology encompassed the design of a model poised to directly generate a sentiment assessment. Initially, our prompt directive was to provide a sentiment score from the full Global Market Wraps text, scaled from 1 (denoting the most pessimistic sentiment) to 10 (signifying the most optimistic sentiment). This straight approach similar in spirit to [Romanko et al. \(2023\)](#) or [Kim et al. \(2023\)](#) is challenging as it can suffer from model *hallucinations*, necessitating careful verification and validation of the output. Notwithstanding, subsequent evaluations revealed that this particular index demonstrated insufficient correlation with prospective stock market movements.

### 5.2. Two steps approach

Subsequently, we employed our Large Language Model (LLM) to scrutinize the news articles, discerning the prevailing thematic elements. In accordance with the recommendations posited in ([Lopez-Lira and Tang, 2023](#)), we devised a more intricate prompt to refine the objectives for ChatGPT, focusing on tasks empirically demonstrated to align well with ChatGPT's capabilities. For each identified theme, the model generated a concise headline that succinctly captured its essence. With these synthesized headlines at our disposal, we further refined our LLM to emulate the analytical capabilities of a financial expert. Given a headline, we tasked the model to categorize its anticipated impact on stock markets as either "positive," "negative," or "neutral." The overall idea of this two steps approach was to ease the task of chatGPT and leverage its amazing capacity to make summaries and therefore identify key headlines. Leveraging this categorization, we have devised an enhanced and more pertinent "Global Equities Sentiment Indicator" as defined below:

**Definition 5.1. Daily Sentiment Score:** Let us denote  $h_i$  as the  $i^{\text{th}}$  headline scanned from the daily news  $n$  and have two scoring functions that are consistent, a positive one  $p(h_i)$  which returns 1 if  $h_i$  is positive, 0 otherwise and a negative one  $n(h_i)$  which returns 1 if  $h_i$  is negative, 0 otherwise.

The sentiment score  $S$  for a day with  $N$  headlines is given by:

$$S = \frac{\sum_{i=1}^N p(h_i) - \sum_{i=1}^N n(h_i)}{\sum_{i=1}^N p(h_i) + \sum_{i=1}^N n(h_i)} \quad (1)$$

The sentiment score  $S$  measures the relative dominance of positive versus negative sentiments

in a day's headlines. It satisfies a couple of simple properties that are trivial to prove.

**Proposition 1.** The sentiment score  $S$  satisfies some canonical properties:

1. **Boundedness:**  $S$  is bounded as  $-1 \leq S \leq 1$ .
2. **Symmetry:** If sentiments of all headlines are reversed, then  $S$  changes its sign.
3. **Neutrality:**  $S = 0$  if there are equal numbers of positive and negative headlines.
4. **Monotonicity:**  $S$  increases as the difference between positive and negative headlines increases.
5. **Scale Invariance:**  $S$  remains the same if we multiply the number of both positive and negative headlines by a constant.
6. **Additivity:** The combined  $S$  for two sets of headlines is the weighted average of their individual  $S$  values.

Figure 2 shows the raw signal and highlights that the signal remains very noisy. Using the raw sentiment score for daily news headlines often results in noisy and less interpretable results. To address this, we propose a *cumulative sentiment score* over a specified period. This score aggregates news sentiments over a duration, offering a more comprehensive measure of the news impact during that period.



Figure 2: Raw signal: it exhibits significant noise.

**Definition 5.2. Cumulative Sentiment Score:** Given:

- $h_{i,t}$  as the  $i^{\text{th}}$  headline on day  $t$ .
- $p(h_{i,t})$  and  $n(h_{i,t})$  as functions returning 1 for positive and negative sentiments of  $h_{i,t}$  respectively, 0 otherwise.
- $d$  as the duration (we use  $d = 20$  business days, approximating a month).

The cumulative sentiment score  $S_d$  over period  $d$  is:

$$S_d = \frac{\sum_{t=1}^d \sum_{i=1}^{N_t} p(h_{i,t}) - \sum_{t=1}^d \sum_{i=1}^{N_t} n(h_{i,t})}{\sum_{t=1}^d \sum_{i=1}^{N_t} p(h_{i,t}) + \sum_{t=1}^d \sum_{i=1}^{N_t} n(h_{i,t})} \quad (2)$$

with  $N_t$  being the number of headlines on day  $t$ .

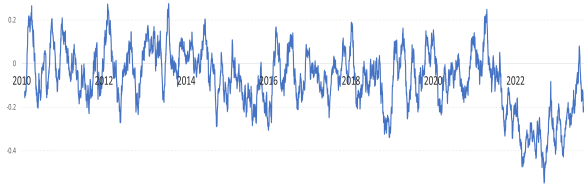


Figure 3: Cumulative sentiment score

The mathematical properties, that is boundedness, symmetry, neutrality, monotonicity, scale invariance remains for the Cumulative Sentiment Score. Figure 3 illustrates how the cumulative process diminishes the noise within the signal.

## 6. Converting to investment strategy

### 6.1. Removing noise

Given the cumulative sentiment score (see definition 5.2), it is crucial to de-trend this score to identify more actionable trading signals. We compute the trend of the sentiment score by calculating the difference between the cumulative sentiment score and its average over a period  $d$ , which we also take as a month.

**Definition 6.1. Detrended Cumulative Sentiment Score:** We call the detrended cumulative sentiment score, the cumulative sentiment score subtracted from its average over  $d$  periods

$$DS(s) = S_d(t) - \frac{1}{d} \sum_{i=1}^d S_d(t-i) \quad (3)$$

### 6.2. Splitting into long and short

From the de-trended score, we can derive two types of trading positions:

$$\text{Long Position} = \max(DS(t), 0) \quad (4)$$

$$\text{Short Position} = \min(DS(t), 0) \quad (5)$$

A long (respectively short) position is the purchase (respectively short) of an asset with the expectation that its value will rise (respectively decline) in the future. Hence, if our detrended score is positive (respectively negative) we take a long (respectively short) position. To backtest our strategy, we use the NASDAQ index as it is well known that US markets tend to react to Macroeconomic News and Outlook (.). We calculate the value of the strategy taking great care of accounting for transaction costs. We apply a linear transaction cost based on the weight difference between time  $t$  and  $t-1$ .

The value of our strategy at time  $t$  is therefore given by the cumulative returns diminished by any

transaction costs:

$$S_t = S_{t_0} \times \prod_{i=1}^t (1 + w_{t-2}r_t - b|w_{t-2} - w_{t-1}|) \quad (6)$$

where  $b$  represents the linear transaction cost and taken to be two basis points for the NASDAQ futures. It is essential to note the two-day lag in our weightings: for day  $t$ , we use the weights computed on  $t-2$ . This lag ensures that the strategy is executed the next day ensuring that our backtest does not suffer from any data leakage.



Figure 4: Long Strategy with Cumulative Sentiment (Blue) & Detrended Score (Orange)

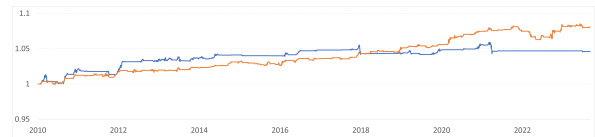


Figure 5: Short Strategy with Cumulative Sentiment (Blue) & Detrended Score (Orange)

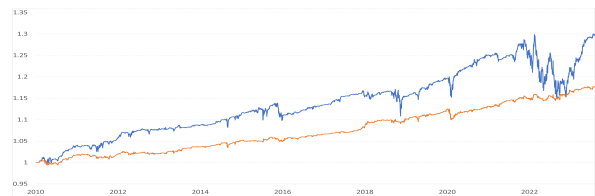


Figure 6: Final strategy (long and short) with Cumulative Sentiment (Blue) & Detrended Score (Orange)

## 7. Results

### 7.1. Descriptive statistics

In order to evaluate the performance of our strategy against a benchmark, such as a simple holding of the NASDAQ index, we consider multiple key financial metrics: Sharpe, Sortino and Calmar ratio presented below.

#### 7.1.1. Sharpe Ratio

The Sharpe (Sharpe, 1964) Ratio is a measure used to understand the return of an investment compared to its risk. It is the average return earned

in excess of the risk-free rate per unit of volatility or total risk. The Sharpe Ratio is defined as:

$$\text{Sharpe Ratio} = \frac{E[R_a - R_f]}{\sigma_a}$$

Where:

- $E[R_a - R_f]$  is the expected return on the asset  $a$  over the risk-free rate  $R_f$ .
- $\sigma_a$  is the standard deviation (volatility) of the return on the asset.

A higher Sharpe Ratio indicates a superior risk-adjusted performance of the asset or strategy.

### 7.1.2. Sortino Ratio

The Sortino Ratio is a modification of the Sharpe Ratio, but it penalizes only negative volatility rather than the overall volatility, making it especially useful for investors and traders who are concerned with downside risk. It is defined as the excess return of an investment portfolio over the risk-free rate, divided by the downside deviation.

Mathematically, the Sortino ratio is given by:

$$\text{Sortino Ratio} = \frac{R_p - R_f}{TDD}$$

where:

- $R_p$  is the portfolio's expected return and  $R_f$  is the risk-free rate.
- $TDD$ , the Total Downside Deviation, measures the volatility of negative asset returns or downside risk. It is computed similarly to the standard deviation but takes into account only negative returns.

A higher Sortino Ratio indicates superior return performance relative to downside risk. Among portfolios, an asset with a higher Sortino is typically favored for its better risk-adjusted reward.

### 7.1.3. Calmar Ratio (Return over Max Drawdown)

The Calmar Ratio, or return over maximum drawdown, gauges the performance of an investment strategy against its potential downside risk. It is defined as:

$$\text{Calmar Ratio} = \frac{CAGR}{MD}$$

Where:

- $CAGR$  is the traditional compound annual growth rate of the investment over its period.

- $MD$  (Max Drawdown) represents the largest single drop from peak to bottom in the value of a portfolio (before a new peak is achieved). Mathematically, it can be defined as:

$$MD = \max \left( \frac{\text{Peak Value} - \text{Trough Value}}{\text{Peak Value}} \right)$$

across all periods under consideration.

The higher the Calmar Ratio, the better the investment's performance relative to the magnitude of its largest drawdown.

## 7.2. Comparison

In our research, we sought to determine the efficacy of various investment strategies derived from the application of the Large Language Model (LLM) versus a passive investment. Our emphasis was on examining strategies based on Cumulative Sentiment Scores (starting by "C") and Detrended Cumulative Sentiment Scores (starting by "DC"), where each strategy is further segregated into three distinct categories: long, short, and total (a combination of the long and short).

### 7.2.1. Comparative Analysis of Strategy Performance Metrics

We present two tables, 1 and 2, detailing the performance metrics of the strategies. In these tables, the best scores are prominently highlighted in blue for easy identification and comparison. Table 1 reveals that:

- The Detrended Cumulative Score (DC) strategy consistently outperforms the baseline across metrics: Sharpe (0.88 vs. 0.79), Sortino (1.06 vs. 1.02), and Calmar (0.52 vs. 0.45). This highlights the DC Total strategy's robustness and Pareto superiority.
- In stark contrast, the naive cumulative score (C) strategies considerably underperform against the baseline. This is particularly noticeable with the C Total, C Long, and C Short strategies which have the lowest ratios across all three metrics.

Table 2 offers a granular insight into the performance by furnishing metrics like annual return, annual volatility, and a tail risk measure computed as the annual return divided by the worst 10% quantile. Mirroring our previous observations, we observe:

- The DC Total strategy has the best 'Return over Worst 10' ratio of 1.71 to compare with the baseline value of 1.03. This implies that DC Total strategy has considerably lower downside risk.

- The Cumulative Sentiment Score strategies again seem less promising with a 'Return Worst 10' ratio of 0.72, further emphasizing the potential pitfalls of a straightforward cumulative score strategy.

Strategy	Sharpe Ratio	Sortino Ratio	Calmar Ratio
DC Total	<b>0.88</b>	<b>1.06</b>	<b>0.52</b>
Baseline	0.79	1.02	0.45
DC Long	0.75	0.76	0.32
DC short	0.56	0.48	0.27
C Total	0.45	0.50	0.17
C Long	0.45	0.27	0.21
C Short	0.38	0.36	0.14

Table 1: Investment Statistics

Strategy	Annual Return	Annual Vol	Return / Worst 10
DC Total	1.2%	1.4%	<b>1.71</b>
Baseline	16.1%	20.4%	1.03
DC Long	0.6%	0.8%	1.12
DC short	0.6%	1.1%	0.68
C Total	1.9%	4.2%	0.72
C Long	0.3%	0.7%	0.28
C Short	1.6%	4.1%	0.60

Table 2: Descriptive Statistics

### 7.2.2. Analysis of Weights Descriptive Statistics

Table 3 presents a comprehensive breakdown of the weights for the four investment strategies: C Long, DC Long, C Short, and DC Short.

The Detrended Cumulative Sentiment (DC) weights display decreased volatility when juxtaposed against their pure Cumulative (C) counterparts. Both DC Long and Short have a volatility of 3.7%. In comparison, C Long and Short are more volatile, recording 4.9% and 11.1%, respectively. When comparing their means, we observe:

- For the long strategies, both C Long and DC Long have comparable mean weights (2.4% and 2.6%, respectively). This similarity indicates a similar average market exposure.
- For the short strategies, there's a notable divergence in mean weights. C Short displays a much stronger bearish sentiment at -9.5%, whereas DC Short is more tempered at -2.7%.

We can conclude that the DC strategies, particularly on the long side, appear more controlled in their weight distribution. When considering the

short strategies, detrending results in a notably milder bearish sentiment in comparison to the undetrended approach. Furthermore, we compare the ratio between the DC and C weights. Table 4 depicts the comparative strength of Detrended Cumulative (DC) weights against their Cumulative (C) counterparts.

- The Long strategy showcases a mean ratio of 1.08. This implies that DC weights for long positions outpace their C weights by approximately 8% on average.
- Conversely, the Short strategy exhibits a starkly lower mean ratio of 0.28. This stark contrast suggests that the DC weights for short positions are only 28% of their C counterparts.
- Observing the standard deviations, the long strategy (0.77) is more variable than the short strategy (0.34), hinting at more pronounced fluctuations in the long strategy's ratio.
- Extreme values indicate possible outliers or peak events, with the long strategy peaking at 0.75 and the short at 0.35. These figures further accentuate the divergence between the long and short strategies.

In essence, detrending amplifies the long strategy weights while significantly dampening the short strategy when juxtaposed against pure Cumulative weights.

To supplement the tabulated insights, we also provide various figures to show the temporal behavior of the six NLP based strategies. Figures 4, 5, and 6 depict the evolution of the long, short, and total strategies for the cumulative and detrended cumulative sentiment respectively.

	Long C	Long DC	Short C	Short DC
mean	2.4%	2.6%	9.5%	2.7%
std	4.9%	3.7%	11.1%	3.7%
max	27.4%	20.6%	54.7%	18.9%

Table 3: Weights descriptive statistics

	Long	Short
mean	1.08	0.28
std	0.77	0.34
extreme	0.75	0.35

Table 4: Ratio between DC and C weights



## 8. Conclusion

This work has explored the potential and adaptability of LLMs, specifically ChatGPT 4, in the challenging realm of financial sentiment analysis.

Using a zero-shot prompting technique, we assessed the model's proficiency in deciphering and producing predictive sentiment scores from financial market summaries without necessitating domain-specific refinement. Our analysis was based on a meticulously assembled dataset of daily Bloomberg market summaries, disseminated via platforms like the Bloomberg terminal to external sources such as Yahoo Finance, CNews, etc.

When comparing the Sharpe and Calmar ratios of a passive investment strategy to our NLP-driven strategy, our findings show notable potential for forecasting future market returns. Additionally, our investigation emphasizes the pivotal role of judicious prompt selection in augmenting the efficacy of sentiment analysis endeavors, as well as the importance of filtering out data noise through mechanisms like cumulative sentiment scores and detrended cumulative sentiment indicators.

Our research highlights ChatGPT's promise in financial sentiment analysis. The prompts' efficacy in summarizing daily news and discerning market sentiment suggests avenues for refining data integration and filtering textual noise.

## 9. Bibliographical References

- Ajay Agrawal, Joshua S. Gans, and Avi Goldfarb. 2019. [Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction](#). *Journal of Economic Perspectives*, 33(2):31–50.
- Douglas W Arner, Janos Barberis, and Ross P Buckley. 2015. The evolution of fintech: A new post-crisis paradigm. *Geo. J. Int'l L.*, 47:1271.
- John L. Campbell, Hsinchun Chen, Dan S. Dhaliwal, Hsin-min min Lu, Logan B. Steele, John L. Campbell, Hsinchun Chen, Hsinchun Chen, Dan S. Dhaliwal, Dan S. Dhaliwal, Hsin-min min Lu, Hsin-min min Lu, Logan B. Steele, and Logan B. Steele. 2014. [The information content of mandatory risk factor disclosures in corporate filings](#). *Review of accounting studies*, 19(1):396–455.
- Tyler Cowen and Alexander T. Tabarrok. 2023. [How to Learn and Teach Economics with Large Language Models, Including GPT](#). *SSRN Electronic Journal*, XXX(XXX):0–0.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, XX(XX):XX.
- Georgios Fatouros, Georgios Makridis, Dimitrios Kotios, John Soldatos, Michael Filippakis, and Dimosthenis Kyriazis. 2023. Deepvar: a framework for portfolio risk assessment leveraging probabilistic deep neural networks. *Digital finance*, 5(1):29–56.
- Diego Garcia. 2013. [Sentiment during Recessions](#). *The Journal of Finance*, 68(3):1267–1300.
- A Shaji George and AS Hovan George. 2023. A review of chatgpt ai's impact on several business sectors. *Partners Universal International Innovation Journal*, 1(1):9–23.
- Abbas Ghaddar and Philippe Langlais. 2020. [Sedar: a large scale french-english financial domain parallel corpus](#). In *Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC)*, pages 3595–3602, LREC. LREC.
- Anne Lundgaard Hansen and Sophia Kazinnik. 2023. [Can ChatGPT Decipher FedSpeak?](#) *SSRN Electronic Journal*, XX(XX):XX.
- Ioan-Bogdan Iordache, Ana Sabina Uban, Catalin Stoean, and Liviu P Dinu. 2022. [Investigating the relationship between romanian financial news and closing prices from the bucharest stock exchange](#). In *Proceedings of the Thirteenth Language Resources and Evaluation Conference (LREC)*, pages 5130–5136, LREC. LREC.
- Ali Jabbari, Olivier Sauvage, Hamada Zeine, and Hamza Chergui. 2020. [A french corpus and annotation schema for named entity recognition and relation extraction of financial news](#). In *Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC)*, pages 2293–2299, LREC. LREC.
- Narasimhan Jegadeesh and Di Wu. 2013. [Word power: A new approach for content analysis](#). *Journal of Financial Economics*, 110(3):712–729.
- Alex Kim, Maximilian Muhn, and Valeri Nikolaev. 2023. Bloated disclosures: Can chatgpt help investors process financial information? *arXiv preprint arXiv:2306.10224*, XXX(0-0):XX.
- Hyungjin Ko and Jaewook Lee. 2023. [Can Chatgpt Improve Investment Decision? From a Portfolio Management Perspective](#). *SSRN Electronic Journal*, XX(XX):XX.
- Jan Kocoń, Igor Cichecki, Oliwier Kaszyca, Mateusz Kochanek, Dominika Szydło, Joanna Baran, Julita Bielaniewicz, Marcin Gruza, Arkadiusz Janz, Kamil Kanclerz, et al. 2023. Chatgpt: Jack of all trades, master of none. *Information Fusion*, page 101861.
- Anton Korinek. 2023. [Language Models and Cognitive Automation for Economic Research](#). *Cambridge, MA*, XX(XX):XX.
- Chenyang Li, Wenbo Ye, and Yilun Zhao. 2022. [Finmath: Injecting a tree-structured solver for question answering over financial reports](#). In *Proceedings of the Thirteenth Language Resources and Evaluation Conference (LREC)*, pages 6147–6152, LREC. LREC.
- Zhuang Liu, Degen Huang, Kaiyu Huang, Zhuang Li, and Jun Zhao. 2021. Finbert: A pre-trained financial language representation model for financial text mining. In *Proceedings of the twenty-ninth international conference on international joint conferences on artificial intelligence*, pages 4513–4519, ICLR. ICLR.
- Alejandro Lopez-Lira and Yuehua Tang. 2023. [Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models](#). *SSRN Electronic Journal*, XXX(XX-XX):XX.

- Tim Loughran and Bill McDonald. 2011. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of finance*, 66(1):35–65.
- Corentin Masson and Patrick Paroubek. 2020. [Nlp analytics in finance with dore: a french 250m tokens corpus of corporate annual reports](#). In *Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC)*, pages 2261–2267, LREC. LREC.
- Antonio Moreno-Ortiz, Javier Fernández-Cruz, and Chantal Pérez Chantal Hernández. 2020. [Design and evaluation of sentiecon: A fine-grained economic/financial sentiment lexicon from a corpus of business news](#). In *Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC)*, pages 5065–5072, LREC. LREC.
- Shakke Noy and Whitney Zhang. 2023. [Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence](#). *SSRN Electronic Journal*, XX(X):XX.
- Joel Oksanen, Abhilash Majumder, Kumar Saunack, Francesca Toni, and Arun Dhondiyal. 2022. [A graph-based method for unsupervised knowledge discovery from financial texts](#). In *Proceedings of the Thirteenth Language Resources and Evaluation Conference (LREC)*, pages 5412–5417, LREC. LREC.
- OpenAI. 2023. [Gpt-4 technical report](#).
- Soujanya Poria, Erik Cambria, Rajiv Bajpai, and Amir Hussain. 2017. A review of affective computing: From unimodal analysis to multimodal fusion. *Information fusion*, 37:98–125.
- Soujanya Poria, Erik Cambria, and Alexander Gelbukh. 2016. Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*, 108:42–49.
- David E Rapach, Jack K Strauss, and Guofu Zhou. 2013. [International stock return predictability: What is the role of the united states?](#) *Journal of Finance*, 68(4):1633–1662.
- Oleksandr Romanko, Akhilesh Narayan, and Roy H Kwon. 2023. Chatgpt-based investment portfolio selection. *arXiv preprint arXiv:2308.06260*, XX(X):XX.
- Robert P Schumaker and Hsinchun Chen. 2009. Textual analysis of stock market prediction using breaking financial news: The azfin text system. *ACM Transactions on Information Systems (TOIS)*, 27(2):1–19.
- William F Sharpe. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19:425–442.
- Paul C. Tetlock. 2007. [Giving Content to Investor Sentiment: The Role of Media in the Stock Market](#). *The Journal of Finance*, 62(3):1139–1168.
- Petter Törnberg. 2023. Chatgpt-4 outperforms experts and crowd workers in annotating political twitter messages with zero-shot learning. *arXiv preprint arXiv:2304.06588*.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017-Decem:5999–6009.
- Michael Webb. 2019. [The Impact of Artificial Intelligence on the Labor Market](#). *SSRN Electronic Journal*.
- Xiang Wei, Xingyu Cui, Ning Cheng, Xiaobin Wang, Xin Zhang, Shen Huang, Pengjun Xie, Jinan Xu, Yufeng Chen, Meishan Zhang, et al. 2023. Zero-shot information extraction via chatting with chatgpt. *arXiv preprint arXiv:2302.10205*.
- Qianqian Xie, Weiguang Han, Yanzhao Lai, Min Peng, and Jimin Huang. 2023. The Wall Street Neophyte: A Zero-Shot Analysis of ChatGPT Over MultiModal Stock Movement Prediction Challenges. *arXiv preprint arXiv:2304.05351*, XX(X):XX.
- Kai-Cheng Yang and Filippo Menczer. 2023. [Large language models can rate news outlet credibility](#). Technical report, arxiv.
- Chaofa Yuan, Yuhan Liu, Rongdi Yin, Jun Zhang, Qinling Zhu, Ruibin Mao, and Ruifeng Xu. 2020. [Target-based sentiment annotation in chinese financial news](#). In *Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC)*, pages 5040–5045, LREC. LREC.
- Thomas Yue, David Au, Chi Chung Au, and Kwan Yuen lu. 2023. Democratizing financial knowledge with chatgpt by openai: Unleashing the power of technology. *Available at SSRN 4346152*, XX(X):XX.
- Nadhem Zmandar, Tobias Daudert, Sina Ahmadi, Mahmoud El-Haj, and Paul Rayson. 2022. [Cofif plus: A french financial narrative summarisation corpus](#). In *Proceedings of the Thirteenth Language Resources and Evaluation Conference (LREC)*, pages 1622–1639, LREC. LREC.