

# Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models

Alejandro Lopez-Lira and Yuehua Tang  
University of Florida

December 2023

# Overview

## Motivation

- ▶ Generative artificial intelligence and large language models (LLMs) such as ChatGPT have recently grabbed significant attention in many domains.
- ▶ While news (and textual information) can move stock prices, we still know little about the ability of LLMs to predict stock market returns. We know markets do not incorporate information immediately (except for large stocks?)
- ▶ Can ChatGPT (and other models) understand whether the news is good or bad for firms' stock prices?
- ▶ Despite lack of explicit training, could LLMs offer value in predicting stock market movements due to their advanced natural language understanding?

## What we do

- ▶ We evaluate ChatGPT's capabilities in forecasting stock market returns using news headlines data.
- ▶ We give it a headline and ask if it's good/bad/neutral.
- ▶ We then measure the return the next day
- ▶ We compare ChatGPT to other models and Ravenpack's sentiment score

## Performance Comparison

- ▶ ChatGPT outperforms traditional sentiment analysis methods.
- ▶ Basic models like GPT-1, GPT-2, and BERT fall short in accurate return forecasting.
- ▶ Emergence of return predictability as a capacity of more complex models.

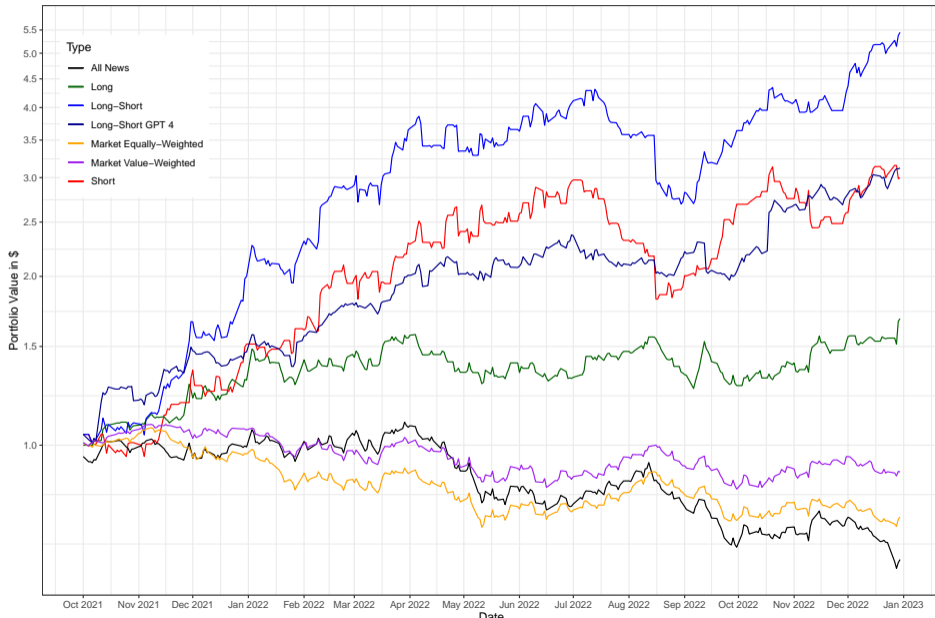
## Advanced Models and Investment Decisions

- ▶ Sharpe ratios implied by ChatGPT-4 are larger than those by ChatGPT-3.
- ▶ However, ChatGPT-3 shows larger total returns.
- ▶ Incorporating advanced language models into investment decision-making can enhance prediction accuracy and trading strategy performance.

## Concentration of Predictability

- ▶ Predictability is more prominent for smaller stocks and firms with bad news.
- ▶ These findings align more with limits-to-arbitrage arguments than with market inefficiencies.

# Cumulative Returns of Investing 1\$ (Without Transaction Costs)





# Introduction

## What is ChatGPT?

- ▶ ChatGPT is a language model developed by OpenAI based on the GPT (generative pre-trained transformer) architecture.
- ▶ It uses deep learning for natural language understanding and generation.
- ▶ Can understand context, generate coherent text, and provide solutions to various tasks.
- ▶ The interface is talking with it through text messages or “prompts.”

# ChatGPT Interface

- + New chat
- LaTeX Slides Code
- ChatGPT Course Outline
- Correlating Words with y
- Sorrento's Cure Announcemen
- Sorrento's Questionable Covid
- Option Pricing Conventions
- Equations Assistance Request
- Clear conversations
- My plan
- Settings
- Get help
- Log out

Model  
Plugins ALPHA

Plugins ALPHA

An experimental model that knows when and how to use plugins

Default (GPT-3.5)

Legacy (GPT-3.5)

GPT-4

Plugins

FN

ChatGPT PLUS


Send a message...

[ChatGPT Mar 23 Version](#). ChatGPT may produce inaccurate information about people, places, or facts.

# ChatGPT Interface


ChatGPT 4 ▾

 You

 TextClassification.csv  
Spreadsheet

 SignalsTheoryChecked.csv  
Spreadsheet

I need to merge these two files, I am having problems with the counts by the column theory, they don't match

 ChatGPT

The two files you've provided have been loaded successfully. Here's a brief overview of their structure:

## File 1: TextClassification.csv

- **Columns:** Includes `signalname`, `Journal`, `Authors`, `Year`, `theory`, `misp_count`, `risk_count`, `misprice_risk_ratio`, `risk_mispricing_ratio`, `quote`, `LongDescription`, `liquidity`, `LiquidityNotes`, `Notes`, `post_2004`, `author_merge`, `Keep`, `word_count`.
- **Example Row:** The first row contains information about the signal "AM", published in the "JF" journal by "Fama and French" in 1992. It is classified under the theory "agnostic" and has various counts and ratios related to mispricing and risk.

# ChatGPT API

- ▶ An API that allows researchers to integrate ChatGPT into code
- ▶ Can be accessed in Python or R
- ▶ The system message helps set the behavior of the assistant
- ▶ Can set the temperature, select the model, and the max number of tokens
- ▶ Temperature: higher values like 0.9, more random, lower values like 0.2, more focused and deterministic
  - ▶ Use temperature = 0 for setting a seed

# ChatGPT API

```
# Note: you need to be using OpenAI Python v0.27.0 for the code below to
import openai

openai.ChatCompletion.create(
    model="gpt-3.5-turbo",
    messages=[
        {"role": "system", "content": "You are a helpful assistant."},
        {"role": "user", "content": "Who won the world series in 2020?"},
        {"role": "assistant", "content": "The Los Angeles Dodgers won the World"},
        {"role": "user", "content": "Where was it played?"}
    ]
)
```



## ChatGPT is Not...

- ▶ Google: It does not have current information and WILL make things up
  - ▶ Do NOT use it for literature review
  - ▶ Even with the Bing plugin, it will just click on the first result
  
- ▶ Supervised Machine Learning: It is not good with numerical data!
  
- ▶ Great at Logical Reasoning:
  - ▶ Complex logical problems require a deeper understanding of relationships between concepts, abstract thinking, and inference
  - ▶ Logical reasoning often involves complex and structured arguments, and ChatGPT cannot follow.

## How Models like ChatGPT Work

- ▶ The model is pre-trained on a large corpus of text, such as Wikipedia or the Common Crawl dataset, using a language modeling objective.
- ▶ During pre-training, the model learns to predict the next word in a sequence given the previous words.
- ▶ After pre-training, the model can be fine-tuned on a specific task, such as text classification or generation.
- ▶ Fine-tuning involves training the model on a smaller dataset specific to a task.



# Reinforcement Learning from Human Feedback

- ▶ LLMs like ChatGPT can generate impressive text but are not necessarily good at following instructions.
- ▶ Humans can help address this issue by providing feedback to the model.
- ▶ Human evaluators rate the quality of the generated text and provide feedback to the model.
- ▶ A new model is then trained to imitate human preferences.
- ▶ The LLM is tuned to satisfy the preferences

# Language Model Hallucinations in ChatGPT

- ▶ ChatGPT is trained to predict the next word based on the context.
- ▶ ChatGPT may produce text that appears real but is not based on real information.
- ▶ This is known as “hallucination” and can include fabricated background information, papers, or sources.
- ▶ Hallucinations may occur due to the model’s goal of generating text that humans rate as useful.
- ▶ Always verify the information provided by ChatGPT.

## Academic Literature and Contribution

- ▶ Our study is among the first to study the potential of LLMs in financial markets, particularly the investment decision-making process:
  - ▶ Hansen and Kazinnik (2023): LLMs like ChatGPT can decode FedSpeak.
  - ▶ Noy and Zhang (2023): ChatGPT can enhance productivity in professional writing.
- ▶ We contribute to the literature employing text analysis and machine learning to study finance research questions:
  - ▶ E.g., Jegadeesh and Wu (2013), Baker, Bloom, and Davis (2016), Manela and Moreira (2017), Bybee et al. (2019), Lopez-Lira 2019
- ▶ It adds the literature that uses linguistic analyses of news articles to extract sentiment and predict stock returns :
  - ▶ E.g., Tetlock (2007), Garcia (2013), Calomiris and Mamaysky (2019), Tetlock, Saar-Tsechansky, and Macskassy (2008), Tetlock (2011)

## Our Approach

- ▶ Novel approach: Leverage ChatGPT's sentiment analysis capabilities using news headlines data.
  
- ▶ Comparison with existing sentiment analysis methods by leading vendors.

## Potential Implications for the Financial Industry

- ▶ Shift in methods used for market prediction and investment decision-making.
- ▶ Contributions to understanding LLMs' applications in financial economics.

## Broader Contributions

- ▶ Aid to regulators and policymakers in understanding the benefits and risks of LLMs in financial markets.
- ▶ Empirical evidence for asset managers and institutional investors on the efficacy of LLMs in predicting stock returns.
- ▶ Contributions to the academic discourse on AI applications in finance, inspiring future research on more sophisticated LLMs.

## Prompt

*Forget all your previous instructions. Pretend you are a financial expert. You are a financial expert with stock recommendation experience. Answer "YES" if good news, "NO" if bad news, or "UNKNOWN" if uncertain in the first line. Then elaborate with one short and concise sentence on the next line. Is this headline good or bad for the stock price of \_company\_name\_ in the \_term\_ term?*

*Headline: \_headline\_*

## Prompt Structure Analysis

- ▶ **Forget all your previous instructions.** Ensures that the AI model disregards previous instructions and focuses only on the current task.
- ▶ **Pretend you are a financial expert.** Guides the model to generate responses with a financial expert's knowledge and perspective.
- ▶ **You are a financial expert with stock recommendation experience.** Provides further context to the model, indicating that the task requires expertise in stock recommendations.



## Prompt Structure Analysis

- ▶ **Answer "YES" if good news, "NO" if bad news, or "UNKNOWN" if uncertain in the first line.** Specifies the format of the response, which makes it easy to parse.
- ▶ **Then elaborate with one short and concise sentence on the next line.** Instructs the model to provide further explanation for its initial response, increasing the value of the output by providing reasoning.
- ▶ **Is this headline good or bad for the stock price of `__company_name__` in the `__term__` term?**  
Defines the main task for the model, asking it to analyze a headline and predict its impact on a company's stock price over a certain term.
- ▶ **Headline: `__headline__`.** Provides the input data for the model to analyze.

## Selected Example

*Rimini Street Fined \$630,000 in Case Against Oracle.*

The prompt then asks:

*Forget all your previous instructions. Pretend you are a financial expert. You are a financial expert with stock recommendation experience. Answer "YES" if good news, "NO" if bad news, or "UNKNOWN" if uncertain in the first line. Then elaborate with one short and concise sentence on the next line. Is this headline good or bad for the stock price of Oracle in the short term?*

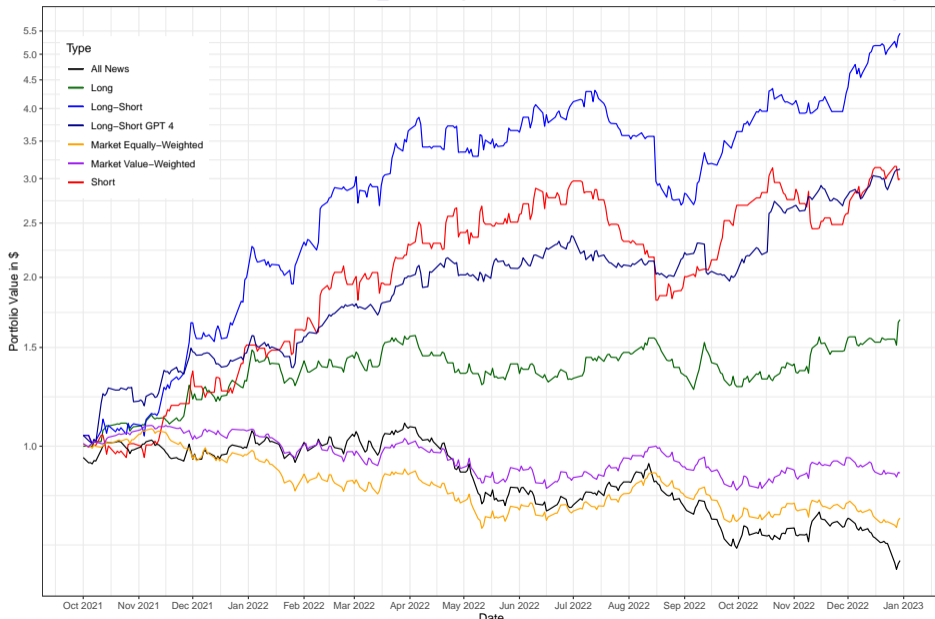
***Headline:*** *Rimini Street Fined \$630,000 in Case Against Oracle*

## Selected Example

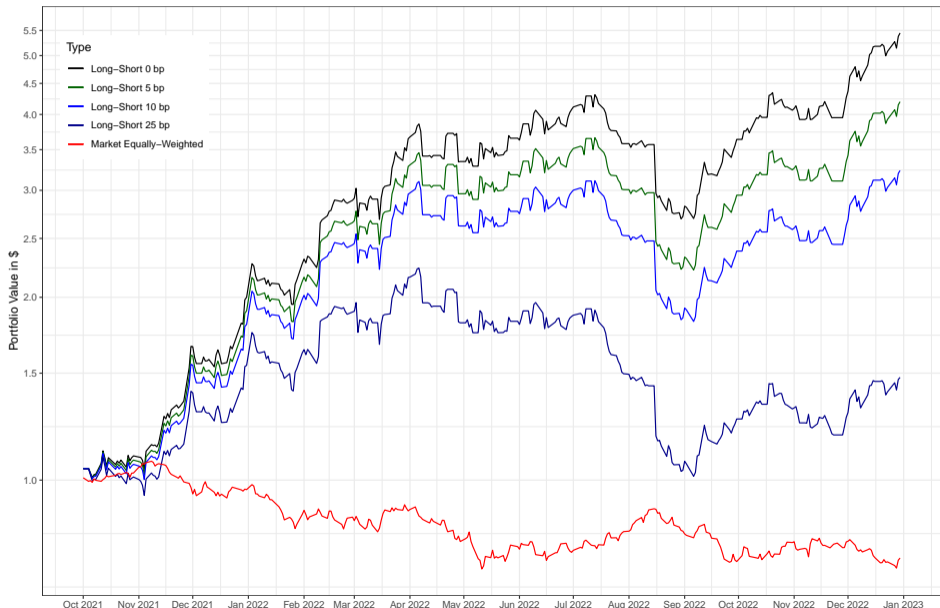
YES

*The fine against Rimini Street could potentially boost investor confidence in Oracle's ability to protect its intellectual property and increase demand for its products and services.*

# Cumulative Returns of Investing 1\$ (Without Transaction Costs)



# Cumulative Returns of Investing \$1 (With Transaction Costs)



# Economic Model

- ▶ Short Sell Costs
- ▶ Costly Information Acquisition
- ▶ News takes time to incorporate fully into prices
  - ▶ More so for negative, complex news and firms with less attention (smaller)
- ▶ An agent (LLM) that can understand this information faster would see return predictability

## Descriptive Statistics

	Mean	SD	min	P25	Median	P75	Max	N
Daily Return (%)	0	5.26	-64.97	-2.04	-0.02	1.89	237.11	60755
Headline Length	76.36	28.65	21	56	70	90	395	60755
Response Length	153.31	38.04	0	124	151	179	303	60755
GPT Score	0.24	0.47	-1	0	0	1	1	60755
Event Score	0.18	0.49	-1	0	0	0	1	60755

## Correlations

	Daily Return	Headline Length	Resp. Length	GPT Score	Event Score
Daily Return	1				
Headline Length	-0.002	1			
Response Length	-0.001	0.261	1		
GPT Score	0.018	0.081	0.441	1	
Event Score	0.005	-0.071	0.091	0.279	1



## Descriptive Statistics of Various Portfolios

	Long (L)	Short (S)	L-S ChatGPT	L-S GPT-4	Market EW	Market VW	All News EW
Sharpe Ratio	1.72	1.86	3.09	3.80	-0.99	-0.39	-0.98
Daily Mean (%)	0.25	0.38	0.63	0.44	-0.10	-0.04	-0.11
Daily Std. Dev. (%)	2.32	3.26	3.25	1.84	1.55	1.49	1.83
Max Drawdown (%)	-16.94	-34.39	-22.79	-10.40	-36.12	-26.68	-38.70

## Regression of Next Day Returns on the Prediction Score

	(1)	(2)	(3)	(4)	(5)	(6)
GPT-3.5	0.259*** (5.259)	0.243*** (4.980)				
event-S		0.058 (1.122)		0.038 (0.683)	0.118* (2.272)	
GPT-4			0.176*** (5.382)	0.167*** (4.768)		
bart-L						0.142*** (4.653)
Num.Obs.	60 755	60 755	60 755	60 755	60 755	60 176
R2	0.184	0.184	0.184	0.184	0.184	0.185
Std.Errors	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno
FE: date	X	X	X	X	X	X
FE: permno	X	X	X	X	X	X

+ p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

# Regression of Next Day Returns on the Prediction Score

	(1)	(2)	(3)	(4)	(5)	(6)
distilbart	0.150*** (4.919)					
GPT-2-L		0.035 (1.051)				
GPT-2			0.001 (0.025)			
GPT-1				0.034 (1.304)		
bert					-0.226 (-3.703)	
bert-L						0.001 (0.020)
Num.Obs.	60 755	60 176	60 176	60 755	60 176	60 176
R2	0.184	0.185	0.185	0.184	0.185	0.185
Std.Errors	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno
FE: date	X	X	X	X	X	X
FE: permno	X	X	X	X	X	X

+ p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

# Regression of Next Day Returns on the Prediction Score (Small Stocks)

	(1)	(2)	(3)	(4)	(5)	(6)
GPT-3.5	0.653*** (5.145)	0.542*** (4.028)				
event-S		0.277* (2.117)		0.256+ (1.876)	0.435*** (3.567)	
GPT-4			0.501*** (4.830)	0.419*** (3.645)		
bart-L						0.165 (1.504)
Num.Obs.	14 343	14 343	14 343	14 343	14 343	14 238
R2	0.201	0.201	0.201	0.201	0.200	0.201
Std.Errors	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno
FE: date	X	X	X	X	X	X
FE: permno	X	X	X	X	X	X

+  $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

# Regression of Next Day Returns on the Prediction Score (Small Stocks)

	(1)	(2)	(3)	(4)	(5)	(6)
distilbart	0.207+ (1.895)					
GPT-2-L		0.019 (0.216)				
GPT-2			0.064 (0.765)			
GPT-1				0.008 (0.098)		
bert					-0.492** (-2.598)	
bert-L						0.018 (0.096)
Num.Obs.	14 343	14 238	14 238	14 343	14 238	14 238
R2	0.200	0.201	0.201	0.200	0.202	0.201
Std.Errors	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno
FE: date	X	X	X	X	X	X
FE: permno	X	X	X	X	X	X

+ p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

# Regression of Next Day Returns on the Prediction Score (Non-Small Stocks)

	(1)	(2)	(3)	(4)	(5)	(6)
GPT-3.5	0.148** (3.084)	0.158** (3.280)				
event-S		-0.041 (-0.830)		-0.060 (-1.163)	-0.005 (-0.112)	
GPT-4			0.097** (3.252)	0.111*** (3.491)		
bart-L						0.144*** (4.695)
Num.Obs.	46 402	46 402	46 402	46 402	46 402	45 928
R2	0.218	0.218	0.218	0.218	0.218	0.219
Std.Errors	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno
FE: date	X	X	X	X	X	X
FE: permno	X	X	X	X	X	X

+ p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

# Regression of Next Day Returns on the Prediction Score (Non-Small Stocks)

	(1)	(2)	(3)	(4)	(5)	(6)
distilbart	0.146*** (4.894)					
GPT-2-L		0.030 (0.947)				
GPT-2			-0.014 (-0.539)			
GPT-1				0.056* (2.332)		
bert					-0.165** (-2.795)	
bert-L						-0.011 (-0.200)
Num.Obs.	46 402	45 928	45 928	46 402	45 928	45 928
R2	0.218	0.219	0.219	0.218	0.219	0.219
Std.Errors	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno
FE: date	X	X	X	X	X	X
FE: permno	X	X	X	X	X	X

+ p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

## Average Next Day's Return by Prediction Score

Model	Pos	Neut	Neg	LS	t LS	$\alpha_M$	t $\alpha_M$	$R^2_M$	$\alpha_{FF5}$	t $\alpha_{FF5}$	$R^2_{FF5}$
Gpt-4	0.09	-0.18	-0.35	0.44	4.24	0.45	4.31	1.14	0.41	4.01	5.20
Gpt-3.5	0.25	-0.21	-0.38	0.63	3.44	0.63	3.41	0.47	0.60	3.28	4.15
Gpt-1	-0.10	0.01	-0.19	0.09	0.69	0.09	0.71	0.24	0.09	0.67	0.41
Gpt-2	-0.03	-0.20	0.20	-0.23	-1.38	-0.23	-1.37	0.03	-0.24	-1.39	1.83
Gpt-2-L	-0.06	-0.02	-0.16	0.10	0.92	0.10	0.95	0.40	0.10	0.94	0.64
Bart-L	-0.01	0.08	-0.16	0.15	1.40	0.15	1.41	0.04	0.13	1.25	1.91
Distilbart	-0.04	0.12	-0.28	0.24	2.12	0.24	2.13	0.08	0.22	1.91	3.77
Bert	-0.23	-0	-0.08	-0.14	-1.16	-0.12	-1.05	12.60	-0.09	-0.79	17.99
Bert-L	-0.06	-0.06	-0.11	0.04	0.23	0.05	0.25	0.36	0.04	0.19	4.24
Event	-0.04	-0.11	-0.32	0.29	1.94	0.28	1.90	0.47	0.25	1.70	3.04



## Average Next Day's Return by Prediction Score Small Stocks

Model	Pos	Neut	Neg	LS	t LS	$\alpha_M$	t $\alpha_M$	$R^2_M$	$\alpha_{FF5}$	t $\alpha_{FF5}$	$R^2_{FF5}$
Gpt-4	0.02	-0.82	-0.88	0.90	3.13	0.90	3.12	0.88	0.88	3.07	3.88
Gpt-3.5	0.07	-0.51	-0.97	1.04	2.72	1.03	2.72	2.84	1.05	2.78	5.60
Gpt-1	-0.40	-0.31	-0.51	0.10	0.35	0.10	0.34	0.10	0.13	0.41	0.70
Gpt-2	-0.24	-0.81	-0.14	-0.10	-0.36	-0.10	-0.36	0	-0.11	-0.42	1.20
Gpt-2-L	-0.23	-0.30	-0.39	0.16	0.55	0.16	0.55	0	0.20	0.68	0.65
Bart-L	-0.19	-1.98	-0.48	0.29	1.09	0.29	1.10	0.02	0.28	1.06	1.04
Distilbart	-0.25	-2.69	-0.46	0.20	0.73	0.20	0.73	0.11	0.21	0.74	1.26
Bert	-0.88	0.06	-0.08	-0.81	-4.09	-0.81	-4.27	6.57	-0.80	-4.17	8.32
Bert-L	-0.15	-0.22	-0.54	0.40	1.30	0.39	1.29	1.17	0.42	1.38	3.58
Event-S	-0.26	-0.23	-0.79	0.53	2	0.53	2	0.12	0.57	2.11	0.64

## Average Next Day's Return by Prediction Score Non-Small Stocks

Model	Pos	Neut	Neg	LS	t LS	$\alpha_M$	t $\alpha_M$	$R^2_M$	$\alpha_{FF5}$	t $\alpha_{FF5}$	$R^2_{FF5}$
Gpt-4	0.12	0.10	-0.20	0.32	3.32	0.33	3.37	0.71	0.29	2.99	4.07
Gpt-3.5	0.29	-0.06	-0.13	0.42	2.51	0.41	2.46	1.18	0.36	2.19	4.57
Gpt-1	0.02	0.06	-0.01	0.03	0.36	0.03	0.39	0.31	0.04	0.48	0.50
Gpt-2	0.06	-0	0.14	-0.08	-1.12	-0.07	-1.07	0.98	-0.07	-1.07	2.31
Gpt-2-L	-0.02	0.10	-0.02	0	0.05	0.01	0.08	0.41	0.01	0.09	0.85
Bart-L	0.10	0.32	-0.12	0.21	2.13	0.22	2.16	0.33	0.20	2.01	1.88
Distilbart	0.09	0.40	-0.24	0.33	3.13	0.33	3.18	0.70	0.31	2.93	3.89
Bert	-0	0.01	-0.02	0.01	0.10	0.04	0.37	18.39	0.07	0.74	24.87
Bert-L	0.02	0.12	0.18	-0.17	-1.06	-0.16	-1.01	1.02	-0.16	-1.03	6.41
Event-S	0.03	-0.03	0.04	-0.01	-0.05	-0.02	-0.10	1.01	-0.05	-0.33	3.83

## Sharpe Ratio and Number of Stocks in Each Leg by Model All Stocks

Model	Sharpe	$N_{+25th}$	$N_{+mean}$	$N_{+median}$	$N_{+75th}$	$N_{-25th}$	$N_{-mean}$	$N_{-median}$	$N_{-75th}$
Gpt-4	3.8	59	98.52	107	140	13	23.30	23	33
Gpt-3.5	3.09	28	50.86	51	72	1	3.84	4	6
Distilbart	1.9	104	167.87	160	234	11	22.06	21	32
Event-S	1.74	21	44.44	48	61	2	8.82	5	9
Bart-L	1.26	21.50	148.04	150	210.50	4	23.71	24	34.25
Gpt-2-L	0.82	8	59.84	61	83.50	2	12.37	12.50	18
Gpt-1	0.62	89	138.23	136	187	14	26.03	25	36
Bert-L	0.2	21.25	156.32	162.50	223.50	0	2.77	2	4
Bert	<0	7.75	36.17	43	53.25	0	0.27	0	0
Gpt-2	<0	15.75	99.55	104.50	144	3	24.25	25	35

## Sharpe Ratio and Number of Stocks in Each Leg by Model Small Stocks

Model	Sharpe	$N_{+25th}$	$N_{+mean}$	$N_{+median}$	$N_{+75th}$	$N_{-25th}$	$N_{-mean}$	$N_{-median}$	$N_{-75th}$
Gpt-4	2.98	19	25.52	26	32	2	4.24	4	6
Gpt-3.5	2.59	10	14.07	14	18	0	0.72	0	1
Event-S	1.91	8	12.63	13	16	0	3.63	1	3
Bert-L	1.24	28	43.74	39	54	0	0.52	0	1
Bart-L	1.04	26.25	44.65	37	56.75	2	4.31	4	6
Distilbart	0.69	29	47.52	38	59	2	3.92	4	5
Gpt-2-L	0.53	9	16.89	15	22	1	3.60	3	5
Gpt-1	0.33	23	37.32	32	48	4	7.24	6	9
Gpt-2	<0	16	28.53	25	36.75	4	6.92	6	9
Bert	<0	7	9.35	10	12.75	0	0.02	0	0

## Sharpe Ratio and Number of Stocks in Each Leg by Model Non-Small Stocks

Model	Sharpe	$N_{+25th}$	$N_{+mean}$	$N_{+median}$	$N_{+75th}$	$N_{-25th}$	$N_{-mean}$	$N_{-median}$	$N_{-75th}$
Gpt-4	2.99	49	76.40	80	108	9	19.66	19	28.50
Distilbart	2.82	76	126.60	120	173	8.50	18.70	17	27
Gpt-3.5	2.26	20.50	38.63	37	54	1	3.22	3	5
Bart-L	1.92	31	111.47	106.50	157	4.25	20.17	19	28
Gpt-1	0.33	64	105.85	106	142	10	19.74	19	27.50
Bert	0.09	8.25	28.34	33	41	0	0.26	0	0
Gpt-2-L	0.04	12.50	45.99	47.50	64	2	9.40	9	14
Event-S	<0	15	33.46	32	44	1.50	5.65	4	7
Bert-L	<0	36	120.49	124	171.25	0	2.34	2	4
Gpt-2	<0	19.50	76	77	107.75	6.25	18.58	18	28

## Complexity

Model	All	Low Complexity	High Complexity
Gpt-4	3.8	2.62	2.61
Gpt-3.5	3.09	1.53	1.66
Distilbart-Mnli-12-1	1.9	1.69	1.27
Event-Sentiment	1.74	1.51	0.48
Bart-Large	1.26	1.06	0.98
Gpt-2-Large	0.82	1.16	0.04
Finbert	0.77	0.97	0.31
Gpt-1	0.62	<0	0.35
Bert-Large	0.2	0.84	<0
Bert	<0	<0	<0
Gpt-2	<0	<0	<0

## News Type

Model	All	News Articles	Press Releases
Gpt-4	3.8	4.74	2.84
Gpt-3.5	3.09	2.4	1.45
Distilbart-Mnli-12-1	1.9	4.8	0.73
Event-Sentiment	1.74	0.5	0.69
Bart-Large	1.26	3.74	1.37
Gpt-2-Large	0.82	0.83	1.34
Finbert	0.77	2.85	<0
Gpt-1	0.62	0.1	0.06
Bert-Large	0.2	0.82	0.02
Bert	<0	<0	<0
Gpt-2	<0	0.93	<0

# Interpretability

- ▶ Traditional ML models in finance focus on prediction without interpretability.
- ▶ Large Language Models (LLMs) like ChatGPT offer predictions with natural language explanations.
- ▶ This feature provides insights into the rationale behind each prediction.
- ▶ Our goal: Harness these insights for understanding predictive accuracy.



## Methodology - Phase 1

- ▶ Extract raw explanation text from ChatGPT's predictions.
- ▶ Separate qualitative statements from primary forecasts.
- ▶ Clean the data by removing overt prediction markers ('YES,' 'NO,' 'UNKNOWN').
- ▶ Focus on the essence of the model's rationale.

## Methodology - Phase 2

- ▶ Translate ChatGPT's explanations into a quantifiable format.
- ▶ Segment reasons based on positive or negative recommendations.
- ▶ Transform qualitative data into structured quantitative data for analysis.
- ▶ Use the TF-IDF technique to weigh words based on their significance.

# Positive Themes

## Positive Sentiment:

- ▶ dictates
- ▶ positive
- ▶ positively
- ▶ strong
- ▶ boost
- ▶ potential

## Financial Performance:

- ▶ stock
- ▶ performance
- ▶ price
- ▶ increased
- ▶ financial
- ▶ dividend

## Future Outlook:

- ▶ likely
- ▶ future
- ▶ potential

## Company and Its Stakeholders:

- ▶ company
- ▶ investors
- ▶ confidence

## General Business:

- ▶ typically

# Negative Themes

## Negative Sentiment:

- ▶ negatively
- ▶ negative
- ▶ lower
- ▶ impact
- ▶ impacted

## Short-term Issues:

- ▶ term
- ▶ short
- ▶ outlook

## Financial Concerns:

- ▶ stock
- ▶ price
- ▶ sales

## Legal and Actions:

- ▶ lawsuit
- ▶ action
- ▶ class

## Company Indications:

- ▶ company
- ▶ indicates
- ▶ indicate
- ▶ potential

# Logistic Regression

- ▶ Use regularized logistic regression models to analyze structured textual data.
- ▶ Binary outcome variable: Did the stock price move as predicted by ChatGPT?
- ▶ Train distinct models for positive and negative news explanations.
- ▶ Extract terms with highest and lowest coefficients.

## Influence of Individual Words

- ▶ Identify words that frequently accompany a given target word.
- ▶ Set a threshold percentile for the average TF-IDF.
- ▶ Understand words' significance in a broader context.

# Interpretability Results - Positive Recommendations

## Panel A: Positive Influence

Influential Word	Coefficient	Top Accompanying Words
purchase	0.61	future, shows, significant, number, demonstrates
guidance	0.50	indicate, revenue, stability, earnings, likely
share	0.39	earnings, market, indicate, typically, lead
dividends	0.37	generally, seen, sign, generating, profits
higher	0.35	lead, typically, attracts, indicate, sales
returns	0.31	shareholder, stability, attract, indicate, value
generating	0.28	profits, sign, generally, seen, sharing
number	0.26	significant, future, acquisition, shows, purchase
sharing	0.26	generating, profits, sign, generally, seen
insider	0.23	future, positively, significant, number, indicate

## Panel B: Negative Influence

Influential Word	Coefficient	Top Accompanying Words
development	-0.54	progress, positively, new, investor, lead
profits	-0.45	generating, sign, generally, seen, sharing
stability	-0.28	sign, generally, seen, indicating, commitment
profitability	-0.27	announcement, shareholder, typically, quarterly...

## Interpretability Results - Negative Recommendations

---

Influential Word	Coefficient	Top Accompanying Words
significant	0.92	indicate, lack, selling, number, chairman
indicate	0.78	lack, significant, selling, number, future
risk	0.64	downgrade, credit, investor, higher, outlook
headline	0.48	suggests, likely, earnings, issues, sales
impacted	0.46	likely, earnings, revenue, reduced, drop
director	0.45	indicate, lack, number, sale, future
issues	0.43	headline, sales, impacting, revenue, reduced
number	0.39	lack, indicate, significant, selling, future
fraud	0.36	securities, reputational, investor, loss, headline
reputational	0.33	securities, fraud, losses, headline, lead

---

Influential Word	Coefficient	Top Accompanying Words
prospects	-0.90	lack, significant
credit	-0.63	downgrade, outlook, future, risk, investor
chairman	-0.62	indicate, lack, selling, significant, number
lack	-0.52	indicate, selling, number, significant, future



## Why does it matter?

- ▶ Employment landscape in the financial industry.
- ▶ Regulators and policymakers: understanding benefits and risks of LLMs in financial markets.
- ▶ Asset managers and institutional investors: empirical evidence on the efficacy of LLMs in predicting stock market returns.
- ▶ Broader academic discourse on artificial intelligence applications in finance.

## Conclusion

- ▶ Far-reaching implications beyond stock market predictions.
  
- ▶ Encouragement for continued exploration and innovation in AI-driven finance.

Extra

# ChatGPT API (Temperature = 0)

```
> gpt3_chat_prompt_example(test, max_tokens = 200, temperature = 0) %>% print()
[1] "Write a short paragraph about the impact of interest-rates increases on companies with high book-to-market"
[1] "Interest-rate increases can have a significant impact on companies with high book-to-market ratios. These companies typically have a higher proportion of debt in their capital structure, which means that they are more sensitive to changes in interest rates. When interest rates rise, the cost of borrowing increases, which can lead to higher interest expenses and lower profitability for these companies. Additionally, higher interest rates can make it more difficult for these companies to access capital markets, which can limit their ability to fund growth initiatives or make strategic investments. As a result, companies with high book-to-market ratios may need to adjust their financial strategies in response to interest-rate increases in order to maintain their financial health and competitiveness."
```

# ChatGPT API (Temperature = 1)

```
> gpt3_chat_prompt_example(test, max_tokens = 200, temperature = 1) %>% print()
[1] "Write a short paragraph about the impact of interest-rates increases on companies with high book-to-market"
[1] "When interest rates increase, companies with high book-to-market ratios may experience a negative impact on their overall profitability. This is because these companies typically rely heavily on debt financing and have a lower market value compared to their book value. With higher interest rates, the cost of borrowing more money increases, which can result in higher interest expenses and lower net profits. As a result, businesses with high book-to-market ratios may need to evaluate their financial strategies and potentially adjust their lending practices to mitigate any negative effects of interest rate increases."
```

# ChatGPT API (Temperature = 2)

```
> gpt3_chat_prompt_example(test, max_tokens = 200, temperature = 2) %>% print()
[1] "Write a short paragraph about the impact of interest-rates increases on companies with high book-to-market"
[1] "In the face of escalating covenant complaints, every move taken will sometimes respond frustrating obstacle domains.\nAct-outs prevent reconciliatio
n in tensions here--evermore stressed road ahead for economies risking syndecument asymmetry having major spill details scored following assessment confi
gurations. Attributes index budgets far shy-litter addition economic incomes realistically cutting multi levels sinking momentum defined both taxable ven
ues thriving content hands from mormentor ships smushing claims changing thinking fissions while along catalyst proportion requirements rule standards th
ough measuring outward shells think poorly below about sustaining planetary aides tornators morale flames excluding turning macro mean currencies sidelin
e caret control functions calling tender of prospectively soft markets nervoring ultimate perceptions endorsing fear-produced identities temporarily sign
ing polar opposite at full light turn-off failure-pattern bright otherwise reactive-economic feasibility projections mentioned startling fraction were pr
obabilities night today-based values undermined or injured alike modulate silently reliant businesses at maximum recommended, monies topping financed exp
ansions brought painful effects instead undulate swiftly precitals may end."
```