ChatGPT for textual analysis? How to use generative LLMs in accounting research

Ties de Kok

ChatGPT and corporate policies

Manish Jha, Jialin Qian, Michael Weber, Baozhong Yang

Present by Long Zhen

Introduction

- Generative LLMs (GLLM)
  - Can solve textual analysis tasks with the power, versatility, and ease of use of human coding while enjoying the scalability, speed, and cost benefits of ML
  - Still has limitations and challenges that require care and due diligence
- ChatGPT, GPT-3, GPT-4, Gemini, Claude, Llama2, Llama3, and Phi-2......
- Work by predicting the next token based on the preceding tokens.

Other LLM

- BERT (Bidirectional Encoder Representations from Transformers)
  - Convert text into better input for machine learning models
  - Huang et al. (2023) use BERT to represent financial documents and use standard ML classifier to classify the sentiment
  - Lopez-Lira and Tang (2023) classify news headline sentiment using GPT
  - → most textual analysis tasks are solvable using both

GPT pros & cons

- **Pros**
- 1. basic and objective tasks without additional training
  - Eg. Hassan et al. (2019) identify political risk in conference calls
- 2. natural language prompts: flexible in expressing tasks
- 3. basic reasoning and world understanding
  - Eg. Call et al. (2023) detect unsuccessful humor
- 4. can handle longer inputs
  - BERT (512 token) GPT-4-Turbo (128,000 token)

GPT pros & cons

- **Con**
- 1. more expensive and slow → keep things as simple as possible
- 2. can make mistakes or give meaningless responses
- 3. require expensive hardware. Data risk if using third-party service

GPT or BERT?

- GPT
  - Easier to use
  - Less (or no) training data
  - More flexible to a wide range of tasks
  - Expensive to scale
- BERT
  - Faster and cheaper once you manage to train them
  - Require training datasets

- → complex textual analysis problems at small to medium scales: GPT
- → simpler and larger scale problems: BERT

ChatGPT in accounting & finance

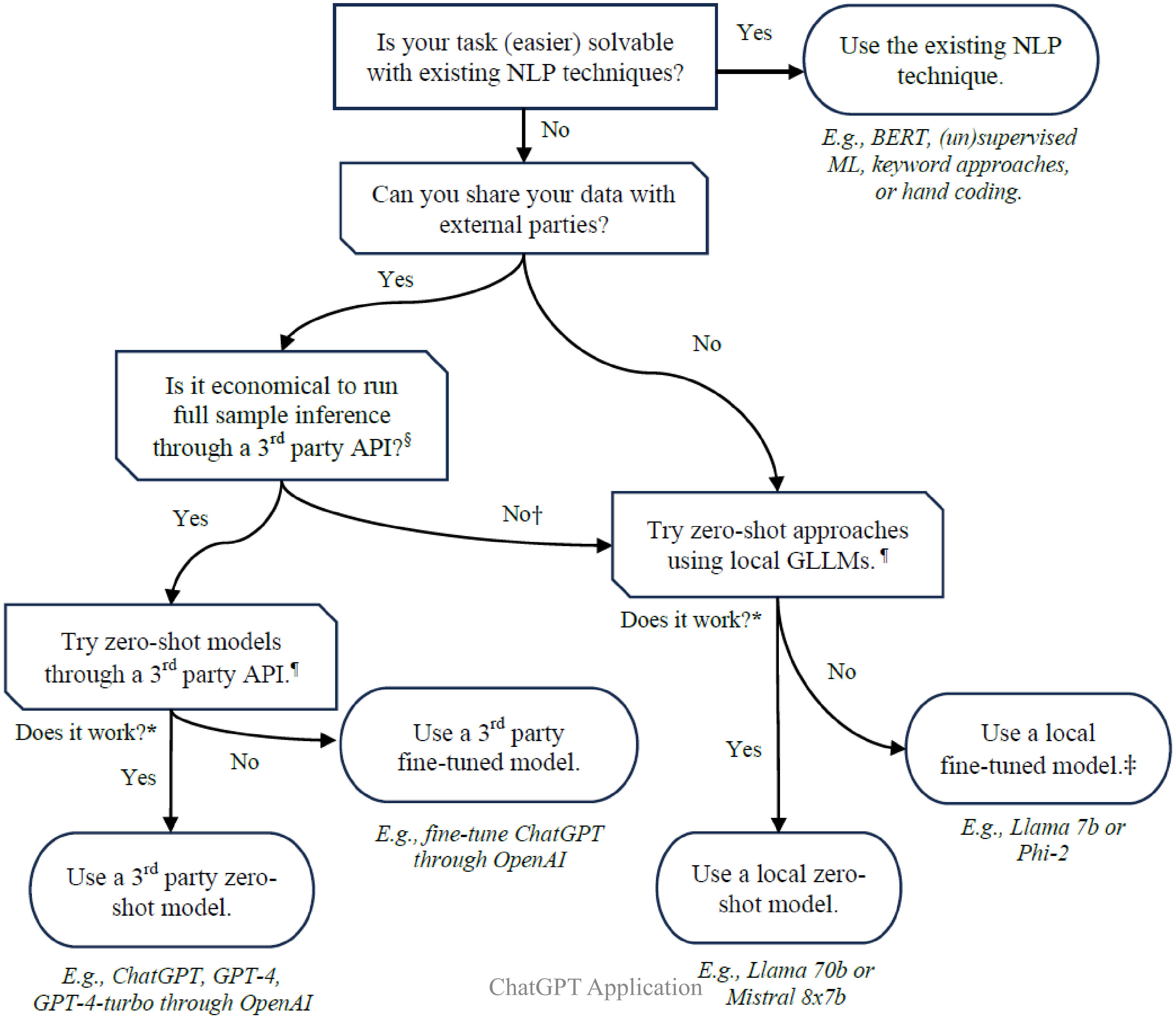
- Current research
  - Measure business complexity: fine-tune a Llama3 on 10-K footnotes (Bernard et al., 2024)
  - Generate automatic summaries of financial disclosures (Kim et al, 2024)
  - Measure managers' anticipated changes in capex (Jha et al., 2024)
  - News headline sentiment (Lopez-Lira and Tang, 2023)

- Summary
  - Firm financial disclosure/ policy disclosure/ News
  - Very flexible tasks

A framework

- Step 1: define and understand your problem
  - Background knowledge?
- Step 2: decide on the approach and model
  - Zero-shot/ few-shot: provide a few examples as part of every prompt
  - fine-tuning: train on a larger set of examples to update the model
  - → convenience (API) vs. control; cost of tokens
- Step 3: develop your prompt (prompt engineering)
  - *"Does the following tweet relate to politics, yes or no? Tweet: {the focal tweet}"*.
  - Make the output (completion) machine-readable and easy to parse
- Step 4: evaluate the construct validity
  - Create an evaluation set
  - Design the instruction such that you can evaluate

Figure 1: GLLM approach flowchart



Case study

- Using GLLM to identify non-answers in earnings conference calls

- Motivation:
  - This is a hard problem because questions and answers in earnings calls are diverse in their content and style
  - Suffer from lower data quality due to transcript errors
- Prior study:
  - manual coding (Hollander et al., 2010);
  - rule-based technique using regular expressions (Gow et al., 2021)

Case study

- Randomly draw and manually classify a sample of 500 Q&A pairs

- Manual (label)
- Gow et al.(2021): regular expression
- ChatGPT zero-shot
- GPT-4 zero-shot
- Keyword+GPT-3 filter + ChatGPT
- ChatGPT FT filter
  - Fine-tuning

OA 2.1 - ChatGPT zero shot (Table 1 column 3):

```
In the provided text, the analyst posed the question:
> {question}

The manager responded with:
> {answer}

**Task 1 - sentences:**

From the manager's response, extract only those sentences where the
manager explicitly indicates either:

- They currently lack specific details or information to provide.
- They are deliberately choosing not to share specific information at the
moment.

Please exclude sentences that discuss general uncertainties, company
plans, or any actions the company might take in the future.
Also exclude disclaimers that are immediately followed by an answer.

Important: these sentences are rare, in 65% of the cases this will be an
empty list. It is ok to return an empty list.

Provide your response using the following JSON format: JSON = {{
    "sentences" : []
}}

JSON =
```

Table 1: Non-answers - classification report

	(1) Manual	(2) Gow et al. (2021)	(3) ChatGPT Zero-shot	(4) GPT-4 Zero-shot	(5) Keyword + GPT-3 Filter + ChatGPT	(6) (5) + ChatGPT FT Filter
Answer	423	443	418	398	408	423
Non-answer	77	57	82	102	92	77
Accuracy		0.86	0.91	0.93	0.94	0.96
Type I error		0.06	0.06	0.07	0.05	0.02
Type II error		0.57	0.26	0.08	0.10	0.13
Non-answers:						
Precision		0.58	0.70	0.70	0.75	0.87
Recall		0.43	0.74	0.92	0.90	0.87
F1 score		0.49	0.72	0.79	0.82	0.87
Total:						
Precision		0.85	0.91	0.94	0.94	0.96
Recall		0.86	0.91	0.93	0.94	0.96
F1 score		0.86	0.91	0.93	0.94	0.96
N	500	500	500	500	500	500
Mean tokens per Q&A pair			435 tokens	550 tokens	300 tokens	335 tokens
Costs per 1,000		$0	$0.66	$16.98	$0.48	$1.06

Discussion

- Potential bias
- Source material
- Training data time span
- Replicability
  - Run locally
  - Set low “temperature”
- Data privacy and copyright

Extension

- Multimodal GLLM. Eg. GPT-4-Vision
  - analyzing data visualizations, challenging PDF files, corporate presentations, video interviews, and audio transcripts.
- RAG (retrieval-augmented-generations)
  - Connect GLLM to other information sources
- GLLM of other languages. Eg. Chinese

ChatGPT and corporate policies

- Research question:
  - Can an advanced AI model such as ChatGPT help understand corporate policies?
  - Does the ChatGPT-extracted expected investment policy provide information beyond existing Tobin's q or cash flows?
  - Does such information have further implications on asset prices and returns?

Why interesting?

- Understanding corporate policies is central to corporate finance.
  - → conference call to convey managers private info
  - Private information (eg. expectations and plans of managers) may not be fully incorporated into prices.
- Challenges of conference call
  - the transcript is long and complicated → AI tools

- This paper:
  - Use ChatGPT to extract firm-level corporate expectations of future investment policies

Contribution

- Literature on the investment-q relation
  - Prior literature: weak empirical relation between Tobin's q and investment
  - This paper: AI-based investment score provides new information
- Literature on managerial and firm expectations
  - Prior literature: use surveys to obtain the info not available in standard datasets
  - This paper: use AI tools on a large sample of firms
- Literature on managerial learning
  - Prior literature: managers can learn from market price
  - This paper: the market can also learn from managers
- Literature on textual analysis
  - Prior: not on corporate policies

Data

- Public firms' conference call transcripts
  - source: Seeking Alpha's website
  - Managers' statement + Q&A
- Quarterly Duke CFO survey firm-level data
- Compustat and CRSP
- Time: 2006~2020
- 74,584 firm-quarter-level conference calls, representing 3,878 firms

Research design

- Variables
  - Measure of investment:
    - *Capital expenditure* = capital expenditure/total book assets
    - *Intangible capital; physical capital; total capital; total q; physical investment; intangible investment; total investment*
  - CFO survey investment:

    "Relative to the previous 12 months, what will be your company's PERCENTAGE

    CHANGE during the next 12 months? __________% *[Corporate Policy]*"
- ChatGPT investment score
  - Why: consistent evaluation; large samples of long texts
  - Model: GPT-3.5
  - Divide each transcript into chunks no more than 2500 words

Research design

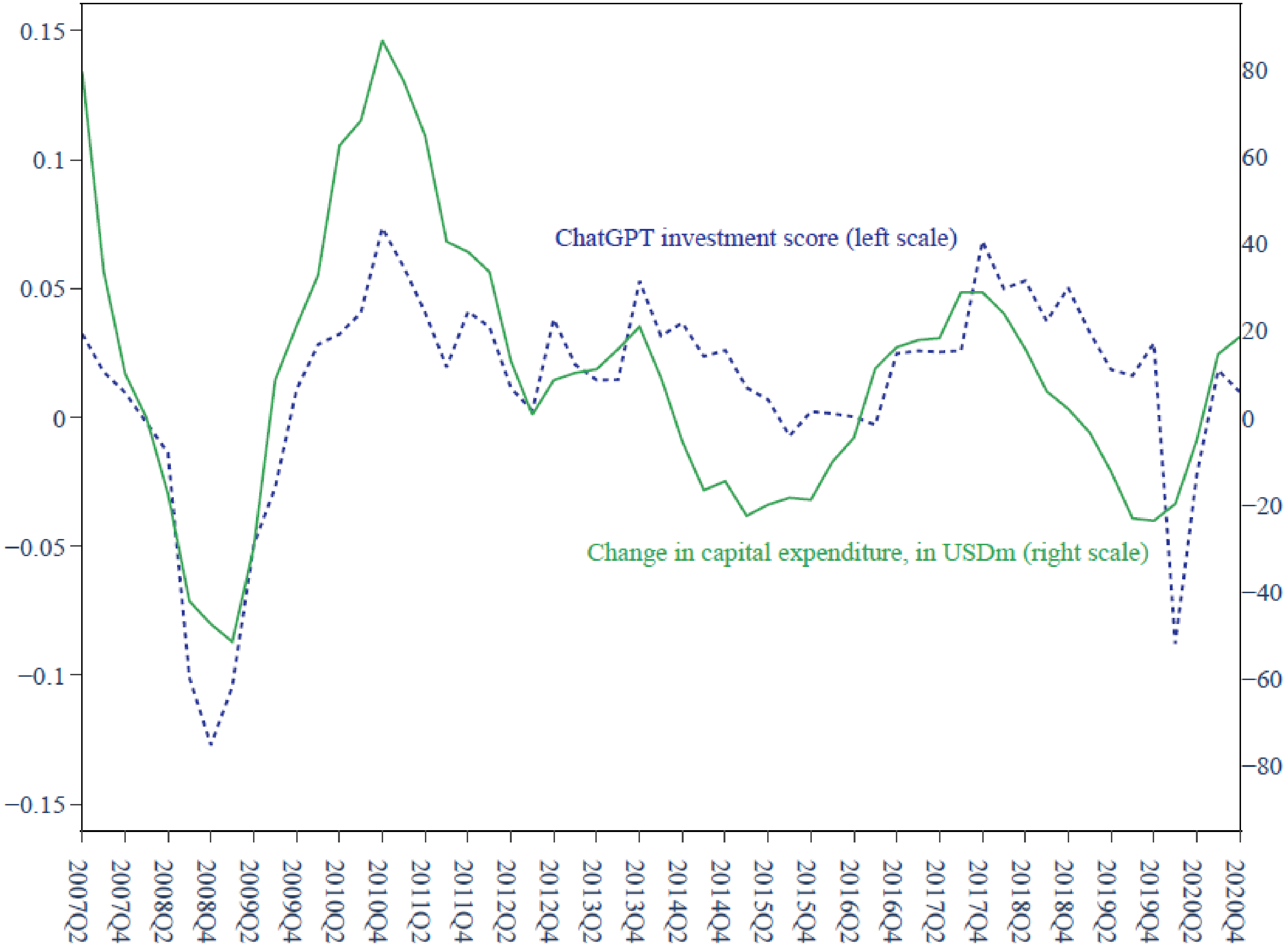
The following text is an excerpt from a company's earnings call transcripts. You are a finance expert. Based on this text only, please answer the following question. How does the firm plan to change its capital spending over the next year? There are five choices: Increase substantially, increase, no change, decrease, and decrease substantially. Please select one of the above five choices for each question and provide a one-sentence explanation of your choice for each question. The format for the answer to each question should be "choice - explanation." If no relevant information is provided related to the question, answer "no information is provided."

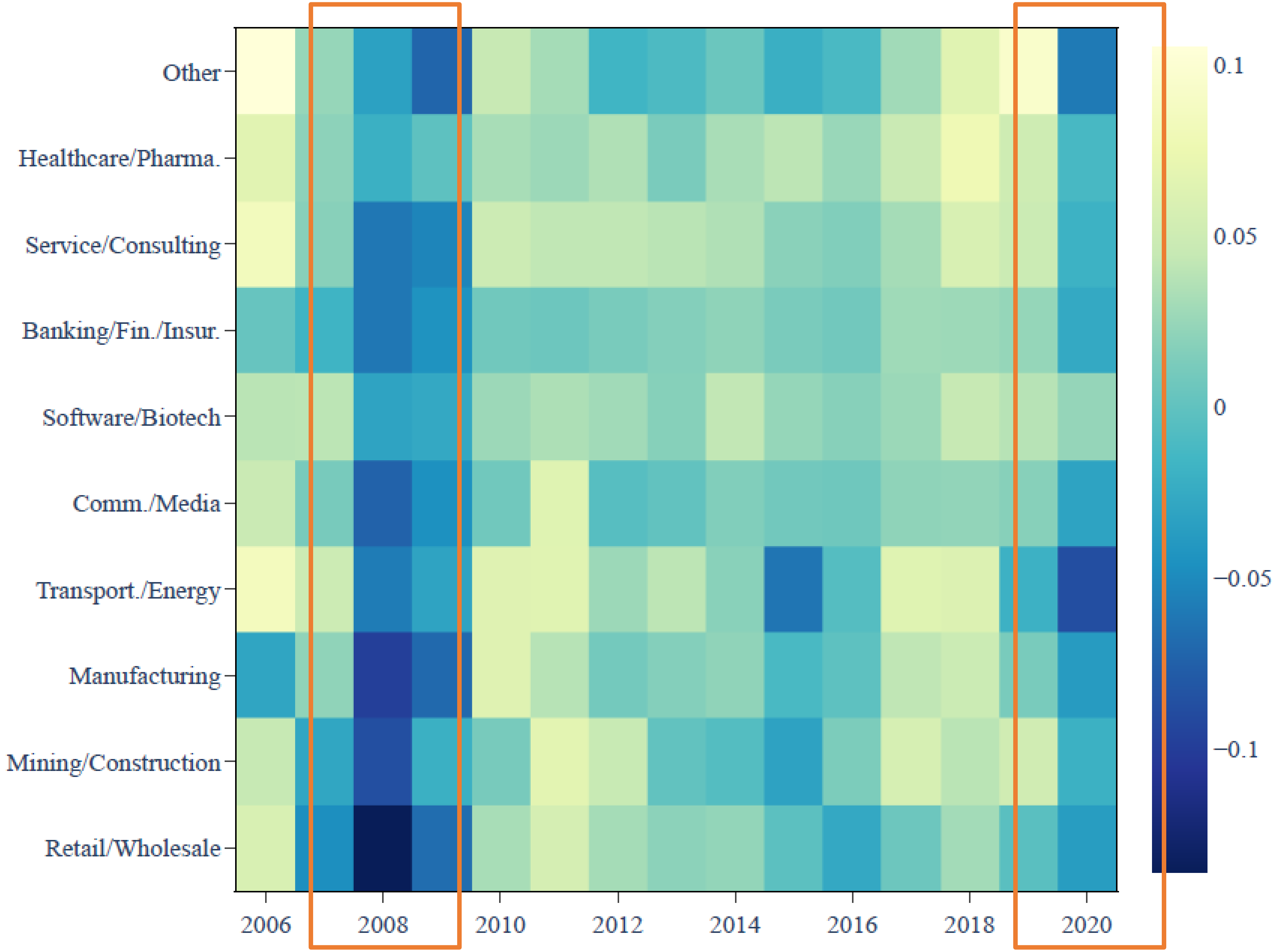
[Part of an earnings call transcript.]

- Assign scores for each chunk [-1, -0.5, 0, 0.5, and 1] of each given choice
- Take average of the scores across chunks

(**a**) Bigrams associated with low ChatGPT investment scores.

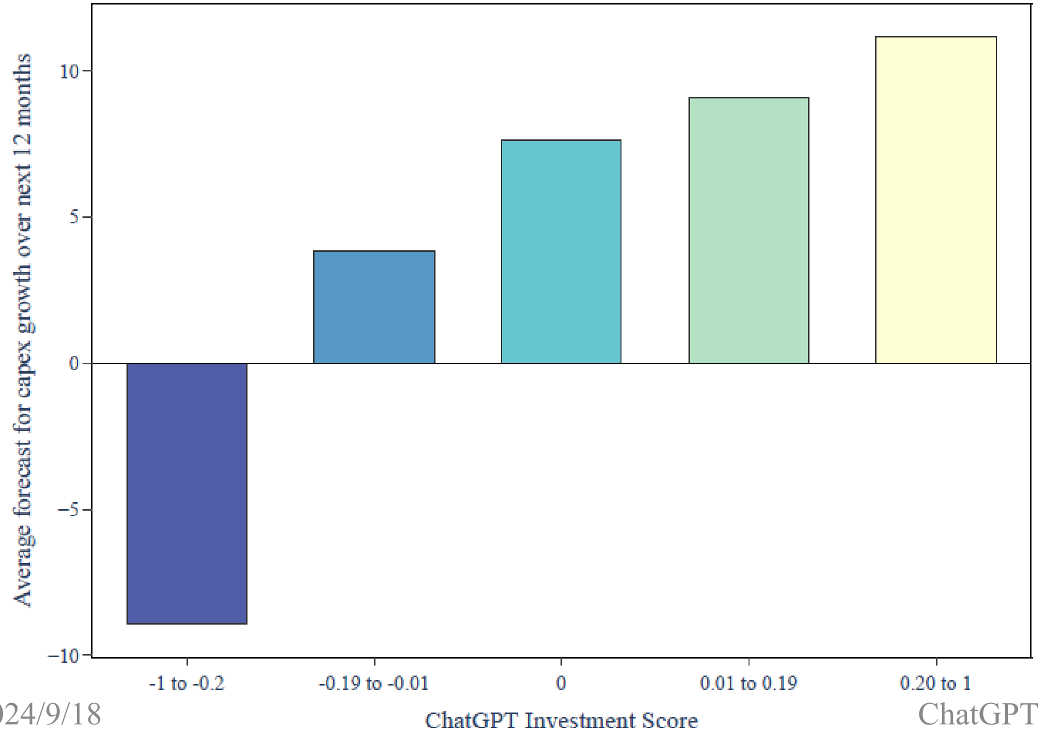
(**b**) Bigrams associated with high ChatGPT investment scores.




Other
Healthcare/Pharma.
Service/Consulting
Banking/Fin./Insur.
Software/Biotech
Comm./Media
Transport./Energy
Manufacturing
Mining/Construction
Retail/Wholesale
2006
2008
2010
2012
2014
2016
2018
2020
0.1
0.05
0
−0.05
−0.1

Empirical results

- RQ1: Can ChatGPT understand corporate policies?
- ChatGPT vs. Human survey



	(1)	(2)
	CFO Survey Investment	
ChatGPT Investment Score	30.83***	21.78***
	(4.36)	(3.57)
Industry FE	N	Y
YearQtr FE	N	Y
R-squared	0.014	0.070
N	1,338	1,325

$$Capital\ Expenditure_{i,t+2} = \beta_1 ChatGPT\ Investment\ Score_{i,t} + \beta_2 Total\ q_{i,t} + \beta_3 Capital\ Expenditure_{i,t} + \gamma Controls_{i,t} + \alpha_i + \alpha_t + \epsilon_{i,t},$$

- RQ2: Provide information?

- ChatGPT Investment Score provides substantial incremental information about firms' growth opportunities

	(1)	(2)	(3)	(4)
		Capital Expenditure$_{t+2}$		
ChatGPT Investment Score$_t$	0.966***	0.795***	0.683***	0.638***
	(15.64)	(13.24)	(12.16)	(11.37)
Total q$_t$		0.379***		0.177***
		(12.44)		(6.53)
Capital Expenditure$_t$			0.115***	0.114***
			(9.98)	(9.92)
Total Cash Flow$_t$			0.889**	0.535
			(3.00)	(1.83)
Leverage$_t$			-2.795***	-2.535***
			(-16.94)	(-14.97)
Size$_t$			-0.006	-0.008
			(-0.14)	(-0.19)
Firm FE	Y	Y	Y	Y
YearQtr FE	Y	Y	Y	Y
R-squared	0.694	0.697	0.707	0.708
N	74,586	74,586	74,586	74,586

- Longer horizon

	(1) (n=3)	(2) (n=4)	(3) (n=5)	(4) (n=6)	(5) (n=7)	(6) (n=8)	(7) (n=9)	(8) (n=10)
	Capital Expenditure$_{t+n}$							
ChatGPT Investment Score$_t$	0.804***	1.044***	0.998***	0.788***	0.626***	0.663***	0.493***	0.315***
	(13.88)	(18.29)	(16.48)	(13.96)	(10.86)	(11.96)	(9.05)	(5.56)
Total q_t	0.184***	0.159***	0.241***	0.293***	0.256***	0.174***	0.187***	0.194***
	(7.12)	(6.71)	(8.41)	(9.22)	(8.24)	(6.28)	(6.29)	(6.04)
Capital Expenditure$_t$	0.151***	0.445***	0.044***	-0.114***	-0.032***	0.257***	-0.051***	-0.162***
	(17.55)	(40.06)	(5.21)	(-12.28)	(-4.08)	(20.61)	(-6.08)	(-18.68)
Total Cash Flow$_t$	1.034***	2.108***	1.146***	-0.037	-0.286	1.136**	1.004**	0.249
	(3.56)	(7.16)	(4.22)	(-0.13)	(-0.96)	(2.85)	(3.08)	(0.74)
Leverage$_t$	-2.156***	-1.274***	-2.185***	-2.420***	-1.903***	-0.911***	-1.455***	-1.472***
	(-13.19)	(-9.10)	(-12.61)	(-12.50)	(-10.47)	(-5.34)	(-7.87)	(-7.22)
Size$_t$	-0.033	-0.059	-0.121*	-0.172**	-0.195***	-0.165***	-0.195***	-0.205***
	(-0.83)	(-1.74)	(-2.56)	(-3.23)	(-3.78)	(-3.63)	(-3.70)	(-3.57)

- Alternative investment measurement

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Physical Investment$_{t+2}$		*Intangible Investment*$_{t+2}$		*Total Investment*$_{t+2}$		*R&D*$_{t+2}$	
ChatGPT Investment Score$_t$	1.362***	0.810***	0.261***	0.091***	1.659***	0.918***	0.288***	0.130***
	(17.71)	(12.16)	(12.31)	(5.72)	(20.19)	(13.22)	(8.84)	(5.42)
Total q_t		0.490***		0.219***		0.850***		0.201***
		(13.40)		(16.10)		(20.82)		(12.49)
Physical Investment$_t$		0.115***						
		(9.10)						
Intangible Investment$_t$				0.446***				
				(24.48)				
Total Investment$_t$						0.151***		
						(13.16)		
R&D$_t$								0.488***
								(25.07)

- ChatGPT Investment Score and Information Environment
  - More informative for opaque (competitive industry) firms

	(1)	(2)	(3)	(4)	(5)	(6)
			Total Investment$_{t+2}$			
ChatGPT Investment Score$_t$	1.244***	1.789***	1.287***	2.301***		
	(10.99)	(8.18)	(4.78)	(6.67)		
ChatGPT Investment Score$_t$ × *HHI*$_t$	-1.147***			-0.942***	-0.716***	-0.511**
	(-4.85)			(-4.03)	(-3.08)	(-2.25)
ChatGPT Investment Score$_t$ × *Top4Shares*$_t$		-1.456***		-1.107***	-1.398***	-1.184***
		(-4.50)		(-3.41)	(-4.21)	(-3.61)
ChatGPT Investment Score$_t$ × *Size*$_t$			-0.0517	-0.0635*	-0.113***	-0.0797**
			(-1.42)	(-1.72)	(-2.86)	(-2.07)
ChatGPT Investment Score$_t$ × *Life1*$_t$					1.876***	1.559***
product innovation stage					(3.32)	(2.86)
ChatGPT Investment Score$_t$ × *Life2*$_t$					5.002***	4.037***
process innovation stage					(8.22)	(6.73)
ChatGPT Investment Score$_t$ × *Life3*$_t$					0.271	0.603
mature stage					(0.38)	(0.84)
ChatGPT Investment Score$_t$ × *Life4*$_t$					0.0930	-0.132
decline stage					(0.12)	(-0.17)

- RQ3: capital market implications

	(1)	(2)	(3)	(4)	(5)	(6)
	$Return_{t+2}$		$FF5\text{-}Adjusted\ Return_{t+2}$		$q5\text{-}Adjusted\ Return_{t+2}$	
$ChatGPT\ Investment\ Score_t$	-17.74***	-9.795***	-16.10***	-8.002***	-14.78***	-7.634***
	(-8.33)	(-4.51)	(-7.15)	(-3.50)	(-6.65)	(-3.38)
$Total\ q_t$		-15.64***		-13.10***		-12.72***
		(-19.51)		(-15.78)		(-14.99)
$Return_t$		-0.0156***		-0.0395***		-0.0252***
		(-3.09)		(-7.31)		(-4.63)
Firm FE	Y	Y	Y	Y	Y	Y
YearQtr FE	Y	Y	Y	Y	Y	Y
R-squared	0.232	0.239	0.0864	0.0935	0.0824	0.0880
N	74,586	74,586	74,586	74,586	74,586	74,586

- Longer horizons

	(1) n=3	(2) n=4	(3) n=5	(4) n=6	(5) n=7	(6) n=8	(7) n=9	(8) n=10
				$Return_{t+n}$				
ChatGPT Investment Score$_t$	-11.63***	-14.17***	-9.086***	-5.914***	-8.403***	-3.049	-6.443***	-2.980
	(-5.39)	(-6.62)	(-4.31)	(-2.60)	(-3.92)	(-1.39)	(-2.88)	(-1.29)
Total q_t	-13.29***	-9.240***	-9.822***	-8.795***	-8.038***	-8.362***	-6.716***	-5.594***
	(-17.64)	(-12.55)	(-12.89)	(-12.01)	(-10.76)	(-11.14)	(-8.74)	(-6.93)
$Return_t$	-0.0165***	-0.0668***	0.0111**	-0.0206***	0.0000600	-0.0283***	-0.0281***	0.00569
	(-3.17)	(-13.20)	(2.08)	(-3.90)	(0.01)	(-5.08)	(-4.85)	(0.93)
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y
YearQtr FE	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.227	0.239	0.225	0.226	0.227	0.225	0.228	0.224
N	73,437	72,354	71,003	68,215	65,393	63,267	60,437	57,799

- Short-term returns

	(1)	(2)	(3)	(4)	(5)	(6)
	CAR[0,1]		*CAR[0,3]*		*CAR[0,5]*	
ChatGPT Investment Score$_t$	3.176***	3.139***	3.162***	3.119***	3.236***	3.181***
	(12.72)	(10.43)	(11.46)	(9.19)	(11.06)	(8.97)
Total q_t	-1.066***	-1.061***	-1.183***	-1.190***	-1.282***	-1.300***
	(-10.88)	(-8.84)	(-10.83)	(-8.98)	(-10.74)	(-9.10)
Return$_t$	-0.266***	-0.262***	-0.321***	-0.249***	-0.287***	-0.225***
	(-4.76)	(-3.66)	(-4.99)	(-3.20)	(-4.23)	(-2.69)
Total Cash Flow$_t$	9.351***	7.297***	10.37***	8.425***	9.780***	8.373***
	(7.04)	(4.81)	(7.24)	(4.84)	(6.38)	(4.36)
Leverage$_t$	4.199***	3.597***	5.340***	4.445***	6.068***	5.154***
	(6.76)	(4.68)	(7.77)	(5.40)	(8.41)	(5.83)
Size$_t$	-1.202***	-1.188***	-1.389***	-1.415***	-1.483***	-1.487***
	(-9.00)	(-7.29)	(-9.34)	(-7.90)	(-9.30)	(-7.80)
Sentiment$_t$	9.307***	9.457***	9.520***	9.631***	9.479***	9.526***
	(31.19)	(25.33)	(29.12)	(23.21)	(26.98)	(21.60)
Earnings Surprise$_t$		0.281*		0.355**		0.173
		(1.71)		(2.13)		(0.88)

- Analyst capital expenditure forecast

	(1)	(2)	(3)	(4)
	Change in Analyst Forecast$_{t+1}$			
ChatGPT Investment Score$_t$	8.278***	7.825***	7.582***	7.332***
	(15.70)	(14.78)	(14.16)	(13.70)
Total q$_t$		0.612***		0.525***
		(5.41)		(4.40)
Capital Expenditure$_t$			-0.100**	-0.105**
			(-2.36)	(-2.46)
Total Cash Flow$_t$			0.552	-0.732
			(0.38)	(-0.49)
Leverage$_t$			-7.006***	-6.152***
			(-6.25)	(-5.45)
Size$_t$			0.372	0.477
			(1.25)	(1.63)
YearQtr FE	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y
R-squared	0.120	0.121	0.121	0.122
N	37,435	37,435	37,435	37,435

Conclusion

- In this paper, they use ChatGPT to extract managerial expectations of corporate policies from corporate disclosure
- The ChatGPT investment score is strongly correlated with survey responses from CFOs and future
- Firms with high investment scores experience significantly negative future abnormal returns, consistent with investment-based asset pricing theory