ChatGPT in Finance and Accounting Research

Presented by Long Zhen

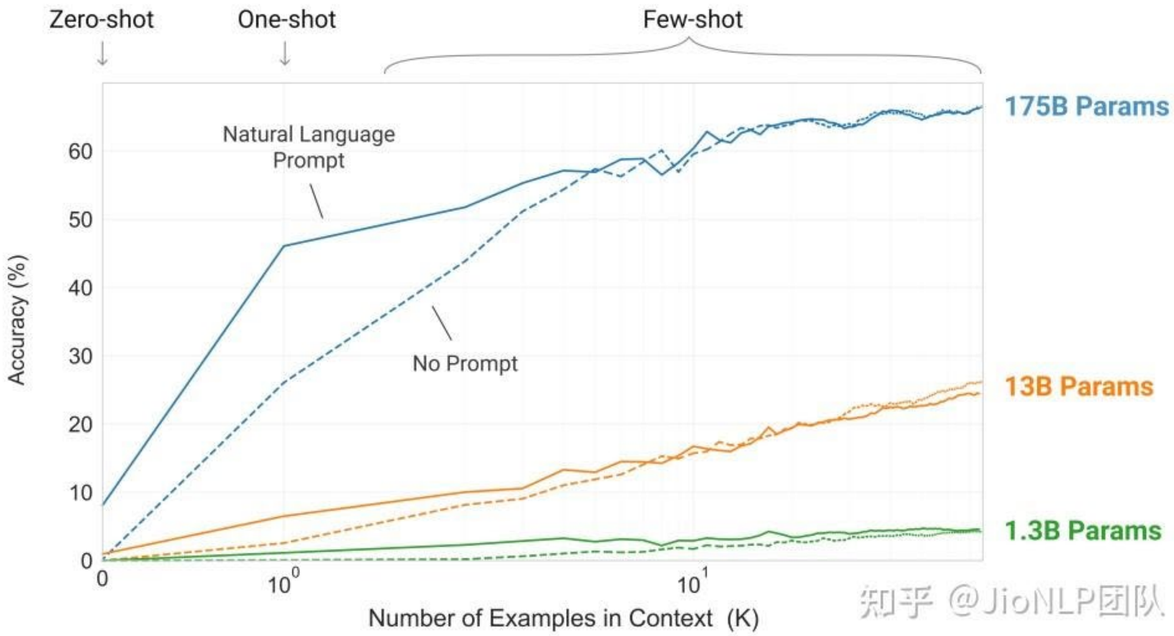
What is ChatGPT

- ChatGPT is an AI-powered conversational agent developed by OpenAI.
  - GPT (Generative Pre-trained Transformer)
- Feature:
  - Natural Language Understanding (NLU): ChatGPT comprehends and responds to human language inputs.
  - Conversational Abilities: It engages in dialogue, mimicking human-like conversation.
  - Adaptability: Capable of learning and adapting to diverse conversational contexts.
  - Large Knowledge Base: Informed by extensive pre-training on vast datasets, allowing it to generate contextually relevant responses.

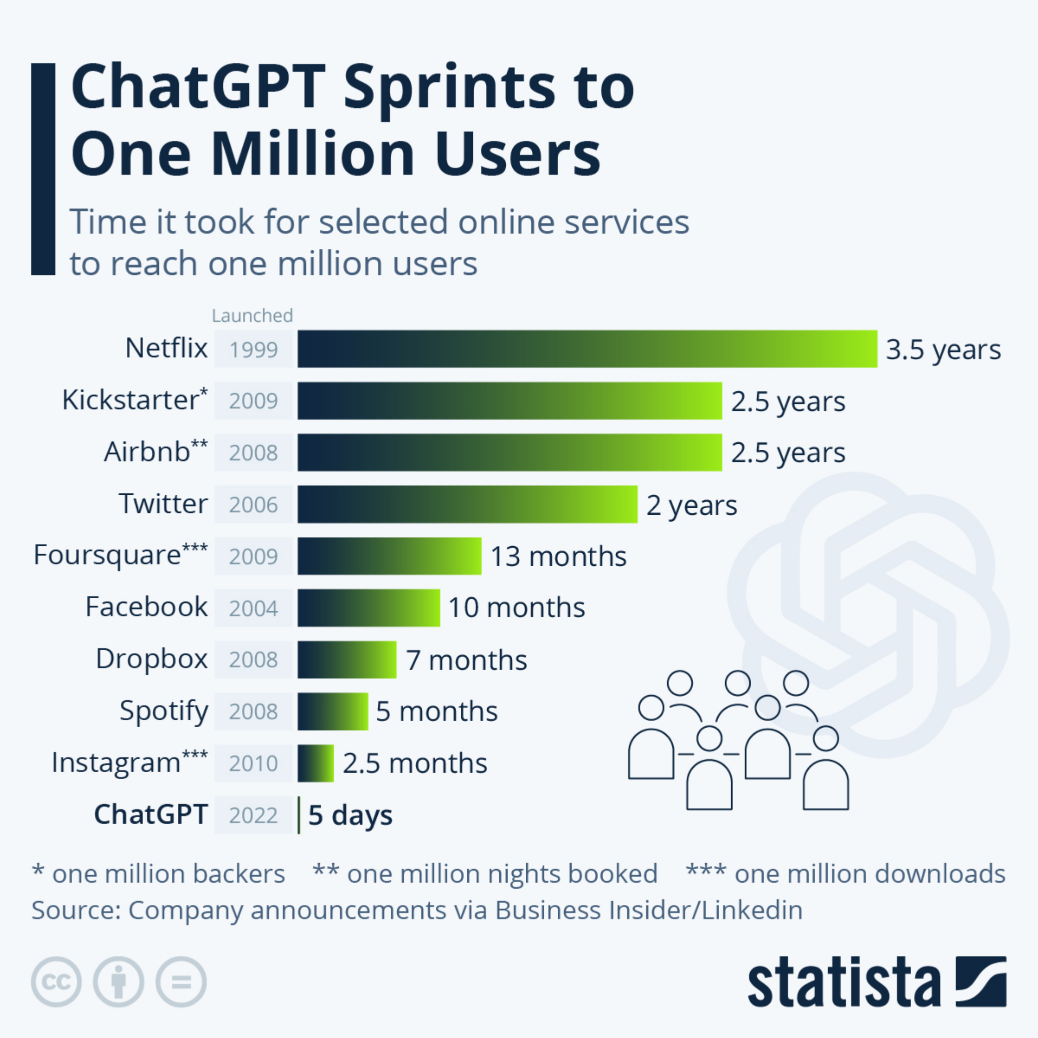
—— produced by ChatGPT 3.5

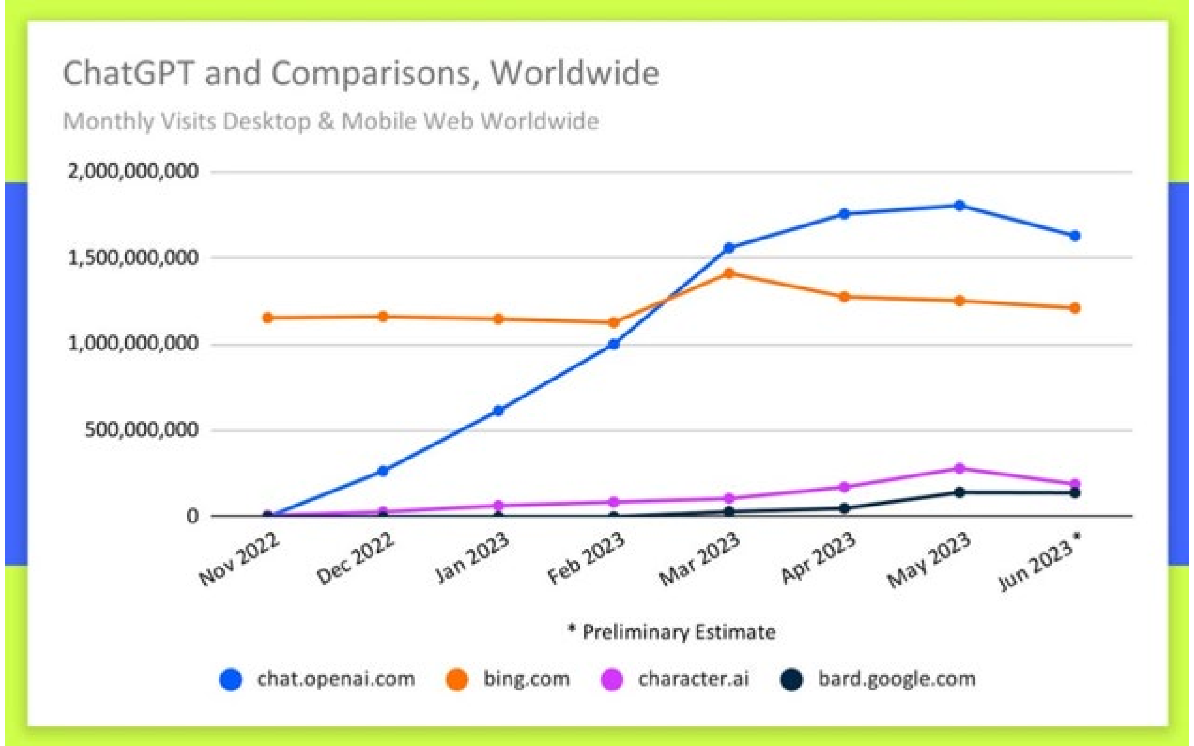
What is ChatGPT

- Based on the transformer architect
- ChatGPT prompts are a way to facilitate conversations between users and their AI-driven chatbot
  - Make responses more relevant



Transforming





…for finance and accounting research

- Ultimate goal: resource allocation ← information
- A framework:
  - Awareness cost
  - Acquisition cost
  - Integration cost
- → How ChatGPT makes a difference in information processing?

Bloated disclosure: Can ChatGPT help investors process information?

Alex G. Kim
Maximilian Muhn
Valeri V. Nikolaev
R&R at JAE

Research question

- Whether ChatGPT help investors process information from corporate disclosure by summarizing the contents?
- What are the determinants and consequences of disclosure bloat?

→ A bigger picture:

- How LLMs and Generative AI shape the information environment of firms and investors?

Why interesting

- The exponential growth of textual data + popularity of LLM
  - The economic usefulness of these tools in information processing remains unclear
- Corporate disclosure provide an ideal platform for understanding the value of language modeling from a user's perspective
  - Information overload in corporate filings
    - Strategic disclosure (length, complexity)
    - SEC regulation: plain English
- GPT-3.5 is well-suited for analyzing corporate disclosure
  - Concise, effective, and humanly understandable

Contribution

- Literature on the economic usefulness of technology in analyzing textual data
  - Prior studies: introduction of EDGAR/XBRL by the SEC
    - e.g., Goldstein et al., 2023; Blankespoor, 2019
  - This study suggest that generative AI can advance financial reporting tech
- Literature on measures of textual information redundancies
  - Prior studies: readability/complexity → linguistic complexity (textual+content)
    - e.g., Li, 2008; Loughran and McDonald, 2014a; Bonsall et al.,2017
  - This study propose a novel measure bloat that avoid the complexity confusion

Contribution

- Literature on machine learning and AI in financial market
  - Contemporaneous works: sentiment analysis/ firm risk assessment
  - This study: the value of generative AI in processing complex disclosure
- Literature on value of language models in extracting information
  - Prior studies: measuring political risk/ climate risk/ climate change exposure…
  - This study measures disclosure information overload

Research design

- Data
  - Two narrative disclosures: MD&As in annual reports, earnings conference call
  - All US non-financial public firms from fiscal years 2009-2020
  - → 8,699 MD&As and 40,362 conference calls
  - Random 20% → 1,790 MD&As by 339 firms and 8,537 cc by 360 firms
  - Other data:
    - CRSP/TAQ/Compustat/IBES/Thomson Reuters 13-F filings
- GPT-3.5-Turbo API
- Prompt:
  - summarize the input text using only the information included in the text and to not restrict the length of the summary

Empirical results

- How effective are the summaries?
  - Length, sentiment, and readability (fog index; plain English measure)

Panel A. MD&A Sample								
Levels:								
		Raw Document			Summarized Document			
	N	Mean	Median	Std	Mean	Median	Std	Diff.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(5) – (2)
Length	1,790	17,901	14,254	13,151	3,779	3,433	1,882	-14,122***
Sentiment	1,790	-0.360	-0.371	0.202	-0.366	-0.438	0.316	-0.006
Fog	1,790	10.025	9.960	1.591	10.175	10.130	0.936	0.150***
Plain_Eng	1,790	0.289	0.286	0.039	0.290	0.290	0.011	0.000

Changes:						
				Percentiles		
	N	Mean	Std	p25	p50	p75
Δ*Length*	1,790	-14,122	11,788	-16,729	-10,462	-7,093
Δ*Sentiment*	1,790	-0.006	0.188	-0.132	-0.039	0.110
Δ*Fog*	1,790	0.150	1.144	-0.600	0.190	0.920
Δ*Plain_Eng*	1,790	0.000	0.036	-0.020	0.004	0.026

- whether the summaries capture the sentiment of the original document in a more definitive (precise) way

Panel A. MD&A Sample

Subsample 1: $Sentiment^{Raw}$ > Median								
		Raw Document			Summarized Document			
	N	Mean	Median	Std	Mean	Median	Std	Diff.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(5) – (2)
Length	897	13,997	12,445	7,493	3,637	3,419	1,602	-10,360***
Sentiment	897	-0.205	-0.240	0.147	-0.182	-0.276	0.322	0.024***
Fog	897	9.690	9.540	1.591	10.310	10.310	0.914	0.620***
Plain_Eng	897	0.284	0.282	0.039	0.289	0.289	0.011	0.005***

Subsample 2: $Sentiment^{Raw}$ < Median								
		Raw Document			Summarized Document			
	N	Mean	Median	Std	Mean	Median	Std	Diff.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(5) – (2)
Length	893	21,821	17,066	16,117	3,922	3,444	2,118	-17,899***
Sentiment	893	-0.515	-0.499	0.110	-0.551	-0.564	0.168	-0.036***
Fog	893	10.362	10.190	1.518	10.040	10.010	0.939	-0.322***
Plain_Eng	893	0.294	0.290	0.038	0.290	0.291	0.011	-0.004***

- Summarization appears to amplify textual sentiment.
- Summarized documents become slightly less readable

- Whether GPT summaries are indeed more informative compared to the originals

$$Abn\_Ret_{[0,1]it} = \beta Sentiment^{j}_{it} + \gamma \mathbf{X}_{it} + \delta_t + \theta_i + \varepsilon_{it}$$

Panel A. MD&A Sample

Dependent Variable = $Abn\_Ret_{[0,1]}$

	Raw Documents				Summarized Documents			
	Full		Pos	Neg	Full		Pos	Neg
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Sentiment^{Raw}$	-0.008	0.001	0.005	-0.028				
	(0.008)	(0.012)	(0.017)	(0.020)				
$Sentiment^{Sum}$					0.025***	0.051***	0.027***	0.099***
					(0.004)	(0.009)	(0.006)	(0.011)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	No	Yes	No	No	No	Yes	No	No
Industry FE	Yes	No	Yes	Yes	Yes	No	Yes	Yes
Cluster	Ind	Ind	Ind	Ind	Ind	Ind	Ind	Ind
N	1,790	1,790	897	893	1,790	1,790	897	893
Adjusted R^2	-0.009	0.041	-0.027	0.001	0.017	0.098	0.001	0.121

- → a flip side: quantify the degree of redundant

- Measuring disclosure bloat
  - *Bloat* is the difference between the length of the original document and that of its summary scaled by the length of the original

$$Bloat_{it} = \gamma \mathbf{X}_{it} + \delta_t + \theta + \varepsilon_{it}$$

- Determinants of disclosure bloat

Dependent Variable =	*Bloat*	
Sample =	MD&A	
	(1)	(2)
Log_ME	-0.012***	-0.004
	(0.001)	(0.004)
Log_BE_ME	-0.005*	-0.000
	(0.003)	(0.003)
N_Analyst	0.000	-0.000
	(0.000)	(0.000)
Inst_Own	-0.008**	-0.001
	(0.003)	(0.005)
Earn_Vol	0.026***	-0.005
	(0.009)	(0.014)
ROA	-0.036*	-0.056**
	(0.021)	(0.026)
Loss	0.008*	0.002
	(0.004)	(0.004)
Sentiment	-0.073***	-0.026*
	(0.010)	(0.014)
Fog	-0.001	0.002
	(0.002)	(0.002)

- Bloat is associated with the financial circumstances of a firm in intuitive ways, which helps to establish its validity.
- managers are more likely to release bloated disclosures when their firm performs worse

- Consequence of Bloat

$$Info\ Friction_{it} = \beta Bloat_{it} + \gamma \mathbf{X}_{it} + \delta_t + \theta + \varepsilon_{it}$$

probability of informed trading　　　　post-filing volatility

Panel A. MD&A Sample

Dep Var =	*PIN*		*Abn_Spread*		*Post_Vol*	
	(1)	(2)	(3)	(4)	(5)	(6)
Bloat	0.099***	0.062***	0.494***	0.544**	0.027***	0.037***
	(0.014)	(0.015)	(0.125)	(0.214)	(0.007)	(0.007)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	No	Yes	No	Yes	No	Yes
Industry FE	Yes	No	Yes	No	Yes	No
Avg. Dep.	0.009	0.009	0.043	0.043	0.018	0.018
Cluster	Industry	Industry	Industry	Industry	Industry	Industry
N	1,790	1,790	1,790	1,790	1,790	1,790
Adjusted R^2	0.146	0.452	0.076	0.310	0.476	0.569

Panel B. Conference Calls Sample

Dep Var =	*PIN*		*Abn_Spread*		*Post_Vol*	
	(1)	(2)	(3)	(4)	(5)	(6)
Bloat	0.034***	0.037***	0.172**	0.226***	0.008***	0.007***
	(0.007)	(0.005)	(0.068)	(0.052)	(0.003)	(0.001)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	No	Yes	No	Yes	No	Yes
Industry FE	Yes	No	Yes	No	Yes	No
Avg. Dep.	0.023	0.023	0.279	0.279	0.021	0.021
Cluster	Industry	Industry	Industry	Industry	Industry	Industry
N	8,537	8,537	8,537	8,537	8,537	8,537
Adjusted R^2	0.284	0.477	0.174	0.368	0.393	0.498

- Disclosure bloat hinders effective information transfer between companies and information users.

Conclusion

- The summary-based sentiment better explains stock market reactions to disclosed information than the original's sentiment
- They construct a novel and easy-to-implement measure of the degree of "bloat" in corporate disclosures
- Bloated disclosures are associated with higher price efficiency and higher information asymmetry, thus implying negative capital market consequences.

From transcripts to insights: Uncovering corporate risks using generative AI

Alex G. Kim, Maximilian Muhn, Valeri V. Nikolaev

Research question

- How generative AI tools (ChatGPT) help investors uncover political-, climate-, and AI-related corporate risk?

Why interesting

- Corporations face multifaceted risks that extend far beyond traditional financial metrics
  - Global political instability/ climate uncertainty/ rapid technological change…
  - → long-term growth and stakeholder value
- Generative language model:
  - Have the general knowledge go beyond the context of a given text
  - Synthesize the info into coherent, understandable narratives

- This paper fill the gap between generative AI and risk assessment by examining the potential of LLMs

Contribution

- Literature on LLM applications
  - Prior studies use LLM for other purposes e.g., forecast, measure complexity…
    - (see Bernard et al., 2023; Lopez-Lira and Tang, 2023; Jha et al., 2023; Eisfeldt et al., 2023; Kim et al., 2023; Chen et al., 2023)
  - This study shows that AI tools are effective at extracting risk categories
- Literature on measuring firm-level risk exposure using disclosures
  - Prior literature: topic-based bigram dictionaries…
  - This study adopt AI-based technology to analyze risks
- Literature on the value of general AI
  - Prior studies focus on the text
  - This study shows LLMs leverage their general knowledge to derive insights

Research design

- Data
  - Earning conference call transcripts
- Prompt
  - Summary: instruct GPT to focus solely on the document contents and avoid making judgments (minimum general knowledge)
  - Assessment: instructs GPT to generate an assessment of a given risk, which is not limited to the transcript. (include general knowledge)

$$RiskSum_{it} = \frac{\sum_{l=1}^{K_{it}} \text{len}(\mathbf{S}(c_{it}^{l}))}{\text{len}(c_{it})}$$

$$RiskAssess_{it} = \frac{\sum_{l=1}^{K_{it}} \text{len}(\mathbf{A}(c_{it}^{l}))}{\text{len}(c_{it})}$$

- Political risk
- Climate risk
- AI risk

An example

B1. Political Risk Summary

The company is subject to political and regulatory risks and uncertainties in Europe and North America. The recent government auction of HS1 in the U.K. is mentioned as an example.

B2. Political Risk Assessment

The firm is subject to political and regulatory risks and uncertainties in Europe and North America. The focus on deficit reduction in these regions may lead to an increased flow of government disposals and potentially PFI (Private Finance Initiative) opportunities. The recent government auction of HS1 in the U.K. is mentioned as an example. Additionally, the flow of non-core disposals by corporate and financial institutions is continuing, as evidenced by the firm's recent investment in Eversholt, which was purchased from HSBC. These political and regulatory factors could impact the firm's operations and investment opportunities in these regions.

Empirical results

- Capital market consequences – political risk as an example

$$Volatility_{it+1} = \beta Risk_{it} + \gamma \mathbf{X}_{it} + \delta_x + \varepsilon_{it},$$

Panel A. Industry and Time Fixed Effects					
Dep Var = Implied Volatility					
	(1)	(2)	(3)	(4)	(5)
PRiskSum	0.888***		0.857***		0.160
	(4.50)		(4.37)		(0.73)
PRiskAssess		0.777***		0.757***	0.676***
		(5.08)		(4.97)	(3.72)
PRiskBigram			0.015	0.014	0.014
			(1.54)	(1.43)	(1.40)
Controls	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Time & Ind	Time & Ind	Time & Ind	Time & Ind	Time & Ind
N	35003	35003	35003	35003	35003
Adjusted R^2	0.477	0.477	0.477	0.477	0.477
Dep Var = Abnormal Volatility					
	(1)	(2)	(3)	(4)	(5)
PRiskSum	1.576***		1.574***		-1.031***
	(6.55)		(6.54)		(-2.96)
PRiskAssess		1.987***		1.995***	2.523***
		(11.42)		(11.44)	(10.00)
PRiskBigram			0.001	-0.006	-0.004
			(0.07)	(-0.48)	(-0.34)
Controls	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Time & Ind	Time & Ind	Time & Ind	Time & Ind	Time & Ind
N	39276	39276	39276	39276	39276
Adjusted R^2	0.353	0.355	0.353	0.355	0.355

- Capital market consequences - Out-of-sample analysis
  - Not attributing to GPT seeing the underlying data during its training phase
  - 2022 - March 2023

Panel A. Political Risks				
	Implied Volatility		Abnormal Volatility	
	(1)	(2)	(3)	(4)
PRiskSum	1.716***		0.430	
	(5.04)		(1.09)	
PRiskAssess		1.246***		0.941***
		(5.12)		(3.44)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
N	9923	9923	9246	9246
Adjusted R^2	0.423	0.423	0.115	0.117

- Firm decisions – investment decisions
  - Higher risk exposure → harder to finance capital investments

$$Investment_{it} = \beta Risk_{it} + \gamma \mathbf{X}_{it} + \delta + \varepsilon$$

Panel A. Political Risk					
Dep Var			Investment		
	(1)	(2)	(3)	(4)	(5)
PRiskSum	-0.792**		-0.792**		0.317
	(-2.08)		(-2.06)		(0.58)
PRiskAssess		-0.918***		-0.922***	-1.081***
		(-3.33)		(-3.32)	(-2.72)
PRiskBigram			-0.000	0.003	0.002
			(-0.01)	(0.13)	(0.11)
Controls	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Ind FE	Yes	Yes	Yes	Yes	Yes
N	35615	35615	35615	35615	35615
Adjusted R^2	0.279	0.280	0.279	0.280	0.280

- Firm decisions – responses to mitigate risk exposures
  - Lobby/ green patents/ AI patents

Panel A. Political Risk					
Dep Var			**1** ($ Lobby Amount > 0)		
	(1)	(2)	(3)	(4)	(5)
PRiskSum	0.057		0.142		-1.258***
	(0.18)		(0.46)		(-4.19)
PRiskAssess		0.489**		0.437*	1.079***
		(2.04)		(1.84)	(4.21)
PRiskBigram			0.041**	0.037*	0.039*
			(2.04)	(1.83)	(1.94)
Controls	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Ind FE	Yes	Yes	Yes	Yes	Yes
N	39937	39937	39937	39937	39937
Adjusted R^2	0.191	0.191	0.191	0.191	0.192

- Relative importance of different types of risk over time

$$Implied\_Volatility_{it} = \beta_{1t} PRiskAssess_{it} + \beta_{2t} CRiskAssess_{it} + \beta_{3t} AIRiskAssess_{it} + \gamma_t \mathbf{X}_{it} + \delta_q + \delta_s + \varepsilon_{it}$$

Panel B. Regressions with Three Risk Measures Together

		PRiskAssess		*CRiskAssess*		*AIRiskAssess*	
		Coeff	t-stat	Coeff	t-stat	Coeff	t-stat
Start	End	(1)	(2)	(3)	(4)	(5)	(6)
2018Q1	2018Q4	0.448	2.04	-0.414	-1.24	0.328	0.82
2018Q2	2019Q1	0.395	1.80	-0.072	-0.21	0.290	0.68
2018Q3	2019Q2	0.525	2.49	-0.075	-0.22	-0.077	-0.18
2018Q4	2019Q3	0.385	1.73	0.292	0.82	-0.200	-0.48
2019Q1	2019Q4	0.446	1.99	0.233	0.65	-0.200	-0.47
2019Q2	2020Q1	0.640	2.63	1.572	4.11	-0.947	-2.00
2019Q3	2020Q2	0.597	2.22	2.331	5.78	-1.540	-3.22
2019Q4	2020Q3	0.430	1.54	2.357	5.89	-1.192	-2.31
2020Q1	2020Q4	0.343	1.18	2.111	5.24	-1.067	-2.08
2020Q2	2021Q1	0.569	2.02	1.350	3.40	-0.306	-0.64
2020Q3	2021Q2	0.577	2.20	0.634	1.62	0.410	0.96
2020Q4	2021Q3	0.686	2.77	0.115	0.33	0.233	0.58
2021Q1	2021Q4	0.672	2.69	0.022	0.06	0.294	0.75
2021Q2	2022Q1	0.415	1.73	0.165	0.50	0.607	1.74
2021Q3	2022Q2	0.500	1.99	0.632	1.88	0.908	2.51
2021Q4	2022Q3	0.633	2.36	1.327	3.73	0.923	2.49
2022Q1	2022Q4	0.719	2.57	1.502	4.08	1.047	2.76
2022Q2	2023Q1	0.864	2.94	1.669	3.95	0.713	2.99

- Equity market pricing

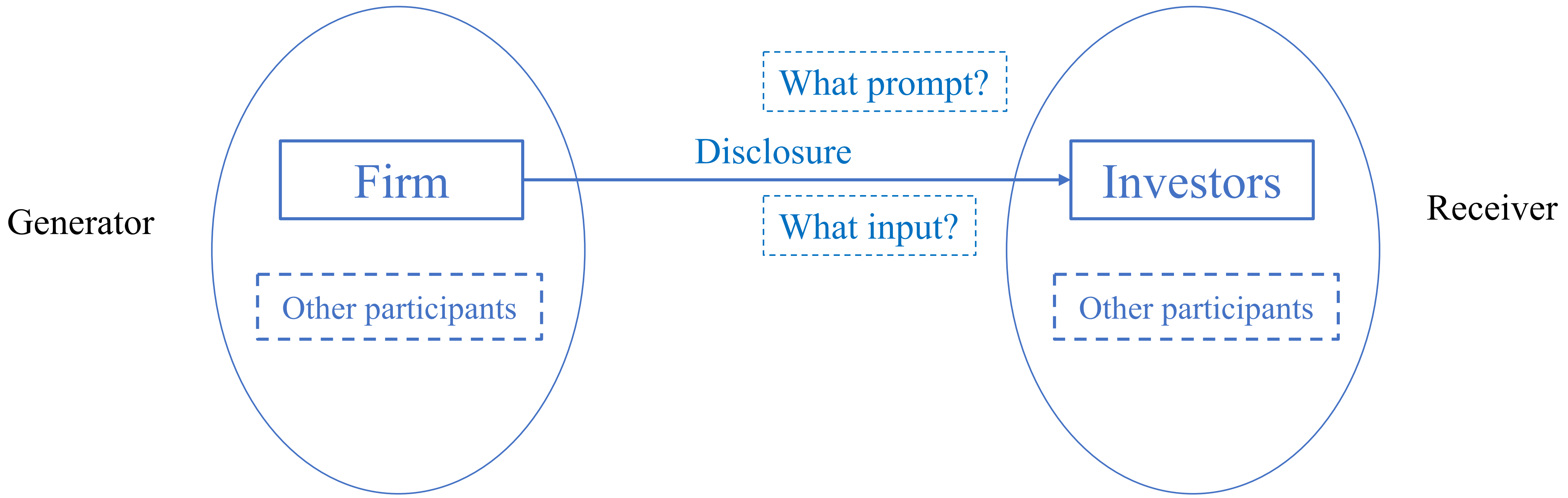
Panel A. Fama and MacBeth (1973) Regressions			
Dep Var		Return	
	(1)	(2)	(3)
$r_{0,1}$	-0.009	-0.001	-0.010
	(-0.67)	(-0.67)	(-0.72)
$r_{2,12}$	0.002	0.002	0.002
	(0.47)	(0.49)	(0.33)
$\log(ME)$	-0.000	-0.000	-0.000
	(-0.34)	(-0.33)	(-0.32)
$\log(BE/ME)$	-0.001	-0.001	-0.001
	(-0.54)	(-0.56)	(-0.44)
Profitability	0.001	0.001	0.001
	(0.54)	(0.60)	(0.62)
Investment	-0.006	-0.008	-0.005
	(-0.65)	(-0.80)	(-0.55)
$PRiskAssess^{ann}$	0.077		
	(1.36)		
$CRiskAssess^{ann}$		0.211*	
		(1.98)	
$AIRiskAssess^{ann}$			0.317*
			(1.72)
Adjusted R^2	0.054	0.054	0.055

Panel B. Single Sorts			
		Fama and French (2015) Five-Factor Alphas	
Sorts On =	$PRiskAssess^{ann}$	$CRiskAssess^{ann}$	$AIRiskAssess^{ann}$
	(1)	(2)	(3)
Low	-0.06	-0.14	-0.11
2	-0.05	-0.06	0.23
3	0.25	0.08	0.05
4	0.30	0.21	0.16
High	0.38	0.42	0.42
High – Low	0.44	0.56*	0.53**
t-value	(1.51)	(1.90)	(2.31)

Conclusion

- This paper evaluate whether recent advances in AI can help investors assess critical aspects of corporate risks.
- They use GPT 3.5 Turbo to develop and validate three proxies for firm-level exposure to political, climate, and AI-related risks.

Summary



What prompt?
Disclosure
Firm
Investors
Generator
Receiver
What input?
Other participants
Other participants