Seminar Presentation

2023 年 10 月 10 日

1 How to Talk When a Machine Is Listening: Corporate Disclosure in the Age of AI

1.1 Motivation

- More and more investors begin to use machine learning tools to process the information in the regular disclosure in the age of AI.
- Companies realize that the target audience of mandatory and voluntary disclosures no longer solely consists of human analysts and investors.
- There has not been an analysis exploring the **feedback effect**: how companies adjust the way they talk knowing that machines are listening.

1.2 Outline

- Finding an effective proxy for AI readership
  - AI reader base and machine-friendlier disclosure (**Machine downloads**)
  - The effect of Machine downloads on trading and information dissemination
    - Time to first trade and Time to first directional trade
- Managing Sentiment and Tone with Machine Readers
  - High machine downloads avoid LM-negative words but **only post-2011**
    - Researchers and practitioners have long relied on the Harvard Dictionary
    - LM (2011) a leading lexicon for algorithms to sort out sentiments
  - Managing other textual tones of LM (2011) with machine readers
    - Litigious, Uncertainty, Weak modal, and Strong modal
- Firms derive and incur heterogeneous **benefits and costs** from managing sentiment.
  - Benefits: External financing needs

  - Costs: Litigation risk

- **Out-of-Sample Tests**: Recent Technology and Audio Tone

  - Managing sentiment in response to recent technology(BERT)

  - Managing audio quality in conference calls with machine readers

1.3 Contribution

- This study builds on an expanding literature on information acquisition and dissemination via SEC-filing downloads (Bernard et al., 2020, Cao et al., 2021, Chen et al., 2020, Crane et al., 2023)

  - A central theme from the rapidly growing literature on textual analysis is that **qualitative information** from and the writing quality of disclosures predict asset returns and corporate performance

  - This study departs from the extant literature as we explore managerial disclosure strategies in **response** to the growing presence of AI analytical tools

- This paper connects to a distinct literature on the "feedback effect"

  - This is an increasing challenge on machine learning to be **"manipulation proof"** in that the algorithms learn to anticipate the strategic behavior of informed agents without observing it in training samples

1.4 Data, Variable Construction, and Sample Overview

1.4.1 Data source

- The primary data source of this study is the Securities and Exchange Commission's (SEC) Electronic Data Gathering, Analysis, and Retrieval **(EDGAR)** system and the associated Log File Data Set.

  - Log File tracks requests and downloads, which comprises all records of the requests of SEC filings from EDGAR from January 2003 to June 2017 (after which the SEC stopped updating the Log Files).
  - Each observation in the original data set contains information on the visitor's Internet Protocol (IP) address, timestamp, and the unique accession number of the filing that the visitor downloads.

- Regulatory filings have the advantage that their audience composition is mostly exogenous to firms' own decisions, which is less true in other settings.
  - For example, managers can invite a selected audience to corporate events, while regulatory filings are open to everyone (Cohen et al., 2020).

1.4.2 Construction of main variables

1. **Machine downloads**

- Matching: match the **accession number** with the SEC master filing index to select all the 10-K and 10-Q filings (MD&A)
- IP address Identification:
  - exclude requests that land on index pages because such requests do not download actual company filing
  - include the requests downloading more than **50 unique** firms' filings on any given date, and that are attributed to **web crawlers** (Lee et al., 2015)
- Aggregation: We aggregate machine requests and other requests for each filing **within 7 days**; the majority of requests occur during this period
- Machine downloads is the natural logarithm of the average number of machine downloads of firm i's filings during the four quarters prior to time t
- It is worth noting that the SEC log files became **available to the public in 2015** with retrospective information from 2003 and quarterly updates forward. (The list

of people/entities who requested the SEC Log files from the SEC website regarding FOIA logs is available at `https://www.sec.gov/foia/docs/foia-logs`.)

2. AI ownership

- We classify an investment company as such if it has AI-related job postings in the past 5 years according to data from Burning Glass (Abis and Veldkamp, 2023).
- AI ownership is the percentage of shares outstanding held by investment companies with AI capabilities from 2011 to 2019 since the BG data are available after 2010.

3. AI talent supply

- we retrieve the number of people between 18 and 64 with college or graduate school degrees in IT, scaled by the population at the state-year level, using data from Integrated Public Use Microdata Series (IPUMS) surveys, from 2010 to 2019.
- For each firm and during the quarter prior to the current filing, we aggregate AI talent supply over all states based on the headquarters of the investors, weighted by their ownership.

4. Machine readability

- We adopt multiple metrics developed in Allee et al. (2018) that we believe to best summarize the important attributes distinctly related to machine readability
  - **Table extraction**, the ease of separating tables from the text
  - **Number extraction**, the ease of extracting numbers from the text
  - **Table format**, the ease of identifying the information contained in the table (e.g., headings, column headings, row separators, and cell separators);
  - **Self-containedness,** whether a filing includes all needed information (i.e., without relying on external exhibits)
  - **Standard characters**, the proportion of characters that are standard ASCII (American Standard Code for Information Interchange) characters

- In our main specification, each attribute is standardized to a **Z-score** before being averaged to form a single-index Machine readability measure.

5. Firm characteristics

Size	Market capitalization in the natural logarithm
Tobin's q	Natural logarithm of the ratio of the sum of market value of equity and book value of debt to that of the book value
ROA	Ratio of EBITDA to assets
Leverage	Ratio of total debt to assets at book value
Growth	Average sales growth of the past 3 years
Industry adjusted return	Monthly average SIC three-digit industry-adjusted stock returns over the past year.
Institutional ownership	Ratio of the shares of institutional ownership to shares outstanding
Analyst coverage	Natural logarithm of 1+number of IBES analysts covering the stock
Idiosyncratic volatility	Annualized idiosyncratic volatility (daily data) from the FF3 model
Turnover	Monthly average of the ratio of trading volume to shares outstanding
Segment	Business segments, following Cohen and Lou (2012)

- All control variables are constructed **annually** using information available at the previous year-end. All unbounded variables are winsorized at the 1% extremes.

1.5 AI Readership and Machine Readability of Disclosures

Top machine downloaders

Rank	Name of institution	#MD	Type of institution
1	Renaissance Technologies	536,753	Quantitative hedge fund
2	Two Sigma Investments	515,255	Quantitative hedge fund
3	Barclays Capital	377,280	Financial conglomerate with asset management
4	JPMorgan Chase	154,475	Financial conglomerate with asset management
5	Point72 Asset Management	104,337	Quantitative hedge fund
6	Wells Fargo	94,261	Financial conglomerate with asset management
7	Morgan Stanley	91,522	Investment bank with asset management
8	Citadel LLC	82,375	Quantitative hedge fund
9	RBC Capital Markets	79,469	Financial conglomerate with asset management
10	D. E. Shaw Co.	67,838	Quantitative hedge fund
11	UBS AG	64,029	Financial conglomerate with asset management
12	Deutsche Bank AG	55,825	Investment bank with asset management
13	Union Bank of California	50,938	Full-service bank with private wealth management
14	Squarepoint Ops	48,678	Quantitative hedge fund
15	Jefferies Group	47,926	Investment bank with asset management
16	Stifel, Nicolaus Company	24,759	Investment bank with asset management
17	Piper Jaffray	18,604	Investment bank with asset management
18	Lazard	18,290	Investment bank with asset management
19	Oppenheimer Co.	15,203	Investment bank with asset management
20	Northern Trust Corporation	11,916	Financial conglomerate with asset management

- Quant funds are major players in integrating big data and unstructured data analyses in making investment decisions. The remaining institutions are mostly brokers and investment banks with significant asset management businesses.

1.5.1 Relation between Machine downloads and Machine readability

A first test is to relate Machine readability to Machine downloads in the cross-section and over time.

$$MachineReadability_{i,j,t} = \beta MachineDownloads_{i,j,t} + \delta OtherDownloads_{i,j,t}$$
$$+ \gamma Control_{i,year} + \alpha_i(\alpha_{SIC3}) + \alpha_{year} + \epsilon_{i,j,t}$$

A. Machine readability

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	*Machine readability*				*MR upgrade*	
Machine downloads	0.076***	0.075***	0.060***	0.078***		
	(13.89)	(17.45)	(10.33)	(15.93)		
ΔMachine downloads					0.005***	0.006***
					(2.90)	(3.40)
Other downloads	0.005	0.002	−0.007	−0.006	0.000	−0.001
	(1.15)	(0.47)	(−1.44)	(−1.33)	(0.20)	(−0.44)

B. Components of Machine readability

	(1)	(2)	(3)	(4)	(5)
			Machine readability		
Dependent variable	*Table extraction*	*Number extraction*	*Table format*	*Self-containedness*	*Standard characters*
Machine downloads	0.051***	0.028***	0.026***	0.161***	0.125***
	(6.02)	(3.47)	(2.88)	(21.80)	(14.68)
Other downloads	0.018**	−0.011	0.022**	−0.036***	−0.040***
	(2.37)	(−1.49)	(2.51)	(−6.69)	(−6.08)

C. Alternative machine-readership measures

	(1)	(2)	(3)	(4)
Dependent variable	*Machine readability*			
AI ownership	0.553***	0.400***		
	(8.53)	(9.56)		
AI talent supply			0.240***	0.349***
			(14.85)	(21.01)

- A larger presence of machine readers in the market will lead firms to increase the readability of their disclosure with respect to machines.

1.5.2 The effect on trading and information dissemination

- When disclosures are read more by machines, and when filings are made more machine readable, trades motivated by the information in the disclosures should materialize faster, and the **speed of information dissemination** should be faster
  - Time to the first trade is the length of time, in seconds, between the EDGAR posting and the first subsequent trade of the issuer's stock
  - Time to the first directional trade adds a requirement that the trade needs to be profitable based on the price at the end of the 15th minute post-filing

$$\begin{aligned} TimetoTrade_{i,j,t} &= \beta_1 MachineDownloads_{i,j,t} \times MachineReadability_{i,j,t} \\ &+ \beta_2 MachineDownloads_{i,j,t} + \beta_3 MachineReadability_{i,j,t} \\ &+ \delta OtherDownloads_{i,j,t} + \gamma Control_{i,year} + \alpha_i(\alpha_{SIC3}) + \alpha_{year} + \epsilon_{i,j,t} \end{aligned}$$

A. Time to the first trade

	(1)	(2)	(3)	(4)
Dependent variable	*Time to first trade*		*Time to first directional trade*	
Machine downloads	−4.857*	−3.398	−7.540***	−7.258**
	(−1.68)	(−1.14)	(−2.71)	(−2.55)
Machine downloads ×		−3.887***		−2.127*
Machine readability		(−2.84)		(−1.67)
Machine readability		−5.980		−8.709
		(−0.92)		(−1.46)
Other downloads	3.499	1.304	3.885*	2.336
	(1.42)	(0.51)	(1.72)	(1.00)
Observations	161,664	144,193	161,664	144,193
R-squared	.269	.272	.285	.286
Control variables	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

B. Effects of machine readership: Bid-ask spread

	(1)	(2)	(3)	(4)
Dependent variable	*Bid-ask spread*		*Bid-ask spread*	
Groups	Entire sample		Low turnover	High turnover
Machine downloads × After	0.055***	0.081***	0.080***	0.089***
	(8.46)	(10.91)	(7.18)	(8.97)
Machine readability × After		0.023	0.010	0.030
		(1.15)	(0.33)	(1.10)
Observations	2,673,992	2,416,151	1,203,653	1,212,498
R-squared	.720	.732	.738	.715
Firm FE	Subsumed	Subsumed	Subsumed	Subsumed
Filing FE	Yes	Yes	Yes	Yes
Minute FE	Yes	Yes	Yes	Yes

- Providing machine traders with a more accurate signal **increases the information asymmetry** between the machine traders and the market maker
  - BAS is the latest pair of lowest ask price and highest bid price within each minute scaled by the midpoint of the bid price and ask price
  - After equals to one if minute m occurs after the filing is posted

$$BidAskSpread_{i,j,m} = \beta MachineDownloads_{i,j,t} \times After_{i,j,m} + \gamma MachineReadability_{i,j} \times After_{i,j,m} + \alpha_{i,j} + \alpha_m + \epsilon_{i,j,m}$$

- The machine-equipped investors are able to update their judgments about a firm's fundamentals more efficiently than others, which worsens information asymmetry.

1.6 Managing Sentiment and Tone with Machine Readers

- If AI readers shape the style and quality of corporate writing, we expect the difference-in-differences coefficient β_1 to be significantly negative for LM sentiment, but not for Harvard sentiment.
- Sorted by the reduction in the average frequency per filing, the ten most avoided words are: "restructuring," "termination," "restatement," "declined," "correction," "misstatement," "terminated," "late," "alleged," and "omitted."

$$NegativeSentiment_{i,j,t} = \beta_1 MachineDownloads_{i,j,t} \times Post_t + \beta_2 MachineDownloads_{i,j,t} + \delta OtherDownloads_{i,j,t} + \gamma Control_{i,year} + \alpha_i(\alpha_{SIC3}) + \alpha_{year} + \epsilon_{i,j,t}$$

Machine downloads and sentiment: Loughran and McDonald (2011) publication

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	*LM – Harvard sentiment*		*LM sentiment*		*Harvard sentiment*	
Machine downloads	−0.072***	−0.079***	−0.062***	−0.050***	0.010	0.029***
× Post	(−6.95)	(−8.94)	(−4.98)	(−4.99)	(0.76)	(2.65)
Machine downloads	−0.007	−0.011**	−0.009	−0.019***	−0.002	−0.008
	(−1.17)	(−2.46)	(−1.18)	(−3.72)	(−0.23)	(−1.43)
Observations	158,578	158,515	158,578	158,515	158,578	158,515
R-squared	.217	.568	.241	.632	.208	.590
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	No	Yes	No	Yes	No	Yes
Industry FE	Yes	No	Yes	No	Yes	No
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

- Managers use negative words from Harvard in place of synonyms from the LM list

Machine downloads and other tones: Loughran and McDonald (2011) publication

Dependent variable	(1) *Litigious*	(2) *Uncertainty*	(3) *Weak modal*	(4) *Strong modal*
Machine downloads × Post	−0.057***	−0.021***	−0.034***	−0.007***
	(−6.02)	(−3.49)	(−8.86)	(−4.39)
Machine downloads	0.007	−0.009***	−0.021***	−0.004***
	(1.44)	(−3.05)	(−10.05)	(−4.98)
Observations	158,515	158,515	158,515	158,515
R-squared	.509	.600	.624	.571
Control variables	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

- The frequency of all four tones was met with negative stock market reactions, we conjecture that managers of firms with high expected machine readership should moderate these words after 2011

1.7 Equilibrium and cross-sectional effects

- The first test explores motives underlying positive disclosures by sorting firms by upcoming **external financing needs** (net total issuance)

- The second test builds on the premise that firms under tighter regulatory scrutiny or higher **litigation risk** are constrained in mincing words (Bertomeu et al., 2021)

Machine readability and sentiment: Cross-sectional effects in terms of costs and benefits

Dependent variable	(1) *Machine readability*	(2)	(3) *LM – Harvard sentiment*	(4)	(5) *LM – Harvard sentiment*	(6)
	External financing needs				Litigation risk	
Groups	Top quartile	Other	Top quartile	Other	Top quartile	Other
Machine downloads × Post			−0.103***	−0.075***	−0.054***	−0.090***
			(−6.62)	(−7.55)	(−3.46)	(−8.81)
Machine downloads	0.107***	0.076***	−0.024***	−0.011**	−0.018**	−0.012**
	(10.28)	(13.37)	(−2.89)	(−1.96)	(−2.54)	(−2.16)
Difference of coefficients	0.031***		−0.028*		0.036**	
p-value	.004		.065		.027	
Observations	35,014	101,242	36,984	106,468	48,457	102,467
R-squared	.439	.365	.635	.572	.598	.591
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

1.8 Out-of-Sample Tests: Recent Technology and Audio Tone

1.8.1 Managing sentiment in response to recent technology (BERT)

- EDGAR Log File Data Set stopped in 2017 and BERT was published in 2018, our Machine downloads variable is not available for this test

$$BERTSentiment_{i,t} = \beta AIReadership_{i,t} \times PostBERT_t$$
$$+ \delta AIReadership_{i,t} + \gamma Control_{i,t} + \alpha_i + \alpha_t + \epsilon_{i,j,t}$$

Managing sentiment in response to recent technology (BERT)

	(1)	(2)	(3)	(4)
Dependent variable	*BERT sentiment*			
	NegSent/TotalSent		NegSent/TotalWords	
AI ownership × Post-BERT	−4.953**		−0.212***	
	(−2.49)		(−2.68)	
AI ownership	2.313		0.103	
	(1.26)		(1.39)	
AI talent supply × Post-BERT		−0.983***		−0.041***
		(−3.61)		(−3.98)
AI talent supply		−0.522		−0.010
		(−1.18)		(−0.65)
Observations	6,627	6,627	6,627	6,627
R-squared	.796	.796	.804	.804
Control variables	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

1.8.2 Managing audio quality in conference calls with machine readers

- If managers are aware that disclosure documents could be parsed by machines, they should have realized that their machine readers may be also using voice analyzers to extract signals from vocal patterns and emotions contained in managers' speeches

- We build a web crawler (Selenium-Python) to obtain audios of conference calls from 2010 to 2016 from EarningsCast and use **pyAudioAnalysis** to process the data
  - Emotion valence: an emotion is positive or negative
  - Emotion arousal: intensity or strength of the associated emotion state
  - Both measures are bounded between –1 and 1.

$$Emotion_{i,k,t} = \beta MachineDownloads_{i,k,t} + \delta OtherDowloads_{i,k,t}$$
$$+ \gamma Control_{i,year} + \alpha_i(\alpha_{SIC3}) + \alpha_{year} + \epsilon_{i,k,t}$$

Machine Downloads and Managers' Emotion during Conference Calls

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	*Emotion valence*			*Emotion arousal*		
Machine downloads	0.043***	0.042***	0.042***	0.004*	0.005**	0.007**
	(11.40)	(11.14)	(8.84)	(1.79)	(2.28)	(2.49)
Other downloads	−0.017***	−0.017***	−0.012***	−0.006***	−0.006***	−0.006***
	(−5.74)	(−5.67)	(−3.12)	(−3.65)	(−3.71)	(−2.92)
Observations	43,336	41,224	27,437	43,336	41,224	27,437
R-squared	0.389	0.383	0.388	0.395	0.395	0.469
Control Variables	No	Yes	Yes	No	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

1.9 Conclusion

- This paper presents the first study showing how corporate disclosure in writing and speaking has been reshaped by machine readership
- Increasing AI readership motivates firms to prepare filings that are friendlier to machine parsing and processing
- Firms manage sentiment and tone perception that caters to AI readers by differentially avoiding words perceived as negative by algorithms. CEOs also aim to present with vocal qualities that are favorably rated by software

2 Copycat Skills and Disclosure Costs: Evidence from Peer Companies' Digital Footprints

2.1 Motivation

- Despite the capital market benefits of greater transparency, **public disclosures** may reveal proprietary information that ultimately harms the disclosing firm and works to the advantage of its competitors.
- Competing firms can view and imitate a disclosing firm's strategies known as "copycatting." The copycatting behavior of peer companies remains relatively unexplored.
- **Investment companies disclose** trading strategies that can be easily quantified and linked to the subsequent actions of copycats and prior studies have documented that trades disclosed in 13F filings contain **proprietary information**
- We examine **whether and when** copycats profit from imitation, and under what conditions copycats cause competitive harm to disclosing companies.

2.2 Outline

- Identifying Copycatting Behavior
  - EDGAR Log File data, which contain information on retrievals (referred to interchangeably as "viewing activities" or "downloads") of SEC filings
  - The trading decision of disclosing peers and subsequent activity of copycats
- The Screening Skills of Copycats
  - The subsequent performance of copycatted trades
  - Return source: sophistication and research intensity of copycats
- The Cost to Disclosing Companies
  - Quantifying proprietary costs, especially opportunity costs

  - Proprietary costs and screening skills
  - Proprietary costs and position horizon
  - Proprietary costs and information asymmetry

2.3 Contribution

- Using the digital footprints of companies, our paper is **the first** to directly examine the process through which copycats obtain information and selectively imitate their peers' strategies.
- Contrary to popular belief, **copycatting is not naïve imitation** but instead involves effort and a well-honed capability for information acquisition and processing
- **Identifying this mechanism** allows us to understand the skills of copycats, and how such skills affect the costs inflicted on disclosing companies.

2.4 Data

2.4.1 Identifying viewing companies

Panel A: Sample selection

Viewers:	Filter	# Companies
	Investment companies covered by the Thomson Reuters 13F database	6,912
Subtract:	Companies whose IP is not registered with ARIN	(4,518)
Subtract:	Companies whose IP registry does not contain an entire block of 256 IP addresses	(1,621)
Subtract:	Matched companies that did not view any EDGAR filings	(123)
Subtract:	Companies that did not view any 13F filings (but viewed other types of filings)	(70)
Subtract:	Companies without data on first-buy and last-sell trades	(17)
	Final sample	563

Filers:	Filter	# Companies
	Investment companies covered by Thomson Reuters 13F database	6,912
Subtract:	Companies whose Thomson identifier (*mgrno*) cannot be matched to a Central Index Key (*cik*)	(1,977)
Subtract:	Companies whose 13F filings were not viewed by any peer	(107)
Subtract:	Companies without data on first-buy and last-sell trades	(265)
	Final sample	4,563

Panel B: Descriptive statistics for investment companies

Mean	Thomson	Viewers	Viewers – Thomson	Robot Viewers	Robot Viewers –Viewers	Filers	Filers – Thomson
AUM	1,679	6,233	4,554***	10,350	4,117***	1,788	109
Age	6.140	11.776	5.636***	12.790	1.010	7.835	1.695***
Return	0.798	0.852	0.054	0.950	0.090	0.790	−0.008
Alpha	0.071	0.011	−0.060	0.097	0.086	−0.026	−0.097

2.5 Data source

- EDGAR Log File data
- Data on investment company portfolio holdings from Thomson Reuters
- Stock market data from CRSP
- Hedge fund returns from a union hedge fund data set

2.6 Identifying Copycatting Behavior

- If a copycat gleans useful information from a peer's filings, the trading decisions of the disclosing peer should inform the subsequent trading decisions of the copycat.
- We define $ViewingActivity_{i,j,t}$ as an indicator variable equal to 1 if a viewer i views a filer j' s 13F filing in a given quarter t, and 0 otherwise
- $DisclosedTrade_{j,t}$ is defined as an indicator variable equal to 1 if a filer j has one or more disclosed trading decisions at the individual stock level that are either first buys or last sells in quarter t,and 0 otherwise
- $ViewerTrade_{i,j,t+1}$ is an indicator variable equal to 1 if viewer i in quarter t + 1 has at least one first buy or last sell at the stock level identical to filer j in quarter t, and 0 otherwise.

$$ViewerTrade_{i,j,t+1} = \beta ViewingActivity_{i,j,t} \times DisclosedTrade_{j,t} + \delta ViewingActivity_{i,j,t} \\ + \gamma DisclosedTrade_{j,t} + \alpha_i + \alpha_j + \alpha_{i,j} + \mu_t + \epsilon_{i,j,t+1}$$

- The coefficients imply that for an average viewer, viewing activity increases the likelihood of imitating a peer's trading decisions by 40.5% (0.135/0.333)

	Viewer Trade$_{i,j,t+1}$			
Dependent Variable	(1)	(2)	(3)	(4)
Viewing Activity × *Disclosed Trade*	0.135***	0.120***	0.089***	0.025**
	(4.31)	(4.98)	(4.16)	(2.30)
Disclosed Trade	0.333***	0.311***	0.108***	0.112***
	(29.23)	(29.08)	(19.77)	(20.61)
Viewing Activity	−0.000	−0.070***	−0.062***	−0.016*
	(−0.01)	(−3.82)	(−3.36)	(−1.80)
Time FE	No	Yes	Yes	Yes
Viewer FE	No	Yes	Yes	–
Filer FE	No	No	Yes	–
Viewer × Filer FE	No	No	No	Yes
N	53,460,288	53,460,288	53,460,288	53,460,288
Adjusted R^2	0.032	0.204	0.345	0.407

2.7 The Screening Skills of Copycats

2.7.1 Subsequent performance of copycatted trades

- We classify the trading activities of two investment companies that trade sequentially into categories: copycatted trades, unfollowed trades, and coincidental trades.
  - A Copycatted trade is a trade made by a disclosing company that is replicated in the next quarter by a company that views its filing.
  - An Unfollowed trade is a trade made by a disclosing company that is not followed by any viewing company.
  - A Coincidental trade is a trade made by both a disclosing company (in the current quarter) and another company (in the following quarter) that does not view the disclosing company's 13F filings.
- We form two value-weighted hedge portfolios that takes a long position in first-buy stocks and a short position in last-sell stocks

- Copycats develop skills to screen for more profitable trades, the returns on the two portfolios are both positive and significant

	Performance of Disclosed Trades				
	(1) *Copycatted*	(2) *Unfollowed*	(3) *Coincidental*	(4) *Copycatted − Unfollowed*	(5) *Copycatted − Coincidental*
Excess Return	0.345***	−0.106**	−0.141***	0.451***	0.486***
	(4.50)	(−2.31)	(−3.01)	(4.17)	(4.35)
CAPM Alpha	0.343***	−0.116**	−0.153***	0.459***	0.496***
	(4.38)	(−2.48)	(−3.21)	(4.16)	(4.35)
FF3 Alpha	0.343***	−0.115**	−0.150***	0.458***	0.493***
	(4.35)	(−2.47)	(−3.17)	(4.13)	(4.31)
FFC4 Alpha	0.345***	−0.116**	−0.150***	0.460***	0.495***
	(4.35)	(−2.48)	(−3.16)	(4.14)	(4.30)
# Months	171	171	171	171	171

2.7.2 Sophistication and research intensity of copycats

- We use the average return of the first-buy and last-sell stocks that it has copycatted over the past two years to measure the sophistication of the viewing company
  - We sort viewing companies into halves based on this measure in each quarter and label the group above (below) the median as "sophisticated" ("naïve").
- Intensity of fundamental research by the viewing company, proxied by its quarterly viewing activities of related SEC filings, including 10-K and 10-Q.
  - We sort viewing companies into halves based on research intensity in each quarter and label the group above the median as "high research intensity."

Panel A: Copycat sophistication

	Performance of Disclosed Trades					
	(1)	(2)	(3)	(4)	(5)	(6)
	Sophisticated		Naïve		Sophisticated – Naïve	
	Copycatted − Unfollowed	*Copycatted − Coincidental*	*Copycatted − Unfollowed*	*Copycatted − Coincidental*	*Copycatted − Unfollowed*	*Copycatted − Coincidental*
Excess Return	1.055***	0.813***	−0.044	0.041	1.099***	0.772***
	(8.43)	(6.39)	(−0.36)	(0.31)	(6.39)	(3.98)
CAPM Alpha	1.146***	0.871***	−0.049	0.010	1.195***	0.861***
	(9.33)	(6.80)	(−0.39)	(0.08)	(6.96)	(4.42)
FF3 Alpha	1.145***	0.870***	−0.055	0.005	1.200***	0.865***
	(9.26)	(6.76)	(−0.45)	(0.04)	(7.03)	(4.46)
FFC4 Alpha	1.133***	0.862***	−0.054	−0.002	1.187***	0.863***
	(9.28)	(6.72)	(−0.44)	(−0.01)	(7.00)	(4.43)
# Months	168	168	168	168	168	168

- Overall, we find strong evidence that copycats possess the ability to discern profitable trades from voluminous peer disclosures. This finding contradicts the popular belief that copycats merely follow all trades naïvely.

Panel B: Copycat research intensity

	Performance of Disclosed Trades					
	(1)	(2)	(3)	(4)	(5)	(6)
	High Research Intensity		Low Research Intensity		High Research Intensity – Low Research Intensity	
	Copycatted − Unfollowed	*Copycatted − Coincidental*	*Copycatted − Unfollowed*	*Copycatted − Coincidental*	*Copycatted − Unfollowed*	*Copycatted − Coincidental*
Excess Return	0.483***	0.449***	0.198	0.214	0.285	0.235
	(4.28)	(3.82)	(1.13)	(1.21)	(1.46)	(1.21)
CAPM Alpha	0.524***	0.476***	0.136	0.129	0.387**	0.347*
	(4.59)	(3.98)	(0.77)	(0.73)	(1.99)	(1.79)
FF3 Alpha	0.528***	0.480***	0.136	0.127	0.392**	0.353*
	(4.61)	(4.00)	(0.78)	(0.72)	(2.04)	(1.82)
FFC4 Alpha	0.533***	0.485***	0.124	0.117	0.409**	0.368*
	(4.66)	(4.04)	(0.72)	(0.66)	(2.15)	(1.92)
# Months	171	171	171	171	171	171

2.8 The Cost to Disclosing Companies

- To measure copycats' impact on disclosing companies, we use **realized return** at the fund company level to capture proprietary costs
  - Suppose a fund purchases $1 million of stock A and plans to purchase another $1 million of the same stock later on, and the fund can obtain 10%
  - However, copycatting may result in higher prices for stock A, forcing the disclosing fund to buy a inferior stock B with 5% return
  - Thus, there is an opportunity cost of 2.5% resulting from the copycats. The opportunity cost can be captured by **company-level portfolio returns**

	Performance of Disclosing Company$_{j,t}$			
Dependent Variable	(1)	(2)	(3)	(4)
Copycatted by Peer$_{j,t}$	−0.216***	−0.217***		
	(−4.55)	(−4.56)		
Coincidental by Peer$_{j,t}$		0.050		
		(0.37)		
# Copycats$_{j,t}$			−0.085***	−0.079***
			(−4.62)	(−3.69)
# Coincidental Peers$_{j,t}$				0.000
				(0.72)
Disclosing Company FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
N	48,969	48,969	48,969	48,969
Adjusted R^2	0.063	0.063	0.063	0.063

- Copycats with greater screening skills impose higher proprietary costs on the disclosing company

Proprietary Costs to the Disclosing Company: Copycat Ability

	Performance of Disclosing Company$_{j,t}$	
Dependent Variable	(1)	(2)
Copycatted by Sophisticated Peer$_{j,t}$	−0.393***	
	(−7.72)	
Copycatted by Naïve Peer$_{j,t}$	0.117**	
	(2.07)	
Copycatted by High-Research-Intensity Peer$_{j,t}$		−0.235***
		(−4.78)
Copycatted by Low-Research-Intensity Peer$_{j,t}$		−0.076
		(−1.04)
Difference of Coefficients	−0.480***	−0.159*
p-value	<0.001	0.072
Disclosing Company FE	Yes	Yes
Time FE	Yes	Yes
N	47,597	48,963
Adjusted R^2	0.063	0.063

- If a disclosing company plans to accumulate its positions over several quarters, copycats who follow its initial trades may jeopardize the disclosing company's prospects for successfully accumulating positions at advantageous price levels.
  - We divide all first-buy positions within a given quarter into two groups: positions completed within one quarter (the first group) and positions that increase for at least one additional quarter (the second group).

- Performance-depressing effect of copycatting is more pronounced when the disclosing company's stock portfolio is characterized by greater information asymmetry.
  - We measure the information asymmetry of a stock by the quarterly average of the daily bid-ask spread

Proprietary Costs to the Disclosing Company: Characteristics of Disclosed Holdings

	Performance of Disclosing Company$_{j,t}$			
Dependent Variable	Position-building Horizon		Information Asymmetry	
	(1)	(2)	(3)	(4)
Subgroup:	Long	Short	High	Low
Copycatted by Peer$_{j,t}$	−0.293***	−0.147**	−0.278***	−0.129**
	(−3.87)	(−2.33)	(−3.51)	(−2.20)
Difference of Coefficients	−0.146*		−0.149*	
p-value	0.069		0.065	
Disclosing Company FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
N	23,198	23,104	23,511	23,402
Adjusted R^2	0.076	0.061	0.071	0.060

2.9 Additional Tests

2.9.1 Information content of disclosure: requests for confidential treatment

- **?** find that hedge funds may hide their private information by requesting confidential treatment for their 13F filings (obtain data on confidentiality requests)
- As such, the original 13F filing contains less valuable information for potential copycats. Therefore, we test this prediction using the subsample of observations in which the disclosing company is a hedge fund company.

Information Content of Disclosure: Requests for Confidential Treatment

Panel A: Viewing activities and subsequent trades					
Dependent Variable	*Viewer Trade$_{i,j,t+1}$*				
	(1)		(2)		(3)
Subgroup	Filings without Requests		Filings with Requests		Filings without Requests by Confidential Filers
Viewing Activity$_{i,j,t}$ × *Disclosed Trade$_{j,t}$*	0.177***		−0.021		0.137***
	(5.82)		(−0.35)		(4.44)
Disclosed Trade$_{j,t}$	0.398***		0.503***		0.434***
	(31.64)		(22.39)		(33.42)
Viewing Activity$_{i,j,t}$	−0.141***		−0.149***		−0.172***
	(−6.00)		(−4.66)		(−6.82)
Difference of Coefficients		(1) − (2)		(3) − (2)	
(Interaction Term)		0.198***		0.158***	
p-value		0.002		0.010	
Viewer FE	Yes		Yes		Yes
Time FE	Yes		Yes		Yes
N	4,898,873		144,570		792,147
Adjusted R^2	0.209		0.236		0.215

2.9.2 The timeliness of viewing activities

- We examine whether copycatting companies can reap greater benefits from and impose greater damages on disclosing companies when they view filings earlier.
- We divide the sample into two subsamples based on whether the human viewing activities occur within the first 12 hours after filings become publicly available ("early viewers") or after the first 12 hours ("late viewers").

2.9.3 The validity of copycatting activities

- Our main analysis uses quarterly holdings data, which do not allow us to pinpoint the intra-quarter dates of copycatted trades.
- To mitigate these issues, we match our sample with data from ANCerno (previously Abel Noser), which contain daily transactions of select investment companies

Viewing Activities and Subsequent Trades: ANCerno Sample

	Viewer Trade$_{i,j,t+1}$	
Dependent Variable	(1)	(2)
Viewing Activity$_{i,j,t}$ × *Disclosed Trade*$_{j,t}$	0.282**	0.320***
	(2.78)	(3.40)
Disclosed Trade$_{j,t}$	0.317***	0.303***
	(8.48)	(9.89)
Viewing Activity$_{i,j,t}$	−0.000	−0.321***
	(−0.00)	(−3.38)
Viewer FE	No	Yes
Time FE	No	Yes
N	298,661	298,661
Adjusted R^2	0.017	0.290

2.9.4 Robot viewing activities

- In our main analyses, we have excluded viewing activities generated by robots and focused on human viewing activities.

- In this section, we examine whether the findings based on human viewing activities carry over to robot-generated viewing activities.

Robot Viewing Activities

Panel A: Viewing activities and subsequent trades

	Viewer Trade$_{i,j,t+1}$	
Dependent Variable	(1)	(2)
Viewing Activity$_{i,j,t}$ × *Disclosed Trade*$_{j,t}$	0.189**	0.190**
	(2.60)	(2.59)
Disclosed Trade$_{j,t}$	0.331***	0.309***
	(29.03)	(28.89)
Viewing Activity$_{i,j,t}$	−0.000	−0.175**
	(−0.01)	(−2.58)
Viewer FE	No	Yes
Time FE	No	Yes
N	53,460,288	53,460,288
Adjusted R^2	0.033	0.204

Panel B: The performance of copycatted trades

	Performance of Disclosed Trades				
	(1) *Copycatted*	(2) *Unfollowed*	(3) *Coincidental*	(4) *Copycatted − Unfollowed*	(5) *Copycatted − Coincidental*
Excess Return	0.414**	−0.068	−0.062*	0.482**	0.476**
	(2.26)	(−1.06)	(−1.71)	(2.14)	(2.40)
CAPM Alpha	0.438**	−0.086	−0.077**	0.523**	0.514**
	(2.34)	(−1.30)	(−2.10)	(2.28)	(2.55)
FF3 Alpha	0.439**	−0.087	−0.078**	0.526**	0.517**
	(2.35)	(−1.32)	(−2.13)	(2.30)	(2.57)
FFC4 Alpha	0.426**	−0.087	−0.079**	0.512**	0.505**
	(2.29)	(−1.31)	(−2.16)	(2.24)	(2.51)
# Months	174	174	174	174	174

参考文献

S. Abis and L. Veldkamp. The Changing Economics of Knowledge Production. *Review of Financial Studies*, 00(0), 2023.

K. D. Allee, M. D. Deangelis, and J. R. Moon. Disclosure "Scriptability". *Journal of Accounting Research*, 56(2):363–430, 2018.

D. Bernard, T. Blackburne, and J. Thornock. Information flows among rivals and corporate investment. *Journal of Financial Economics*, 136(3):760–779, 2020.

J. Bertomeu, E. Cheynel, E. Floyd, and W. Pan. Using machine learning to detect misstatements. *Review of Accounting Studies*, 26(2):468–519, 2021.

S. S. Cao, K. Du, B. Yang, and A. L. Zhang. Copycat Skills and Disclosure Costs: Evidence from Peer Companies' Digital Footprints. *Journal of Accounting Research*, 59(4):1261–1302, 2021.

H. Chen, L. Cohen, U. Gurun, D. Lou, and C. Malloy. IQ from IP: Simplifying search in portfolio choice. *Journal of Financial Economics*, 138(1):118–137, 2020.

L. Cohen and D. Lou. Complicated firms. *Journal of Financial Economics*, 104(2): 383–400, May 2012.

L. Cohen, D. Lou, and C. J. Malloy. Casting Conference Calls. *Management Science*, 66 (11):5015–5039, 2020.

A. Crane, K. Crotty, and T. Umar. Hedge Funds and Public Information Acquisition. *Management Science*, 69(6):3241–3262, 2023.

C. M. Lee, P. Ma, and C. C. Wang. Search-based peer firms: Aggregating investor perceptions through internet co-searches. *Journal of Financial Economics*, 116(2): 410–431, 2015.