HOW TO TALK WHEN A MACHINE IS LISTENING: CORPORATE DISCLOSURE IN THE AGE OF AI

Sean Cao , Wei Jiang , Baozhong Yang , Alan L. Zhang

Working Paper 27950

王健

2020-03-11

content

- Introduction
  - Background & Motivation
  - Question
  - Research content
  - Related researches
  - Contribution
- Data
- Empirical results
- Conclusion

1. Introduction

Background & Motivation

- "Be fearful when others are greedy and greedy when others are fearful"; "When it's raining gold, reach for a bucket, not a thimble" -Warren Buffet
- More and more companies realize that the age of AI is coming.
- While the literature is guiding investors to applying computational tools to extract information from disclosure, there has no been an analysis exploring the feedback effect.

1. Introduction

Question: how companies adjust the way they talk knowing that machines are listening?

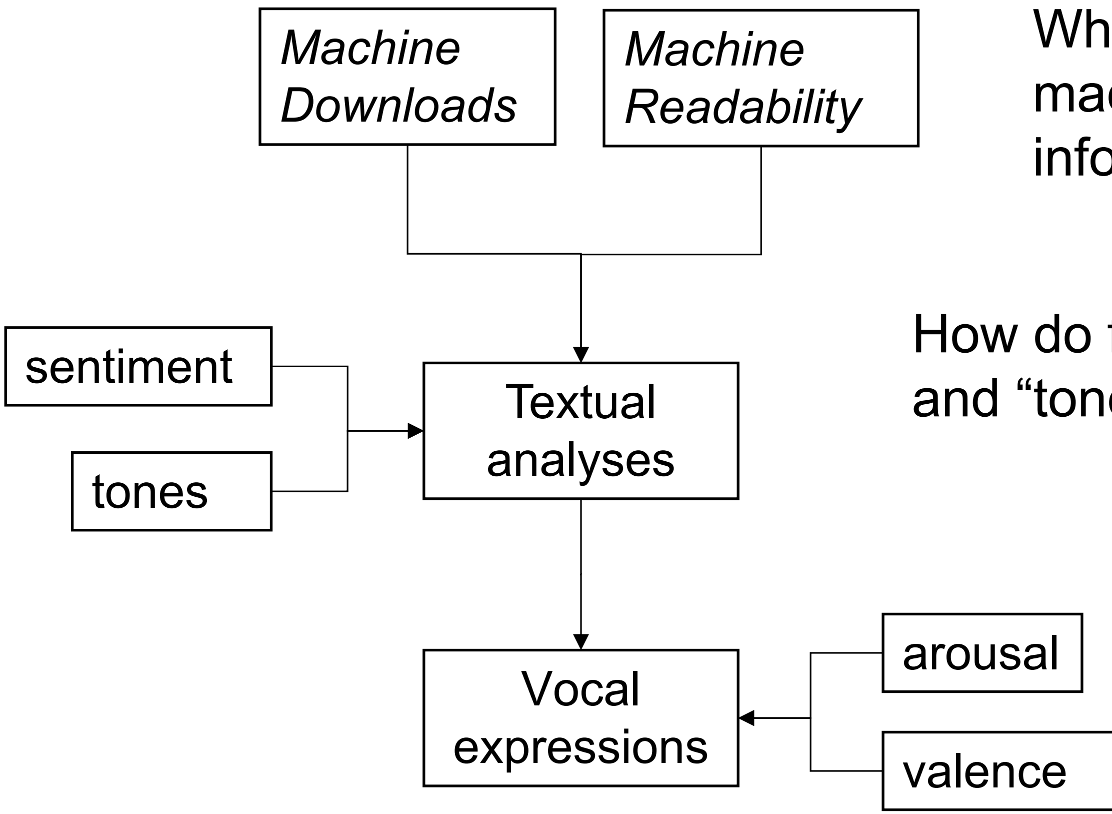
- What is the real impact of machine-process on information dissemination?
- How do firms manage “sentiment” and “tones” perceived by machines?
- How do managers improve their vocal performances along the quantifiable metrics?

1. Introduction

Research contents

- We start with a diagnostic test that connects the expected extent of AI readership and how machine friendly the company composes its disclosure.
- We use event study and regression to show how firms' sentiment expression changes to cater to their AI readers
- Apart from the textual information, we also consider how other application of computational technology affects the managers' vocal performance.

1. Introduction



What is the real impact of machine-process on information dissemination?

How do firms manage "sentiment" and "tones" perceived by machines?

How do managers improve their vocal performances along the quantifiable metrics?

1.Introduction

Related researches

- Crane, Crotty, and Umar (2020) : find that hedge funds that employ robotic downloads perform better than those that do not.
- Allee, DeAngelis and Moon (2018): studied the proposed metrics of information processing costs.
- Loughran and McDonald (2011): create dictionaries of positive and negative words that are specific to the context of financial documents.
- Mayew and Ventakachalam (2012); Hu and Ma (2020): use commercial Layered Voice Analysis (LVA) software to study the vocal expressions of managers and their implications on capital markets.

1.Introduction

Contribution

- Our study builds on an expanding literature on information acquisition and dissemination via SEC filings downloads, opting in a new angle on the consequences of machine downloads and potentially machine processing.
- Our paper also contributes to the rapidly growing literature on modern machine learning techniques with a central theme that qualitative information from text or non-text data predicts asset returns and corporate performance.
- Our study departs from the existent literature as we explore managerial disclosure strategies in response to the growing presence of AI analytical tools in both the industry and academia.

2.Data

A. Data sources

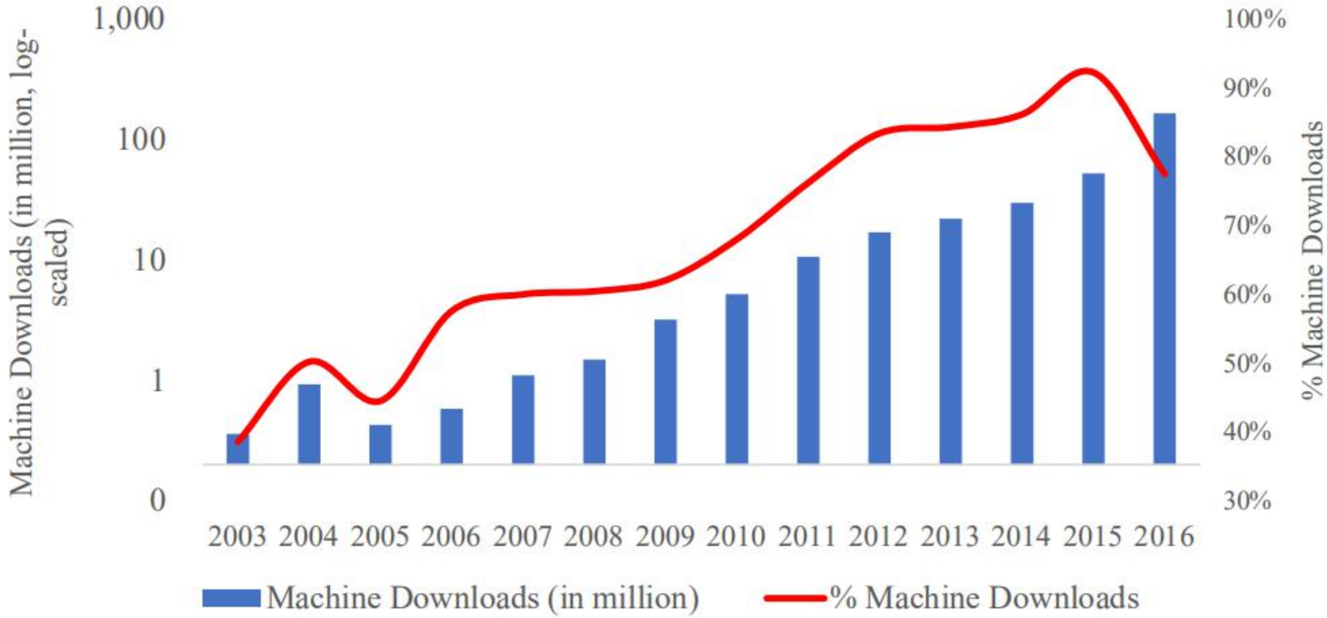
- Text data and download information: Securities and Exchange Commission's (SEC) Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system and the associated Log File Data Set.

- Calls :EarningsCast, a commercial aggregator for company earnings calls.
- Firms character: CRSP/Compustata and Thomson Reuters Ownership Database

2.Data

B. Construction of main variables

1) Machine Downloads: IP address downloading more than 50 unique firms‘ filings on any given date as a machine visitor （machine downloads, other downloads）



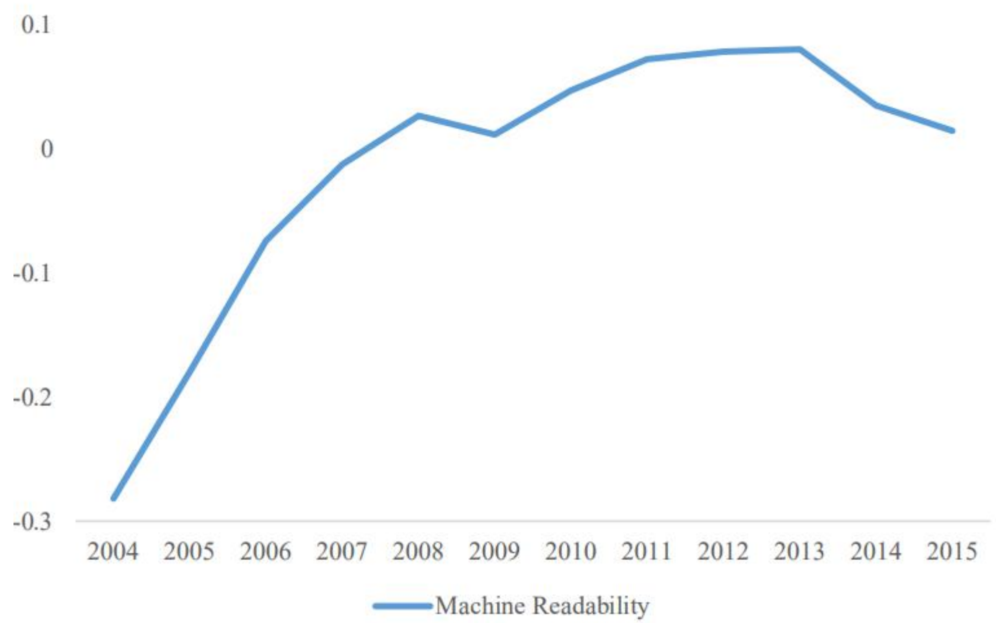
2.Data

B. Construction of main variables

2) Machine Readability:

We summarize the most important attributes: (i)Table Extraction, (ii) Number Extraction, (iii) Table Format, (iv) Self-Containedness, (v) Standard Characters.

Each attribute is standardized to a Z-score before being averaged to form a single index *Machine Readability*



2.Data

B. Construction of main variables

3) (Negative) sentiment:

The first lexicon is the Harvard General Inquirer IV-4 psychological dictionary, which is termed *Harvard Sentiment.*

The second lexicon is developed by Loughran and McDonald (2011),which is named *LM Sentiment.*

Finally, we form the difference, LM – Harvard Sentiment, to capture the contrast

4) Additional sentiment measures:

It builds on Loughran and McDonald (2011)'s list of measures for broader sentiment, including *litigiousness*, *uncertainty*, *weak modal* and *strong modal* words, all in financial contexts.

2.Data

B. Construction of main variables

5) Vocal emotions:

It is based on the Circumplex model (Russell, 1980), which suggests that emotions are distributed in a two-dimensional space defined by *valence* and *arousal*.

Emotion Valence described the extent to which an emotion is positive or negative, with a larger value indicating greater positivity.

Emotion Arousal refers to the intensity or the strength of the associated emotion state and a greater (lower) value suggests that the speaker is more excited (calmer).

6) Firm characteristics: size, Tobin's Q and *ROA* and so on.

3. Empirical results

I. AI Readership and Machine Readability of Corporate Disclosure

1) Determinants of machine downloads

- We first try to understand what factors drive *machine downloads* variation.

Dependent Variable	(1)	(2) Machine Downloads	(3)	(4)	(5) % Machine Downloads	(6)	(7) Machine Downloads	(8)
Size	0.135***	0.139***	0.040***	-0.028***	-0.029***	-0.008***	0.139***	0.040***
	(40.29)	(45.62)	(7.05)	(-26.71)	(-28.91)	(-6.01)	(45.62)	(7.05)
Tobin's Q	-0.048***	-0.066***	-0.022***	0.010***	0.011***	0.002	-0.066***	-0.022***
	(-9.41)	(-13.24)	(-3.38)	(6.51)	(8.09)	(1.25)	(-13.24)	(-3.38)
ROA	-0.011	-0.031***	-0.002	-0.013***	0.012***	0.013***	-0.031***	-0.002
	(-0.94)	(-2.68)	(-0.14)	(-3.78)	(3.15)	(3.46)	(-2.68)	(-0.14)
Leverage	0.085***	0.122***	0.055***	-0.019***	-0.020***	-0.010**	0.122***	0.055***
	(6.58)	(9.39)	(3.37)	(-5.25)	(-5.75)	(-2.48)	(9.39)	(3.37)
Growth	-0.078***	-0.068***	-0.024***	-0.004**	-0.008***	-0.011***	-0.068***	-0.024***
	(-13.69)	(-12.21)	(-3.63)	(-2.50)	(-5.31)	(-6.76)	(-12.21)	(-3.63)
IndAdjRet	-0.847***	-0.729***	-0.322***	0.217***	0.188***	0.084***	-0.729***	-0.322***
	(-15.75)	(-13.97)	(-6.00)	(15.44)	(14.22)	(7.31)	(-13.97)	(-6.00)
InstOwnership	-0.005	-0.024*	-0.026	0.038***	0.045***	0.028***	-0.024*	-0.026
	(-0.32)	(-1.66)	(-1.24)	(8.47)	(11.19)	(6.48)	(-1.66)	(-1.24)
Log(#analyst)	-0.008	-0.008	-0.021***	-0.004**	-0.005***	0.000	-0.008	-0.021***
	(-1.52)	(-1.54)	(-2.92)	(-2.46)	(-3.38)	(0.07)	(-1.54)	(-2.92)
IdioVol	0.091***	0.060***	-0.062***	-0.080***	-0.073***	-0.028***	0.060***	-0.062***
	(6.07)	(4.32)	(-4.37)	(-17.94)	(-18.71)	(-8.68)	(4.32)	(-4.37)
Turnover	0.022***	0.019***	0.022***	-0.006***	-0.005***	-0.006***	0.019***	0.022***
	(13.20)	(12.08)	(12.11)	(-12.11)	(-10.90)	(-14.55)	(12.08)	(12.11)
Segment	0.007***	0.007***	0.007***	-0.000	-0.001***	-0.001***	0.007***	0.007***
	(6.81)	(6.97)	(3.89)	(-0.55)	(-3.06)	(-2.94)	(6.97)	(3.89)
AIHedgeFund							0.728***	0.417**
							(4.52)	(2.54)

This suggests that machines tend to download filings from more mature firms with more firm-specific developments and firms attract more machine downloads when it exhibits lower than usual idiosyncratic volatility and have lower analyst coverage.

3. Empirical results

2) Relation between machine downloads and machine readability of reports

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable			*Machine Readability*			
Machine Downloads	0.076***	0.075***	0.060***	0.078***		
	(13.89)	(17.45)	(10.33)	(15.93)		
Other Downloads	0.005	0.002	-0.007	-0.006		
	(1.15)	(0.47)	(-1.44)	(-1.33)		
% Machine Downloads					0.121***	0.173***
					(3.91)	(6.39)
Total Downloads					0.053***	0.074***
					(10.27)	(16.26)

machine downloads, whether measured as the volume or percentage of machine downloads, significantly (at the 1% level) predicts more machine-reading friendly reports across all specifications, while other firm characteristics have little confounding effect.

3. Empirical results

2) Relation between machine downloads and machine readability of reports

- We use alternative specifications to show the results

	(1)	(2)	(3)	(4)
Dependent Variable	*PCA Machine Readability*		*Machine Readability*	
Machine Downloads	0.131***	0.162***		
	(11.18)	(16.14)		
Other Downloads	-0.047***	-0.046***		
	(-4.75)	(-5.88)		
Machine Downloads (Alt. def.)			0.052***	0.064***
			(9.51)	(13.72)
Other Downloads (Alt. def.)			-0.010	-0.000
			(-1.51)	(-0.05)

All four specifications show that *Machine Downloads* is significantly (at the 1% level) associated with, but *Other Downloads* exhibits no positive relation with *Machine Readability.*

3. Empirical results

2) Relation between machine downloads and machine readability of reports

- We use alternative specifications to show the results

	(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)

Again, the coefficients of Other Downloads do not have consistent signs across the five attributes.

3. Empirical results

3) Cross validation of Machine Downloads and Machine Readability as empirical proxies.

- First, we connect Machine Downloads to its primary suspect, hedge funds who adopt AI strategies.
- Second, we conjecture and test that machine readers should lead to faster trades after a filing is posted.
- Moreover, such an advantage should be elevated when the files are composed to be machine friendly

$$Time\ to\ Trade_{i,j,t} = \beta_1 Machine\ Downloads_{i,j,t} \times Machine\ Readability_{i,j,t} +$$

$$\beta_2 Machine\ Downloads_{i,j,t} + \beta_3 Machine\ Readability_{i,j,t} +$$

$$\delta Other\ Downloads_{i,j,t} + \boldsymbol{\gamma}\ Control_{i,year} + \alpha_i(\alpha_{SIC3}) + \alpha_{year} + \varepsilon_{i,t}.$$

Dependent Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Time to the First Trade				*Time to the First Directional Trade*		
Machine Downloads	-8.353**	-4.857*	-7.347**	-3.398	-12.365***	-7.540***	-12.374***	-7.258**
	(-2.56)	(-1.68)	(-2.19)	(-1.14)	(-3.94)	(-2.71)	(-3.87)	(-2.55)
Machine Downloads × *Machine Readability*			-3.761**	-3.887***			-2.815*	-2.127*
			(-2.46)	(-2.84)			(-1.87)	(-1.67)
Machine Readability			-6.540	-5.980			-5.695	-8.709
			(-0.99)	(-0.92)			(-0.91)	(-1.46)
Other Downloads	15.342***	3.499	15.151***	1.304	13.961***	3.885*	13.436***	2.336
	(5.29)	(1.42)	(5.06)	(0.51)	(4.95)	(1.72)	(4.67)	(1.00)

The results support the prediction that high *Machine Downloads* are associated with faster trades after a filing becomes publicly available.

Moreover, the relation between Machine Downloads and Time to Trade is indeed significantly stronger when Machine Readability is higher.

Answer question:

To the extent that faster market reaction to corporate disclosure is a sign of effective communication of firms' financial health, which results in efficient information dissemination

3. Empirical results

II. Managing Sentiment and Tones with Machine Readers

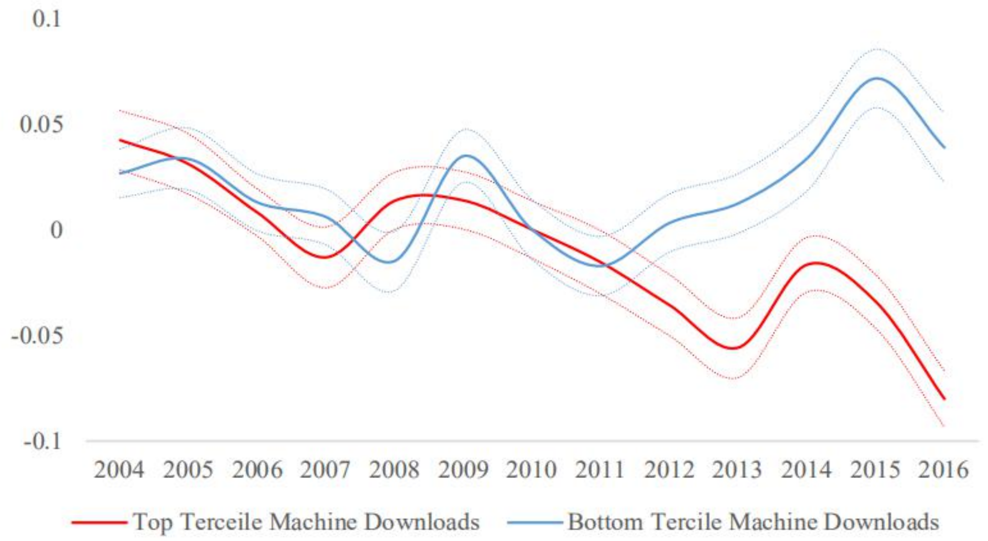
1) Textual sentiment

$$Negative\ Sentiment_{i,j,t} = \beta_1 Machine\ Downloads_{i,j,t} \times Post_t + \beta_2 Machine\ Downloads_{i,j,t} + \delta Other\ Downloads_{i,j,t} + \boldsymbol{\gamma}\ Control_{i,year} + \alpha_i(\alpha_{SIC3}) + \alpha_{year} + \varepsilon_{i,t}.$$

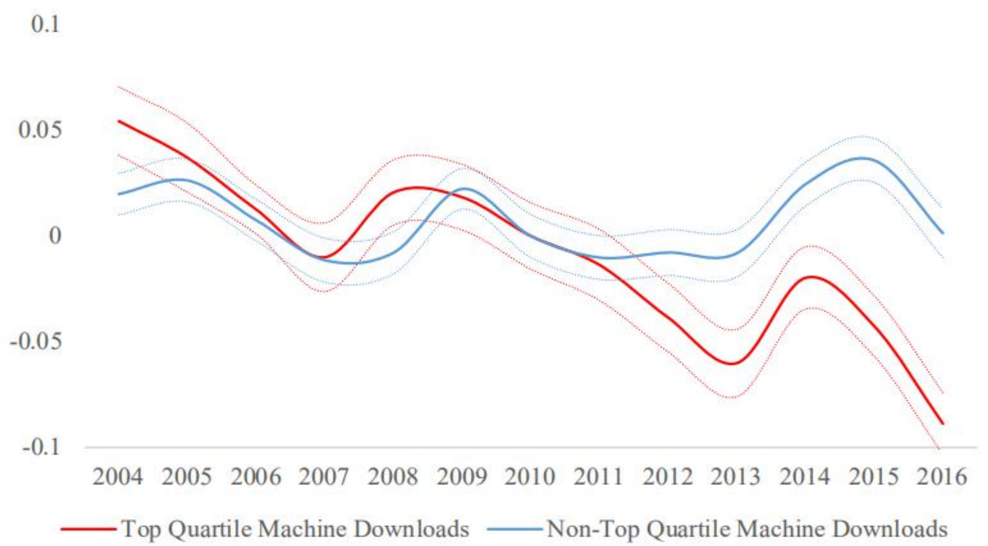
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	*LM – Harvard Sentiment*		*LM Sentiment*		*Harvard Sentiment*	
Machine Downloads × Post	-0.072***	-0.079***	-0.062***	-0.050***	0.010	0.029***
	(-6.95)	(-8.94)	(-4.98)	(-4.99)	(0.76)	(2.65)
Machine Downloads	-0.007	-0.011**	-0.009	-0.019***	-0.002	-0.008
	(-1.17)	(-2.46)	(-1.18)	(-3.72)	(-0.23)	(-1.43)

Above table and figure thus support the hypothesis that corporate writing has been adjusted to serve machine readers, which was impacted by the availability of the LM dictionary.

Panel A: Top tercile machine downloads vs. bottom tercile machine downloads



Panel B: Top quartile machine downloads vs. the rest



3. Empirical results

2) Managing other textual tones with machine readers

$$Tone_{i,j,t} = \beta_1 Machine\ Downloads_{i,j,t} \times Post_t + \beta_2 Machine\ Downloads_{i,j,t} +$$

$$\delta Other\ Downloads_{i,j,t} + \boldsymbol{\gamma}\ Control_{i,year} + \alpha_i (or\ \alpha_{SIC3}) + \alpha_{year} + \varepsilon_{i,t}.$$

Dependent Variable	(1) Litigious	(2)	(3) Uncertainty	(4)	(5) Weak Modal	(6)	(7) Strong Modal	(8)
Machine Downloads × *Post*	-0.056***	-0.057***	-0.016**	-0.021***	-0.028***	-0.034***	-0.008***	-0.007***
	(-5.38)	(-6.02)	(-2.01)	(-3.49)	(-4.85)	(-8.86)	(-4.39)	(-4.39)
Machine Downloads	0.011*	0.007	-0.006	-0.009***	-0.018***	-0.021***	-0.003**	-0.004***
	(1.71)	(1.44)	(-1.33)	(-3.05)	(-5.39)	(-10.05)	(-2.19)	(-4.98)

3. Empirical results

3) Managing audio quality in conference calls with machine readers

$$Emotion_{i,k,t} = \beta Machine\ Downloads_{i,k,t} + \delta Other\ Downloads_{i,k,t} +$$

$$\gamma\ Control_{i,year} + \alpha_i(\alpha_{SIC3}) + \alpha_{year} + \varepsilon_{i,k,t}.$$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	*Emotion-Valence*				*Emotion-Arousal*			
Machine Downloads	0.043***	0.035***	0.042***	0.042***	0.004*	0.003	0.005**	0.007**
	(11.40)	(8.13)	(11.14)	(8.84)	(1.79)	(0.94)	(2.28)	(2.49)
Other Downloads	-0.017***	-0.014***	-0.017***	-0.012***	-0.006***	0.000	-0.006***	-0.006***
	(-5.74)	(-4.32)	(-5.67)	(-3.12)	(-3.65)	(0.19)	(-3.71)	(-2.92)

Our findings support the hypothesis that managers may manipulate their vocal expressions to achieve a more favorable effect on investors that rely on machine processing

4. Conclusion

- Our findings indicate that increasing AI readership motivates firms to prepare filings that are more friendly to machine parsing and processing.
- Firms manage sentiment and tone perception that is catered to AI readers by differentially avoiding words that are perceived as negative by algorithms.
- Our study is the first to identify and analyze the feedback effect