Dissecting Corporate Culture Using Generative AI*

Kai Li, Feng Mai, Rui Shen, Chelsea Yang, and Tengfei Zhang

First version: August 2023 This version: January 2025

Abstract

This paper conducts the first large-scale study of how analysts and corporate insiders differ in their assessment of corporate culture and quantifies the economic implications of these differences. We employ generative AI to analyze analyst reports, earnings call transcripts, and employee reviews, and organize extracted information into a knowledge graph that links a culture type to its perceived causes and effects. We document systematic differences between analysts' perspectives on culture and those of executives and employees. We further show that analysts' culture analyses are incorporated into stock recommendations and target prices, and that investors react to a report's coverage of culture. Our findings suggest that analysts' perspectives on corporate culture contain value-relevant information not captured by insider views.

Keywords: corporate culture; equity analysts; analyst reports; earnings calls; employee reviews; large language models; generative AI; ChatGPT; cause-effect knowledge graph

JEL classifications: C45; C55; G32; G34; Z1

---

* Kai Li: Sauder School of Business, University of British Columbia. Email: kai.li@sauder.ubc.ca. Feng Mai: Tippie College of Business, University of Iowa. Email: feng-mai@uiowa.edu. Rui Shen: Shenzhen Finance Institute, School of Management and Economics, The Chinese University of Hong Kong, Shenzhen. Email: ruishen@cuhk.edu.cn. Chelsea Yang: Sauder School of Business, University of British Columbia. Email: chunxue.yang@sauder.ubc.ca. Tengfei Zhang: Rutgers School of Business-Camden, Rutgers University. Email: tengfei.zhang@rutgers.edu. We thank helpful comments from Hoa Briscoe-Tran, Francesco D'Acunto (discussant), Jin-Chuan Duan, Roby Lehavy (discussant), Alejandro Lopez-Lira (discussant), Johan Sulaeman, Baozhong Yang, Bernard Yeung, Frank Yu, seminar participants at CEIBS, Chinese University of Hong Kong, Chinese University of Hong Kong - Shenzhen, Fanhai International School of Finance, Fudan University, Monash University, Nanyang Technological University, National University of Singapore, NYU – Shanghai, NYU Stern Corporate Governance Research Luncheon, Renmin University, Rutgers University, Shanghai Jiaotong University, Shanghai Advanced Institute of Finance, Singapore Management University, Southern University of Science and Technology, Stanford University (Finance and Accounting), Stevens Institute of Technology, Tsinghua University, UC Davis, and conference participants at the First New Finance Conference (Toronto), the New Idea Session at the Financial Management Association Meetings (Chicago), the HKUST Accounting Research Symposium (Hong Kong), the Culture Connect Conference (Berkeley), the Finance Down Under Conference (Melbourne), the Weinberg/ECGI Corporate Governance Symposium, and the VSB Mid-Atlantic Research Conference (MARC) in Finance. We acknowledge financial support from the Social Sciences and Humanities Research Council of Canada (Grant Number: 435-2022-0285). Li acknowledges financial support from the Canada Research Chair in Corporate Governance. Zhang acknowledges data support from the Whitcomb Center for Research in Financial Services at Rutgers University. All errors are our own.

Dissecting Corporate Culture Using Generative AI

Abstract

This paper conducts the first large-scale study of how analysts and corporate insiders differ in their assessment of corporate culture and quantifies the economic implications of these differences. We employ generative AI to analyze analyst reports, earnings call transcripts, and employee reviews, and organize extracted information into a knowledge graph that links a culture type to its perceived causes and effects. We document systematic differences between analysts' perspectives on culture and those of executives and employees. We further show that analysts' culture analyses are incorporated into stock recommendations and target prices, and that investors react to a report's coverage of culture. Our findings suggest that analysts' perspectives on corporate culture contain value-relevant information not captured by insider views.

Keywords: corporate culture; equity analysts; analyst reports; earnings calls; employee reviews; large language models; generative AI; ChatGPT; cause-effect knowledge graph

JEL classifications: C45; C55; G32; G34; Z1

1. Introduction

Since "the 'cultural revolution' in finance" began a decade ago (Zingales 2015), researchers have gained a deeper understanding of corporate culture gleaned from values presented on corporate websites, and through surveys and interviews with corporate executives, earnings conference calls, employee reviews, and job postings (Guiso, Sapienza, and Zingales 2015; Grennan 2019; Li, Mai, Shen, and Yan 2021; Graham, Grennan, Harvey, and Rajgopal 2022a, 2022b; Huang, Pacelli, Shi, and Zou 2024; Li, Chen, and Shen 2024); and using proxies for corporate culture (see, for example, Ahern, Daminelli, and Fracassi 2015; Liu 2016; also see surveys by Gorton, Grennan, and Zentefis 2022; Grennan and Li 2023). This body of work as a whole, however, leaves a number of important gaps unexplored; in particular, given that corporate insiders may have skewed perspectives on culture because of their vested interests, how capital market participants (e.g., equity analysts) assess culture, and how different stakeholders perceive and analyze the mechanisms through which culture emerges and shapes business outcomes. Our study is among the first in finance, accounting, and economics to apply generative artificial intelligence (AI) models as reasoning agents on different data sets of analyst reports, earnings call transcripts, and employee reviews to address these gaps.

Sell-side equity analysts play an important role in processing, producing, and disseminating information about publicly listed companies to capital market participants (see, for example, Womack 1996; Brav and Lehavy 2003; Asquith, Mikhail, and Au 2005; Derrien and Kecskés 2013; Huang, Lehavy, Zang, and Zheng 2018; Birru, Gokkaya, Liu, and Stulz 2022). These analysts gain in-depth knowledge of the firms they follow through both formal and informal channels: reading financial statements, attending earnings conference calls, conducting site visits, and engaging directly with top and divisional managers (Soltes 2014; Brown, Call, Clement, and Sharp 2015). A number of prior papers show that analysts possess value-relevant non-financial information about the companies they cover, for example, management quality and innovation (Previts, Bricker, Robinson, and Young 1994; Huang, Zang, and Zheng 2014; Brown et al. 2015; Bellstam, Bhagat, and Cookson 2021). It is thus natural for us to investigate whether and how analysts conduct analyses on corporate culture, given their information intermediary role. Our research addresses the following questions from the vantage points of 2.4 million analyst reports, 243 thousand earnings calls, and 5.3 million employee reviews: 1) What cultural values are important to different stakeholders in a modern corporation? 2) What events, people, and/or systems shape corporate culture? 3) How does corporate culture matter according to different stakeholders? 4) What is the relationship between corporate culture and price formation?

Answering these questions using three very different large textual data sets presents significant challenges. Using the report corpus as an example, an analyst report contains quantitative analyses of recent and estimated firm performance (e.g., earnings per share) and qualitative interpretations of information signals (e.g., management quality) (Asquith, Mikhail, and Au 2005; Soltes 2014). With an average report comprising approximately 60 sentences over eight pages (Huang, Zang, and Zheng 2014), manually coding reports in search of answers to our research questions is infeasible. To overcome this challenge, we introduce a novel framework that applies generative artificial intelligence (AI) models, such as OpenAI's ChatGPT, as reasoning agents to automatically extract and analyze cause-effect relations pertaining to corporate culture from the reports (Blanco, Castell, and Moldovan 2008; Radinsky, Davidovich, and Markovitch 2012; Heindorf et al. 2020) and the tones of the discussion. These relations reflect the reasoning processes analysts undertake to dissect corporate culture, connecting different culture types (e.g., innovation and adaptability) to their perceived causes and effects (e.g., management turnover and customer satisfaction).

To illustrate, in the report featuring Cisco Systems Inc. by Walter Piecyk from the brokerage firm Painewebber Inc., released on April 14, 2000, the analyst says, *"…Cisco promotes a highly entrepreneurial culture within its organization, which has enabled the company to lure and hire about 3,000 new employees per quarter to its 26,100-employee base.*" Our method extracts a cause-effect relation in a triple as *('innovation and adaptability culture,' 'enables,' 'luring and hiring about 3,000 new employees per quarter'*) with a positive tone. To assess the accuracy of our approach, we compare the output of three generative AI models (ChatGPT, Claude, and Gemini) against human annotations in a sample of 200 culture-related segments. We find that all three models exhibit high levels of accuracy, precision, recall, and F1. In particular, ChatGPT achieves accuracy rates of 98.5%, 84.0%, 91.0%, and 97.5% for culture types, causes, effects, and tones, respectively, and outperforms other models. Given the task's complexity, these performance metrics are indicative of generative AI models' reasoning capabilities. We subsequently use the output from ChatGPT to answer the research questions.

To provide a bird's-eye view of corporate culture from analysts' perspective, we first normalize extracted relations into six culture types, eighteen causes, and seventeen business outcomes, then aggregate them into a cause-effect knowledge graph. The graph shows that, according to equity analysts, the top two culture types prevailing in a modern corporation are innovation and adaptability and collaboration and people-focused cultures; the top two drivers of cultural changes are business strategy and management team; the top two business outcomes that culture shapes are market share and growth and profitability. Our method further allows us to pinpoint the events, people, and/or systems that shape specific culture types, and identify which culture types have the most impact in driving specific business outcomes. For example, analysts identify that management team (people) and employee hiring and retention (a system) are the top two factors shaping the collaboration and people-focused culture, and that the collaboration and people-focused culture in turn bolsters market share and growth and drives employee satisfaction.

How do analysts' perspectives on corporate culture compare with insiders'? We adapt the same generative AI method to the call and review data sets and have the following observations. Both analysts and management identify innovation and adaptability, collaboration and people-focused, and performance-oriented as the most important to them. In contrast, employees view collaboration and people-focused culture as the most important to them. Moreover, both analysts and management believe business strategy is the most important driver for cultural changes, whereas employees view the management team as primarily responsible for cultural changes. In terms of culture's effects, both analysts and management believe market share and growth as the number one outcome of firms with a strong culture, whereas employees view their satisfaction as the most important outcome. These systematic differences align with the distinct roles and incentives of each stakeholder group. They also provide evidence of our approach's effectiveness in distinguishing between insiders' and outsiders' perspectives on corporate culture.

To gain a better understanding of why analysts are interested in corporate culture, we employ regression analysis relating firm and analyst characteristics to the likelihood of analysts' featuring culture discussions in their reports and the specific culture type discussed. We find that firm size, sales growth, profitability, and major events such as top management turnover and deal-making, firms whose executives discuss culture in calls, firms that host/participate in investor/broker conferences are positively and significantly associated with, whereas firm age, leverage, tangibility, earnings volatility, ownership by large shareholders, and board independence are negatively and significantly associated with, analysts' discussing culture in their reports. In terms of specific culture types discussed, we see some interesting variations. For example, tangibility is positively (negatively) and significantly associated with analysts' discussing customer-oriented culture (innovation and adaptability culture), and earnings volatility is only negatively and significantly associated with analysts' discussing collaboration and people-focused culture. Our findings on the negative influences of large shareholders and the positive influences of key corporate events (such as management turnover and M&As) are largely consistent with prior literature (e.g., Guiso, Sapienza, and Zingales 2015; Li et al. 2021). Our findings on the positive association between management discussing culture in calls (the number of meetings firms host/participate in investor/broker conferences) and analysts' featuring culture in their reports, suggesting that analysts do pay attention to value-relevant intangibles such as corporate culture in their research.

In terms of analyst characteristics, we show that analysts who are highly ranked, CFA charter holders, have masters' degrees, are women, more experienced, and make frequent forecasts are more likely to discuss culture in their reports. The associations between certain analyst characteristics – star status, credential, gender, experience, and effort – and coverage of culture help assuage concerns about analysts' indifference to or lack of insights into corporate culture.

Our cumulative evidence above suggests that equity analysts do have unique and significant insights into corporate culture from outside a corporation compared to insights gleaned largely from stakeholders within a corporation.

We next explore the determinants of any divergence in insiders' and outsiders' perspectives on culture. Given our multi-dimensional output generated from our information extraction methods – culture types, causes, effects, and tones, we employ two sets of measures to capture the degree of divergence in perspectives. Our first set is based on the Jensen-Shannon (JS) divergence measure (Lin 1991) between analyst reports and earnings calls (Glassdoor reviews), which quantifies the dissimilarity between probability distributions and always has a finite value between 0 and 1. A divergence of 0 means the distributions are identical, while 1 indicates the maximum difference. This set of measures does not have a sign. Our second set is the signed measures based on the difference between analysts' and management's (employees') perspectives.

Using the JS measures as summary measures of divergence in perspectives, we show that firm size is negatively and significantly associated with any divergence in insiders' and outsiders' perspectives on culture constructs, whereas ROA volatility is positively and significantly associated with divergent views. The first result suggests that there is less difference in insiders' and outsiders' perspectives on culture when firms are large. Interestingly, we note that CEO risk-taking incentives, as captured by their equity portfolio's vega (Guay 1999), are negatively and significantly associated with any divergence between analysts' and management's perspectives on culture.

Using the signed measures of divergence in perspectives on specific culture types and tones, we show that firm size, ROA, CEO equity incentives, and the number of broker/investor conferences participated/hosted are positively and significantly, whereas board independence is negatively and significantly, associated with the gap in analysts' focus on corporate culture compared to that of management. In contrast, we show that sales growth and leverage are positively and significantly, whereas firm size, ROA volatility, large institutional ownership, and the number of broker/investor conferences participated/hosted are negatively and significantly, associated with the gap in analysts' focus on corporate culture compared to that of employees. Taken as a whole, our results suggest that in large firms or firms participating/hosting more broker/investor conferences, analysts pay more attention to culture than management, whereas in small firms or firms participating/hosting more broker/investor conferences, employees pay more attention to culture than analysts.

Finally, we explore whether there is any relationship between corporate culture and price formation. We find that analysts' positive tones in discussing culture are positively and significantly associated with their stock recommendations and target prices. In terms of economic significance, a change in tones from neutral to positive is associated with a 3.2 percentage point-increase in the probability of analysts upgrading their recommendations and a 2.0 percentage point-increase in target price forecast relative to the mean. Moreover, the above effect is further substantiated if the report also contains culture-related cause (effect) analysis. Importantly, we also find that investors react positively and significantly to the positive tone in report text on culture only if the focal report exhibits very different tones from other analysts following the same firm, controlling for a report's quantitative and qualitative characteristics and analyst/firm characteristics. In terms of economic significance, a change in tones from neutral to positive by an analyst with one-standard-deviation away in tone the rest of the pack, results in an additional three-day abnormal return of 23.5 basis points around the release date of a report, corresponding to an $87.80 million increase in market value for an average firm in the sample. We conclude that analysts' research on culture offers new insights into its causes and effects, and that we are one step closer than prior work to establishing the culture-firm value link.

We contribute to the literature in three ways. First, our study provides new insights into corporate culture from the vantage points of different stakeholders of a modern corporation – equity analysts, management, and employees. By applying generative AI to rich and granular textual data for information extraction, we reveal the causes and effects of different culture types that have been difficult if not impossible to uncover in prior work. As such, our paper extends prior work that studies corporate culture (see, for example, Guiso, Sapienza, and Zingales 2015; Grennan 2019; Au, Dong, and Tremblay 2021; Li et al. 2021; Briscoe-Tran 2022; Graham et al. 2022a, 2022b; Huang et al. 2024; Li, Chen, and Shen 2024) by conducting the first large-scale study of insiders' (e.g., management) and outsiders' (e.g., equity analysts) perspectives on culture using a uniform methodological approach and trying to gain some understanding of why such divergence occurs. Our novel findings of different insiders' and outsiders' perspectives on culture facilitate a better understanding of the mechanisms through which culture affects business outcomes, and have a wide range of management implications.

Second, our study contributes to the literature on big data and machine learning in finance, accounting, and economics (Gentzkow, Kelly, and Taddy 2019; Goldstein, Spatt, and Ye 2021). Contemporaneous research demonstrates the power of generative AI models in enabling capital market participants to derive new information and glean valuable insights from large quantities of textual data (Bai et al. 2023; Bybee 2023; Jha, Qian, Weber, and Yang 2023; Kim, Muhn, and Nikolaev 2023; Lopez-Lira and Tang 2023; Li, Tu, and Zhou 2023). Several papers explore the impact of generative AI on firm value and stock returns (Bertomeu, Lin, Liu, and Ni 2023; Eisfeldt, Schubert, and Zhang 2023; Babina, Fedyk, He, and Hodson 2024). Our study differs from current research in two important ways. First, we apply generative AI for complicated information extraction rather than text classification or prediction tasks, and demonstrate its effectiveness. Our approach yields a multi-faceted output that encompasses different culture types and their respective causes or effects. Second, we combine the versatility of generative AI models such as ChatGPT, which are constrained by such factors as speed, cost, and context length limitations, with the efficiency of smaller large language models such as Bidirectional Encoder Representation from Transformers (BERT) models (Devlin, Chang, Lee, and Toutanova 2018), to effectively filter and retrieve the most relevant information.

On that note, our study highlights some unique considerations and design elements that must be addressed in order to harness the full potential of generative AI models in the context of financial text analysis. For instance, a step-by-step, chain-of-thought prompting strategy is beneficial for extracting perceived cause-effect relations – a task that requires high-level reasoning (Wei et al. 2022a). Furthermore, most generative AI models face inherent challenges when analyzing long documents.[1] We demonstrate that feeding smaller segments related to corporate culture, while allowing for the dynamic augmentation of input segments by searching for relevant information from a full report, can enhance the overall capability of these models.

Third, our study contributes to the literature on equity analysts, particularly the strand applying textual analysis to research reports in order to gain insights into their information discovery and interpretation roles (e.g., Asquith, Mikhail, and Au 2005; Twedt and Rees 2012; Huang, Zang, and Zheng 2014; Huang et al. 2018; Bellstam, Bhagat, and Cookson 2021).[2] By applying generative AI to one of the largest report samples available, our research sheds light on the "black box" of analysts' fundamental research by not only underscoring culture as an integral input, but also elucidating the deductive processes analysts employ to transform qualitative, soft information into actionable insights (e.g., stock recommendations). As such, our big data-based research complements case study/interview/survey approaches

[1] Due to how these models process text (using the attention mechanism where every word needs to interact with every other word), the computational cost grows quadratically with text length. In addition, although some models can theoretically process longer inputs, their performance degrades when accessing and analyzing relevant information in the middle of a longer input (Liu et al. 2023).

[2] Using a sample of 1,126 reports by 56 All-America analysts over the period 1997-1999, Asquith, Mikhail, and Au (2005) construct a measure for the strength of arguments and show that such measure reduces, and sometimes eliminates the significance of the information available in earnings forecast or recommendation revisions. Using a sample of 2,057 reports in 2006 and a dictionary approach to measure tones in reports, Twedt and Rees (2012) find that the tone in a report contains significant information incremental to its quantitative content (e.g., earnings forecasts). Using a sample of 363,952 reports issued for S&P 500 firms over the period 1995-2008, Huang, Zang, and Zheng (2014) show that investors react more strongly to analyst reports that emphasize non-financial topics more than financial topics. Applying topic modeling, Huang et al. (2018) find that analysts both provide new information and interpret information released by corporate managers in their reports. Using a similar methodology and 665,714 analyst reports issued for S&P 500 firms over the period 1990-2012, Bellstam, Bhagat, and Cookson (2021) show that their textual-based measure of corporate innovation is more comprehensive and accurate than standard metrics of innovation output through multiple validation tests. Yet little is known about analysts' deductive process to produce their research output.

(e.g., Soltes 2014; Brown et al. 2015; Chi, Hwang, and Zheng 2023) to gain a better understanding of analysts' information production process.[3]

2. Our Information Extraction Method

2.1. Overview

We employ generative AI to analyze corporate culture through the lens of analyst reports (earnings calls and employee reviews). By extracting cause-effect reasoning between culture types and business outcomes, we introduce a novel framework for representing the analytical logic underlying analysts' perspectives on culture in a structured format. The main outputs are triples capturing reasoning chains like "innovation and adaptability (culture type), leads to (direction), higher profitability (business outcome)." To examine whether analysts provide distinct analytical perspectives compared to other stakeholders, we construct divergence measures between analyst reports and earnings calls (employee reviews). This approach allows us to map analysts' reasoning about corporate culture while systematically quantifying how their perspectives differ from those of management (employees).

Our approach builds on recent social science studies that demonstrate the capability of generative AIs to process complex sociocultural constructs from unstructured text. For example, Rathje et al. (2024) show large language models (LLMs) excel at measuring sentiment and psychological constructs from text. They often outperform traditional machine learning approaches and have higher test-retest reliability than humans. Ziems et al. (2024) demonstrate that LLMs can reliably process qualitative information related to linguistic,

[3] After examining a set of proprietary records from a large-cap firm, Soltes (2014) concludes that analysts' private interactions with management help them interpret firm news and better understand a firm's operations. Based on surveys and interviews of analysts, Brown et al. (2015) find that analysts' private communications with management are more helpful to their earnings forecasts and stock recommendations than their own primary research, recent earnings performance, and 10-K/10-Q filings, suggesting that analysts may be better-positioned to assess culture prevailing in an organization. Chi, Hwang, and Zheng (2023) show that analysts utilizing alternative data (such as job postings and employee reviews) produce more accurate forecasts, and their forecasts generate greater stock market reactions.

psychological, and cultural categories for computational social sciences. They also can help parse unstructured text to a more structured format to express relationships between entities. These capabilities to analyze text at scale with minimal manual input make generative AI particularly suitable for our study.

This section describes our methodological approach to extracting cultural cause-effect relations from analyst reports. Figure 1 presents a flowchart of how we apply generative AI to extract analysts' perspectives on corporate culture from their research reports. We provide an overview of the key steps, including data preprocessing, triple extraction, and output validation. The implementation details, including modifications made to extract and analyze earnings calls and employee reviews, are provided in the Internet Appendix, and the code is available upon request.

2.2. Culture-related segments in reports

Our primary data consists of 2,451,766 analyst reports from Thomson One's Investext database covering S&P 1500 firms from 2000 to 2020. We convert the reports from PDF to plain text. To address the loss of paragraph structure during conversion, we implement the C99 algorithm developed by Choi (2000) to reconstruct coherent text segments.[4] We then implement a machine learning model to filter out boilerplate content from reports.[5]

We employ a three-stage approach to find culture-related discussions in reports. First, we conduct a keyword search using terms explicitly related to corporate culture (e.g., "corporate culture" or "workplace culture") to find the report segments. Second, we

---

[4] The C99 algorithm is a text segmentation method that identifies coherent segments within a document. It calculates a similarity score between pairs of sentences and uses these scores to determine where to place segment boundaries. Intuitively, the algorithm groups together consecutive sentences that discuss similar topics or concepts, thereby reconstructing the logical paragraph-like structure of a report.

[5] The data set for training consists of segments in reports produced by the top 20 brokers, with positive examples identified as the most frequently repeated segments and negative examples as those least repeated. We fine-tune a BERT model to classify boilerplate segments automatically, and the trained model demonstrates high accuracy. Table IA1 in the Internet Appendix lists predicted boilerplate probabilities and boilerplate examples, sorted by decile. We retain segments with a boilerplate probability of 0.22 (the sample median) or lower.

implement a hybrid machine learning classifier to capture segments that discuss culture without using these keywords. The classifier uses a BERT model trained on segments containing clear cultural keywords to identify segments with similar semantic patterns. Third, we use generative AI to filter these candidate segments and retain only those that substantively discuss organizational culture (Table IA2 in the Internet Appendix lists the prompts used). This approach captures both overt cultural discussions and segments where analysts examine culture through descriptions of organizational practices or values. Our final data set comprises 138,545 culture-related segments in 86,112 reports.

Figure 2 plots the intensity of culture-related segments in analyst reports over time and by report section over the period 2000-2020. The horizontal axis indicates report year, and the vertical axis indicates report section, binned into 20 equal sections from the start to the end of a report. The color gradient depicts the intensity of culture-related segments, computed as the number of these segments normalized by the number of reports in a year (we multiply this variable by a hundred).

We make two observations. First, there is a shift in the location of culture-related segments in reports. At the beginning of the sample period, these segments are generally scattered throughout a report. In more recent years, they appear more concentrated in the first half of a report. This shift suggests that, over time, analysts are more aware of the importance of culture, and hence they position their culture-related analyses in the front end of their reports. Second, there is a marked rise in the intensity of culture-related segments in reports in recent years following a discernable dip during and in the aftermath of the Great Recession (2008-2012); this dip could potentially be attributed to analysts’ heightened focus on financial performance and cost-cutting in a period of economic uncertainty.

2.3. Cause-effect relation extraction

We extract structured information about culture types and their causal relationships from analyst reports. For each report, we combine any culture-related segments into a single input document. We use GPT-4 (Achiam et al. 2023), specifically the *GPT-4o mini* model, to analyze these inputs. This choice follows an evaluation of multiple models based on accuracy and computational efficiency. We then use a prompt (shown in Table 1 Panel A) that guides the GPT-4 model to identify the specific culture type being discussed and categorize that value into one of six predefined culture types. The prompt also instructs the model to assess the overall sentiment or tone of the culture-related discussion, determine if the segment contains a detailed causal analysis—and if it does, extract the specific factors that influence the culture type (i.e., causes) and the specific business outcomes affected by the culture type (i.e., effects), and summarize the causal relationships in a structured triple format.

The six predefined culture types in the prompt (*collaboration and people-focused, customer-oriented, innovation and adaptability, integrity and risk management, performance-oriented,* and *miscellaneous*) are derived through a two-stage process. In a pilot study, we manually review the most frequent culture-related phrases across reports. Through discussion and reference to the original reports, we group those phrases into broad culture types guided by prior literature (see Table IA3 in the Internet Appendix for a survey of prior work). We then validate and refine these initial categories using a data-driven approach. We embed all extracted culture types into a vector space using a BERT model and cluster them using hierarchical clustering. By inspecting these clusters and mapping them to the manually defined categories, we arrive at the final set of six culture types that capture the major themes in analyst discussions.

After extracting the phrases representing causes and effects of each culture type, we conduct a similar two-stage process to map them to standardized entity names. We first embed all extracted causes and effects into a vector space and group them into 50 clusters.

Upon manual review of these clusters, we arrive at a set of 18 standardized cause (e.g., mergers and acquisitions, management change) and 17 effect entities (e.g., profitability, resilience) that capture the key themes in analyst culture-related discussions. Each standardized entity is associated with 5-10 representative phrases to illustrate its scope. We then create a prompt (shown in Table 1 Panel B) that provides the standardized cause/effect entities and their associated examples. For each extracted cause or effect phrase, the prompt instructs the model to map it to the most appropriate standardized entity based on similarity to the provided examples. Finally, we canonicalize the extracted cause-effect relation triples by assigning one end of each triple to a culture type, another end to a standardized cause/effect, and classify the relationship's direction (i.e., -> for cause to effect, <- for effect to cause, or <-> for bidirectional). This approach allows us to extract granular yet standardized triples about the antecedents and consequences of corporate culture from unstructured analyst discussions. Table IA4 in the Internet Appendix lists some representative examples of the extracted culture types and their causes and effects.

To improve the model's ability to accurately extract complex causal relationships, we implement two key techniques. First, we use chain-of-thought prompting (Wei et al. 2022) to break down the model's reasoning process into discrete steps. By providing a detailed step-by-step prompt, we enable the model to systematically analyze different aspects of the culture-related discussion before integrating the information into a final causal graph.

Second, we employ retrieval augmented generation (Lewis et al. 2020), which allows the model to request additional context if needed to complete the causal analysis. If the model outputs "I need more context" for any of the prompted questions, we dynamically retrieve other relevant segments in the report and provide them as additional input. To ensure the additional context remains focused on corporate culture, we filter the retrieved segments to retain only those exceeding the 75th percentile of culture-relevance probabilities based on our

BERT culture classification model. Importantly, we only provide extra context upon the model's request to avoid biasing the initial analysis with potentially irrelevant information. Table IA5 in the Internet Appendix provides examples illustrating how retrieved additional context helps the model make inferences.

2.4. Model performance and discussion

In this section, we evaluate the performance of three mainstream generative AI models. The first model belongs to the ChatGPT model series developed by OpenAI: GPT-4o-mini. The second model belongs to the Claude model series developed by Anthropic: Claude-3.5-Haiku. The third model belongs to Alphabet's Gemini model series, specifically Gemini-Flash-1.5. These three models share a similar cost per token. Our evaluation involves manually annotating culture types, causes, effects, and tones in 200 randomly selected culture-related segments, and comparing them with the relations and tones extracted by generative AI models. Table 2 presents the model performance results using accuracy, precision, recall, and F1 scores.

In terms of accuracy, GPT-4o-mini achieves 98.5%, 84.0%, 91.0% for culture types, causes, and effects, respectively. Claude-3.5-Haiku achieves 93.0%, 71.5%, and 77.0%, respectively. Gemini-Flash-1.5 achieves 94.0%, 75.5%, and 78.5%, respectively. We find similar results using precision, recall, and F1.

In terms of tones in culture-related segments, we show that GPT-4o-mini achieves an accuracy of 97.5%, slightly outperforming Claude-3.5-Haiku and Gemini-Flash-1.5, which achieve an accuracy of 96.5% and 95.0%, respectively.

In summary, we find all three models exhibit good model performance when extracting culture types, causes, effects, and tones in textual data. GPT-4o-mini has an edge compared to Claude-3.5-Haiku and Gemini-Flash-1.5, especially for identifying causes and effects. In the remainder of this paper, we employ GPT-4o-mini as our primary AI model to conduct analyses.

There are two general concerns when applying generative AI: look-ahead bias and hallucination. Look-ahead bias refers to using future information not available at the time of prediction. Our application focuses on extracting analysts' perspectives (or executives' or employees' perspectives) of corporate culture rather than making out-of-sample predictions. Moreover, our regression analyses primarily use the extracted cultural information based on information available at the time of a report's publication. In short, our approach and research design by construction mitigate look-ahead bias concerns.

Hallucination refers to generated text that is unfaithful or ignores source material (Maynez, Narayan, Bohnet, and McDonald 2020). A key strategy to mitigate hallucination is to improve the alignment between the input and the generated output (Ji et al. 2023). Our approach ensures that the input segments align closely with what the prompt asks for because the segments either contain explicit culture-related keywords or are selected through a multi-stage filtering process. Moreover, we implement additional safeguards in our prompt design. The prompt requires explicit textual evidence for each causal relationship by instructing the model to "*focus on specific results, business outcomes, or tangible impacts*" and "*avoid implicit or indirect outcomes*." For each relationship, the model must provide a "*reason for identifying the causal relation, citing specific words or phrases or logical reasoning*" from the text to support the causality. Our chain-of-thought prompting guides the model to reason strictly within a report's content in discrete steps. We also give the model the explicit option to output "N/A" or empty arrays when information is absent, or causal relationships cannot be confidently determined from the text. These design choices help constrain the model to extract only relationships that are directly supported by the source material.

2.5. Divergence in insiders' and outsiders' perspectives on culture

Our method for extracting and analyzing culture-related information extends beyond analyst reports to earnings conference calls and employee reviews on Glassdoor. We apply the same information extraction approach to these different data sets. For methodological consistency and to facilitate direct comparisons across different corpora, we employ the same taxonomies of culture types, causes, and effects when analyzing all three corpora. The specific adaptations are detailed in the Technical Appendix A in the Internet Appendix.

To quantify differences in cultural discussions across stakeholders (outsiders like analysts versus insiders like management and employees), we construct Jensen-Shannon (JS) divergence measures between analyst reports, earnings calls, and Glassdoor reviews. The JS divergence (Lin 1991) quantifies the dissimilarity between probability distributions and always has a finite value between 0 and 1. A divergence of 0 means the distributions are identical, while 1 indicates the maximum difference. For two probability distributions p_1 and p_2, it is defined as:

$$JS(p_1, p_2) \ = \ H(\frac{p_1 + p_2}{2}) \ - \frac{H(p_1) + H(p_2)}{2}$$

where H is the Shannon entropy $H(p) = -\sum_i p_\mathrm{i} \log(p_\mathrm{i})$.

We compute the divergence measures separately for four aspects of how corporate culture is discussed: culture types, causes, consequences, and tone. For each firm-year observation, we first create probability distributions of how frequently each cultural aspect is mentioned by analysts, management, and employees over the past three years. For example, consider culture type, where we have defined six main categories. Suppose in a given year, analyst reports for a firm mention “collaboration and people-focused” 40% of the time, “innovation and adaptability” 30% of the time, and “integrity and risk management” the remaining 30%. This would yield a 1 × 6 probability distribution of [0.4, 0, 0.3, 0.3, 0, 0, 0]. The earnings calls focus more heavily on collaboration and people-focused, with a 1 × 6 distribution of [0.6, 0, 0.2, 0.2, 0, 0, 0]. Meanwhile, Glassdoor reviews by employees emphasize ant innovation and adaptability culture, with a distribution of [0.2, 0, 0.5, 0.2, 0.1, 0]. The JS divergence between the analyst and management distributions is 0.029, indicating a relatively small difference in how they discuss corporate culture types. In contrast, the divergence between the analyst and employee distributions is 0.1, suggesting a more substantial difference.

We follow the same process for the 18 cause categories, 17 consequence categories, and 3 tone categories (positive, neutral, and negative). This yields four divergence measures for each firm-year: types, causes, consequences, and tones between any two corpora. These divergence measures help identify cases where analysts (outsiders) provide distinct perspectives on corporate culture. They complement our primary analysis by quantifying differences in how key stakeholders perceive and discuss corporate culture. High divergence values suggest analysts emphasize different aspects of corporate culture than what management communicates or employees experience.

3. Revealing Analysts' Perspectives on Corporate Culture from Reports

3.1. The cause-effect knowledge graph

Our generative AI-fueled method allows us to identify the events, people, and/or systems that significantly influence a specific culture type, and to determine which culture types are most impactful for various business outcomes. Figure 3 plots the cause-effect knowledge graph capturing analysts' perspectives on corporate culture.

The graph is divided into three columns of entities: 1) the left column lists the seventeen drivers of culture grouped by events, people, and systems suggested by Guiso, Sapienza, and Zingales (2015), Graham et al. (2022a, 2022b), and Grennan and Li (2023) (omitting the miscellaneous category); 2) the center column lists the five culture types (omitting the miscellaneous category); and 3) the right column lists the sixteen effects of culture (omitting the miscellaneous category). The height of each entity (a culture type, a cause, or an effect) denotes the number of relevant segments, thereby providing an intuitive visual representation of each entity's prominence. The width of each link denotes the number of segments mentioning a particular cause or effect relation. For an uncluttered depiction, we only retain the top two most frequent links for each entity in Figure 3.

Focusing on the links coming out of the cause column (on the left side of Figure 3), analysts identify that business strategy is the top influencing factor for three culture types – performance-oriented, innovation and adaptability, and customer-oriented, and management team for performance-oriented, innovation and adaptability, and collaboration and people-focused. Focusing on the links coming into the culture type column (in the middle of Figure 3), analysts identify that management team (people) and employee hiring and retention (a system) are the top two factors shaping the collaboration and people-focused culture, and business strategy (a system) and management change (people) are the top two factors shaping the innovation and adaptability culture. Focusing on the links coming out of the culture type column (in the middle of Figure 3), analysts identify the collaboration and people-focused culture as having an impact on almost all aspects of business operations, ranging from market share and growth to risk management, whereas the customer-oriented culture has less impact. Focusing on the links coming into the effect column (on the right side of Figure 3), analysts identify performance-oriented and innovation and adaptability as the top two culture types shaping market share and growth and profitability. Notably, the above discussions are merely the proverbial tip of the iceberg, with many additional granular insights into the complex relations among the five culture types, seventeen causes, and sixteen effects available from these links. To the best of our knowledge, such insights are wholly absent from the existing literature on corporate culture.

In the effect column, we utilize a color-coding scheme to represent tones in culture-related segments, with darker shades signifying more negative tones. We find that market share and growth, customer satisfaction, and innovation are viewed positively, whereas misconduct, internal conflicts, and risk management are viewed negatively in analysts' analyses of business outcomes relating to culture. In our sample, 61.3% of analysts have positive tones when discussing culture. By comparison, Graham et al. (2022a) note that 63% of the surveyed executives express positive sentiments when discussing culture.

Overall, Figure 3 provides a comprehensive visualization of analysts' perspectives on corporate culture spanning different culture types, their causes, and their effects. The natural question to ask is whether and how insiders' and outsiders' perspectives on culture differ.

3.2. Comparing insiders' and outsiders' perspectives on culture

To shed light on the extent to which analysts' perspectives on corporate culture align with or deviate from those of company insiders, we visualize the frequency of different culture types, causes, and effects across analyst reports, earnings calls, and Glassdoor employee reviews in Figure 4.

Panel A presents the frequency of culture types using vertical bar charts. There is generally high agreement between analysts' and management's views on the types. Yet, the relative frequency of integrity and risk management is markedly higher in analyst reports compared to the other two sources, which aligns with analysts' monitoring role. Compared to other types, collaboration and people-focused culture dominates employee reviews on Glassdoor. This likely reflects employees' natural focus on the day-to-day experience of organizational culture and how it affects them personally.

Panels B and C present the frequency of different causes and effects of corporate culture. Two general observations emerge. First, we see a general agreement in the ranking of the causes and consequences across the three sources. This strong agreement lends validity to our approach of extracting cultural insights from unstructured text data using generative AI. Second, for the majority of causes and effects, the frequency in analyst reports falls between those in earnings calls and employee reviews. This balanced perspective indicates that analysts incorporate insights from both management's strategic framing and employees' lived experiences in their assessments of the drivers and consequences of corporate culture.

Nevertheless, there are some noteworthy differences. Business strategy, customer relations, disruptive technology, business relationships, and COVID-19 are more frequently cited as causes of cultural change in earnings calls. Management team, employee hiring and retention, internal conflicts, and workplace safety feature more prominently in employee reviews. These differences suggest that management focuses on high-level drivers and external events as key influencers of corporate culture, while employees view people and internal elements as the primary drivers.

In terms of effects, market share and growth and profitability are discussed much more extensively in earnings calls and by analysts, likely reflecting management's focus on financial performance. Employee satisfaction and internal conflicts, in contrast, are more salient effects in Glassdoor reviews.

In summary, we find that while analysts' cultural narratives largely balance the perspectives of senior management and rank-and-file employees, they also exhibit some distinctive emphases, such as a greater focus on integrity and risk management and a lesser focus on performance-oriented culture. Previous studies on corporate culture have typically relied on a single internal source, either employee reviews or management discussions (e.g., Li et al. 2021; Graham et al. 2022a, 2022b; Huang et al. 2024; Li, Chen, and Shen 2024). Clearly, these singular perspectives inherently reflect the unique vantage points and priorities of their respective stakeholder groups. The insights of equity analysts have the potential to provide a new understanding of corporate culture—a proposition that we formally explore in the rest of the paper.

4. Understanding Analysts' Perspectives on Corporate Culture

Unlike numbers and information gleaned from financial statements, the notion of corporate culture is somewhat nebulous and thus raises a number of questions regarding our research premise: What firm and analyst characteristics are associated with them covering corporate culture in their research reports? We have shown that analysts do not simply reiterate management's narratives in earnings conference calls (or employees' reviews of their organizational culture) in the analysis above. It would be interesting to investigate whether any firm characteristics, including managerial incentives, are associated with the divergence in insiders' and outsiders' perspectives on corporate culture. We address these questions in this section.

4.1. Sample formation and overview

We download from Thomson One's Investext 2,451,766 reports covering S&P 1500 constituent firms over the period 2000–2020. We obtain report date, gvkey, lead analyst name (including last name and first name initial), and broker name from the meta file. Our earnings call data over the period 2004–2020 is from Capital IQ Transcripts database. Our Glassdoor employee review data over the period 2008–2020 is from Revelio Labs. Section A.7 of the Internet Appendix describes sample formation steps. Section B of the Internet Appendix describes how we match analyst name in a report to analyst ID in the Institutional Brokers Estimates System (I/B/E/S) database, in order to construct analyst characteristic variables.

Table 3 Panel A presents the coverage of culture-related discussions (culture types, causes, and effects) in analyst reports, earnings calls, and employee reviews. We note that out of 86,112 reports with culture mentioning, analysts provide both cause and effect analysis in almost 40% of those reports. Similarly, out of 101,533 earnings calls in which management discusses culture, executives provide both cause and effect analysis in 56% of those calls. In contrast, out of 866,796 employee reviews with culture mentioning, less than a tenth of those reviews contain both cause and effect analysis. Moreover, employees seem to say more about the effect of their culture, and say far less about any cause of a cultural change, compared to management and analysts. Panel B presents the coverage of divergence measures between analysts and executives and between analysts and employees (using a three-year window to maximize the sample size).

Table IA7 in the Internet Appendix provides the summary statistics for the different samples used in our analyses. All continuous variables are winsorized at the 1st and 99th percentiles, and the dollar values are in 2020 dollars. Variable definitions are provided in the Appendix. Panel A presents the summary statistics for the firm-year sample. At the firm-year level, our measure of the frequency of analysts' discussions of culture, *Culture discussion*, has a mean of 0.563, suggesting that there is at least one analyst discussing culture in 56 percent of firm-year observations. Panel B presents the summary statistics for the firm-analyst-year sample. Panels C and D present the summary statistics for the firm-year sample relating to divergence measures. Table IA8 in the Internet Appendix presents the Pearson correlation matrices of the firm-year samples and the firm-analyst-year sample. Examination of the correlation matrices suggests that multicollinearity is unlikely to be an issue.

4.2. Firm characteristics and analysts' discussing corporate culture

To examine whether and how firm characteristics are related to the likelihood of analysts' research on culture in their reports, we employ the following regression specification at the firm-year level:

$$Y_{i,t} = \alpha + \beta \times Firm\ characteristics_{i,t-1} + Ind\ FE + Year\ FE + \varepsilon_{i,t}, \quad (1)$$

where the dependent variables $Y_{i,t}$ is whether analysts discuss culture (or not). Firm characteristics largely follow prior work (Guiso, Sapienza, and Zingales 2015; Li et al. 2021; Li, Chen, and Shen 2024). We further include the number of broker/investor conferences participated/hosted by a firm, as a proxy for the extent of interaction between analysts and their covered firm. In one specification, we include industry fixed effects to control for the effect of time-invariant industry factors on analysts' discussions of culture, and year fixed effects to control for changing trends in analysts' awareness of culture. In another specification, we include firm and year fixed effects, the former is to control for the effect of time-invariant firm factors on analysts' discussions of culture.

Table 4 Panel A presents the results when the dependent variable is whether a firm's analysts discuss culture or not. Columns (1) and (4) present the regression results including the basic set of firm-level controls with different sets of fixed effects. Columns (2) and (5) further add culture and number of meetings between a firm and its analysts. Columns (3) and (6) further add two Glassdoor review-related variables: employee culture rating and number of employee reviews. We find that firm size, sales growth, profitability, and major events such as top management turnover and deal-making, firms whose management emphasizes culture, firms that host/participate in investor/broker conferences are positively and significantly associated with, whereas firm age, leverage, tangibility, earnings volatility, ownership by large shareholders, and board independence are negatively and significantly associated with, analysts' discussing culture in their reports.

Limiting to a subsample of firm-year observations with culture-related segments in analyst reports, we examine the determinants of analysts' discussing different culture types. Panel B presents the results. We see some interesting variations in analysts' coverage of different culture types. For example, tangibility is positively (negatively) and significantly associated with analysts' discussing customer-oriented culture (innovation and adaptability culture), earnings volatility is only negatively and significantly associated with analysts' discussing collaboration and people-focused culture. Our findings on the negative influences of large shareholders and the positive influences of key corporate events (such as management turnover and M&As) are largely consistent with prior literature (e.g., Guiso, Sapienza, and Zingales 2015; Li et al. 2021). Our findings on the positive association between management discussing culture in calls (the number of meetings firms host/participate in investor/broker conferences) and analysts' featuring culture in their reports, suggesting that analysts do pay attention to value-relevant intangibles such as corporate culture in their research.

4.3. Analyst characteristics and their discussing corporate culture

To examine what analyst characteristics are associated with the likelihood of their writing about culture in reports, we employ the following regression specification at the firm-analyst-year level:

$$Y_{i,j,t} = \alpha + \beta \times Analyst\ characteristics_{j,t-1} + Firm \times Year\ FE + Broker\ FE + \varepsilon_{i,j,t}, \quad (2)$$

where the dependent variables $Y_{i,j,t}$ are: whether an analyst discusses culture (or not), the number of culture types discussed, the number of causes discussed, the number of effects discussed, and her tone in discussing culture. Our analyst characteristics largely follow prior literature (e.g., Clement and Tse 2005). In addition to analyst characteristics, we include firm × year and/or broker fixed effects to control for time-varying unobservable firm and/or broker characteristics that may affect analysts' coverage decisions and/or their decisions to discuss culture in reports. Table 4 Panel C presents the regression results.

Column (1) presents the regression results when the dependent variable is whether an analyst discusses culture or not. We show that analysts who are star analysts, are CFA charter holders, have postgraduate degrees, are women; who have more general experience and make more frequent forecasts; and who are affiliated with large brokers are each more likely to discuss corporate culture. In contrast, analysts who follow more firms are less likely to discuss culture.

Conditional on an analyst discussing culture in her report, columns (2)-(5) present the regression results when the dependent variables capture the scope and depth of her culture coverage and the tone used. We note that star analysts, analysts with more general and firm-specific experiences, and analysts who make more frequent forecasts are associated with discussing more culture types in their reports. Analysts with more general experience and less firm-specific experience are associated with more positive tone when discussing culture. We show little explanatory power from other analyst characteristics considered.

We conclude that the significant associations between certain analyst characteristics – star status, gender, experience, and effort – and coverage of culture help assuage concerns about analysts' indifference to or lack of insights into corporate culture.[6]

In summary, our analyses thus far suggest that generative AI models have the potential to reveal new insights into corporate culture from the vantage points of sell-side equity analysts, and that those insights differ from corporate insiders' perspectives on culture. This begs the question of what explains those divergent perspectives on corporate culture.

4.4. Explaining different perspectives on culture by corporate insiders and outsiders

[6] In Table IA9 in the Internet Appendix, we explore what firm and analyst characteristics are associated with divergence in perspectives on culture among fellow analysts. At the firm-year level, we note that firm size, loss year, CEO who is close to retirement are positively and significantly, whereas CEO equity incentives are negatively and significantly, associated with the divergence in perspectives on culture among equity analysts. At the firm-analyst-year level, including firm times year fixed effects and broker fixed effects, we note that analysts with a postgraduate degree or analysts following more industries are negatively and significantly, whereas analysts following more firms are positively and significantly, associated with the divergence in tone about culture. At the firm-analyst-year level, using the signed measure of divergence in perspectives on a specific culture type or tone, we see that education, gender, and the scope of coverage matter somewhat. All in all, it seems that after removing time-varying firm characteristics (with the firm times year fixed effects) and time-invariant broker effects, there are not many analyst characteristics that could explain inter-analyst differences in perspectives on culture.

We next explore the determinants of any divergence in insiders' and outsiders' perspectives on culture. Given our multi-dimensional output generated from our information extraction methods – culture types, causes, effects, and tones, we employ two sets of measures to capture the degree of divergence in views. Our first set is based on Jensen-Shannon (JS) divergence measures between analyst reports and earnings calls (Glassdoor reviews). The JS divergence (Lin 1991) quantifies the dissimilarity between probability distributions and always has a finite value between 0 and 1. A divergence of 0 means the distributions are identical, while 1 indicates the maximum difference. This set of measures does not have a sign. Our second set is signed measures based on the difference between analysts' and management's (employees') perspectives. Table 5 presents the regression results.

Using the JS measures as summary measures of divergence in perspectives, we show that firm size is negatively and significantly associated with any divergence in insiders' and outsiders' perspectives on culture constructs, whereas ROA volatility is positively and significantly associated with divergent views. The first result suggests that there is less difference in insiders' and outsiders' perspectives on culture when firms are large. Interestingly, we note that CEO risk-taking incentives, as captured by their equity portfolio's vega (Guay 1999), are negatively and significantly associated with any divergence between analysts' and management's perspectives on culture.

Using the signed measures of divergence in perspectives on specific culture types and tones, we show that firm size, ROA, CEO equity incentives, and the number of broker/investor conferences participated/hosted are positively and significantly, whereas board independence is negatively and significantly, associated with the gap in analysts' focus on corporate culture compared to that of management. In contrast, we show that sales growth and leverage are positively and significantly, whereas firm size, ROA volatility, large institutional ownership, and the number of broker/investor conferences participated/hosted are negatively and significantly, associated with the gap in analysts' focus on corporate culture compared to that of employees. Taken as a whole, our results suggest that in large firms or firms participating/hosting more broker/investor conferences, the gap between analysts and management paying attention to culture is larger than in small firms or firms participating/hosting fewer broker/investor conferences. Similarly, in small firms or firms participating/hosting more broker/investor conferences, the gap between analysts and employees paying attention to culture is smaller than in large firms or firms participating/hosting fewer broker/investor conferences.

5. Analysts' Perspectives on Corporate Culture and Stock Price Implications

Prior studies show that analysts' fundamental research contributes to stock price formation (see, for example, Womack 1996; Brav and Lehavy 2003; Loh and Stulz 2011; Huang, Zang, and Zheng 2014; Kecskés, Michaely, and Womack 2017). In this section, we examine whether and how analysts' perspectives on culture impact price formation. The analysis is at the report level. Section C of the Internet Appendix describes how we match reports from Investext in our sample to the I/B/E/S forecast data.

5.1. Analysts' perspectives on corporate culture and their research output

To examine the relationship between analysts' perspectives on culture and their research output, we focus on stock recommendations and target prices because both measures capture a firm's long-term prospects (e.g., Brav and Lehavy 2003; Loh and Stulz 2011), which aligns well with the role of corporate culture in long-term value creation (e.g., Guiso, Sapienza, and Zingales 2015; Li, Liu, Mai, and Zhang 2021). Moreover, a strong culture could over time boost a firm's cash flows and/or lower its discount rate; either or both outcomes would be reflected in stock recommendations and target prices. Given the importance of sentiment in textual data (see, for example, Antweiler and Frank 2004; Tetlock 2007; Loughran and McDonald 2011; Huang, Zang, and Zheng 2014), the key variable of interest is ChatGPT's classification of analysts' tones in culture-related segments, *Tone*. We also explore whether an analyst's more in-depth culture analysis involving causes and effects could affect her research output. We further control for tones in the remainder of a report, *Non-culture tone*. Table 6 Panel A presents the summary statistics for the key variables. We note the mean (median) number of causes and effects discussed in a report is 1.4 (2.0). Other statistics are largely consistent with prior literature (e.g., Bradshaw, Brown, and Huang 2013; Huang, Zang, and Zheng 2014; Kecskés, Michaely, and Womack 2017). Panel B presents the regression results. We include firm × year and analyst fixed effects to control for time-varying unobservable firm characteristics that may affect analysts' coverage decisions and analyst innate skill or preferences relating to their discussions of corporate culture, respectively.

We show that *Tone* is positively and significantly associated with stock recommendations and target prices, suggesting that analysts' sentiments on culture play a significant role in their stock recommendations and target price forecasts. In terms of economic significance, a change in *Tone* from negative to neutral (or from neutral to positive) is associated with a 3.2 percentage point-increase in the (linear) probability of analysts upgrading their recommendations and a 2.0 percentage point-increase in target price forecast relative to the mean.[7] Moreover, the above effect is further substantiated if the report also contains culture-related cause (effect) analysis. The coefficient on the interaction term *Tone × Number of cause and effect* is positive and significant at the 5% level in columns (2) and (4). In terms of economic significance, a one-standard-deviation increase in the number of

[7] The 3.2 percentage point-increase is calculated from 1 × 0.127 × 100/4 where the denominator 4 is the range of stock recommendation from -2 to 2, and the 2.0 percentage point-increase is calculated from 1 × 0.024 × 100/1.173 where the denominator 1.173 is the sample average target price relative to the mean.

cause and effect discussed in a report adds a further .52 basis points increase in stock recommendation, and a further 0.42 percentage point-increase in target price forecast relative to the mean.[8]

For comparison, a change in *Non-culture Tone* from negative to neutral (or from neutral to positive) is associated with a 14.1 percentage point-increase in the (linear) probability of analysts upgrading their recommendations and a 9.2 percentage point-increase in target price forecast relative to the mean.[9] Given that the non-culture segments regarding a firm's financial performance are more directly related to stock recommendations and target price forecasts, the effect of culture-related *Tone* is noteworthy.

5.2. The information content of analysts' perspectives on corporate culture

To investigate the information content of analysts' perspectives on corporate culture in reports, we employ an event study relating three-day cumulative abnormal returns (CAR) around the report date, to measures of analysts' perspectives on culture controlling quantitative and qualitative summary measures of a report, and analyst and firm characteristics (Huang, Zang, and Zheng 2014; Huang et al. 2018). Table 7 Panel A presents the summary statistics of the key variables used. Panel B presents the regression results.

We show that in column (1), *Tone* is positively and significantly associated with *CAR[-1,+1],* suggesting that culture discussions in a report provide information beyond that provided by its quantitative and qualitative measures. However, after adding the interaction term *Tone* × *Tone divergence with other analysts,* the standalone term *Tone* loses its significance. In contrast, the coefficient on the interaction term is positive and significant. In terms of economic significance, a change in *Tone* from negative to neutral (or from neutral to

[8] The 0.52 percentage point-increase is calculated from 1.225 × 0.017 × 100/4 where the denominator 4 is the range of stock recommendation from -2 to 2, and the 0.42 percentage point-increase is calculated from 1.225 × 0.004 × 100/1.173 where the denominator 1.173 is the sample average target price relative to the mean.
[9] The 14.1 percentage point-increase is calculated from 1 × 0.562 × 100/4, and the 9.2 percentage point-increase is calculated from 1 × 0.108 /1.173.

positive) by an analyst whose tone is one-standard-deviation away from her peers results in an additional three-day abnormal return of 23.5 basis points around the report date, corresponding to an $87.8 million increase in market value for an average firm in the sample.[10] For comparison, a change in *Non-culture tone* from negative to neutral (or from neutral to positive) results in an additional three-day abnormal return of 199.4 basis points, corresponding to a $745.0 million increase in market value for an average firm in the sample.[11] It is worth noting that the effect documented above is the direct information effect of analysts' perspectives on culture in reports, and that there are also indirect effects via stock recommendation and target price revisions shown in Table 6.

We conclude that analysts incorporate their perspectives on corporate culture in their research output, and that investors significantly react to report text on culture.

6. Conclusions

Our study is among the first in finance, accounting, and economics to apply generative AI models as reasoning agents on analyst reports, earnings call transcripts, and employee reviews to gain insights into corporate insiders' and outsiders' perspectives on corporate culture.

We employ generative AI (ChatGPT) to analyze 2.4 million analyst reports between 2000 and 2020, 243 thousand calls between 2004 and 2020, and 5.3 million employee reviews between 2008 and 2020. Generative AI organizes analysts' views into a knowledge graph that links different cultural values to their perceived causes and effects. We show that both analysts and management identify innovation and adaptability, collaboration and people-

[10] The 23.5 basis points increase is calculated from 0.215 × 1.093 ×100, and the $87.8 million increase in market value of equity is calculated from 0.235% × $37.4 billion where $37.4 billion is the sample average market capitalization.
[11] The 113.6 basis points increase is calculated from 1 × 1.994 × 100, and the $745.0 million increase in market value of equity is calculated from 1.136% × $37.4 billion.

focused, and performance-oriented as the most important to them. In contrast, employees view collaboration and people-focused culture as the most important to them. Moreover, both analysts and management believe business strategy is the most important driver for cultural changes, whereas employees view the management team as primarily responsible for cultural changes. In terms of culture's effects, both analysts and management believe market share and growth as the number one outcome of firms with a strong culture, whereas employees view their satisfaction as the most important outcome. These systematic differences align with the distinct roles and incentives of each stakeholder group. They also provide evidence of our approach's effectiveness in distinguishing between insiders' and outsiders' perspectives on corporate culture.

Finally, we show that analysts' perspectives on corporate culture are reflected in their stock recommendations and target price forecasts as well as impact price reactions to the release of their reports. We conclude that analysts' research on corporate culture offers new insights into its causes and effects, and that we are closer to establishing the culture-firm value link.

Our paper highlights the tremendous potential of generative AI in extracting cause-effect relations, and offers a roadmap for applying generative AI to finance and accounting research.

Appendix
Variable definitions

All continuous variables are winsorized at the 1st and 99th percentiles. All dollar values are in 2020 dollars.

Variable	Definition
Firm-year level	
Culture discussion	An indicator variable that takes the value of one if corporate culture is discussed in analyst reports in a year, and zero otherwise.
Collaboration and people-focused	An indicator variable that takes the value of one if collaboration and people-focused culture is discussed in analyst reports in a year, and zero otherwise. Other culture type-specific indicators (e.g., customer-oriented) are defined analogously.
Number of types (causes/effects)	Number of culture types (causes/effects of culture) discussed in analyst reports in a year.
Tone	The average tone of culture-related segments in analyst reports in a year. ChatGPT classifies each segment as negative (-1), neutral (0), or positive (1).
Type (Cause/Effect) divergence	The Jensen-Shannon (JS) divergence measure between analysts and management (employees) quantifies the dissimilarity between probability distributions and always has a finite value between 0 and 1. A divergence of 0 means the distributions are identical, while 1 indicates the maximum difference. Specifically, for each firm-year observation, we first create a probability distribution of how frequently each culture type is mentioned by analysts (management/employees) over the past three years into a six by one vector (for causes, it will be an eighteen by one vector; for effects, it will be a seventeen by one vector). Once we have two probability distributions p_1 and p_2 representing analysts' and management's perspectives, respectively, the type divergence measures is computed as: $$JS(p_1, p_2) = H(\frac{p_1 + p_2}{2}) - \frac{H(p_1) + H(p_2)}{2}$$ where H is the Shannon entropy $H(p) = -\sum_i p_i \log(p_i)$.
(Analyst-Executive) Collaboration and people-focused	The signed difference in perspectives by analysts and management regarding collaboration and people-focused culture, computed as the difference in the frequency count in analyst reports and earnings calls over the past three years. Other univariate signed differences are defined analogously.
Total assets	Book value of total assets (in millions of dollars).
Firm size	Natural logarithm of total assets.
Firm age	Number of years since a firm first appears in Compustat.
Sales growth	One year sales changes divided by last year sales.
ROA	Operating income before interest and taxes divided by total assets.
Leverage	Book value of debt divided by total assets.
Tangibility	Net property, plant, and equipment divided by total assets.
ROA volatility	Standard deviation of annual ROA, calculated over last three years, multiplied by one hundred.
Large institutional ownership	The fraction of shares outstanding held by institutional investors with at least 5% of ownership of a firm. Missing values are assigned zero.
Board independence	The share of independent directors on a board.
CEO duality	An indicator variable that takes the value of one if the CEO is also Chairman of the Board, and zero otherwise.
Number of key people changes	Number of top executive and board member changes in a year. The data is from Capital IQ Key Developments database.
Number of M&As	Number of announcements related to mergers and acquisitions in a year. The data is from Capital IQ Key Developments database.
Strong culture	An indicator variable that take the value of one, if a firm's cultural score is above the top quartile in a year, and zero otherwise. A firm's cultural score is the sum of five cultural values (innovation, integrity, quality, respect, and teamwork), extracted

	from earnings conference calls in a year. The data is from Li et al. (2021).
Employee culture rating	The average of employee ratings of culture & values (1-5) from all (i.e., current and former) employees in a year. The data is from Glassdoor.
Number of employee reviews	The number of all (i.e., current and former) employees providing culture & values ratings in a year. The data is from Glassdoor.
Loss year	An indicator variable that takes the value of one if a firm experiences a negative ROA in a year, and zero otherwise.
CEO tenure	The number of years a CEO is in office. The data is from ExecuComp.
CEO Delta	The change in the dollar value of a CEO's wealth for a one percentage point change in stock price (Guay 1999). The data is from ExecuComp.
CEO Vega	The change in the dollar value of a CEO's wealth for a 0.01 change in the annualized standard deviation of stock returns (Guay 1999). The data is from ExecuComp.
CEO close-to-retire	An indicator variable that takes the value of one if the age of a CEO is over 61 years old, and zero otherwise. The data is from ExecuComp.
Number of meetings	The sum of the number of investor meetings organized by a firm and the number of broker-held meetings that a firm is invited to attend in a year. Investor meetings include analyst/investor day (with a key development id of 192) and shareholder/analyst calls (with a key development id of 50). Broker-held meetings are company conference presentations (with a key development id of 51). The data is from Capital IQ Key Development.
Firm-analyst-year level	
Culture discussion	An indicator variable that takes the value of one if corporate culture is discussed by an analyst in her reports in a firm-year, and zero otherwise.
Number of types (causes/effects)	The number of culture types (causes/effects of culture) discussed by an analyst in her reports in a firm-year.
Tone	The average tone of culture-related segments in an analyst's reports in a firm-year.
Star analyst	An indicator variable that takes the value of one if an analyst is accredited to All-America research team status, and zero otherwise.
CFA	An indicator variable that takes the value of one is an analyst is a Chartered Financial Analyst (CFA) charter holder, and zero otherwise.
Postgraduate	An indicator variable that takes the value of one if an analyst possesses a postgraduate degree, and zero otherwise.
Female	An indicator variable that takes the value of one if an analyst is a female, and zero otherwise.
Local analyst	An indicator variable that takes the value of one if an analyst is based in an office that is within driving distance from a focal firm's headquarters (100 miles), and zero otherwise. The data for analyst office address is from Capital IQ. The data for firm headquarters address is sourced from their SEC 10-X filing headers. To determine the coordinates (latitudes and longitudes) of both addresses, we use Google Map API. We then calculate the distance between analyst office and corporate headquarters using the Haversine formula.
General experience	The number of years for which an analyst makes at least one forecast of any firm.
Firm experience	The number of years for which an analyst makes at least one forecast of a given firm.
Number of industries followed	Number of two-digit SIC industries in which an analyst makes at least one forecast of any firm in that industry.
Number of firms followed	Number of firms for which an analyst makes at least one forecast.
Forecast frequency	Number of forecasts that an analyst makes of a given firm.
Broker size	Natural logarithm of number of analysts making at least one forecast at a given broker.
Forecast horizon	The average of forecast horizons (in terms of the number of years based on FPI in I/B/E/S) that an analyst employs when making forecasts in a firm-year.

Report-level	
Recommendation	Stock recommendation in a report using a five-tier rating system where 2 represents "strong buy," 1 represents "buy," 0 represents "hold," -1 represents "underperform," and -2 represents "sell."
Target price	Target price in a report divided by the stock price 50 days before the report date, following Huang, Zang, and Zheng (2014).
CAR[-3,+3]	Cumulative seven-day abnormal return (in percentage points) centered around the report date (day 0) based on a market model in which the market portfolio is the CRSP value-weighted market index.
Tone	The average tone of culture-related segments in a report.
Number of cause and effect	The number of causes and effects of culture in a report.
Non-culture tone	The average tone of non-culture-related segments in a report. FinBERT classifies each segment as negative (-1), neutral (0), or positive (1).
Report length	Natural logarithm of number of segments in a report.
Earnings forecast revision	Earnings forecast in a report minus the last earnings forecast in I/B/E/S issued by the same analyst for the same firm, divided by the stock price 50 days before the report date.
Recommendation revision	Recommendation in a report minus the last recommendation in I/B/E/S issued by the same analyst for the same firm.
Target price revision	Target price in a report minus the last target price in I/B/E/S issued by the same analyst for the same firm, divided by the stock price 50 days before the report date.
Prior CAR	Cumulative ten-day abnormal return (in percentage points) ending two trading days before the report date based on a market model in which the market portfolio is the CRSP value-weighted market index.

References

Achiam, Josh, et al., 2023. GPT-4 technical report, arXiv preprint arXiv:2303.08774.

Ahern, Kenneth R., Daniele Daminelli, and Cesare Fracassi, 2015, Lost in translation? The effect of cultural values on mergers around the world, *Journal of Financial Economics* 117, 165–189.

Antweiler, Werner, and Murray Z. Frank, 2004, Is all that talk just noise? The information content of internet stock message boards, *Journal of Finance* 59, 1259–1294.

Asquith, Paul, Michael B. Mikhail, and Andrea S. Au, 2005, Information content of equity analyst reports, *Journal of Financial Economics* 75, 245–282.

Au, Shiu-Yik, Ming Dong, and Andreanne Tremblay, 2021, Employee flexibility, exogenous risk, and firm value, *Journal of Financial and Quantitative Analysis* 56, 853–884.

Babina, Tania, Anastassia Fedyk, Alex He, and James Hodson, 2024, Artificial intelligence, firm growth, and product innovation, *Journal of Financial Economics* in press.

Bai, John Jianqiu, Nicole M. Boyson, Yi Cao, Miao Liu, and Chi Wan, 2023, Executives vs. chatbots: Unmasking insights through human-AI differences in earnings conference Q&A, Northeastern University working paper.

Bellstam, Gustaf, Sanjai Bhagat, and J. Anthony Cookson, 2021, A text-based analysis of corporate innovation, *Management Science* 67, 4004–4031.

Bertomeu, Jeremy, Yupeng Lin, Yibin Liu, and Zhenghui Ni, 2023, Capital market consequences of generative AI: Early evidence from the ban of ChatGPT in Italy, Washington University in St. Louis working paper.

Birru, Justin, Sinan Gokkaya, Xi Liu, and René M. Stulz, 2022, Are analyst short-term trade ideas valuable? *Journal of Finance* 77, 1829–1875.

Blanco, Eduardo, Nuria Castell, and Dan Moldovan, 2008, Causal relation extraction, *Proceedings of the 6th International Conference on Language Resources and Evaluation, LREC*.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan, 2003, Latent dirichlet allocation, *Journal of Machine Learning Research* 3, 993–1022.

Bradshaw, Mark T., Lawrence D. Brown, and Kelly Huang, 2013, Do sell-side analysts exhibit differential target price forecasting ability? *Review of Accounting Studies* 18, 930–955.

Brav, Alon, and Reuven Lehavy, 2003, An empirical analysis of analysts' target prices: Short-term informativeness and long-term dynamics, *Journal of Finance* 58, 1933–1967.

Briscoe-Tran, Hoa, 2022, Diversity and inclusion, operating flexibility, and financial policies, Ohio State University working paper.

Brown, Lawrence D., Andrew C. Call, Michael B. Clement, and Nathan Y. Sharp, 2015, Inside the "black box" of sell-side financial analysts, *Journal of Accounting Research* 53, 1–47.

Bybee, Leland, 2023, The ghost in the machine: Generating beliefs with large language models, ArXiv preprint arXiv:2305.02823.

Chi, Feng, Byoung-Hyoun Hwang, and Yaping Zheng, 2023, The use and usefulness of big data in finance: Evidence from financial analysts, Nanyang Business School Research Paper No. 22-01.

Choi, Freddy Y.Y., 2000, Advances in domain independent linear text segmentation, In *Proceedings of the 1st North American Chapter of the Association for Computational Linguistics Conference, NAACL 2000*, USA: Association for Computational Linguistics, 26–33.

Clement, Michael B., and Senyo Y. Tse, 2005, Financial analyst characteristics and herding behavior in forecasting, *Journal of Finance* 60, 307–341.

Derrien, François, and Ambrus Kecskés, 2013, The real effects of financial shocks: Evidence from exogenous changes in analyst coverage, *Journal of Finance* 68, 1407–1440.

Devlin, Jacob, Ming-wei Chang, Kenton Lee, and Kristina Toutanova, 2018, BERT: Pre-training of deep bi-directional transformers for language understanding, ArXiv preprint arXiv:1810.04805.

Eisfeldt, Andrea L., Gregor Schubert, and Miao Ben Zhang, 2023, Generative AI and firm values, University of California, Los Angeles working paper.

Etzioni, Oren, Michele Banko, Stephen Soderland, and Daniel S. Weld, 2008, Open information extraction from the web, *Communications of the ACM* 51, 68–74.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy, 2019, Text as data, *Journal of Economic Literature* 57, 535–574.

Goldstein, Itay, Chester S. Spatt, and Mao Ye, 2021, Big data in finance, *Review of Financial Studies* 34, 3213–3225.

Gorton, Gary B., Jillian Grennan, and Alexander K. Zentefis, 2022, Corporate culture, *Annual Review of Financial Economics* 14, 535–561.

Graham, John R., Jillian Grennan, Campbell R. Harvey, and Shivaram Rajgopal, 2022a, Corporate culture: Evidence from the field, *Journal of Financial Economics* 146, 552–593.

Graham, John R., Jillian A. Grennan, Campbell R. Harvey, and Shivaram Rajgopal, 2022b, Corporate culture: The interview evidence, *Journal of Applied Corporate Finance* 34, 22–41.

Grennan, Jillian, 2019, A corporate culture channel: How increased shareholder governance reduces firm value, Duke University working paper.

Grennan, Jillian, and Kai Li, 2023, Corporate culture: A review and directions for future research, In *Handbook of Financial Decision Making*, edited by Giles Hilary and David McLean, United Kingdom: Edward Elgar Publishing Ltd., 112–132.

Guay, Wayne, 1999. The sensitivity of CEO wealth to equity risk: An analysis of the magnitude and determinants, *Journal of Financial Economics* 53, 43–71.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, 2015, The value of corporate culture, *Journal of Financial Economics* 117, 60–76.

Heindorf, Stefan et al., 2020, CauseNet: Towards a causality graph extracted from the Web, *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, October 2020, 3023–3030.

Huang, Allen H., Reuven Lehavy, Amy Y. Zang, and Rong Zheng, 2018, Analyst information discovery and interpretation roles: A topic modeling approach, *Management Science* 64, 2833–2855.

Huang, Allen H., Amy Y. Zang, and Rong Zheng, 2014, Evidence on the information content of text in analyst reports, *Accounting Review* 89, 2151–2180.

Huang, Allen H., Hui Wang, and Yi Yang, 2023, FinBERT—A large language model approach to extracting information from financial text, *Contemporary Accounting Research* 40, 806–841.

Jha, Manish, Jialin Qian, Michael Weber, and Baozhong Yang, 2023, ChatGPT and corporate policies, Georgia State University working paper.

Ji, Ziwei et al., 2023, Survey of hallucination in natural language generation, *ACM Computing Surveys* 55, 1–38.

Kecskés, Ambrus, Roni Michaely, and Kent L. Womack, 2017, Do earnings estimates add value to sell-side analysts' investment recommendations? *Management Science* 63, 1855–1871.

Kim, Alex G., Maximilian Muhn, and Valeri V. Nikolaev, 2023, Bloated disclosures: Can ChatGPT help investors process information? University of Chicago working paper.

Lewis, Patrick et al., 2020, Retrieval-augmented generation for knowledge-intensive NLP tasks, In *Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS'20*, Red Hook, NY, USA: Curran Associates Inc., 9459–9474.

Li, Edward Xuejun, Zhiyuan Tu, and Dexin Zhou, 2023, The promise and peril of generative AI: Evidence from ChatGPT as sell-side analysts, City University of New York working paper.

Li, Feng, 2010, The information content of forward-looking statements in corporate filings—A Naïve Bayesian machine learning approach, *Journal of Accounting Research* 48, 1049–1102.

Li, Hongyi, and Eric Van den Steen, 2021, Birds of a feather … enforce social norms? Interactions among culture, norms, and strategy, *Strategy Science* 6, 166–189.

Li, Kai, Xing Liu, Feng Mai, and Tengfei Zhang, 2021, The role of corporate culture in bad times: Evidence from the COVID-19 pandemic, *Journal of Financial and Quantitative Analysis* 56, 2545–2583.

Li, Kai, Feng Mai, Rui Shen, and Xinyan Yan, 2021, Measuring corporate culture using machine learning, *Review of Financial Studies* 34, 3265–3315.

Li, Xina, Guoli Chen, and Rui Shen, 2024, Steering the intangible wheel: CEO succession and corporate cultural change, *Organization Science* forthcoming.

Lin, Jianhua, 1991. Divergence measures based on the Shannon Entropy, *IEEE Transactions on Information Theory* 37, 145–151.

Liu, Nelson F. et al., 2023. Lost in the middle: How language models use long contexts, ArXiv preprint arXiv:2307.03172.

Liu, Xiaoding, 2016, Corruption culture and corporate misconduct, *Journal of Financial Economics* 122, 307–327.

Loh, Roger K., and René M. Stulz, 2011, When are analyst recommendation changes influential? *Review of Financial Studies* 24, 593–627.

Lopez-Lira, Alejandro, and Yuehua Tang, 2023, Can ChatGPT forecast stock price movements? Return predictability and large language models, University of Florida working paper.

Loughran, Tim, and Bill McDonald, 2011, When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks, *Journal of Finance* 66, 35–65.

Maynez, Joshua, Shashi Narayan, Bernd Bohnet, and Ryan McDonald, 2020, On faithfulness and factuality in abstractive summarization, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 1906–1919.

Mihăilă, Claudiu, Tomoko Ohta, Sampo Pyysalo, and Sophia Ananiadou, 2013, BioCause: Annotating and analysing causality in the biomedical domain, *BMC Bioinformatics* 14, 1–18.

Oh, Jong-Hoon et al., 2013, Why-question answering using intra-and inter-sentential causal relations, *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1733–1743.

Ouyang, Long et al., 2022, Training language models to follow instructions with human feedback, *Advances in Neural Information Processing Systems 35.*

Huang, H. Allen, Joseph Pacelli, Tianshuo Shi, and Yuan Zou, 2024, Communicating corporate culture in labor markets: Evidence from job postings, Harvard Business School working paper.

Previts, Gary John, Robert J. Bricker, Thomas R. Robinson, and Stephen J. Young, 1994. A content analysis of sell-side financial analyst company reports, *Accounting Horizons* 8, 55–70.

Radinsky, Kira, Sagie Davidovich, and Shaul Markovitch, 2012, Learning causality for news events prediction, *Proceedings of the 21st International Conference on World Wide Web*, 909–918.

Rathje, Steve, et al., 2024. GPT is an effective tool for multilingual psychological text analysis, *PNAS* 121, e2308950121.

Reimers, Nils, and Iryna Gurevych, 2019, Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 3982–3992.

Soltes, Eugene, 2014, Private interaction between firm management and sell-side analysts, *Journal of Accounting Research* 52, 245–272.

Tetlock, Paul C., 2007, Giving content to investor sentiment: The role of media in the stock market, *Journal of Finance* 62, 1139–1168.

Twedt, Brady, and Lynn Rees, 2012, Reading between the lines: An empirical examination of qualitative attributes of financial analysts' reports, *Journal of Accounting and Public Policy* 31, 1–21.

Vaswani, Ashish et al., 2017, Attention is all you need, *Advances in Neural Information Processing Systems* 30.

Wei, Jason et al., 2022a, Chain-of-thought prompting elicits reasoning in large language models, *Advances in Neural Information Processing Systems* 35, 24824–24837.

Wei, Jason et al., 2022b, Emergent abilities of large language models, *Transactions on Machine Learning Research*.

Womack, Kent L., 1996. Do brokerage analysts' recommendations have investment value? *Journal of Finance* 51, 137–167.

Ziems, Caleb, et al., 2024. Can large language models transform computational social science? *Computational Linguistics* 50, 237–291.

Zingales, Luigi, 2015, The "cultural revolution" in finance, *Journal of Financial Economics* 117, 1–4.

Figure 1
Flowchart of our information extraction method

This figure presents a flowchart that shows how our generative AI method extracts information about corporate culture from 2.4 million analyst reports over the period 2000–2020.

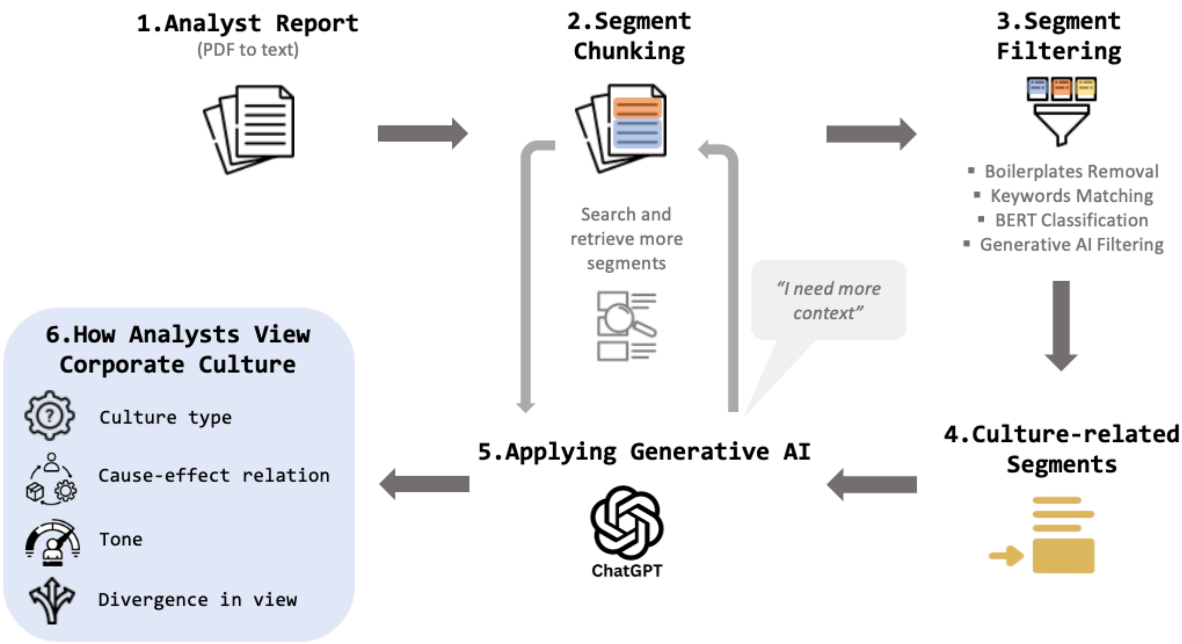


Figure 2
Intensity of analysts' culture-related discussions by year and report section

This heatmap depicts the intensity of analysts' culture-related discussions across years and report sections. Our sample comprises 2.4 million analyst reports over the period 2000–2020. The horizontal axis indicates report year, and the vertical axis indicates report section, binned into 20 equal sections from the start to the end of a report. The color gradient, ranging from light to dark, signifies the intensity of analysts' culture-related discussions. Intensity is computed as 100 × the number of culture-related segments in a section divided by the number of reports in a year.

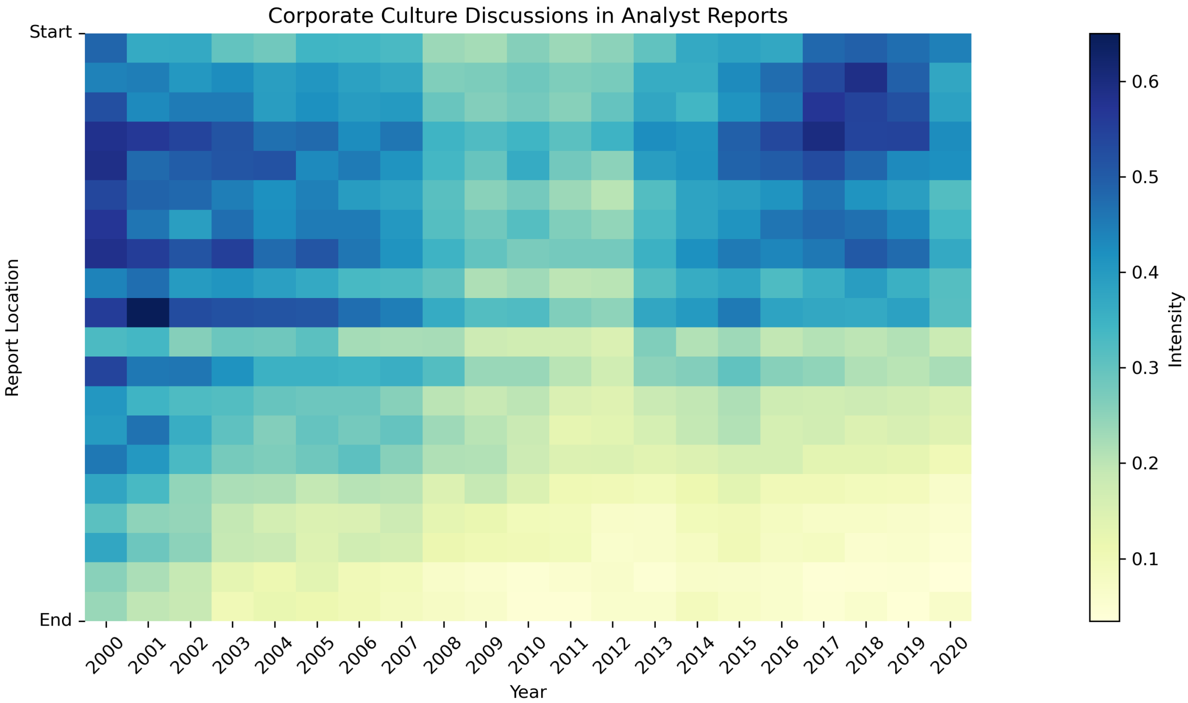


Figure 3
Cause-effect knowledge graph of corporate culture

This figure summarizes major cause-effect relations involving corporate culture extracted from analyst reports (omitting the miscellaneous category). Our sample comprises 2.4 million analyst reports over the period 2000–2020. In the left column, we group the 17 causes into three groups: events, people, and systems. In the center column, we list the five culture types. In the right column, we color-code the 16 business outcomes by tone. The height of each entity (cause, culture type, or effect) corresponds to the number of relevant segments. The width of each link corresponds to the number of segments mentioning a particular cause or effect relation.

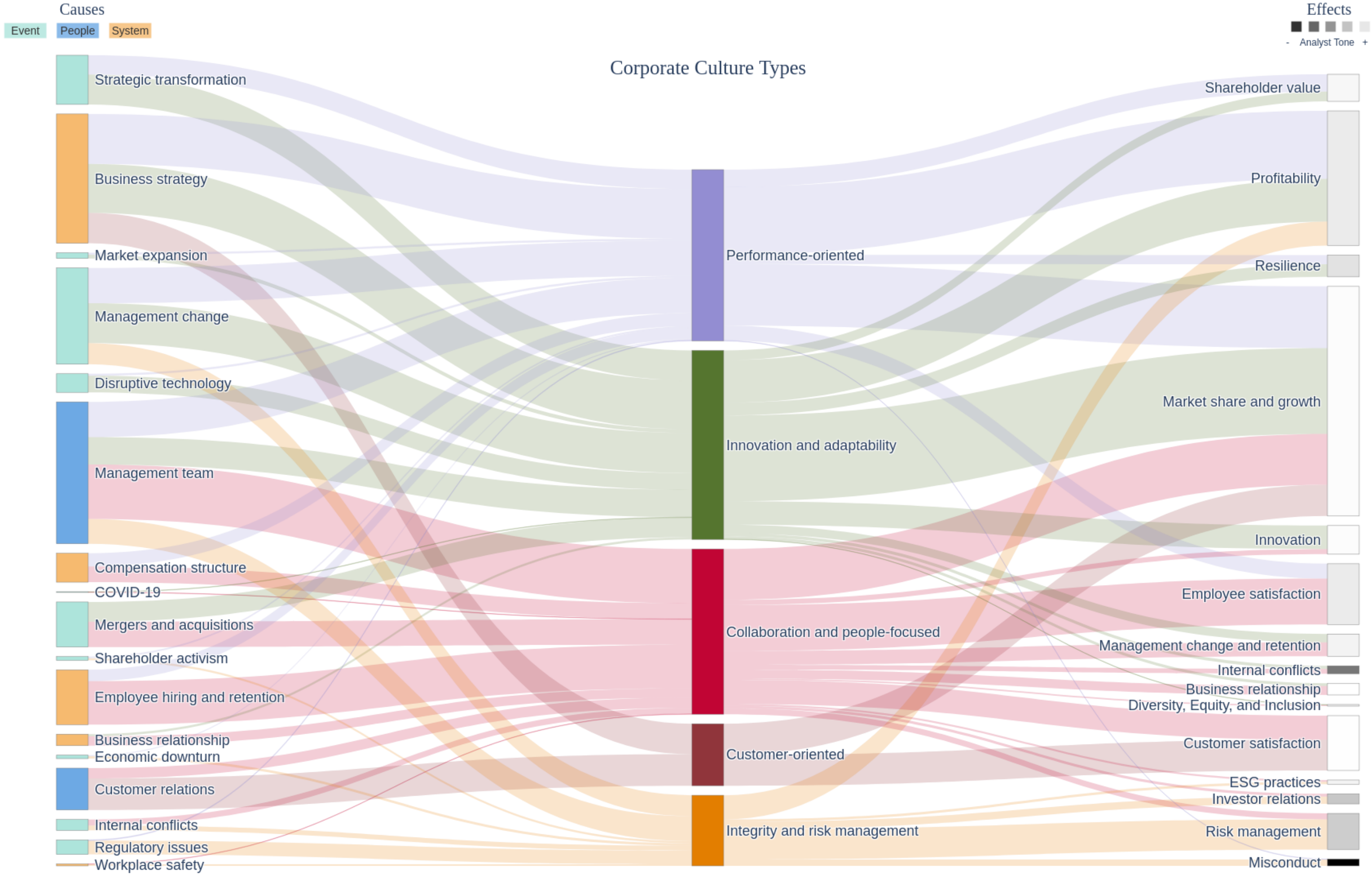
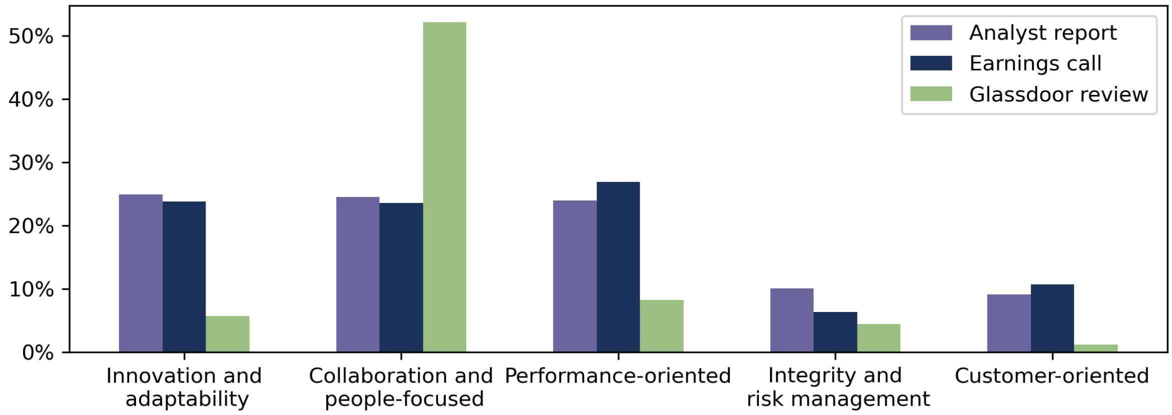


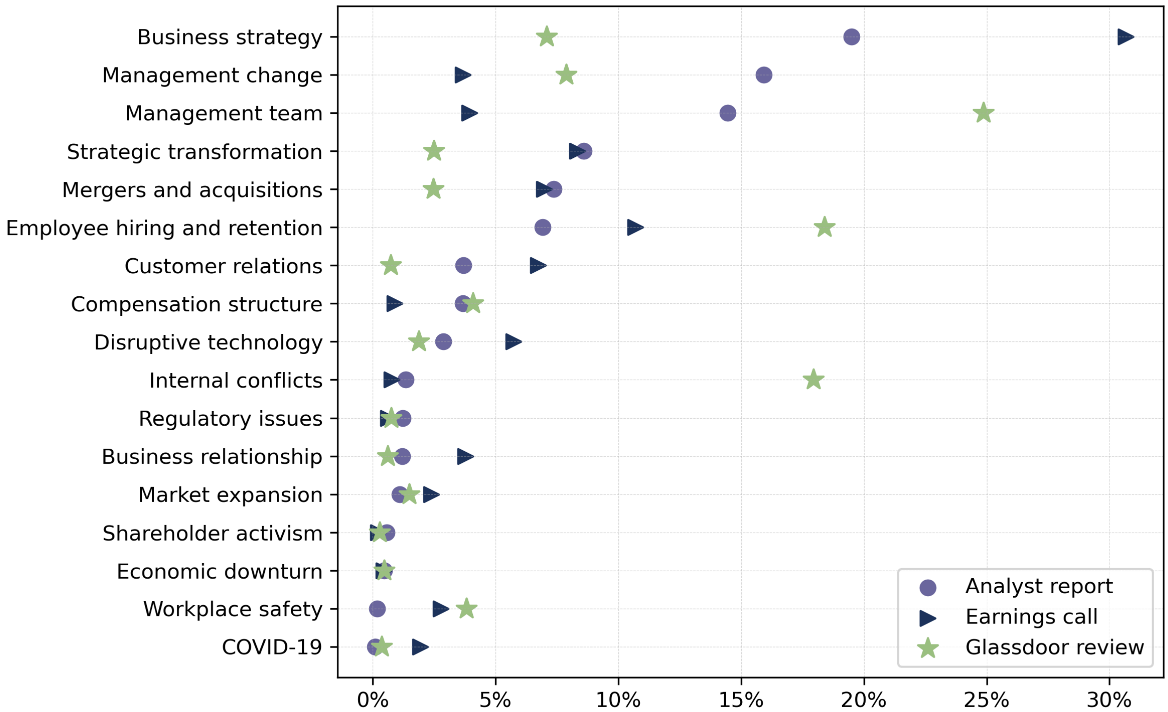
Figure 4
Insiders' versus outsiders' perspectives on corporate culture

This figure compares the frequency of different culture types, causes, and effects in analyst reports, earnings conference calls, and Glassdoor employee reviews. The Misc./Other categories are omitted. Panel A compares the frequency of the five culture types in different corpora. The horizontal axis lists different culture types, and the vertical axis indicates the percentage of segments from each corpus that mention a given culture type. Panels B and C compare the frequency of the 17 causes and 16 effects of corporate culture, respectively. The horizontal axis indicates the percentage of segments that mention a given cause (effect), and the vertical axis lists different causes (effects). Each corpus is represented by a different colored dot: purple circles for analyst reports, dark blue triangles for earnings calls, and green stars for Glassdoor employee reviews.

Panel A: Insiders' versus outsiders' perspectives on culture types



Panel B: Insiders' versus outsiders' perspectives on causes of cultural change



Panel C: Insiders' versus outsiders' perspectives on effects of culture

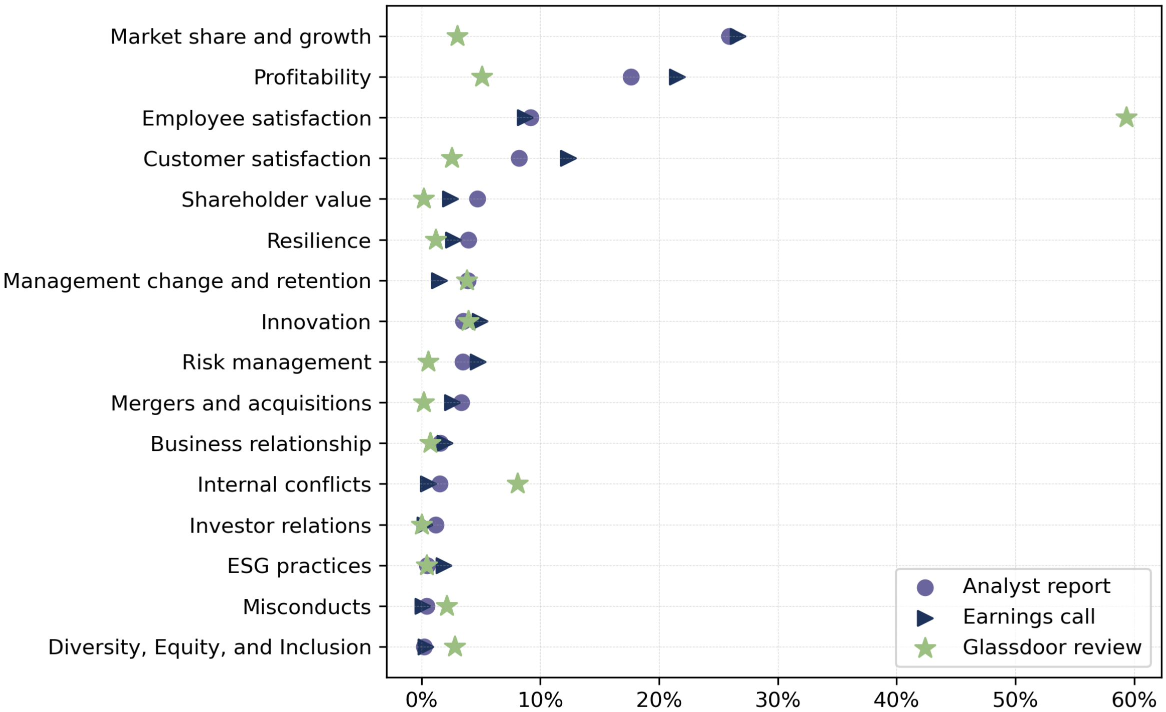


Table 1
Prompts for generative AI models

This table presents detailed instructions given to generative AI models to analyze and canonicalize information from corporate culture-related segments in analyst reports. Panel A shows the main prompt outlining the step-by-step process that generative AI models follow to extract information from each culture-related segment. It starts with identifying a culture type (e.g., innovation and adaptability), followed by extracting information on its cause, effect, and tone. The final step involves generating cause-effect triples with their representation in a standard digital format called JSON (JavaScript Object Notation). When more context is asked by generative AI models and is provided through retrieval augmented generation (RAG), the underlined sections in Panel A are omitted. Panel B shows the prompt to canonicalize causes and effects of a culture type.

Panel A: Chain-of-thought (CoT) prompt

```
As an expert specializing in corporate culture and causal reasoning, your task is to analyze
segments from analyst reports about corporate culture. Your goal is to extract and interpret
information about a company's corporate culture and identify cause-effect relationships. Present
your findings in a structured JSON format. Let's think step by step.

Step-by-Step Instructions:

1. Summarize the Corporate Culture (Q1):
   - Task: Determine the specific corporate culture being discussed in the segment.
   - Action: Summarize it in a short phrase starting with an adjective. If corporate culture is not
explicitly mentioned, infer it from the context.
   - Note: Avoid using generic adjectives such as strong/weak or positive/negative culture. If more
context from the report is needed for the analysis, output "I need more context" in the relevant
JSON field.

2. Classify the Corporate Culture (Q2):
   - Task: Categorize the identified corporate culture into one of the following six types:
     * Collaboration and People-Focused: Focusing on (or deficient in) collaboration, cooperation,
teamwork, supportive, low levels of conflict, community, communication within an organization,
employee well-being, employee equity sharing and compensation, diversity, inclusion, empowerment, or
talent.
     * Customer-Oriented: Focusing on (or deficient in) sales, customer, customer service, listening
to the customer, customer retention, customer experience, customer satisfaction, user experience,
client service, being brand-driven, quality of product, quality of service, quality of solution, or
taking pride in service.
     * Innovation and Adaptability: Focusing on (or deficient in) innovation, creativity,
technology, entrepreneurship, adaptability, transformations, flexibility, agility, willingness to
experiment, beyond tradition, disruption, fast-moving, quick to take advantage of opportunities,
resilience to change, or taking initiative.
     * Integrity and Risk Management: Focusing on (or deficient in) integrity, high ethical
standards, being honest, being transparent, accountability, do the right thing, fair practices,
being trustworthy, risk management, risk control, compliance, discipline, or financial prudence.
     * Performance-Oriented: Focusing on (or deficient in) high expectations for performance, sales
growth, achievement, competitiveness, results, hard work, efficiency, productivity, consistency in
executing tasks, setting clear goals, following best practices, striving for operational excellence,
or exceeding benchmarks.
     * Miscellaneous: Non-specific corporate culture, or corporate culture that does not easily fit
into the above types.  For example, "strong culture", "weak culture", "positive culture", "negative
culture", or "cultural change" (without details on the company's culture).

3. Identify Detailed Causal Analysis (Q3):
   - Task: Determine if the segment contains a detailed causal analysis of corporate culture.
   - Criteria: Look for explicit causal reasoning statements with trigger words like affect, cause,
influence, lead to, result in, fosters, driven by.
   - Action:
     - If YES, provide a short reason demonstrating the in-depth analysis.
     - If NO, the answers for Questions 4 and/or 5 should be an empty array [].
     - If more context is needed, output "I need more context" in the relevant JSON field.

4. Identify Causes of Corporate Culture (Q4):
   - Task: If a detailed causal analysis exists, identify explicitly mentioned events or factors
that have shaped, changed, or will change the corporate culture.
   - Action: List the most important causes or return an empty array [] if not applicable.
   - Note: Do not list other corporate culture types as causes. Focus on specific people, systems,
or events. Avoid implicit or indirect causes.

5. Identify Outcomes from Corporate Culture (Q5):
```

```
   - Task: If a detailed causal analysis exists, identify explicitly mentioned past, present, or
future outcomes or impacts of the corporate culture on the company.
   - Action: List the most important outcomes or return an empty array [] if not applicable.
   - Note: Do not list other corporate culture types as outcomes. Focus on specific results,
business outcomes, or tangible impacts. Avoid implicit or indirect outcomes.

6. Determine the Tone (Q6):
   - Task: Assess the tone of the discussion about corporate culture.
   - Options: "positive", "negative", "neutral".
   - Note: If the tone is unclear, mark it as "neutral".

7. Extract Causal Graph Triples (Q7):
   - Task: Based on the answers from Q1 (the specific corporate culture), Q4 (causes of corporate
culture), and Q5 (outcomes from corporate culture), extract causal graph triples related to that
specific culture.
   - Format: For each triple, provide:
     - Triple: ["entity_1", "relation", "entity_2"]
     - Explanation: A brief reason citing specific words or phrases or logical reasoning.
   - Criteria:
     - "entity_1" or "entity_2": Must be the specific corporate culture identified from Q1.
     - "relation": A clear and simple verb phrase conveying the cause-effect direction.
     - The Other Entity: Should be a cause (people, systems, or events) or outcome (result, impact)
for the specific corporate culture, not another corporate culture.
     - Avoid: Both entities being corporate culture.

JSON Output Structure:

{
    "all_results": [
        {
            "input_id": "XXXX",
            "identified_corporate_culture": "adjective + specific corporate culture" or "I need more
context",
            "corporate_culture_type": "one of the six types" or "I need more context",
            "detailed_causal_analysis": "YES" or "NO" or "I need more context",
            "causes_of_culture": ["cause_1", "..."] or [],
            "outcomes_from_culture": ["outcome_1", "..."] or [],
            "tone": "positive" / "negative" / "neutral",
            "causal_graph_triples": [
                {
                    "triple": ["entity_1", "relation", "entity_2"],
                    "explanation": "Reason for identifying the causal relation, citing specific
words or phrases or logical reasoning."
                }
                // ... additional triples
            ] or []
        },
        {
            "input_id": "YYYY",
            "identified_corporate_culture": "adjective + specific corporate culture" or "I need more
context",
            "corproate_culture_type": "one of the six types" or "I need more context",
            "detailed_causal_analysis": "YES" or "NO",
            "causes_of_culture": ["cause_1", "..."] or [],
            "outcomes_from_culture": ["outcome_1", "..."] or [],
            "tone": "positive" / "negative" / "neutral",
            "causal_graph_triples": [
                {
                    "triple": ["entity_1", "relation", "entity_2"],
                    "explanation": "Reason for identifying the causal relation, citing specific
words or phrases or logical reasoning."
                }
                // ... additional triples
            ] or []
        }
        // ... other inputs
    ]
}
```

Panel B: Prompt to canonicalize causes and effects of a culture type

```
As an expert specializing in corporate culture and causal reasoning, conduct entity canonicalization
on selected phrases taken from analyst reports. These phrases are considered to be *causes
(consequences) of corporate culture*. Your objective is to map each phrase to the most appropriate
standardized entity name from the predefined list below.

Instructions:
 - Review the provided examples for each standardized entity to grasp their scope and nuances.
 - Carefully analyze each phrase and assign it to the most fitting standardized entity based on its
underlying meaning.
 - Assign the phrase to "Other" only if it does not align with any of the predefined entities.
Before assigning a phrase to "Other", ensure that it does not reasonably fit any existing
categories.

The available standardized entity names and examples are:

----
[Categories and examples of causes (see Table IA4 for details)]
----

JSON Output Structure:

{
    "all_results": [
        {
            "input_phrase_id": 1,
            "input_phrase": "",
            "canonical_entity": "One of the predefined standardized entity names or 'Other'"
        },
        {
            "input_phrase_id": 2,
            "input_phrase": "",
            "canonical_entity": "One of the predefined standardized entity names or 'Other'"
        }
        ...
    ]
}
```

Table 2
Performance evaluation of different generative AI models

This table presents our performance evaluation of different generative AI models in terms of extracting culture types, causes, effects, and tones. Tone refers to the tone (negative, neutral, or positive) of culture-related information. We compare the performance of GPT-4o-mini, Claude-3.5-Haiku, and Gemini-Flash-1.5 in terms of accuracy, precision, recall, and F1 against human annotations of a randomly chosen set of 200 culture-related segments. We group our fact-checking into four scenarios. True positive denotes a scenario in which generative AI extracts similar information about cultural values/causes/effects relations as we do. False positive denotes a scenario in which generative AI extracts different (false) information from what we do. True negative denotes a scenario in which generative AI extracts no information, and neither do we. False negative denotes a scenario in which generative AI extracts false information while we do not find relevant information. We compute four performance metrics. Accuracy is defined as (#True Positive + #True Negative)/(#True Positive + #False Positive + #False Negative + #True Negative), and measures how accurate a model is at correctly classifying culture-related information out of the 200 segments. Precision is defined as (#True Positive)/(#True Positive + #False Positive) and measures how accurate a model is at identifying correct (positive) culture-related information out of all culture-related information that is predicted to be positive. Recall is defined as (#True Positive)/(#True Positive + #False Negative), and measures how accurate a model is at identifying correct (positive) culture-related information out of all identified culture-related information. F1 is the harmonic mean of Precision and Recall.

Performance metric	AI model	Culture type	Cause	Effect	Tone
Accuracy	GPT-4o-mini	98.5%	84.0%	91.0%	97.5%
	Claude-3.5-Haiku	93.0%	71.5%	77.0%	96.5%
	Gemini-Flash-1.5	94.0%	75.5%	78.5%	95.0%
Precision	GPT-4o-mini	98.5%	81.0%	98.4%	97.5%
	Claude-3.5-Haiku	93.0%	77.4%	76.7%	96.5%
	Gemini-Flash-1.5	94.0%	76.9%	84.4%	95.5%
Recall	GPT-4o-mini	100.0%	79.0%	88.5%	100.0%
	Claude-3.5-Haiku	100.0%	52.7%	91.2%	100.0%
	Gemini-Flash-1.5	100.0%	71.4%	83.8%	99.5%
F1	GPT-4o-mini	99.2%	80..0%	93.2%	98.7%
	Claude-3.5-Haiku	96.4%	62.7%	83.3%	98.2%
	Gemini-Flash-1.5	96.9%	74.1%	84.1%	97.4%

Table 3
Sample overview

This table reports firm-year observations with different coverage of culture types, causes, and effects. Our analyst report sample comprises 2.4 million reports over the period 2000–2020. Our earnings call sample comprises 243 thousand calls over the period 2004–2020. Our Glassdoor review sample comprises 5.3 million reviews over the period 2008–2020. Panel A presents firm-year observations with different coverage of corporate culture in analyst reports, earnings calls, and Glassdoor reviews. Panel B presents firm-year observations with divergence measures.

Panel A: Culture coverage in different corpora

Analyst reports	#report	#firm-year	#firm
Raw data	2,451,766	41,672	3,079
with culture	86,112	21,242	2,747
with cause	33,512	13,491	2,385
with effect	53,455	16,694	2,553
with either cause or effect	54,018	16,745	2,554
with both cause and effect	32,949	13,415	2,384
Earnings calls	**#call**	**#firm-year**	**#firm**
Raw data	243,501	72,749	12,006
with culture	101,533	48,192	10,238
with cause	58,535	34,485	8,835
with effect	80,930	41,239	9,551
with either cause or effect	82,008	41,514	9,576
with both cause and effect	57,170	34,016	8,783
Glassdoor reviews	**#review**	**#firm-year**	**#firm**
Raw data	5,343,864	41,969	5,187
with culture	866,796	29,081	4,485
with cause	81,154	15,998	3,240
with effect	151,689	18,959	3,508
with either cause or effect	157,680	19,153	3,527
with both cause and effect	75,163	15,576	3,200

Panel B: Samples with divergence measures

	Analyst-Executive	Analyst-Employee
#firm-year with culture in reports	21,242	21,242
#firm-year after merging with culture in calls (reviews)	10,372	6,335
#firm after merging with culture in calls (reviews)	1,844	1,010

Table 4
Determinants of analysts' discussing corporate culture in reports

This table examines the determinants of analysts' discussing corporate culture in their reports. Panel A examines the relationships between firm characteristics and analysts' discussing culture at the firm-year level. Our firm-year sample consists of 24,250 firm-year observations, representing 2,318 unique firms over the period 2002-2020. The dependent variable, *Culture discussion*, is an indicator variable that takes the value of one if corporate culture is discussed in analyst reports in a year, and zero otherwise. Panel B examines the relationships between firm characteristics and analysts' discussing a specific culture type at the firm-year level. Panel C examines the relationships between analyst characteristics and their discussing culture at the firm-analyst-year level. Our firm-analyst-year sample consists of 160,332 firm-analyst-year observations (smaller sample in regressions due to fixed effects), representing 2,471 unique firms followed by 4,096 analysts. Industry fixed effects are based on Fama-French 12-industry classifications. Definitions of the variables are provided in the Appendix. Heteroskedasticity-consistent standard errors (in parentheses) are clustered at the firm level. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A: Firm characteristics and analysts' discussing corporate culture in their reports

	Culture discussion					
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Firm size	0.095***	0.091***	0.049***	0.100***	0.103***	0.095***
	(0.004)	(0.004)	(0.007)	(0.009)	(0.011)	(0.020)
Ln(Firm age + 1)	-0.037***	-0.029***	-0.031**	-0.056**	-0.064*	-0.034
	(0.008)	(0.009)	(0.013)	(0.027)	(0.033)	(0.065)
Sales growth	0.037***	0.041**	0.126***	0.041***	0.045***	0.073**
	(0.014)	(0.016)	(0.028)	(0.014)	(0.016)	(0.030)
ROA	0.399***	0.382***	0.219***	0.159***	0.126**	0.012
	(0.047)	(0.051)	(0.079)	(0.044)	(0.050)	(0.084)
Leverage	-0.137***	-0.112***	-0.041	-0.091***	-0.077**	-0.077
	(0.026)	(0.028)	(0.040)	(0.032)	(0.036)	(0.061)
Tangibility	-0.073**	-0.065**	-0.066	-0.019	-0.042	0.248*
	(0.030)	(0.031)	(0.042)	(0.058)	(0.065)	(0.128)
ROA volatility	-0.162**	-0.189**	-0.367***	-0.053	-0.047	-0.032
	(0.067)	(0.075)	(0.132)	(0.067)	(0.077)	(0.146)
Large institution ownership	-0.040**	-0.034**	-0.067***	-0.005	-0.006	-0.057*
	(0.016)	(0.017)	(0.026)	(0.017)	(0.019)	(0.031)
Board independence	-0.158***	-0.177***	-0.313***	-0.042	-0.078*	-0.176**
	(0.038)	(0.042)	(0.065)	(0.041)	(0.047)	(0.075)
CEO duality	0.004	0.004	-0.009	-0.020**	-0.031***	-0.052***
	(0.009)	(0.009)	(0.014)	(0.009)	(0.010)	(0.017)
Ln(Number of key people changes + 1)	0.066***	0.059***	0.053***	0.026***	0.023***	0.022**
	(0.005)	(0.006)	(0.009)	(0.005)	(0.005)	(0.009)
Ln(Number of M&As + 1)	0.024***	0.018**	-0.002	0.001	0.001	-0.004
	(0.007)	(0.007)	(0.011)	(0.006)	(0.007)	(0.011)
Strong culture		0.106***	0.064***		0.011	0.009
		(0.011)	(0.016)		(0.010)	(0.017)
Ln(Number of meetings + 1)		0.036***	0.018*		0.007	-0.012
		(0.008)	(0.010)		(0.008)	(0.011)
Employee culture rating			0.003			0.004
			(0.008)			(0.007)

Ln(Number of employee reviews + 1)			0.049***			-0.006
			(0.006)			(0.009)
Constant	YES	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	NO	NO	NO
Firm FE	NO	NO	NO	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Adjusted R^2	0.169	0.176	0.181	0.337	0.343	0.367
Observations	29,385	24,250	9,371	29,385	24,250	9,371

Panel B: Firm characteristics and analysts' discussing a specific culture type in their reports

	Collaboration and People-focused	Customer-oriented	Innovation and adaptability	Integrity and risk management	Performance-oriented	Miscellaneous
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Firm size	0.055***	0.032***	0.063***	0.050***	0.067***	0.038***
	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)
Ln(Firm age + 1)	-0.020***	-0.016***	-0.019***	-0.002	0.002	-0.007
	(0.008)	(0.006)	(0.007)	(0.005)	(0.008)	(0.005)
Sales growth	0.052***	0.014	0.017	0.014	0.011	0.002
	(0.013)	(0.009)	(0.014)	(0.010)	(0.013)	(0.009)
ROA	0.250***	0.180***	0.326***	0.026	0.189***	0.057**
	(0.039)	(0.030)	(0.041)	(0.026)	(0.040)	(0.026)
Leverage	-0.067***	-0.050***	-0.091***	-0.092***	-0.038	-0.042***
	(0.024)	(0.017)	(0.022)	(0.016)	(0.024)	(0.014)
Tangibility	-0.014	0.084***	-0.110***	-0.005	-0.022	0.013
	(0.027)	(0.022)	(0.025)	(0.016)	(0.027)	(0.015)
ROA volatility	-0.254***	-0.014	0.001	0.044	-0.124**	0.049
	(0.053)	(0.037)	(0.059)	(0.033)	(0.055)	(0.036)
Large institution ownership	-0.043***	-0.027***	-0.040***	-0.031***	-0.044***	-0.041***
	(0.014)	(0.010)	(0.014)	(0.009)	(0.014)	(0.009)
Board independence	-0.135***	-0.060**	-0.170***	-0.086***	-0.168***	-0.057**
	(0.037)	(0.026)	(0.034)	(0.025)	(0.037)	(0.023)
CEO duality	0.013	0.011*	0.002	0.011*	-0.001	0.001
	(0.008)	(0.006)	(0.008)	(0.006)	(0.008)	(0.005)
Ln(Number of key people changes + 1)	0.039***	0.019***	0.049***	0.019***	0.049***	0.026***
	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)
Ln(Number of M&As + 1)	0.017**	-0.002	0.024***	0.012**	0.023***	0.024***
	(0.007)	(0.005)	(0.007)	(0.006)	(0.008)	(0.006)
Strong culture	0.121***	0.081***	0.111***	0.021***	0.069***	0.046***
	(0.011)	(0.008)	(0.010)	(0.007)	(0.010)	(0.007)
Ln(Number of meetings + 1)	0.025***	0.009	0.049***	0.006	0.027***	0.014***
	(0.008)	(0.006)	(0.007)	(0.006)	(0.008)	(0.005)
Constant	YES	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Adjusted R^2	0.124	0.098	0.142	0.106	0.124	0.069

Observations	24,250	24,250	24,250	24,250	24,250	24,250

Panel C: Analyst characteristics and their discussing corporate culture in reports

Variable	Culture discussion (1)	Number of types (2)	Number of causes (3)	Number of effects (4)	Tone (5)
Star analyst	0.027***	0.094***	-0.009	-0.021	-0.015
	(0.005)	(0.026)	(0.018)	(0.030)	(0.021)
CFA	0.003*	0.019*	-0.001	0.003	-0.005
	(0.002)	(0.010)	(0.006)	(0.011)	(0.008)
Postgraduate	0.012***	0.011	-0.008	0.008	0.004
	(0.002)	(0.009)	(0.008)	(0.013)	(0.009)
Female	0.017***	0.022	0.007	-0.009	0.023
	(0.004)	(0.019)	(0.014)	(0.024)	(0.017)
General experience	0.001***	0.002*	0.001	0.000	-0.000
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Firm experience	-0.000	0.005***	-0.001	-0.002	0.000
	(0.000)	(0.002)	(0.001)	(0.002)	(0.002)
Number of industries followed	-0.001	-0.000	0.003	-0.002	0.009***
	(0.001)	(0.003)	(0.003)	(0.005)	(0.003)
Number of firms followed	-0.001***	0.000	-0.002*	-0.002	-0.002**
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Forecast frequency	0.007***	0.021***	-0.001	-0.002	-0.002
	(0.001)	(0.003)	(0.002)	(0.004)	(0.003)
Ln(Broker size)	0.009*	0.024	0.009	0.030	0.023
	(0.005)	(0.031)	(0.027)	(0.045)	(0.030)
Ln(Number of meetings + 1)	-0.001	0.000	-0.005	-0.000	0.001
	(0.001)	(0.006)	(0.004)	(0.005)	(0.004)
Constant	YES	YES	YES	YES	YES
Firm × Year FE	YES	YES	YES	YES	YES
Broker FE	YES	YES	YES	YES	YES
Adjusted R^2	0.154	0.058	0.046	0.047	0.164
No. of observations	156,598	20,150	20,150	20,150	20,150

Table 5
Determinants of divergence in outsiders' and insiders' perspectives on culture

This table examines the determinants of divergence in outsiders' and insiders' perspectives on culture. Panel A examines the relationships between firm characteristics and divergences between analysts and executives (employees) about corporate culture. Our analyst-executive firm-year sample consists of 8,369 firm-year observations, representing 1,581 unique firms over the period 2004-2020. Our analyst-employee firm-year sample consists of 5,216 firm-year observations, representing 909 unique firms over the period 2008-2020. The dependent variable, *Type divergence*, is the JS divergence regarding culture types between outsiders (analyst reports) and insiders (executives' discussions in earnings calls or employees' reviews from Glassdoor). *Analyst-Executive* represents the divergence between analysts and executives. *Analyst-Employee* represents the divergence between analysts and employees. Cause divergence and Effect divergence are defined analogously. Panel B examines the relationships between firm characteristics and the difference in analysts' and executives' perspectives on a specific culture type and tone. The dependent variable, *Type difference*, is the difference in the frequency of the discussion of a specific culture type (tone) by analysts and executives. Panel C examines the relationships between firm characteristics and the difference in analysts' and employees' perspectives on a specific culture type and tone. Industry fixed effects are based on Fama-French 12-industry classifications. Definitions of the variables are provided in the Appendix. Heteroskedasticity-consistent standard errors (in parentheses) are clustered at the firm level. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A: Firm characteristics and divergences in viewing culture between analysts and executives/employees.

	Analyst-Executive	Analyst-Employee	Analyst-Executive	Analyst-Employee	Analyst-Executive	Analyst-Employee
	Type divergence	Type divergence	Cause divergence	Cause divergence	Effect divergence	Effect divergence
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Firm size	-0.006**	-0.025***	-0.007**	-0.021***	-0.003	-0.015***
	(0.003)	(0.004)	(0.003)	(0.003)	(0.002)	(0.003)
Ln(Firm age + 1)	-0.003	0.015**	0.003	0.000	-0.002	0.004
	(0.006)	(0.007)	(0.005)	(0.005)	(0.005)	(0.005)
Sales growth	0.023	-0.012	0.000	0.003	-0.011	-0.002
	(0.015)	(0.018)	(0.012)	(0.014)	(0.011)	(0.012)
ROA	0.041	0.032	-0.024	0.003	-0.040	-0.017
	(0.060)	(0.062)	(0.049)	(0.052)	(0.042)	(0.043)
Leverage	-0.013	0.022	0.004	0.011	-0.004	0.032**
	(0.020)	(0.024)	(0.016)	(0.016)	(0.015)	(0.015)
Tangibility	-0.043**	0.041*	-0.021	0.052***	0.003	0.030*
	(0.020)	(0.021)	(0.018)	(0.017)	(0.017)	(0.017)
ROA volatility	0.154*	-0.038	0.179***	-0.115	0.208***	-0.039
	(0.091)	(0.087)	(0.057)	(0.071)	(0.054)	(0.057)
Large institution ownership	0.002	0.032**	0.011	-0.004	0.012	0.004
	(0.014)	(0.015)	(0.011)	(0.011)	(0.010)	(0.010)
Board independence	0.017	0.041	0.029	0.078***	0.062**	0.059**
	(0.031)	(0.035)	(0.027)	(0.027)	(0.026)	(0.026)
Loss year	0.007	-0.013	-0.004	-0.010	0.000	-0.015*
	(0.011)	(0.013)	(0.009)	(0.011)	(0.008)	(0.009)
CEO duality	0.003	0.009	-0.001	0.005	-0.004	0.005
	(0.007)	(0.008)	(0.006)	(0.007)	(0.006)	(0.006)
CEO tenure	0.001*	-0.001	0.001	0.000	0.002***	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
CEO Delta	0.004	0.003	0.000	-0.005*	-0.000	0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)

CEO Vega	-0.005***	0.001	-0.004***	0.000	-0.004***	-0.000
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
CEO close-to-retire	-0.007	-0.020**	-0.004	-0.001	0.012**	0.011*
	(0.008)	(0.010)	(0.007)	(0.007)	(0.006)	(0.006)
Ln(Number of meetings + 1)	-0.005	-0.003	-0.001	-0.001	-0.007*	-0.005
	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Constant	YES	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Adjusted R^2	0.033	0.054	0.046	0.118	0.090	0.081
Observations	8,369	5,216	7,524	4,441	8,654	5,073

Panel B: Firm characteristics and divergences in viewing a specific culture type or tone between analysts and executives

	Analyst-Executive						
	Collaboration and People-focused	Customer-oriented	Innovation and adaptability	Integrity and risk management	Performance-oriented	Misc.	Tone
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Firm size	0.032***	0.015**	0.042***	0.026***	0.020***	0.049***	-0.025***
	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.006)
Ln(Firm age + 1)	-0.017	-0.020	-0.000	-0.003	0.002	-0.018	0.013
	(0.013)	(0.013)	(0.013)	(0.011)	(0.013)	(0.012)	(0.013)
Sales growth	-0.045	-0.011	0.039	0.018	-0.073***	-0.041	0.014
	(0.028)	(0.027)	(0.031)	(0.024)	(0.027)	(0.027)	(0.030)
ROA	0.419***	0.086	0.235**	0.066	0.295***	0.069	0.153
	(0.110)	(0.099)	(0.106)	(0.089)	(0.112)	(0.098)	(0.103)
Leverage	0.058	0.021	-0.025	0.002	-0.008	-0.007	-0.016
	(0.040)	(0.038)	(0.044)	(0.035)	(0.039)	(0.038)	(0.039)
Tangibility	-0.118***	-0.025	-0.040	-0.123***	-0.060	-0.050	0.027
	(0.043)	(0.045)	(0.045)	(0.043)	(0.041)	(0.040)	(0.041)
ROA volatility	0.208	0.149	0.082	0.267**	0.224	0.429***	-0.353**
	(0.143)	(0.108)	(0.151)	(0.117)	(0.166)	(0.133)	(0.141)
Large institution ownership	-0.036	-0.011	0.019	-0.039*	-0.062***	-0.025	-0.014
	(0.026)	(0.023)	(0.025)	(0.021)	(0.024)	(0.023)	(0.025)
Board independence	-0.116*	-0.015	-0.121*	-0.154***	-0.249***	0.017	-0.124*
	(0.064)	(0.058)	(0.065)	(0.056)	(0.068)	(0.059)	(0.064)
Loss year	0.033	0.014	0.017	-0.010	0.045**	0.020	-0.070***
	(0.021)	(0.019)	(0.021)	(0.017)	(0.021)	(0.019)	(0.021)
CEO duality	0.001	-0.001	-0.002	-0.003	-0.010	-0.008	0.013
	(0.015)	(0.014)	(0.015)	(0.014)	(0.015)	(0.014)	(0.014)
CEO tenure	0.001	0.002	-0.000	0.001	0.001	-0.002	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
CEO Delta	0.014**	0.013*	0.003	0.013**	0.012*	-0.001	0.018**
	(0.007)	(0.007)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)
CEO Vega	0.001	-0.006	-0.003	-0.002	-0.001	-0.004	-0.006
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)

CEO close-to-retire	0.011	-0.012	0.030	-0.017	0.007	-0.002	0.016
	(0.017)	(0.016)	(0.018)	(0.013)	(0.017)	(0.015)	(0.017)
Ln(Number of meetings + 1)	0.033***	0.012	0.022*	0.004	0.029**	0.030***	0.024**
	(0.012)	(0.011)	(0.012)	(0.010)	(0.012)	(0.010)	(0.010)
Constant	YES	YES	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES
Adjusted R^2	0.133	0.067	0.098	0.059	0.108	0.080	0.104
Observations	8,369	8,369	8,369	8,369	8,369	8,369	8,369

Panel C: Firm characteristics and divergences in viewing a specific culture type or tone between analysts and employees

	Analyst-Employee						
	Collaboration and People-focused	Customer-oriented	Innovation and adaptability	Integrity and risk management	Performance-oriented	Misc.	Tone
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Firm size	-0.025**	-0.051***	-0.055***	-0.048***	-0.048***	-0.012	-0.024**
	(0.010)	(0.008)	(0.009)	(0.009)	(0.010)	(0.008)	(0.010)
Ln(Firm age + 1)	0.020	-0.015	-0.006	-0.023	0.025	-0.037**	0.031
	(0.019)	(0.014)	(0.020)	(0.019)	(0.020)	(0.016)	(0.021)
Sales growth	0.180***	0.033	0.045	0.091**	0.109**	0.094***	0.115**
	(0.041)	(0.029)	(0.041)	(0.044)	(0.043)	(0.035)	(0.049)
ROA	-0.170	-0.042	-0.123	-0.141	-0.267*	-0.272**	-0.321*
	(0.143)	(0.119)	(0.150)	(0.156)	(0.155)	(0.125)	(0.170)
Leverage	0.133**	0.125***	0.145**	0.078	0.133**	0.079*	0.149**
	(0.059)	(0.047)	(0.061)	(0.061)	(0.061)	(0.048)	(0.067)
Tangibility	-0.077	0.162***	0.018	0.121*	0.141**	0.054	-0.093
	(0.055)	(0.040)	(0.057)	(0.063)	(0.057)	(0.045)	(0.071)
ROA volatility	-0.694***	-0.070	-0.381**	-0.014	-0.316	-0.010	-0.593**
	(0.175)	(0.141)	(0.189)	(0.203)	(0.211)	(0.161)	(0.234)
Large institution ownership	-0.117***	0.013	-0.067*	-0.099**	-0.129***	-0.133***	0.009
	(0.038)	(0.029)	(0.037)	(0.039)	(0.040)	(0.032)	(0.046)
Board independence	-0.110	-0.009	0.009	-0.056	-0.032	-0.042	-0.131
	(0.085)	(0.067)	(0.089)	(0.094)	(0.101)	(0.076)	(0.106)
Loss year	-0.022	-0.020	-0.014	-0.030	-0.058*	-0.010	-0.070**
	(0.029)	(0.024)	(0.030)	(0.030)	(0.031)	(0.024)	(0.035)
CEO duality	-0.028	-0.004	-0.050**	0.001	-0.002	-0.013	0.004
	(0.020)	(0.017)	(0.021)	(0.023)	(0.022)	(0.018)	(0.026)
CEO tenure	-0.001	-0.000	-0.004**	-0.000	-0.002	-0.000	-0.004*
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)
CEO Delta	-0.011	0.001	0.006	-0.011	-0.026***	-0.017**	0.004
	(0.009)	(0.007)	(0.009)	(0.010)	(0.009)	(0.008)	(0.012)
CEO Vega	-0.002	0.005	-0.003	-0.007	-0.002	-0.009**	-0.000
	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)	(0.004)	(0.006)

CEO close-to-retire	0.081***	-0.004	0.034	0.027	0.033	0.055***	0.068***
	(0.023)	(0.019)	(0.026)	(0.023)	(0.024)	(0.018)	(0.025)
Ln(Number of meetings + 1)	-0.028*	-0.025**	-0.030*	-0.035**	-0.033**	-0.010	-0.029*
	(0.015)	(0.011)	(0.016)	(0.016)	(0.016)	(0.012)	(0.017)
Constant	YES	YES	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES
Adjusted R^2	0.155	0.155	0.146	0.156	0.158	0.178	0.061
Observations	5,216	5,216	5,216	5,216	5,216	5,216	5,216

Table 6
Analysts' perspectives on corporate culture and their research output

This table examines the relationships between analysts' perspectives on corporate culture and their stock recommendations and target prices at the report level. Panel A presents the summary statistics for the key variables. Panel B examines the relationships between analysts' tones in culture-related segments and their stock recommendations and target prices. The dependent variable in column (1), *Recommendation*, is a report's stock recommendation using a five-tier rating system where 2 represents "strong buy," 1 represents "buy," 0 represents "hold," -1 represents "underperform," and -2 represents "sell." The dependent variable in column (2), *Target price*, is a report's target price divided by the stock price 50 days before the report date (in percentage points). Industry fixed effects are based on Fama-French 12-industry classifications. Definitions of the variables are provided in the Appendix. Heteroskedasticity-consistent standard errors (in parentheses) are double-clustered at the firm and analyst levels. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A: Summary statistics for the key variables

	Mean	25th Percentile	Median	75th Percentile	SD
Stock recommendation sample					
Recommendation	0.734	0.000	1.000	1.000	0.853
Tone	0.504	0.000	1.000	1.000	0.681
Number of cause and effect	1.432	0.000	2.000	2.000	1.225
Target price sample					
Target price	1.173	1.045	1.167	1.286	0.218

Panel B: The tone in culture-related segments, stock recommendations, and target prices

	Recommendation	Recommendation	Target price	Target price
Variable	(1)	(2)	(3)	(4)
Tone	0.127***	0.102***	0.024***	0.019***
	(0.011)	(0.014)	(0.003)	(0.003)
Tone × Number of cause and effect		0.017***		0.004**
		(0.006)		(0.001)
Number of cause and effect		-0.012**		-0.003**
		(0.005)		(0.001)
Star analyst	-0.014	-0.014	-0.004	-0.004
	(0.037)	(0.037)	(0.008)	(0.008)
Female	0.010	0.011	-0.011	-0.011
	(0.044)	(0.044)	(0.008)	(0.008)
Forecast horizon	-0.099**	-0.099**	0.002	0.002
	(0.039)	(0.039)	(0.008)	(0.008)
General experience	-0.001	-0.001	-0.000	-0.000
	(0.003)	(0.003)	(0.000)	(0.000)
Firm experience	0.002	0.002	0.002**	0.002**
	(0.003)	(0.003)	(0.001)	(0.001)
Number of industries followed	0.031***	0.031***	0.001	0.001
	(0.008)	(0.008)	(0.002)	(0.002)
Number of firms followed	-0.004*	-0.004*	0.000	0.000
	(0.002)	(0.002)	(0.000)	(0.000)
Forecast frequency	-0.002	-0.002	-0.001	-0.001
	(0.005)	(0.005)	(0.001)	(0.001)
Ln(Broker size)	-0.175***	-0.176***	-0.014***	-0.014***
	(0.017)	(0.017)	(0.003)	(0.003)
Non-culture tone	0.562***	0.560***	0.108***	0.108***
	(0.029)	(0.029)	(0.008)	(0.008)

Ln(Report length)	0.000	0.001	-0.004	-0.004
	(0.013)	(0.013)	(0.003)	(0.003)
Constant	YES	YES	YES	YES
Firm × Year FE	YES	YES	YES	YES
Analyst FE	YES	YES	YES	YES
Adjusted R^2	0.457	0.457	0.423	0.423
No. of observations	28,880	28,880	29,146	29,146

Table 7
Information content of analysts' perspectives on corporate culture

This table examines the information content of analysts' perspectives on corporate culture at the report level. The sample comprises 13,223 reports that contain culture discussions and are not issued at the same time as any other major corporate announcements. Panel A presents the summary statistics for the key variables. Panel B presents the regression results. The dependent variable, *CAR[-3,+3]*, is the cumulative abnormal return (in percentage points) centered around the report date (day 0) based on a market model. Industry fixed effects are based on Fama-French 12-industry classifications. Definitions of the variables are provided in the Appendix. Heteroskedasticity-consistent standard errors (in parentheses) are double-clustered at the firm and analyst levels. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A: Summary statistics for the key variables

	Mean	25th Percentile	Median	75th Percentile	SD
CAR[-3,+3] (%)	0.193	-1.950	0.122	2.367	5.291
Tone	0.504	0.000	1.000	1.000	0.681
Non-culture tone	0.202	0.040	0.200	0.371	0.260
Tone divergence with other analysts	0.455	0.309	0.465	0.614	0.215

Panel B: Price reactions to analyst reports

	CAR[-3,+3]	CAR[-3,+3]
Variable	(1)	(2)
Tone	0.220***	-0.346
	(0.068)	(0.271)
Tone × Tone divergence with other analysts		1.093***
		(0.412)
Tone divergence with other analysts		-0.182
		(0.385)
Non-culture tone	1.855***	1.994***
	(0.181)	(0.187)
Ln(Report length)	0.053	0.040
	(0.064)	(0.064)
Earnings forecast revision	11.630*	11.458
	(6.828)	(7.247)
Recommendation revision	1.621***	1.615***
	(0.276)	(0.283)
Target price revision	0.149***	0.131***
	(0.036)	(0.037)
Prior CAR	-0.027***	-0.028**
	(0.010)	(0.011)
Other analyst/firm controls	YES	YES
Constant	YES	YES
Industry × Year FE	YES	YES
Adjusted R^2	0.037	0.040
No. of observations	13,223	12,665

Internet Appendix for

Dissecting Corporate Culture Using Generative AI

A. Technical Appendix

In this appendix, we describe technical details of how we preprocess analyst reports, remove boilerplate segments, identify culture-related segments via word sense disambiguation, a machine learning model, and ChatGPT, and implement canonicalization of culture types and causes/effects. Figure 1 provides a flowchart of our information extraction method built on generative AI models.

1. Converting Reports from PDF to Text

We download 2,434,782 reports over the period 2000–2020 from Thomson One's Investext database. The reports are in PDF format. We use GROBID (https://github.com/kermitt2/grobid), an open-source software, to extract structured information from PDF documents and transform this information into XML documents. The XML documents are then stripped of information identified as tables, annexes, notes, and author information; the main content is converted to plain text. We further split text into sentences using OpenNLP's sentence segment module, a built-in function in GROBID.

2. Segment Chunking

An inherent challenge resulting from the conversion process described above is the loss of paragraph structure in reports. Moreover, even if the structure could have been maintained, differentiating between headers, bullet points, and coherent paragraphs in a report is not straightforward.

To address these issues, we employ the C99 algorithm, a common text segmentation technique developed by Choi (2000). The C99 algorithm is a domain-independent, unsupervised method for linear text segmentation. Its defining principle is that topics in a text document are coherent, and topic shifts can be identified by a sharp decline in coherence. The algorithm quantifies coherence through pairwise similarity of sentences. It builds a matrix of cosine similarities between the TF-IDF representations of sentences in a document. Segmentation points are identified wherever there are sudden drops in the average similarity of sentences. The C99 algorithm enables us to coalesce individual sentences into larger, more meaningful segments that align more closely with coherent thoughts or ideas in the text. Consequently, we use these segmented units, rather than individual sentences, as the unit of analysis for our study.

3. Removing Boilerplate Segments

To identify and remove boilerplate segments in analyst reports, we employ a machine learning model specifically designed and trained for this purpose.

To construct the training data set for our model, we first identify the top 20 brokers producing the highest volume of reports each year. From each of those brokers in each sample year, we sample 1,000 of their reports. We then identify the top 10% most frequently repeated segments within those reports. For a segment to be classified as a positive example, it must satisfy two criteria: it is among the top 10% most frequently repeated segments, and it is repeated at least five times by the same broker within the same year.

Negative examples, or segments least likely to be boilerplates, are identified by randomly selecting 10 segments with no repetition in each broker-year sample. To ensure balance within our data set, we randomly sample from the remaining non-boilerplate segments to achieve a one-to-ten ratio of positive to negative examples. This results in a training data set of 547,790 examples, comprising

54,779 positive examples of boilerplate segments and 493,011 negative examples of non-boilerplate segments. The data set is split into training, validation, and testing sets, using an 80/10/10 ratio.

Our approach to identifying boilerplate segments in reports makes use of the SentenceTransformer model, specifically the all-mpnet-base-v2 variant. This model builds on an architecture similar to BERT, but focuses on creating high-quality sentence-level embeddings instead of token-level embeddings. This is particularly beneficial for our task as it views sentences and segments as distinct units of meaning, and thereby generates more effective and contextually relevant embeddings. To generate these embeddings, the SentenceTransformer model employs a mean pooling operation on the output of the transformer network, i.e., it creates a fixed-length sentence embedding by averaging all token embeddings. This operation gives us a representation of each sentence in a 768-dimensional vector space. For a segment containing multiple sentences, we compute the mean of all the sentence embeddings within that segment to yield a representative vector. This is an essential step because it allows us to convert segments of varying lengths into fixed-length representations, which can then be directly fed into our classification model. We find that this simple aggregation method performs well in capturing the overall semantic context of a segment.

BERT, with its bi-directional context understanding, is known to be effective for a broad range of NLP tasks (Devlin et al. 2018). A standard practice is to fine-tune the pre-trained BERT model on a specific task, which adjusts all the model parameters. This approach typically achieves high performance as it enables the pre-trained BERT model to learn from the specifics of the task, capitalizing on its general language understanding capabilities while adapting to task-specific nuances. However, given the context of our research, we adopt a different strategy that leverages both the representation power of the pre-trained BERT model and the efficiency of a classification head. Rather than fine-tuning and adjusting all parameters of the model, we freeze the parameters of the BERT model (i.e., the embeddings are fixed) and add a classification head to the model. The classification head takes embeddings from the BERT model as inputs and processes them with two hidden layers: the first layer contains 16 neurons, and the second layer contains 8 neurons. Each layer applies a Rectified Linear Unit (ReLU) activation to introduce non-linearity. Following the processing of the embeddings by the two layers, the resulting output vector is directed towards a softmax layer, which computes the probability of each segment as boilerplate or not.

The choice of the above strategy (architecture) is motivated by a number of considerations. First, our main generative AI model leverages retrieval augmented generation (RAG) that uses the all-mpnet-base-v2 embedding to help retrieve more context for culture-related information extraction. The chosen architecture allows us to maintain the consistency of embeddings across models. Second, we find that identifying boilerplate text is a relatively straightforward task that does not require the full-scale fine-tuning of the model, which would be computationally expensive and time-consuming. By freezing the parameters of the segment representation model and deploying only the classification head, we optimize computational efficiency and streamline the training process.

The trained classification model achieves good performance, with an Area Under the Curve (AUC) of 0.966 on the test set. The false positive rate is 0.093 and the false negative rate is 0.073.

Table IA1 in the Internet Appendix lists predicted boilerplate probabilities and boilerplate examples, sorted by decile. We retain segments with a boilerplate probability of 0.22 (the sample median) or lower.

4. Identifying Culture-related Segments

We identify segments related to corporate culture through a three-step procedure.

In step 1, we start with an exhaustive text search using two sets of keywords. The initial set of keywords is based on the word set explicitly about corporate culture, identifying a total of 5,541 relevant segments.[1] We also employ a second, more flexible set of keywords, which match all segments containing the word "culture(s)" or "cultural," excluding those already identified to avoid duplication. This second search results in a larger set of 46,795 segments. These segments contain potentially relevant mentions of corporate culture, although their meaning could be ambiguous. The word "culture(s)," and to a lesser extent the word "cultural," may refer to biological or social context.

To address these ambiguities, we employ generative AI for word sense disambiguation (WSD). Table IA2 Panel A shows the prompt. Our method matches the word "culture(s)" or "cultural" with one of the three definitions from dictionary.com.[2] The following examples illustrate our application of WSD:

> 1. "*Organization structure, talent model and deep bench of UNH make for a strong competitive advantage. The passion for excellence, humility, restlessness and desire to win is a culture that keeps UNH at the top of its game, and we expect will make it hard for fast-followers in the Large Cap MCO space and new Big Tech entrants such as AMZN and AAPL to catch up.*" -> Organizational
>
> 2. "*Demand for specialty proteins, probiotics, and cultures supported pricing gains. Continued strength in demand should contribute to a 1% and 4% YoY increase in sales for 1Q14 and 2014, respectively.*" -> Biological
>
> 3. "*Cultural hurdles more relative than price. There's no question that web conferencing is significantly less expensive than face-to-face meetings that require corporate travel (although it doesn't always replace a face-to-face meeting). To our knowledge, no one is questioning the value proposition of web conferencing.*" -> Societal

We exclude segments in which the discussion of culture is classified as in a biological or societal context, resulting in a final set of 41,038 segments.

It is possible that there are segments about corporate culture without mentioning explicit words or phrases. Consider the following segment "*One word we have heard from BBY's management team, a word that has led to the highest service levels at retailers and often the most successful ones is empowerment. From Wal-Mart in its heyday to Home Depot to Costco to Bed Bath and Beyond, empowering employees has been a critical element to success among retailers*." This segment, while not mentioning 'culture,' a human reader will conclude that it discusses a key aspect of corporate culture: employee empowerment.

In step 2, we fine-tune a BERT model to identify culture-related segments that lack specific keywords. The construction of our training set involves using segments, identified in step 1 as containing relevant keywords, as positive examples (culture = 1). Conversely, we include randomly selected segments without those keywords as negative examples (culture = 0). This training set is used to fine-tune the model, which is then deployed across all segments (excluding those identified in step 1). Based on the model's predictions, we sort these segments by percentile rankings of predicted probabilities. We focus on the top 5% of segments with the highest predicted probabilities of relating to culture.

In step 3, we use ChatGPT to screen the segments. Although the trained model achieves a high AUC (at 0.981), we observe a significant number of false positives. The segments predicted with high probabilities often pertain to other intangible aspects such as leadership or strategy, rather than

---

[1] We use the following phrases for this exact matching process: "corporate culture," "company culture," "company's culture," "firm culture," "firm's culture," "organizational culture," "workplace culture," "business culture," and "culture in the company."

[2] For the segments containing the word "cultural," the definition is simply "Cultural: of or relating to culture, defined as …", followed by the corresponding definition for "culture."

corporate culture. To address this issue, we integrate the capabilities of ChatGPT for an additional layer of filtering. Table IA2 Panel B shows the prompt. The prompt instructs ChatGPT to assess whether each input segment is relevant to corporate culture, based on any of the following four definitions of corporate or organizational culture: principles and values guiding employees (Guiso, Sapienza, and Zingales 2015), shared beliefs, assumptions, values, or preferences driving group behaviors (Li and Van den Steen 2021), norms and values widely shared and strongly held in the organization (O'Reilly and Chatman 1996), or an informal institution characterized by behavioral patterns reinforced by events, people, and systems (Grennan and Li 2023).

For each input segment, ChatGPT is asked to provide a brief explanation (not exceeding 50 words) justifying whether the segment discusses corporate or organizational culture topics as per the provided definitions, and a classification of the segment as either "Culture" or "No Culture." Only those segments classified as "Culture" are retained. This step adds 97,507 segments. Our final data set comprises 138,545 culture-related segments (41,038 segments from step 1 + 97,507 segments from steps 2 and 3). For reports containing multiple culture-related segments, these segments are combined into a single consolidated segment. This consolidation facilitates subsequent analysis by treating each report as a coherent unit of observation. The resulting data set consists of 86,112 segments. Each of them represents the aggregated culture-related content within a single report.

5. Cause-effect Relation Extraction (RE)

Our approach leverages the power of generative AI combined with two key techniques: chain-of-thought (CoT) prompting and retrieval augmented generation (RAG). We apply these techniques in a two-stage process.

In the first stage, we employ CoT prompting to provide a structured reasoning framework for the model. The CoT prompt (Table 1 Panel A) guides the generative AI model to break down the task of relation extraction into a series of discrete steps. For each input segment, the model first identifies and extracts any mentioned culture type. It then looks for causal factors that influenced a culture type (causes) as well as downstream impacts or outcomes that the culture type had on the organization (effects). Finally, the model composes the extracted culture type, causes, and effects into standardized cause-effect triples. Depending on the richness of the discussion, a single segment may yield multiple causal triples or none at all. In cases where the segment text alone lacks sufficient context for the model to confidently discern any culture type or causal relations, it outputs "I need more context" and the segment is passed to the second stage.

The second stage handles segments where the model requests additional context. We employ retrieval augmented generation (RAG) to dynamically integrate relevant information from other parts of the report. For each segment needing more context, we perform semantic search using cosine similarity between pre-computed embeddings to identify related segments. The search retrieves up to five most similar segments, filtered by a probability threshold set at the 75th percentile of culture-relevance probabilities (as determined by the BERT-based culture probability model in A.4) across all segments. We also include segments immediately preceding and following the focal segment, as these often contain contextual information. The total context is constrained by a maximum token count of 128,000 tokens to prevent exceeding model context windows. When selecting context segments, we prioritize retaining segments with higher similarity scores while removing less similar ones until the token constraint is satisfied. Non-adjacent segments in the final context are separated by skip markers ([...]) to maintain logical flow. The model then receives both the focal segment and this filtered additional context as input. With this augmented input, the model re-attempts relation extraction using the same chain-of-thought prompting approach. If no cause (effect) relations can be confidently extracted even with additional context, the model produces an empty output for that segment.

To optimize efficiency and minimize cost when leveraging ChatGPT, we implement batch prompting (Cheng, Kasai, and Yu 2023) and parallel processing through multithreading. Batch prompting allows the model to generate responses for multiple samples in one batch during a single inference run, reducing the total number of API calls needed from N to N/b, where N represents the total number of samples and b signifies the number of samples per batch. We set b = 5 following Cheng, Kasai, and Yu (2023), who demonstrate this provides an optimal balance between cost and model performance. The multithreading implementation uses a thread pool executor to process multiple batches concurrently across separate CPU threads. This parallel architecture substantially reduces total processing time compared to sequential execution. Each thread independently handles API calls for its assigned batch while maintaining thread safety through synchronously writing results to an SQLite database. We set the temperature parameter to 0 and random seed to 1 to ensure deterministic output.

6. Canonicalization of Culture Types, Causes, and Effects

A key challenge in extracting corporate culture-related insights from text is the linguistic diversity in how cultural concepts are discussed. Analysts may refer to the same culture type, cause, or effect using a wide variety of phrasings and terminologies. To enable meaningful aggregation and analysis of the extracted cause or effect relations, we implement a canonicalization process for normalizing the extracted culture types, causes, and effects to a standardized taxonomy.

For culture types, we employ a two-stage approach combining manual and AI-driven categorization.

In the first stage, we focus on the most frequently mentioned culture types, causes, and consequences — specifically, the unique phrases that appear at least ten times across our corpus of analyst reports. We manually review and categorize each of these phrases into a taxonomy of six broad culture types, 18 causes, and 17 consequences drawn from prior literature. This manual categorization involves each author independently reviewing and categorizing the phrases, with disagreements or ambiguities resolved through discussion and by referring back to the original report context in which the phrase was used. The prototypical examples are used in our prompt in Table 1 Panels A and B.

In the second stage, because we find that the miscellaneous/other category from stage one contains potentially relevant cultural phrases that warrant more precise classification. In other words, the model has false positives in classifying categories as miscellaneous/other. To further refine this category, we perform an additional round of generative AI categorization on only the phrases initially categorized as miscellaneous/other. We first conduct a manual inspection of the most frequent phrases initially classified as miscellaneous/other. After examining these phrases, we document representative examples and use them as prompts in Table IA2 Panels C-E to help the model perform another pass on the miscellaneous/other category.

The above two steps give us culture types, causes, and effects in a standardized taxonomy, but we still need to canonicalize the full cause-effect triples into a consistent format. This involves the following steps:

1. We map one of the entities in each extracted triple to one of our standardized culture types, leveraging the phrase-to-category mapping developed earlier. This entity becomes the normalized culture type for the triple.

2. We use generative AI to classify the specific causal relationship between the two entities as either forward causality (->, the first entity causes the second), backward causality (<-, the second entity causes the first), or bidirectional causality (<->, the entities mutually influence each other). For example, “provides opportunity for” is canonicalized as ->, “threatened by” as <-, and “align with” as <->.

3. We then determine which of the two entities is the culture entity and which is the “other” (i.e., cause or effect) entity. To do so, we calculate the similarity between each entity and the raw extracted culture type phrase associated with the triple, using fuzzy string matching. Specifically, we use the partial_token_set_ratio metric from the fuzzywuzzy library, which computes the similarity between two strings based on their shared token sets while allowing for partial token matches. The entity with

the higher similarity score is designated as the "culture" entity, while the other is designated as the "other" entity (representing a cause or effect). In rare cases where both similarity scores are lower than 80 (less than 80% of the tokens in the extracted cultural phrase match the tokens in either entity after accounting for partial matches), we perform an additional check for the presence of culture-related keywords (e.g., "culture," "cultural") to make the designation.

4. Finally, by examining the directionality of the classified causal relation, we then determine whether the "other" entity is a cause (if pointing to the culture entity) or an effect (if originating from the culture entity).

The end result is a fully standardized causal triple of the form (Culture_Type, Relation_Direction, Cause_or_Effect) for each extracted relation. Table 1 Panel B shows the prompt used to canonicalize the reasons analysts discuss culture.

7. Relation Extraction and Canonicalization of Earnings Call Transcripts and Glassdoor Employee Reviews

The identification of culture-related segments, relation extraction, and canonicalization of earnings call transcripts and Glassdoor employee reviews largely follow our process for analyst reports.

Our earnings call data over the period 2004-2020 is from Capital IQ Transcripts database. We take a number of steps to clean the data, including removing firms without Capital IQ company ID, keep earnings conference calls only (removing other types of calls), dropping duplicated calls, using the last copy of a call, and retaining calls that can merge with gvkey. Our final call sample comprises 243,501 calls by 12,006 firms (corresponding to 72,749 firm-year observations).

Our Glassdoor employee review data over the period 2008-2020 is from Revlio Lab. After matching employer to gvkey, our final employee review sample comprises 5,343,864 reviews by employees from 5,187 firms (corresponding to 41,969 firm-year observations).

Our core methodology remains consistent across these two different corpora, with four specific adaptations to account for their distinct features.

First, the text segmentation process differs. Earnings call components and employee reviews serve as natural units of analysis, which eliminates the need of the C99 algorithm for text chunking. For earnings calls, each component is usually a paragraph of presentation or a complete answer from an executive to an analyst's question during Q&A. Each employee review represents a self-contained segment that typically focuses on a specific culture type.

Second, the context augmentation requirement differs. We do not apply retrieval augmented generation (RAG) to earnings call components or employee reviews. These text units are sufficiently self-contained, rendering additional context retrieval unnecessary. The culture type, causes, and effects are typically within the segment itself, when applicable.

Third, the prompt specification requires customization. We tailor the prompts to reflect each corpus' different context. The earnings call prompt refers to "executives' discussions during earnings calls about corporate culture." For Glassdoor data, the prompt specifies "employee reviews of companies from Glassdoor.com about corporate culture." These contextual markers help direct the model's interpretive framework.

B. Matching Analyst Name in Reports to Analyst ID (AMASKCD) in I/B/E/S

We match lead analyst (i.e., the first author of a report) name to analyst ID (AMASKCD) in the I/B/E/S database as follows.

First, to unmask abbreviated broker names and analyst names from I/B/E/S, we manually search each broker's full name and its analysts from Capital IQ. Our matching process involves three steps: 1) we match abbreviated broker names in I/B/E/S (ESTIMID) to full broker names in Capital IQ by resemblance; 2) we ascertain the match in Step 1 by matching analyst names (ANALYST) in I/B/E/S with those in Capital IQ using the last name and first name initial; and 3) we supplement the above two steps by checking whether Capital IQ analysts' stock coverage is the same as that by matched I/B/E/S analysts. Of the 1,075 broker names in I/B/E/S, we are able to unmask full names for 928 brokers (an 86.3% matching rate).

We then obtain analyst information, including biography and prefix (Mr. versus Ms.), from their employment history in Capital IQ. In the end, we are able to unmask 13,164 out of the 14,909 analysts in the I/B/E/S Detail Recommendations file (an 88.3% matching rate).

Second, to match each analyst in the report sample to analyst ID (AMASKCD) in the I/B/E/S data set, we match each analyst's name in Investext to our unmasked broker names and analyst names in the I/B/E/S-Capital IQ merged sample as described above. Our matching proceeds as follows: 1) we match each broker in Investext to broker name and ID (EMASKCD) in the I/B/E/S-Capital IQ merged file; of the 1,006 unique brokers in Investext, we can link 443 brokers with EMASKCD – analysts affiliated with these 443 brokers produce 91% of the reports in our report sample; and 2) for cases in which Investext has lead analyst's full first name and full last name, we match analyst name in Investext to analyst name and ID (AMASKCD) in the I/B/E/S-Capital IQ merged file; we further verify this match if there is also a match between broker name and EMASKCD established above. In the end, we are able to uncover AMASKCD for 7,921 analysts, representing 78% of the analysts affiliated with the 443 brokers in our analyst report sample.

Our final sample comprises 1,744,540 reports covering 38,530 firm-year observations for 2,988 unique firms over the period 2000-2020.

C. Matching Analyst Reports to I/B/E/S Forecast Data

When examining whether and how analysts' perspectives on culture impact price formation at the report level, we need to control for each report's quantitative and qualitative output (i.e., earnings forecast, stock recommendation, and target price). As a result, we need to match analyst reports from Investext in our sample with I/B/E/S forecast data following prior work (e.g., Huang, Zang, and Zheng 2014).

We employ a similar approach to link each report with its earnings forecast, stock recommendation, and target price in I/B/E/S. Here is an illustration of the process using earnings forecasts as an example.

Each report in our sample (from Section B) has a report date (DATE), a firm ID from I/B/E/S (CUSIP), an analyst ID from I/B/E/S (AMASKCD), and a broker ID from I/B/E/S (EMASKCD). Each earnings forecast in the I/B/E/S Detail file has an announcement date (ANNDATS), a review date (REVDATS), a firm ID (CUSIP), an analyst ID (AMASKCD), and a broker ID (EMASKCD). The announcement date is the day when an analyst revises her estimate and provides a forecast in a report. The review date is the day when an analyst confirms to I/B/E/S that her outstanding forecast is current. In I/B/E/S, a forecast is considered valid during the period from its announcement date until its review date. Within the period, analysts may issue multiple reports reiterating a forecast, but these reiterations have no separate entries in I/B/E/S.

We use a matching window, which is two days before a report's announcement date to two days after its review date, to match a report from Investext in our sample to an earnings forecast from I/B/E/S if the report date (from Investext) is within the "matching window," and there is a match of CUSIP-AMASKCD-EMASKCD between a report in our sample and an I/B/E/S earnings forecast.

Of the 1,744,540 reports in our sample (from Section B), we are able to match stock recommendations from I/B/E/S for 1,413,260 reports (an 81.0% matching rate), representing 36,100 firm-year observations associated with 2,898 unique firms; we are able to match target prices for 1,402,233 reports (an 80.4% matching rate), representing 35,673 firm-year observations and 2,887 unique firms. The samples used in Table 8 are smaller due to data availability for control variables.

Finally, we are able to match earnings estimates, stock recommendations, and target prices from I/B/E/S for 1,089,760 reports (a 62.5% matching rate), representing 34,314 firm-year observations associated with 2,858 firms. The sample used in Table 9 is smaller due to data availability for control variables.

References

Cheng, Z., J. Kasai, and T. Yu, 2023. Batch prompting: Efficient inference with large language model APIs, *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track*, 792–810.

Table IA1
Examples of the boilerplate segments

The table provides examples of the boilerplate segments in analyst reports, sorted by the predicted probability of a segment being boilerplate using a fine-tuned BERT model (in descending order). We retain segments with the predicted probability at 0.22 (the sample median) or lower.

Decile	Probability	Example
10	0.998	The investments or services contained or referred to in this report may not be suitable for you and it is recommended that you consult an independent investment advisor if you are in doubt about such investments or investment services. Nothing in this report constitutes investment, legal, accounting or tax advice or a representation that any investment or strategy is suitable or appropriate to your individual circumstances or otherwise constitutes a personal recommendation to you. CS does not offer advice on the tax consequences of investment and you are advised to contact an independent tax adviser.
9	0.979	EEA -The securities and related financial instruments described herein may not be eligible for sale in all jurisdictions or to certain categories of investors.
8	0.925	Distribution of ratings: See the distribution of ratings disclosure above. Price Chart: See the price chart, with changes of ratings and price targets in prior periods, above, or, if electronic format or if with respect to multiple companies which are the subject of this report, on the DBSI website at http://gm.db.com.
7	0.871	Suspended -the company rating, target price and earnings estimates have been temporarily suspended. For disclosure purposes, Evercore Group's prior "Overweight," "Equal-Weight" and "Underweight" ratings were viewed as "Buy," "Hold" and "Sell," respectively. Evercore ISI utilizes an alternate rating system for companies covered by analysts who use a model portfolio-based approach to determine a company's investment recommendation.
6	0.858	As a result, investors should be aware that the firm may have a conflict of interest that could affect the objectivity of the report and investors should consider this report as only a single factor in making their investment decision.
5	0.264	He also noted the Fed now has much more robust tools to guard against systemic risk vs. overseeing individual institutions. There were no questions about the history of the supervision process of large bank holding companies within the Atlanta Fed (district six) or more generally the troubled Georgia and Florida real estate markets.
4	0.054	These contracts usually have 1-5 year terms. Since residential customers generally fall under local government jurisdictions, the contracts are negotiated with the municipality -which in turn, will bill the residential customer through taxes. For larger commercial and industrial customers, the company will negotiate directly with the end-user.
3	0.028	Integration of recent acquisitions. Our downside risk for the shares is $10, which represents just over 10x our 2012 EPS estimate of $0.95 and is typically the bottom end of the company's historical multiple range.
2	0.003	We include these gains in our calculation of ongoing earnings because they are not entirely one-time though they are infrequent items. We have raised our 2005 EPS estimate to $1.50 from $1.40 to include an expected $0.11 fourth quarter investment gain on the sale of an interest in a coal-fired power plant in Georgia.
1	0.000	"Harman is still in a transitional phase with its highly anticipated scalable infotainment backlog not launching until fiscal 2013/2014. And while we acknowledge the company has made significant progress on the cost side, Harman will have to consistently execute on those cost cutting initiatives for the next several quarters to help prop-up its low-price and low-margin customized business."

Table IA2
Prompts to filter culture-related segments in analyst reports and refine classification of culture types/causes/effects

This table presents detailed instructions given to generative AI models to filter corporate culture-related segments in analyst reports and to refine classification of miscellaneous/other culture types/causes/effects. Panel A shows the prompt for word sense disambiguation relating to the words “culture(s)” and “cultural.” Panel B shows the prompt to determine whether a segment with a high predicted probability of relating to culture is really about corporate culture. Panel C shows the prompt to further classify miscellaneous culture types. Panel D shows the prompt to further classify other causes. Panel C shows the prompt to further classify other effects.

Panel A: Prompt to perform word sense disambiguation

```
Perform word sense disambiguation on the word 'culture(s)' or 'cultural' in the input segments.
These segments are from sell-side equity analyst research reports on companies. Classify each input
segment into one of the three categories: 'organizational,' 'societal,' or 'biological.' Definitions
for these categories are as follows:

* organizational culture: 'The values, typical practices, and goals of a business or other
organization, especially a large corporation. e.g., Their corporate "culture" frowns on avoiding
risk. We recognize a cost-cutting "culture". '
* societal culture: 'The behaviors and beliefs characteristic of a particular group of people, as a
social, ethnic, professional, or age group (usually used in combination). e.g., the youth "culture";
the drug "culture".
* biological culture: 'Biology. the cultivation of microorganisms, as bacteria, or of tissues, for
scientific study, medicinal use, etc. or the product or growth resulting from such cultivation.
e.g., cell "culture".'

For each input segment, provide:
    1. The input ID (input_id).
    2. The classification of the word 'culture(s)' or 'cultural' in input text as 'organizational,'
'societal,' or 'biological.'

Format your response in JSON. Make sure you process all of the inputs. Response format:

{"all_results":
    [
        {
            "input_id": "XXXX",
            "classification": "organizational/societal/biological",
        },
        {
            "input_id": "YYYY",
            "classification": "organizational/societal/biological",
        },
        ... (other inputs)
    ]
}
```

Panel B: Prompt to determine if a segment is really about corporate culture

```
Assess whether each of the following input segments is relevant to corporate or organizational
culture. The segments are from sell-side equity analyst research reports on companies. An input
segment is relevant to corporate or organizational culture if it discusses topics consistent with
any of the following definitions for corporate or organizational culture:

* "principles and values that should inform the behavior of all the firms’ employees."
* "a group’s shared beliefs, assumptions, values, or preferences that then drive that group’s
behaviors."
* "a set of norms and values that are widely shared and strongly held throughout the organization."
* "an informal institution typified by patterns of behavior and reinforced by events, people, and
systems."

For each input segment, provide:
    1. The input ID (input_id).
    2. Brief reasons (in 50 words or less) that explain if the segment contains discussions about
corporate or organizational culture topics using ANY of the definitions provided above.
    3. Classification of the segment as 'Culture' if it is relevant to corporate or organizational
culture, or as 'No Culture' if it does not.

Format your response in JSON. Make sure you process all of the inputs. Response format:
```

```
{"all_results":
    [
        {
            "input_id": "XXXX",
            "explanation": "This segment is relevant to corporate/organizational culture
because...",
            "classification": "Culture",
        },
        {
            "input_id": "YYYY",
            "explanation": "This segment is not relevant to corporate/organizational culture because
it mainly discusses...",
            "classification": "No Culture",
        },
        ... (other segments)
    ]
}
```

Panel C: Prompt to further classify "miscellaneous" culture types

```
Task: Categorize the phrases on corporate culture into one of the following six types:

* Collaboration and People-Focused: Focusing on (or deficient in) collaboration, cooperation,
teamwork, supportive, low levels of conflict, community, communication within an organization,
employee well-being, employee equity sharing and compensation, diversity, inclusion, empowerment, or
talent.
* Customer-Oriented: Focusing on (or deficient in) sales, customer, customer service, listening to
the customer, customer retention, customer experience, customer satisfaction, user experience,
client service, being brand-driven, quality of product, quality of service, quality of solution, or
taking pride in service.
* Innovation and Adaptability: Focusing on (or deficient in) innovation, creativity, technology,
entrepreneurship, adaptability, transformations, flexibility, agility, willingness to experiment,
beyond tradition, disruption, fast-moving, quick to take advantage of opportunities, resilience to
change, or taking initiative.
* Integrity and Risk Management: Focusing on (or deficient in) integrity, high ethical standards,
being honest, being transparent, accountability, do the right thing, fair practices, being
trustworthy, risk management, risk control, compliance, discipline, or financial prudence.
* Performance-Oriented: Focusing on (or deficient in) high expectations for performance, sales
growth, achievement, competitiveness, results, hard work, efficiency, productivity, consistency in
executing tasks, setting clear goals, following best practices, striving for operational excellence,
or exceeding benchmarks.
* Miscellaneous: Non-specific corporate culture, or corporate culture that does not easily fit into
the above types.  For example, "strong culture", "weak culture", "positive culture", "negative
culture", or "cultural change" (without details on the company's culture).

JSON Output Structure:

{
    "all_results": [
        {
            "input_phrase_id": 1,
            "input_phrase": "",
            "canonical_entity": "One of the predefined standardized entity names or 'Miscellaneous'"
        },
        ...
    ]
}
```

Panel D: Prompt to further classify "other" causes

```
As an expert specializing in corporate culture and causal reasoning, conduct entity canonicalization
on selected phrases taken from analyst reports. These phrases are considered to be *causes of
corporate culture*. Your objective is to map each phrase to the most appropriate standardized entity
name from the predefined list below.

Instructions:
 - Review the provided examples for each standardized entity to grasp their scope and nuances.
 - Carefully analyze each phrase and assign it to the most fitting standardized entity based on its
underlying meaning.
 - Assign the phrase to "Other" only if it does not align with any of the predefined entities.
Before assigning a phrase to "Other", ensure that it does not reasonably fit any existing
categories.
```

```
The available standardized entity names and examples are:
----
* COVID-19:

* Disruptive technology:

* Economic downturn:

* Internal conflicts:

* Mergers and acquisitions:
Examples: "integration challenges", "integration risks", "cultural integration challenges"
* Management change:
Examples: "appointment of chief diversity officer"
* Market expansion:
Examples: "increased competition", "improved market position"
* Regulatory issues:
Examples: "food safety incidents", "SEC investigation"
* Shareholder activism:

* Strategic transformation:
Examples: "new organizational structure", "organizational change", "transition to a more solutions-
based focus", "new product launches", "organizational structure changes"
* Customer relations:
Examples: "strong brand reputation", "strong Midwestern franchise", "strong brand recognition",
"strong brand"
* Management team:
Examples: "management initiatives", "management decisions", "key individuals", "management's focus
on culture"
* Compensation structure:
Examples: "correct makeup of human capital", "high cost of living in California"
* Employee hiring and retention:

* Business relationship:
Examples: "long-standing personal relationships with management", "local decision makers"
* Business strategy:
Examples: "Danaher Business System (DBS)", "decentralized structure", "investment in technology",
"strong balance sheet", "management's conservative approach", "investment in R&D", "decentralized
organizational structure", "decentralized management structure", "commitment to R&D", "One Ford
strategy", "My Macy's initiative", "no-haggle pricing model", "management's focus on efficiency",
"focus on R&D", "decentralized operating structure", "cost-cutting measures", "strong defensible
competitive position", "flat organizational structure", "focus on operational excellence", "cost
savings initiatives", "custom solutions", "increased R&D spending", "heavy reliance on internal
development", "focus on execution", "guiding principles", "focus on cost control", "increased
accountability", "Honeywell Operating System (HOS)", "Fortive Business System", "cost reductions",
"cost reduction initiatives", "strong financial commitment to drug development", "lean initiatives",
"internal initiatives", "long history as an asset manager", "implementation of MAGIC Selling
program", "focus on cost management", "flattened organizational structure", "first-mover advantage",
"financial commitment to drug development", "focus on efficiency", "ethos prohibiting
preservatives", "decentralization", "establishing guiding principles", "decentralized operating
model", "conservative credit culture"
* Workplace safety:

----

JSON Output Structure: {
    "all_results": [
        {
            "input_phrase_id": 1,
            "input_phrase": "",
            "canonical_entity": "One of the predefined standardized entity names or 'Other'"
        },
        {
            "input_phrase_id": 2,
            "input_phrase": "",
            "canonical_entity": "One of the predefined standardized entity names or 'Other'"
        }
        ...
    ]
}
```

Panel E: Prompt to further classify “other” effects

```
As an expert specializing in corporate culture and causal reasoning, conduct entity canonicalization
on selected phrases taken from analyst reports. These phrases are considered to be *consequences of
corporate culture*. Your objective is to map each phrase to the most appropriate standardized entity
name from the predefined list below.

Instructions:
 - Review the provided examples for each standardized entity to grasp their scope and nuances.
 - Carefully analyze each phrase and assign it to the most fitting standardized entity based on its
underlying meaning.
 - Assign the phrase to "Other" only if it does not align with any of the predefined entities.
Before assigning a phrase to "Other", ensure that it does not reasonably fit any existing
categories.

The available standardized entity names and examples are:

----
* Market share and growth:
Examples: "competitive advantage", "sustainable competitive advantage", "improved competitive
positioning", "narrow economic moat", "wider economic moats", "improved competitive position",
"differentiation from competitors", "increased competitiveness", "competitive differentiation",
"enhanced competitive positioning", "differentiation from peers", "differentiation from
competition", "enhanced competitive position", "potential for multiple expansion", "success
attributed to local decision makers", "competitive advantages", "strong financial position",
"competitive positioning", "key competitive advantage", "pricing power", "better job of gathering
and retaining assets", "better asset retention", "sustained competitive advantages", "better-run
restaurants", "attractive investment opportunity", "concerns about straying from core competencies",
"competitive edge"

* Profitability:
Examples: "improved operational efficiency", "increased productivity", "improved productivity",
"improved execution", "improved performance", "improved efficiency", "improved operations",
"consistent investment performance", "better execution", "improved results", "enhanced operational
efficiency", "strong performance", "operational efficiency", "improved balance sheet", "operational
excellence", "operational improvements", "enhanced productivity", "streamlined operations", "stable
moat trend", "positive free cash flow", "operational improvement", "leaner organization", "improved
financial profile", "improved store operations", "strong cash flow generation", "superior results",
"productivity gains", "operational challenges", "faster decision-making", "increased efficiencies",
"stronger cash flows", "improved store productivity", "healthier balance sheet", "improved decision-
making", "improved operational execution", "solid performance", "improved inventory management",
"improved sales force productivity", "increased operational efficiency", "improved working capital",
"improved overall performance", "improved fund performance", "high degree of consistency of
operations", "financial results", "better performance", "better decision-making", "improved
restaurant operations", "efficiency gains", "sharing of best practices", "significant free cash
flow", "streamlined operations", "improve design and quality", "maximized yields for farmers",
"consistent execution", "enhanced operational performance", "increased agent productivity",
"disciplined capital allocation", "above-average execution", "enhanced productivity"

* Customer satisfaction:
Examples: "improved product quality", "improved brand perception", "increased brand awareness",
"improved quality of care", "sales disruptions", "improved company image", "improved brand equity",
"increased brand equity", "improved market perception", "improved patient outcomes", "increased
brand recognition", "improved visibility", "improved user experience", "improved reputation",
"successful brand management", "improved health outcomes", "diminished reputation", "elevated brand"

* Employee satisfaction:
Examples: "improved collaboration", "cultural transformation", "enhanced collaboration", "effective
teamwork", "improved communication", "employees", "improved internal collaboration"

* Mergers and acquisitions:
Examples: "successful cultural integration"

* Internal conflicts:
Examples: "organizational changes", "decentralized management structure", "decentralized
organizational structure", "improved organizational structure", "cultural changes", "difficulty in
changing culture", "corporate culture shock", "decentralization"

* Innovation

* Diversity, Equity, and Inclusion

* Management change and retention

* Risk management:
```

```
Examples: "increased accountability", "improved accountability", "improved credit quality",
"improved transparency", "pristine credit quality", "improved internal controls", "improved asset
quality", "greater accountability", "solid credit quality", "strong asset quality", "stronger than
peer asset quality numbers", "debt reduction", "minimized franchise risk", "improved credit
profile", "strong credit quality", "greater transparency", "accountability", "better
accountability", "stable credit quality", "strong credit culture"

* Misconducts

* Resilience:
Examples: "successful turnaround", "earnings volatility", "business model uncertainty", "consistent
results", "successful transition", "financial flexibility", "successful transition to a dynamic
competitor", "turnaround success", "potential turnaround", "increased flexibility", "significant
negative impact on business", "successful cultural integration", "potential for improved
performance", "financial stability", "successful execution of strategy"

* Business relationship:
Examples: "top ranking in agent surveys", "increased collaboration", "local decision makers"

* ESG practices

* Shareholder value:
Examples: "long-term success", "improved corporate governance", "effective capital allocation",
"share price volatility", "stock price volatility", "continued success", "consistent results",
"improved capital allocation", "major contributor to company's success", "stock underperformance",
"stock price decline", "disciplined capital allocation"

* Investor relations
----


JSON Output Structure:

{
    "all_results": [
        {
            "input_phrase_id": 1,
            "input_phrase": "",
            "canonical_entity": "One of the standardized entity names."
        },
        {
            "input_phrase_id": 2,
            "input_phrase": "",
            "canonical_entity": "One of the standardized entity names."
        }
        ...
    ]
}
```

Table IA3
A list of culture types identified in prior work

This table lists culture types examined by prior work in the literature. The last column lists the five culture types that analysts/executives/employees refer to when discussing culture extracted by generative AI in our paper.

	Guiso, Sapienza, and Zingales (2015)	Grennan (2019)	Li et al. (2021)	Graham et al. (2022a, 2022b)	Our paper
	(1)	(2)	(3)	(4)	(5)
Data source(s)	Corporate website	Employee reviews	Earnings calls	Executive surveys and interviews	Analyst reports Earnings calls Glassdoor reviews
Culture type	Communication	Adaptability	Innovation	Adaptability	Collaboration and people-focused
	Community	Collaboration	Integrity	Collaboration	Customer-oriented
	Hard work	Customer-orientation	Quality	Community	Innovation and adaptability
	Innovation	Detail-orientation	Respect	Customer-orientation	Integrity and risk management
	Integrity	Integrity	Teamwork	Detail-orientation	Performance-oriented
	Quality	Results-orientation		Integrity	
	Respect	Transparency		Results-orientation	
	Safety				
	Teamwork				

Table IA4
Representative examples of the extracted culture types, their causes, and their effects

The table provides some representative examples of the extracted culture types, and their causes and effects. The causes are grouped into three categories: events, people, and systems suggested by Guiso, Sapienza, and Zingales (2015), Graham et al. (2022a, 2022b), and Grennan and Li (2023).

Panel A: Different culture types

Culture type	Example
Collaboration and people-focused	collaborative culture, integration-oriented culture, team-oriented culture, cohesive corporate culture, partnership-oriented culture, cooperative culture, team-based culture, silos culture, alignment-oriented culture, collegial culture, employee-centric culture, inclusive culture, diverse corporate culture, family-oriented culture, people-centric culture, talent-focused culture, people-focused culture, empowering culture, internal-promotion culture, supportive culture
Customer-oriented	customer-centric culture, sales-driven culture, sales-oriented culture, service-oriented culture, customer-focused culture, customer-centric, client-focused culture, client-centric culture, client-oriented culture, consumer-centric culture, customer-obsessed culture, customer support-focused culture, customer-friendly culture, user experience centric culture, culture of customer satisfaction, brand-centric culture, brand-driven culture
Innovation and adaptability	innovative culture, entrepreneurial culture, growth-oriented culture, innovative corporate culture, data-driven culture, technology-driven culture, innovation-driven culture, knowledge-driven culture, creative and innovative culture, entrepreneurial and decentralized culture, adaptive culture, adaptive corporate culture, change-oriented culture, resilient culture, proactive culture, continuous improvement culture, evolving culture, transformative culture, transformational culture, agile culture
Integrity and risk management	accountable culture, community-oriented culture, ethical corporate culture, socially responsible culture, accountability culture, accountability-driven culture, values-driven culture, integrity-based culture, transparent culture, integrity-driven culture, disciplined culture, conservative culture, risk-averse culture, cautious culture, risk-aware culture, safety-oriented culture, compliance-oriented culture, financially disciplined culture, prudent culture, risk management culture
Performance-oriented	performance-driven culture, results-oriented culture, competitive culture, aggressive culture, profit-driven culture, goal-oriented culture, high-performance culture, shareholder-focused culture, winning culture, success-driven culture, decentralized culture, cost-conscious culture, efficiency-driven culture, efficiency-oriented culture, quality-focused culture, cost-cutting culture, detail-oriented culture, centralized culture, process-oriented culture, operational culture
Miscellaneous	challenging corporate culture, acquisitive culture, ambitious culture, stable corporate culture, dedicated corporate culture, experienced corporate culture, focused culture, unique culture, traditional corporate culture, long-term focused culture

Panel B: Different causes of culture

Category	Cause	Example
Event	COVID-19	covid-19 pandemic, disruption from covid-19, response to covid-19 pandemic, uncertainty related to the covid-19 pandemic, covid disruptions
Event	Disruptive technology	disruptive products, next-generation technology, focus on disruptive innovation, phase of disruptive technology, breakthrough innovations through AI
Event	Economic downturn	financial crisis, economic cycle, economic downturn, economic environment deterioration, weak macro environment and jobs market
Event	Internal conflicts	Bureaucracy, internal politics, disconnect between it department and senior management, history of frequent strike activity, internal power

		struggles
Event	Management change	leadership change, new management team, management turnover, new CEO, separation of chairman and CEO roles
Event	Market expansion	international expansion, expansion into new markets, global presence, rapid expansion, seeking new areas of growth
Event	Mergers and acquisitions	acquisitions, strategic acquisitions, mergers and acquisitions, M&A activities, integration of acquired businesses
Event	Regulatory issues	regulatory actions, challenging regulatory environments, unknowns within consumer regulatory agency, regulatory changes, SEC investigation
Event	Shareholder activism	shareholder pressure, significant insider ownership, interaction with activist investors, agreement with activist investor for board changes, proxy fight with activist investor
Event	Strategic transformation	organizational restructuring, strategic initiatives, reorganization, transition to solutions-based focus, strategic changes, business transformation
People	Customer relations	focus on customer service, best-of-breed customer service, deep client relationships, decades of high-quality service, direct relationships with end-users
People	Management team	experienced management team, strong management team, CEO's leadership, visionary leadership, long-tenured management team
System	Business relationship	local management with deep roots in each community, valuable commercial client relationships, establishment of unique relationship with independent agents, ability to attract and foster close and long-lasting business relationships, strategic partnerships
System	Business strategy	training salesforce in value-over-volume strategy, focus on cost management, differentiated merchandising strategy, management's aggressive expansion initiative
System	Compensation structure	compensation structure, competitive compensation programs, compensation structure emphasizing incentive pay, incentive compensation structure, employee stock ownership
System	Employee hiring and retention	promotion from within, extensive training programs, long tenure of employees, workforce reduction, resisting layoffs during recession
System	Workplace safety	desire to minimize personal risk for employees, industry's effort to improve safety practices, independent safety oversight committee, efforts to improve safety practices, fair hearing and remedy process for workers' grievances
	Miscellaneous	structural attributes, resources, competitive environment, deep and broad industry expertise, accounting practices

Panel C: Different effects of culture

Effect	**Example**
Business relationship	strong customer relationships, cross-selling opportunities, critical industry relationships, retaining valuable commercial client relationships, stronger franchisee alignment
Customer satisfaction	improved customer service, improved customer experience, customer satisfaction, customer loyalty, improved customer satisfaction
Diversity, equity, and inclusion	improved diversity of leadership team, development of a diverse talent base, lack of diversity in board composition, promotion of more women, toxic culture of sexual harassment
Employee satisfaction	employee turnover, employee retention, low employee turnover, employee satisfaction, employee ownership
ESG practices	corporate governance weaknesses, environmental sustainability efforts, esg practices, enhanced governance practices, development of environmentally and ethically responsible products

Innovation	accelerated development of desirable new products, product innovation, new product development, technological leadership, focus on innovation
Internal conflicts	potential for business conflicts, wrestling with production planning, resistance to change, potential muddled strategy and infighting, management distraction
Investor relations	attractiveness to investors, rebuilding investor confidence, alignment of management and shareholder interests, shareholder friendliness, improved communication with investors
Management change and retention	loss of key personnel, management resource strain, smooth leadership transition, strong management team, management turnover
Market share and growth	revenue growth, market share gains, increased market share, expansion into new markets, establishing a strong presence in key markets around the world
Mergers and acquisitions	successful integration of acquisitions, successful acquisitions, challenges in integration, lower-than-expected synergies, M&A strategy
Misconduct	management protecting their own interests over investors, unusual accounting moves, massive legal liabilities, multiple scandals, legal troubles
Miscellaneous	Competitive advantage, long-term success, improved operations, continued success, positive geographic mix
Profitability	margin expansion, improved profitability, cost savings, increased profitability, more stable levels of profitability
Resilience	business resilience, resilience in the next downturn, resilience during recession, successful weathering of recent market volatility and macroeconomic uncertainty, persistent corporate momentum
Risk management	focus on risk management, focus and importance placed on risk management, handling credit risk well, improved credit quality, minimized franchise risk
Shareholder value	consistently above-average returns, strong balance sheet, returns exceeding cost of capital for longer periods, enhanced shareholder value, alignment of interests with shareholders

Table IA5
Examples of Retrieval Augmented Generation

This table illustrates how retrieval augmented generation (RAG) improves ChatGPT's analysis of analyst reports. The highlighted segments in each example are the input to ChatGPT (without additional context), while the entire passage in each example are the input to ChatGPT after applying RAG.

Example 1.

Obviously the company has tremendous visibility into 1Q04 since the quarter is 2/3 complete. We expect guidance to point to sequential revenue and earnings growth, and FY04 revenue growth of about 14 - 15%. • We reiterate our view, that Symbol is emerging from 2003 leaner, with a now stable channel, and a fuller pipeline. We believe revenue growth will come from a broadbased cyclical recovery, and that earnings will surge due to operational efficiences that follow from 2003 restructuring activities, and the elimination of expenses related to the SEC investigation and internal audits. [...] We believe Symbol remains a high risk stock in view of the on-going SEC investigation and potential shareholder lawsuit awards against the company following revelations regarding accounting malpractices at the company in the period 1999 -2002. A number of possible events could prompt us to view the stock more negatively and cause our $22 price target not to be achieved within the 12 month time frame. There is the risk that the Smart Media ruling ($218 million award) could weaken the balance sheet, although we think that this possibility is not very likely. Although Symbol is technically not in compliance with NYSE listing rules due to the delay in publishing audited results, we believe delisting risk is minimal given the ongoing and proactive dialogue between the company and the NYSE. Although the results and restatements presented were summary, unaudited figures, we believe they went a long way towards reassuring investors and mitigating delisting risk. In addition, the company has changed its supplier strategy, implemented a new channel strategy, centralized and relocated functions, installed new systems, processes and controls. While the strategy appears to be paying off, it is possible that the company is doing too much too quickly and that execution problems may arise, or that the improvements will prove fleeting in the absence of a more substantial change of culture. Symbol faces credible and emerging competition in each of its segments, including Cisco (wireless LANs), Proxim (Wireless LANs), Motorola (Wireless WANs), Unova (data capture solutions, ruggedized handhelds), Handheld Products (ruggedized computing), Metrologic (data capture solutions) and others. Symbol's leadership in laser scanning is threatened marginally by the introduction of low-cost imaging solutions using CCD or CMOS sensors.

Extracted relation(s):

- **Ongoing SEC investigation (Regulatory issues) → Risk-aware culture (Integrity and risk management).**
- **Risk-aware culture (Integrity and risk management) → Potential shareholder lawsuit awards (Misconducts).**

Explanation: The original segment only mentions "improvements will prove fleeting in the absence of a more substantial change of culture." It provides no basis for determining culture type or relationships. The new context revealing “accounting malpractices” and an “ongoing SEC investigation” allows the model to properly classify this as integrity and risk management culture and extract two key relationships: the SEC investigation influencing risk-aware culture, and this culture leading to potential shareholder lawsuits.

This report was written by Paul Coster from JPMorgan for Symbol Technologies, Inc. released on 11/5/2003.

Example 2.

Boeing CEO Dennis Muilenburg and BCA Chief Engineer John Hamilton provided testimony yesterday to the U.S. Senate's Commerce, Science & Transportation Committee and today to the House Committee on Transportation & Infrastructure.
Congress is investigating the two MAX crashes, and the MAX's certification process in an effort to make any needed changes to regulations to improve aircraft certification and enhance aviation safety.
We think investors have approached these hearings with an eye toward their short term impact, if any, on the return to service (RTS) of the MAX, which yielded +2% reaction yesterday and 1% reversal today.

Heated, but no major surprises: Committee members were prepared with pointed but reasonable questions for BA, and some displayed clear anger with what they viewed as obfuscation of the culture behind decisions that undermined MCAS and the MAX.
That said, we found the exchange to be largely as expected: hardline questioning from members with a sympathetic but tactical responses from Boeing.
One notable exhibit, a 2015 employee email questioning the single AOA sensor architecture was explained away as not triggering the required level of criticality for redundancy.
Ultimately, CEO Muilenburg defended Boeing's efforts in developing and certifying the MAX, but acknowledged Boeing made some mistakes.
When asked to cite the top three, he called out (1) the implementation of the AOA disagree alert, (2) MCAS architecture, and (3) communication and documentation shortfalls.

He emphasized BA's commitment to aviation safety and to ensuring such accidents never occur again.
That said, there was no clear admission of willful concealment of MCAS or an attempt to avoid incremental training.
Some members questioned Mr. Muilenburg's accountability given his pay-levels after the first accident.

Extracted relation(s):

- **Investigations into crashes (Regulatory issues) → Accountability-driven safety culture (Integrity and risk management).**
- **Accountability-driven safety culture (Integrity and risk management) → Focus on improving safety regulations (Risk management).**

Explanation: The original segment mentions “obfuscation of the culture behind decisions that undermined MCAS and the MAX,” but provides insufficient context to determine specific cultural relationships. The retrieved context revealing Congressional investigations into “two MAX crashes” and “the MAX’s certification process” enables identification of cause-effect patterns: investigations driving an accountability-focused safety culture, and this culture fostering improved safety regulations. Specifically, the wider context detailing “Congress is investigating…to make needed changes to regulations to improve aircraft certification and enhance aviation safety” establishes the link between accountability culture and safety improvements.

This report was written by Robert Spingarn from Credit Suisse for Boeing Company released on 10/30/2019.

Example 3.

TAG Take: The overall tone of the event was positive, with Whole Foods emphasizing its unique position in the market.

The company acknowledged the tough FY14 in which an increasing number of competitors encroached on its space.
And, Whole Foods presented its case for still being a highly differentiated and evolving business, with initiatives underway to maintain or grow its market share and better engage customers.
Whole Foods emphasized its unique in-store experience, beyond simply being a grocer, and its increased use of technology and social media.
[...]
We continue to view Whole Foods as being in a transition period to mid-growth from high-growth and to somewhat defensiveminded as competitors encroach on its natural/organic niche.
We maintain our Market Perform rating and 12-month price target of $52, based on applying an EV/EBITDA multiple of 12.5x to our CY15 EBITDA estimate of $1.48B.

[...]

Whole Foods follows an opportunistic real estate strategy, as it looks for the best sites it can find as opposed to following a set strategy by market.
In addition, all capital expenditure projects require an EVA analysis and must clear a five-year hurdle rate based on conservative forecasts for sales, capital, and rent.
Whole Foods has seen its rent increase 10% over the last five years, as the commercial real estate price index has risen 52%.
At the same time, the company has reduced the cost to build by 10% on a per square foot basis.

[...]

For example, seven locations were opened in Boston during FY13, providing weak comps the following year, but Boston is now delivering same-store sales above the company average.
With regard to competition, Whole Foods noted some stores might see a 10% headwind when a competitor opens, but after the first year Whole Foods starts to comp positively again.
Refreshing Older Stores and Lowering Cost Structure: Whole Foods refreshed 40 stores during 1QF15, and is on track to reach 200 by the end of 2015.
Décor refreshes are the least expensive and enhance the logos, colors, and lighting of the store, while the next level of refresh consists of bringing in new fixtures and cases, and lastly, some stores receive a full remodel.
Beyond store refreshes, the company is working to lower the in-store cost structure.

[...]

Price Investments: Whole Foods continues to be about quality, service, selection, and store experience, but the company realizes it needs to offer relevant pricing.
With that in mind, Whole Foods has made some successful produce investments, but is continuing to search for the best pricing strategy.
Currently, lower produce pricing experiments are being conducted in six or seven markets, and some are expected to roll out even further in 2H15.
However, given the tests started in October and including the high-volume holiday season, it is too early to extrapolate the results.
Competitive produce pricing should expand the overall value perception, given that produce tends to be the first department customers shop when entering the store and can set the tone for the shopping experience.
So far, the company is pleased with the unit lift it has seen in the markets with produce pricing investments, and Whole Foods intends to fund the price investments through cost efficiency.
In addition, the company noted that promotions are equally as promising as lower everyday pricing.

[...]

Raising Food Standards: Although Whole Foods is widely known as a premier natural and organic grocer, the company is intent on continuing to set the pace for raising food standards and transparency.
In recent years, the company has implemented five-step animal welfare ratings for its meat sales, non-GMO and organic labeling, and sustainability ratings.

The pursuit of higher food standards has required Whole Foods to create unique partnerships with special growers around the world, such as bananas from Costa Rica, non-antibiotic farm-raised salmon in Norway, and pasture-raised chicken.
Evolving Store Experience: The company wants to give consumers a reason to come in the store beyond just food shopping by creating a unique experience at each store.
For example, Whole Foods now brews its own beers in two of its stores and roasts coffee in two markets.

The company also has one location with a spa and another with a restaurant that accepts reservations and offers table service.
Some stores offer a cooking studio to teach customers how to prepare healthier meals and other locations offer different community events.
A key piece of the Whole Foods story is its unique and dynamic culture.

The culture is one that can adapt, shift, and evolve quickly, while allowing its core values to guide the decision-making process.
The company is very proud of the fact that it is consistently rated one of the best places to work, and views the high morale of its employees as being essential to its success.

Extracted relation(s):

- **Dynamic adaptive culture (Innovation and adaptability) → Unique in-store experience initiatives (Customer satisfaction).**
- **Dynamic adaptive culture (Innovation and adaptability) → High employee morale (Employee satisfaction).**

Explanation: The original segment mentions "unique and dynamic culture." It does not provide sufficient context to establish specific relationships. The context discussing how "Whole Foods emphasized its unique in-store experience" and detailing various innovations (brewing beer, roasting coffee, cooking studios) enables identification of key patterns: adaptive culture fostering unique store initiatives. Moreover, the wider context noting Whole Foods is "consistently rated one of the best places to work" and views "high morale of its employees as being essential to its success" establishes the link between adaptable culture and employee satisfaction.

This report was written by Joseph Feldman from Telsey Advisory Group for Whole Foods Market Inc. released on 3/2/2015.

Example 4.

Rice Brothers on the Cusp of Deposing EQT's Management Team
EQT reported the results of its AGM, with shareholders electing all of the seven Rice nominees as well as the five nominees supported by the Rice brothers and EQT.
We believe that it is likely that the company will announce later today that Toby Rice will replace Robert McNally as the CEO.
During the past several months, the Rice team has progressively released details of its plan to transform EQT, which were met with rebuttals from the management team of EQT.
Rice had been critical of EQT's high well costs and suggested this resulted in weaker-than-optimal capital efficiencies.
The Rice team highlighted the $300 million cost over-run in H2/18 as evidence of a lack of capital discipline.
The Rice brothers pointed out that the well costs could be significantly brought down by drilling longer laterals, thus decreasing the costs on a per foot basis.

Recently, the Rice brothers had broadened their reform agenda by providing a more detailed plan.
This included a comprehensive 100day plan, which include a plethora of technological initiatives to enhance capital efficiencies.
A key focus of the Rice plan will be to improve the current planning and scheduling of various operations.

A vital area for improvement identified by the Rice team is the optimization of rig mobilization to reduce rig movement while drilling wells by ~75%.
The Terminal Objective of the Rice Plan is to generate an incremental FCF of $500 million.
■ Our View: The successful execution of the plan by the Rice team remains to be seen.
However, the involvement of the Rice team has certainly put capital discipline and efficiencies under a microscope.

Looking forward, we believe this renewed focus on efficiency, improvement, and FCF is ultimately in the best interest of equity holders.
Practically speaking, in our experience, organizational transitions can take longer to realize than originally anticipated -especially when dealing with a transformation of corporate culture.
We will revisit our rating and target price upon evidence that the proposed plan is resulting in notable improvements.

EQT is the largest natural gas producer in North America, accounting for 4% of North American production.
It is primarily focused in the Appalachian Basin.
The company recently (2017) acquired its competitor Rice Energy.
The successful integration of this business has the potential to drive significant operational synergies going forward.

Extracted relation(s):

- **Rice brothers' reform agenda (Business strategy) → Transformational efficiency culture (Performance-oriented culture).**
- **Transformational efficiency culture (Performance-oriented culture) → Improved capital efficiencies (Profitability).**

Explanation: The original segment mentions “transformation of corporate culture” no specific relationships. The broader context revealing the “Rice brothers’ reform agenda” with its “100-day plan” and “technological initiatives to enhance capital efficiencies” enables identification of key patterns: reform agenda driving transformation toward efficiency culture, and this culture leading to improved capital efficiencies. Specifically, the wider context detailing criticism of “high well costs” and plans for “optimization of rig mobilization” establishes the link between cultural transformation and operational improvements.

This report was written by Aaron Bilkoski from TD Securities for EQT Corporation released on 7/10/2019.

Table IA6
Examples of the extracted cause or effect relations

This table provides examples of the extracted cause or effect relations in analyst reports by generative AI. In each example, a snippet of the culture-related segment is provided, the extracted terms with corresponding canonicalized terms are in parentheses, the cause or effect relations are highlighted in boldface, and generative AI's explanation for such extraction is provided.

Example 1: Coca-Cola's CEO Neville Isdell presented for the company at the CAGNY conference in Arizona this morning. While there was not much new news in the presentation, the company's tone has changed meaningfully from Isdell's presentation at CAGNY two years ago (when KO was promising the market very little) to today, when the company has greater confidence that its long-term algorithm is both working and sustainable. We highlight what we think were a few of the key takeaways below:
• The company highlighted the improving performance in 2006 as KO has moved through its Manifesto For Growth strategy and now enters a phase of likely sustainable growth with a focus on growing the core brands, capturing emerging platforms by establishing a culture of innovation, and providing franchise leadership to the bottlers (which was consistent with what we heard from CCE yesterday, with a greater focus on increasing collaboration with the bottling system).
• KO is entering 2007 with some of the strongest growth momentum in the last several years, and the company has demonstrated that they can continue to post solid growth despite underperformance in key markets.NA should remain a challenge in 2007, but KO has clearly shown an ability to turn around problem markets and seems very comfortable with its long-term growth model given strength in the balance of the business. While it's not yet clear what stage in the turnaround we have entered for key markets such as Japan, Germany, and India, KO expressed confidence in the recent improvements in these areas and we are encouraged by the results…

Extracted relation(s):

- **Manifesto for growth strategy (Business strategy) → Innovative and growth-oriented culture (Innovation).** Explanation: The culture is shaped by the strategic focus on growth and innovation.
- **Innovative and growth-oriented culture (Innovation) → Sustainable growth (Market share and growth).** Explanation: The culture fosters sustainable growth as indicated by the company's performance.

Ref: This report was written by John A. Faucher from JP Morgan for Coca-Cola Co. released on 2/22/2007.

Example 2: Our $46 Dec-12 Price Target is based on 12.0x our 2013 EPS of $3.85 (above M's trailing 3-year 10.5x avg, but 230bps below its dept store peer average of 14.3x). Focused on company specific initiatives (My Macy's, Magic Selling, and Direct/Omni-channel) M has separated itself from moderate peers (JCP/KSS) executing on a three-tiered strategy (brands, fashion, price). With the turn in the selling culture taking place just last summer (according to CEO Lundgren) and with gross margin drivers on the horizon (Omnichannel and price optimization) we see double digit earnings growth through 2015.

Extracted relation(s):

- **My Macy's initiative (Business strategy) → Focused selling culture (Customer-oriented).** Explanation: The segment mentions the My Macy's initiative as a specific strategy that has contributed to the focused selling culture.
- **Focused selling culture (Customer-oriented) → Double digit earnings growth through 2015 (Market share and growth).** Explanation: The report indicates that the focused selling culture is expected to result in double digit earnings growth.

Ref: This report was written by Matthew R. Boss from JP Morgan for Macy's Inc. released on 3/12/2012.

Example 3: It was championed by Starbucks' chairman Howard Schultz since early in Starbucks' existence. Health care costs have increased dramatically over the past several years, and continue to pressure the company's operating margins. However, we do not expect Starbucks to move away from the health benefit program, as it is an important part of its culture and its ability to attract good employees...

Extracted relation(s):

- **Employee-focused culture (Collaboration and People-oriented) → Ability to attract good employees (Employee satisfaction).** Explanation: A strong focus on employee benefits helps in attracting talent.

Ref: This report was written by Ashley R. Woodruff from Bear, Stearns & Co., Inc. for Starbucks Corporation released on 5/18/2005.

Example 4: … XPO has also instilled a strong performance-based culture (salary plus incentive compensation tied to gross margin dollars and gross margin dollars per load), which we believe has contributed to the recent increase in productivity and should help drive revenue growth as the salesforce continues to mature (average tenure is a little over one year).We believe XPO is positioned well to gain market share as it has focused on high customer service while broadening its solution set-something few competitors have been able to achieve successfully. … XPO has instilled a performance-driven culture focused on delivering high-quality service to all customers, with a goal of zero service failures (overall service levels are currently around 97% for pickup and delivery).

Extracted relation(s):

- **Strong performance-driven culture (Performance-oriented) → Increase in productivity (Market share and growth).** Explanation: The performance-driven culture established by Mr. Jacobs contributes to increased productivity as noted in the report.
- **Strong performance-driven culture (Performance-oriented) → High customer service (Customer satisfaction).** Explanation: The focus on performance drives high customer service levels, as indicated by the company's service goals.

Ref: This report was written by Nathan Brochmann from William Blair & Company for XPO Inc. released on 6/19/2015.

Example 5: … Demand Remains Solid; Raising FVE to $88 17 Feb 2017. Arista reported strong results in its fourth quarter, with revenue increasing above our expectations. We are impressed by another year of stellar revenue growth, as the company's strategic focus on large customers' needs and its culture of product innovation are paying off. …

Extracted relation(s):

- **Innovative product culture (Innovation and Adaptability) → Stellar revenue growth (Market share and growth).** Explanation: The focus on product innovation is directly associated with stellar revenue growth.

Ref: This report was written by Ilya Kundozerov from Morningstar Inc. for Arista Networks Inc. released on 5/8/2017.

Table IA7
Summary statistics

This table presents the summary statistics for samples used in different regression analyses. In Panel A, our firm-year sample consists of 24,250 firm-year observations, representing 2,318 unique firms over the period 2002-2020. In Panel B, our firm-analyst-year sample consists of 160,332 firm-analyst-year observations, representing 2,471 unique firms followed by 4,096 analysts. In Panel C, our analyst-executive firm-year sample consists of 8,369 firm-year observations, representing 1,581 unique firms over the period 2004-2020. In Panel D, our analyst-employee firm-year sample consists of 5,216 firm-year observations, representing 909 unique firms over the period 2008-2020.

Panel A: Summary statistics for Table 4 Panel A

	Observations	Mean	P25	Median	P75	STD
Culture discussion	24,250	0.563	0.000	1.000	1.000	0.496
Total assets	24,250	12516.700	749.830	2356.565	8275.834	33757.550
Firm age	24,250	26.951	13.000	22.000	40.000	17.655
Sales growth	24,250	0.098	-0.006	0.071	0.168	0.222
ROA	24,250	0.038	0.012	0.043	0.082	0.094
Leverage	24,250	0.235	0.067	0.213	0.354	0.196
Tangibility	24,250	0.239	0.053	0.149	0.362	0.238
ROA volatility	24,250	0.040	0.008	0.018	0.043	0.062
Large institutional ownership	24,250	0.387	0.127	0.373	0.599	0.293
Board independence	24,250	0.722	0.615	0.700	0.857	0.134
CEO duality	24,250	0.479	0.000	0.000	1.000	0.500
Number of key people changes	24,250	3.646	1.000	3.000	5.000	3.622
Number of M&As	24,250	0.709	0.000	0.000	1.000	1.224
Strong culture	24,250	0.194	0.000	0.000	0.000	0.395
Number of meetings	24,250	1.163	0.000	0.000	1.000	2.294
Employee culture rating	9,371	2.205	0.000	2.700	3.250	1.444
Number of employee reviews	9,371	85.069	4.000	18.000	66.000	203.381

Panel B: Summary statistics for Table 4 Panel C

	Observations	Mean	P25	Median	P75	STD
Star analyst	160,322	0.074	0.000	0.000	0.000	0.262
CFA	160,322	0.481	0.000	0.000	1.000	0.714
Postgraduate	160,322	0.619	0.000	1.000	1.000	0.660
Female	160,322	0.103	0.000	0.000	0.000	0.304
General experience	160,322	11.141	6.000	10.000	16.000	7.022
Firm experience	160,322	5.532	2.000	4.000	8.000	4.438
Number of industries followed	160,322	4.624	3.000	4.000	6.000	2.627
Number of firms followed	160,322	19.134	14.000	18.000	24.000	8.412
Forecast frequency	160,322	4.572	3.000	4.000	6.000	2.520
Broker size	160,322	64.430	26.000	57.000	103.000	42.506
Analyst-firm distance	160,322	1967.957	509.720	1348.400	2941.930	1935.344

Panel C: Summary statistics for Table 5 Panel A (Analyst-Executive)

	Observations	Mean	P25	Median	P75	STD
Type divergence	8,369	0.419	0.274	0.405	0.562	0.219
Cause divergence	8,369	0.633	0.515	0.638	0.833	0.161

Effect divergence	8,369	0.520	0.406	0.508	0.620	0.153
Total assets	8,369	22015.100	1692.296	5417.000	17948.490	47140.110
Firm age	8,369	30.110	16.000	25.000	45.000	18.397
Sales growth	8,369	0.080	-0.004	0.059	0.135	0.183
ROA	8,369	0.049	0.014	0.047	0.086	0.077
Leverage	8,369	0.253	0.100	0.232	0.365	0.191
Tangibility	8,369	0.222	0.049	0.134	0.330	0.226
ROA volatility	8,369	0.029	0.006	0.015	0.032	0.044
Large institutional ownership	8,369	0.407	0.186	0.398	0.603	0.281
Board independence	8,369	0.734	0.643	0.706	0.875	0.124
Loss year	8,369	0.120	0.000	0.000	0.000	0.325
CEO duality	8,369	0.456	0.000	0.000	1.000	0.498
CEO tenure	8,369	7.612	2.611	5.663	10.501	6.855
CEO Delta	8,369	859.995	102.624	283.269	764.191	1874.316
CEO Vega	8,369	170.772	5.436	64.900	220.681	255.287
CEO close-to-retire	8,369	0.208	0.000	0.000	0.000	0.406
Number of meetings	8,369	1.991	0.000	1.000	3.000	2.848

Panel D: Summary statistics for Table 5 Panel A (Analyst-Employee)

	Observations	Mean	P25	Median	P75	STD
Type divergence	5,216	0.535	0.407	0.550	0.662	0.182
Cause divergence	5,216	0.658	0.560	0.659	0.772	0.132
Effect divergence	5,216	0.675	0.590	0.681	0.769	0.116
Total assets	5,216	23089.980	1899.613	5700.878	18918.910	47797.210
Firm age	5,216	32.248	17.000	26.000	49.000	18.762
Sales growth	5,216	0.079	-0.002	0.059	0.132	0.180
ROA	5,216	0.057	0.023	0.055	0.094	0.075
Leverage	5,216	0.256	0.106	0.240	0.367	0.190
Tangibility	5,216	0.229	0.066	0.146	0.329	0.223
ROA volatility	5,216	0.029	0.007	0.016	0.033	0.042
Large institutional ownership	5,216	0.427	0.229	0.420	0.610	0.270
Board independence	5,216	0.728	0.643	0.692	0.857	0.123
Loss year	5,216	0.111	0.000	0.000	0.000	0.314
CEO duality	5,216	0.444	0.000	0.000	1.000	0.497
CEO tenure	5,216	7.797	2.666	5.748	10.752	7.035
CEO Delta	5,216	1042.531	126.017	357.691	925.764	2153.757
CEO Vega	5,216	187.784	4.484	80.402	257.474	264.185
CEO close-to-retire	5,216	0.212	0.000	0.000	0.000	0.409
Number of meetings	5,216	2.396	0.000	1.000	4.000	3.091

Table IA8
Correlation matrices for the firm-year and firm-analyst-year samples

This table presents the correlation matrices for samples used in different regression analyses. In Panel A, our firm-year sample consists of 24,250 firm-year observations, representing 2,318 unique firms over the period 2002-2020. In Panel B, our firm-analyst-year sample consists of 160,332 firm-analyst-year observations, representing 2,471 unique firms followed by 4,096 analysts. In Panel C, our analyst-executive firm-year sample consists of 8,369 firm-year observations, representing 1,581 unique firms over the period 2004-2020. In Panel D, our analyst-employee firm-year sample consists of 5,216 firm-year observations, representing 909 unique firms over the period 2008-2020.

Panel A: Correlation matrix for Table 4 Panel A

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Culture discussion	1.000															
Firm size	0.282	1.000														
Firm age	0.035	0.325	1.000													
Sales growth	0.034	-0.069	-0.187	1.000												
ROA	0.123	0.067	0.050	0.149	1.000											
Leverage	0.017	0.249	0.101	-0.057	-0.129	1.000										
Tangibility	-0.059	0.159	0.194	-0.098	-0.048	0.254	1.000									
ROA volatility	-0.124	-0.265	-0.129	0.037	-0.391	0.002	0.009	1.000								
Large institutional ownership	-0.095	-0.239	-0.059	-0.004	-0.039	0.026	-0.106	-0.017	1.000							
Board independence	-0.202	-0.384	-0.055	0.026	-0.134	-0.016	-0.034	0.041	0.234	1.000						
CEO duality	0.021	0.117	0.080	-0.015	0.044	-0.015	0.044	-0.104	-0.068	-0.077	1.000					
Number of key people changes	0.183	0.321	0.127	-0.069	-0.080	0.031	-0.005	-0.014	-0.101	-0.074	-0.051	1.000				
Number of M&As	0.133	0.377	0.078	0.023	0.001	0.108	-0.031	-0.052	-0.151	-0.187	0.032	0.189	1.000			
Strong culture	0.073	-0.153	-0.189	0.082	-0.019	-0.156	-0.193	0.061	-0.010	-0.018	-0.037	0.070	0.006	1.000		
Number of meetings	0.185	0.340	0.049	0.062	0.056	0.126	-0.060	-0.040	0.017	-0.142	-0.018	0.115	0.168	0.060	1.000	
Employee culture rating	0.069	0.053	0.060	-0.006	0.007	0.109	0.003	-0.056	0.202	0.134	-0.086	-0.008	-0.022	0.039	0.343	1.000
Number of employee reviews	0.320	0.466	0.156	-0.067	0.108	0.111	-0.061	-0.156	0.016	-0.198	0.004	0.257	0.200	0.128	0.423	0.457

Panel B: Correlation matrix for Table 4 Panel C

Variable	1	2	3	4	5	6	7	8	9	10	11
Star analyst	1.000										
CFA	0.022	1.000									

Postgraduate	0.006	-0.064	1.000								
Female	-0.021	-0.030	-0.055	1.000							
General experience	0.115	0.037	0.082	-0.053	1.000						
Firm experience	0.114	0.008	0.037	-0.037	0.600	1.000					
Number of industries followed	0.041	0.052	0.017	-0.008	0.152	0.069	1.000				
Number of firms followed	0.121	0.064	0.078	-0.070	0.288	0.170	0.388	1.000			
Forecast frequency	0.118	0.036	0.015	0.011	0.048	0.132	-0.031	0.113	1.000		
Ln(Broker size)	0.227	0.037	-0.079	0.037	-0.021	0.018	-0.068	0.084	0.114	1.000	
Ln(Analyst-firm distance + 1)	-0.003	-0.010	-0.007	-0.014	-0.043	-0.044	-0.041	-0.026	-0.014	0.046	1.000

Panel C: Correlation matrix for Table 5 Panel A (Analyst-Executive)

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Type divergence	1.000																	
Cause divergence	0.221	1.000																
Effect divergence	0.268	0.320	1.000															
Firm size	-0.045	-0.067	-0.002	1.000														
Ln(Firm age + 1)	-0.066	-0.044	-0.023	0.393	1.000													
Sales growth	0.054	0.019	-0.008	-0.083	-0.226	1.000												
ROA	-0.002	-0.056	-0.094	-0.052	0.001	0.111	1.000											
Leverage	-0.034	-0.021	-0.004	0.118	0.094	-0.057	-0.093	1.000										
Tangibility	-0.068	-0.002	0.041	-0.009	0.131	-0.064	0.023	0.234	1.000									
ROA volatility	0.024	0.062	0.049	-0.235	-0.068	-0.037	-0.332	0.067	0.084	1.000								
Large institutional ownership	0.008	0.018	-0.023	-0.251	-0.125	-0.006	-0.076	0.070	-0.043	0.014	1.000							
Board independence	0.005	0.037	0.028	-0.319	-0.110	0.047	-0.156	0.008	-0.047	0.042	0.218	1.000						
Loss year	0.009	0.043	0.061	-0.137	-0.048	-0.106	-0.618	0.094	0.054	0.408	0.064	0.071	1.000					
CEO duality	-0.007	-0.005	0.033	0.161	0.128	-0.015	0.048	-0.001	0.033	-0.099	-0.055	-0.076	-0.089	1.000				
CEO tenure	0.043	0.023	0.072	-0.034	-0.034	0.056	0.041	-0.063	-0.052	-0.101	-0.002	-0.001	-0.078	0.358	1.000			
CEO Delta	0.001	-0.063	-0.048	0.326	0.093	0.091	0.320	0.038	-0.031	-0.195	-0.179	-0.299	-0.273	0.295	0.412	1.000		
CEO Vega	-0.058	-0.079	-0.082	0.212	0.135	-0.047	0.169	0.044	-0.011	-0.054	-0.156	-0.219	-0.111	0.162	-0.027	0.499	1.000	
CEO close-to-retire	-0.010	-0.013	0.058	0.075	0.054	-0.018	-0.009	-0.003	0.014	-0.042	0.000	0.000	-0.031	0.150	0.357	0.148	-0.004	1.000
Ln(Number of meetings + 1)	-0.010	-0.067	-0.102	0.314	0.131	0.009	0.051	0.146	-0.060	-0.038	0.086	-0.058	-0.056	-0.009	-0.004	0.198	0.044	0.041

Panel D: Correlation matrix for Table 5 Panel A (Analyst-Employee)

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Type divergence	1.000																	
Cause divergence	0.174	1.000																
Effect divergence	0.165	0.329	1.000															
Firm size	-0.163	-0.214	-0.126	1.000														
Ln(Firm age + 1)	-0.009	-0.063	0.002	0.376	1.000													
Sales growth	-0.010	0.016	-0.004	-0.086	-0.232	1.000												
ROA	0.001	-0.015	0.001	-0.045	0.010	0.095	1.000											
Leverage	0.018	-0.028	0.017	0.173	0.114	-0.091	-0.129	1.000										
Tangibility	0.044	0.087	0.090	0.083	0.147	-0.086	0.003	0.207	1.000									
ROA volatility	0.001	0.016	-0.010	-0.188	-0.072	0.007	-0.280	0.042	0.038	1.000								
Large institutional ownership	0.119	0.027	0.027	-0.290	-0.125	-0.002	-0.099	0.079	-0.086	0.008	1.000							
Board independence	0.093	0.157	0.106	-0.351	-0.096	0.018	-0.175	0.045	-0.026	0.014	0.263	1.000						
Loss year	0.009	0.025	-0.013	-0.139	-0.068	-0.067	-0.625	0.094	0.037	0.394	0.083	0.091	1.000					
CEO duality	-0.008	0.016	0.044	0.154	0.122	-0.026	0.030	-0.024	0.060	-0.126	-0.057	-0.029	-0.097	1.000				
CEO tenure	-0.029	0.021	0.034	-0.097	-0.091	0.070	0.072	-0.095	-0.024	-0.076	-0.044	-0.038	-0.066	0.342	1.000			
CEO Delta	-0.075	-0.135	-0.057	0.319	0.047	0.114	0.284	-0.015	-0.033	-0.173	-0.233	-0.322	-0.249	0.300	0.454	1.000		
CEO Vega	-0.030	-0.064	-0.036	0.226	0.145	-0.049	0.128	0.050	-0.039	-0.045	-0.174	-0.220	-0.095	0.151	-0.034	0.466	1.000	
CEO close-to-retire	-0.058	-0.012	0.026	0.046	0.014	-0.009	-0.001	-0.003	0.068	-0.014	-0.020	-0.022	-0.020	0.163	0.393	0.183	-0.030	1.000
Ln(Number of meetings + 1)	-0.081	-0.155	-0.128	0.280	0.075	0.044	0.028	0.152	-0.080	-0.013	0.073	-0.084	-0.046	-0.005	-0.015	0.166	0.040	0.015

Table IA9
Explaining divergence among analysts of their perspectives on corporate culture

This table examines the determinants of divergence among analysts of their perspective on corporate culture. Panel A examines the relationships between firm characteristics and divergences among analysts about culture types (their causes, effects, or tone). Our firm-year sample consists of 13,023 firm-year observations, representing ??? unique firms over the period 2000-2020. The dependent variable, Type divergence, is the firm-year average of each individual analyst following the focal firm, her number of culture types mentioned minus the average of other fellow analysts' number of culture types mentioned in their reports. Other divergence measures are defined analogously. Panel B examines the relationships between analyst characteristics and divergences among analysts about culture types (their causes, effects, or tone). Our firm-analyst-year sample consists of 19,741 firm-analyst-year observations, representing ??? firm-year observations and ??? unique firms over the period 2000-2020. The dependent variable, Type divergence, for each individual analyst following the focal firm, is her number of culture types mentioned minus the average of other fellow analysts' number of culture types mentioned in their reports. Other divergence measures are defined analogously. Panel C examines the relationships between analyst characteristics and the differences among analysts regarding a specific culture type (tone used when discussing culture). The dependent variable, Collaboration and people-focused, for each individual analyst following the focal firm, is her number of times discussing this culture type minus the average number of times by other fellow analysts discussing the same culture type. Other divergence measures are defined analogously. Industry fixed effects are based on Fama-French 12-industry classifications. Definitions of the variables are provided in the Appendix. Heteroskedasticity-consistent standard errors (in parentheses) are clustered at the firm level. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A: Firm characteristics and divergences among analysts

	Analyst-Analyst			
	Type divergence	Cause divergence	Effect divergence	Tone divergence
Variable	(1)	(2)	(3)	(4)
Firm size	0.004**	0.001	0.006***	0.016***
	(0.002)	(0.002)	(0.002)	(0.002)
Ln(Firm age + 1)	-0.000	-0.002	-0.000	-0.005
	(0.004)	(0.004)	(0.003)	(0.005)
Sales growth	0.005	0.002	0.007	-0.013
	(0.010)	(0.012)	(0.008)	(0.012)
ROA	0.032	-0.077*	-0.006	0.025
	(0.033)	(0.041)	(0.030)	(0.044)
Leverage	0.002	-0.029*	-0.000	0.006
	(0.011)	(0.016)	(0.010)	(0.015)
Tangibility	0.013	0.003	0.012	0.021
	(0.012)	(0.015)	(0.011)	(0.015)
ROA volatility	0.070	-0.085	-0.024	0.241***
	(0.047)	(0.065)	(0.047)	(0.059)
Large institution ownership	0.003	-0.007	-0.010	-0.004
	(0.008)	(0.010)	(0.007)	(0.010)
Board independence	0.006	-0.014	0.011	-0.025
	(0.019)	(0.024)	(0.017)	(0.024)
Loss year	0.012	-0.010	0.017**	0.037***
	(0.008)	(0.010)	(0.007)	(0.009)
CEO duality	-0.002	0.005	-0.002	-0.007
	(0.004)	(0.005)	(0.004)	(0.005)

CEO tenure	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
CEO Delta	-0.000	-0.000	-0.004**	-0.007***
	(0.002)	(0.002)	(0.002)	(0.002)
CEO Vega	0.002*	-0.000	-0.000	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
CEO close-to-retire	0.011**	-0.005	0.007*	0.013**
	(0.005)	(0.006)	(0.004)	(0.006)
Ln(Number of meetings + 1)	-0.004	0.008*	0.000	-0.006
	(0.004)	(0.004)	(0.003)	(0.005)
Constant	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Adjusted R-squared	0.006	0.007	0.012	0.019
Observations	13,023	7,723	9,941	13,023

Panel B: Analyst characteristics and divergences among analysts

	Analyst-Analyst			
	Type divergence	Cause divergence	Effect divergence	Tone divergence
Variable	(1)	(2)	(3)	(4)
Star analyst	-0.002	-0.001	0.004	0.001
	(0.004)	(0.007)	(0.005)	(0.005)
CFA	0.001	0.003	-0.001	0.000
	(0.002)	(0.003)	(0.002)	(0.002)
Postgraduate	-0.001	0.005	-0.001	-0.006**
	(0.002)	(0.003)	(0.002)	(0.002)
Female	-0.003	0.011*	0.005	0.001
	(0.004)	(0.006)	(0.004)	(0.004)
General experience	-0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Firm experience	-0.000	-0.001*	0.000	0.000
	(0.000)	(0.001)	(0.000)	(0.000)
Number of industries followed	-0.001	0.001	-0.000	-0.002***
	(0.001)	(0.001)	(0.001)	(0.001)
Number of firms followed	0.000	-0.000	0.000	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Forecast frequency	-0.001	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Ln(Broker size)	-0.008	0.007	-0.003	0.002
	(0.007)	(0.010)	(0.007)	(0.008)
Ln(Number of meetings + 1)	-0.001	0.002	-0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Constant	YES	YES	YES	YES
Firm × Year FE	YES	YES	YES	YES
Broker FE	YES	YES	YES	YES
Adjusted R^2	0.246	0.207	0.240	0.361

No. of observations	19,741	7,599	12,331	19,741

Panel C: Analyst characteristics and divergences among analysts viewing a specific culture type or tone

	Analyst-Analyst Difference								
	Collaboration and People-focused	Customer -oriented	Innovation and adaptability	Integrity and risk management	Performance -oriented	Misc.	Cause	Effect	Tone
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Star analyst	0.013	-0.009	0.008	-0.010	-0.003	0.003	-0.017	-0.018	-0.022
	(0.015)	(0.009)	(0.016)	(0.010)	(0.015)	(0.009)	(0.020)	(0.033)	(0.023)
CFA	-0.001	0.005	-0.005	-0.005	0.003	0.003	-0.005	0.004	-0.009
	(0.006)	(0.004)	(0.005)	(0.004)	(0.006)	(0.003)	(0.007)	(0.012)	(0.009)
Postgraduate	0.000	-0.002	-0.015**	-0.000	0.013*	0.004	-0.004	0.011	0.006
	(0.006)	(0.004)	(0.007)	(0.004)	(0.007)	(0.004)	(0.008)	(0.014)	(0.010)
Female	0.036***	0.002	0.001	-0.008	-0.018	-0.012*	0.005	-0.009	0.025
	(0.012)	(0.008)	(0.012)	(0.007)	(0.013)	(0.006)	(0.015)	(0.026)	(0.019)
General experience	-0.000	0.000	0.000	0.000	0.000	-0.001**	0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.002)	(0.001)
Firm experience	-0.000	-0.001	-0.001	0.000	0.001	0.000	-0.000	-0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Number of industries followed	-0.004*	0.001	0.006**	-0.002	-0.000	-0.000	0.003	-0.002	0.010***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.001)	(0.003)	(0.006)	(0.004)
Number of firms followed	0.001**	-0.000	-0.001	-0.000	-0.001	0.000	-0.001	-0.003*	-0.002**
	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.002)	(0.001)
Forecast frequency	0.001	-0.001	0.002	0.000	-0.002	-0.001	-0.001	-0.002	-0.001
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.004)	(0.003)
Ln(Broker size)	0.005	0.000	0.004	0.009	0.002	-0.021	0.005	0.030	0.021
	(0.021)	(0.014)	(0.021)	(0.012)	(0.022)	(0.013)	(0.029)	(0.049)	(0.033)
Ln(Number of meetings + 1)	0.000	-0.002	0.004	-0.002	0.001	-0.001	-0.007*	-0.001	0.001
	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.004)	(0.006)	(0.005)
Constant	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm × Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Broker FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adjusted R^2	0.024	0.027	0.034	0.084	0.041	0.027	0.019	0.025	0.088
No. of observations	19,719	19,719	19,719	19,719	19,719	19,719	19,719	19,719	19,719