

Essays on Corporate Culture

Ashok Kumar

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for the
Doctorate in Philosophy degree in Digital Transformation and Innovation

Telfer School of Management
University of Ottawa

© Ashok Kumar, Ottawa, Canada, 2026

Abstract

In this dissertation, I examine important questions in the measurement of organizational culture using computational textual analysis of corporate earnings conference calls. The research topics include contextual measurement of corporate integrity using a multi-agent LLM architecture in the first essay and measuring organizational culture through the Competing Values Framework using the same methodology in the second essay. Both of my essays are co-authored with Professor Shantanu Dutta with first essay currently under a revise-and-resubmit decision at Information Systems Frontier.

The first essay addresses a fundamental measurement challenge in the corporate culture literature: dictionary-based approaches to measuring corporate integrity from corporate disclosures count every occurrence of culture-related keywords as evidence of the underlying construct, regardless of the context in which those words appear. We develop and validate a multi-agent large language model (LLM) pipeline, based on a blackboard architecture with eleven specialized agents, to measure four components of corporate integrity (i.e., Accountability, Compliance, Moral Integrity, and Honesty) from the Q&A sections of earnings conference call transcripts. Our contextual approach reveals that fewer than 30% of the segments containing integrity-related dictionary terms are meaningfully related to the integrity construct, demonstrating that traditional dictionary methods substantially overcount meaningful cultural discourse. The multi-agent classification system achieves 95.67% accuracy and a Cohen's kappa of 0.913 against human annotations, with five independent stability runs producing an inter-run mean kappa of 0.945. To validate the resulting measure empirically, we examine its association with established indicators of corporate governance quality and reporting integrity. We find that firms with greater board independence and higher female board representation engage in more integrity-related discourse, consistent with prior research linking these governance characteristics to stronger ethical oversight. We also find that firms with higher contextual integrity scores are more likely to disclose non-fraudulent restatements, consistent with the view that integrity-oriented cultures promote transparent correction of reporting problems rather than increasing the incidence of misreporting. Furthermore, higher integrity scores are associated with lower discretionary accruals, higher dividend payouts, and greater access to trade credit, in line with theoretical predictions that integrity culture constrains opportunistic reporting, enhances credibility with capital providers, and reduces counterparty risk. Across these analyses, the dictionary-based measure exhibits systematically weaker and, in several cases, statistically insignificant associations, underscoring that contextual interpretation is not merely methodological refinement but a necessary step for recovering economically meaningful integrity signals from corporate discourse.

The second essay implements the methodological innovations of the first essay to the Competing Values Framework (CVF), the most widely adopted theoretical model of organizational culture. Despite its theoretical maturity, empirical measurement of the CVF has

relied predominantly on dictionary-based word counts from 10-K filings, an approach that is both context-insensitive and confined to heavily curated corporate documents. We adapt the multi-agent blackboard architecture to classify earnings call Q&A discourse into the four CVF cultural orientations (i.e., Clan, Adhocracy, Hierarchy, and Market) using over 183,000 transcripts spanning 2008 to 2024. Our classification identifies approximately 20% of total segments as containing genuine cultural content, with Market discourse predominating, followed by Hierarchy, Adhocracy, and Clan, a distribution that reflects the externally oriented communicative demands of the earnings call setting. We conduct the first systematic construct-validation exercise that pairs each CVF dimension with the specific firm outcomes its own theoretical logic predicts, and we benchmark our contextual scores against the widely used dictionary-based measures on the same validation outcomes. Our contextual CVF scores produce theoretically consistent and statistically significant associations across all four dimensions (i.e., Clan with labor intensity, Adhocracy with innovation, Market with profitability, and Hierarchy with regulatory discipline) whereas the dictionary-based scores yield inconsistent and often economically negligible results. These findings demonstrate that all four dimensions of the CVF carry distinct economic content when measured with sufficient contextual precision, and that the apparent irrelevance of certain CVF dimensions in prior work likely reflects measurement noise rather than a genuine absence of cultural influence on firm behavior.

Keywords: *Corporate culture, Corporate integrity, Competing Values Framework, Contextual Measure, Earnings conference calls, Artificial Intelligence (AI) · Generative AI (GenAI) · Large Language Models (LLM), Multi-agent systems, Textual analysis, Dictionary methods, Organizational culture measurement*

Dedicated to my mom and dad

Acknowledgements

Completing this dissertation has been one of the most challenging and rewarding experiences of my life, and I certainly did not do it alone. There are many people whose support, kindness, and guidance helped me get to this point, and I want to take a moment to thank them.

First, I would like to thank my supervisor, Professor Shantanu Dutta, for his steady guidance and unwavering support throughout my PhD. He helped me think more clearly and practically, do better work, and grow as a researcher. His feedback was always thoughtful, and he was remarkably generous with his time throughout my PhD journey, for which I am truly grateful. Thank you, Professor, for being not just an excellent supervisor but also a true mentor. The conversations we had and lessons I learned from you go well beyond this dissertation.

To my mom and dad – thank you for everything. You may never fully know or understand what exactly your son has been doing all these years, but honestly, that never mattered. What mattered was that you were always there, believing in me without question. Your hard work and dedication in your own lives have always been my biggest inspiration. Watching you push through difficulties taught me more about perseverance than any book ever could. This achievement is not just mine it is yours too. I am grateful for every sacrifice you made and for giving me more than I could ever ask for.

I would also like to sincerely thank my committee members, Professor Greg Richards and Professor Harshit Rajaiya, for their time and valuable input. Their insightful questions and feedback during my proposal defense pushed me to look at my work from different angles and made this dissertation much stronger than it would have been otherwise.

Lastly, I want to thank all the friends and colleagues who were part of this journey. Whether it was a helpful conversation, a word of encouragement on a tough day, or just being there to share a laugh, you all made this experience so much more memorable. I am lucky to have had such wonderful people around me.

Table of Contents

Essay 1

Developing a Contextual Corporate Integrity Measure Using Large Language Models

1. Introduction.....	2
2. Literature Review.....	8
2.1 The Evolution of Textual Analysis in Management Research.....	8
2.2 Corporate Integrity Culture and Its Components.....	10
2.2.1 Accountability, Governance, and Ownership (Accountability).....	12
2.2.2 Regulatory Compliance and Adherence (Compliance).....	13
2.2.3 Moral Integrity and Fairness (Moral Integrity).....	14
2.2.4 Transparency, Trust, and Honesty (Honesty).....	16
2.3 Earnings Calls and the Measurement of Corporate Integrity.....	17
2.3.1 From Manual to AI-Driven Earnings Call Analysis.....	17
2.3.2 Measurement of Corporate Integrity Culture Using Textual Analysis.....	18
3. Methodology.....	19
3.1 Data Collection and Sample Definition.....	19
3.2 Text Preprocessing and Data Preparation.....	21
3.3 LLM-Based Classification.....	22
3.3.1 Model Selection.....	22
3.3.2 Multi-Agent Classification Architecture.....	24
3.3.3 Addressing LLM Concerns.....	27
3.4 Contextual Corporate Integrity Score Construction.....	28
4. Results.....	29
4.1 Overall Integrity Detection Performance and Validation.....	29
4.1.1 Validation Design for the Phase I Integrity Detection Gate.....	30
4.1.2 Classification Performance of the Phase I Integrity Detection Gate.....	30
4.1.3 Prediction Stability of the Phase I Integrity Detection Gate.....	31
4.2 Contextual Understanding versus Dictionary-Based Approaches.....	32
4.2.1 Beyond Keyword Matching.....	33
4.2.2 Dictionary Word Distribution Analysis.....	33
4.3 Contextual Integrity Measure Validation: Associations with Recognized Corporate Integrity Indicators.....	35
4.3.1 Corporate Governance and Corporate Integrity Culture.....	35

4.3.2 Corporate Integrity Culture and Accounting Restatement Disclosure	37
4.3.3 Corporate Integrity Culture and Earnings Management	40
4.4 Implications of Corporate Integrity Culture: Dividend Policy and Trade Credit	41
4.4.1 Corporate integrity culture and dividend policy	42
4.4.2 Corporate integrity culture and trade credit	44
5. Discussion	46
5.1 Methodological Contributions	46
5.2 Theoretical Contributions	48
5.3 Practical Implications.....	50
6. Conclusion	52
Appendix A: Variable description, Descriptive Statistics and Supplementary Discussions.....	54
Table A1. Variable description	54
Table A2. Descriptive Statistics.....	55
Supplementary Discussions	55
SD1. Validation of the BGE-M3 segmentation variant against the C99 TF-IDF baseline	55
SD2. Integrity Components Identification	60
SD3. Rationale and Design Trade-Offs of the Phase I Gating Mechanism	60
References.....	62

Essay 2

Beyond Keywords: Measuring Corporate Culture Through the Competing Values Framework Using Large Language Models

1. Introduction.....	105
2. Literature Review.....	113
2.1 The Evolution of Textual Analysis in Management Research.....	113
2.2 The Competing Values Framework: Theoretical Foundations and Components	115
2.2.1 Origins and Development of the CVF	116
2.2.2 Integration of Cultural Values from Prior Literature with CVF Dimensions	118
2.2.3 Clan (Collaborate) Culture.....	119
2.2.4 Adhocracy (Create) Culture	120
2.2.5 Hierarchy (Control) Culture.....	121
2.2.6 Market (Compete) Culture	123
2.3 Earnings Calls and the Measurement of Organizational Culture.....	124

2.3.1 Earnings Calls as a Data Source for Cultural Analysis	124
2.3.2 From Manual to AI-Driven Analysis of Earnings Call Discourse	130
2.3.3 Measurement of CVF Culture Using Textual Analysis	131
2.4 Research Gap and Motivation.....	133
3. Methodology	136
3.1 Data Collection and Sample Definition	136
3.2 Text Preprocessing and Data Preparation	138
3.3 LLM-Based Classification	140
3.3.1 Model Selection	140
3.3.2 Multi-Agent Classification Architecture.....	142
3.3.3 Addressing LLM Concerns.....	146
3.4 Contextual CVF Culture Score Construction.....	148
4. Classification Performance	149
4.1 Overall CVF Detection Performance and Validation	150
4.1.1 Validation Design for the Phase I CVF Detection Gate	150
4.1.2 Classification Performance of the Phase I CVF Detection Gate.....	151
4.1.3 Prediction Stability of the Phase I CVF Detection Gate	152
4.2 Contextual Understanding versus Dictionary-Based Approaches	153
4.2.1 Beyond Keyword Matching.....	154
4.2.2 Dictionary Word Distribution Analysis	155
5. Validation of the Four CVF Dimensions	157
5.1 Collaborate culture and labor intensity	158
5.1.1 Employees, labor-related spending, and collaborate culture.....	158
5.1.2 Empirical evidence.....	159
5.2 Compete culture and product-market performance.....	160
5.2.1 Profitability, market share, and compete culture	160
5.2.2 Empirical evidence.....	161
5.3 Create culture and corporate innovation	162
5.3.1 Innovation input, innovation output, and create culture	162
5.3.2 Empirical evidence.....	163
5.4 Control culture and tax and reporting discipline.....	164
5.4.1 Cash ETR, error-restatement disclosure (restatement filing), and control culture.....	164
5.4.2 Empirical evidence.....	165
5.5 Overall assessment.....	166
6. Discussion.....	166

6.1 Theoretical Contributions	167
6.2 Methodological Contributions	171
6.3 Practical Implications.....	173
6.4 Limitations and Future Research	175
7. Conclusion	177
Appendix A: Variable Definitions	180
Table A1. Variable Definitions.....	180
References.....	181

List of Figures

Essay 1

Figure 1. Four components of corporate integrity	71
Figure 2. Process model.....	71
Figure 3. Multi agent integrity culture classification pipeline	72

Essay 2

Figure 1. Competing Values Framework (CVF)	188
Figure 2. Mapping Cultural Values from Prior Literature to CVF Dimensions	189
Figure 3. Process Model	189
Figure 4. Multi agent organizational culture classification pipeline.....	190

List of Tables

Essay 1

Table 1: Classification performance of the Phase I integrity detection gate.....	73
Table 2a: Per-class metrics for the Phase I integrity detection gate.....	73
Table 2b: Confusion matrix (main run)	73
Table 3: Inter-run agreement (Cohen’s κ) – DeepSeek-R1	73
Table 4: Examples of False Keyword Signals	74
Table 5: Integrity Term Classification Rates	75
Table 6: Corporate governance and corporate integrity culture.....	76
Table 7: Corporate integrity and non-fraudulent restatement disclosure.....	77
Table 8: Corporate integrity and earnings management	78
Table 9: Corporate integrity culture and dividend policy	79
Table 10: Corporate integrity culture and trade credit	80

Essay 2

Table 1: Distribution of CVF Components.....	191
Table 2: Classification performance of the Phase I overall CVF culture detection gate	191
Table 3a: Per-class metrics for the Phase I CVF culture detection gate	191
Table 3b: Confusion matrix (main run)	191
Table 4: Inter-run agreement (Cohen’s κ) – DeepSeek-R1	192
Table 5: Examples of False Keyword Signals	193
Table 6: CVF Term Classification Rates	194
Table 7. Descriptive Statistics.....	195
Table 8. Validation of the Collaborate Dimension	196
Table 9. Validation of the Compete Dimension.....	197
Table 10. Validation of the Create Dimension.....	198
Table 11. Validation of the Control Dimension.....	199

List of Appendix B

Essay 1

Appendix B: Additional Analysis Material	81
Table B1: Paired-sample t-tests of BGE-M3 vs. TF-IDF segmentation	81
Table B2: Benchmark performance of candidate models	81
Table B3: Confusion matrices.....	82
Table B4: Pairwise Inter-Run Agreement.....	82
Table B5: Yes Segment % for Integrity Wordlist.....	83
Table B6: Distribution of Integrity Components	90
Table B7: Examples of Integrity Components Identified in the Earnings Call Discourse.....	91
Table B8: Corporate governance and corporate integrity culture	99
Table B9: Corporate integrity and non-fraudulent restatement disclosure	100
Table B10: Corporate integrity and earnings management.....	101
Table B11: Corporate integrity culture and dividend policy	102
Table B12: Corporate integrity culture and trade credit	103

Essay 2

Appendix B: Additional Analysis Material	200
Table B1: Benchmark performance of candidate models	200
Table B2: Examples of CVF Components Identified in the Earnings Call Discourse	201
Table B3: Confusion matrices.....	209
Table B4: Pairwise Inter-Run Agreement.....	209
Table B5: Yes Segment % for CVF Wordlist (based on Fang et al., 2023)	210

Essay 1

Developing a Contextual Corporate Integrity Measure Using Large Language Models

Abstract

This study develops a new Contextual Corporate Integrity Score using large language models (LLMs) to measure firms' integrity culture more accurately. We analyze the unscripted question-and-answer sections of 183,043 quarterly earnings call transcripts from 2008 to 2024, where executives respond spontaneously to analysts. Unlike traditional dictionary-based methods that rely on keyword counts, our approach interprets language in context and identifies genuine integrity-related discourse. The score captures four dimensions of integrity: Accountability (responsibility and governance), Compliance (adherence to laws and rules), Moral Integrity (fairness and ethical conduct), and Honesty (transparency, trust, and truthful communication). These components are combined into an overall integrity measure. We validate the score in two ways. First, the model shows high agreement with human judgments, indicating strong reliability. We also find that over 70% of text segments containing potential integrity-related keywords (suggested in earlier studies) do not actually discuss integrity when context is considered. Second, the score is meaningfully associated with important firm outcomes. Firms with higher integrity scores tend to have more independent boards and greater female board representation, are more likely to disclose and correct accounting issues transparently, have lower discretionary accruals, pay higher dividends, and enjoy greater access to trade credit. Overall, the findings show that contextual AI-based measures provide a more informative view of corporate culture than traditional keyword methods. The measure offers practical value for investors, boards, regulators, and researchers seeking to monitor corporate integrity and understand its effects on governance, reporting quality, and financial decisions.

1. Introduction

Corporate culture—often defined as the shared values and norms that shape behavior within organizations—has long been recognized as a critical driver of ethical conduct and firm value (Guiso et al., 2015; Graham et al., 2022). Within this broad domain, integrity stands out as a core component of corporate culture—encompassing honesty, consistency in word and deed, and adherence to moral and regulatory standards—and is widely regarded as “quite possibly the most commonly cited morally desirable trait” in corporate values (Audi & Murphy, 2006, p. 3). Firms that foster an integrity-centered culture tend to enjoy greater stakeholder trust and sustainable success, whereas lapses in integrity can precipitate reputational damage and financial misconduct (Guiso et al., 2015).

Measuring integrity-related culture, however, presents a significant empirical challenge. Unlike tangible governance mechanisms or financial metrics, culture is intangible and typically observable only through its manifestations in language and behavior (Graham et al., 2022; Kreps, 1990). Early research proxied cultural values using surveys and employee feedback, with Graham et al. (2022) gauging cultural priorities through executive interviews and other studies inferring ethical climate from employee reviews such as Glassdoor ratings (Ji et al., 2017). These approaches yield valuable insights but cover limited samples, are subject to self-report bias, and may not capture the tone at the top set by senior management. An alternative strand of literature turned to textual analysis of corporate communications as a window into corporate culture (Fang et al., (2023); Fiordelisi & Ricci, 2014). However, earlier textual approaches primarily relied on a bag-of-words methodology, which lacks contextual sensitivity and may introduce subjective judgment bias (Hoitash & Hoitash, 2022).

Advances in natural language processing (NLP) have enabled more context-aware measurement, a direction that aligns with the IS discipline's longstanding emphasis on developing computational methods that transform unstructured data into decision-relevant signals. While traditional approaches often relied on raw word counts, newer methods apply machine learning models (such as Word2Vec) to uncover semantic relationships in corporate texts. For example, an influential and well-executed article by Li et al. (2021) still rely on word frequency analysis but enhance it by using Word2Vec-trained embeddings to inform the selection and grouping of words based on their semantic proximity, thereby providing a more content-sensitive foundation for dictionary-based analysis. Although such semi-supervised machine learning methods represent a methodological improvement, they still do not fully address the contextuality challenges inherent in language understanding and may misclassify narratives (i.e., sentences or segments).

These methodological limitations point to a deeper concern. The reliability of existing large-scale, text-based integrity measures (such as dictionary-based and word-level embedding approaches) rests on a common underlying assumption: if a term from an integrity lexicon appears in corporate communication text, its occurrence is counted as genuine evidence of an integrity-oriented culture presence. Yet the same words can carry very different meanings depending on context, and without the ability to distinguish meaningful integrity discourse from incidental word use, widely used dictionary-based proxies may be recovering something other than the cultural construct researchers intended. If this is the case, prior findings built on these proxies may understate or mischaracterize the economic significance of corporate integrity culture. This concern elevates the measurement problem from a purely methodological one to a phenomenon-level question: does a genuinely contextual approach to measuring integrity lead to

a different, and more empirically meaningful, understanding of corporate integrity culture and its implications for firm behavior? An LLM-driven, contextually grounded integrity measure is uniquely suited to investigate this question, because it can help us evaluate whether the linguistic markers that existing approaches rely on actually co-occur with meaningful cultural content. Yet there is no systematic application of LLM in the context of measuring corporate integrity. There is also an important gap in existing integrity measures that goes beyond methodological refinement. Corporate integrity is inherently multidimensional, encompassing accountability, compliance, moral commitment, and transparent communication. Therefore, a robust measure should explicitly capture this multidimensionality and be developed accordingly.

We address these gaps by developing a Contextual Corporate Integrity Score derived from the unscripted question-and-answer (Q&A) portion of quarterly earnings calls, using a reasoning large language model (LLM) embedded in a multi-agent classification architecture. Earnings call Q&A sessions are a uniquely informative source for this purpose: unlike prepared remarks, the Q&A portion exposes executives to unanticipated questions from analysts, producing spontaneous responses that are widely regarded as more revealing of managerial attitudes, tone at the top, and organizational values than rehearsed disclosures (Li et al., 2021; Price et al., 2012). We assemble 183,043 earnings call transcripts from 2008 to 2024, extract the executive responses in the Q&A portion, segment them into coherent discourse units using a semantically aware adaptation of the C99 algorithm, and route each segment through a multi-agent pipeline inspired by blackboard architecture (Salemi et al., 2025). Our classification pipeline begins by assigning each segment a binary label indicating the presence of integrity-related discourse. Segments identified as relevant are then analyzed by domain-specific agents aligned with a four-component integrity framework: Accountability, Governance, and Ownership

(hereafter Accountability); Regulatory Compliance and Adherence (hereafter Compliance); Moral Integrity and Fairness (hereafter Moral Integrity); and Transparency, Trust, and Honesty (hereafter Honesty). Aggregating segment-level classifications to the firm-quarter level produces an overall integrity measure and four component-specific measures that can be tracked through time.

We validate the measure in two complementary ways. First, we evaluate the classification performance of the overall integrity detection against a human-annotated set of 300 segments, drawn through a stratified design that ensures balanced representation of the integrity-related and non-integrity-related classes. The classifier achieves 95.67% accuracy, a macro-F1 of 95.67%, and Cohen's kappa of 0.913 against human annotations, with five independent stability runs producing an inter-run mean kappa of 0.945. Comparison against the dictionary-based specification of Li et al. (2021) further shows that keyword presence is a weak indicator of genuine integrity discourse: overall, fewer than 30% of the segments containing dictionary terms are identified by our pipeline as substantively related to corporate integrity culture, indicating that for more than 70% of the time these terms appear in contexts that bear no relation to integrity culture. Qualitative inspection corroborates this finding, showing that dictionary terms frequently surface in technical, organizational, or otherwise incidental contexts. This evidence directly tests and challenges the prevailing assumption that keyword presence alone reliably signals meaningful cultural discourse.

Second, we validate the resulting contextual measure empirically by examining its association with established indicators of corporate governance quality and reporting integrity. We find that firms with greater board independence and higher female board representation—two governance characteristics that prior research has linked to stronger ethical oversight

(Beasley, 1996; Sun et al., 2021)—engage in more integrity-related discussions. We also find that firms with higher corporate integrity scores are more likely to disclose non-fraudulent restatements, consistent with the view that integrity-oriented cultures strengthen internal detection and promote transparent correction of reporting problems, rather than increasing the underlying incidence of misreporting (Ahluwalia et al., 2018; Kaptein, 2011). Furthermore, we document that higher integrity scores are associated with lower discretionary accruals, higher dividend payouts, and greater access to trade credit, in line with theoretical predictions that integrity culture constrains opportunistic reporting, enhances credibility with capital providers, and reduces counterparty risk in supply relationships (Hasan & Habib, 2019; Jensen, 1986; Sinha & Damle, 2025). Across these analyses, the Li et al. (2021) dictionary-based or semi-supervised specification exhibits systematically weaker and, in several cases, statistically insignificant associations. This finding underscores that contextual interpretation is not only a methodological refinement but a necessary step for recovering economically meaningful integrity signals from corporate discourse.

Our study makes three contributions. First, we advance the IS literature at the intersection of unstructured data analytics and behavioral research by introducing an interpretable, multi-agent LLM pipeline for large-scale context-sensitive textual analysis of corporate communications. This contribution builds on prior IS work that has applied sequential neural networks to financial risk prediction (Yang et al., 2023), topic modeling to organization-generated content on electric vehicles (Sharma et al., 2025), and a range of NLP methods to social media discourse (Bhatt et al., 2023; Kaur et al., 2021; Kurumathur et al., 2025; Singh et al., 2022). By embedding a reasoning LLM within a multi-agent architecture, we show how generative AI can function not only as a classification tool but as a disciplined reasoning

framework that separates evidence extraction from contextual judgment, responding to Storey et al.'s (2025) call for IS research that links generative AI innovation to organizational and societal impact. To our knowledge, this is among the first studies to deploy generative AI for the integrity dimension of corporate culture, extending recent work on disclosure summarization (Kim et al., 2025) and investment score prediction (Jha et al., 2023).

Second, we contribute to the corporate culture literature by introducing a validated, context-aware measure of corporate integrity. Our four-component framework extends foundational work by Guiso et al. (2015) and Li et al. (2021) by articulating integrity as a construct with distinct, measurable facets whose joint and separate variation can be traced across firms and time. Unlike survey- or analyst-report-based measures, our approach captures integrity discourse directly from senior management in unscripted settings and yields both aggregate and component-level scores, a level of granularity that existing proxies do not provide.

Third, we contribute to the emerging literature on generative AI in management and accounting research. While generative AI has been applied to stock price forecasting (Lopez-Lira & Tang, 2023), analyst report analysis (Li et al., 2026), and detection of disclosure “bloat” (Kim et al., 2025), its use to operationalize corporate integrity is novel. By showing that a context-aware integrity signal relates to governance structure, restatement behavior, earnings management, and payout and credit decisions in ways that a widely used dictionary proxy does not, we provide evidence that LLM-based measurement of unscripted managerial discourse can reveal decision-relevant information that existing tools struggle to capture, reinforcing the strategic role of IS in transforming unstructured communication into actionable organizational knowledge.

The remainder of the paper is organized as follows. Section 2 reviews the relevant literature. Section 3 details our methodology for identifying integrity-related discourse in earnings calls. Section 4 presents the validation results, including classification performance for the Phase I integrity detection gate, the role of contextual understanding relative to dictionary-based approaches, and empirical associations with established governance and reporting outcomes. Section 5 discusses methodological, theoretical, and practical implications. Section 6 concludes.

2. Literature Review

We structure the literature review around three areas that together highlight the relevance of IS capabilities in analyzing earnings calls to construct corporate integrity scores. Section 2.1 traces the evolution of textual analysis within the management domain. Section 2.2 develops the concept of corporate integrity culture and proposes a four-component framework grounded in organizational, governance, and ethics literature. Section 2.3 focuses on earnings call Q&A and textual analysis techniques that have been used to extract cultural signals from them.

2.1 The Evolution of Textual Analysis in Management Research

Over the past two decades, textual analysis in management research has evolved in step with advances in natural language processing (NLP). Early work relied on dictionary-based methods, exemplified by Tetlock's (2007) finding that negative news tone predicts market downturns. Loughran and McDonald (2011) developed finance-specific lexicons that enabled predictions of returns, volatility, and profitability from disclosure tone (Henry & Leone, 2015). While

foundational, these approaches treated text as an unordered collection of words and inherently overlooks context, syntactic structure, and meaning.

To move beyond surface-level lexical matching, researchers adopted machine learning techniques capable of capturing broader patterns in text (Li, 2010; Patel et al., 2014). A further shift came with neural networks, beginning with word embeddings (Mikolov et al., 2013) that represented semantic relationships, recognizing, for example, that “bankruptcy” and “Chapter 11” occupy adjacent regions of a learned vector space. Li et al. (2021) applied this approach to construct culture dictionaries from earnings call corpora, while Dutta et al. (2023) used word2vec to study contextualized COVID-related disclosures in 10-K filings. These embedding-based methods captured more semantic structure than dictionaries but still operated at the word level, limiting their ability to interpret the full context in which a cultural signal appears.

Most recently, transformer-based models such as BERT (Devlin et al., 2018), FinBERT (Huang et al., 2023), and generative LLMs have transformed NLP. Trained on very large corpora, LLMs perform diverse downstream tasks with minimal supervision through few- and zero-shot learning (de Kok, 2025), and contemporary reasoning models can articulate the basis for their decisions through structured chains of reasoning (Huang & Chang, 2022; Shojaei-Mend et al., 2024; Wang et al., 2022). These capabilities have opened new avenues for financial text analysis, including document parsing (Kim et al., 2025), management speech analysis (de Kok, 2025), investment prediction (Jha et al., 2023), deception detection (Beckmann et al., 2024), and stock forecasting (Lopez-Lira & Tang, 2023). The trajectory from word counts to contextual reasoning motivates our use of a reasoning LLM to study corporate integrity, a construct whose manifestation in managerial discourse is inherently context-dependent.

2.2 Corporate Integrity Culture and Its Components

Integrity in corporate culture refers to the degree to which a firm's shared values, norms, and practices promote ethical conduct, honesty, and adherence to moral principles (Graham et al., 2022). Often distilled as "doing the right thing" (Guiso et al., 2015), corporate integrity extends well beyond mere legal compliance to encompass transparency, accountability, and, crucially, consistency between an organization's stated values and its actual behavior (Audi & Murphy, 2006; Paine, 1994). As a cultural construct, integrity is inherently multi-faceted. Audi & Murphy (2006) identify integrality, moral purpose, and moral courage as its defining philosophical attributes, while Palanski and Yammarino (2007) synthesize the organizational literature around word-action alignment, value congruence, and moral trustworthiness. Becker (1998) defines integrity as the commitment in action to a morally justified set of principles and values, encompassing independence, justice, and rationality. Solomon (1992) position integrity as a "master virtue" that integrates other business virtues (honesty, trust, fairness, and excellence) within a community context. Kaptein & Wempe (2002) present a stakeholder-based framework arguing that corporate integrity requires balancing economic performance, social responsibility, and ethical values simultaneously. Erhard et al. (2009), develop a positive model in which integrity, defined as being "whole, complete, and sound," necessarily incorporates but is irreducible to morality, ethics, and legality. These perspectives converge on a common insight: integrity is not a single trait but rather a constellation of reinforcing properties that, together, shape how organizations relate to their stakeholders.

Building on the above discussions, we propose a four-component framework (as shown in Figure 1) that conceptualizes corporate integrity as manifesting through four distinct but mutually reinforcing dimensions: (1) Accountability, Governance, and Ownership; (2)

Regulatory Compliance and Adherence; (3) Moral Integrity and Fairness; and (4) Transparency, Trust, and Honesty. For brevity, we refer to these components in the remainder of the paper as Accountability, Compliance, Moral Integrity, and Honesty, respectively. Each component represents a distinct facet of corporate integrity culture.

[Insert Figure 1]

The framework's organizing logic derives from crossing two well-established theoretical distinctions. The primary axis draws on Tenbrunsel et al.'s (2003) model of "ethical infrastructure" which distinguishes between the formal systems that enforce integrity (rules, surveillance, sanctioning mechanisms) and the informal systems that cultivate it (ethical climate, justice climate, communication norms). The first two components (Accountability and Compliance) operate primarily through formal institutional structures, while the latter two (Moral Integrity and Honesty) operate primarily through informal cultural processes. This formal-informal distinction is reinforced by Weaver and Treviño's (1999) empirical finding that formal compliance and informal values orientations are complementary rather than substitutable, and provides the primary organizing axis for the framework.

A second axis, grounded in the distinction between internal and external governance mechanisms (Aguilera et al., 2008), further differentiates the components within each formal-informal category. Aguilera et al. (2008) classify governance practices according to whether they are internally directed (board oversight, ownership structures, managerial incentives) or externally directed (regulatory frameworks, market disciplines, stakeholder activism), and we extend this internal-external distinction to the ethical domain through Victor and Cullen's (1988) ethical climate theory, which locates shared moral norms at the local (organizational) level while recognizing that organizations also operate within cosmopolitan (extra-organizational) normative

environments. Accountability structures are internally directed, designed by organizational leadership to govern the firm's own decision-making processes (Davis et al., 1997; Jensen & Meckling, 1976). Compliance is externally directed, imposed by regulatory authorities, industry bodies, and contractual counterparties (DiMaggio & Powell, 1983; Suchman, 1995). Moral integrity constitutes the organization's internal normative environment, the shared ethical climate and justice norms that shape how members reason about moral questions (Trevino, 1986; Victor & Cullen, 1988). Honesty constitutes the organization's external communicative interface, the mechanisms through which integrity is made visible and credible to outside stakeholders (Healy & Palepu, 2001; Schnackenberg & Tomlinson, 2014). When these two axes are crossed, they produce a 2×2 structure in which each component occupies a unique cell: formal-internal (Accountability), formal-external (Compliance), informal-internal (Moral Integrity), and informal-external (Honesty). Following sections discuss each component in detail, grounding their contribution to corporate integrity in the relevant literature.

2.2.1 Accountability, Governance, and Ownership (Accountability)

The first component captures the foundational requirement that an integrity-oriented firm accept ownership of its outcomes and maintain the structural systems that render it answerable to stakeholders. Accountability, governance, and ownership are consolidated because they operate as a tightly linked causal system: governance structures exist to solve the accountability problems created by the separation of ownership and control (Fama & Jensen, 1983; Jensen & Meckling, 1976), and integrity emerges when managers internalize psychological ownership over organizational outcomes (Pierce et al., 2001). Within Tenbrunsel et al.'s (2003) ethical infrastructure, this component operates through formal systems, including sanctioning mechanisms, monitoring structures, and oversight processes. On the framework's second axis,

this component is internally directed: governance and accountability structures are designed by organizational leadership and oriented inward toward the firm's own decision-making processes (Aguilera et al., 2008), distinguishing them from the externally imposed requirements of regulatory compliance.

Accountability enhances integrity because the anticipation of justifying decisions to legitimate audiences improves their quality and ethical defensibility (Lerner & Tetlock, 1999). Hall et al. (2017), synthesizing over fifty empirical studies, document that felt accountability promotes ethical behavior, reduces moral disengagement, and fosters organizational learning. Governance provides the institutional architecture that makes accountability enforceable at scale: Shleifer and Vishny (1997) define governance as the mechanisms through which capital providers ensure returns, framing it explicitly as an internal accountability architecture, while Arjoon (2005) argues that governance achieves its integrity-enhancing potential only when it embeds moral purpose within institutional design. Furthermore, Pierce et al. (2001, 2003) show that ownership generates felt responsibility and stewardship behavior that extend beyond what formal monitoring alone can produce, and Hernandez (2012) integrates accountability, relational cognition, and long-term perspective as interrelated processes that jointly drive pro-organizational behavior. Together, these constructs represent the formal, internally directed stewardship architecture of corporate integrity.

2.2.2 Regulatory Compliance and Adherence (Compliance)

If Accountability represents the internal stewardship architecture of integrity, Compliance represents its external institutional foundation: the non-negotiable baseline of rule-following without which claims of integrity lack credibility. An organization that espouses ethical values but fails to meet its legal obligations undermines the entire integrity construct; one that achieves

perfect compliance but nothing more has satisfied only the minimum condition (Fuerst et al., 2023; Paine, 1994). Like Accountability, Compliance operates through Tenbrunsel et al.'s (2003) formal ethical infrastructure, but its distinguishing feature is its external direction: compliance requirements are imposed by regulatory authorities, industry bodies, and contractual counterparties rather than generated internally (DiMaggio & Powell, 1983; Suchman, 1995). Paine's (1994) distinction between compliance-oriented and integrity-oriented ethics management positions compliance as necessary but insufficient. Weaver and Treviño (1999) validate this empirically: the interaction of compliance and values orientations uniquely predicts employees' willingness to report misconduct, an outcome neither orientation produces alone. DiMaggio & Powell (1983) explain how coercive institutional pressures compel organizations to adopt compliance structures, and Suchman (1995) argues that regulatory compliance maintains institutional legitimacy. In the financial reporting domain, Ashbaugh-Skaife et al. (2008) demonstrate that remediation of control deficiencies improves accrual quality. Yet compliance becomes counterproductive when isolated from the organization's broader ethical system: Stansbury and Barry (2007) show that surveillance can displace intrinsic moral motivation, and Kaptein (2009) demonstrates that formal programs are insufficient without a supportive ethical culture. Compliance thus establishes the institutional floor of integrity, but its effects depend on being embedded within a broader system of accountability, moral commitment, and transparent communication.

2.2.3 Moral Integrity and Fairness (Moral Integrity)

Where the first two components establish what the firm must do through formal institutional structures, this third component captures what the firm chooses to do: its voluntary commitment to ethical conduct and equitable treatment of stakeholders. Within Tenbrunsel et al.'s (2003)

ethical infrastructure, this component operates through informal systems, including ethical climate, justice climate, and shared normative expectations. Its internal orientation is well established in the literature. Victor & Cullen (1988) demonstrate that ethical climates are properties of the organization's internal environment, emerging from local norms, socialization processes, and leadership behavior rather than external mandates. Trevino (1986) reinforces this positioning by showing that individual moral development interacts with the organization's internal culture to determine ethical conduct, and Arnaud and Schminke (2012) confirm that ethical climate operates as an organization-level phenomenon shaping behavior from within.

Ethical decision-making contributes to integrity by embedding moral reasoning into the organization's cognitive architecture. Rest's (1986) four component model establishes that ethical behavior unfolds through moral awareness, judgment, motivation, and character, with each stage shaped by the internal organizational context. At the leadership level, Brown et al. (2005) show that ethical leaders shape behavior through social learning, reducing deviance and increasing organizational citizenship (Brown & Treviño, 2006). Treviño et al. (2000) elaborate that leaders must both be ethical personally and actively manage the firm's internal ethical environment. Fairness is constitutive of this internal moral architecture: Greenberg's (1987) foundational taxonomy established organizational justice as a multi-dimensional construct rooted in internal organizational processes, and Colquitt (2001) demonstrates that perceived procedural justice strongly predicts rule compliance and organizational commitment. Together, ethical reasoning and fairness constitute the informal, internally directed moral core of integrity, generating normative commitments that differentiate genuine integrity from compliance baseline.

2.2.4 Transparency, Trust, and Honesty (Honesty)

The preceding components establish what an integrity-oriented firm does: it accepts responsibility, meets its obligations, and acts on moral principle. This fourth component addresses how the firm communicates those commitments to its stakeholders and whether they have reason to believe them. Transparency, honesty, and trust form a single causal chain: transparency and honesty produce stakeholder trust (Schnackenberg & Tomlinson, 2014), and trust validates the organization's integrity in the eyes of those it serves. Like Moral Integrity, this component operates through Tenbrunsel et al.'s (2003) informal ethical infrastructure, but its defining characteristic is its external orientation: it governs the interface between the organization and outside constituencies. Healy and Palepu (2001) frame corporate disclosure as inherently outward-facing, directed at reducing information asymmetry between corporate insiders and external stakeholders, and Bushman et al. (2004) empirically distinguish financial transparency from governance transparency, both oriented toward external audiences. Without this externally directed component, the other three dimensions would remain invisible and non-credible to those outside the firm.

Schnackenberg and Tomlinson (2014) conceptualize transparency along three dimensions: disclosure (availability), clarity (comprehensibility), and accuracy (truthfulness), establishing that transparency requires information quality and not merely volume. From a signaling perspective, voluntary disclosure of unfavorable information serves as a costly signal of integrity directed at external stakeholders (Connelly et al., 2010; Spence, 1973). Simons (2002) defines behavioral integrity as perceived word-deed alignment, and Simons et al.'s (2015) meta-analysis shows that it exerts strong effects on trust and performance. Mayer et al. (1995) identify integrity as a core antecedent of trustworthiness, and Dirks and Ferrin (2002) establish that trust in leadership predicts satisfaction, commitment, and performance. Together,

transparency, honesty, and trust constitute the informal, externally directed communicative fabric of corporate integrity.

The four components described above are four different facets of a single underlying construct: corporate integrity. Having established the conceptual foundations of corporate integrity, the following section discusses earnings calls as a textual data source and the textual analysis techniques that have been used to extract integrity-related signals from corporate discourse.

2.3 Earnings Calls and the Measurement of Corporate Integrity

2.3.1 From Manual to AI-Driven Earnings Call Analysis

Earnings calls are key information events for public firms and have attracted extensive research attention (Bochkay et al., 2020; Bochkay et al., 2023; Loughran & McDonald, 2016). A typical call consists of a management presentation followed by a Q&A session, with the latter widely regarded as more revealing of managerial attitudes, tone at the top, and underlying firm culture (Guiso et al., 2015; Li et al., 2021). This section traces how the analysis of earnings calls has evolved from manual coding and simple text metrics toward advanced algorithms and LLMs.

Early research established that Q&A tone carries market-relevant information. Frankel et al. (1999) showed that earnings calls affect stock prices, and Price et al. (2012) found that spontaneous Q&A sentiment predicts short-term returns more reliably than prepared remarks. Larcker and Zakolyukina (2012) linked Q&A linguistic cues to subsequent financial restatements, underscoring the diagnostic value of the Q&A portion for assessing managerial integrity. Li et al. (2021) used dictionary-based method to measure corporate culture, arguing that Q&A discourse best reveals values of the organization.

Textual analysis of earnings calls has since advanced rapidly through AI-driven techniques. Bai et al. (2023) compared management Q&A responses to AI-generated answers to

detect evasiveness, which in turn predicted weaker future performance. Related innovations include vocal stress analysis (Mayew & Venkatachalam, 2012), political risk metrics derived from call transcripts (Hassan et al., 2019), and broader applications of AI in corporate communications (Cao et al., 2023). This progression (from hand-coding to dictionaries to LLMs), combined with prior evidence on the informativeness of Q&A textual analysis, motivates our focus on the Q&A portion of earnings calls as the setting in which corporate integrity can be detected, and sets up the methodological gap that our study addresses.

2.3.2 Measurement of Corporate Integrity Culture Using Textual Analysis

Corporate integrity culture has been studied through surveys, employee reviews, and textual analysis, each with strengths and limitations that motivate our methodological contribution. Surveys provide valuable insider perspectives (Graham et al., 2022) but are limited by impression management and reproducibility concerns (Keiser & Payne, 2019). Employee reviews offer large-scale data (Ji et al., 2017) but suffer from selection bias and inconsistent coverage across firms (Li et al., 2021). Textual analysis of corporate disclosures is more scalable and draws on the linguistic cues that reflect underlying cultural values (Guiso et al., 2015; Li et al., 2021). Early dictionary-based approaches linked ethics language to corporate social responsibility outcomes (Webley & Werner, 2008) and created integrity proxies from disclosure text (Audi et al., 2016), but these measures remain static and context-insensitive (Hoitash & Hoitash, 2022). More advanced word2vec-based methods infer culture from context yet frequently conflate contextually distinct uses of the same term, as we document in Section 4.2. Recent work has begun to leverage LLMs, for example by extracting cultural insights from analyst reports (Li et al., 2026), but has not produced a contextual, firm-quarter integrity score derived from earnings call discourse.

Our study fills this gap by combining the strengths of prior approaches while mitigating their shortcomings. We analyze spontaneous, unscripted earnings call Q&A, which is widely regarded as candid and culturally revealing (Li et al., 2021), using a reasoning LLM embedded in a multi-agent architecture rather than static dictionaries or word-level embedding models. This design yields two key advantages. First, the LLM interprets meaning across sentences, capturing subtle integrity cues that word-level methods miss (de Kok, 2025). Second, analyzing extended segments rather than isolated words enables a dynamic, holistic integrity measure that reflects both content and discursive context. Building on this foundation, the next section details how we operationalize this approach and structure the analysis of Q&A transcripts.

3. Methodology

This section describes the systematic approach used to collect, process, and analyze textual data extracted from corporate earnings call transcripts to construct Contextual Corporate Integrity Scores. Figure 2 presents a process model of the entire pipeline, from data collection to the construction of corporate integrity measures. Our methodology moves beyond traditional keyword-based analyses by leveraging a multi-agent large language model (LLM) architecture to capture the context-dependent nature of integrity culture discourse.

[Insert Figure 2]

3.1 Data Collection and Sample Definition

Following established practices in the literature (Allee & Deangelis, 2015; Bochkay et al., 2020), we sourced full transcripts of quarterly earnings conference calls spanning the period from 2008 to 2024 (fiscal year). Each transcript file contains the complete body of the earnings call

conversation along with metadata that enables matching with the COMPUSTAT database and facilitate systematic data organization for subsequent steps. These metadata elements include the company ticker symbol, company name, transcript title, date, and transcript identifier. Following data matching with COMPUSTAT, our earnings call dataset consists of 183,043 earnings call transcripts. We then extracted the Q&A sections from these transcripts, as this section contains particularly valuable information on organizational culture (Li et al., 2021). To ensure our analysis focused specifically on corporate leadership responses, we filtered the Q&A content using regex-based parsing in Python. This approach was feasible because the transcripts follow a relatively consistent structural convention: each speaker turn is preceded by a standardized speaker identifier and role designation (e.g., participant name, title, and affiliation), while the Q&A portion is separated from the presentation portion by identifiable section headers. We exploited these recurring patterns to construct regular expressions that (i) identify the Q&A section boundary, (ii) segment the dialogue into individual speaker turns, and (iii) classify each turn as belonging to a company executive or an external analyst based on the associated role and affiliation. We then excluded analyst questions and retained responses from company participants, including CEOs, CFOs, and other senior executives. We validated the extraction on a random subset of transcripts through manual inspection and refined the parsing rules to accommodate formatting variations across transcript vendors and years.

This filtering approach (i.e., keeping only company executive discourse) serves multiple purposes. First, it ensures that our analysis focuses exclusively on corporate leadership communication, which most directly reflects organizational values and cultural orientations (Guiso et al., 2015; Li et al., 2021). Second, it eliminates potential noise from analyst questions that might contain integrity-related terminology without reflecting the company's actual cultural

stance. Third, it creates a more homogeneous corpus of executive communication. This filtered corpus constitutes our fiscal-quarter matched sample of public firms for the study period and lays the foundation for subsequent steps.

3.2 Text Preprocessing and Data Preparation

Our data preparation follows a segmentation approach to identify coherent text segments in earnings call transcripts. The transcript content used in this process is the same as described in Section 3.1 and will be referred to as earnings calls throughout the remainder of the manuscript.

We segment Q&A sections into coherent chunks through a three-step process. In the first step, we decompose the relevant content of each Q&A section into individual sentences using the Python library spaCy (Dutta et al., 2023). In the second step, we convert each sentence into high-dimensional vector representations using the BGE-M3 embedding model, recognized for producing semantically meaningful sentence representations suitable for similarity analysis (Chen et al., 2024). Sentence-embedding approaches can be less effective when inputs are long, syntactically complex, or context-dependent, as is often the case with executive discourse. We therefore use BGE-M3, which supports inputs of up to 8,192 tokens and is trained to capture semantic relationships across different levels of granularity (Chen et al., 2024). These capabilities make BGE-M3 well suited for representing the longer and more information-dense sentences in our corpus and provide an appropriate basis for the similarity-based segmentation procedure that follows. The third step implements a semantically aware adaptation of the C99 algorithm (Choi, 2000) for text segmentation. Rather than relying on traditional bag-of-words representations (TF-IDF) employed in the original C99 implementation, our approach leverages the dense sentence embeddings from the BGE-M3 model to compute cosine similarity between sentences. The C99 algorithm identifies coherent text segments by calculating similarity scores

between sentence pairs and using these scores to determine optimal segment boundary placement (Choi, 2000; Li et al., 2026). Intuitively, this approach groups adjacent sentences that discuss similar concepts, effectively reconstructing the document’s logical, topically cohesive, and coherent paragraph-like structure (Li et al., 2026). However, replacing the TF-IDF sentence representations in C99 with BGE-M3 dense embeddings does not alter transcript-level integrity measurements relative to the established TF-IDF-based C99 baseline. This conclusion is supported by paired-sample t-tests conducted across three independent samples of 1,000 earnings calls each; the full discussion is presented in Appendix A Supplementary Discussion SD1. The segmentation process yields a corpus comprising 10,277,650 text segments distributed across 183,043 earnings call transcripts, which provides the foundation for our LLM-based classification pipeline.

3.3 LLM-Based Classification

3.3.1 Model Selection

Following text segmentation, we organized each segment into a structured JSON format, including the text and metadata (e.g., transcript ID, year, company name). Selecting an appropriate LLM for classification involved balancing performance, consistency, accessibility, and cost. Recent studies have raised concerns about using LLMs for academic research: as noted by Chen et al. (2023) and de Kok (2025), LLMs can produce inconsistent outputs across models, versions, and even repeated runs with the same prompt, posing challenges for reproducibility. Chen et al. (2023) documented major output differences between GPT-4’s March and June 2023 versions, highlighting temporal instability, and de Kok (2025) cautions against relying on proprietary model APIs that can be modified or discontinued without notice. OpenAI’s GPT-4o update in early 2025, for example, significantly altered model behavior, including increased

sycophancy (OpenAI, 2025). To address reproducibility concerns, we followed de Kok’s (2025) recommendation to avoid third-party APIs by deploying open-source models on local or cloud infrastructure (e.g., RunPod, VastAI, AWS). Open-source models provide full control over inference, allow rapid iteration, avoid API limits, and prevent disruptions from unexpected updates. They are also more cost-efficient without major quality trade-offs (Irugalbandara et al., 2023).

Building on these advantages, our next step was to identify an open-source LLM capable of reliably classifying corporate integrity discourse. We adopted a benchmark-driven model selection strategy, evaluating candidate models on two benchmarks relevant to our task: the Multistep Soft Reasoning (MuSR) benchmark (Sprague et al., 2023) and the BigBench Hard (BBH) benchmark (Suzgun et al., 2022). MuSR evaluates a model’s language understanding and multistep reasoning capabilities in complex textual settings, while BBH assesses common-sense reasoning and general world knowledge. Given that our classification task requires contextual interpretation and multi-step reasoning over extended discourse segments drawn from earnings call transcripts, MuSR is particularly relevant as a selection criterion. Benchmark performance for all candidate models is reported in Appendix B Table B2. Among the evaluated models, DeepSeek-R1-Distill-Qwen-14B (hereafter DeepSeek-R1) demonstrated the strongest overall performance and, notably, the highest MuSR score, making it the most suitable model for our classification task. Several other candidate models appeared to perform competitively when evaluated on the full imbalanced corpus, but their performance dropped substantially on a balanced validation set, suggesting that their apparent strength reflected exploitation of the severe class imbalance (integrity-positive segments account for only approximately 4.5% of the corpus) rather than genuine discriminative capability. Beyond benchmark performance, our

selection was further motivated by DeepSeek-R1’s reasoning capabilities: it belongs to a class of large language models explicitly designed to engage in structured, multistep reasoning before generating responses (Guo et al., 2025). This reasoning ability adds an interpretability layer to text classification (Huang & Chang, 2022; Wang et al., 2022), as the model can articulate the basis for its decisions. The model’s suitability is further validated through classification performance on a human-annotated validation set, as detailed in Section 4.1.

3.3.2 Multi-Agent Classification Architecture

A key design challenge in LLM-based classification of complex constructs such as corporate integrity is the tension between specialization and coherence. Multi-agent architectures can assign domain-specific experts to individual facets of a construct but risk fragmenting the broader context when agents operate in isolation. Single-agent monolithic designs preserve global coherence but sacrifice the granularity required to reason over distinct conceptual dimensions. Wang et al. (2024) characterize this tension through the concept of cognitive synergy, whereby the collective reasoning capacity of collaborating agents can exceed the sum of their individual contributions, provided appropriate coordination mechanisms are in place. Achieving this synergy, however, demands an architecture that enables agents to share information without sacrificing their domain-level specialization. To address this trade-off, we draw on the blackboard architecture (Salemi et al., 2025), a well-established framework for cooperative problem-solving in which independent knowledge sources contribute to a shared information space under the supervision of a central controller. We adapt its core principle of indirect coordination through shared memory to implement a multi-agent classification system in which specialized agents collaborate through a shared memory structure (the Blackboard) that evolves across four sequential processing phases. The full architecture is illustrated in Figure 3.

[Insert Figure 3]

The classification of each text segment proceeds as follows. In Phase I (Integrity Detection), a dedicated Corporate Integrity Agent evaluates whether the segment contains any signal of integrity-related corporate culture, performing a binary classification (Yes/No). This agent operates as an execution gate: segments classified as “No” receive their final label and processing terminates, while segments classified as “Yes” are forwarded to the next reasoning stages. For segments that pass the gate, the system initializes a shared memory space (the Blackboard) and proceeds to Phase II. A more detailed discussion of the rationale underlying the Phase I design, along with its associated trade-offs, is provided in Appendix A Supplementary Discussion SD3.

In Phase II (Observation Pass), four domain-specific agents analyze the segment in parallel, each aligned with one dimension of the corporate integrity framework: Accountability, Compliance, Moral Integrity, and Honesty. Drawing from the construct conceptualizations formalized in the theoretical framework (Section 2.2), each agent’s prompt embeds the relevant definitions along with chain-of-thought reasoning instructions. In this phase, agents operate under strict constraint: they are prohibited from issuing judgments and must produce only objective, factual observations about the segment as it pertains to their respective dimension. These observations are written to the shared memory space, resulting in an updated state we refer to as Blackboard v_1 , forming a shared fact base. By enforcing a clean separation between evidence gathering and interpretation, this phase reduces the risk of premature or biased judgments and anchors all subsequent reasoning in explicitly extracted evidence.

In Phase III (Verdict Pass), the same four agents are reactivated with read access to the complete Blackboard v_1 , enabling cross-agent awareness. Each agent now produces

contextualized verdicts and reasoning, informed not only by its own prior observations but also by the factual evidence contributed by peer agents. This shared-state design realizes the cognitive synergy described by Wang et al. (2024): each agent retains its domain expertise while its reasoning is enriched by peer observations, achieving collective awareness without sacrificing modularity. The outputs from this phase are written back to the shared memory space, forming Blackboard v_2 , which layers these reasoned verdicts atop the Blackboard v_1 fact base, producing a fully contextualized system state. The indirect coordination enabled by the blackboard aligns with the asynchronous collaboration paradigm described by Salemi et al. (2025).

In Phase IV (Synthesis), a Synthesizer Agent consolidates all outputs from Blackboard v_1 into a final integrated classification. Where clear integrity categories are detected, it produces the final output directly. In cases where ambiguity persists across agent verdicts, a Closest-Fit Resolution Agent is invoked to assign the single most appropriate category based on the available evidence. Illustrative examples of segments assigned to each of the four components are reported in Appendix B Table B7.

In total, our classification pipeline engages eleven agent roles across the four phases: one Corporate Integrity Agent (Phase I), four domain agents in observation mode (Phase II), four agents in reasoning and verdict mode (Phase III), and two adjudication agents (Phase IV). The architecture offers several properties that strengthen classification quality. The strict separation of observation (Phase II) from judgment (Phase III) reduces hallucination risk by grounding verdicts in a shared fact base rather than unconstrained generation. The multi-phase, shared-memory design enables more calibrated classifications that resemble committee-based deliberation rather than single-pass binary decisions. Finally, the blackboard provides a persistent

reasoning trace across all phases, constituting a verifiable audit trail that supports the interpretability and credibility of each classification decision.

3.3.3 Addressing LLM Concerns

Building on the challenges identified by Li et al. (2026), we address two critical concerns associated with the use of generative LLMs: look-ahead bias and hallucination.

Look-ahead bias arises when a model’s output is influenced by information that would not have been available at the time of analysis. This is particularly relevant for LLMs trained on large-scale corpora that may include retrospective or forward-looking knowledge, potentially compromising temporal integrity. In our application, however, the model is not employed for predictive forecasting. Instead, it evaluates whether a given text segment reflects corporate integrity culture based solely on the information contained within that segment. This backward-looking, text-contained design inherently reduces the risk of look-ahead bias. To further safeguard against unintended information leakage or overfitting to specific entities, we follow the method used by Jha et al. (2023) and Li et al. (2026) and conduct an entity-masking test. Using the spaCy NLP library, we anonymized entities in the text: company names were replaced with “[ORGANIZATION]”, individual names with “[PERSON]”, and product names with “[PRODUCT]”. On a randomly selected subset of 2,000 text segments, we compared the model’s classification decisions on unmasked versus masked texts. The resulting Cohen’s kappa coefficient was approximately 0.87, indicating a high level of agreement. This suggests that our results are not primarily driven by entity-specific cues or associations the model may have learned across time, thereby reinforcing the robustness of our classification approach.

Hallucination occurs when models produce content not grounded in the input text (Maynez et al., 2020). Our architecture addresses this concern through multiple safeguards. First,

the multi-phase protocol enforces a strict separation between observation and reasoning, ensuring that agent verdicts are anchored in explicitly extracted evidence rather than unconstrained generation. Second, the blackboard structure provides an auditable chain of reasoning, making it possible to trace any classification back to the factual observations that support it. Third, structured chain-of-thought prompting guides logical reasoning and minimizes unsubstantiated outputs (Li et al., 2026; Wei et al., 2022). The prompts direct agents to analyze the full context of each segment, identify relevant details, and explain their reasoning with explicit reference to the source text. Together, these mechanisms reduce the risk of hallucinated or unfounded classifications.

3.4 Contextual Corporate Integrity Score Construction

After classifying each text segment through the multi-agent pipeline, we construct Contextual Corporate Integrity Scores for each company-quarter. For each earnings call, we compute the following indicators. (i) The Integrity-Related Segment Count represents the total number of segments within the transcript that the pipeline classified as relevant to corporate integrity culture, serving as a direct measure of the frequency of integrity-related discourse during the call. (ii) Component-Specific Segment Counts, based on the final category assignments produced by the Phase IV synthesis stage, capture the number of segments classified under each of the four integrity dimensions: Accountability, Compliance, Moral Integrity, and Honesty. Each count reflects how often a specific integrity dimension is identified in the classified segments, enabling granular analysis of cultural emphasis. The distribution of these components across integrity-related segments is reported in Appendix B Table B6, with accompanying illustrative examples provided in Appendix B Table B7.

Using these segment-level outputs, we derive both an overall integrity score and its component-specific scores at the firm-quarter level. Relevant variables are described in Table A1 provided in Appendix. These measures serve as inputs to the empirical validation analyses reported in Section 4.3, including associations with governance characteristics, restatement filings, earnings management, dividend policy, and trade credit.

4. Results

To establish the reliability and effectiveness of our LLM-based classification pipeline, we adopt a two-pronged validation strategy that combines technical performance assessment with empirical validation. First, we evaluate the classification performance of the Phase I integrity detection gate, the component that determines which segments enter the downstream reasoning stages. We focus validation effort on this gate because it sits at a critical junction where any misclassification would propagate through all subsequent stages of the pipeline. Second, we validate our contextual integrity measure by examining its empirical associations with well-established indicators of governance quality, reporting integrity, and financial policy.

4.1 Overall Integrity Detection Performance and Validation

To establish the reliability of the Phase I integrity detection gate in identifying overall corporate integrity culture in earnings call discourse, we conducted validation tests across multiple dimensions. Our validation strategy focused on three areas: (i) standard classification metrics benchmarked against manually labeled data, (ii) prediction stability across multiple runs to ensure robustness, and (iii) comparative performance relative to dictionary-based approaches.

4.1.1 Validation Design for the Phase I Integrity Detection Gate

Following the evaluation framework outlined by Demers et al. (2024) and Li et al. (2026), we constructed a manually annotated validation set to assess the Phase I integrity detection gate. A total of 300 text segments were drawn and subsequently hand-coded to serve as ground-truth labels. The sample was generated using a stratified sampling design based on the Phase I predicted labels, comprising 150 segments classified as “Yes” (integrity-related), and 150 segments classified as “No” (not integrity-related) by the gate. The rationale for this stratified design is twofold. First, the raw corpus exhibits severe class imbalance, with integrity-positive segments constituting approximately 4.5% of the total. A purely random sample of moderate size would yield very few positive instances, thereby limiting the ability to rigorously evaluate performance on the minority class. Second, sampling 150 segments from each category increases the likelihood that both classes are adequately represented in the manually coded validation set relative to a purely random sample. This enables a meaningful assessment of precision, recall, F1 score, Matthews correlation coefficient (MCC), and overall classification quality.

4.1.2 Classification Performance of the Phase I Integrity Detection Gate

The validation results indicate consistently strong performance across all evaluation metrics. Table 1 presents the performance of the Phase I gate for the main classification run alongside five independent stability runs, all evaluated against the human-annotated validation set of 300 samples. The main classification run refers to the execution that labels all segments in the full corpus, from which the firm-quarter integrity measures used in the empirical analyses are ultimately constructed; the five stability runs, by contrast, were conducted exclusively on the validation set in order to assess run-to-run reproducibility.

[Insert Table 1]

The main run achieved an accuracy of 95.67%, with macro-averaged precision and recall both at 95.67% and 95.68% respectively. The Cohen’s kappa coefficient of 0.913 indicates near-perfect agreement between Phase I predictions and human annotations. According to established benchmarks, kappa values above 0.80 are considered to reflect strong agreement, while values above 0.90 approach near-perfect concordance (Cohen, 1960; Landis & Koch, 1977). The Matthews correlation coefficient of 0.914 corroborates this assessment, providing a balanced measure that accounts for all four cells of the confusion matrix.

[Insert Table 2a and 2b]

Per-class metrics (Table 2a) reveal balanced performance across both categories. In the main run, the gate achieves high recall for integrity-positive segments (96.60%) and high precision for integrity-negative segments (96.67%), indicating that it is both sensitive to genuine integrity discourse and disciplined in avoiding false positives. The confusion matrix for the main run (Table 2b) shows only 13 misclassifications out of 300 segments, confirming the practical reliability of the gate. Detailed confusion matrices for all runs are provided in Appendix B Table B3.

4.1.3 Prediction Stability of the Phase I Integrity Detection Gate

To ensure the robustness and reproducibility of the Phase I gate, we conducted five independent runs of DeepSeek-R1 on the validation set using the same parameters and prompts. This stability analysis is important for establishing confidence in the reproducibility of results, particularly given the stochastic nature of LLM inference.

As shown in Table 1, the five stability runs yield accuracy scores ranging from 92.67% to 94.33%, with macro-averaged F1 scores clustered tightly between 92.65% and 94.33% (SD = 0.6%). Cohen’s kappa values range from 0.853 to 0.887 (mean = 0.872, SD = 0.012), all falling

within the “almost perfect” agreement range. These results confirm consistent behavior of the Phase I gate across independent replications, directly addressing the concerns raised by de Kok (2025) regarding the tendency of LLMs to produce varying outputs from identical inputs.

The inter-run agreement analysis, summarized in Table 3, further validates the gate’s reproducibility. All pairwise Cohen’s kappa scores among the five runs exceed 0.91, with a mean of 0.945 and a standard deviation of 0.013. As subset size increases, the mean kappa remains stable at 0.945 while the standard deviation decreases (from 0.013 at $k=2$ to 0.005 at $k=4$), reflecting convergent behavior. This level of self-consistency provides strong empirical support for the reliability of the Phase I gate under repeated application.

[Insert Table 3]

Taken together, the classification performance and stability results affirm that the Phase I integrity detection gate delivers consistent, high-quality performance across multiple executions. Because all downstream component classifications are conditioned on a segment passing this gate, these findings offer strong empirical support for the reliability of the pipeline as a whole in integrity-related textual analysis. A companion table to Table 3, reporting detailed pairwise kappa values for all run combinations, is provided in Appendix B Table B4.

4.2 Contextual Understanding versus Dictionary-Based Approaches

A central methodological contribution of this study lies in demonstrating the value of contextual understanding over traditional dictionary-based approaches for measuring corporate culture constructs. While prior studies using dictionary methods have provided valuable insights into corporate discourse patterns, our analysis reveals significant limitations in keyword-based approaches that modern LLM capabilities can address.

4.2.1 Beyond Keyword Matching

Our analysis confirms that the mere presence of integrity-related keywords is insufficient for identifying meaningful integrity discourse. Table 4 presents illustrative examples in which integrity-related terms from the Li et al. (2021) dictionary appear in contexts that are not meaningfully related to integrity culture, yet would be counted as integrity signals under a dictionary-based approach. The examples illustrate several distinct patterns of contextual misuse that recur throughout the corpus. In Example 1, the term “Accountable” appears as part of the proper noun “Accountable Care Organization” in a discussion of U.S. healthcare policy priorities, a technical reference to a specific health-delivery arrangement that carries no implication about the speaker’s firm’s integrity orientation. In Example 2, the term “trust” refers to a category of financial institution (“trust banks”) in a discussion of leverage-ratio adjustments, again bearing no cultural meaning. Example 3 features the word “honestly” used as a routine discourse marker punctuating an upbeat commentary on new market participants, rather than as a signal of honesty-oriented culture. In Example 4, “integrity” appears as a technical descriptor in “signal integrity,” an engineering concept in the electronics domain, and in Example 5 the same term denotes “inspection and integrity services,” the name of a service line offered to customers. In each case, the LLM’s ability to interpret semantic context allows it to distinguish incidental keyword usage from a meaningful integrity-related discourse, a distinction that dictionary-based approaches, by construction, cannot make.

[Insert Table 4]

4.2.2 Dictionary Word Distribution Analysis

To further demonstrate the limitations of dictionary-based approaches, we examined how frequently words from the integrity wordlist of Li et al. (2021) appeared across the segments

classified by our pipeline. Specifically, we compared their distribution across segments classified as integrity-related (“Yes”) versus non-integrity-related (“No”). Table 5 presents the top 40 words (or n-grams) sorted by the total number of segments in which they appear, while a full comparison of all dictionary terms is provided in Appendix B Table B5.

[Insert Table 5]

The results reveal a pattern: many dictionary words appear more frequently in segments that our pipeline classified as non-integrity related. The Yes/Total ratios in Table 5 show that dictionary words often appear in contexts disconnected from genuine discussions of integrity culture. For example, “management” appears in integrity-related segments only 15.38% of the time, and “control” only 13.54%. This pattern extends to the full dictionary (Appendix B Table B5), where a large share of terms appear in less than 50% percent of integrity-related segments, and overall, integrity keywords occur in integrity-relevant contexts only about 30% of the time. Terms such as “transparency” (approximately 90%) and “compliance” (approximately 80%) do exhibit high percentages, but such words are relatively few in number.

This finding underscores a fundamental limitation of dictionary-based approaches: they capture keyword frequency rather than meaningful discourse about integrity values and practices. Importantly, the low Yes/Total ratios observed for most dictionary terms should not be read in isolation as evidence in favor of our approach. Rather, they should be interpreted alongside two complementary pieces of evidence already established in the manuscript. First, the qualitative examples in Table 4 demonstrate that integrity-related keywords frequently appear in contexts that are demonstrably unrelated to integrity culture, such as references to corporate entities, technical product features, or routine business operations. Second, the validation results in Section 4.1 confirm that the Phase I gate’s classifications align closely with human judgment,

achieving near-perfect agreement on a rigorously annotated ground-truth set. Taken together, these three lines of evidence (the contextual misuse of dictionary terms documented in Table 4, the low keyword-to-classification correspondence reported here, and the high human-model agreement established through validation) indicate that the divergence between keyword presence and our pipeline’s classifications reflects genuine contextual discrimination rather than arbitrary disagreement. Our approach therefore captures a more meaningful measure of corporate integrity culture, one that distinguishes substantive cultural discourse from keyword occurrence.

4.3 Contextual Integrity Measure Validation: Associations with Recognized Corporate Integrity Indicators

Given that our approach to measuring corporate integrity is novel, it is important to assess the validity of the measure by examining its association with widely recognized indicators of best practices in the corporate sector. Accordingly, we use a broad set of indicators to evaluate both the overall corporate integrity measure and its individual subcomponents.

4.3.1 Corporate Governance and Corporate Integrity Culture

Board independence and corporate integrity: Prior literature suggests that stronger corporate governance, particularly greater board independence, is associated with a stronger corporate integrity culture within firms. From an agency perspective, independent directors enhance monitoring of managerial decisions and strengthen accountability because they have fewer ties to management and face reputational incentives to promote ethical conduct and transparency (Fama, 1980; Fama & Jensen, 1983). Consistent with this view, empirical evidence shows that firms with more independent boards exhibit higher financial reporting quality and lower incidences of financial misreporting and fraud (Beasley, 1996; Klein, 2002). Independent directors also help shape governance structures, ethical policies, and compliance systems that

reinforce organizational norms emphasizing integrity and responsible leadership (Kaptein, 2011; Schwartz et al., 2005). More broadly, governance research finds that firms with more independent boards experience fewer instances of corporate misconduct and regulatory violations, suggesting that board independence strengthens ethical oversight and fosters a corporate culture characterized by greater integrity and accountability (Agrawal & Chadha, 2005; Beasley et al., 2000).

Female board members and corporate integrity: Prior research suggests that greater female representation on corporate boards is associated with a stronger corporate integrity culture. Female directors are often linked to stronger ethical orientations, greater transparency, and more responsible oversight in corporate decision-making. Sun et al. (2021) show that female insiders trade less profitably than male insiders and interpret this as consistent with higher ethical standards; they also find that firms with more female-dominated boards are less associated with future financial misconduct. Similarly, Gul et al. (2011) document that board gender diversity improves stock price informativeness, indicating enhanced transparency and information environments. Evidence from sustainability and disclosure research further suggests that female directors promote stronger ethical and socially responsible practices, as firms with greater female board representation are more likely to engage in voluntary environmental disclosure (Ben-Amar et al., 2017) and exhibit higher levels of corporate social responsibility activities and disclosures (Harjoto & Jo, 2011; Harjoto et al., 2015; Harjoto & Rossi, 2019; McGuinness et al., 2017). Collectively, these findings suggest that greater female board representation can strengthen integrity-oriented values and ethical practices within firms.

[Insert Table 6]

Empirical evidence: Table 6 examines the relation between (i) the independent director ratio and the integrity ratio and its subcomponents, and (ii) the female director ratio and the integrity ratio and its subcomponents. All variables are described in Table A1 presented in Appendix. Following the approach of Li et al. (2021), we employ OLS regressions with industry and year fixed effects. We have excluded utility and financial firms from our analysis. Model (1) focuses specifically on the independent director ratio, while Model (2) focuses on the female director ratio. The remaining models (Models (3) to (9)) include both the independent director ratio and the female director ratio simultaneously in the regression specifications. Models (4) to (7) present results for the four integrity subcategories: Accountability, Compliance, Moral Integrity, and Honesty. Models (8) and (9) report results using the integrity ratio constructed following the specification of Li et al. (2021). Models (1) to (3) show a highly significant positive relation between the independent (and female) director ratios and the integrity ratio ($p < 0.01$). The Accountability, Moral Integrity, and Honesty ratios exhibit similar patterns. In contrast, the integrity ratio constructed using the Li et al. (2021) specification shows only a weak relation in Model (8) ($p < 0.10$) and a statistically insignificant relation in Model (9). Overall, these results are consistent with the expectation that firms with a higher proportion of independent and female directors tend to exhibit a stronger orientation toward an integrity-based corporate culture.

4.3.2 Corporate Integrity Culture and Accounting Restatement Disclosure

Non-fraudulent restatement disclosure: An important perspective in examining the relation between corporate integrity culture and accounting restatement disclosure is to distinguish between the incidence of misreporting and the disclosure of previously undetected errors. Firms with strong integrity cultures are not necessarily more prone to misreporting; rather, they are more likely to detect, escalate, and transparently correct reporting problems once identified.

Because restatements can arise from accounting errors, aggressive earnings management, internal control weaknesses, or complex reporting issues, a positive association between integrity culture and restatement disclosure may reflect stronger detection and correction mechanisms rather than poorer reporting quality. Prior research supports this view: the post-SOX environment strengthened ethics and compliance structures, encouraging transparency and corrective behavior (Ahluwalia et al., 2018), while stronger governance and internal controls improve the detection and correction of reporting problems (Ashbaugh-Skaife et al., 2008; Doyle et al., 2007). Evidence also shows that many restatements arise from unintentional errors or internal control deficiencies rather than fraud (Hennes et al., 2008; Palmrose et al., 2004). Ethical culture further promotes internal reporting and escalation of concerns (Kaptein, 2011), and insiders often play a key role in uncovering misconduct (Dyck et al., 2010). Overall, the literature suggests that corporate integrity culture strengthens the detection, escalation, and transparent correction of financial reporting problems, implying that restatements in such firms may reflect transparency and accountability rather than uniformly poor reporting quality.

Fraud-related restatement disclosure: In contrast to error-based restatements, restatements associated with irregularities or accounting fraud should not be positively related to corporate integrity culture. Irregularity restatements typically reflect intentional misreporting rather than transparent correction, and prior research links such misconduct to weak monitoring, deficient controls, and problematic managerial behavior (Beasley, 1996; Hogan et al., 2008; Zahra et al., 2005). Because integrity culture promotes compliance, trustworthiness, and ethical restraint in financial reporting (Altamuro et al., 2021), it should reduce the likelihood of opportunistic manipulation that culminates in fraud-related restatements. Consistent with this view, evidence on executives' integrity commitments shows improvements in reporting quality and reductions in

income-increasing discretionary accruals following integrity-based reforms (Heese et al., 2023). Accordingly, while integrity culture may be positively associated with the disclosure of non-fraudulent errors, it is unlikely to be positively associated with restatements stemming from accounting irregularities or fraud.

[Insert Table 7]

Empirical evidence: We source restatement data from Audit Analytics. As Bao et al. (2019) note, major restatement and financial-misconduct databases differ in their coverage and are suited to different research questions. Because our analysis focuses on the disclosure of non-fraudulent restatements, including accounting errors and internal-control corrections, we use Audit Analytics because of its broad coverage of restatement disclosures, including the types of non-fraudulent corrections central to our analysis.

Employing logit regressions, Table 7 examines the relation between the integrity ratio and its subcomponents and the probability of a non-fraudulent restatement filing (“restat filing”). In the regression setup, *restat_filing* is a dummy variable equal to 1 if the firm files a non-fraudulent restatement in the subsequent year, and 0 otherwise. All other variables are described in Table A1 presented in Appendix. Model (1) focuses specifically on the integrity ratio. Models (2) to (5) present results for the four integrity subcategories: Accountability, Compliance, Moral Integrity, and Honesty. Models (6) and (7) report results using the integrity ratio constructed following the specification of Li et al. (2021).

Model (1) shows a highly significant positive relation between the integrity ratio and the probability of a non-fraudulent restatement filing. Most subcategories exhibit similar results (Models (2) to (5)), with the exception of Moral Integrity. In contrast, the integrity ratios constructed using the Li et al. (2021) specification do not show any statistically significant

relation (Models (6) and (7)). Overall, these findings are consistent with the expectation that firms with a stronger orientation toward an integrity-based corporate culture are more likely to disclose and file non-fraudulent restatements.

We further examine the relation between the integrity ratio and its subcomponents and the probability of fraudulent restatement filings. We do not find any statistically significant relation, which is consistent with the theoretical discussion above.

4.3.3 Corporate Integrity Culture and Earnings Management

A growing body of research suggests that strong corporate integrity culture is associated with lower discretionary earnings management because integrity-oriented cultures emphasize ethical values, transparency, and accountability, thereby discouraging opportunistic reporting behavior (Graham et al., 2022; Kaptein, 2011; Treviño et al., 2006). Prior research shows that ethical culture shapes employee behavior and organizational norms (Kaptein, 2011; Treviño et al., 2006), while corporate culture more broadly influences firm decisions when formal monitoring is incomplete (Guiso et al., 2015). Consistent with this view, executives report that strong ethical cultures discourage earnings manipulation even under short-term performance pressure (Graham et al., 2022). Evidence from the corporate misconduct literature further highlights the role of leadership and culture: firms led by CEOs involved in unethical practices are more likely to engage in financial fraud and opportunistic earnings management (Biggerstaff et al., 2015). Integrity culture may also constrain discretionary earnings management through stronger internal monitoring and governance, including compliance systems and oversight structures that limit managerial discretion. Stronger governance has been shown to reduce abnormal accruals (Klein, 2002), and post-Sarbanes-Oxley ethics and compliance systems have strengthened transparency and accountability in financial reporting (Ahluwalia et al., 2018). Collectively, these findings

suggest that corporate integrity culture, when reinforced through ethical leadership, internal norms, and governance mechanisms, is negatively associated with discretionary earnings management.

[Insert Table 8]

Empirical evidence: Table 8 examines the relation between discretionary accruals (which indicate earnings management or opportunistic accounting practices) and the integrity ratio and its subcomponents. All variables are described in Table A1 presented in Appendix. Following the approach of Li et al. (2021), we employ OLS regressions with industry and year fixed effects. Model (1) focuses on the integrity ratio. Models (2) to (5) present results for the four integrity subcategories: Accountability, Compliance, Moral Integrity, and Honesty. Models (6) and (7) report results using the integrity ratio constructed following the specification of Li et al. (2021). Model (1) shows a significant negative relation between the integrity ratio and discretionary accruals. Among the subcategories, Honesty exhibits a similar result, suggesting that the Honesty component of integrity is primarily driving this finding. The integrity ratios constructed using the Li et al. (2021) specification do not show any statistically significant relation (Models (6) and (7)). Overall, these results are consistent with the expectation that firms with a stronger orientation toward an integrity-based corporate culture are less likely to engage in earnings management.

4.4 Implications of Corporate Integrity Culture: Dividend Policy and Trade Credit

Survey evidence from North American CEOs and CFOs indicates that corporate culture is widely regarded as a key driver of firm value. More than half of the surveyed executives rank culture among the three most influential determinants of firm value, and over 90% believe that strengthening corporate culture can enhance firm value (Graham et al., 2022). Graham et al.

(2022) report that corporate leaders describe culture as “a beliefs system,” “a coordination mechanism,” “an invisible hand,” “how employees interact with one another,” “a standard of behavior,” “norms around how people treat people,” “part work ethic, part ambiance of the work environment,” “how the company really works, the operating style,” and “the tone for what type of company this is.” Building on this view, Li et al. (2021) examine several dimensions of corporate culture in the context of firms’ risk-taking incentives related to earnings management, executive compensation, firm value, and M&A activity, treating corporate integrity as a main constituent of corporate culture. In our study, we concentrate on corporate integrity culture and employ a novel methodology to develop a contextual measure. We now examine the implications of this measure for dividend policy and trade credit.

4.4.1 Corporate integrity culture and dividend policy

Prior research suggests that corporate culture influences firms’ financial policies by shaping managerial behavior and organizational norms. Corporate integrity culture reflects shared values emphasizing ethics, accountability, and transparency within the firm (Li et al., 2021). Firms with stronger integrity cultures are more likely to emphasize credibility and fairness in their interactions with investors, which may encourage policies that visibly return value to shareholders. Dividends represent a direct mechanism through which firms distribute cash to investors and signal commitment to shareholder interests (Guiso et al., 2015; Li et al., 2021).

Integrity culture may also influence dividend policy through its effects on firm performance, financial constraints, and agency conflicts. Firms that promote ethical conduct and transparency often exhibit stronger operational efficiency and improved profitability (Guiso et al., 2015; Hsu, 2007; Jiang et al., 2019), while investors and creditors perceive high-integrity firms as more trustworthy, improving reputation and reducing financing frictions (Bao et al., 2024; Mazzola et

al., 2006). Consequently, firms facing lower financial constraints are better able to distribute cash to shareholders (DeAngelo & DeAngelo, 1990). Consistent with these arguments, recent evidence shows that corporate integrity culture increases dividend payouts; specifically, a one-standard deviation increase in integrity culture is associated with approximately a 9–10% increase in dividend yield (Sinha & Damle, 2025).

Integrity culture may further influence dividend policy by reducing firm risk and agency conflicts. Firms with stronger integrity norms tend to exhibit higher financial reporting quality and lower opportunistic behavior (Bhandari et al., 2022; Biggerstaff et al., 2015), while ethical environments discourage managerial misconduct and agency problems (Balachandran et al., 2025; Obloj & Zemsky, 2015). By promoting transparent disclosure and discouraging opportunistic behavior, integrity culture reduces information asymmetry and agency costs, thereby lowering firm risk (Sinha & Damle, 2025). Managers typically adopt conservative dividend policies because dividend cuts are costly and negatively perceived by investors. When firm risk is lower, managers face less uncertainty about future cash flows and are therefore more willing to commit to stable dividend payments (Hoberg & Prabhala, 2009). From an agency perspective, dividends themselves also serve as a governance mechanism by reducing free cash flow under managerial discretion (Jensen, 1986) and subjecting managers to greater external monitoring through capital market discipline (Easterbrook, 1984). Taken together, these arguments suggest that firms characterized by stronger corporate integrity cultures are likely to distribute a larger share of cash flows to shareholders through dividends.

Table 9 examines the relation between the integrity ratio and its subcomponents and dividends scaled by total assets. All variables are described in Table A1 presented in Appendix. Following the approach of Li et al. (2021), we employ OLS regressions with industry and year

fixed effects. Model (1) focuses specifically on the integrity ratio. Models (2) to (5) present results for the four integrity subcategories: Accountability, Compliance, Moral Integrity, and Honesty. Models (6) and (7) report results using integrity ratio constructed following the approach of Li et al. (2021).

Models (1) to (5) show highly significant positive relations between dividends to total assets and the integrity ratio and its subcategories. In contrast, the integrity ratio constructed using the Li et al. (2021) specification shows only a weak relation, and only in Model (7). Overall, these results are consistent with the expectation that firms with a stronger orientation toward an integrity-based corporate culture tend to pay higher dividends.

[Insert Table 9 and 10]

4.4.2 Corporate integrity culture and trade credit

Corporate integrity culture can play an important role in shaping firms' access to trade credit by influencing the level of trust and perceived reliability in supplier–customer relationships. Trade credit is fundamentally based on suppliers' willingness to allow delayed payment, which exposes them to the risk of opportunistic behavior or non-payment by customers (Petersen & Rajan, 1997; Smith, 1987). In this context, a strong integrity-based corporate culture, characterized by ethical conduct, honesty, and adherence to formal codes of behavior, can serve as an informal governance mechanism that reassures suppliers about the firm's commitment to honoring contractual obligations (Guiso et al., 2015; Kreps, 1990). Firms with high integrity culture are less likely to engage in opportunistic practices, financial misreporting, or stakeholder violations, thereby reducing perceived counterparty risk and transaction uncertainty. As a result, suppliers may be more willing to extend trade credit and offer more favorable credit terms to such firms. Moreover, integrity culture promotes transparency and ethical interactions with stakeholders,

which can enhance the credibility of financial and operational information shared with suppliers (Guiso et al., 2015). Consistent with these arguments, recent empirical evidence shows that firms characterized by stronger integrity-oriented corporate cultures tend to receive higher levels of trade credit from suppliers, as integrity strengthens trust and long-term relational contracting within supply chains (Hasan & Habib, 2019; Ullah et al., 2025).

Table 10 examines the relation between the integrity ratio and its subcomponents and net credit (a proxy for trade credit). Net credit is defined as the difference between accounts payable and accounts receivable, divided by total sales. All variables are described in Table A1 presented in Appendix. Following the approach of Li et al. (2021), we employ OLS regressions with industry and year fixed effects. Model (1) focuses specifically on the integrity ratio. Models (2) to (5) present results for the four integrity subcategories: Accountability, Compliance, Moral Integrity, and Honesty. Models (6) and (7) report results using the integrity ratio constructed following the specification of Li et al. (2021).

Models (1) to (5) show highly significant positive relations between net credit and the integrity ratio and its subcategories. The integrity ratio constructed using the Li et al. (2021) specification also shows a significant relation in Model (6) and a weaker but still statistically significant relation in Model (7). Overall, these results are consistent with the expectation that firms with a stronger orientation toward an integrity-based corporate culture tend to have better access to trade credit. At the same time, trade-credit terms may also reflect a firm's negotiating power and relational "soft skills," such as its ability to build and sustain supplier relationships, rather than integrity alone; although our specifications control for firm size, which captures some variation in bargaining power, other sources of negotiating leverage may remain. Disentangling these channels represents an interesting avenue for future research.

It is also worth noting that across almost all regressions, the magnitude of the coefficients and the relatively low R^2 values are broadly comparable to those reported in prior research, including Li et al. (2021), and are not unexpected given the scale and nature of the variables examined. We therefore focus primarily on the direction, statistical significance, and consistency of the estimated associations across specifications.

5. Discussion

5.1 Methodological Contributions

This study advances IS research on corporate culture by introducing a contextual, multi-agent LLM-based pipeline inspired by blackboard architectural model (Salemi et al., 2025) for measuring corporate integrity culture from earnings call transcripts. Traditional dictionary approaches, reliant on static keyword matching, often misinterpret context and produce false positives whenever integrity-related terms appear as part of proper nouns (such as program or service names), as descriptors of technical or product-related concepts, or as routine discourse markers detached from any cultural meaning (see Table 4). Our findings show that well-structured agentic architecture, organized around the core idea of blackboard architecture, can overcome these limitations by separating fact extraction from contextual judgment and by enabling specialized agents to reason collectively over a shared evidence base. The strict separation of observation from verdict across Phases II and III ensures that classification decisions are anchored in explicitly extracted evidence rather than unconstrained generation, while the shared blackboard provides a persistent reasoning trace that constitutes a verifiable audit trail for every classification. Positioned within the broader evolution of IS methods, this approach marks a shift from non-contextual lexical markers toward deep contextual

interpretation enabled by generative AI. In doing so, we answer Storey et al.'s (2025) call for IS researchers to harness generative AI's potential by developing applications that link technical innovation with organizational and societal impact. Our design demonstrates how generative AI can serve not only as a computational instrument but also as a disciplined reasoning framework for producing actionable insights into corporate culture.

Empirically, the validation results reinforce the robustness of this approach as reported in Section 4. The comparison against the Li et al. (2021) dictionary approach in Section 4.2 provides direct evidence that contextual interpretation is not merely incrementally better but qualitatively different in what it captures. Our dictionary word distribution analysis reveals that more than 70% of the time, these terms appear in contexts that bear no relation to integrity culture. This finding challenges the foundational assumption underlying dictionary-based measurement that keyword presence reliably signals the presence of the intended cultural construct, and underscores why a context-aware approach is necessary for recovering meaningful integrity signals from corporate discourse. More importantly, the external validation analyses in Section 4.3 demonstrate that this contextual discrimination has economically significant information and implications: our measure yields consistent, statistically significant associations with governance quality, restatement disclosure, earnings management, dividend policy, and trade credit access, whereas the Li et al. (2021) dictionary-based measure exhibits systematically weaker and, in most cases, statistically insignificant associations across the same outcome domains. Together, these results establish a scalable, reliable approach for analyzing complex organizational constructs and demonstrate how a multi-agent LLM architecture can expand the methodological repertoire of IS research.

5.2 Theoretical Contributions

Our work advances the theoretical understanding of corporate integrity culture measurement in several important ways. First, we operationalize corporate integrity through a four-component framework encompassing Accountability, Compliance, Moral Integrity, and Honesty, whose organizing logic is derived from crossing two well-established theoretical distinctions: the formal-informal axis of Tenbrunsel et al.'s (2003) ethical infrastructure model and the internal-external governance axis of Aguilera et al. (2008), extended to the ethical domain through Victor and Cullen's (1988) ethical climate theory. This structure provides a parsimonious yet theoretically grounded basis for conceptualizing corporate integrity, equipping researchers with a refined instrument for examining how integrity culture manifests and evolves within organizations.

Second, our findings contribute to the theoretical understanding of how corporate integrity culture manifests in organizational discourse. A foundational premise in the text-based culture measurement literature is that cultural constructs leave lexical traces: firms with stronger integrity cultures use integrity-related words more frequently, and the presence of such words can therefore serve as a reliable indicator of the underlying construct (Li et al., 2021; Loughran & McDonald, 2011). Our analysis provides the first systematic, discourse-level test of this premise for the integrity dimension of corporate culture and we find it does not hold. Across the full Li et al. (2021) integrity wordlist, fewer than 30% of segments containing dictionary terms are classified as meaningfully related to integrity culture, and qualitative inspection confirms that these terms routinely appear as proper nouns, technical descriptors, or conversational fillers devoid of cultural content (Table 4). The external validation analyses in Section 4.3 reinforce this finding from the opposite direction: our contextual measure yields consistent, statistically

significant associations with governance quality, restatement disclosure, earnings management, dividend policy, and trade credit, whereas the dictionary-based specification is insignificant in most models. Together, these results suggest that corporate integrity culture is not a lexical phenomenon — it does not reside in word choice — but a contextual and discursive one, encoded in the meaning, framing, and communicative function of extended passages rather than in the frequency of individual terms. This distinction carries theoretical weight: it implies that the relationship between corporate communication and underlying culture is mediated by context in ways that prior measurement frameworks have not accounted for, and that empirical conclusions drawn from lexical proxies may understate the economic significance of integrity culture by conflating substantive cultural discourse with incidental word use.

Third, we reinforce the theoretical proposition that corporate integrity culture is observable through spontaneous executive discourse. Q&A segments of earnings calls serve as a valuable lens into the “tone at the top,” providing real-time, unscripted expressions of leadership values that are arguably more authentic than prepared remarks. As Li et al. (2021) note, the candid nature of executive responses in this setting makes such discourse particularly informative when cultural themes do arise. Our empirical results lend substantive weight to this proposition. The fact that integrity discourse measured from unscripted Q&A associates significantly with observable indicators of governance quality and financial reporting integrity (Section 4.3) provides evidence that spontaneous managerial communication contains genuine cultural signals, rather than performative or incidental language. This finding contributes to IS research by framing executive communication as a key site where cultural values are enacted and can be systematically analyzed using AI-driven contextual methods, and it suggests that future work on other cultural constructs such as innovation orientation, collaboration orientation etc.

(Fang et al., 2023) may similarly benefit from context-aware analysis of unscripted corporate discourse.

5.3 Practical Implications

Beyond its methodological and theoretical contributions, the Contextual Corporate Integrity Score has practical implications that follow directly from the empirical findings reported in Section 4.3. Because the measure is grounded in the unscripted Q&A discourse of senior management and decomposes integrity into four interpretable components, it offers a tool that boards, investors, and researchers can use to make sense of how a firm's integrity orientation relates to observable governance characteristics and reporting outcomes. Our empirical analyses point to concrete channels through which the measure can inform decision-making. We find that our score is meaningfully related to firms' governance structures, non-fraudulent restatement filings, discretionary accruals, dividend policy, and trade credit access. In each of these settings, the signal carried by our context-aware measure is stronger and more consistent than that of a widely used dictionary-based proxy, indicating that contextual refinement is not merely a technical improvement but one with economic content. For investors, this suggests that integrating the integrity measure into existing screening and risk assessment workflows can surface governance-relevant information that keyword-based approaches miss. For boards and audit committees, the component-level decomposition (Accountability, Compliance, Moral Integrity, and Honesty) enables targeted diagnostics, allowing firms to identify which facets of their integrity culture are most or least emphasized in unscripted managerial communication. For regulators, the ability to track integrity discourse longitudinally across firms and industries offers a complementary lens to traditional compliance metrics, enabling oversight that accounts not only for rule-following but also for the underlying cultural drivers of ethical behavior. More

broadly, our study illustrates how generative AI, when embedded within a structured multi-agent reasoning framework, can be operationalized to inform both organizational governance and broader institutional decision-making.

5.4 Future Research Directions

Our study also suggests several avenues for future research. First, future work could compare open-source and proprietary LLMs to examine whether advances in model reasoning and contextual understanding yield further improvements in integrity detection. Future research could also move beyond the binary identification of integrity discourse by incorporating measures of intensity, sentiment, or the extent of discussion, and by extending the analysis beyond earnings calls to other corporate disclosures. A particularly interesting extension would be to compare integrity discourse in unscripted executive Q&A with more carefully curated disclosures, such as 10-K filings. Differences between these scripted and unscripted settings could provide insights into how firms' espoused integrity compares with how integrity is expressed in spontaneous communication.

Future research could further examine whether CEO and CFO transitions are associated with changes in firms' integrity culture. Such analyses could help distinguish persistent organizational characteristics from the influence of individual executives and their "tone at the top". Another promising avenue is to develop a complementary measure of corruption-oriented corporate culture. Liu (2016) shows that firm-level corruption culture is associated with corporate misconduct, suggesting that integrity and corruption may represent complementary dimensions of firms' ethical orientation. Examining whether a contextual corruption-culture measure constructed using our framework exhibits systematically opposite associations with

governance and reporting outcomes could provide further evidence on the validity and broader scope of the integrity construct.

Finally, future research could examine the extent to which integrity culture remains informative under conditions that create stronger incentives for opportunistic behavior, such as financial distress. Studying whether the documented relation between integrity culture and reporting outcomes persists in such settings could provide further insight into the stability and behavioral relevance of the construct.

6. Conclusion

This study introduces a novel approach to measuring corporate integrity culture by embedding a reasoning LLM within a multi-agent architecture and applying it to the unscripted Q&A portions of quarterly earnings calls. The resulting Contextual Corporate Integrity Score is a theoretically grounded, firm-quarter measure built on a four-component framework (Accountability, Compliance, Moral Integrity, and Honesty) whose logical completeness is established by crossing the formal-informal distinction of Tenbrunsel et al. (2003) with the internal-external governance distinction of Aguilera et al. (2008). We further inform the internal-external distinction to the ethical domain through Victor and Cullen's (1988) ethical climate theory. Unlike survey- or dictionary-based approaches, our measure captures integrity discourse directly from senior management in real-time, context-dependent communication, and provides both aggregate and component-level scores that can be tracked over time.

We validate the resulting measure against a set of well-recognized corporate integrity indicators. Firms with stronger integrity cultures exhibit stronger governance characteristics, are more likely to transparently disclose non-fraudulent restatements, and engage in less

discretionary earnings management. Extending the analysis to firms' financial policies, we further find that integrity culture is associated with higher dividend payouts and greater access to trade credit. Across these analyses, our contextual measure produces consistent, theoretically predicted associations, whereas the dictionary-based benchmark yields systematically weaker and, in most cases, statistically insignificant results. This pattern establishes that contextual discrimination carries genuine economic content.

The study contributes to IS research by showing how generative AI can move from a tool for pattern recognition to a disciplined reasoning framework that links technical innovation with organizational and societal impact. It contributes to the corporate culture literature by introducing a validated, component-level integrity measure that extends Guiso et al. (2015) and Li et al. (2021), and to the growing literature on generative AI in management and accounting research by operationalizing an intangible governance construct at scale.

This work lays a theoretical and methodological foundation for corporate integrity research via LLM-driven textual analysis. Using open-source models ensures reproducibility and supports future extensions. By operationalizing corporate integrity in this way, we open avenues to explore its antecedents, effects on firm outcomes, and role in stakeholder relationships, and we highlight the strategic role of IS in transforming unstructured corporate communication into decision-relevant organizational knowledge.

Appendix A: Variable description, Descriptive Statistics and Supplementary Discussions

Table A1. Variable description

Variable	Description
Integrity ratio	Ratio of integrity related earnings call Q&A segments as identified by the LLM model to total earnings call Q&A segments.
Accountability	Ratio of accountability integrity related earnings call Q&A segments as identified by the LLM model to total earnings call Q&A segments.
Compliance	Ratio of compliance integrity related earnings call Q&A segments as identified by the LLM model to total earnings call Q&A segments.
Moral	Ratio of moral integrity related earnings call Q&A segments as identified by the LLM model to total earnings call Q&A segments.
Honesty	Ratio of honesty integrity related earnings call Q&A segments as identified by the LLM model to total earnings call Q&A segments.
Integrity Li tfidf	Integrity ratio in the spirit of Li et al. (2021). It represents weighted-frequency count of integrity-related words in the Q&A section of earnings calls. We use both 1-year measure and past 3-year average measure in our regression set-up. The variable is measured as follows: $\text{score}_{\text{TFIDF}}(d) = \frac{100}{L_d} \sum_{w \in D} \text{tf}_{w,d} \cdot \ln\left(\frac{N}{\text{df}_w}\right)$ Where, d denote one transcript's concatenated Q&A document D is the integrity dictionary w a dictionary word tf_{w,d} the raw count of w in d L_d = len(document.split()) is the document's token/word count (or document length). N is the corpus size (number of documents with L_d > 0) df_w is the number of corpus documents in which w appears at least once
Integrity Li wtfidf	It is a another modified measure of integrity ratio as presented in Li et al. (2021). It represents weighted-frequency count of integrity-related words in the Q&A section of earnings calls. However, it does sublinear scaling of the term-frequency component. We use both 1-year measure and past 3-year average measure in our regression set-up. The variable is measured as follows: $\text{score}_{\text{WTFIDF}}(d) = \frac{100}{L_d} \sum_{w \in D} (1 + \ln(\text{tf}_{w,d})) \cdot \ln\left(\frac{N}{\text{df}_w}\right)$ All components used in the formula are described above.
Independent director ratio	Ratio of independent directors of a board to the total number of directors.
Female director ratio	Ratio of female directors of a board to the total number of directors.
Restat filing	It is a dummy variable; it is equal to '1' if the firm files for a non-fraudulent restatement in the subsequent year, or '0' otherwise.
Disc. accrual	Estimated using modified Jones' (1991) model. In each year, we run the following regression within each two-digit SIC industry: $\text{Accrual/TA} = \alpha + \beta \times 1/\text{TA} + \gamma \times \Delta \text{Sales/TA} + \theta \times \text{PPE/TA}$
Div to total asset	Ratio of total dividend to total assets
Net credit	Difference between accounts payable and accounts receivable, divided by total sales.
Firm size	Natural logarithm of market value of equity
Leverage	Ratio of total debt to total asset
R&D to asset	Ratio of R&D expense to total asset
ROA	Ratio of EBITDA to total asset

Sales growth	(Salest/ Salest-1) - 1
Standard deviation ROA	Standard deviation of last five years' ROA

Table A2. Descriptive Statistics

This table reports descriptive statistics for the dependent variables, segment-based integrity measures, Li et al. (2021) benchmark measures, and control variables used in the validation tests.

Variables	N	Mean	SD	P25	Median	P75
Integrity ratio	38922	0.0417	0.0316	0.0200	0.0357	0.0564
Accountability	38922	0.0268	0.0235	0.0106	0.0217	0.0370
Compliance	38922	0.0116	0.0161	0.0000	0.0068	0.0162
Moral	38922	0.0149	0.0170	0.0000	0.0106	0.0214
Honesty	38922	0.0222	0.0196	0.0087	0.0183	0.0309
Integrity Li tfidf	38922	0.3549	0.3023	0.1688	0.2890	0.4594
Integrity Li wtfidf	38922	0.3135	0.2464	0.1577	0.2641	0.4099
Independent director ratio	33196	0.7209	0.1718	0.6667	0.7500	0.8333
Female director ratio	33196	0.1854	0.1363	0.1000	0.1818	0.2857
Restat filing	35001	0.0825	0.2751	0.0000	0.0000	0.0000
Disc. accrual	32327	0.0024	0.1180	-0.0359	0.0066	0.0470
Div to total asset	38740	0.0373	0.0770	0.0000	0.0105	0.0458
Net credit	37930	-0.0415	0.8058	-0.1357	-0.0654	-0.0005
Firm size	38765	7.2785	1.9976	5.9032	7.3058	8.6267
Leverage	38621	0.2493	0.2216	0.0442	0.2172	0.3883
R&D to asset	38761	0.0582	0.1196	0.0000	0.0020	0.0643
ROA	38752	0.0023	0.2313	-0.0070	0.0552	0.1043
Sales growth	36510	0.1409	0.4990	-0.0321	0.0649	0.1925
Standard deviation ROA	38223	0.0289	0.0323	0.0086	0.0171	0.0346

Supplementary Discussions

SD1. Validation of the BGE-M3 segmentation variant against the C99 TF-IDF baseline

The C99 algorithm introduced by Choi (2000) is a widely used method for unsupervised linear text segmentation. In its original formulation, C99 represents each sentence as a TF-IDF bag-of-words vector, constructs a pairwise inter-sentence similarity matrix, applies a rank transform that replaces raw similarity values with their local rank, and then recovers segment boundaries through divisive clustering. The procedure is domain independent and has proved robust enough to be adopted in recent finance research as a principled way to decompose unstructured corporate disclosures into topically coherent units. In particular, Li et al., (2026) rely on C99 in their study

of corporate culture using generative AI, which gives the algorithm a well-established place in the corporate-culture measurement literature that sits closest to our own work.

Our pipeline preserves the full C99 mechanism and modifies only the sentence representation. We replace the TF-IDF vectors used to construct the similarity matrix with BGE-M3 dense sentence embeddings. The motivation is conceptual rather than ad hoc. TF-IDF captures surface lexical co-occurrence and is sparse, context-free, and sensitive to paraphrase. Two sentences that convey the same idea using different vocabulary will have low TF-IDF similarity, whereas two sentences that share stopword-heavy phrasing but discuss entirely different topics can have deceptively high similarity. Contextual dense embeddings such as BGE-M3 place sentences in a continuous semantic space in which paraphrases and domain-specific figurative language cluster together, and in which superficially similar but topically distinct sentences are separated. For earnings calls, where managers and analysts paraphrase freely across long turns, a representation that encodes meaning rather than lexical overlap is a priori better suited to identifying topical boundaries.

This conceptual advantage is not self-justifying, however. C99 is an established benchmark and the literature that uses it has produced a body of calibrated findings. Any modification to its representation layer has to be checked against the benchmark before downstream results can be read with confidence. The question addressed in this section is therefore narrower than a full robustness check of our integrity-classification pipeline. It is whether our BGE-M3 variant of C99 produces transcript-level integrity measures that are statistically and practically indistinguishable from those produced by the original TF-IDF version of C99. A null result in that comparison is the outcome that validates the variant: it establishes that our modification has not drifted the measurement away from the established baseline, and that we gain the interpretive benefits of a contextual embedding without losing comparability to prior C99-based work.

Empirical design

From the full sample of transcripts, we drew 3,000 transcript identifiers at random and partitioned them into three non-overlapping sub-samples of 1,000 transcripts each. Every transcript in every sub-sample was segmented twice, once with the BGE-M3 variant and once with the TF-IDF version of C99 algorithm, and each resulting set of segments was passed through the same multi-agent integrity-classification pipeline. For each transcript we therefore

have paired observations on the outcome measures under the two segmentation methods, and we test for equality of method means using two-tailed paired-sample t-tests.

The test is paired because each transcript serves as its own control: the two methods are applied to the same underlying text, so within-transcript differences reflect the method rather than across-transcript heterogeneity in firm coverage, call length, or industry context. The test is two-tailed because we impose no prior on the direction of the effect. We ask whether the population mean difference between BGE-M3 and TF-IDF equals zero in either direction, which is the appropriate null when the goal is to establish equivalence. Within this design, the outcome we want is a failure to reject the null: a non-significant result is what signals agreement between the two methods. Replicating the test on three non-overlapping sub-samples of 1,000 transcripts provides evidence that any non-rejection is not the artifact of a single chance draw; consistent agreement across three independent sub-samples is a considerably stronger basis for inference than a single-sample result with the same total size.

Three transcript-level outcomes (or variables) assess whether the two segmentation methods agree on the integrity culture signal extracted from the earnings call transcripts. The Yes-segment count measures the absolute volume of integrity-relevant material each method identifies. The Yes / Total ratio normalizes that count by the full number of segments produced and is therefore invariant to mechanical differences in segment granularity between the two methods. The Yes / No ratio expresses the odds of integrity-relevant classification among segments that were classified definitively. Together, these three measures capture both the volume and the density of integrity discussion detected by the pipeline.

Results

Appendix B Table B1 reports the paired-sample t-test results for the three integrity-relevant outcome measures across the three sub-samples. The sample size N is exactly 1,000 for the Yes-segment count, which is defined for every transcript, but falls slightly below 1,000 for the ratio measures. This occurs because the Yes / No ratio is mathematically undefined when a transcript has no No-classified segments (division by zero), so such transcripts are dropped pairwise from the row for that ratio; the Yes / Total ratio excludes transcripts with zero total segments (i.e., those whose Q&A section is missing) on the same logic. These pairwise exclusions explain the N values of 987 to 991 observed across sub-samples for the ratio rows, and they remain close to the original sample size of 1,000. The Mean columns summarize the transcript-level distribution of

the outcome under each method. The Mean Difference column reports the paired difference (BGE-M3 minus TF-IDF) averaged across transcripts. The 95% confidence interval is the range within which the true population mean difference is expected to lie with 95% probability under repeated sampling; a confidence interval that straddles zero is consistent with the null of no difference between methods.

The t statistic is the mean difference divided by its standard error, and the number in parentheses is the degrees of freedom (df), equal to $N - 1$ for a paired-sample design; the loss of one degree of freedom reflects the fact that one parameter (the sample mean difference) has been estimated from the data. The two-tailed p value is the probability of observing a mean difference at least as extreme in absolute value as the one observed if the null of no method difference were true. Finally, Cohen's d for paired samples is the mean difference scaled by the standard deviation of the paired differences and provides a unit-free effect-size measure. Following Cohen (1988), absolute values below roughly 0.20 are typically treated as negligible, values around 0.20 as small, around 0.50 as medium, and 0.80 or above as large.

Discussion of findings

Across all three outcomes, presented in Panel A, B, and C of Appendix B Table B1, and across all three sub-samples, the pattern is uniform. The mean differences are below 0.03 segments per transcript for the Yes-segment count (see Panel A) and in the third or fourth decimal place for the ratio measures (see Panel B and C). Every 95% confidence interval contains zero. No p value falls anywhere near the 5% threshold; the smallest of the nine p values is 0.3388. Every Cohen's d is below 0.03 in absolute value, comfortably within the negligible band. Perhaps most telling, the sign of the mean difference alternates between positive and negative across sub-samples for both ratio measures, which is precisely what one would expect under the null of no systematic difference: random sampling noise pushes the point estimate around zero without any consistent direction. If the methods were in fact producing different integrity signals we would expect the sign to be stable across independent samples, and it is not.

The reading of this evidence is unambiguous. The BGE-M3 and TF-IDF variants of C99 produce transcript-level integrity measures that are statistically and practically indistinguishable. Both methods identify essentially the same number of integrity-relevant segments per transcript, and the proportion of Yes segments, whether normalized by the total segment count or expressed as odds against No segments, is effectively equal under the two methods in every sub-sample.

The reason to emphasize this finding is narrower than a general claim of pipeline invariance. It is that our variant of C99 has been validated against the original C99 in its TF-IDF form, which is the benchmark used by Li et al. (2026) and a large prior body of segmentation research. We are not asserting that the integrity measure would be unchanged under an arbitrary segmentation method; we are asserting that replacing TF-IDF with BGE-M3 inside the C99 framework does not move the integrity measure away from the established baseline. This provides a defensible basis for reporting our subsequent results: the reader can interpret them alongside prior C99-based findings without the concern that a non-standard segmentation has silently altered the measurement.

Equivalence at the integrity-measure level does not, however, imply that the two methods produce identical segmentations of the underlying text. Qualitative inspection of the BGE-M3 output reveals a pattern worth noting. BGE-M3 produces a non-trivial number of single-sentence segments that correspond to earnings-call boilerplate: opening salutations, operator hand-offs, brief acknowledgments, thank-you sentences, and closing courtesies. These occurrences do not cohere topically with the surrounding discussion, and a contextual embedding correctly places them at a semantic distance from their neighbors, causing the C99 boundary step to isolate them as standalone segments. TF-IDF, which is dominated by common function words and generic vocabulary in such sentences, tends to absorb them into an adjacent multi-sentence segment. The dense embedding is therefore doing something the sparse representation cannot: it is recognizing when a sentence does not belong with its neighbors, even when the two sentences share surface vocabulary. This is the substantive upside we hoped the embedding replacement would deliver, and it is consistent with the broader finding in the segmentation literature that more informative sentence representations yield cleaner boundaries.

Crucially, this upside does not distort the integrity measure. The isolated boilerplate segments carry no integrity content and are uniformly classified as “No” by the downstream classifier. Whether the algorithm records them as a standalone one-sentence segment or folds them into a longer neighbor segment therefore has no measurable effect on the Yes count or on either of the Yes ratios. The integrity measure is invariant to the point along the segmentation continuum at which this non-topical material is handled. It follows that our BGE-M3 variant gains the interpretive benefit of isolating boilerplate without paying any cost in the fidelity of the integrity measurement.

Taken together, the evidence in Appendix B Table B1 supports treating the BGE-M3 variant of C99 as a defensible replacement for the original TF-IDF version in this measurement context. Integrity-relevant outcomes are statistically indistinguishable between the two methods across three independent sub-samples of 1,000 transcripts each, with every Cohen's d in the negligible range and every p value well above conventional significance thresholds. Our modification of C99 is therefore not a departure from the established segmentation benchmark. It is at worst an equivalent reformulation, and at best a modestly sharper one that isolates non-topical boilerplate into short standalone segments without compromising the identification of integrity-relevant discussion. The replacement of the TF-IDF similarity matrix with a BGE-M3 similarity matrix inside the C99 framework is, on this evidence, methodologically justified.

SD2. Integrity Components Identification

Guided by the theoretical framework developed in Section 2.2, our multi-agent pipeline operationalizes corporate integrity along four dimensions. Appendix B Table B6 reports the resulting distribution, detailing the number and percentage of integrity-related segments in which each component is identified by the pipeline. Because a single segment may simultaneously invoke multiple dimensions of integrity culture, and because the multi-agent design explicitly accommodates such co-occurrence through independent agent verdicts, the cumulative percentage across components exceeds 100%. Appendix B Table B7 provides illustrative examples of how the multi-agent pipeline identified integrity-related discourse across different components. These examples illustrate the system's capability to recognize various dimensions of integrity culture, from explicit ethical statements to discussions of transparency, accountability, and responsible business practices. The diversity of examples across integrity components confirms that our approach captures both the explicit "talk" and the implicit "walk" of integrity culture as revealed through managerial communications.

SD3. Rationale and Design Trade-Offs of the Phase I Gating Mechanism

The rationale for the Phase I gating design, described in Section 3.3.2, merits explicit discussion, as it reflects a deliberate trade-off rather than a constraint imposed by the architecture. In principle, the downstream phases of the pipeline are sufficiently comprehensive to allow all segments to be routed directly through Phases II–IV, bypassing Phase I altogether. Under such a

configuration, the Closest-Fit Resolution Agent in Phase IV could be retained as an optional ambiguity-resolution mechanism, invoked only when the Synthesizer Agent identifies persistent disagreement among the domain agents, or omitted entirely depending on the use case. This alternative design would maximize the contextual depth of each classification decision.

In practice, however, two considerations make Phase I gating the preferable default in our study. First, as Li et al. (2021) observe, corporate culture surfaces in earnings calls as an unintended byproduct of largely informational discussions of firm operations and performance by senior executives, rather than as the focal subject of the conversation. The empirical implication is that integrity-related discourse occupies only a small fraction of the overall corpus, rendering uniform processing of all segments computationally inefficient and disproportionately costly relative to the marginal analytical value gained. Second, because Phase I occupies a critical upstream position, any misclassification at this stage has the potential to propagate through all subsequent phases. This makes its performance particularly consequential and necessitates rigorous validation, a focus to which we return in Section 4.1.

Accordingly, we adopt Phase I gating as the default operating mode for large-scale corpora, with particular emphasis placed on ensuring its reliability. Researchers working with smaller samples, or those prioritizing maximal contextual fidelity for each segment, may instead choose to bypass Phase I and route all segments through the rest of the agentic pipeline proposed in this study. In such cases, the Closest-Fit Resolution Agent in Phase IV may be retained or omitted depending on the desired balance between computational efficiency and exhaustive multi-agent deliberation.

References

- Agrawal, A., & Chadha, S. (2005). Corporate governance and accounting scandals. *The Journal of Law and Economics*, 48(2), 371–406.
- Aguilera, R. V., Filatotchev, I., Gospel, H., & Jackson, G. (2008). An organizational approach to comparative corporate governance: costs, contingencies, and complementarities. *Organization Science*, 19(3), 475–492.
- Ahluwalia, S., Ferrell, O. C., Ferrell, L., & Rittenburg, T. L. (2018). Sarbanes–Oxley section 406 code of ethics for senior financial officers and firm behavior. *Journal of Business Ethics*, 151(3), 693–705.
- Allee, K. D., & Deangelis, M. D. (2015). The Structure of Voluntary Disclosure Narratives: Evidence from Tone Dispersion. *Journal of Accounting Research*, 53(2), 241–274.
- Altamuro, J. L. M., Gray, J. V., & Zhang, H. (2021). Corporate Integrity Culture and Compliance: A study of the pharmaceutical industry*. *Contemporary Accounting Research*, 39(1), 428–458.
- Arjoon, S. (2005). Corporate Governance: an Ethical perspective. *Journal of Business Ethics*, 61(4), 343–352.
- Arnaud, A., & Schminke, M. (2011). The Ethical Climate and Context of Organizations: A Comprehensive Model. *Organization Science*, 23(6), 1767–1780.
- Ashbaugh-Skaife, H., Collins, D. W., Kinney, W. R., & LaFond, R. (2008). The effect of SOX internal control deficiencies and their remediation on accrual quality. *The Accounting Review*, 83(1), 217–250.
- Audi, R., Loughran, T., & McDonald, B. (2016). Trust, but Verify: MD&A Language and the Role of Trust in Corporate Culture. *Journal of Business Ethics*, 139, 551–561.
- Audi, R., & Murphy, P. E. (2006). The many faces of integrity. *Business Ethics Quarterly*, 16(1), 3–21.
- Bai, J., Boyson, N. M., Cao, Y., Liu, M., & Wan, C. (2023). Executives vs. Chatbots: Unmasking Insights through Human-AI Differences in Earnings Conference Q&A. *SSRN Electronic Journal*.
- Balachandran, B., Faff, R. W., Mishra, S., & Shams, S. (2025). Target firm’s integrity culture and M&A performance. *Journal of Business Finance & Accounting*, 52(1), 433–471.
- Bao, Y., Ke, B., Li, B., Yu, Y. J., & Zhang, J. (2019). Detecting accounting fraud in publicly traded U.S. firms using a machine learning approach. *Journal of Accounting Research*, 58(1), 199–235.
- Bao, X., Han, M., Lau, R., & Xu, X. (2024). Corporate integrity culture and credit rating assessment. *Journal of International Financial Markets Institutions and Money*, 93, 102007.
- Beasley, M. S., Carcello, J. V., Hermanson, D. R., & Lapides, P. D. (2000). Fraudulent Financial Reporting: consideration of industry traits and corporate governance mechanisms. *Accounting Horizons*, 14(4), 441–454.

- Beasley, M. S. (1996). An Empirical Analysis of the Relation between the Board of Director Composition and Financial Statement Fraud. *The Accounting Review*, 71(4), 443–465.
- Becker, T. E. (1998). Integrity in organizations: beyond honesty and conscientiousness. *Academy of Management Review*, 23(1), 154–161.
- Beckmann, L., Beckmeyer, H., Filippou, I., Menze, S., & Zhou, G. (2024). Unusual Financial Communication - Evidence from ChatGPT, Earnings Calls, and the Stock Market. *SSRN Electronic Journal*.
- Ben-Amar, W., Chang, M., & McIlkenny, P. (2017). Board Gender Diversity and Corporate Response to Sustainability Initiatives: Evidence from the Carbon Disclosure Project. *Journal of Business Ethics*, 142(2), 369–383.
- Bhandari, A., Mammadov, B., Thevenot, M., & Vakilzadeh, H. (2022). Corporate culture and financial reporting quality. *Accounting Horizons*, 36(1), 1–24.
- Bhatt, P., Vemprala, N., Valecha, R., Hariharan, G., & Rao, H. R. (2023). User Privacy, Surveillance and Public Health during COVID-19 – An Examination of Twitterverse. *Information Systems Frontiers*, 25(5), 1667–1682.
- Biggerstaff, L., Cicero, D. C., & Puckett, A. (2015). Suspect CEOs, unethical culture, and corporate misbehavior. *Journal of Financial Economics*, 117(1), 98–121.
- Bochkay, K., Brown, S. V., Leone, A. J., & Tucker, J. W. (2023). Textual analysis in accounting: What's next? *SSRN Electronic Journal*.
- Bochkay, K., Hales, J., & Chava, S. (2020). Hyperbole or reality? Investor response to extreme language in earnings conference calls. *The Accounting Review*, 95(2), 31–60.
- Brown, M. E., & Treviño, L. K. (2006). Ethical leadership: A review and future directions. *The Leadership Quarterly*, 17(6), 595–616.
- Brown, M. E., Treviño, L. K., & Harrison, D. A. (2005). Ethical leadership: A social learning perspective for construct development and testing. *Organizational Behavior and Human Decision Processes*, 97(2), 117–134.
- Bushman, R. M., Piotroski, J. D., & Smith, A. J. (2004). What determines corporate transparency? *Journal of Accounting Research*, 42(2), 207–252.
- Cao, S., Jiang, W., Yang, B., & Zhang, A. L. (2023). How to talk when a machine is listening: Corporate Disclosure in the age of AI. *Review of Financial Studies*, 36(9), 3603–3642.
- Chen, J., Xiao, S., Zhang, P., Luo, K., Lian, D., & Liu, Z. (2024). M3-Embedding: Multi-Linguality, Multi-Functionality, Multi-Granularity text embeddings through Self-Knowledge distillation. *arXiv (Cornell University)*.
- Chen, L., Zaharia, M., & Zou, J. (2023). *How is ChatGPT's behavior changing over time?* arXiv.org.
- Choi, F. Y. Y. (2000). *Advances in domain independent linear text segmentation*. arXiv.org.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.

- Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology*, 86(3), 386–400.
- Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2010). Signaling Theory: A Review and assessment. *Journal of Management*, 37(1), 39–67.
- Davis, J. H., Schoorman, F. D., & Donaldson, L. (1997). TOWARD a STEWARDSHIP THEORY OF MANAGEMENT. *Academy of Management Review*, 22(1), 20–47.
- DeAngelo, H., & DeAngelo, L. (1990). Dividend policy and Financial distress: An Empirical investigation of troubled NYSE firms. *The Journal of Finance*, 45(5), 1415–1431.
- Demers, E., Wang, V. X., & Wu, K. (2024). Measuring Corporate Human capital disclosures: lexicon, data, code, and research opportunities. *Journal of Information Systems*, 38(2), 163–186.
- Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. arXiv.org.
- de Kok, T. (2025). ChatGPT for textual Analysis? How to use Generative LLMs in Accounting Research. *Management Science*, 71(9), 7888–7906.
- DiMaggio, P. J., & Powell, W. W. (1983). The Iron Cage revisited: institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48(2), 147.
- Dirks, K. T., & Ferrin, D. L. (2002). Trust in leadership: Meta-analytic findings and implications for research and practice. *Journal of Applied Psychology*, 87(4), 611–628.
- Doyle, J. T., Ge, W., & McVay, S. (2007). Accruals Quality and Internal Control over Financial Reporting. *The Accounting Review*, 82(5), 1141–1170.
- Dutta, S., Kumar, A., Pant, P., Walsh, C., & Dutta, M. (2023). Using 10-K text to gauge COVID-related corporate disclosure. *PLoS ONE*, 18(3), e0283138.
- Dyck, A., Morse, A., & Zingales, L. (2010). Who blows the whistle on corporate fraud? *The Journal of Finance*, 65(6), 2213–2253.
- Easterbrook, F. H. (1984). Two Agency-Cost Explanations of Dividends. *The American Economic Review*, 74(4), 650–659.
- Erhard, W., Jensen, M. C., & Zaffron, S. (2009). Integrity: A Positive Model that Incorporates the Normative Phenomena of Morality, Ethics and Legality. *SSRN Electronic Journal*.
- Fama, E. F. (1980). Agency problems and the theory of the firm. *Journal of Political Economy*, 88(2), 288–307.
- Fama, E. F., & Jensen, M. C. (1983). Separation of ownership and control. *The Journal of Law and Economics*, 26(2), 301–325.
- Fang, Y., Fiordelisi, F., Hasan, I., Leung, W. S., & Wong, G. (2023). Corporate culture and firm value: Evidence from crisis. *Journal of Banking & Finance*, 146, 106710.
- Fiordelisi, F., & Ricci, O. (2014). Corporate culture and CEO turnover. *Journal of Corporate Finance*, 28, 66–82.

- Frankel, R., Johnson, M., & Skinner, D. J. (1999). An empirical examination of conference calls as a voluntary disclosure medium. *Journal of Accounting Research*, 37(1), 133.
- Fuerst, M. J., Luetge, C., Max, R., & Kriebitz, A. (2023). Toward organizational integrity measurement: Developing a theoretical model of organizational integrity. *Business and Society Review*, 128(3), 417–435.
- Graham, J. R., Grennan, J., Harvey, C. R., & Rajgopal, S. (2022). Corporate culture: Evidence from the field. *Journal of Financial Economics*, 146(2), 552–593.
- Greenberg, J. (1987). A taxonomy of Organizational Justice theories. *Academy of Management Review*, 12(1), 9–22.
- Guiso, L., Sapienza, P., & Zingales, L. (2015). The value of corporate culture. *Journal of Financial Economics*, 117(1), 60–76.
- Gul, F. A., Srinidhi, B., & Ng, A. C. (2011). Does board gender diversity improve the informativeness of stock prices? *Journal of Accounting and Economics*, 51(3), 314–338.
- Guo, D., Yang, D., Zhang, H., Song, J., Wang, P., Zhu, Q., Xu, R., Zhang, R., Ma, S., Bi, X., Zhang, X., Yu, X., Wu, Y., Wu, Z. F., Gou, Z., Shao, Z., Li, Z., Gao, Z., Liu, A., . . . Zhang, Z. (2025). DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning. *Nature*, 645(8081), 633–638.
- Hall, A. T., Frink, D. D., & Buckley, M. R. (2017). An accountability account: A review and synthesis of the theoretical and empirical research on felt accountability. *Journal of Organizational Behavior*, 38(2), 204–224.
- Harjoto, M. A., & Jo, H. (2011). Corporate Governance and CSR Nexus. *Journal of Business Ethics*, 100(1), 45–67.
- Harjoto, M. A., & Rossi, F. (2019). Religiosity, female directors, and corporate social responsibility for Italian listed companies. *Journal of Business Research*, 95, 338–346.
- Harjoto, M., Laksmana, I., & Lee, R. (2015). Board Diversity and Corporate Social Responsibility. *Journal of Business Ethics*, 132(4), 641–660.
- Hassan, T. A., Hollander, S., Van Lent, L., & Tahoun, A. (2019). Firm-Level Political Risk: Measurement and Effects. *The Quarterly Journal of Economics*, 134(4), 2135–2202.
- Healy, P. M., & Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics*, 31(1–3), 405–440.
- Heese, J., Pérez-Cavazos, G., & Peter, C. D. (2023). When Executives pledge integrity: The effect of the accountant's oath on firms' financial reporting. *The Accounting Review*, 98(7), 261–288.
- Hennes, K. M., Leone, A. J., & Miller, B. P. (2008). The Importance of Distinguishing Errors from Irregularities in Restatement Research: The Case of Restatements and CEO/CFO Turnover. *The Accounting Review*, 83(6), 1487–1519.

- Henry, E., & Leone, A. J. (2015). Measuring Qualitative Information in Capital Markets Research: Comparison of alternative methodologies to measure disclosure tone. *The Accounting Review*, 91(1), 153–178.
- Hernandez, M. (2012). Toward an understanding of the psychology of stewardship. *Academy of Management Review*, 37(2), 172–193.
- Hoberg, G., & Prabhala, N. R. (2009). Disappearing dividends, catering, and risk. *Review of Financial Studies*, 22(1), 79–116.
- Hogan, C. E., Rezaee, Z., Riley, R. A., & Velury, U. K. (2008). Financial Statement Fraud: Insights from the Academic Literature. *Auditing a Journal of Practice & Theory*, 27(2), 231–252.
- Hoitash, R., & Hoitash, U. (2022). A measure of firm complexity: data and code. *Journal of Information Systems*, 36(2), 161–172.
- Hsu, D. H. (2007). Experienced entrepreneurial founders, organizational capital, and venture capital funding. *Research Policy*, 36(5), 722–741.
- Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A Large Language Model for Extracting Information from Financial Text*. *Contemporary Accounting Research*, 40(2), 806–841.
- Huang, J., & Chang, K. C. (2022). *Towards Reasoning in large Language Models: A survey*. arXiv.org.
- Irugalbandara, C., Mahendra, A., Daynauth, R., Arachchige, T. K., Dantanarayana, J., Flautner, K., Tang, L., Kang, Y., & Mars, J. (2023). *Scaling Down to Scale Up: A Cost-Benefit Analysis of Replacing OpenAI's LLM with Open Source SLMs in Production*. arXiv.org.
- Jensen, M. C. (1986). Agency costs of free cash flow, corporate finance, and takeovers. *The American Economic Review*, 76(2), 323–329.
- Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4), 305–360.
- Jha, M., Qian, J., Weber, M., & Yang, B. (2023). ChatGPT and Corporate Policies. *SSRN Electronic Journal*.
- Ji, Y., Rozenbaum, O., & Welch, K. T. (2017). Corporate Culture and Financial reporting risk: Looking through the glassdoor. *SSRN Electronic Journal*.
- Jiang, F., Kim, K. A., Ma, Y., Nofsinger, J. R., & Shi, B. (2019). Corporate Culture and Investment–Cash Flow sensitivity. *Journal of Business Ethics*, 154(2), 425–439.
- Kaptein, M. (2009). Ethics Programs and Ethical Culture: A next step in unraveling their Multi-Faceted relationship. *Journal of Business Ethics*, 89(2), 261–281.
- Kaptein, M. (2011). Understanding unethical behavior by unraveling ethical culture. *Human Relations*, 64(6), 843–869.
- Kaptein, M., & Wempe, J. (2002). The balanced company: A Theory of Corporate Integrity. In *Oxford University Press eBooks*. Oxford University Press.

- Kaur, H., Ahsaan, S. U., Alankar, B., & Chang, V. (2021). A proposed Sentiment Analysis deep learning algorithm for analyzing COVID-19 tweets. *Information Systems Frontiers*, 23(6), 1417–1429.
- Keiser, N. L., & Payne, S. C. (2019). Are employee surveys biased? Impression management as a response bias in workplace safety constructs. *Safety Science*, 118, 453–465.
- Kim, A., Muhn, M., & Nikolaev, V. (2025). Bloated disclosures: Can ChatGPT help investors process information? *arXiv (Cornell University)*.
- Klein, A. (2002). Audit committee, board of director characteristics, and earnings management. *Journal of Accounting and Economics*, 33(3), 375–400.
- Kreps, D. M. (1990). Corporate culture and economic theory. In: J. E. Alt & K. A. Shepsle (Eds.), *Perspectives on positive political economy* (pp. 90–143). Cambridge University Press.
- Kurumathur, S. K., Bhatt, P., Valecha, R., Hariharan, G., & Rao, H. R. (2025). Examination of Polarization in Social Media in Aggressor-Oriented and Victim-Oriented Discourse Following Vigilantism. *Information Systems Frontiers*, 27(4), 1393–1412.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159.
- Larcker, D. F., & Zakolyukina, A. A. (2012). Detecting deceptive discussions in conference calls. *Journal of Accounting Research*, 50(2), 495–540.
- Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, 125(2), 255–275.
- Li, F. (2010). The information Content of Forward-Looking Statements in Corporate Filings—A naïve Bayesian Machine Learning approach. *Journal of Accounting Research*, 48(5), 1049–1102.
- Li, K., Mai, F., Shen, R., & Yan, X. (2021). Measuring corporate culture using machine learning. *Review of Financial Studies*, 34(7), 3265–3315.
- Li, K., Mai, F., Shen, R., Yang, C., & Zhang, T. (2026). Dissecting corporate culture using generative AI. *Review of Financial Studies*, 39(1), 253–296.
- Liu, X. (2016). Corruption culture and corporate misconduct. *Journal of Financial Economics*, 122(2), 307–327.
- Lopez-Lira, A., & Tang, Y. (2023). Can ChatGPT forecast stock price movements? return predictability and large language models. *SSRN Electronic Journal*.
- Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1), 35–65.
- Loughran, T., & McDonald, B. (2016). Textual Analysis in Accounting and Finance: a survey. *Journal of Accounting Research*, 54(4), 1187–1230.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709.

- Mayew, W. J., & Venkatachalam, M. (2012). The power of voice: managerial affective states and future firm performance. *The Journal of Finance*, 67(1), 1–43.
- Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). *On faithfulness and factuality in abstractive summarization*. arXiv.org.
- Mazzola, P., Ravasi, D., & Gabbioneta, C. (2006). How to build reputation in financial markets. *Long Range Planning*, 39(4), 385–407.
- McGuinness, P. B., Vieito, J. P., & Wang, M. (2017). The role of board gender and foreign ownership in the CSR performance of Chinese listed firms. *Journal of Corporate Finance*, 42, 75–99.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space*. arXiv.org.
- Obloj, T., & Zemsky, P. (2015). Value creation and value capture under moral hazard: Exploring the micro-foundations of buyer– supplier relationships. *Strategic Management Journal*, 36(8), 1146–1163.
- OpenAI. (2025, April 29). Sycophancy in GPT-4o: what happened and what we’re doing about it.
- Paine, L. S. (1994). Managing for organizational integrity. *Harvard Business Review*, 72(2), 106–117.
- Palanski, M. E., & Yammarino, F. J. (2007). Integrity and leadership: *European Management Journal*, 25(3), 171–184.
- Palmrose, Z.-V., Richardson, V. J., & Scholz, S. (2004). Determinants of market reactions to restatement announcements. *Journal of Accounting and Economics*, 37(1), 59–89.
- Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock market index using fusion of machine learning techniques. *Expert Systems With Applications*, 42(4), 2162–2172.
- Petersen, M. A., & Rajan, R. G. (1997). Trade Credit: Theories and evidence. *Review of Financial Studies*, 10(3), 661–691.
- Pierce, J. L., Kostova, T., & Dirks, K. T. (2001). Toward a theory of psychological ownership in organizations. *Academy of Management Review*, 26(2), 298.
- Pierce, J. L., Kostova, T., & Dirks, K. T. (2003). The State of Psychological Ownership: Integrating and extending a century of research. *Review of General Psychology*, 7(1), 84–107.
- Price, S. M., Doran, J. S., Peterson, D. R., & Bliss, B. A. (2012). Earnings conference calls and stock returns: The incremental informativeness of textual tone. *Journal of Banking & Finance*, 36(4), 992–1011.
- Rest, J. R. (1986). *Moral development: Advances in research and theory*. Praeger.
- Salemi, A., Parmar, M., Goyal, P., Song, Y., Yoon, J., Zamani, H., Pfister, T., & Palangi, H. (2025). LLM-Based Multi-Agent Blackboard System for information discovery in data science. *ArXiv.org*.
- Schnackenberg, A. K., & Tomlinson, E. C. (2014). Organizational transparency. *Journal of Management*, 42(7), 1784–1810.

- Schwartz, M. S., Dunfee, T. W., & Kline, M. J. (2005). Tone at the top: An ethics code for directors? *Journal of Business Ethics*, 58(1–3), 79–100.
- Sharma, G., Chatterjee, S., Rani, V., Tatavarthy, A. D., & Bose, I. (2025). Driving Sustainability: unravelling the dynamics of user engagement on organization generated social media content on electric vehicles. *Information Systems Frontiers*, 27(6), 2431–2461.
- Shleifer, A., & Vishny, R. W. (1997). A survey of Corporate Governance. *The Journal of Finance*, 52(2), 737–783.
- Shojaee-Mend, H., Mohebbati, R., Amiri, M., & Atarodi, A. (2024). Evaluating the strengths and weaknesses of large language models in answering neurophysiology questions. *Scientific Reports*, 14(1), 10785.
- Simons, T. (2002). Behavioral integrity: the perceived alignment between managers' words and deeds as a research focus. *Organization Science*, 13(1), 18–35.
- Simons, T., Leroy, H., Collewaert, V., & Masschelein, S. (2015). How leader alignment of words and deeds Affects followers: A Meta-analysis of Behavioral Integrity Research. *Journal of Business Ethics*, 132(4), 831–844.
- Singh, J. P., Kumar, A., Rana, N. P., & Dwivedi, Y. K. (2022). Attention-Based LSTM network for Rumor Veracity Estimation of Tweets. *Information Systems Frontiers*, 24(2), 459–474.
- Sinha, R. K., & Damle, H. (2025). Corporate integrity culture and dividend policy. *Journal of Behavioral and Experimental Finance*, 46, 101028.
- Smith, J. K. (1987). Trade credit and informational asymmetry. *The Journal of Finance*, 42(4), 863–872.
- Solomon, R. C. (1992). Ethics and excellence: Cooperation and integrity in business. Oxford University Press.
- Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, 87(3), 355–374.
- Sprague, Z., Ye, X., Bostrom, K., Chaudhuri, S., & Durrett, G. (2023). *MuSR: Testing the Limits of Chain-of-thought with Multistep Soft Reasoning*. arXiv.org.
- Stansbury, J., & Barry, B. (2007). Ethics programs and the paradox of control. *Business Ethics Quarterly*, 17(2), 239–261.
- Storey, V. C., Yue, W. T., Zhao, J. L., & Lukyanenko, R. (2025). Generative Artificial intelligence: evolving technology, growing societal impact, and opportunities for information systems research. *Information Systems Frontiers*, 27(5), 2081–2102.
- Suchman, M. C. (1995). Managing Legitimacy: strategic and institutional approaches. *Academy of Management Review*, 20(3), 571–610.
- Sun, F., Dutta, S., Zhu, P., & Ren, W. (2021). Female insiders' ethics and trading profitability. *International Review of Financial Analysis*, 74, 101710.

- Suzgun, M., Scales, N., Schärli, N., Gehrmann, S., Tay, Y., Chung, H. W., Chowdhery, A., Le, Q., V., Chi, E. H., Zhou, D., & Wei, J. (2022). *Challenging BIG-Bench tasks and whether Chain-of-Thought can solve them*. arXiv.org.
- Tenbrunsel, A. E., Smith-Crowe, K., & Umphress, E. E. (2003). Building houses on rocks: The role of the ethical infrastructure in organizations. *Social Justice Research*, 16(3), 285–307.
- Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3), 1139–1168.
- Trevino, L. K. (1986). Ethical Decision Making in Organizations: A Person-Situation Interactionist model. *Academy of Management Review*, 11(3), 601–617.
- Treviño, L. K., Hartman, L. P., & Brown, M. (2000). Moral Person and Moral Manager: How Executives Develop a reputation for ethical leadership. *California Management Review*, 42(4), 128–142.
- Treviño, L. K., Weaver, G. R., & Reynolds, S. J. (2006). Behavioral Ethics in Organizations: A review. *Journal of Management*, 32(6), 951–990.
- Ullah, G. M. W., Khan, I., & Abdullah, M. (2025). Corporate culture and trade credit. *International Review of Financial Analysis*, 103, 104178.
- Victor, B., & Cullen, J. B. (1988). The organizational bases of ethical work climates. *Administrative Science Quarterly*, 33(1), 101–125.
- Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., Chowdhery, A., & Zhou, D. (2022). *Self-Consistency improves chain of thought reasoning in language models*. arXiv.org.
- Wang, Z., Mao, S., Wu, W., Ge, T., Wei, F., & Ji, H. (2024). *Unleashing the Emergent Cognitive Synergy in Large Language Models: A Task-Solving Agent through Multi-Persona Self-Collaboration* (pp. 257–279). NAACL.
- Weaver, G. R., & Treviño, L. K. (1999). Compliance and Values Oriented Ethics Programs: Influences on employees' attitudes and behavior. *Business Ethics Quarterly*, 9(2), 315–335.
- Webley, S., & Werner, A. (2008). Corporate codes of ethics: necessary but not sufficient. *Business Ethics a European Review*, 17(4), 405–415.
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2022). *Chain-of-Thought prompting elicits reasoning in large language models*. arXiv.org.
- Yang, Y., Qin, Y., Fan, Y., & Zhang, Z. (2023). Unlocking the Power of Voice for Financial Risk Prediction: A Theory-Driven Deep Learning Design Approach. *MIS Quarterly*, 47(1), 63–96.
- Zahra, S. A., Priem, R. L., & Rasheed, A. A. (2005). The antecedents and consequences of top management fraud. *Journal of Management*, 31(6), 803–828.

Figures

	Internal Orientation	External Orientation
Formal	Accountability, Governance, and Ownership <i>(Accountability)</i>	Regulatory Compliance and Adherence <i>(Compliance)</i>
Informal	Moral Integrity and Fairness <i>(Moral Integrity)</i>	Transparency, Trust, and Honesty <i>(Honesty)</i>

Figure 1. Four components of corporate integrity

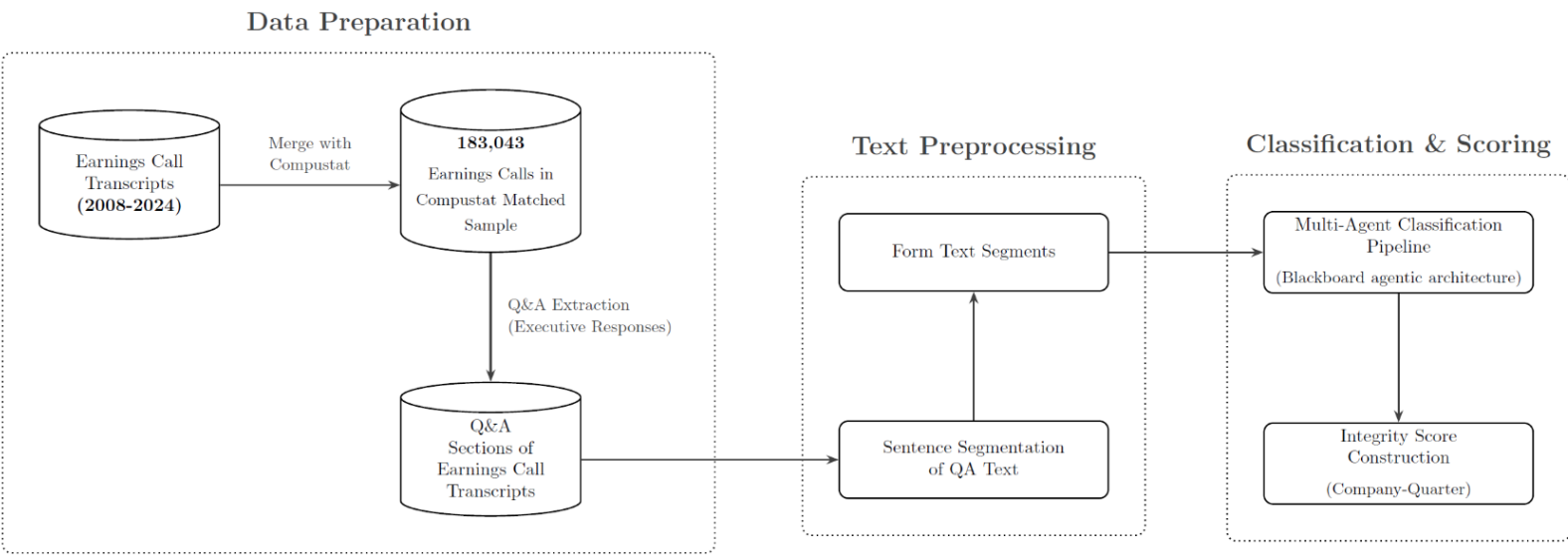


Figure 2. Process model

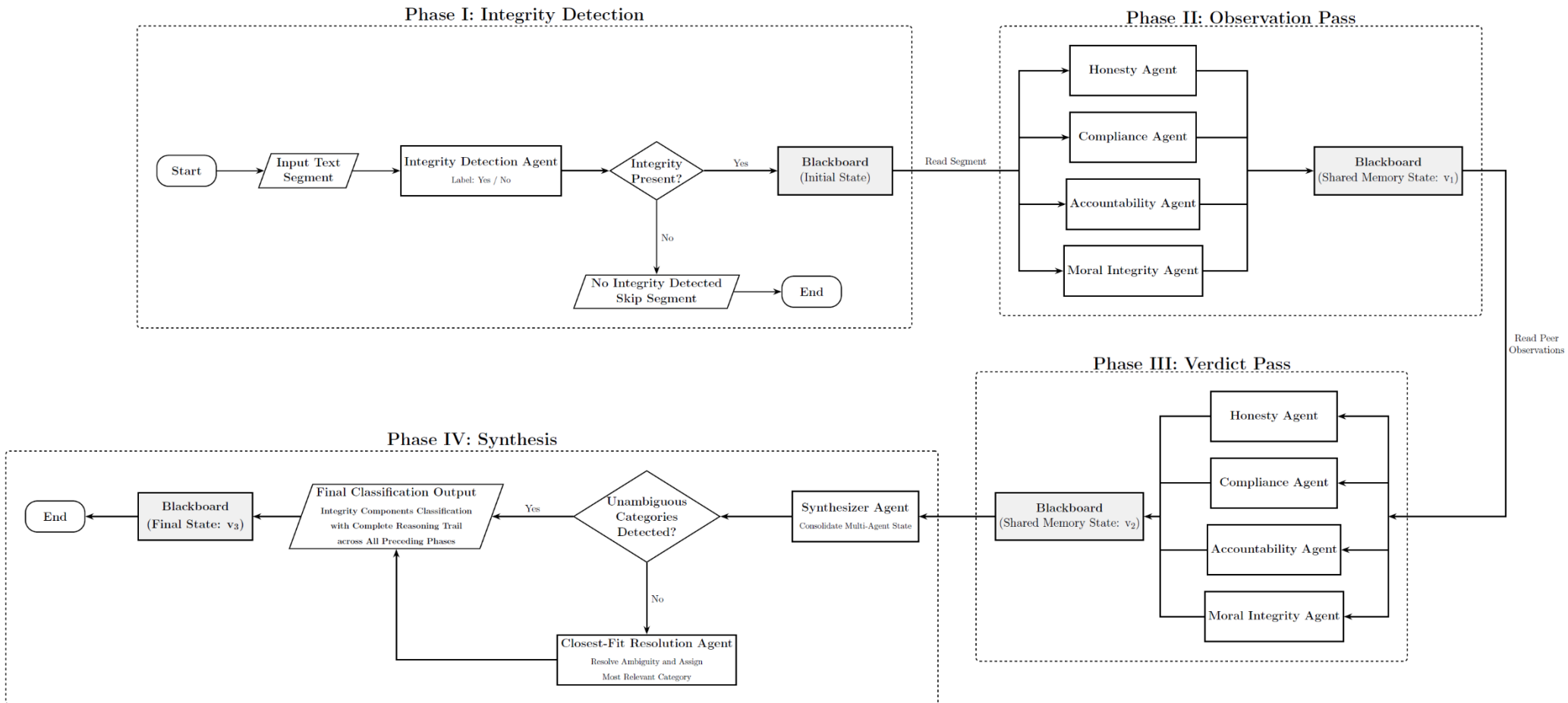


Figure 3. Multi agent integrity culture classification pipeline

Tables

Table 1: Classification performance of the Phase I integrity detection gate

	Accuracy	Precision (Macro)	Recall (Macro)	F1 (Macro)	F1 (Weighted)	Cohen's κ	MCC
Main Run	95.67%	95.67%	95.68%	95.67%	95.67%	0.913	0.914
Run 1	94.33%	94.36%	94.31%	94.33%	94.33%	0.887	0.887
Run 2	92.67%	92.82%	92.61%	92.65%	92.66%	0.853	0.854
Run 3	93.67%	93.79%	93.62%	93.65%	93.66%	0.873	0.874
Run 4	93.33%	93.43%	93.29%	93.32%	93.33%	0.867	0.867
Run 5	94.00%	94.10%	93.96%	93.99%	94.00%	0.880	0.881
Mean (Runs 1–5)	93.60%	93.70%	93.56%	93.59%	93.60%	0.872	0.873
SD (Runs 1–5)	–	–	–	0.6%	0.6%	0.012	–

Table 2a: Per-class metrics for the Phase I integrity detection gate

	Precision	Recall	F1-Score	Support
Main Run – No	96.67%	94.77%	95.71%	153
Main Run – Yes	94.67%	96.60%	95.62%	147
Runs 1–5 Avg – No	92.08%	95.69%	93.85%	153
Runs 1–5 Avg – Yes	95.32%	91.43%	93.33%	147

Table 2b: Confusion matrix (main run)

	Predicted: No	Predicted: Yes
True: No	145	8
True: Yes	5	142

Table 3: Inter-run agreement (Cohen's κ) – DeepSeek-R1

Subset Size (k)	N Combinations	Mean Pairwise κ	SD κ	Min κ	Max κ
2	10	0.945	0.013	0.920	0.967
3	10	0.945	0.008	0.933	0.960
4	5	0.945	0.005	0.941	0.951

Table 4: Examples of False Keyword Signals

Note: Bold and underlined terms denote words drawn from the corporate integrity dictionary developed by Li et al. (2021). The examples illustrate cases in which keyword presence does not correspond to meaningful integrity-related cultural content.

Sl. No.	Ticker	Publish Year	Segment Text	Model Output	Model Reason
1	SEM	2021	Do not know. I think it's a good question and they're putting in place, all the new heads of CMS and we just don't know at this point where their priorities are going to be. If you can recall AJ, back in the Obama administration, most of the efforts through CMS were really on the <u>Accountable</u> Care Organization and getting more people enrolled in that, really more on the insurance and coverage side.	No	The text discusses uncertainty about CMS priorities and historical focus on Accountable Care Organizations, but does not explicitly mention or imply any of the four integrity categories. There are no direct references to accountability, governance, compliance, moral integrity, transparency, or social responsibility.
2	STT	2017	So I think there's a real desire to fix it. The fix, the conversation around a fix is kind of two dimensional. One way to adjust the excess deposit issue for the <u>trust</u> banks would be simply to rescale the leverage ratio and that's one of the conversations out there so that you'd give the custody banks a little bit more headroom to accept these deposits. And at the same time, put in place some kind of a provision so that in a crisis or in a market event that we get relief from the leverage ratio. So that's one kind of conversation, which would tend to just deal with the <u>trust</u> bank issues of the leverage ratio. And then the broader thought out there is to reduce the denominator of the leverage ratio to accommodate central bank deposits, and that's, remember, that Bank of England took that approach in their leverage ratio.	No	The text discusses technical adjustments to the leverage ratio and deposit issues, focusing on regulatory and operational aspects without addressing accountability, governance, ethical decision-making, or transparency. There are no direct mentions of integrity-related concepts, making it difficult to infer a presence of integrity culture.
3	NDAQ	2017	And as the market makers and the other participants have been getting used to these fees, we've seen some moderation of volumes, but <u>honestly</u> , we continue to see new participants joining. We've had a couple of new firms joining as partners in the last couple of weeks and definitely, there still is very strong demand to continue to develop the platform and bring on new products and new activities. So we see this as maybe a little bit of a plateauing just based on them getting used to new fee-based, but we're pretty – we're still very encouraged and optimistic about the business.	No	The text segment does not contain any direct or indirect references to integrity-related categories. It discusses market participants, volume moderation, new partners, and business optimism but does not address accountability, governance, compliance, ethical decision-making, transparency, or social responsibility. The absence of key signals from the defined categories leads to the classification of 'No'.
4	ANSS	2017	We don't really break that out. We don't really break that out. But I think it's fair to say when you look at the -- when you look at a strategy, our strategy of course, is around multiphysics, and it's about being able to sell the whole portfolio. When you look at some of the initiatives that customers are driving, there is a lot of excitement, of course, about electronics and electronics within a product that previously didn't have smart controls, and that's something that we are actively involved with, and that's where some of our capabilities on antenna design, signal <u>integrity</u> and so forth come in. And then of course, when you start to think about chip package system, that's what brings together our semiconductor products and our electronics products as well as some of the mechanical and fluid products when you start to think about issues of thermal, et cetera.	No	The text segment does not contain any direct or indirect references to accountability, governance, ownership, regulatory compliance, moral integrity, fairness, social responsibility, transparency, trust, or honesty. The focus is on technical aspects of product strategy and capabilities without addressing integrity-related themes.
5	CELPQ	2017	It's a good question. I think it's a combination of things. Number one is, I think we're doing a much better job of evangelizing the story about our capabilities and broadening the number of customers, we offer our inspection and <u>integrity</u> services to and we've only scratched the surface there, Marshall.	No	The text segment discusses efforts to broaden the reach of inspection and integrity services but does not explicitly mention accountability, governance, compliance, ethical decision-making, fairness, transparency, or social responsibility. The term 'integrity' here refers to the services offered rather than the company's culture or practices related to integrity. There are no direct or indirect references to the defined categories, making it insufficient to classify as 'Yes'.

Table 5: Integrity Term Classification Rates

Note: The table reports, for each of the 40 most frequent terms in the corporate integrity dictionary developed by Li et al. (2021), the percentage of segments containing that term that were classified as “Yes”.

Li et al. (2021) Integrity Word	Total number of segments	#Yes #No segments	Yes/Total segment (%)	Li et al. (2021) Integrity Word	Total number of segments	#Yes #No segments	Yes/Total segment (%)
management	136233	20946 115287	15.38%	satisfaction	5120	1331 3789	26.00%
control	55964	7580 48384	13.54%	informed	4550	926 3624	20.35%
safety	32825	7537 25288	22.96%	risk_management	4215	1892 2323	44.89%
disciplined	22755	8503 14252	37.37%	mindset	4073	843 3230	20.70%
careful	19985	4518 15467	22.61%	responsive	4014	699 3315	17.41%
compliance	12296	9910 2386	80.60%	expert	3677	436 3241	11.86%
trust	12084	5811 6273	48.09%	principle	3522	683 2839	19.39%
thoughtful	11920	3316 8604	27.82%	integrity	3439	2703 736	78.60%
honestly	11685	4093 7592	35.03%	advice	3433	725 2708	21.12%
transparent	8092	7531 561	93.07%	authority	3284	641 2643	19.52%
committee	7311	2601 4710	35.58%	truth	3221	722 2499	22.42%
diligence	7311	2188 5123	29.93%	keen	2998	396 2602	13.21%
responsible	7274	4132 3142	56.81%	diligent	2937	1344 1593	45.76%
consistency	6889	955 5934	13.86%	thorough	2874	917 1957	31.91%
transparency	6541	5889 652	90.03%	rigorous	2837	1566 1271	55.20%
proper	6381	1434 4947	22.47%	identity	2811	518 2293	18.43%
candidly	6053	3479 2574	57.48%	credibility	2705	1255 1450	46.40%
responsibility	5737	3174 2563	55.33%	honest	2622	2287 335	87.22%
seriously	5519	2188 3331	39.64%	regulator	2607	1204 1403	46.18%
assure	5120	1469 3651	28.69%	society	2551	741 1810	29.05%

Table 6: Corporate governance and corporate integrity culture

This table presents the relation between (i) independent director ratio and integrity ratio and its subcomponents, and (ii) female director ratio and integrity ratio and its subcomponents. All variables are defined in **Table A1** presented in the Appendix. All models include industry and year fixed effects. White heteroscedastic-consistent robust standard errors in parentheses are clustered at the firm level (White, 1980). Asterisks (*, **, and ***) indicate the significance levels of 10 %, 5 %, and 1 %, respectively. A detailed companion table to **Table 6**, including control variables, is provided in Appendix B **Table B8**.

Variables	(1) Integrity ratio	(2) Integrity ratio	(3) Integrity ratio	(4) Accountability	(5) Compliance	(6) Moral	(7) Honesty	(8) Integrity Li tfidf lyr	(9) Integrity Li wtfidf lyr
Independent director ratio	0.0086*** (0.002)		0.0078*** (0.002)	0.0056*** (0.001)	0.0011 (0.001)	0.0015* (0.001)	0.0039*** (0.001)	0.0305* (0.016)	0.0168 (0.014)
Female director ratio		0.0108*** (0.003)	0.0093*** (0.003)	0.0057*** (0.002)	-0.0019* (0.001)	0.0048*** (0.001)	0.0053*** (0.002)	0.0333 (0.023)	0.0221 (0.019)
Firm size	0.0007*** (0.000)	0.0007*** (0.000)	0.0005** (0.000)	0.0001 (0.000)	0.0000 (0.000)	0.0008*** (0.000)	0.0001 (0.000)	0.0050*** (0.002)	0.0010 (0.001)
Leverage	-0.0006 (0.001)	-0.0006 (0.001)	-0.0006 (0.001)	0.0009 (0.001)	-0.0014** (0.001)	-0.0010 (0.001)	0.0006 (0.001)	-0.0405*** (0.012)	-0.0301*** (0.010)
Dividend to asset	0.0135*** (0.005)	0.0131*** (0.005)	0.0128*** (0.005)	0.0082** (0.003)	0.0035 (0.002)	0.0063*** (0.002)	0.0060** (0.003)	0.0582 (0.043)	0.0629* (0.033)
R&D to asset	-0.0422*** (0.004)	-0.0417*** (0.004)	-0.0422*** (0.004)	-0.0308*** (0.003)	-0.0060*** (0.002)	-0.0150*** (0.002)	-0.0206*** (0.002)	0.0489 (0.036)	0.0111 (0.028)
ROA	-0.0127*** (0.002)	-0.0125*** (0.002)	-0.0126*** (0.002)	-0.0076*** (0.001)	-0.0056*** (0.001)	-0.0057*** (0.001)	-0.0072*** (0.001)	-0.0928*** (0.019)	-0.0750*** (0.015)
Sales growth	-0.0017*** (0.000)	-0.0017*** (0.000)	-0.0016*** (0.000)	-0.0017*** (0.000)	0.0001 (0.000)	0.0001 (0.000)	-0.0008*** (0.000)	0.0072* (0.004)	0.0046 (0.003)
Standard deviation ROA	0.0537*** (0.012)	0.0535*** (0.012)	0.0539*** (0.012)	0.0392*** (0.008)	0.0254*** (0.006)	0.0271*** (0.006)	0.0303*** (0.007)	0.2917*** (0.096)	0.2690*** (0.078)
Constant	0.0350*** (0.002)	0.0401*** (0.002)	0.0354*** (0.002)	0.0267*** (0.002)	0.0085*** (0.001)	0.0071*** (0.001)	0.0216*** (0.001)	0.2265*** (0.021)	0.2417*** (0.017)
Year fixed effect	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry fixed effect	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	31,197	31,197	31,197	31,197	31,197	31,197	31,197	31,272	31,272
R-squared	0.028	0.027	0.029	0.028	0.065	0.048	0.023	0.045	0.042

Table 7: Corporate integrity and non-fraudulent restatement disclosure

This table presents the relation between integrity ratio and its subcomponents and non-fraudulent restatement filing. In the regression set-up, ‘restat-filing’ is a dummy variable; it is equal to ‘1’ if the firm files for a non-fraudulent restatement in the subsequent year, or ‘0’ otherwise. All variables are defined in **Table A1** presented in the Appendix. All models include industry and year fixed effects. White heteroscedastic-consistent robust standard errors in parentheses are clustered at the firm level (White, 1980). Asterisks (*, **, and ***) indicate the significance levels of 10 %, 5 %, and 1 %, respectively. A detailed companion table to **Table 7**, including control variables, is provided in Appendix B **Table B9**.

VARIABLES	(1) Restat filing	(2) Restat filing	(3) Restat filing	(4) Restat filing	(5) Restat filing	(6) Restat filing	(7) Restat filing
Integrity ratio	2.2172*** (0.727)						
Accountability		3.5249*** (0.953)					
Compliance			4.3115*** (1.476)				
Moral				1.4396 (1.420)			
Honesty					3.1403*** (1.130)		
Integrity Li tfidf 3yr						-0.1291 (0.109)	
Integrity Li wtfidf 3yr							-0.1406 (0.130)
Other controls	Y	Y	Y	Y	Y	Y	Y
Industry fixed effect	Y	Y	Y	Y	Y	Y	Y
Observations	28,430	28,430	28,430	28,430	28,430	32,329	32,329
Pseudo r-squared	0.0312	0.0315	0.0312	0.0307	0.0311	0.0292	0.0291

Table 8: Corporate integrity and earnings management

This table presents the relation between integrity ratio and its subcomponents and discretionary accruals (earnings management). All variables are defined in **Table A1** presented in the Appendix. All models include industry and year fixed effects. White heteroscedastic-consistent robust standard errors in parentheses are clustered at the firm level (White, 1980). Asterisks (*, **, and ***) indicate the significance levels of 10 %, 5 %, and 1 %, respectively. A detailed companion table to **Table 8**, including control variables, is provided in Appendix B **Table B10**.

VARIABLES	(1) Disc. Accrual	(2) Disc. Accrual	(3) Disc. Accrual	(4) Disc. Accrual	(5) Disc. Accrual	(6) Disc. Accrual	(7) Disc. Accrual
Integrity ratio	-0.0547** (0.027)						
Accountability		-0.0449 (0.035)					
Compliance			-0.0710 (0.062)				
Moral				-0.0405 (0.050)			
Honesty					-0.1156** (0.048)		
Integrity Li tfidf 3yr						0.0000 (0.003)	
Integrity Li wtfidf 3yr							0.0007 (0.004)
Other controls	Y	Y	Y	Y	Y	Y	Y
Year fixed effect	Y	Y	Y	Y	Y	Y	Y
Industry fixed effect	Y	Y	Y	Y	Y	Y	Y
Observations	26,675	26,675	26,675	26,675	26,675	30,506	30,506
R-squared	0.011	0.011	0.011	0.011	0.011	0.012	0.012

Table 9: Corporate integrity culture and dividend policy

This table presents the relation between integrity ratio and its subcomponents and dividend to total asset ratio. All variables are defined in **Table A1** presented in the Appendix. All models include industry and year fixed effects. White heteroscedastic-consistent robust standard errors in parentheses are clustered at the firm level (White, 1980). Asterisks (*, **, and ***) indicate the significance levels of 10 %, 5 %, and 1 %, respectively. A detailed companion table to **Table 9**, including control variables, is provided in Appendix B **Table B11**.

VARIABLES	(1) Div to asset	(2) Div to asset	(3) Div to asset	(4) Div to asset	(5) Div to asset	(6) Div to asset	(7) Div to asset
Integrity ratio	0.0588*** (0.020)						
Accountability		0.0618*** (0.022)					
Compliance			0.0988** (0.049)				
Moral				0.0939*** (0.036)			
Honesty					0.0828** (0.032)		
Integrity Li tfidf 3yr						0.0037 (0.003)	
Integrity Li wtfidf 3yr							0.0059* (0.003)
Other controls	Y	Y	Y	Y	Y	Y	Y
Year fixed effect	Y	Y	Y	Y	Y	Y	Y
Industry fixed effect	Y	Y	Y	Y	Y	Y	Y
Observations	31,197	31,197	31,197	31,197	31,197	35,577	35,577
R-squared	0.147	0.146	0.146	0.146	0.146	0.129	0.129

Table 10: Corporate integrity culture and trade credit

This table presents the relation between integrity ratio and its subcomponents and net credit. All variables are defined in **Table A1** presented in the Appendix. All models include industry and year fixed effects. White heteroscedastic-consistent robust standard errors in parentheses are clustered at the firm level (White, 1980). Asterisks (*, **, and ***) indicate the significance levels of 10 %, 5 %, and 1 %, respectively. A detailed companion table to **Table 10**, including control variables, is provided in Appendix B **Table B12**.

VARIABLES	(1) Net credit	(2) Net credit	(3) Net credit	(4) Net credit	(5) Net credit	(6) Net credit	(7) Net credit
Integrity ratio	0.8678*** (0.271)						
Accountability		0.7615** (0.299)					
Compliance			2.7355*** (0.863)				
Moral				0.9313** (0.396)			
Honesty					0.9967** (0.389)		
Integrity Li tfidf 3yr						0.0835** (0.042)	
Integrity Li wtfidf 3yr							0.0892* (0.053)
Other controls	Y	Y	Y	Y	Y	Y	Y
Year fixed effect	Y	Y	Y	Y	Y	Y	Y
Industry fixed effect	Y	Y	Y	Y	Y	Y	Y
Observations	30,885	30,885	30,885	30,885	30,885	35,208	35,208
R-squared	0.069	0.068	0.070	0.068	0.068	0.070	0.070

Appendix B: Additional Analysis Material

Table B1: Paired-sample t-tests of BGE-M3 vs. TF-IDF segmentation

This table presents paired-sample t-tests comparing BGE-M3 and TF-IDF segmentation on integrity-relevant outcomes across three non-overlapping subsamples of 1,000 earnings-call transcripts each. S1, S2, S3 denote the three non-overlapping subsamples, each drawn at random from a pool of 183,043 transcripts and each containing 1,000 transcripts. All tests are two-tailed paired-sample t -tests with the null hypothesis that the population mean difference between BGE-M3 and TF-IDF equals zero. Mean Diff is computed as BGE-M3 minus TF-IDF; a positive value therefore indicates a higher value under BGE-M3. The 95% confidence interval is for the mean difference. Degrees of freedom equal $N - 1$. Sample sizes below 1,000 for the ratio rows reflect pairwise exclusion of transcripts for which the ratio is undefined (Yes / No is undefined when No = 0; Yes / Total is undefined when Total = 0). Cohen’s d is computed as the mean paired difference divided by the standard deviation of the paired differences.

Sample	N	BGE-M3 Mean	TF-IDF Mean	Mean Diff	95% CI	t-stat (df)	p-value	Cohen’s d
Panel A: Yes segments per transcript (count of integrity-relevant segments)								
S1	1000	2.0430	2.0340	+0.0090	[-0.0521, 0.0701]	0.289 (999)	0.7726	0.0091
S2	1000	1.9930	1.9650	+0.0280	[-0.0294, 0.0854]	0.957 (999)	0.3388	0.0303
S3	1000	1.9850	1.9660	+0.0190	[-0.0419, 0.0799]	0.612 (999)	0.5406	0.0194
Panel B: Yes / Total ratio (proportion of Yes segments among all segments)								
S1	987	0.0419	0.0422	-0.0003	[-0.0017, 0.0011]	-0.382 (986)	0.7028	-0.0121
S2	991	0.0423	0.0425	-0.0002	[-0.0016, 0.0013]	-0.231 (990)	0.8176	-0.0073
S3	988	0.0421	0.0429	-0.0007	[-0.0026, 0.0011]	-0.815 (987)	0.4154	-0.0259
Panel C: Yes / No ratio (odds of Yes classification against No among classified segments)								
S1	987	0.0467	0.0467	-0.0000	[-0.0020, 0.0019]	-0.035 (986)	0.9721	-0.0011
S2	991	0.0469	0.0468	+0.0001	[-0.0017, 0.0019]	0.089 (990)	0.9291	0.0028
S3	987	0.0459	0.0468	-0.0009	[-0.0038, 0.0019]	-0.646 (986)	0.5181	-0.0206

Table B2: Benchmark performance of candidate models

In this table, BBH refers to BigBench Hard (Suzgun et al. 2022), and MuSR denotes Multistep Soft Reasoning (Sprague et al. 2023). Boldface values indicate the best performance within each benchmark. Notably, DeepSeek-R1-Distill-Qwen-14B attains the highest score on MuSR, which is the benchmark most closely aligned with our long-context, multistep classification task.

Model	BBH	MuSR
DeepSeek-R1-Distill-Qwen-14B	40.69%	28.71%
DeepSeek-R1-Distill-Qwen-32B	17.15%	16.14%
Qwen/QwQ-32B	2.87%	11.05%
Qwen2.5-14B-Instruct	48.36%	10.16%
Qwen2.5-32B-Instruct	56.49%	13.50%
Microsoft Phi-4	52.58%	23.72%
Mistral-Small-24B-Base-2501	48.54%	10.89%

Table B3: Confusion matrices

This table reports the confusion matrices for all runs of the Phase I integrity detection gate.

Model	Run	True Negative (TN)	False Positive (FP)	False Negative (FN)	True Positive (TP)
Original Model	Main Run	145	8	5	142
DeepSeek-R1-14B	1	146	7	10	137
	2	146	7	15	132
	3	147	6	13	134
	4	146	7	13	134
	5	147	6	12	135

Table B4: Pairwise Inter-Run Agreement

This table presents pairwise inter-run agreement statistics (Cohen’s kappa) for the Phase I integrity detection gate.

Subset Size (k)	Runs Included	Average Pairwise κ	Interpretation
2	R1, R2	0.9398	Almost Perfect
	R1, R3	0.9599	Almost Perfect
	R1, R4	0.9399	Almost Perfect
	R1, R5	0.9666	Almost Perfect
	R2, R3	0.9531	Almost Perfect
	R2, R4	0.9196	Almost Perfect
	R2, R5	0.9330	Almost Perfect
	R3, R4	0.9398	Almost Perfect
	R3, R5	0.9531	Almost Perfect
	R4, R5	0.9465	Almost Perfect
3	R1, R2, R3	0.9509	Almost Perfect
	R1, R2, R4	0.9331	Almost Perfect
	R1, R2, R5	0.9465	Almost Perfect
	R1, R3, R4	0.9465	Almost Perfect
	R1, R3, R5	0.9599	Almost Perfect
	R1, R4, R5	0.9510	Almost Perfect
	R2, R3, R4	0.9375	Almost Perfect
	R2, R3, R5	0.9464	Almost Perfect
	R2, R4, R5	0.9331	Almost Perfect
	R3, R4, R5	0.9465	Almost Perfect
4	R1, R2, R3, R4	0.9420	Almost Perfect
	R1, R2, R3, R5	0.9509	Almost Perfect
	R1, R2, R4, R5	0.9409	Almost Perfect
	R1, R3, R4, R5	0.9510	Almost Perfect
	R2, R3, R4, R5	0.9409	Almost Perfect

Table B5: Yes Segment % for Integrity Wordlist

This table reports, for each integrity-related term in the Li et al. (2021) integrity culture dictionary, the percentage of segments containing that term that were classified as “Yes”.

Li et al. (2021) Integrity Word	Number of Yes	Number of No Segments	Number of Total	Yes segment
management	20946	115287	136233	15.38%
control	7580	48384	55964	13.54%
safety	7537	25288	32825	22.96%
disciplined	8503	14252	22755	37.37%
careful	4518	15467	19985	22.61%
compliance	9910	2386	12296	80.60%
trust	5811	6273	12084	48.09%
thoughtful	3316	8604	11920	27.82%
honestly	4093	7592	11685	35.03%
transparent	7531	561	8092	93.07%
committee	2601	4710	7311	35.58%
diligence	2188	5123	7311	29.93%
responsible	4132	3142	7274	56.81%
consistency	955	5934	6889	13.86%
transparency	5889	652	6541	90.03%
proper	1434	4947	6381	22.47%
candidly	3479	2574	6053	57.48%
responsibility	3174	2563	5737	55.33%
seriously	2188	3331	5519	39.64%
assure	1469	3651	5120	28.69%
satisfaction	1331	3789	5120	26.00%
informed	926	3624	4550	20.35%
risk_management	1892	2323	4215	44.89%
mindset	843	3230	4073	20.70%
responsive	699	3315	4014	17.41%
expert	436	3241	3677	11.86%
principle	683	2839	3522	19.39%
integrity	2703	736	3439	78.60%
advice	725	2708	3433	21.12%
authority	641	2643	3284	19.52%
truth	722	2499	3221	22.42%
keen	396	2602	2998	13.21%
diligent	1344	1593	2937	45.76%
thorough	917	1957	2874	31.91%
rigorous	1566	1271	2837	55.20%
identity	518	2293	2811	18.43%
credibility	1255	1450	2705	46.40%
honest	2287	335	2622	87.22%
regulator	1204	1403	2607	46.18%
society	741	1810	2551	29.05%

strict	1026	1430	2456	41.78%
governance	2113	319	2432	86.88%
assurance	515	1804	2319	22.21%
continuity	403	1900	2303	17.50%
spirit	455	1779	2234	20.37%
comply	1674	496	2170	77.14%
convince	320	1802	2122	15.08%
investment_community	457	1638	2095	21.81%
conflict	378	1593	1971	19.18%
freedom	216	1717	1933	11.17%
accountable	1612	186	1798	89.66%
confidentiality	1260	521	1781	70.75%
counsel	614	1167	1781	34.48%
privacy	1109	523	1632	67.95%
motivation	288	1337	1625	17.72%
rigor	798	788	1586	50.32%
council	405	1172	1577	25.68%
deserve	503	1051	1554	32.37%
oversight	801	723	1524	52.56%
accountability	1470	44	1514	97.09%
credible	397	1084	1481	26.81%
honor	720	708	1428	50.42%
faith	353	1041	1394	25.32%
stand_up	190	1078	1268	14.98%
candid	1062	203	1265	83.95%
independence	213	1038	1251	17.03%
paramount	295	941	1236	23.87%
frustration	313	890	1203	26.02%
pragmatic	188	1005	1193	15.76%
advocacy	240	911	1151	20.85%
adherence	427	662	1089	39.21%
mantra	223	813	1036	21.53%
advocate	295	720	1015	29.06%
responsibly	862	151	1013	85.09%
harm	355	628	983	36.11%
decentralized	134	761	895	14.97%
fiduciary	457	432	889	51.41%
autonomy	94	747	841	11.18%
police	150	625	775	19.35%
convincing	76	693	769	9.88%
adhere	429	333	762	56.30%
enforce	310	450	760	40.79%
applaud	180	579	759	23.72%
prudence	245	492	737	33.24%
wisdom	78	568	646	12.07%

sincerely	144	491	635	22.68%
stakeholder	272	340	612	44.44%
supervision	158	444	602	26.25%
prescriptive	94	505	599	15.69%
level_playing_field	210	380	590	35.59%
finance_team	153	436	589	25.98%
genuinely	120	442	562	21.35%
citizen	159	393	552	28.80%
steadfast	191	331	522	36.59%
attentive	109	410	519	21.00%
sincere	176	342	518	33.98%
crime	146	327	473	30.87%
democratic	54	412	466	11.59%
utmost	173	293	466	37.12%
core_value	142	309	451	31.49%
insist	87	334	421	20.67%
steward	151	269	420	35.95%
stewardship	256	148	404	63.37%
safeguard	127	274	401	31.67%
humble	127	272	399	31.83%
unequivocally	69	324	393	17.56%
reassure	136	255	391	34.78%
courage	153	236	389	39.33%
bureaucracy	63	304	367	17.17%
reputable	96	268	364	26.37%
criticism	118	243	361	32.69%
ethical	302	57	359	84.12%
patient_safety	204	148	352	57.95%
corporate_governance	335	13	348	96.26%
supervisory	131	216	347	37.75%
fairness	270	72	342	78.95%
ingrained	84	218	302	27.81%
audit_committee	266	31	297	89.56%
abide	198	97	295	67.12%
employee_engagement	142	153	295	48.14%
resolute	59	232	291	20.27%
executive_committee	102	188	290	35.17%
adhering	167	107	274	60.95%
quality_assurance	124	146	270	45.93%
bureaucratic	51	206	257	19.84%
irresponsible	184	60	244	75.41%
acutely_aware	66	175	241	27.39%
management_approach	53	185	238	22.27%
religious	44	192	236	18.64%
constituent	31	196	227	13.66%

wholeheartedly	48	174	222	21.62%
suitability	59	163	222	26.58%
journalism	47	175	222	21.17%
forthright	186	34	220	84.55%
exemplary	61	158	219	27.85%
custodian	31	186	217	14.29%
seriousness	74	142	216	34.26%
harmonized	37	172	209	17.70%
quality_care	99	107	206	48.06%
visionary	26	177	203	12.81%
candor	130	69	199	65.33%
business_practice	56	139	195	28.72%
onus	63	131	194	32.47%
thoughtfulness	70	119	189	37.04%
unbiased	80	104	184	43.48%
arrogant	31	153	184	16.85%
moral	120	61	181	66.30%
conscientious	118	62	180	65.56%
fiduciary_responsibility	176	3	179	98.32%
risk_control	100	75	175	57.14%
instill	83	92	175	47.43%
information_flow	41	132	173	23.70%
unanimously	64	107	171	37.43%
franchisor	27	135	162	16.67%
traceability	81	80	161	50.31%
corruption	130	31	161	80.75%
vested_interest	47	114	161	29.19%
praise	36	124	160	22.50%
honesty	143	17	160	89.38%
passionately	55	101	156	35.26%
trustworthy	137	15	152	90.13%
money_laundering	113	35	148	76.35%
sympathetic	28	115	143	19.58%
management_philosophy	41	92	133	30.83%
unequivocal	17	112	129	13.18%
management_committee	54	75	129	41.86%
humility	73	55	128	57.03%
compliance_program	117	11	128	91.41%
meticulous	51	72	123	41.46%
communicative	65	54	119	54.62%
finance_department	23	94	117	19.66%
sovereignty	43	72	115	37.39%
empathy	58	53	111	52.25%
entrust	29	81	110	26.36%
performance_culture	78	31	109	71.56%

instruct	10	98	108	9.26%
delegate	21	84	105	20.00%
unwavering_commitment	82	22	104	78.85%
politician	12	91	103	11.65%
governance_structure	90	9	99	90.91%
consumer_protection	70	29	99	70.71%
legitimacy	26	72	98	26.53%
impeccable	34	64	98	34.69%
company_management	28	68	96	29.17%
democracy	26	70	96	27.08%
mission_statement	34	60	94	36.17%
institutionalize	13	81	94	13.83%
equality	43	50	93	46.24%
disrespect	14	79	93	15.05%
objectivity	42	46	88	47.73%
company_culture	60	27	87	68.97%
legislator	26	61	87	29.89%
shareholder_interest	54	31	85	63.53%
fiduciary_duty	82	2	84	97.62%
ego	19	64	83	22.89%
ethically	69	12	81	85.19%
empathetic	33	47	80	41.25%
partisan	12	65	77	15.58%
credo	19	57	76	25.00%
preach	24	52	76	31.58%
governance_process	71	4	75	94.67%
conscience	51	24	75	68.00%
conformity	29	45	74	39.19%
management_board	38	36	74	51.35%
subcommittee	39	35	74	52.70%
paramount_importance	22	52	74	29.73%
management_style	23	51	74	31.08%
embarrassment	15	58	73	20.55%
executive_board	27	45	72	37.50%
ethic	39	32	71	54.93%
humbly	20	50	70	28.57%
honorable	51	17	68	75.00%
management_discipline	24	44	68	35.29%
faithful	28	40	68	41.18%
risk_committee	54	13	67	80.60%
safety_culture	59	8	67	88.06%
secrecy	37	28	65	56.92%
supervise	18	47	65	27.69%
risk_manager	21	43	64	32.81%
core_principle	26	37	63	41.27%

decentralize	13	49	62	20.97%
compliance_team	60	1	61	98.36%
authoritative	15	45	60	25.00%
principled	47	9	56	83.93%
judiciary	16	39	55	29.09%
obsessive	14	39	53	26.42%
quality_organization	21	31	52	40.38%
safety_management	33	19	52	63.46%
quality_management_system	31	18	49	63.27%
meritocracy	35	14	49	71.43%
proprietor	3	46	49	6.12%
philanthropy	26	23	49	53.06%
sincerity	28	19	47	59.57%
governance_perspective	45	2	47	95.74%
compassion	21	26	47	44.68%
fiduciary_obligation	43	2	45	95.56%
corrupt	30	15	45	66.67%
socially_responsible	44	1	45	97.78%
value_system	25	19	44	56.82%
intellectually_honest	44	0	44	100.00%
zeal	14	29	43	32.56%
unethical	37	6	43	86.05%
thoroughness	19	24	43	44.19%
reputational_risk	33	9	42	78.57%
solidarity	14	25	39	35.90%
governance_standpoint	36	2	38	94.74%
governance_model	34	4	38	89.47%
compliance_department	38	0	38	100.00%
procurement_department	9	28	37	24.32%
harassment	31	5	36	86.11%
let_you_down	26	10	36	72.22%
insistence	12	24	36	33.33%
frugality	4	29	33	12.12%
impartial	18	14	32	56.25%
esteem	5	27	32	15.63%
vouch	8	22	30	26.67%
shareholder_community	12	18	30	40.00%
governmental_agency	5	24	29	17.24%
nominating_committee	18	11	29	62.07%
trustworthiness	25	2	27	92.59%
caretaker	7	20	27	25.93%
risk_management_group	18	9	27	66.67%
misconduct	24	2	26	92.31%
journalistic	4	22	26	15.38%
disinterested	8	17	25	32.00%

anonymity	11	14	25	44.00%
resourcefulness	7	18	25	28.00%
citizenship	15	9	24	62.50%
scrupulous	19	3	22	86.36%
audit_function	18	4	22	81.82%
oversight_committee	14	6	20	70.00%
safety_environment	10	10	20	50.00%
courteous	4	16	20	20.00%
decision_rights	7	12	19	36.84%
risk_oversight	13	6	19	68.42%
compliance_function	18	1	19	94.74%
controllership	9	9	18	50.00%
inculcate	8	10	18	44.44%
governance_framework	17	0	17	100.00%
litigator	6	11	17	35.29%
nonpartisan	2	15	17	11.76%
act_responsibly	15	2	17	88.24%
renown	0	16	16	0.00%
board_director	8	8	16	50.00%
governance_issue	14	1	15	93.33%
individualistic	0	15	15	0.00%
constitutionally	5	10	15	33.33%
risk_management_function	10	4	14	71.43%
policy maker	2	12	14	14.29%
moral_obligation	13	0	13	100.00%
dishonest	12	1	13	92.31%
authorship	3	9	12	25.00%
decency	6	6	12	50.00%
bureaucrat	1	10	11	9.09%
business_conduct	7	4	11	63.64%
ombudsman	6	5	11	54.55%
oversight_responsibility	7	3	10	70.00%
correctness	3	7	10	30.00%
condone	7	2	9	77.78%
egalitarian	1	8	9	11.11%
compliance_aspect	7	2	9	77.78%
overseer	1	8	9	11.11%
business_principle	6	3	9	66.67%
remuneration_committee	4	5	9	44.44%
board_oversight	7	2	9	77.78%
customer_champion	7	1	8	87.50%
hold_accountable	7	1	8	87.50%
policy_maker	2	4	6	33.33%
demonize	0	6	6	0.00%
risk_organization	3	3	6	50.00%

suasion	2	3	5	40.00%
probity	4	1	5	80.00%
rigor_discipline	2	3	5	40.00%
governance_system	5	0	5	100.00%
illegality	3	1	4	75.00%
governance_practice	4	0	4	100.00%
community_member	0	4	4	0.00%
dictatorship	0	4	4	0.00%
safety_soundness	4	0	4	100.00%
derelict	3	1	4	75.00%
reproach	0	3	3	0.00%
communication_style	2	1	3	66.67%
unselfish	1	2	3	33.33%
paternalistic	0	3	3	0.00%
government_regulator	2	1	3	66.67%
instil	1	2	3	33.33%
stakeholder_management	1	2	3	33.33%
advice_counsel	1	1	2	50.00%
reprehensible	0	2	2	0.00%
brand_champion	0	2	2	0.00%
nondiscrimination	2	0	2	100.00%
core_responsibility	0	2	2	0.00%
ethic_integrity	2	0	2	100.00%
governance_standard	1	0	1	100.00%
business_ethic	1	0	1	100.00%
operation_council	0	1	1	0.00%
board_function	1	0	1	100.00%

Table B6: Distribution of Integrity Components

Component	Number of Segments	Presence in Integrity Segments (%)
Accountability, Governance, and Ownership	260,246	64.57%
Regulatory Compliance and Adherence	116,796	28.98%
Moral Integrity and Fairness	145,641	36.13%
Transparency, Trust, and Honesty	212,820	52.80%

Table B7: Examples of Integrity Components Identified in the Earnings Call Discourse

Integrity Components Present	Segment Text	Synthesizer Agent Reasoning	Full Agentic Reasoning Trail
Accountability	This is something that is recurring in our discussions with our Board. We are always looking at ways of maximizing the return in capital and believe me, if we get to a conclusion that we don't have that capacity, the Board will act immediately and we will do something as far as returning capital is concerned. But this is not in our horizon anytime soon because what we need to keep in perspective is that we are coming from maybe one of the most negative credit cycles for Latin America with the drop in commodity prices that we experienced after 2015, both oil prices and commodities overall, soft commodities and metals that impacted tremendously the region. We had, as you rightly commented, to adjust our activities to the risk-reward in the most efficient way.	The Synthesizer Agent consolidates the verdicts from all four agents. The Compliance Agent and the Honesty Agent both voted NO, as there was no explicit mention of adherence to regulations or transparent communication, respectively. The Moral Integrity Agent also voted NO, noting the absence of ethical decision-making or CSR initiatives. However, the Accountability Agent voted YES, highlighting the presence of governance structures, the Board's immediate action on capital returns, and the responsibility towards stakeholders. The text emphasizes the Board's role in strategic decisions and adjustments, reflecting a culture of accountability and ownership. While other categories were not strongly supported, the evidence for accountability is compelling, making it the sole category present.	[Accountability Agent] Verdict: YES Observation: The text mentions recurring discussions with the Board, indicating the presence of governance structures. The Board is referenced as making decisions regarding capital returns, showing accountability. The firm acknowledges the need to adjust activities based on risk-reward, reflecting a sense of responsibility towards stakeholders. The Board's immediate action upon reaching a conclusion demonstrates a governance framework in place. The discussion about returning capital highlights the firm's ownership of outcomes and accountability to shareholders. Reasoning: The text discusses recurring Board meetings focused on maximizing return on capital, indicating a governance structure. The Board's immediate action upon capacity assessment shows accountability. Adjustments to activities based on risk-reward reflect responsibility towards stakeholders, aligning with the Accountability, Governance, & Ownership culture. [Compliance Agent] Verdict: NO Observation: The text discusses recurring discussions with the Board about maximizing return on capital and the potential for immediate action if capacity is lacking. It mentions adjustments to activities based on risk-reward, referencing the impact of commodity price drops post-2015. However, there are no explicit mentions of adherence to external laws, government regulations, or internal contractual obligations. The focus is on financial performance and strategic adjustments rather than compliance-related activities. Reasoning: The text discusses strategic adjustments and financial performance but does not mention adherence to external regulations or contractual obligations. There are no references to specific laws, compliance programs, or regulatory frameworks. The focus is on maximizing returns and operational efficiency, not on compliance culture. [Honesty Agent] Verdict: NO Observation: The text discusses recurring discussions with the Board about maximizing return on capital, indicating a focus on financial performance. It mentions that if the company lacks capacity, the Board will act immediately to return capital, suggesting a commitment to addressing financial obligations. The segment references the impact of the 2015 commodity price drop on Latin America, highlighting the need to adjust activities to risk-reward efficiently. However, there is no explicit mention of transparent communication, protection of sensitive information, or mechanisms to build trust with stakeholders. The emphasis is on financial strategies and adjustments due to external economic factors rather

			than on fostering a culture of honesty or trust. Reasoning: The text discusses financial performance and strategic adjustments but lacks explicit mentions of transparent communication, protection of sensitive information, or mechanisms to build trust with stakeholders. While there is a commitment to addressing financial obligations, there is no clear evidence of a culture emphasizing honesty or trust. [Moral Integrity Agent] Verdict: NO Observation: The text discusses the Board's focus on maximizing return on capital and acting immediately if capacity is lacking. It mentions adjustments to activities based on risk-reward efficiency, particularly in the context of Latin America's economic challenges post-2015. There is no explicit mention of ethical decision-making, equitable treatment of stakeholders, or contributions to social good beyond economic adjustments. The emphasis is on financial returns and operational adjustments rather than moral considerations or CSR initiatives. The recurring discussions with the Board indicate a focus on profitability and efficiency without clear evidence of integrating ethical or social responsibility aspects into decision-making processes. Reasoning: The text segment discusses the Board's focus on maximizing return on capital and making adjustments based on risk-reward efficiency, particularly in the context of Latin America's economic challenges post-2015. However, there is no explicit mention of ethical decision-making, equitable treatment of stakeholders, or contributions to social good beyond economic adjustments. The emphasis is on financial returns and operational adjustments rather than moral considerations or CSR initiatives. The recurring discussions with the Board indicate a focus on profitability and efficiency without clear evidence of integrating ethical or social responsibility aspects into decision-making processes.
Moral Integrity	I guess, the power train decision ongoing from a combustion engine to EVs or hybrids is not really controlling their -- the level of autonomously in the vehicles. So I think their focus and their vision is to have no one killed in a car from 2020 and onwards. And that's their target that we're looking to support in terms of a safer car through a better ADAS system and through a high level of autonomously in Zenuity.	The text segment discusses a company's focus on safety and preventing fatalities through advanced driver-assistance systems (ADAS) and autonomous driving technology. The Compliance Agent noted a lack of explicit regulatory compliance, the Accountability Agent pointed to no clear ownership or governance structures, and the Honesty Agent found no evidence of transparency or trust. However, the Moral_Integrity Agent identified a strong commitment to ethical decision-making and societal well-being, aligning with CSR principles. The emphasis on safety and preventing harm reflects a clear moral stance, making Moral_Integrity the only category present.	[Accountability Agent] Verdict: NO Observation: The text discusses a focus on safety, specifically aiming to prevent deaths in cars starting from 2020. It mentions the development of a safer car through a better ADAS system and a high level of autonomy in Zenuity. However, there are no explicit references to accountability, governance, or ownership culture. The emphasis is on technological advancements and safety outcomes without addressing responsibility, structural frameworks, or systems ensuring accountability to stakeholders. Reasoning: The text emphasizes safety and technological advancements but does not explicitly mention accountability, governance, or ownership. The Accountability Agent's observations highlight the absence of references to responsibility, structural frameworks, or systems ensuring accountability to stakeholders. While the focus on safety aligns with ethical considerations, it does not directly address the core elements of the Accountability, Governance, & Ownership culture as defined. [Compliance Agent] Verdict: NO Observation: The text discusses a focus on safety, specifically aiming to prevent deaths in cars starting from 2020. It mentions supporting this goal through improved ADAS systems and high levels of autonomous driving in Zenuity. However, there are no explicit references to regulatory compliance, adherence to external laws, government regulations, or internal

			contractual obligations. The emphasis is on safety and technological advancement rather than compliance with specific legal or contractual requirements. Reasoning: The text focuses on safety and technological advancements but does not explicitly mention regulatory compliance, adherence to laws, or contractual obligations. There are no references to specific regulations or compliance programs, which are key indicators for Regulatory Compliance & Adherence culture. [Honesty Agent] Verdict: NO Observation: The text discusses the company's focus on improving ADAS systems and achieving high levels of autonomous driving to enhance safety and prevent fatalities. It mentions a target to ensure no one is killed in a car starting from 2020. However, there are no explicit references to transparency, trust, or honesty culture as defined. The content focuses on technological advancements and safety goals without addressing open communication, truthful disclosure, or the protection of sensitive information. Reasoning: The text segment discusses the company's focus on safety and technological advancements, particularly in ADAS systems and autonomous driving, to prevent fatalities. However, there are no explicit references to transparency, trust, or honesty. While the content emphasizes safety, it does not address open communication, truthful disclosure, or the protection of sensitive information, which are key aspects of the Transparency, Trust, & Honesty culture. The observations from other agents also highlight the absence of these elements, further supporting the conclusion that the culture is not clearly present in this segment. [Moral Integrity Agent] Verdict: YES Observation: The text discusses a focus on safety, aiming to prevent deaths in cars from 2020 onwards. It mentions supporting this goal through advanced driver-assistance systems (ADAS) and high levels of autonomy in Zenuity. The emphasis on preventing harm aligns with ethical decision-making. The commitment to safety reflects a consideration of societal well-being, which is part of CSR. The mention of autonomous driving systems suggests an integration of technology for public benefit, indicating a focus on social responsibility. Reasoning: The text emphasizes a clear commitment to ethical decision-making by focusing on safety and preventing harm, which aligns with moral integrity. The mention of ADAS systems and autonomous driving for societal benefit reflects a CSR approach. The target to prevent fatalities demonstrates a fairness and equity consideration towards all stakeholders, indicating a strong sense of social responsibility.
Accountability, Honesty	We frankly dropped the ball of the process and there is some kind of mitigating circumstances where we had terminated a couple of quality employees out at On-X at the facility there, because they weren't up to our standard. In the process that all happening, it kind of fell through the cracks. And what was surprising to us –	The text segment indicates a strong presence of Accountability and Honesty. The acknowledgment of failure and the governance structure, as highlighted by the Accountability Agent, demonstrate a clear sense of responsibility and ownership. The honest admission of dropping the ball and the candor about employee	[Accountability Agent] Verdict: YES Observation: The text mentions that the company 'dropped the ball of the process,' indicating an acknowledgment of failure. It references the termination of employees due to not meeting standards, which relates to accountability for upholding quality. The text notes the involvement of a Senior Vice President with 35 years of experience, suggesting a governance structure with experienced oversight. The phrase 'fell through the cracks' implies a lapse in systems or processes, highlighting a potential governance oversight. The mention of 'mitigating circumstances' suggests an attempt to explain the failure, which could relate to

	and again, our Head of Senior Vice President of our operations, quality and regulatory is a 35-year experienced executive and is very well-versed in this.	terminations, as noted by the Honesty Agent, underscore a commitment to transparency. While the Compliance Agent and Moral Integrity Agent found insufficient evidence, the explicit statements about process lapses and internal standards support the presence of Accountability and Honesty. The overall narrative reflects an acceptance of failure and a straightforward communication style, which aligns with these two categories.	ownership of the issue. Reasoning: The text explicitly acknowledges failure by stating ‘We frankly dropped the ball of the process,’ which demonstrates accountability. The mention of a Senior Vice President with 35 years of experience indicates a governance structure in place. The discussion of mitigating circumstances and the lapse in process handling (‘fell through the cracks’) suggests ownership of the issue and a governance oversight, aligning with the Accountability, Governance, & Ownership culture. [Compliance Agent] Verdict: NO Observation: The text mentions the termination of quality employees at the On-X facility due to not meeting standards. It references the Head of Senior Vice President of operations, quality, and regulatory, who has 35 years of experience and is well-versed in this area. The phrase ‘fell through the cracks’ suggests a lapse in process handling, potentially indicating a compliance oversight. The mention of ‘regulatory’ in the VP’s title indicates a focus on regulatory matters. The text does not explicitly state adherence to external laws or contractual obligations beyond the internal standards mentioned. Reasoning: The text mentions the termination of employees due to not meeting standards and references a Senior Vice President with regulatory expertise. However, there is no explicit mention of adherence to external laws, government regulations, or contractual obligations. The focus is on internal standards and quality, which aligns more with accountability and ethical practices rather than formal regulatory compliance. The phrase ‘fell through the cracks’ suggests a process lapse but does not directly indicate compliance adherence. Therefore, the evidence is insufficient to confirm the presence of a Regulatory Compliance & Adherence culture. [Honesty Agent] Verdict: YES Observation: The text segment includes the statement ‘We frankly dropped the ball of the process,’ which indicates an acknowledgment of failure, a component of transparency. The phrase ‘there is some kind of mitigating circumstances’ suggests an attempt to explain the situation, which could relate to clear communication. The admission of terminating employees ‘because they weren’t up to our standard’ shows honesty about internal actions. The mention of a senior executive with 35 years of experience implies a commitment to quality and regulatory standards, which can contribute to trust. The overall narrative reflects an honest account of events and actions taken, aligning with transparency and honesty principles. Reasoning: The text includes explicit statements about dropping the ball and being honest about the situation, which aligns with transparency. The mention of mitigating circumstances and the candor about employee terminations demonstrate clear communication. Additionally, the acknowledgment of a senior executive’s experience in quality and regulatory matters suggests a commitment to maintaining trust and honest reporting. [Moral Integrity Agent] Verdict: NO Observation: The text mentions that the company terminated employees at the On-X facility because they were not up to standard. This indicates a focus on maintaining quality and
--	---	---	--

			standards, which relates to ethical decision-making. The mention of a Senior Vice President with 35 years of experience suggests a leadership role in operations, quality, and regulatory matters, implying a structured approach to ethical considerations. The phrase ‘dropped the ball of the process’ and ‘fell through the cracks’ indicates potential lapses in oversight, which could relate to fairness in handling employee terminations. The text does not explicitly address CSR initiatives or broader societal contributions, nor does it provide details on equitable treatment beyond the mentioned terminations. The emphasis on standards and experienced leadership points to a commitment to ethical practices, though specific instances of fairness or CSR are not elaborated upon. Reasoning: The text discusses the termination of employees due to not meeting standards and acknowledges a lapse in process handling, which relates to ethical decision-making. However, there is no explicit mention of fairness in stakeholder treatment or CSR initiatives. The focus is on internal standards and quality, without clear evidence of equitable treatment or contributions to societal good beyond maintaining quality. Therefore, while there is an indication of ethical considerations, the evidence is insufficient to fully support the presence of Moral Integrity, Fairness, & CSR culture as defined.
Moral Integrity, Honesty	Please go ahead. Well, we’ve always been I would say focused on sustainability, the impact we have on the environment around us in terms of our operation, but also the positive impact we live in the various communities that we are part of around the world. What we haven’t done up until the last couple of years is to really document this and put this together into something that we can present to governments, as well as to the investor base. And I think that’s the main focus that we’ve had now over the past few years. There’s also a certain report and certain things that we do become part of in order to get certain certifications and rankings in terms of what our efforts are. So over the past couple years we have put that together into the report, which is then available both to local communities, governments and to the investment community where we really summarize and demonstrate what is long-term commitment is all about. And we understand the importance of this for the investment community for instance around what part of the	The text discusses the company’s focus on sustainability and its efforts to document and present this information to stakeholders. The Honesty Agent highlights the company’s commitment to transparency by creating reports and making them available to various stakeholders, which builds trust. The Moral_Integrity Agent notes the company’s alignment with ethical values through its sustainability efforts, including certifications and fairness. While the Compliance and Accountability Agents found insufficient evidence for their respective categories, the emphasis on ethical practices and transparency strongly supports Honesty and Moral_Integrity.	[Accountability Agent] Verdict: NO Observation: The text discusses the company’s focus on sustainability and its impact on the environment and communities. It mentions the importance of documenting and presenting this information to governments and the investment community. The company has created a report summarizing their long-term commitment to sustainability, which is available to local communities, governments, and investors. The report includes certifications and rankings related to their efforts. The text highlights the company’s understanding of the investment community’s interest in sustainability as a key decision criterion. There is no explicit mention of accountability, governance, or ownership structures within the organization. Reasoning: The text discusses the company’s focus on sustainability and its efforts to document and present this information to stakeholders, including governments and investors. However, there is no explicit mention of accountability, governance, or ownership structures within the organization. While the company highlights its commitment to sustainability and the importance of these efforts for investors, the absence of specific details about how the firm takes responsibility for outcomes, the structural systems in place to ensure accountability, or clear ownership of these initiatives makes it difficult to conclude that Accountability, Governance, & Ownership culture is clearly present. The observations from the Accountability Agent note that these aspects are not explicitly addressed. [Compliance Agent] Verdict: NO Observation: The text discusses the company’s focus on sustainability and its impact on the environment and communities. It mentions the importance of documenting these efforts to present to governments and investors. The company has created reports to obtain certifications and rankings, which are shared with local communities, governments, and the investment community. The reports summarize their long-term commitment to sustainability. The investment community’s interest in sustainability is highlighted as a key decision criterion.

	investment community are interested -- invested in the company where the focus on sustainability is obviously a key decision criteria.		Reasoning: The text discusses sustainability and creating reports for governments and investors but does not explicitly mention adherence to specific laws, regulations, or contractual obligations. While there is a focus on certifications and rankings, these are more about ethical standards and community relations rather than compliance with external legal requirements. There is no direct evidence of strict adherence to rules or legal obligations as defined in the Regulatory Compliance & Adherence culture. [Honesty Agent] Verdict: YES Observation: The text discusses the company's focus on sustainability and its impact on the environment and communities. It mentions the importance of documenting and presenting this information to governments and the investment community. The company has compiled reports to demonstrate its long-term commitment to sustainability, which is available to local communities, governments, and investors. The emphasis is on summarizing and showcasing their efforts to meet certifications and rankings. The investment community's interest in sustainability is highlighted as a key decision criterion. The company's actions, such as creating reports and ensuring availability of information, suggest efforts to maintain transparency and build trust with stakeholders. Reasoning: The text emphasizes the importance of documenting and presenting sustainability efforts to stakeholders, including investors and governments. The company has created reports to obtain certifications and rankings, which are made available to local communities, governments, and the investment community. This demonstrates a commitment to transparency by openly sharing their sustainability practices and efforts, thereby building trust with stakeholders. The emphasis on clear communication and availability of information aligns with the core concepts of Transparency, Trust, & Honesty culture. [Moral Integrity Agent] Verdict: YES Observation: The text mentions a focus on sustainability, indicating an aspect of CSR. The company has started documenting their environmental impact and community involvement, suggesting a commitment to ethical decision-making. They aim to present this information to governments and investors, highlighting transparency and accountability. The mention of certifications and rankings shows efforts to meet external standards, aligning with fairness and integrity. The company emphasizes a long-term commitment to sustainability, reflecting a broader ethical approach beyond legal requirements. Reasoning: The text explicitly discusses the company's commitment to sustainability, which is a key aspect of Corporate Social Responsibility (CSR). The company has documented its environmental and community impacts, indicating a deliberate effort to exceed legal requirements and align with ethical values. Additionally, the emphasis on presenting this information to investors and governments highlights transparency and fairness, aligning with the principles of Moral Integrity and Fairness. The mention of certifications and rankings further demonstrates a commitment to ethical standards and equitable treatment of stakeholders.
Moral Integrity, Accountability, Honesty	We are trying to be much more relationship driven and how we look at	The text segment demonstrates a strong commitment to several key principles.	[Accountability Agent] Verdict: YES

	pricing. I would say we are getting close to stability. And obviously, quite frankly, I will be honest and say that while it's frustrating to see that number come down, I think it's the right thing that happened.	The Accountability Agent highlights the speaker's ownership and acceptance of responsibility, indicating a culture of accountability. The Honesty Agent notes the explicit commitment to honesty and transparency, which is further supported by the speaker's candid language. The Moral Integrity Agent emphasizes ethical decision-making and fairness, aligning with principles of social responsibility. While the Compliance Agent did not find evidence of regulatory adherence, the other agents' verdicts collectively show a focus on ethical and responsible business practices.	Observation: The text segment discusses the company's shift towards being more relationship-driven and stable in pricing. It mentions frustration with a decreasing number but acknowledges it as the right outcome, indicating acceptance of responsibility for actions and outcomes. The speaker takes ownership of the situation and admits the difficulty in accepting the change, showing accountability. There is no explicit mention of governance structures or systems, such as boards or oversight committees. The focus is on the company's approach to pricing and stakeholder relationships rather than formal accountability mechanisms. Reasoning: The Accountability Agent observes that the speaker takes ownership of the situation, admits the difficulty in accepting the change, and acknowledges the outcome as the right thing. This indicates acceptance of responsibility for actions and outcomes, which aligns with the concept of accountability. While there is no explicit mention of governance structures, the focus on honesty and transparency suggests a commitment to ethical standards, reinforcing the presence of accountability, governance, and ownership culture. [Compliance Agent] Verdict: NO Observation: The text segment discusses the company's focus on being relationship-driven and stable in pricing. It mentions frustration with a number coming down but acknowledges it as the right thing. There is no explicit mention of external laws, government regulations, or contractual obligations. No specific compliance-related terms or phrases are present. The emphasis is on business strategy and customer relationships rather than adherence to regulatory standards. The absence of any compliance-related content is notable. Reasoning: The text segment does not mention any external laws, government regulations, or contractual obligations. It focuses on business strategy and customer relationships, with no compliance-related terms or evidence of adherence to regulatory standards. The emphasis is on honesty and ethical decision-making rather than legal or regulatory compliance. [Honesty Agent] Verdict: YES Observation: The speaker mentions being 'relationship driven' and 'getting close to stability,' which suggests a focus on maintaining trust and transparency in their approach. They explicitly state, 'I will be honest,' indicating a commitment to honesty. The phrase 'frankly' also implies a straightforward and candid communication style. The speaker acknowledges frustration about a number coming down but affirms it as the 'right thing that happened,' showing accountability and transparency in discussing challenges. Overall, the language used emphasizes honesty and openness in addressing the situation. Reasoning: The speaker explicitly states, 'I will be honest,' and uses the phrase 'quite frankly,' indicating a commitment to candor. They acknowledge the challenges and frustrations with the situation, showing transparency about the difficulties faced. Additionally, the focus on being 'relationship driven' and maintaining 'stability' suggests a dedication to building trust with stakeholders. These elements align with the Transparency, Trust, & Honesty culture as defined, which emphasizes open communication and the protection of stakeholder relationships. [Moral Integrity Agent] Verdict: YES
--	---	--	--

			Observation: The text mentions a commitment to being ‘relationship driven’ in pricing, which suggests a focus on ethical decision-making. The speaker expresses frustration about a number coming down but acknowledges it as the ‘right thing that happened,’ indicating a moral stance. The phrase ‘I will be honest’ reflects a commitment to transparency, a key aspect of moral integrity. The emphasis on doing what is ‘right’ aligns with ethical decision-making. The speaker’s acknowledgment of the situation as ‘frustrating’ but still morally correct underscores the firm’s commitment to fairness and integrity. Reasoning: The text segment demonstrates a clear commitment to moral integrity, fairness, and social responsibility. The speaker emphasizes being ‘relationship driven’ and acknowledges the ‘right thing’ that happened, indicating ethical decision-making. The use of phrases like ‘I will be honest’ and ‘frankly’ underscores a commitment to transparency, which is a key aspect of moral integrity. Additionally, the focus on fairness in pricing and stakeholder relationships aligns with the principles of CSR culture.
--	--	--	---

Table B8: Corporate governance and corporate integrity culture

This table presents the relation between (i) independent director ratio and integrity ratio and its subcomponents, and (ii) female director ratio and integrity ratio and its subcomponents. All variables are defined in **Table A1** presented in the Appendix. All models include industry and year fixed effects. White heteroscedastic-consistent robust standard errors in parentheses are clustered at the firm level (White, 1980). Asterisks (*, **, and ***) indicate the significance levels of 10 %, 5 %, and 1 %, respectively.

Variables	(1) Integrity ratio	(2) Integrity ratio	(3) Integrity ratio	(4) Accountability	(5) Compliance	(6) Moral	(7) Honesty	(8) Integrity Li tfidf lyr	(9) Integrity Li wtfidf lyr
Independent director ratio	0.0086*** (0.002)		0.0078*** (0.002)	0.0056*** (0.001)	0.0011 (0.001)	0.0015* (0.001)	0.0039*** (0.001)	0.0305* (0.016)	0.0168 (0.014)
Female director ratio		0.0108*** (0.003)	0.0093*** (0.003)	0.0057*** (0.002)	-0.0019* (0.001)	0.0048*** (0.001)	0.0053*** (0.002)	0.0333 (0.023)	0.0221 (0.019)
Firm size	0.0007*** (0.000)	0.0007*** (0.000)	0.0005** (0.000)	0.0001 (0.000)	0.0000 (0.000)	0.0008*** (0.000)	0.0001 (0.000)	0.0050*** (0.002)	0.0010 (0.001)
Leverage	-0.0006 (0.001)	-0.0006 (0.001)	-0.0006 (0.001)	0.0009 (0.001)	-0.0014** (0.001)	-0.0010 (0.001)	0.0006 (0.001)	-0.0405*** (0.012)	-0.0301*** (0.010)
Dividend to asset	0.0135*** (0.005)	0.0131*** (0.005)	0.0128*** (0.005)	0.0082** (0.003)	0.0035 (0.002)	0.0063*** (0.002)	0.0060** (0.003)	0.0582 (0.043)	0.0629* (0.033)
R&D to asset	-0.0422*** (0.004)	-0.0417*** (0.004)	-0.0422*** (0.004)	-0.0308*** (0.003)	-0.0060*** (0.002)	-0.0150*** (0.002)	-0.0206*** (0.002)	0.0489 (0.036)	0.0111 (0.028)
ROA	-0.0127*** (0.002)	-0.0125*** (0.002)	-0.0126*** (0.002)	-0.0076*** (0.001)	-0.0056*** (0.001)	-0.0057*** (0.001)	-0.0072*** (0.001)	-0.0928*** (0.019)	-0.0750*** (0.015)
Sales growth	-0.0017*** (0.000)	-0.0017*** (0.000)	-0.0016*** (0.000)	-0.0017*** (0.000)	0.0001 (0.000)	0.0001 (0.000)	-0.0008*** (0.000)	0.0072* (0.004)	0.0046 (0.003)
Standard deviation ROA	0.0537*** (0.012)	0.0535*** (0.012)	0.0539*** (0.012)	0.0392*** (0.008)	0.0254*** (0.006)	0.0271*** (0.006)	0.0303*** (0.007)	0.2917*** (0.096)	0.2690*** (0.078)
Constant	0.0350*** (0.002)	0.0401*** (0.002)	0.0354*** (0.002)	0.0267*** (0.002)	0.0085*** (0.001)	0.0071*** (0.001)	0.0216*** (0.001)	0.2265*** (0.021)	0.2417*** (0.017)
Year fixed effect	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry fixed effect	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	31,197	31,197	31,197	31,197	31,197	31,197	31,197	31,272	31,272
R-squared	0.028	0.027	0.029	0.028	0.065	0.048	0.023	0.045	0.042

Table B9: Corporate integrity and non-fraudulent restatement disclosure

This table presents the relation between integrity ratio and its subcomponents and non-fraudulent restatement filing. In the regression set-up, 'restat-filing' is a dummy variable; it is equal to '1' if the firm files for a non-fraudulent restatement in the subsequent year, or '0' otherwise. All variables are defined in **Table A1** presented in the Appendix. All models include industry and year fixed effects. White heteroscedastic-consistent robust standard errors in parentheses are clustered at the firm level (White, 1980). Asterisks (*, **, and ***) indicate the significance levels of 10 %, 5 %, and 1 %, respectively.

VARIABLES	(1) Restat filing	(2) Restat filing	(3) Restat filing	(4) Restat filing	(5) Restat filing	(6) Restat filing	(7) Restat filing
Integrity ratio	2.2172*** (0.727)						
Accountability		3.5249*** (0.953)					
Compliance			4.3115*** (1.476)				
Moral				1.4396 (1.420)			
Honesty					3.1403*** (1.130)		
Integrity Li tfidf 3yr						-0.1291 (0.109)	
Integrity Li wtfidf 3yr							-0.1406 (0.130)
Female director ratio	0.1832 (0.244)	0.1817 (0.244)	0.2148 (0.244)	0.1977 (0.244)	0.1889 (0.244)	0.0673 (0.232)	0.0665 (0.232)
Independent director ratio	0.2618 (0.170)	0.2591 (0.170)	0.2732 (0.170)	0.2769 (0.169)	0.2664 (0.170)	0.2920* (0.161)	0.2908* (0.161)
Firm size	-0.0455*** (0.016)	-0.0443*** (0.017)	-0.0444*** (0.017)	-0.0460*** (0.017)	-0.0449*** (0.017)	-0.0525*** (0.016)	-0.0529*** (0.016)
Leverage	0.3426** (0.135)	0.3394** (0.135)	0.3466*** (0.134)	0.3407** (0.135)	0.3382** (0.135)	0.2929** (0.129)	0.2939** (0.129)
Dividend to asset	-1.6939*** (0.498)	-1.6992*** (0.499)	-1.6845*** (0.497)	-1.6709*** (0.497)	-1.6845*** (0.497)	-1.5491*** (0.479)	-1.5481*** (0.479)
R&D to asset	-1.9517*** (0.429)	-1.9303*** (0.429)	-2.0237*** (0.431)	-2.0309*** (0.432)	-1.9828*** (0.431)	-1.9020*** (0.374)	-1.9069*** (0.374)
ROA	-0.2603 (0.187)	-0.2607 (0.187)	-0.2592 (0.187)	-0.2773 (0.187)	-0.2639 (0.187)	-0.2801* (0.163)	-0.2792* (0.163)
Sales growth	-0.0954* (0.051)	-0.0921* (0.051)	-0.1009** (0.051)	-0.1015** (0.051)	-0.0972* (0.051)	-0.1152** (0.046)	-0.1154** (0.046)
Standard deviation ROA	2.1540** (0.970)	2.1347** (0.973)	2.1746** (0.968)	2.2338** (0.969)	2.1691** (0.969)	2.8101*** (0.891)	2.8076*** (0.891)
Constant	-2.2702*** (0.216)	-2.2882*** (0.215)	-2.2268*** (0.214)	-2.1965*** (0.214)	-2.2571*** (0.215)	-2.1201*** (0.207)	-2.1160*** (0.209)
Year fixed effect	Y	Y	Y	Y	Y	Y	Y
Industry fixed effect	Y	Y	Y	Y	Y	Y	Y
Observations	28,430	28,430	28,430	28,430	28,430	32,329	32,329
Pseudo r-squared	0.0312	0.0315	0.0312	0.0307	0.0311	0.0292	0.0291

Table B10: Corporate integrity and earnings management

This table presents the relation between integrity ratio and its subcomponents and discretionary accruals (earnings management). All variables are defined in **Table A1** presented in the Appendix. All models include industry and year fixed effects. White heteroscedastic-consistent robust standard errors in parentheses are clustered at the firm level (White, 1980). Asterisks (*, **, and ***) indicate the significance levels of 10 %, 5 %, and 1 %, respectively.

VARIABLES	(1) Disc. Accrual	(2) Disc. Accrual	(3) Disc. Accrual	(4) Disc. Accrual	(5) Disc. Accrual	(6) Disc. Accrual	(7) Disc. Accrual
Integrity ratio	-0.0547** (0.027)						
Accountability		-0.0449 (0.035)					
Compliance			-0.0710 (0.062)				
Moral				-0.0405 (0.050)			
Honesty					-0.1156** (0.048)		
Integrity Li tfidf 3yr						0.0000 (0.003)	
Integrity Li wtfidf 3yr							0.0007 (0.004)
Female director ratio	0.0072 (0.006)	0.0069 (0.006)	0.0065 (0.006)	0.0068 (0.006)	0.0072 (0.006)	0.0064 (0.006)	0.0064 (0.006)
Independent director ratio	-0.0083** (0.004)	-0.0084** (0.004)	-0.0087** (0.004)	-0.0086** (0.004)	-0.0083** (0.004)	-0.0061 (0.004)	-0.0061 (0.004)
Firm size	-0.0022*** (0.000)	-0.0022*** (0.000)	-0.0022*** (0.000)	-0.0022*** (0.000)	-0.0022*** (0.000)	-0.0019*** (0.000)	-0.0019*** (0.000)
Leverage	0.0158*** (0.004)	0.0158*** (0.004)	0.0157*** (0.004)	0.0157*** (0.004)	0.0158*** (0.004)	0.0176*** (0.004)	0.0176*** (0.004)
R&D to asset	0.0424*** (0.013)	0.0433*** (0.013)	0.0443*** (0.013)	0.0440*** (0.013)	0.0423*** (0.013)	0.0524*** (0.013)	0.0523*** (0.013)
ROA	0.0708*** (0.008)	0.0711*** (0.008)	0.0711*** (0.008)	0.0712*** (0.008)	0.0706*** (0.008)	0.0722*** (0.007)	0.0722*** (0.007)
Sales growth	0.0030 (0.003)	0.0031 (0.003)	0.0032 (0.003)	0.0031 (0.003)	0.0030 (0.003)	0.0045* (0.002)	0.0045* (0.002)
Standard deviation ROA	0.0784** (0.035)	0.0774** (0.035)	0.0773** (0.035)	0.0769** (0.035)	0.0789** (0.035)	0.0467 (0.033)	0.0466 (0.032)
Constant	0.0181*** (0.006)	0.0173*** (0.006)	0.0167*** (0.006)	0.0164*** (0.006)	0.0186*** (0.006)	0.0122** (0.006)	0.0120** (0.006)
Year fixed effect	Y	Y	Y	Y	Y	Y	Y
Industry fixed effect	Y	Y	Y	Y	Y	Y	Y
Observations	26,675	26,675	26,675	26,675	26,675	30,506	30,506
R-squared	0.011	0.011	0.011	0.011	0.011	0.012	0.012

Table B11: Corporate integrity culture and dividend policy

This table presents the relation between integrity ratio and its subcomponents and dividend to total asset ratio. All variables are defined in **Table A1** presented in the Appendix. All models include industry and year fixed effects. White heteroscedastic-consistent robust standard errors in parentheses are clustered at the firm level (White, 1980). Asterisks (*, **, and ***) indicate the significance levels of 10 %, 5 %, and 1 %, respectively.

VARIABLES	(1) Div to asset	(2) Div to asset	(3) Div to asset	(4) Div to asset	(5) Div to asset	(6) Div to asset	(7) Div to asset
Integrity ratio	0.0588*** (0.020)						
Accountability		0.0618*** (0.022)					
Compliance			0.0988** (0.049)				
Moral				0.0939*** (0.036)			
Honesty					0.0828** (0.032)		
Integrity Li tfidf 3yr						0.0037 (0.003)	
Integrity Li wtfidf 3yr							0.0059* (0.003)
Female director ratio	0.0160*** (0.006)	0.0161*** (0.006)	0.0167*** (0.006)	0.0160*** (0.006)	0.0161*** (0.006)	0.0141** (0.006)	0.0141** (0.006)
Independent director ratio	0.0037 (0.004)	0.0039 (0.004)	0.0041 (0.004)	0.0041 (0.004)	0.0039 (0.004)	0.0053 (0.004)	0.0054 (0.004)
Firm size	0.0051*** (0.000)	0.0051*** (0.000)	0.0052*** (0.000)	0.0051*** (0.000)	0.0051*** (0.000)	0.0053*** (0.000)	0.0054*** (0.000)
Leverage	0.0112* (0.006)	0.0112* (0.006)	0.0114* (0.006)	0.0113* (0.006)	0.0112* (0.006)	0.0085 (0.006)	0.0085 (0.006)
R&D to asset	0.1513*** (0.014)	0.1508*** (0.014)	0.1495*** (0.014)	0.1503*** (0.014)	0.1506*** (0.014)	0.1262*** (0.016)	0.1263*** (0.016)
ROA	0.1418*** (0.010)	0.1416*** (0.010)	0.1417*** (0.010)	0.1417*** (0.010)	0.1417*** (0.010)	0.1209*** (0.011)	0.1210*** (0.011)
Sales growth	-0.0090*** (0.001)	-0.0090*** (0.001)	-0.0091*** (0.001)	-0.0091*** (0.001)	-0.0090*** (0.001)	-0.0092*** (0.001)	-0.0092*** (0.001)
Standard deviation ROA	0.1715*** (0.024)	0.1723*** (0.024)	0.1722*** (0.024)	0.1722*** (0.024)	0.1722*** (0.024)	0.1787*** (0.026)	0.1782*** (0.026)
Constant	-0.0099* (0.006)	-0.0095* (0.006)	-0.0086 (0.006)	-0.0085 (0.006)	-0.0096* (0.006)	-0.0076 (0.006)	-0.0082 (0.006)
Year fixed effect	Y	Y	Y	Y	Y	Y	Y
Industry fixed effect	Y	Y	Y	Y	Y	Y	Y
Observations	31,197	31,197	31,197	31,197	31,197	35,577	35,577
R-squared	0.147	0.146	0.146	0.146	0.146	0.129	0.129

Table B12: Corporate integrity culture and trade credit

This table presents the relation between integrity ratio and its subcomponents and net credit. All variables are defined in **Table A1** presented in the Appendix. All models include industry and year fixed effects. White heteroscedastic-consistent robust standard errors in parentheses are clustered at the firm level (White, 1980). Asterisks (*, **, and ***) indicate the significance levels of 10 %, 5 %, and 1 %, respectively.

VARIABLES	(1) Net credit	(2) Net credit	(3) Net credit	(4) Net credit	(5) Net credit	(6) Net credit	(7) Net credit
Integrity ratio	0.8678*** (0.271)						
Accountability		0.7615** (0.299)					
Compliance			2.7355*** (0.863)				
Moral				0.9313** (0.396)			
Honesty					0.9967** (0.389)		
Integrity Li tfidf 3yr						0.0835** (0.042)	
Integrity Li wtfidf 3yr							0.0892* (0.053)
Female director ratio	-0.1086 (0.085)	-0.1049 (0.085)	-0.0955 (0.085)	-0.1050 (0.085)	-0.1058 (0.085)	-0.1103 (0.084)	-0.1095 (0.084)
Independent director ratio	-0.0119 (0.068)	-0.0095 (0.068)	-0.0079 (0.068)	-0.0067 (0.069)	-0.0091 (0.069)	0.0197 (0.073)	0.0207 (0.073)
Firm size	0.0312*** (0.007)	0.0315*** (0.007)	0.0316*** (0.007)	0.0309*** (0.007)	0.0315*** (0.007)	0.0306*** (0.007)	0.0309*** (0.007)
Leverage	-0.2508*** (0.068)	-0.2519*** (0.068)	-0.2475*** (0.067)	-0.2503*** (0.068)	-0.2518*** (0.068)	-0.2414*** (0.066)	-0.2422*** (0.066)
Dividend to asset	0.2334* (0.121)	0.2384** (0.121)	0.2351* (0.121)	0.2388** (0.121)	0.2386** (0.121)	0.2267* (0.124)	0.2258* (0.124)
R&D to asset	0.6424*** (0.142)	0.6291*** (0.143)	0.6215*** (0.143)	0.6198*** (0.143)	0.6264*** (0.143)	0.6318*** (0.142)	0.6349*** (0.142)
ROA	-0.6596*** (0.075)	-0.6647*** (0.076)	-0.6554*** (0.074)	-0.6652*** (0.076)	-0.6633*** (0.076)	-0.6543*** (0.072)	-0.6552*** (0.072)
Sales growth	-0.1068*** (0.019)	-0.1069*** (0.019)	-0.1087*** (0.019)	-0.1083*** (0.019)	-0.1073*** (0.019)	-0.1072*** (0.018)	-0.1069*** (0.018)
Standard deviation ROA	1.0031** (0.392)	1.0198*** (0.392)	0.9825** (0.390)	1.0244*** (0.393)	1.0196*** (0.393)	0.8498** (0.366)	0.8509** (0.367)
Constant	-0.1610** (0.070)	-0.1506** (0.069)	-0.1537** (0.070)	-0.1369** (0.067)	-0.1517** (0.069)	-0.1661** (0.070)	-0.1686** (0.069)
Year fixed effect	Y	Y	Y	Y	Y	Y	Y
Industry fixed effect	Y	Y	Y	Y	Y	Y	Y
Observations	30,885	30,885	30,885	30,885	30,885	35,208	35,208
R-squared	0.069	0.068	0.070	0.068	0.068	0.070	0.070

Essay 2

Beyond Keywords: Measuring Corporate Culture Through the Competing Values Framework Using Large Language Models

Abstract

This study develops new Contextual Competing Values Framework (CVF) Culture Scores using a multi-agent large language model (LLM) architecture to measure organizational culture more accurately. We analyze the unscripted question-and-answer sections of 183,043 quarterly earnings call transcripts from 2008 to 2024, where executives respond spontaneously to analysts. Unlike traditional dictionary-based methods that rely on keyword counts, our approach interprets language in context and identifies genuine culture-related discourse. The scores capture the four CVF cultural orientations: Collaborate (teamwork and human development), Create (innovation and adaptability), Control (efficiency and adherence to rules), and Compete (market share and profitability). We validate the scores in two ways. First, the classifier achieves 96.00% accuracy and a Cohen's kappa of 0.920 against human judgments, indicating near-perfect reliability. We also find that many of the most frequently occurring dictionary terms from prior studies appear in culture-related segments less than half the time, demonstrating that keyword presence alone is a weak indicator of genuine cultural discourse. Second, we pair each CVF dimension with the firm outcomes its own theoretical logic predicts. Firms with higher Collaborate scores maintain significantly larger workforces; higher Create scores are associated with greater R&D intensity and patent output; higher Compete scores correspond to stronger profitability and market share; and higher Control scores relate to greater regulatory discipline. When benchmarked against the widely used dictionary-based measures of Fang et al. (2023), our contextual scores produce theoretically consistent and statistically significant associations across all four dimensions, whereas the dictionary-based scores yield inconsistent and often economically negligible results. Overall, the findings demonstrate that contextual AI-based and reasoning-based measures provide a more informative view of organizational culture than traditional keyword methods, and that the apparent irrelevance of certain CVF dimensions in prior work likely reflects measurement noise rather than a genuine absence of cultural influence on firm behavior.

1. Introduction

Organizational culture—the shared values, beliefs, and norms that guide behavior within firms—has long been recognized as a fundamental determinant of organizational effectiveness, strategic decision-making, and economic performance (Guiso et al., 2015; Graham et al., 2022; O'Reilly & Chatman, 1996). Among the many frameworks developed to conceptualize and classify organizational culture, the Competing Values Framework (CVF) introduced by Quinn and Rohrbaugh (1983) has emerged as the most widely adopted and empirically validated typology in the organizational sciences (Cameron & Quinn, 2006; Hartnell et al., 2011). The CVF organizes organizational culture along two fundamental tensions: internal versus external orientation and flexibility versus stability; yielding four distinct cultural archetypes: Clan (collaborative, people-centered), Adhocracy (innovative, entrepreneurial), Market (competitive, results-oriented), and Hierarchy (controlled, process-driven). Meta-analytic evidence confirms that these four orientations relate to employee attitudes, operational effectiveness, and financial performance in theoretically predictable ways (Hartnell et al., 2011), making the CVF a particularly attractive foundation for large-scale empirical research on how culture shapes firm behavior and outcomes.

Despite the theoretical maturity of the CVF, measuring organizational culture at scale remains an empirical challenge. Culture is inherently intangible: unlike governance mechanisms or financial ratios, it is not directly observable and must be inferred from its manifestations in language, decisions, and behavior (Graham et al., 2022; Kreps, 1990). Early approaches relied on survey instruments such as the Organizational Culture Assessment Instrument (OCAI), which, while psychometrically validated, are costly to administer, limited in sample coverage, and susceptible to social-desirability bias (Cameron & Quinn, 2006). Recognizing these constraints, a

growing body of research has turned to textual analysis of corporate communications, particularly annual reports (10-Ks) and earnings call transcripts as a scalable window into corporate culture (Fiordelisi & Ricci, 2014; Li et al., 2021; Fang et al., 2023). These text-based approaches have significantly expanded the scope of culture research, enabling studies that cover thousands of firms over extended time periods.

However, the dominant methodology in this literature: dictionary-based word counting; rests on an assumption that is rarely tested: that the occurrence of a word from a culture-specific lexicon constitutes reliable evidence of the corresponding cultural orientation. This assumption is problematic because the same words can carry very different meanings depending on context. The stem “partner*” may reflect a collaborative orientation when an executive discusses joint decision-making with employees, but it signals nothing about Clan culture when it refers to a joint-venture counterparty or a limited partnership structure. Similarly, “revenue*” appears in the Market (competitive) dictionary yet overwhelmingly surfaces in routine financial reporting rather than discussions of competitive strategy or market dominance. Without the ability to distinguish substantive cultural discourse from incidental keyword occurrence, dictionary-based proxies may recover something other than the cultural construct researchers intend to measure. If this is the case, prior findings that rely on these proxies may understate, or even mischaracterize, the economic significance of organizational culture.

Fang et al. (2023) make an important contribution by introducing CVF-based culture measures to the corporate finance literature, constructing updated dictionaries for the four CVF dimensions and documenting the stabilizing role of controlling (Hierarchy) culture during the 2008–09 financial crisis. However, their analysis is focused almost exclusively on the Control dimension and on crisis-period outcomes. The other three CVF dimensions, i.e., Collaborate,

Compete, and Create are consistently insignificant in their main specifications, and the paper does not investigate whether this reflects genuine economic irrelevance or measurement limitations. More broadly, the dictionary-based methodology they employ inherits the fundamental context-insensitivity problem described above: their wordlists contain highly generic stems such as “cost*,” “system*,” “new*,” and “value*” that appear in virtually every corporate filing for reasons entirely unrelated to organizational culture (Koch & Pasch, 2023). Whether the inability of their Collaborate, Compete, and Create measures to produce consistent empirical associations reflects a genuine absence of economic content in these cultural dimensions or an artifact of noisy measurement remains an open question, one that has significant implications for how researchers interpret the role of culture in shaping firm behavior.

We address these gaps by developing Contextual CVF Culture Scores derived from the unscripted question-and-answer (Q&A) portion of quarterly earnings calls, using a reasoning large language model (LLM) embedded in a multi-agent classification architecture. Earnings call Q&A sessions are a uniquely informative source for measuring organizational culture: unlike prepared remarks or annual filings, the Q&A portion exposes executives to unanticipated questions from analysts, producing spontaneous responses that are widely regarded as more revealing of managerial attitudes, strategic orientations, and organizational values than rehearsed disclosures (Li et al., 2021; Price et al., 2012). We assemble 183,043 earnings call transcripts spanning 2008 to 2024, extract executive responses from the Q&A portion, and segment them into coherent discourse units using a semantically aware adaptation of the C99 algorithm, yielding a corpus of over 10.2 million text segments. Each segment is routed through a multi-agent pipeline inspired by blackboard architecture (Salemi et al., 2025). The pipeline begins with a binary classification gate that identifies segments containing genuine organizational culture

discourse. Segments that pass the gate are then analyzed by four domain-specific agents, each aligned with one dimension of the CVF (i.e., Clan, Adhocracy, Hierarchy, and Market) operating through sequential phases that separate evidence extraction from contextual judgment. Aggregating segment-level classifications to the firm-quarter level produces an overall culture measure and four component-specific scores that can be tracked across firms and over time.

An important empirical regularity that emerges from this classification exercise concerns the prevalence of culture-related discourse in earnings calls. Organizational culture is not a topic that executives address in every exchange; the vast majority of earnings call discourse is devoted to financial results, operational metrics, and forward guidance. Our method identifies culture-positive segments as constituting approximately 20% of the total corpus, confirming that genuine cultural discourse represents a minority of earnings call content and that the resulting class imbalance makes contextual discrimination especially important—a purely dictionary-based approach that counts every keyword occurrence as cultural evidence would substantially overcount the frequency of meaningful cultural discourse.

We validate the measure in two complementary ways. First, we evaluate the classification performance of the overall culture detection gate against a human-annotated set of 300 segments, drawn through a stratified design that ensures balanced representation of culture-related and non-culture-related classes. The classifier achieves 96.00% accuracy, a Cohen's kappa of 0.920, and a Matthews correlation coefficient of 0.923 against human annotations, with five independent stability runs producing consistently strong agreement. Comparison against the dictionary-based specification of Fang et al. (2023) further demonstrates that keyword presence is a weak indicator of genuine cultural discourse: across the four CVF components, many of the most frequently occurring dictionary terms appear in culture-related segments less than half the time,

with some of the most common stems such as “detail*” in the Hierarchy dictionary and “partic*” in the Clan dictionary registering culture-relevance rates below 35%. Qualitative inspection corroborates this finding, showing that dictionary terms surface in technical, organizational, or otherwise incidental contexts that bear no relation to organizational culture. This evidence directly tests and challenges the prevailing assumption that keyword presence alone reliably signals meaningful cultural discourse.

Second, we validate the resulting contextual measures empirically by conducting a systematic construct-validation exercise that pairs each CVF dimension with the specific firm outcomes that its own theoretical logic predicts. We find that firms with higher Collaborate scores maintain a significantly larger employee base relative to their assets and incur greater spending on labor and coordination related activities, consistent with the CVF’s characterization of Clan culture as centering on employee participation and human development (Cameron & Quinn, 2006; Edmans, 2011). Firms with higher Compete scores exhibit significantly stronger operating profitability and larger product-market shares, aligning with the Market quadrant’s emphasis on competitiveness, goal achievement, and market positioning (Campello, 2003; He & Huang, 2017). Higher Create scores are associated with significantly greater R&D intensity and patent output, consistent with the Adhocracy quadrant’s emphasis on innovation, experimentation, and entrepreneurial risk-taking (Li et al., 2021; Fiordelisi et al., 2019). Finally, firms with higher Control scores display more conservative tax behavior and a greater tendency to formally identify and correct reporting errors through restatement filings, in line with the Hierarchy quadrant’s emphasis on procedural discipline, monitoring, and rule-based compliance (Dyreng et al., 2008; Ashbaugh-Skaife et al., 2008). Across these analyses, the Fang et al. (2023) dictionary-based scores capture some of these patterns, particularly for product-market

performance and patent output but produce statistically insignificant associations for several key paired validation outcomes, including employee intensity for the Collaborate dimension, R&D intensity for the Create dimension, and both tax behavior and restatement filing for the Control dimension. This pattern—consistent, theoretically grounded associations across all four dimensions for our contextual measures, versus incomplete and selectively significant associations for the dictionary-based benchmark—indicates that context-aware approach to gauge organizational culture is not merely a methodological refinement but a necessary condition for recovering the distinct economic content embedded in each cultural dimension.

Our study makes three contributions. First, we advance the IS literature at the intersection of unstructured data analytics and organizational research by introducing an interpretable, multi-agent LLM pipeline for large-scale, context-sensitive measurement of organizational culture from corporate communications. This contribution builds on prior IS work that has applied deep learning to financial prediction (Yang et al., 2023), topic modeling to organization-generated content (Sharma et al., 2025), and a range of NLP methods to analyze textual data in organizational contexts (Bhatt et al., 2023; Kaur et al., 2021; Kurumathur et al., 2025; Singh et al., 2022). By embedding a reasoning LLM within a multi-agent architecture that separates evidence extraction from contextual judgment, we demonstrate how generative AI can function not only as a classification tool but as a disciplined reasoning framework for measuring complex, intangible constructs at scale, responding to calls in the IS literature for research that connects generative AI innovation to organizational and societal impact (Storey et al., 2025). To our knowledge, this is among the first studies to deploy generative AI for measuring the full multidimensional structure of organizational culture as conceptualized by the CVF, extending the

application of methodological innovations introduced in the previous essay on corporate integrity to a complementary but distinct cultural framework.

Second, we contribute to the corporate culture literature by introducing validated, context-aware measures of all four CVF dimensions and providing the first systematic construct-validation exercise that benchmarks these measures against theoretically predicted firm outcomes on a dimension-by-dimension basis. Existing research has either validated only a subset of CVF dimensions (Fang et al., 2023), relied on survey-based operationalizations with limited sample coverage (Cameron & Quinn, 2006), or employed dictionary-based methods without testing whether keyword presence corresponds to genuine cultural discourse. Our four-component framework extends the foundational work of Quinn and Rohrbaugh (1983) and Cameron and Quinn (2006) by demonstrating that all four cultural archetypes carry distinct, measurable economic content when measured with a context aware methodology—a finding that the dictionary-based approach of Fang et al. (2023) does not consistently support. By conducting both four-culture and single-dimension regressions, and by benchmarking against the Fang et al. (2023) scores on the same validation outcomes across a panel of up to approximately 39,000 firm-year observations, we provide a more rigorous and complete assessment of construct validity than has previously been available in the literature. This design reveals that the apparent economic irrelevance of certain CVF dimensions in prior work likely reflects measurement noise rather than a genuine absence of cultural influence on firm behavior.

Third, we contribute to the emerging literature on generative AI in management, accounting, and finance research. While generative AI has been applied to stock price forecasting (Lopez-Lira & Tang, 2023), analyst report analysis (Li et al., 2026), and detection of disclosure inefficiencies (Kim et al., 2025), its application to operationalizing the multidimensional

structure of organizational culture is novel. By showing that context-aware CVF scores relate to labor intensity, profitability, innovation, and regulatory discipline in ways that a widely used dictionary-based proxy does not, we provide evidence that LLM-based measurement of unscripted managerial discourse can reveal decision-relevant information about organizational culture that existing tools struggle to capture. This finding reinforces the strategic role of IS in transforming unstructured communication into actionable organizational knowledge, and it demonstrates that the value of contextual reasoning is not confined to a single cultural dimension; rather, it extends across the broader framework of competing values through which organizations navigate tensions between flexibility and stability, as well as internal integration and external differentiation.

The remainder of the paper is organized as follows. Section 2 reviews the relevant literature on organizational culture, the Competing Values Framework, and computational approaches to culture measurement. Section 3 details our methodology for identifying and classifying culture-related discourse in earnings calls, including the text segmentation procedure, model selection, multi-agent classification architecture, and score construction. Section 4 presents the classification performance results, including classification performance for the Phase I culture detection gate, the role of contextual understanding relative to dictionary-based approaches. Section 5 presents the empirical associations with theoretically predicted firm-level outcomes for each CVF dimension. Section 6 discusses methodological, theoretical, and practical implications. Section 7 concludes.

2. Literature Review

We structure the literature review around three areas that together establish the conceptual and methodological foundations for using IS capabilities to construct a contextual measure of organizational culture from earnings call discourse. Section 2.1 traces the evolution of textual analysis within the management domain, outlining the trajectory from dictionary-based word counts to context-aware large language models. Section 2.2 develops the Competing Values Framework (CVF) as the theoretical lens for our study, reviews its origins, discusses how the cultural dimensions identified in prior literature map systematically onto the CVF's four quadrants, and elaborates each quadrant in detail. Section 2.3 focuses on earnings calls as a data source for cultural measurement, examines how textual analysis techniques have been applied to extract cultural signals from corporate communications, and identifies the limitations of existing approaches. Section 2.4 synthesizes these strands to articulate the research gap and motivation for the present study.

2.1 The Evolution of Textual Analysis in Management Research

Over the past two decades, textual analysis in management research has evolved in step with advances in natural language processing (NLP). Early work relied on dictionary-based methods, exemplified by Tetlock's (2007) finding that negative news tone predicts market downturns. Loughran and McDonald (2011) subsequently developed finance-specific lexicons that addressed the domain mismatch inherent in general-purpose dictionaries, enabling more accurate predictions of returns, volatility, and profitability from disclosure tone (Henry & Leone, 2015). These lexicons represented an important advance in measurement precision, demonstrating that

word meaning is domain-dependent. However, the underlying methodology remained fundamentally context-insensitive: text was treated as an unordered collection of independent word tokens, with no capacity to account for syntactic structure, negation, irony, or the broader discursive context in which terms appear.

To move beyond surface-level lexical matching, researchers adopted machine learning techniques capable of capturing broader patterns in text (Li, 2010; Patel et al., 2015). A further paradigm shift came with neural language models, beginning with word embeddings such as Word2Vec (Mikolov et al., 2013) that represented semantic relationships in continuous vector spaces, recognizing, for example, that “bankruptcy” and “Chapter 11” occupy adjacent regions of a learned semantic space. Li et al. (2021) applied this approach to construct culture dictionaries from earnings call corpora, using Word2Vec-trained embeddings to inform the selection and grouping of seed words based on their semantic proximity, thereby providing a more content-sensitive foundation for dictionary-based analysis. Dutta et al. (2023) similarly used Word2Vec to study contextualized COVID-related disclosures in 10-K filings. Although these embedding-based methods captured more semantic structure than static dictionaries, they still operated at the word level, limiting their ability to interpret the full discursive context in which a cultural signal appears. A word embedding can identify that “collaboration” and “teamwork” are semantically proximate, but it cannot determine whether a specific mention of “teamwork” reflects a genuine cultural orientation or merely describes a project’s organizational structure.

Most recently, transformer-based models such as BERT (Devlin et al., 2018), FinBERT (Huang et al., 2023), and generative large language models (LLMs) have transformed NLP. Trained on very large corpora, LLMs perform diverse downstream tasks with minimal supervision through few- and zero-shot learning (de Kok, 2025), and contemporary reasoning

models can articulate the basis for their decisions through structured chains of reasoning (Huang & Chang, 2022; Shojae-Mend et al., 2024; Wang et al., 2022; Wei et al., 2022). These capabilities have opened new avenues for financial text analysis, including document parsing (Kim et al., 2025), management speech analysis (de Kok, 2025), investment prediction (Jha et al., 2023), deception detection (Beckmann et al., 2024), and stock forecasting (Lopez-Lira & Tang, 2023). The trajectory from word counts to contextual reasoning motivates our use of a reasoning LLM to measure organizational culture through the CVF lens, a construct whose manifestation in managerial discourse is inherently context-dependent and requires interpretation that transcends individual word frequencies.

2.2 The Competing Values Framework: Theoretical Foundations and Components

Organizational culture has been defined as the shared values, beliefs, and assumptions that shape behavior within organizations (Schein, 2010). Despite its acknowledged importance, culture remains difficult to operationalize empirically because it is intangible and typically observable only through its manifestations in language and behavior (Graham et al., 2022; Kreps, 1990). Several prominent studies have sought to identify and catalogue the specific cultural values that matter for organizational outcomes. Guiso et al. (2015) examined how “advertised” corporate values (integrity, teamwork, innovation, respect, and quality) relate to firm performance and stakeholder outcomes. Graham et al. (2022) surveyed over 1,300 executives and identified adaptability, collaboration, results-orientation, and detail-orientation, among others, as core cultural priorities. Grennan (2019) developed a cultural profile measure using similar dimensions drawn from corporate filings and analyst reports. Li et al. (2021) operationalized five cultural

values (innovation, integrity, quality, respect, and teamwork) through a Word2Vec-enhanced dictionary approach applied to earnings call transcripts.

While these studies have made significant contributions to understanding the role of specific cultural values, the resulting cultural typologies have been developed inductively from different data sources and methods, and as a result they vary in their component dimensions and lack a unifying theoretical structure. The Competing Values Framework, by contrast, provides an empirically derived, theoretically grounded organizing structure that can accommodate and integrate these diverse cultural dimensions within a single coherent model. As Fang et al. (2023) observed, the CVF's two-axis model effectively captures and integrates the cultural values identified across different frameworks. We adopt the CVF as our theoretical lens precisely because of this integrative capacity: it allows us to measure organizational culture comprehensively and to examine how competing cultural orientations coexist and interact within the same firm, rather than treating individual values in isolation.

2.2.1 Origins and Development of the CVF

The Competing Values Framework (CVF) is a widely recognized and empirically derived model that provides a comprehensive map of organizational culture, as illustrated in Figure 1 (Cameron & Quinn, 2006; Quinn & Rohrbaugh, 1983). Its origins trace to Quinn and Rohrbaugh's (1983) foundational work, in which they conducted a statistical analysis of organizational effectiveness indicators to identify the underlying value dimensions that structure organizational life. Through extensive factor analysis of effectiveness criteria rated by organizational theorists, their research revealed two fundamental axes of tension that define the cultural landscape of any organization (Fiordelisi & Ricci, 2014; Nguyen et al., 2019).

[Insert Figure 1]

The first axis contrasts an Internal Focus and Integration with an External Focus and Differentiation. Organizations on the internal end prioritize cohesion, employee development, and process control, while those on the external end prioritize competition, market positioning, and growth. The second axis contrasts Flexibility and Discretion with Stability and Control. Organizations on the flexible end value innovation, adaptation, and autonomy, while those on the stable end value efficiency, predictability, and order (Cameron et al., 2006; Fang et al., 2023). The theoretical power of the CVF lies in the recognition that these tensions are not pathological but constitutive: effective organizations do not resolve them in favor of one pole but manage the inherent tensions between competing demands (Cameron & Quinn, 2006). This insight distinguishes the CVF from frameworks that treat cultural values as independent attributes and positions it as a relational model in which each quadrant acquires meaning partly through its contrast with the others.

The intersection of these two axes creates four distinct cultural quadrants, each representing a coherent set of values, leadership styles, and effectiveness criteria: Clan (Collaborate), located in the internal-flexible quadrant; Adhocracy (Create), located in the external-flexible quadrant; Hierarchy (Control), located in the internal-stable quadrant; and Market (Compete), located in the external-stable quadrant. Each quadrant operates on distinct theoretical assumptions about what makes organizations effective, and each implies a characteristic approach to leadership, strategy, and stakeholder engagement. The following sections elaborate each component in detail.

2.2.2 Integration of Cultural Values from Prior Literature with CVF Dimensions

A key rationale for adopting the CVF is its capacity to serve as an empirically derived, integrative structure that organizes and contextualizes the various cultural dimensions identified in prior research. As Fang et al. (2023) noted, the CVF's two-axis model effectively subsumes the cultural values catalogued across multiple influential studies. A systematic mapping exercise demonstrates that the seemingly ad hoc lists of cultural values identified in recent prominent studies align systematically within the four CVF quadrants, as shown in Figure 2.

[Insert Figure 2]

The Clan quadrant, defined by an internal focus and flexibility, subsumes people-oriented values found across the literature. These include “collaboration” (Grennan, 2019; Graham et al., 2022), “teamwork” (Guiso et al., 2015; Li et al., 2021), “respect” (Guiso et al., 2015; Li et al., 2021), “community” (Guiso et al., 2015; Graham et al., 2022), and “collaboration and people-focused” orientations (Li et al., 2026). The Adhocracy quadrant, characterized by an external focus and flexibility, aligns with innovation-centric dimensions such as “innovation” (Guiso et al., 2015; Li et al., 2021), “adaptability” (Grennan, 2019; Graham et al., 2022), and “innovation and adaptability” (Li et al., 2026). The Hierarchy quadrant, defined by an internal focus and stability, groups values associated with process reliability, including “quality” (Guiso et al., 2015; Li et al., 2021), “detail-orientation” (Grennan, 2019; Graham et al., 2022), and “safety” (Guiso et al., 2015). The Market quadrant, combining an external focus with stability, serves as the logical home for performance-oriented values such as “results-orientation” (Grennan, 2019; Graham et al., 2022), “performance-oriented” (Li et al., 2026), “customer-orientation” (Grennan, 2019; Graham et al., 2022), and “hard work” (Guiso et al., 2015).

2.2.3 Clan (Collaborate) Culture

The Clan culture, occupying the internal-flexible quadrant, is characterized by an orientation toward internal maintenance and human development with a high degree of flexibility (Cameron et al., 2006; Fiordelisi & Ricci, 2014). Organizationally, it functions like an extended family, bound by loyalty, tradition, and mutual commitment. Key value drivers include teamwork, cohesion, employee participation, empowerment, and long-term investment in human capital. The strategic emphasis is on creating a humane work environment where employees feel valued, involved, and psychologically safe. Leadership in a Clan culture is expressed through mentoring, facilitating, and team-building, with effectiveness defined by employee morale, satisfaction, and organizational commitment (Fang et al., 2023; Luu et al., 2023). The underlying theory of effectiveness holds that high morale, commitment, and participation lead to organizational success through enhanced human resource development and collective problem-solving.

The Clan quadrant draws on a substantial body of organizational behavior research. The emphasis on participation and empowerment reflects Kanter's (1977) foundational work on structural empowerment, which demonstrated that employee access to information, resources, and support predicts both satisfaction and productivity. Denison and Mishra (1995), in their study of cultural traits and organizational effectiveness, found that the involvement trait, encompassing empowerment, team orientation, and capability development, predicted both employee satisfaction and financial performance. More recently, research on psychological safety (Edmondson, 1999) has reinforced the importance of clan-type cultural attributes: organizations that foster psychological safety encourage knowledge sharing, error reporting, and innovation, all of which depend on the participative and trust-based climate that defines this quadrant.

In the context of earnings calls, Clan culture manifests through executives' emphasis on employee well-being, team achievements, collaborative decision-making processes, and long-term human capital development. References to employee satisfaction, retention strategies, talent development programs, and participative management approaches signal this cultural orientation. The challenge for measurement is that terms such as "team" or "collaboration" frequently appear in operational contexts (e.g., "management team," "cross-functional collaboration on the product launch") without necessarily reflecting the underlying cultural values of participation, empowerment, and human development that define the Clan quadrant. This distinction between cultural discourse and incidental word use is precisely what motivates our LLM-based contextual classification approach.

2.2.4 Adhocracy (Create) Culture

The Adhocracy culture, located in the external-flexible quadrant, is a dynamic, entrepreneurial, and creative workplace (Fiordelisi et al., 2019; Hasan et al., 2024). It is oriented toward external positioning and thrives on adaptability and change. Key value drivers include innovation, risk-taking, visionary leadership, and the pursuit of cutting-edge products and services. The strategic emphasis is on fostering creativity, pioneering new initiatives, and rapidly adapting to market shifts. Leaders in an Adhocracy are innovators, entrepreneurs, and visionaries who are willing to challenge the status quo and tolerate calculated failures. The theory of effectiveness posits that constant innovation and the creation of new resources are the keys to sustainable market success (Cameron & Quinn, 2006).

The theoretical underpinnings of the Adhocracy quadrant connect to a rich body of research on organizational innovation and entrepreneurship. Burns and Stalker (1961) were

among the first to identify organic organizational structures, characterized by flexibility, decentralized authority, and lateral communication, as essential for innovation in rapidly changing environments. March (1991) formalized the tension between exploitation and exploration, arguing that organizations must balance the refinement of existing competencies with the search for new possibilities, and the Adhocracy quadrant corresponds directly to the exploration pole of this framework. Teece et al. (1997) further elaborated this logic through the dynamic capabilities perspective, arguing that competitive advantage in turbulent environments depends on the firm's capacity to sense, seize, and reconfigure resources, capacities that are cultivated within an Adhocratic cultural orientation.

During earnings calls, Adhocracy culture would be evident in discussions of R&D investments, breakthrough innovations, market disruption strategies, and tolerance for calculated risk-taking. Executives might emphasize experimentation, learning from failures, and the organization's ability to adapt rapidly to changing competitive dynamics. As with other cultural dimensions, the measurement challenge is contextual: the word "innovation" may appear in generic product descriptions without reflecting an underlying cultural commitment to creative exploration. An LLM-based approach can distinguish between a statement like "we launched an innovative new product feature" (operational usage) and "our culture rewards experimentation and we encourage teams to take risks and learn from failure" (genuine cultural discourse).

2.2.5 Hierarchy (Control) Culture

The Hierarchy culture, situated in the internal-stable quadrant, represents a formalized and structured environment oriented toward internal maintenance with a primary focus on stability and control (Fang et al., 2023; Nguyen et al., 2019). Key value drivers include efficiency,

reliability, predictability, consistency, and adherence to formal rules and standardized procedures. The strategic emphasis is on smooth operations, quality control, and error reduction. Leaders in a Hierarchy culture act as coordinators, monitors, and organizers who ensure that processes are followed and standards are met. The theory of effectiveness holds that control and efficiency, achieved through capable and reliable processes, produce stable and predictable outcomes (Barth & Mansouri, 2021; Cameron & Quinn, 2006).

The Hierarchy quadrant draws on Weber's (1947) classical bureaucratic model, which holds that organizational effectiveness derives from rational-legal authority, formal rules, and hierarchical specialization. In the management literature, the emphasis on process control and reliability finds support in the high-reliability organization (HRO) tradition (Weick & Sutcliffe, 2007), which demonstrates that organizations operating in high-consequence environments achieve remarkable safety and consistency through cultures that prioritize procedural discipline, deference to expertise, and preoccupation with failure. The quality management literature, particularly the work of Deming (1986) and the total quality management movement, further reinforces this quadrant's emphasis on continuous process improvement, standardization, and measurement as pathways to operational excellence.

In earnings call discourse, Hierarchy culture manifests through emphasis on operational efficiency, risk management systems, compliance frameworks, internal controls, and standardized processes. Executives might highlight quality assurance measures, the predictability of business operations, and improvements to operational infrastructure. The contextual measurement challenge here mirrors those of other quadrants: terms such as "efficiency" or "quality" are ubiquitous in corporate communications and frequently appear in purely financial or operational contexts (e.g., "cost efficiency improved by 200 basis points") without reflecting

the cultural emphasis on process discipline, standardization, and procedural reliability that characterizes the Hierarchy archetype.

2.2.6 Market (Compete) Culture

The Market culture, occupying the external-stable quadrant, is a results-oriented and competitive environment focused on external positioning, market dominance, and financial performance (Fiordelisi & Ricci, 2014; Luu et al., 2023). Key value drivers include market share, goal achievement, profitability, productivity, and aggressive competition. The strategic emphasis is on winning in the marketplace by outperforming competitors and satisfying customers. Leaders are hard-drivers, producers, and competitors who are demanding and results-focused. The theory of effectiveness holds that aggressively competing and achieving ambitious goals lead to market leadership and sustained profitability (Cameron & Quinn, 2006).

The Market quadrant connects to the competitive strategy literature initiated by Porter (1980, 1985), which frames organizational effectiveness in terms of competitive positioning, cost leadership, differentiation, and sustained competitive advantage. Williamson's (1975) transaction cost economics provides additional theoretical grounding: Market-oriented cultures emphasize arm's-length relationships, explicit performance metrics, and contractual discipline, all of which align with transaction cost reasoning about efficient governance of exchange relationships. Empirical research has documented the consequences of strong Market cultures: Barth and Mansouri (2021) found that Market culture in banking is associated with higher risk-taking and more aggressive competitive behavior, while Luu et al. (2023) showed that the relative strength of Market versus Clan orientations predicts bank stability outcomes.

During earnings calls, Market culture is reflected in discussions of competitive positioning, market share gains, customer acquisition strategies, financial performance metrics, and shareholder value maximization. Executives emphasize beating competitors, achieving targets, and delivering superior returns. The measurement challenge is again contextual: virtually every earnings call discusses financial results and competitive dynamics, but not all such discussions reflect a Market cultural orientation as defined by the CVF. Reporting quarterly revenue figures is informational; framing every organizational decision through the lens of competitive dominance and aggressive goal achievement reflects an underlying cultural commitment. Distinguishing these requires the interpretive capacity that an LLM-based approach provides.

2.3 Earnings Calls and the Measurement of Organizational Culture

2.3.1 Earnings Calls as a Data Source for Cultural Analysis

Earnings calls are key information events for public firms and have attracted extensive research attention in accounting, finance, and information systems (Bochkay et al., 2020; Loughran & McDonald, 2016). A typical call consists of two distinct segments: a scripted management presentation followed by an unscripted question-and-answer (Q&A) session with financial analysts. Although both segments convey information, the research literature has increasingly converged on the view that the Q&A portion is the more diagnostic segment for inferring managerial attitudes, organizational values, and underlying firm culture. In this section, we develop a systematic case for the Q&A portion of earnings calls as the primary data source for measuring organizational culture through the CVF, drawing on the theoretical distinction between espoused and enacted culture, the empirical evidence on the information content of

Q&A discourse, and the well-documented limitations of alternative data sources, particularly 10-K filings.

The conceptual foundation for our data source choice rests on a distinction that is central to the study of corporate values: the difference between espoused culture and enacted culture. Guiso et al. (2015), in their influential study of the value of corporate culture, drew a sharp distinction between “advertised” corporate values, those that firms publicly espouse through mission statements, codes of conduct, and formal disclosures, and the values that actually guide organizational behavior in practice. Their analysis demonstrated that advertised values bear little systematic relation to firm outcomes; it is only when culture is measured through channels that capture enacted values, such as employee perceptions, that meaningful associations with performance, productivity, and ethical behavior emerge. This finding has a direct methodological implication: data sources that are more likely to capture enacted rather than espoused values will yield more valid measures of organizational culture.

The Q&A portion of earnings calls is particularly well suited for capturing enacted culture because its spontaneous, interactive format limits the opportunity for strategic impression management. Unlike prepared remarks, which are drafted, reviewed, and polished in advance, Q&A responses arise in real time as executives field unanticipated questions from analysts (Li et al., 2021; Price et al., 2012). Analysts frequently probe areas of concern, challenge management narratives, and introduce topics that executives would not voluntarily address, forcing responses that draw on the cognitive frameworks, priorities, and normative assumptions that are deeply embedded in organizational experience. When a CEO responds to a pointed question about a strategic setback, the language, framing, and emphasis in that response are more likely to reflect the values that actually guide decision-making than a carefully curated paragraph in an annual

report. Li et al. (2021), in their influential study on corporate culture measurement, explicitly chose the Q&A portion of earnings calls over prepared remarks and 10-K filings on precisely this basis, arguing that the Q&A portion best reveals the values of the organization and provides the most authentic window into the cultural norms and priorities that guide executive behavior.

Empirical research consistently supports the view that the Q&A portion carries information content that is distinct from and often superior to that of scripted communications. Price et al. (2012) found that spontaneous Q&A sentiment predicts short-term abnormal returns more reliably than the tone of the scripted management presentation, indicating that the market recognizes the incremental information value of unscripted managerial discourse. Larcker and Zakolyukina (2012) linked linguistic cues from the Q&A portion to subsequent financial restatements, demonstrating that the way executives respond to probing questions under time pressure can reveal deceptive intent that is effectively concealed in prepared statements. Mayew and Venkatachalam (2012) extended this line of evidence beyond text to vocal properties, showing that managerial affect conveyed through voice during Q&A responses carries information that predicts future firm performance. Together, these findings establish that the Q&A portion provides a uniquely revealing window into managerial cognition, and the values that underlie managerial cognition are precisely the manifestations of organizational culture that this study seeks to measure.

Against the advantages of Q&A discourse, the limitations of formal filings as a data source for cultural measurement come into sharp relief. The predominant alternative in the CVF literature has been 10-K reports (Fang et al., 2023; Fiordelisi & Ricci, 2014), but these documents are subject to several sources of measurement error that collectively undermine their suitability for capturing authentic organizational culture. First, 10-K reports are static, backward-

looking documents that undergo extensive review and editing by legal counsel and public relations professionals. The language that appears in a 10-K is the product of a drafting process that prioritizes regulatory compliance, legal risk minimization, and strategic self-presentation, not the authentic expression of cultural values. As Guiso et al. (2015) observed, the cultural signals in such documents may reflect an idealized or “advertised” culture rather than the operational reality and values enacted in daily practice.

Second, and relatedly, the textual approach to measuring culture from 10-K reports rests on the assumption that organizational culture, which develops gradually over time, determines word choice and expressions in written filings (Fang et al., 2023). However, this assumption is problematic in several respects. Such an approach captures corporate culture only to the extent that cultural traits can manifest themselves in the specific genre of formal written reports. If certain cultural attributes are inherently difficult to express in the structured, compliance-oriented prose of a 10-K, proxy variables derived from these documents will be subject to measurement errors and will never fully capture the underlying cultural construct (Fang et al., 2023). This is particularly concerning for culture types such as Clan, which emphasizes relational values like mentorship, psychological safety, and participative decision-making. These values are enacted through interpersonal interactions and informal organizational processes that may leave little trace in the formal, legalistic language of annual filings.

Third, substantial portions of 10-K reports are drafted or heavily influenced by firms’ in-house legal departments, and the extent and nature of this legal influence varies systematically across companies (Fang et al., 2023). If legal departments differ in their stylistic conventions, risk tolerance, and editorial practices, the word choice and expressions in 10-K filings will be shaped by this heterogeneity in ways that are unrelated to the firm’s actual culture. This

introduces a source of measurement noise that is particularly insidious because it is systematic rather than random: firms with more conservative legal departments may appear to have different cultural profiles simply because of differences in drafting conventions rather than differences in underlying values. The culture-related synonyms extracted from these reports may thus capture the preferences of legal reviewers as much as the values of the organization's operational leadership.

Fourth, the temporal properties of 10-K reports limit their usefulness for studying cultural dynamics. These documents are filed annually and describe the preceding fiscal year, providing at best a single data point per firm per year with a backward-looking orientation. Culture, however, is not static. It responds to leadership transitions, strategic pivots, competitive shocks, and organizational crises. Earnings calls, by contrast, occur quarterly, providing a panel structure with four observations per firm per year and a more contemporaneous window into the cultural orientations that executives bring to their strategic discussions. This temporal granularity enables researchers to track how cultural emphasis shifts in response to changing conditions and to examine whether firms' cultural profiles exhibit the stability over time that would be expected if the measure is capturing a genuine organizational attribute rather than transient linguistic variation.

Earnings call Q&A sessions also address limitations inherent in alternative data sources for cultural measurement. Surveys, while providing valuable insider perspectives, are limited by impression management, self-report bias, and reproducibility concerns (Graham et al., 2022; Keiser & Payne, 2019). The standard survey-based operationalization of the CVF, Cameron and Quinn's (2006) Organizational Culture Assessment Instrument, requires respondents to allocate points across competing cultural descriptions, a method that yields rich individual-level data but

is impractical for large-scale longitudinal research spanning thousands of firms over many years. Employee reviews on platforms such as Glassdoor provide large-scale data but suffer from selection bias, inconsistent coverage across firms and time periods, and the absence of the senior management perspective that sets the “tone at the top” (Ji et al., 2017; Li et al., 2021). Earnings call Q&A sessions uniquely combine the advantages of spontaneity, coverage, regularity, and executive-level perspective: they provide unscripted discourse from senior leadership across a large cross-section of publicly traded firms on a consistent quarterly basis.

Finally, the conversational nature of Q&A discourse provides a richer linguistic environment for LLM-based cultural analysis than any of the alternative data sources discussed above. When executives elaborate on topics under questioning, they produce extended narrative segments in which cultural values are not merely mentioned but articulated, justified, and connected to strategic decisions. A CEO explaining why the organization invests heavily in employee development, or why the firm tolerates short-term losses in pursuit of breakthrough innovation, generates discourse that is qualitatively different from a keyword appearing in a boilerplate filing. This discursive richness is precisely what an LLM-based approach can exploit: rather than counting isolated words, a contextual classifier can evaluate whether an extended segment of Q&A discourse genuinely reflects a cultural orientation such as collaboration, innovation, process control, or competitive intensity. The combination of spontaneity, executive-level perspective, temporal regularity, and discursive depth makes earnings call Q&A the most suitable data source for a contextual, LLM-driven measurement of organizational culture through the CVF.

2.3.2 From Manual to AI-Driven Analysis of Earnings Call Discourse

The textual analysis of earnings calls has progressed through several methodological generations. Early research established that the linguistic properties of earnings calls carry market-relevant information. Frankel et al. (1999) showed that earnings calls affect stock prices, laying the groundwork for viewing these events as information-rich. Subsequent studies applied dictionary-based methods to quantify the tone and content of call transcripts. Price et al. (2012) used sentiment dictionaries to demonstrate that Q&A tone predicts abnormal returns, while Bochkay et al. (2020) examined how the language characteristics of earnings calls influence investor processing of information. These studies relied on lexicons such as the Loughran and McDonald (2011) finance-specific dictionaries, which improved upon general-purpose alternatives but remained limited by their inability to interpret context.

A second generation of studies introduced more sophisticated textual analysis techniques. Li et al. (2021) combined dictionary-based methods with Word2Vec embeddings to measure corporate culture from earnings call Q&A sessions. Their approach used Word2Vec to identify words that are semantically proximate to seed terms representing five cultural values (including innovation, integrity, quality, respect, and teamwork), thereby expanding the dictionary in a data-driven manner. While this semi-supervised approach represented an important methodological advance over purely manual dictionary construction, it remained fundamentally anchored to word-level analysis. The resulting measure counts the frequency of dictionary terms in the Q&A portion, without evaluating whether each occurrence actually reflects the cultural construct it is intended to represent.

More recently, AI-driven techniques have opened new frontiers in earnings call analysis. Bai et al. (2023) compared management Q&A responses to AI-generated answers to detect

evasiveness, finding that managerial deviation from AI-predicted responses predicts weaker future performance. Hassan et al. (2019) developed political risk metrics derived from call transcripts. Li et al. (2026) used generative AI to extract cultural insights from analyst reports. de Kok (2025) demonstrated that LLMs can extract economically meaningful signals from management speech patterns that were previously inaccessible to automated methods. These developments illustrate a trajectory from static, decontextualized word counting toward dynamic, context-aware interpretation of corporate discourse, a trajectory that motivates the present study's application of a reasoning LLM to the measurement of organizational culture through the CVF.

2.3.3 Measurement of CVF Culture Using Textual Analysis

The measurement of the CVF in the management and finance literature has traditionally relied on two primary methods: surveys and dictionary-based textual analysis. Survey-based approaches provide direct insight from organizational members but are prone to self-report bias and impression management (Keiser & Payne, 2019), are difficult to scale across large samples of firms, and are typically limited to cross-sectional snapshots that cannot capture cultural dynamics over time (Graham et al., 2022). Cameron and Quinn's (2006) Organizational Culture Assessment Instrument (OCAI), the standard survey-based operationalization of the CVF, requires respondents to allocate points across competing cultural descriptions, a method that yields rich individual-level data but is impractical for large-scale empirical research in finance and accounting.

To achieve scalability, a strand of the literature has turned to dictionary-based textual analysis of corporate filings. Fang et al. (2023) and Fiordelisi and Ricci (2014) operationalized

the CVF by conducting bag-of-words analysis on corporate 10-K reports. Their approach involves creating lists of synonyms, or “bags of words,” corresponding to each of the four cultural dimensions, derived from foundational management texts such as Cameron et al. (2006) and independent lexical resources like the Harvard IV-4 Psychosocial Dictionary. The frequency of these words in a firm’s 10-K report is then used to calculate a score for each cultural dimension. This methodology has been influential, with subsequent studies applying the dictionary-based CVF measure to research in banking (Barth & Mansouri, 2021; Luu et al., 2023), mergers and acquisitions (Hertel et al., 2024), and tax avoidance (Hasan et al., 2024).

Despite the scalability of this approach, the combination of 10-K reports and dictionary-based methods presents three significant limitations that collectively motivate the present study. First, 10-K reports are static, backward-looking documents that undergo extensive review and editing by legal counsel and public relations professionals. The language used may reflect an idealized or “advertised” culture rather than the operational reality and values enacted in daily practice (Guiso et al., 2015). Second, dictionary-based methods are inherently context-insensitive. They cannot distinguish between a substantive discussion of a cultural value and a superficial or incidental mention of a keyword, leading to a high potential for false positives and measurement noise. For instance, the word “team” might appear in a 10-K in the context of “management team” or “legal team” without reflecting a collaborative, clan-type culture, yet dictionary-based methods would count this occurrence as evidence of clan culture. Similarly, “efficiency” might appear in the phrase “capital efficiency ratio” as a financial metric rather than as a cultural value emphasizing process discipline and standardization. Third, and relatedly, these methods implicitly assume that the mere presence of a word from a cultural lexicon in a corporate text constitutes evidence of the underlying cultural orientation it is intended to

represent. This assumption is problematic for cultural constructs, which are inherently complex and context-dependent. As documented in the previous essay on corporate integrity, when dictionary terms are subjected to contextual evaluation by a reasoning LLM, fewer than 30% of the segments containing integrity dictionary terms are identified as meaningfully related to the cultural construct. If a similar pattern holds for CVF dimensions, then existing dictionary-based CVF measures may be recovering something other than the cultural constructs researchers intended, potentially understating or mischaracterizing the economic significance of organizational culture.

This concern elevates the measurement problem from a purely methodological one to a phenomenon-level question: does a genuinely contextual approach to measuring organizational culture through the CVF lead to a different, and more empirically meaningful, understanding of culture and its implications for firm behavior? An LLM-driven, contextually grounded CVF measure is uniquely suited to investigate this question because it can evaluate whether the linguistic markers that existing approaches rely on actually co-occur with meaningful cultural content. Yet there is no systematic application of LLMs to the measurement of organizational culture through the CVF.

2.4 Research Gap and Motivation

The preceding review reveals a critical gap at the intersection of IS capabilities, organizational culture measurement, and the application of generative AI. While prior research has made significant strides in analyzing corporate communications and developing culture measures, several limitations persist that together motivate the present study.

First, there is a construct-method mismatch. The CVF is a nuanced theoretical model that captures the competing tensions inherent in organizational values, but its empirical measurement has not kept pace with methodological advances in NLP. The predominant use of dictionary-based methods (Fang et al., 2023; Fiordelisi & Ricci, 2014) fails to capture the contextual subtleties essential for understanding cultural discourse. The mere presence of a keyword like “innovate” is insufficient to classify a statement as reflecting an Adhocracy culture; the context in which the word is used, whether it refers to genuine cultural values or merely describes operational activities, is paramount. As the integrity essay demonstrates, the vast majority of dictionary keyword occurrences appear in contexts that are unrelated to the construct the dictionary is designed to measure. If this pattern generalizes to the CVF, then existing dictionary-based measures are likely subject to substantial measurement noise that attenuates their ability to detect meaningful cultural variation across firms.

Second, data source limitation exists. The application of CVF measurement has been overwhelmingly concentrated on static, formal documents such as 10-K reports, which reflect espoused rather than enacted culture. This neglects more dynamic and spontaneous data sources such as earnings call Q&A sessions, which prior research has shown to be more revealing of the authentic cultural orientations demonstrated by leadership under pressure (Guiso et al., 2015; Li et al., 2021; Larcker & Zakolyukina, 2012; Price et al., 2012). The shift from the formal script of the 10-K to the live, interactive discourse of the Q&A session represents a methodological choice that directly addresses the espoused-versus-enacted culture distinction that Guiso et al. (2015) identified as central to the study of corporate values.

Third, there is a technological opportunity. The emergence of powerful, reasoning-capable LLMs offers a novel path to overcome the context-insensitivity of dictionary methods

and to analyze the complex language of earnings calls at scale. These models can interpret context, infer speaker intent, and provide justifications for their classifications, enabling a more accurate and reliable assessment of intangible constructs like culture. Unlike dictionary-based or even Word2Vec-enhanced approaches, an LLM can evaluate extended discourse segments holistically, determining whether a passage reflects genuine cultural content or merely contains incidental mentions of culture-related keywords.

Fourth, there is a dimensionality gap. Existing CVF measures based on dictionary approaches treat each cultural dimension independently, computing separate word frequency counts for each quadrant. They do not address the possibility that the same textual segment may contain signals relevant to multiple quadrants, nor do they model the relative intensity of cultural discourse across the four dimensions within a given firm-quarter. A multi-agent LLM architecture, in which specialized agents evaluate each CVF dimension with domain-specific expertise, can address this gap by producing both aggregate cultural profiles and component-level scores that capture the distribution of cultural emphasis across the four quadrants.

This research is motivated by the need to bridge these gaps. By applying a reasoning LLM methodology within a multi-agent classification architecture to measure the well-established CVF from earnings call Q&A transcripts, this study aims to develop a more sophisticated, reliable, and dynamic measure of organizational culture. The proposed approach creates a hybrid methodology that acknowledges the value of established constructs for organizing cultural analysis while leveraging the LLM to overcome the fundamental weakness of existing operationalizations: the inability to distinguish substantive cultural discourse from incidental keyword occurrence. This approach extends the methodological innovations developed in the companion essay on corporate integrity to a complementary but distinct cultural

framework, recognizing that while integrity represents an ethical dimension, the CVF provides a more comprehensive map of the competing cultural tensions that shape organizational effectiveness.

3. Methodology

This section describes the systematic approach used to collect, process, and analyze textual data extracted from corporate earnings call transcripts to construct Contextual CVF Culture Scores. Figure 3 presents a process model of the entire pipeline, from data collection to the construction of organizational culture measures. Our methodology moves beyond traditional keyword-based analyses by leveraging a multi-agent large language model (LLM) architecture to capture the context-dependent nature of organizational culture discourse. By adapting the blackboard-based multi-agent classification system developed and validated in the companion essay on corporate integrity, this study applies the same architectural principles to the distinct conceptual domain of the Competing Values Framework (CVF), thereby enabling contextual measurement of the four CVF cultural orientations (i.e., Clan, Adhocracy, Hierarchy, and Market) from earnings call transcripts.

[Insert Figure 3]

3.1 Data Collection and Sample Definition

Following established practices in the literature (Allee & Deangelis, 2015; Bochkay et al., 2020), we sourced full transcripts of quarterly earnings conference calls spanning the period from 2008 to 2024 (fiscal year). Each transcript file contains the complete body of the earnings call

conversation along with metadata that enables matching with the COMPUSTAT database and facilitates systematic data organization for subsequent steps. These metadata elements include the company ticker symbol, company name, transcript title, date, and transcript identifier. Following data matching with COMPUSTAT, our earnings call dataset consists of 183,043 earnings call transcripts. We then extracted the Q&A sections from these transcripts, as this section contains particularly valuable information on organizational culture (Li et al., 2021). The Q&A section is especially relevant for CVF measurement because it captures spontaneous, less-rehearsed executive discourse in which cultural orientations are more likely to surface authentically than in the scripted presentation portions of the call (Larcker & Zakolyukina, 2012). To ensure our analysis focused specifically on corporate leadership responses, we filtered the Q&A content using regex-based parsing in Python. This was feasible because the transcripts exhibit a relatively consistent structure, with speaker turns identified by standardized names, roles, and affiliations, and the Q&A section separated from the prepared remarks by identifiable section headers. We used these recurring patterns to (i) identify the Q&A section, (ii) divide the dialogue into individual speaker turns, and (iii) distinguish company executives from external analysts based on their role and affiliation. Analyst questions were excluded, while responses from CEOs, CFOs, and other senior company executives were retained. We manually inspected a random subset of transcripts to validate the extraction and refined the parsing rules to accommodate formatting differences across transcript vendors and years.

Filtering out analyst questions serves multiple purposes. First, it ensures that our analysis focuses exclusively on corporate leadership communication, which most directly reflects organizational values and cultural orientations (Guiso et al., 2015; Li et al., 2021). Senior executives are the primary architects and carriers of organizational culture, and their spontaneous

responses during Q&A sessions offer a window into the cultural values that shape their strategic thinking and decision-making. Second, it eliminates potential noise from analyst questions that might contain culture-related terminology without reflecting the company's actual cultural stance. An analyst asking about competitive positioning, for example, does not indicate a Market culture orientation within the firm. Third, it creates a more homogeneous corpus of executive communication, ensuring that the textual data subjected to classification reflects a consistent source of cultural signals. This filtered corpus constitutes our fiscal-quarter matched sample of public firms for the study period and lays the foundation for subsequent steps.

3.2 Text Preprocessing and Data Preparation

Our data preparation follows a segmentation approach to identify coherent text segments in earnings call transcripts. The transcript content used in this process is the same as described in Section 3.1 and will be referred to as earnings calls throughout the remainder of the manuscript.

We segment Q&A sections into coherent chunks through a three-step process. In the first step, we decompose the relevant content of each Q&A section into individual sentences using the Python library `spaCy` (Dutta et al., 2023). Sentence-level decomposition is a necessary prerequisite for the subsequent embedding and segmentation stages, as it establishes the atomic units upon which semantic similarity computations are performed. In the second step, we convert each sentence into high-dimensional vector representations using the BGE-M3 embedding model, recognized for producing semantically meaningful sentence representations suitable for similarity analysis (Chen et al., 2024). These dense embeddings capture the semantic content of each sentence in a continuous vector space, enabling the identification of topical coherence across adjacent sentences. The third step implements a semantically aware adaptation of the C99

algorithm (Choi, 2000) for text segmentation. Rather than relying on traditional bag-of-words representations (TF-IDF) employed in the original C99 implementation, our approach leverages the dense sentence embeddings from the BGE-M3 model to compute cosine similarity between sentences. The C99 algorithm identifies coherent text segments by calculating similarity scores between sentence pairs and using these scores to determine optimal segment boundary placement (Choi, 2000; Li et al., 2026). Intuitively, this approach groups adjacent sentences that discuss similar concepts, effectively reconstructing the document’s logical, topically cohesive, and coherent paragraph-like structure (Li et al., 2026).

The choice of segmentation granularity is particularly important for CVF measurement. Organizational culture is rarely expressed through isolated sentences; rather, cultural orientations emerge across multi-sentence passages in which executives elaborate on values, strategic priorities, and organizational practices. A CEO describing collaborative decision-making processes, for example, may develop this theme across several adjacent sentences that collectively signal a Clan orientation. By grouping such semantically related sentences into coherent segments, the segmentation procedure preserves the discursive context that is essential for accurate cultural classification. However, as established in the first essay of this dissertation, replacing the TF-IDF sentence representations in C99 with BGE-M3 dense embeddings does not alter transcript-level culture measurements relative to the TF-IDF-based C99 baseline. The detailed evaluation, based on paired-sample t-tests across three independent samples of 1,000 earnings calls each, is presented in Appendix A, Supplementary Discussion SD1. The segmentation process yields a corpus comprising 10,277,650 text segments distributed across 183,043 earnings call transcripts, which provides the foundation for our LLM-based classification pipeline.

3.3 LLM-Based Classification

3.3.1 *Model Selection*

Following text segmentation, we organized each segment into a structured JSON format, including the text and metadata (e.g., transcript ID, year, company name). Selecting an appropriate LLM for classification involved balancing performance, consistency, accessibility, and cost. Recent studies have raised concerns about using LLMs for academic research: as noted by Chen et al. (2023) and de Kok (2025), LLMs can produce inconsistent outputs across models, versions, and even repeated runs with the same prompt, posing challenges for reproducibility. Chen et al. (2023) documented major output differences between GPT-4's March and June 2023 versions, highlighting temporal instability, and de Kok (2025) cautions against relying on proprietary model APIs that can be modified or discontinued without notice. OpenAI's GPT-4o update in early 2025, for example, significantly altered model behavior, including increased sycophancy (OpenAI, 2025). To address reproducibility concerns, we followed de Kok's (2025) recommendation to avoid third-party APIs by deploying open-source models on local or cloud infrastructure (e.g., RunPod, VastAI, AWS). Open-source models provide full control over inference, allow rapid iteration, avoid API limits, and prevent disruptions from unexpected updates. They are also more cost-efficient without major quality trade-offs (Irugalbandara et al., 2023).

Building on these advantages, our next step was to identify an open-source LLM capable of reliably classifying organizational culture discourse along the four CVF dimensions. We adopted a benchmark-driven model selection strategy, evaluating candidate models on two benchmarks relevant to our task: the Multistep Soft Reasoning (MuSR) benchmark (Sprague et

al., 2023) and the BigBench Hard (BBH) benchmark (Suzgun et al., 2022). MuSR evaluates a model’s language understanding and multistep reasoning capabilities in complex textual settings, while BBH assesses common-sense reasoning and general world knowledge. Given that our classification task requires contextual interpretation and multi-step reasoning over extended discourse segments drawn from earnings call transcripts, MuSR is particularly relevant as a selection criterion. The CVF classification task is arguably more demanding than a binary construct classification because it requires the model to distinguish among four competing cultural orientations that may co-occur within the same discourse segment, necessitating nuanced reasoning about which cultural archetype is most prominently expressed. Benchmark performance for all candidate models is reported in Appendix B Table B1.

Among the evaluated models, DeepSeek-R1-Distill-Qwen-14B (hereafter DeepSeek-R1) demonstrated the strongest overall performance and, notably, the highest MuSR score, making it the most suitable model for our classification task¹. Beyond benchmark performance, our selection was further motivated by DeepSeek-R1’s reasoning capabilities: it belongs to a class of large language models explicitly designed to engage in structured, multistep reasoning before generating responses (Guo et al., 2025). This reasoning ability adds an interpretability layer to text classification (Huang & Chang, 2022; Wang et al., 2022), as the model can articulate the basis for its decisions. For CVF classification specifically, this reasoning transparency is valuable because it allows researchers to verify whether the model’s cultural categorizations are grounded in theoretically appropriate cues rather than superficial lexical patterns. The model’s suitability is further validated through classification performance on a human-annotated validation set, as detailed in Section 4.1.

¹ Several other candidate models appeared to perform competitively when evaluated on the full imbalanced corpus, but their performance dropped substantially on a balanced validation set, suggesting that their apparent strength reflected exploitation of the severe class imbalance rather than genuine discriminative capability.

3.3.2 Multi-Agent Classification Architecture

A key design challenge in LLM-based classification of complex constructs such as organizational culture is the tension between specialization and coherence. Multi-agent architectures can assign domain-specific experts to individual facets of a construct but risk fragmenting the broader context when agents operate in isolation. Single-agent monolithic designs preserve global coherence but sacrifice the granularity required to reason over distinct conceptual dimensions. Wang et al. (2024) characterize this tension through the concept of cognitive synergy, whereby the collective reasoning capacity of collaborating agents can exceed the sum of their individual contributions, provided appropriate coordination mechanisms are in place. Achieving this synergy, however, demands an architecture that enables agents to share information without sacrificing their domain-level specialization.

This tension is particularly acute for CVF classification. The four cultural quadrants are defined relationally: each acquires meaning partly through its contrast with the others (Cameron & Quinn, 2006). A passage about efficiency and standardization, for instance, may signal Hierarchy culture, but the same passage viewed in the context of competitive benchmarking may instead reflect a Market orientation. Agents specialized in individual quadrants must therefore be aware of one another's assessments to produce accurate classifications. To address this trade-off, we draw on the blackboard architecture (Salemi et al., 2025), a well-established framework for cooperative problem-solving in which independent knowledge sources contribute to a shared information space under the supervision of a central controller. We adapt its core principle of indirect coordination through shared memory to implement a multi-agent classification system in which specialized agents collaborate through a shared memory structure (the Blackboard) that evolves across four sequential processing phases. The full architecture is illustrated in Figure 4.

[Insert Figure 4]

The classification of each text segment proceeds as follows. In Phase I (Culture Detection), a dedicated Organizational Culture Agent evaluates whether the segment contains any signal of culture-related organizational discourse, performing a binary classification (Yes/No). This agent is designed to identify passages where executives discuss values, norms, practices, leadership approaches, strategic orientations, or organizational characteristics that reflect underlying cultural assumptions, rather than purely financial or operational content devoid of cultural signal. This agent operates as an execution gate: segments classified as “No” receive their final label and processing terminates, while segments classified as “Yes” are forwarded to the next reasoning stages. The gating mechanism serves an important practical function: given that the vast majority of earnings call discourse addresses financial results, operational metrics, and forward guidance rather than cultural themes, the Phase I gate ensures that computational resources in the subsequent multi-agent phases are directed exclusively toward segments that contain genuine cultural content. For segments that pass the gate, the system initializes a shared memory space (the Blackboard) and proceeds to Phase II.

In Phase II (Observation Pass), four domain-specific agents analyze the segment in parallel, each aligned with one dimension of the Competing Values Framework: Clan, Adhocracy, Hierarchy, and Market. Drawing from the construct conceptualizations formalized in the theoretical framework (Section 2.2), each agent’s prompt embeds the relevant definitions along with chain-of-thought reasoning instructions. The Clan agent, for instance, is prompted with the theoretical characteristics of collaborative culture—teamwork, mentorship, employee development, participative decision-making, and cohesion—drawn from Cameron and Quinn (2006) and Hartnell et al. (2011). Similarly, the Adhocracy agent is attuned to signals of

innovation, risk-taking, entrepreneurial leadership, and adaptive flexibility; the Hierarchy agent to formal procedures, efficiency, standardization, and control mechanisms; and the Market agent to competitive positioning, goal achievement, market share, and results orientation. In this phase, agents operate under strict constraint: they are prohibited from issuing judgments and must produce only objective, factual observations about the segment as it pertains to their respective CVF dimension. These observations are written to the shared memory space, resulting in an updated state we refer to as Blackboard v_1 , forming a shared fact base. By enforcing a clean separation between evidence gathering and interpretation, this phase reduces the risk of premature or biased judgments and anchors all subsequent reasoning in explicitly extracted evidence.

The observation-only constraint in Phase II is particularly important for CVF classification because of the inherent ambiguity in cultural discourse. Many terms and phrases used in earnings calls can plausibly signal more than one cultural orientation. References to “efficiency,” for example, might reflect a Hierarchy emphasis on process control or a Market emphasis on competitive cost advantages. By requiring agents to record observations without rendering verdicts, the architecture ensures that such ambiguous signals are preserved for cross-agent deliberation in the subsequent phase rather than being prematurely resolved by a single domain agent.

In Phase III (Verdict Pass), the same four agents are reactivated with read access to the complete Blackboard v_1 , enabling cross-agent awareness. Each agent now produces contextualized verdicts and reasoning, informed not only by its own prior observations but also by the factual evidence contributed by peer agents. This shared-state design realizes the cognitive synergy described by Wang et al. (2024): each agent retains its domain expertise while its

reasoning is enriched by peer observations, achieving collective awareness without sacrificing modularity. For example, the Market agent may revise its assessment of a competitive discourse segment upon reading the Clan agent's observation that the same passage emphasizes team-based achievement rather than individual performance metrics, thereby shifting the classification toward Clan. The outputs from this phase are written back to the shared memory space, forming Blackboard v_2 , which layers these reasoned verdicts atop the Blackboard v_1 fact base, producing a fully contextualized system state. The indirect coordination enabled by the blackboard aligns with the asynchronous collaboration paradigm described by Salemi et al. (2025).

In Phase IV (Synthesis), a Synthesizer Agent consolidates all outputs from Blackboard v_2 into a final integrated classification. Where clear cultural categories are detected, it produces the final output directly. In cases where ambiguity persists across agent verdicts—a situation that is more common in CVF classification than in binary construct detection because competing cultural values frequently co-occur within the same organizational discourse—a Closest-Fit Resolution Agent is invoked to assign the single most appropriate category based on the available evidence. This resolution agent weighs the strength and specificity of the evidence contributed by each domain agent, giving precedence to the cultural orientation for which the most concrete and contextually grounded evidence has been assembled across the observation and verdict phases. Illustrative examples of segments assigned to each of the four CVF components are reported in Appendix B Table B2.

In total, our classification pipeline engages eleven agent roles across the four phases: one Organizational Culture Agent (Phase I), four domain agents in observation mode (Phase II), four agents in reasoning and verdict mode (Phase III), and two adjudication agents (Phase IV). The architecture offers several properties that strengthen classification quality. The strict separation of

observation (Phase II) from judgment (Phase III) reduces hallucination risk by grounding verdicts in a shared fact base rather than unconstrained generation. The multi-phase, shared-memory design enables more calibrated classifications that resemble committee-based deliberation rather than single-pass binary decisions. Finally, the blackboard provides a persistent reasoning trace across all phases, constituting a verifiable audit trail that supports the interpretability and credibility of each classification decision. This architectural transparency is especially valuable for organizational culture measurement, where the subjective and contested nature of cultural constructs demands that classification decisions be traceable to explicit textual evidence and articulated reasoning.

3.3.3 Addressing LLM Concerns

Building on the challenges identified by Li et al. (2026), we address two critical concerns associated with the use of generative LLMs: look-ahead bias and hallucination.

Look-ahead bias arises when a model's output is influenced by information that would not have been available at the time of analysis. This is particularly relevant for LLMs trained on large-scale corpora that may include retrospective or forward-looking knowledge, potentially compromising temporal integrity. In our application, however, the model is not employed for predictive forecasting. Instead, it evaluates whether a given text segment reflects organizational culture based solely on the information contained within that segment. This backward-looking, text-contained design inherently reduces the risk of look-ahead bias. The concern is worth addressing explicitly for CVF classification because cultural labels such as “innovative” or “bureaucratic” are sometimes used retrospectively to characterize firms after major events (e.g., describing a firm's culture as “rigid” after a strategic failure), and an LLM with access to such

retrospective characterizations in its training data could potentially allow post-hoc knowledge to influence its segment-level classifications.

To further safeguard against unintended information leakage or overfitting to specific entities, we follow the method used by Jha et al. (2023) and Li et al. (2026) and conduct an entity-masking test. Using the spaCy NLP library, we anonymized entities in the text: company names were replaced with “[ORGANIZATION],” individual names with “[PERSON],” and product names with “[PRODUCT].” On a randomly selected subset of 2,000 text segments, we compared the model’s classification decisions on unmasked versus masked texts. The resulting Cohen’s kappa coefficient was approximately 0.85, indicating a high level of agreement. This suggests that our results are not primarily driven by entity-specific cues or associations the model may have learned across time, thereby reinforcing the robustness of our classification approach. The high agreement is particularly reassuring for CVF measurement because certain firms are widely associated with specific cultural archetypes in popular business discourse (e.g., Google with innovation, Toyota with process control), and the entity-masking test confirms that the model’s classifications are driven by the content of the discourse rather than by pre-existing associations with firm identities.

Hallucination occurs when models produce content not grounded in the input text (Maynez et al., 2020). Our architecture addresses this concern through multiple safeguards. First, the multi-phase protocol enforces a strict separation between observation and reasoning, ensuring that agent verdicts are anchored in explicitly extracted evidence rather than unconstrained generation. Second, the blackboard structure provides an auditable chain of reasoning, making it possible to trace any classification back to the factual observations that support it. Third, structured chain-of-thought prompting guides logical reasoning and minimizes unsubstantiated

outputs (Li et al., 2026; Wei et al., 2022). The prompts direct agents to analyze the full context of each segment, identify relevant details, and explain their reasoning with explicit reference to the source text. Together, these mechanisms reduce the risk of hallucinated or unfounded classifications. The risk of hallucination is particularly salient for cultural classification because culture is an inherently interpretive construct; without the disciplining mechanisms embedded in our architecture, an LLM might “see” cultural signals where none exist, imposing cultural interpretations on mundane operational discourse. The observation-first protocol and the requirement for textual grounding serve as important guardrails against this tendency.

3.4 Contextual CVF Culture Score Construction

After classifying each text segment through the multi-agent pipeline, we construct Contextual CVF Culture Scores for each company-quarter. These scores operationalize the four CVF dimensions as continuous measures derived from the frequency and distribution of classified discourse segments within each earnings call transcript. For each earnings call, we compute the following indicators.

(i) The Culture-Related Segment Count represents the total number of segments within the transcript that the pipeline classified as relevant to organizational culture, serving as a direct measure of the frequency of culture-related discourse during the call. This aggregate count captures the overall salience of cultural themes in executive communication, reflecting the extent to which organizational values and cultural orientations permeate leadership discourse during earnings calls.

(ii) Component-Specific Segment Counts, based on the final category assignments produced by the Phase IV synthesis stage, capture the number of segments classified under each of the four

CVF cultural dimensions: Clan, Adhocracy, Hierarchy, and Market. Each count reflects how often a specific cultural orientation is identified in the classified segments, enabling granular analysis of cultural emphasis. For instance, a firm whose earnings call transcripts consistently yield high Adhocracy segment counts and low Hierarchy counts would be characterized as exhibiting a cultural profile skewed toward innovation and flexibility over formal control and standardization. The distribution of these components across culture-related segments is reported in Table 1, with accompanying illustrative examples provided in Appendix B Table B2.

[Insert Table 1]

Using these segment-level outputs, we derive both an overall culture score and its component-specific scores at the firm-quarter level. The component-specific scores enable researchers to construct cultural profiles that capture not only the intensity of cultural discourse but also its distributional composition across the four CVF quadrants. This distributional perspective is theoretically grounded in the CVF’s foundational insight that organizational culture is defined by the relative emphasis placed on competing values rather than by the absolute level of any single cultural dimension (Cameron & Quinn, 2006; Quinn & Rohrbaugh, 1983). Relevant variables are described in Table A1 provided in Appendix A. These measures serve as inputs to the empirical validation analyses reported in Section 4.3, including associations with theoretically predicted firm-level outcomes for each CVF dimension.

4. Classification Performance

To establish the reliability and effectiveness of our LLM-based classification pipeline, we adopt a two-pronged validation strategy that combines technical performance assessment with

empirical validation. First, we evaluate the classification performance of the Phase I CVF detection gate, the component that determines which segments enter the downstream reasoning stages. We focus validation effort on this gate because it sits at a critical junction where any misclassification would propagate through all subsequent stages of the pipeline. Second, we validate our contextual CVF measure by examining its empirical associations with well-established indicators of governance quality, organizational effectiveness, and strategic orientation.

4.1 Overall CVF Detection Performance and Validation

To establish the reliability of the Phase I CVF detection gate in identifying overall organizational culture discourse in earnings call transcripts, we conducted validation tests across multiple dimensions. Our validation strategy focused on three areas: (i) standard classification metrics benchmarked against manually labeled data, (ii) prediction stability across multiple runs to ensure robustness, and (iii) comparative performance relative to dictionary-based approaches.

4.1.1 Validation Design for the Phase I CVF Detection Gate

Following the evaluation framework outlined by Demers et al. (2024) and Li et al. (2026), we constructed a manually annotated validation set to assess the Phase I CVF detection gate. A total of 300 text segments were drawn and subsequently hand-coded to serve as ground-truth labels. The sample was generated using a stratified sampling design based on the Phase I predicted labels, comprising 150 segments classified as “Yes” (CVF-related), and 150 segments classified as “No” (not CVF-related) by the gate. The rationale for this stratified design is twofold. First, the raw corpus exhibits severe class imbalance, with culture-positive segments constituting a

relatively small share of the total. A purely random sample of moderate size would yield very few positive instances, thereby limiting the ability to rigorously evaluate performance on the minority class. Second, sampling 150 segments from each category increases the likelihood that both classes are adequately represented in the manually coded validation set relative to a purely random sample. This enables a meaningful assessment of precision, recall, F1 score, Matthews correlation coefficient (MCC), and overall classification quality.

4.1.2 Classification Performance of the Phase I CVF Detection Gate

The validation results indicate consistently strong performance across all evaluation metrics. Table 2 presents the performance of the Phase I gate for the main classification run² alongside five independent stability runs, all evaluated against the human-annotated validation set of 300 samples.

[Insert Table 2]

The main run achieved an accuracy of 96.00%, with macro-averaged precision and recall at 96.00% and 96.30%, respectively. The Cohen's kappa coefficient of 0.920 indicates near-perfect agreement between Phase I predictions and human annotations. According to established benchmarks, kappa values above 0.80 are considered to reflect strong agreement, while values above 0.90 approach near-perfect concordance (Cohen, 1960; Landis & Koch, 1977). The Matthews correlation coefficient of 0.923 corroborates this assessment, providing a balanced measure that accounts for all four cells of the confusion matrix.

[Insert Table 3a and 3b]

² The main classification run refers to the execution that labels all segments in the full corpus, from which the firm-quarter CVF measures used in the empirical analyses are ultimately constructed; the five stability runs, by contrast, were conducted exclusively on the validation set in order to assess run-to-run reproducibility.

Per-class metrics (Table 3a) reveal strong and balanced performance across both categories. In the main run, the gate achieves perfect recall for CVF-positive segments (100%) and perfect precision for CVF-negative segments (100%), indicating that the model successfully identifies all genuine instances of cultural discourse while exercising strict discipline in avoiding false positives among non-cultural segments. Notably, the gate produces no false negatives in the main run, meaning that every segment manually coded as CVF-related by human annotators was correctly identified by the pipeline. The confusion matrix for the main run (Table 3b) shows only 12 misclassifications out of 300 segments, all of which are false positives (non-CVF segments incorrectly classified as CVF-related), confirming the practical reliability of the gate. Detailed confusion matrices for all runs are provided in Appendix B Table B3.

4.1.3 Prediction Stability of the Phase I CVF Detection Gate

To ensure the robustness and reproducibility of the Phase I gate, we conducted five independent runs of DeepSeek-R1 on the validation set using the same parameters and prompts. This stability analysis is important for establishing confidence in the reproducibility of results, particularly given the stochastic nature of LLM inference.

As shown in Table 2, the five stability runs yield accuracy scores ranging from 94.00% to 95.67%, with macro-averaged F1 scores clustered tightly between 93.99% and 95.65% (SD = 0.62%). Cohen’s kappa values range from 0.880 to 0.913 (mean = 0.896, SD = 0.012), all falling within the “almost perfect” agreement range. These results confirm consistent behavior of the Phase I gate across independent replications, directly addressing the concerns raised by de Kok (2025) regarding the tendency of LLMs to produce varying outputs from identical inputs.

The inter-run agreement analysis, summarized in Table 4, further validates the gate's reproducibility. All pairwise Cohen's kappa scores among the five runs are at or above 0.900, with a mean of 0.931 and a standard deviation of 0.020 at the pairwise level. As subset size increases from $k = 2$ to $k = 4$, the mean kappa remains stable at 0.931 while the standard deviation decreases (from 0.020 at $k = 2$ to 0.007 at $k = 4$), reflecting convergent behavior. This level of self-consistency provides strong empirical support for the reliability of the Phase I gate under repeated application.

[Insert Table 4]

Taken together, the classification performance and stability results affirm that the Phase I CVF detection gate delivers consistent, high-quality performance across multiple executions. Because all downstream component classifications are conditioned on a segment passing this gate, these findings offer strong empirical support for the reliability of the pipeline as a whole in culture-related textual analysis. A companion table to Table 4, reporting detailed pairwise kappa values for all run combinations, is provided in Appendix B Table B4.

4.2 Contextual Understanding versus Dictionary-Based Approaches

A central methodological contribution of this study lies in demonstrating the value of contextual understanding over traditional dictionary-based approaches for measuring corporate culture constructs. While prior studies using dictionary methods have provided valuable insights into corporate discourse patterns, our analysis reveals significant limitations in keyword-based approaches that modern LLM capabilities can address.

4.2.1 Beyond Keyword Matching

Our analysis confirms that the mere presence of CVF-related keywords is insufficient for identifying meaningful organizational culture discourse. Table 5 presents illustrative examples in which culture-related terms from the Fang et al. (2023) CVF dictionary appear in contexts that are not meaningfully related to organizational culture yet would be counted as culture signals under a dictionary-based approach. The examples illustrate several distinct patterns of contextual misuse that recur throughout the corpus.

In Example 1, the Clan keyword “collab*” appears in the compound product name “Video Collaboration,” embedded within a discussion of strong financial growth across product categories and a decision to raise guidance. The segment centers on revenue performance and strategic confidence rather than on any collaborative cultural orientation; the keyword’s appearance is incidental to a product label, not indicative of shared values or teamwork. In Example 2, the Clan keywords “partner*” and “team*” surface in a passage about production allocation discussions with Kazatomprom and the relative size of that business relationship. Although “partner” and “team” could signal collaborative culture, here they refer to a joint-venture arrangement and an internal group conducting negotiations, respectively, with no reference to cultural norms of participation, cohesion, or employee development. In Example 3, the Adhocracy keyword “creat*” appears in “create inflation,” a macroeconomic observation about tariff-driven price pressures and Colombia’s trade balance with the United States. The term carries no connotation of innovation, experimentation, or creative risk-taking. In Example 4, the Adhocracy keywords “idea*” and “pioneer*” appear through the discourse marker “ideally” and the proper noun “Pioneer,” the name of an acquired company whose mill assets are being used as loan collateral. Neither usage relates to visionary thinking or entrepreneurial culture. In Example

5, the Market keyword “revenue*” appears in a straightforward operational update on viewership declines and their impact on category-level revenue, reflecting routine financial reporting rather than a cultural emphasis on competitiveness, goal achievement, or market dominance.

In each case, the LLM’s ability to interpret semantic context allows it to distinguish incidental keyword usage from meaningful culture-related discourse, a distinction that dictionary-based approaches, by construction, cannot make.

[Insert Table 5]

4.2.2 Dictionary Word Distribution Analysis

To further demonstrate the limitations of dictionary-based approaches, we examined how frequently words from the CVF wordlist of Fang et al. (2023) appeared across the segments classified by our pipeline. Specifically, we compared their distribution across segments classified as CVF-related (“Yes”) versus non-CVF-related (“No”). Table 6 presents the top 40 words (or n-grams), comprising the ten most frequent keywords for each of the four CVF components (Clan, Market, Hierarchy, and Adhocracy), sorted by total segment count within each component. A full comparison of all dictionary terms across all four components is provided in Appendix B Table B5.

[Insert Table 6]

The results reveal a pattern: many dictionary words appear frequently in segments that our pipeline classified as not related to organizational culture. The Yes/Total ratios in Table 6 show that dictionary words often appear in contexts disconnected from genuine discussions of cultural values and orientations. For the Clan component, “partic*” appears in CVF-related segments

only 33.73% of the time, while “help*” registers at 44.96%. For the Market component, “revenue*” appears in culture-relevant contexts only 34.42% of the time, and “strong*” at 43.71%. Similarly, among Hierarchy keywords, “detail*” appears in culture-relevant segments only 27.88% of the time and “inform*” at 34.45%. For the Adhocracy component, “trend*” appears in culture-relevant contexts 32.81% of the time, while “start*” registers at only 38.46%.

This pattern extends to the full dictionary (Appendix B Table B5), where a substantial share of terms across all four CVF components appear in culture-related segments less than 50% of the time. Certain terms do exhibit high culture-relevance percentages; for instance, “innovat*” in the Adhocracy component reaches approximately 92%, “compet*” in the Market component reaches approximately 81%, and “collab*” in the Clan component reaches approximately 81%. However, such terms are relatively few in number, and even high-performing keywords are not immune to occasional contextual misuse.

This finding underscores a fundamental limitation of dictionary-based approaches: they capture keyword frequency rather than meaningful discourse about organizational culture and values. Importantly, the low Yes/Total ratios observed for most dictionary terms should not be read in isolation as evidence in favor of our approach. Rather, they should be interpreted alongside two complementary pieces of evidence already established in the manuscript. First, the qualitative examples in Table 5 demonstrate that CVF-related keywords frequently appear in contexts that are demonstrably unrelated to organizational culture, such as references to product names, macroeconomic conditions, financial reporting metrics, or specific business entities. Second, the validation results in Section 4.1 confirm that the Phase I gate’s classifications align closely with human judgment, achieving near-perfect agreement on a rigorously annotated ground-truth set. Taken together, these three lines of evidence (the contextual misuse of

dictionary terms documented in Table 5, the low keyword-to-classification correspondence reported here, and the high human-model agreement established through validation) indicate that the divergence between keyword presence and our pipeline’s classifications reflects genuine contextual discrimination rather than arbitrary disagreement. Our approach therefore captures a more meaningful measure of corporate culture, one that distinguishes substantive cultural discourse from keyword occurrence.

5. Validation of the Four CVF Dimensions

Given that our approach to constructing the four contextual CVF measures is new, it is important to assess the validity of these measures by examining whether they are associated with firm outcomes that reflect the underlying logic of each quadrant. Accordingly, we validate each dimension against outcomes that should be most closely related to the economic content of that cultural orientation. We further compare our newly constructed segment-based contextual CVF measures with benchmark measures based on the corresponding Fang et al. (2023) culture ratios. The purpose of this comparison is to evaluate whether our measures deliver consistent validation evidence for the intended cultural dimensions. Overall, these tests are designed to assess the construct validity of the newly generated CVF-based culture measures. Following Fang et al. (2023), our validation regressions include a common set of firm-level controls: the natural logarithm of market capitalization, leverage, cash holdings scaled by net assets, dividend yield, and the natural logarithm of firm age. All specifications further include industry fixed effects and fiscal-year fixed effects, with standard errors clustered at the firm level.

[Insert Table 7]

Table 7 reports descriptive statistics for the variables used in the validation tests. The sample includes up to 39,389 firm-year observations. The segment-based CVF ratios have means of 0.026 for Collaborate, 0.190 for Compete, 0.077 for Create, and 0.106 for Control. The Fang benchmark ratios are smaller in magnitude, with means ranging from 0.006 for Collaborate_fang and Control_fang to 0.017 for Compete_fang. The validation outcomes also display substantial variation: the mean Emp/AT is 0.004, SGA/Sales is 0.442, ROA is -0.018, Market Share is 0.017, RD is 0.069, $\ln(1+\text{Patent})$ is 0.949, Cash ETR is 0.249, and the mean Restatement Filing (‘Restat filing’) indicator is 0.090.

5.1 Collaborate culture and labor intensity

5.1.1 Employees, labor-related spending, and collaborate culture

In the CVF, collaborative culture is the internally oriented and flexible “people” quadrant. It emphasizes teamwork, shared goals, employee participation, feedback, cohesion, and commitment to human development (Quinn & Rohrbaugh, 1983; Quinn, 1988; Cameron & Quinn, 2006; Cameron et al., 2006; Hartnell et al., 2011; Fang et al., 2023). A firm organized around these values places employees at the center of production and coordination. It follows that collaborative firms should maintain a larger employee base and spend more on the internal organizational systems that support teamwork and human-capital development. Prior finance research provides a natural empirical basis for these predictions. Edmans (2011) shows that employee-related intangible capital is value relevant. Eisfeldt and Papanikolaou (2013) define organization capital as a production factor embodied in key talent. Peters and Taylor (2017) argue that SG&A partly reflects investment in organization capital, and Chan et al. (2022) show that SG&A includes expenditures such as recruiting, training, marketing, and information

technology that build such capital. Taken together, these arguments imply that firms with stronger collaborate cultures should employ more people per unit of assets and should spend more on the organizational infrastructure needed to sustain teamwork and employee participation. To validate the collaborative dimension, we therefore examine whether firms with stronger collaborative values maintain a larger employee base relative to the assets they employ and incur greater spending on labor- and coordination-related activities, captured by employees scaled by assets ('Emp/AT') and SG&A scaled by sales ('SGA/Sales').

5.1.2 Empirical evidence

Table 8 examines the relation between the Collaborate score and two outcomes that should be closely linked to collaborative culture: employees scaled by assets ('Emp/AT') and SG&A scaled by sales ('SGA/Sales'). Columns (1)-(4) use our segment-based CVF measures, while columns (5)-(8) use the Fang et al. (2023) benchmark measures. For each dependent variable and measurement approach, the first column includes only the focal culture dimension, and the second includes the four CVF dimensions jointly. All specifications include the control variables shown in the table, industry fixed effects, and fiscal-year fixed effects, with standard errors clustered at the firm level.

[Insert Table 8]

The estimates strongly support the interpretation that collaborative culture is related to labor intensity and labor-related organizational spending. Columns (1) and (2) show that our Collaborate measure is positively and significantly associated with a greater employee base per asset employed: the coefficient is 0.006 in the collaborate-only specification ($p < 0.01$) and 0.005 in the four-culture specification ($p < 0.05$). Columns (3) and (4) show the same pattern for labor-

related operating spending. The coefficient on Collaborate is 1.520 in the collaborate-only specification ($p < 0.01$) and 0.724 in the four-culture specification ($p < 0.01$). The Fang benchmark provides mixed evidence for this dimension. Collaborate_fang is not statistically significant for Emp/AT in columns (5) and (6), but it is positive and significant for SGA/Sales in columns (7) and (8), with coefficients of 16.598 and 16.452, respectively (both $p < 0.01$). Overall, the segment-based Collaborate measure validates on both labor intensity and SGA spending, whereas the Fang benchmark is concentrated in the SGA/Sales outcome.

5.2 Compete culture and product-market performance

5.2.1 Profitability, market share, and compete culture

In the CVF, competitive culture is the externally oriented and stability-focused “market” quadrant. It emphasizes competitiveness, market position, productivity, profitability, goal achievement, and market share (Quinn & Rohrbaugh, 1983; Quinn, 1988; Cameron & Quinn, 2006; Cameron et al., 2006; Hartnell et al., 2011; Fang et al., 2023). Firms organized around these values should be more successful in converting operating resources into profits and more effective in securing product-market demand. This prediction also aligns with product-market finance research, which treats profitability and market-share gains as direct manifestations of competitive positioning (Campello, 2003). He and Huang (2017) likewise document that stronger product-market positioning is associated with higher market-share growth and improved operating profitability. Overall, prior research also suggests that firms with stronger competitive cultures should exhibit stronger operating performance and greater product-market success. To validate the competitive dimension, we therefore examine whether firms with stronger competitive values exhibit higher operating profitability and a stronger realized position in

product markets, captured by return on assets ('ROA', measured as EBIT scaled by total assets) and firm sales as a share of total sales in the same industry ('Market Share'). Collectively, these arguments imply that a valid Compete measure should be positively related to both profitability and market share.

5.2.2 Empirical evidence

Table 9 examines the relation between the Compete score and two outcomes that should reflect market-oriented culture: Return on Asset ('ROA') and market share ('Market Share'). Columns (1)-(4) use our segment-based CVF measures, while columns (5)-(8) use the Fang et al. (2023) benchmark measures. For each dependent variable and measurement approach, the first column includes only the focal culture dimension, and the second includes the four CVF dimensions jointly. All specifications include the control variables shown in the table, industry fixed effects, and fiscal-year fixed effects, with standard errors clustered at the firm level.

[Insert Table 9]

The evidence supports the expected product-market interpretation. Columns (1) and (2) show that competitive firms are positively and significantly associated with higher operating profitability. The coefficient on Compete is 0.065 in the compete-only specification ($p < 0.01$) and 0.318 in the four-culture specification ($p < 0.01$). Columns (3) and (4) show that competitive firms also hold larger product-market shares. The coefficient on Compete is 0.025 in the compete-only specification ($p < 0.01$) and 0.016 in the four-culture specification ($p < 0.01$). The Fang benchmark also loads positively on these product-market outcomes. Compete_fang is positive and significant for ROA in columns (5) and (6), with coefficients of 3.536 and 4.194, respectively (both $p < 0.01$), and positive and significant for Market Share in columns (7) and

(8), with coefficients of 0.207 and 0.154, respectively (both $p < 0.01$). Thus, both measures validate the Compete dimension, while the segment-based measure delivers the expected positive relation in both the focal-only and four-culture specifications.

5.3 Create culture and corporate innovation

5.3.1 Innovation input, innovation output, and create culture

In the CVF, creative culture is the externally oriented and flexible “adhocracy” quadrant. It emphasizes experimentation, entrepreneurship, pioneering initiatives, adaptability, and risk-taking (Quinn & Rohrbaugh, 1983; Quinn, 1988; Cameron & Quinn, 2006; Cameron et al., 2006; Hartnell et al., 2011; Fang et al., 2023). Firms organized around these values should commit more resources to innovation and generate more innovative output. The innovation literature uses R&D spending and patenting outcomes to capture these two aspects of innovative activity. Li et al. (2021) validate innovation culture using R&D spending and number of patents, while Fiordelisi et al. (2019) show directly that creative corporate culture is an important driver of innovation, measured by patent quantity and patent importance. Overall, prior literature further suggests that firms with stronger creative cultures should commit more resources to innovation and generate more innovative output. To validate the creative dimension, we therefore examine whether firms with stronger creative values commit more resources to innovation and generate more innovative output, captured by R&D intensity ($\text{'R\&D/AT'}$) and the natural logarithm of one plus the number of patents granted in a year ('ln(1+Patent')). The patent data come from Kogan et al. (2017). Taken together, these findings imply that a valid Create measure should load on both innovation input and innovation output.

5.3.2 Empirical evidence

Table 10 examines the relation between the Create score and two outcomes that should be most closely linked to creative culture: R&D intensity ('R&D/AT') and patent output ('ln(1+Patent)'). Columns (1)-(4) use our segment-based CVF measures, while columns (5)-(8) use the Fang et al. (2023) benchmark measures. For each dependent variable and measurement approach, the first column includes only the focal culture dimension, and the second includes the four CVF dimensions jointly. All specifications include the control variables shown in the table, industry fixed effects, and fiscal-year fixed effects, with standard errors clustered at the firm level.

[Insert Table 10]

The estimates provide strong support for the creative dimension. Columns (1) and (2) show that our Create measure is positively and significantly associated with greater R&D intensity. The coefficient on Create is 0.174 in the create-only specification ($p < 0.01$) and 0.371 in the four-culture specification ($p < 0.01$). Columns (3) and (4) show an equally strong pattern for innovative output, measured by $\ln(1+\text{Patent})$. The coefficient on Create is 3.349 in the create-only regression ($p < 0.01$) and 4.300 in the four-culture regression ($p < 0.01$). The Fang benchmark also captures part of the innovation pattern. Create_fang is statistically insignificant for R&D intensity in the create-only specification but becomes positive and significant in the four-culture specification, with a coefficient of 0.630 ($p < 0.05$). For patent output, Create_fang is positive and significant in both specifications, with coefficients of 27.087 and 16.612 (both $p < 0.01$). Overall, the segment-based Create measure consistently loads on both innovation input and output, while the Fang benchmark is especially strong for patent output and less stable for R&D intensity in the focal-only specification.

5.4 Control culture and tax and reporting discipline

5.4.1 Cash ETR, error-restatement disclosure (restatement filing), and control culture

In the CVF, control culture is the internally oriented and stability-focused “hierarchy” quadrant. It emphasizes reliability, predictability, efficiency, standardization, monitoring, and formalized procedures (Quinn & Rohrbaugh, 1983; Quinn, 1988; Cameron & Quinn, 2006; Cameron et al., 2006; Hartnell et al., 2011; Fang et al., 2023). Firms organized around these values should be more rule-based in activities that depend heavily on internal processes and compliance. This logic maps naturally into tax outcomes: firms that place greater weight on monitoring and procedural discipline should be less inclined to pursue aggressive tax avoidance, implying a higher cash effective tax rate (‘Cash ETR’) (Dyreng et al., 2008; Afzali & Thor, 2025). The same logic also extends to the formal reporting of accounting errors. Ashbaugh-Skaife et al. (2008) show that observed disclosures of internal control problems reflect not only underlying control failures but also firms’ incentives to discover and report those problems. Doyle et al. (2007) link internal control weaknesses to poorer reporting quality, and Hennes et al. (2008) show the importance of distinguishing error restatements from irregularities. We therefore interpret the filing of an error-restatement (‘Restat Filing’), defined as an indicator equal to one if the firm files a restatement that discloses and corrects an identified reporting error in the subsequent year, and zero otherwise, as an observed act of formal error identification and correction. Under this interpretation, a positive association between control culture and error-restatement filing is consistent with stronger monitoring and corrective discipline. Overall, we examine whether firms with stronger control values display more rule-based tax behavior and a stronger tendency to formally identify and correct reporting errors, captured by the cash effective tax rate (‘Cash

ETR') and an indicator for whether the firm issues an error-restatement ('Restat Filing') announcement.

5.4.2 Empirical evidence

Table 11 examines the relation between the Control score and two outcomes that should be closely linked to hierarchical culture: 'Cash ETR' and error-restatement ('restat filing') disclosure. Columns (1)-(4) use our segment-based CVF measures, while columns (5)-(8) use the Fang et al. (2023) benchmark measures. For 'Cash ETR', we use OLS regressions; for 'Restat Filing', we use logit regressions. All specifications include the control variables shown in the table, industry fixed effects, and fiscal-year fixed effects, with standard errors clustered at the firm level.

[Insert Table 11]

The estimates fit the expected interpretation. Columns (1) and (2) show that firms with stronger control cultures are positively and significantly associated with higher cash effective tax rates. The coefficient on Control is 0.137 in the control-only specification ($p < 0.01$) and 0.205 in the four-culture specification ($p < 0.01$). Columns (3) and (4) show that control firms are also more likely to issue an error-restatement announcement. The coefficient on Control is 1.639 in the control-only specification ($p < 0.01$) and 1.574 in the four-culture specification ($p < 0.01$). By contrast, the Fang benchmark provides little evidence for the Control dimension in this setting. Control_fang is statistically insignificant for 'Cash ETR' in columns (5) and (6), with coefficients of 13.805 and 3.708, and statistically insignificant for 'Restat Filing' in columns (7) and (8), with coefficients of 7.999 and 8.595. Taken together, these results suggest that the

segment-based Control measure captures rule-based tax behavior and formal error-identification discipline more consistently than the Fang benchmark.

5.5 Overall assessment

Overall, the results provide broad support for the validity of our newly constructed CVF measures. Higher Collaborate scores are associated with a larger employee base and greater SGA spending, higher Compete scores are associated with stronger operating profitability and larger market shares, higher Create scores are associated with greater R&D intensity and patent output, and higher Control scores are associated with higher cash effective tax rates and a greater likelihood of formally filing error-restatement announcements. The Fang benchmark captures several related patterns, including SGA spending, product-market performance, and patent output. However, the benchmark evidence is no significant for some paired validation outcomes, particularly ‘Emp/AT’ for the Collaborate dimension, ‘R&D/AT’ for the Create dimension, and both ‘Cash ETR’ and ‘Restat Filing’ for the Control dimension. By contrast, our measures produce consistent validation results across the four focal cultural dimensions. Collectively, these findings suggest that our segment-based contextual CVF measures capture distinct cultural dimensions in a way that maps naturally into observable firm policies and outcomes.

6. Discussion

This study set out to develop a context-sensitive, LLM-based methodology for measuring the four dimensions of the Competing Values Framework from the unscripted Q&A portions of earnings conference calls. The results reported in Section 4 demonstrate that the proposed multi-

agent pipeline achieves strong classification performance against human annotations, reveals substantial divergence from dictionary-based keyword counts, and produces contextual culture scores that associate consistently and significantly with theoretically predicted firm outcomes, as presented in Section 5. In this section, we discuss the theoretical, methodological, and practical contributions of these findings.

6.1 Theoretical Contributions

This research advances the theoretical understanding and measurement of organizational culture in several important ways. First, by providing a new, context-sensitive operationalization of the Competing Values Framework, it refines how this foundational construct is measured. The CVF has long been recognized as a powerful tool for understanding organizational culture (Cameron & Quinn, 2006; Hartnell et al., 2011), but existing measurement approaches have struggled to capture its nuanced, competing tensions. A foundational premise in the text-based culture measurement literature is that cultural constructs leave lexical traces: firms with stronger cultural orientations use culture-related words more frequently, and the presence of such words can therefore serve as a reliable indicator of the underlying construct (Fang et al., 2023; Fiordelisi & Ricci, 2014; Loughran & McDonald, 2011). Our analysis provides the first systematic, discourse-level test of this premise for the four dimensions of the CVF, and the results challenge its validity. Across the Fang et al. (2023) CVF wordlists, many of the most frequently occurring dictionary terms appear in culture-related segments less than half the time, with some of the most common stems such as “detail*” in the Hierarchy dictionary and “partic*” in the Clan dictionary registering culture-relevance rates below 35%. Qualitative inspection corroborates this finding, showing that dictionary terms appear devoid of meaningful cultural content (Table 5). This

evidence directly challenges the prevailing assumption in the textual analysis literature that the mere presence of keywords reliably signals meaningful cultural discourse and underscores the necessity of adopting context-sensitive analytical methods.

The empirical validation analyses in Section 5 reinforce this finding from the opposite direction. Our contextual measures yield consistent, statistically significant associations with the theoretically predicted outcomes for each CVF dimension—employee intensity and SGA spending for Collaborate, operating profitability and product-market share for Compete, R&D intensity and patent output for Create, and tax behavior and restatement filing for Control—whereas the Fang et al. (2023) dictionary-based specification captures some of these patterns but produces statistically insignificant associations for several key paired outcomes, including employee intensity for Collaborate, R&D intensity for Create, and both tax behavior and restatement filing for Control. Together, these results suggest that organizational culture as captured by the CVF is not a lexical phenomenon—it does not reside in word choice—but a contextual and discursive one, encoded in the meaning, framing, and communicative function of extended passages rather than in the frequency of individual terms. This distinction carries theoretical weight: it implies that the relationship between corporate communication and underlying culture is mediated by context in ways that prior measurement frameworks have not accounted for, and that empirical conclusions drawn from lexical proxies may understate the economic significance of organizational culture by conflating meaningful cultural discourse with incidental word use.

These findings also position our work as building on and extending Fang et al. (2023), who make an important contribution by introducing CVF-based culture measures to the corporate finance literature and documenting the stabilizing role of controlling culture during the

2008–09 crisis. However, their analysis is focused almost exclusively on the Control dimension and on crisis-period outcomes; the other three CVF dimensions (i.e., Collaborate, Compete, and Create) are consistently insignificant in their main specifications, and the paper does not investigate whether this reflects genuine economic irrelevance or measurement limitations inherent in the dictionary approach. We address this gap in three ways. First, we conduct a systematic construct-validation exercise that pairs each CVF dimension with the specific firm outcomes that its own theoretical logic predicts, rather than regressing a broad set of firm characteristics on all four culture scores simultaneously. Second, we benchmark our newly constructed contextual scores against Fang et al. (2023)’s dictionary-based scores on the same validation outcomes, revealing that their approach produces inconsistent and, for several key outcomes, economically negligible associations. Third, our panel-based validation covers a substantially larger sample and broader set of firm outcomes, establishing that all four dimensions of the CVF carry distinct economic content when measured with sufficient contextual precision. Together, these results suggest that the insignificance of Collaborate, Compete, and Create dimensions in prior work is more likely an artifact of measurement limitations than evidence of economic irrelevance.

Second, our findings reinforce the theoretical proposition that organizational culture is observable through spontaneous executive discourse. By shifting the analytical lens from static 10-K reports to the dynamic Q&A of earnings calls, this study provides evidence that culture is enacted by leadership in real-time, unscripted communication. As Li et al. (2021) note, the candid nature of executive responses in the Q&A setting makes such discourse particularly informative when cultural themes do arise. Our empirical results lend meaningful weight to this proposition. The fact that cultural discourse measured from unscripted Q&A associates

significantly with observable indicators across all four CVF dimensions—employee outcomes, competitive performance, innovation activity, and governance quality—provides evidence that spontaneous managerial communication contains genuine cultural signals, rather than performative or incidental language. This finding contributes to IS and management research by framing executive discourse as a key site where cultural values are performed and can be systematically analyzed using AI-driven contextual methods, moving the measurement of culture from what is espoused in formal filings to what is practiced in unscripted communication. The distinction between espoused and enacted culture is particularly important for the CVF, as organizations may claim to balance competing values in their formal documents while actually privileging certain dimensions over others in practice (Graham et al., 2022; Guiso et al., 2015).

Third, the study provides new empirical evidence on the relative prevalence and co-occurrence of CVF dimensions in contemporary corporate discourse. While the original CVF was developed through analysis of organizational effectiveness criteria (Quinn & Rohrbaugh, 1983), relatively little is known about how these four cultural orientations are distributed in the spontaneous communication of senior executives. Our classification results reveal a theoretically interpretable pattern: Market discourse predominates, followed by Hierarchy, Adhocracy, and Clan. This ordering aligns with the communicative demands of the earnings call setting, where the audience of analysts and investors naturally elicits discussion of competitive performance and operational discipline more frequently than employee-centered or innovation-centered themes. Importantly, however, all four quadrants are detectable at meaningful frequencies, confirming that the earnings call Q&A captures variation across the full CVF even in a setting oriented toward external stakeholders. Furthermore, a substantial share of segments (~48%) reflect more than one cultural orientation, consistent with the CVF's foundational premise that

organizations simultaneously exhibit competing values in varying degrees (Cameron & Quinn, 2006). This co-occurrence pattern suggests that the four cultural dimensions are not experienced or communicated as discrete alternatives but rather as overlapping orientations that coexist within the same organizational discourse, and that future research should attend to the interplay among cultural orientations rather than treating them as mutually exclusive categories.

6.2 Methodological Contributions

This study makes significant methodological contributions to IS research on organizational culture. It introduces a contextual, multi-agent LLM-based pipeline inspired by blackboard architecture (Salemi et al., 2025) for measuring the four dimensions of the CVF from earnings call transcripts, overcoming the limitations of traditional dictionary approaches that often misinterpret context and produce false positives whenever culture-related terms appear as part of proper nouns, technical descriptors, financial metrics, or routine discourse markers detached from any meaningful cultural meaning (Table 5). Building on the methodological innovations demonstrated in the companion essay on corporate integrity, this work extends those techniques to a more comprehensive cultural framework. While the integrity work focused on a single cultural construct with four components, this study tackles the additional complexity of measuring four competing cultural orientations that exist in tension with one another—a classification challenge that is more demanding because the four CVF quadrants are defined relationally, each acquiring meaning partly through its contrast with the others (Cameron & Quinn, 2006).

The architecture’s design offers several properties that strengthen classification quality and distinguish this approach from prior computational methods. The strict separation of

observation from judgment across Phases II and III ensures that classification decisions are anchored in explicitly extracted evidence rather than unconstrained generation, reducing hallucination risk by grounding verdicts in a shared fact base. The multi-phase, shared-memory design enables more calibrated classifications that resemble committee-based deliberation rather than single-pass binary decisions. The shared blackboard provides a persistent reasoning trace that constitutes a verifiable audit trail for every classification—a property that is especially valuable for organizational culture measurement, where the subjective and contested nature of cultural constructs demands that classification decisions be traceable to explicit textual evidence and articulated reasoning.

The validation evidence reported in Section 4 and Section 5 reinforces these architectural claims empirically. The strong and stable classification performance across independent replications directly addresses the reproducibility concerns raised by de Kok (2025) and Chen et al. (2023) regarding the use of LLMs in academic research, demonstrating that a well-structured multi-agent pipeline can produce consistent outputs even given the stochastic nature of LLM inference. The comparison against the Fang et al. (2023) dictionary approach further demonstrates that the methodological gains from contextual interpretation are not marginal but qualitative: the pipeline recovers a fundamentally different and empirically more valid signal than keyword counting, as evidenced by the consistent divergence in external validation outcomes across the two approaches. Critically, the validation design itself constitutes a methodological contribution. By combining classification performance assessment against human annotations with external empirical validation that benchmarks contextual scores against an established dictionary-based alternative on theory-matched outcomes, this study offers a

multi-layered validation blueprint that future IS research can adopt when developing LLM-based measures of other complex organizational constructs.

The proposed methodology demonstrates how generative AI can serve not only as a computational tool but also as a disciplined reasoning framework for producing actionable insights into complex organizational phenomena. Our agentic classification pipeline offers a practical and scalable blueprint for future research in computational social science. Furthermore, this work is positioned within the broader evolution of IS methods, marking a shift from non-contextual lexical markers toward deep contextual interpretation enabled by generative AI. In doing so, we answer Storey et al.'s (2025) call for IS researchers to harness generative AI's potential by developing applications that link technical innovation with organizational and societal impact. By establishing a reliable and reproducible approach for analyzing complex organizational constructs, this study expands the methodological repertoire of IS research.

6.3 Practical Implications

Beyond its theoretical and methodological contributions, the Contextual CVF Culture Scores have practical implications that follow directly from the empirical findings reported in Section 4.3. Because the measures are grounded in the unscripted Q&A discourse of senior management and decompose organizational culture into four interpretable components aligned with a well-established theoretical framework, they offer tools that boards, investors, regulators, and researchers can use to make sense of how a firm's cultural orientation relates to observable outcomes.

For investors, this IS-driven analytical approach functions as a novel decision support tool, translating unstructured managerial discourse into quantifiable measures of culture. Our

empirical analyses demonstrate that contextual culture scores carry meaningful associations with employee outcomes, competitive performance, innovation activity, and governance quality—in each case, the signal carried by our context-aware measures is stronger and more consistent than that of the widely used dictionary-based proxy. This suggests that integrating contextual culture measures into existing screening and risk assessment workflows can surface governance-relevant information that keyword-based approaches miss. The scores can enhance risk assessment, inform M&A due diligence by identifying potential cultural clashes or synergies (Hertel et al., 2024), and enrich ESG-focused investment strategies. Understanding the cultural profiles of organizations through the CVF lens can help investors anticipate how firms will respond to different types of challenges and opportunities—whether a firm’s cultural orientation predisposes it toward innovation-driven growth, operational discipline, people-centered resilience, or competitive aggressiveness.

For corporate boards and managers, the component-level decomposition (Clan, Adhocracy, Hierarchy, and Market) enables targeted diagnostics, allowing firms to identify which facets of their organizational culture are most or least emphasized in unscripted managerial communication. The ability to track cultural discourse patterns longitudinally enables organizations to assess whether their intended cultural transformation initiatives are actually taking hold in executive thinking and communication. Organizations can use these insights to identify gaps between espoused and enacted values, enabling more targeted cultural interventions. The CVF framework’s emphasis on competing values also helps leaders understand the trade-offs inherent in cultural choices and make more informed decisions about cultural priorities, for example, recognizing when an emphasis on Market-oriented

competitiveness may be crowding out the Clan-oriented employee investment that sustains long-term capability.

For regulators and policymakers, the approach provides a scalable method to analyze cultural patterns across industries, offering a new form of supervisory intelligence that accounts not only for compliance metrics but also for the underlying cultural drivers of organizational behavior. Understanding whether firms in a particular sector tend toward Hierarchy (emphasizing control and compliance) versus Adhocracy (emphasizing innovation and risk-taking) can inform regulatory approaches and risk assessments. The ability to track cultural discourse longitudinally across firms and industries offers a complementary lens to traditional compliance metrics, enabling oversight that accounts for the cultural conditions that make rule-following more or less likely. More broadly, this study illustrates how generative AI, when embedded within a structured multi-agent reasoning framework, can be operationalized within IS to inform both organizational governance and broader institutional decision-making, bridging the gap between technological innovation and practical applications.

6.4 Limitations and Future Research

Our research has several limitations that also suggest directions for future work. First, we relied on open-source models rather than state-of-the-art proprietary models (e.g., GPT-4.5, Gemini 2.5 Pro), which may offer stronger reasoning and contextual understanding. While DeepSeek-R1 performed well; with F1 and Cohen's kappa values indicating near-perfect agreement with human annotations, proprietary models could yield marginal improvements. Future work should compare open-source and proprietary LLMs in CVF classification to understand the trade-offs between accessibility, reproducibility, and performance. The choice of open-source models was

deliberate, following de Kok's (2025) recommendation to avoid proprietary APIs that can be modified or discontinued without notice, but this trade-off should be explicitly acknowledged.

Second, our approach detects the presence of cultural discourse and classifies it into CVF dimensions but does not assess its degree or sentiment. A segment classified as reflecting Adhocracy culture is treated equivalently whether the executive is enthusiastically championing innovation or reluctantly acknowledging a failed initiative. Future work could incorporate finer-grained scoring, sentiment, or intensity measures to capture not only whether a cultural orientation is expressed but also the extent and manner in which it is communicated.

Third, focusing solely on the Q&A portions of earnings call transcripts limits the scope of cultural discourse captured. While the Q&A setting offers the advantage of spontaneous, unscripted communication, extending the analysis to other corporate channels such as prepared remarks, annual reports, investor presentations, internal communications, or social media could provide a more comprehensive view of organizational culture. Cross-channel studies could examine whether cultural values are communicated differently to various stakeholder groups and whether the alignment or misalignment across channels carries additional informational content.

Finally, the methodology could be extended to examine the antecedents and consequences of cultural change. Because our measure can be constructed at the firm-quarter level, it enables longitudinal tracking of cultural shifts in response to leadership transitions, strategic pivots, regulatory changes, or exogenous shocks. Understanding the dynamics of cultural evolution how quickly organizations shift their cultural emphasis and whether such shifts predict future performance represents a promising frontier for IS and management research.

7. Conclusion

This study introduces a novel approach to measuring organizational culture through the theoretical lens of the Competing Values Framework by embedding a reasoning LLM within a multi-agent blackboard architecture and applying it to the unscripted Q&A portions of quarterly earnings conference calls. The resulting Contextual CVF Culture Scores are theoretically grounded, firm-quarter measures built on the four cultural archetypes (i.e., Clan, Adhocracy, Hierarchy, and Market) that the CVF identifies as capturing the fundamental tensions in organizational life (Quinn & Rohrbaugh, 1983; Cameron & Quinn, 2006). Unlike survey- or dictionary-based approaches, our measures capture cultural discourse directly from senior management in real-time, context-dependent communication, and provide both aggregate and component-level scores that can be tracked across firms and over time.

We assemble 183,043 earnings call transcripts spanning 2008 to 2024, extract executive responses from the Q&A portion, and segment them into over 10.2 million discourse units. Each segment is routed through an eleven-agent classification pipeline that separates evidence extraction from contextual judgment, producing classifications that are traceable to explicit textual evidence and articulated reasoning. The pipeline achieves 96.00% accuracy, a Cohen's kappa of 0.920, and a Matthews correlation coefficient of 0.923 against human annotations, with strong stability across independent replications. Comparison with the Fang et al. (2023) dictionary-based approach reveals that many of the most common CVF dictionary stems appear in culture-related contexts less than half the time, confirming that contextual interpretation is a necessary condition for accurate culture measurement.

We validate the resulting measures through a systematic construct-validation exercise that pairs each CVF dimension with the specific firm outcomes that its own theoretical logic predicts. Firms with higher Collaborate scores maintain larger workforces relative to their assets and incur greater spending on labor-related activities. Firms with higher Compete scores exhibit stronger operating profitability and larger product-market shares. Higher Create scores are associated with greater R&D intensity and patent output. Higher Control scores relate to lower effective tax rates and fewer restatement filings. Across these analyses, our contextual measures produce consistent, theoretically grounded associations, while the dictionary-based benchmark captures some patterns but produces insignificant associations for several key outcomes. This pattern establishes that contextual discrimination is not merely a methodological refinement but a meaningful improvement with economic content.

The study contributes to IS research by demonstrating how generative AI can move from a tool for pattern recognition to a disciplined reasoning framework that links technical innovation with organizational and societal impact. It contributes to the corporate culture literature by providing a validated, component-level CVF measure that extends the work of Fang et al. (2023) and Fiordelisi and Ricci (2014), and to the growing literature on generative AI in management and accounting research by operationalizing an important but previously difficult-to-quantify organizational construct at scale. For practitioners, the measures offer actionable tools for monitoring, benchmarking, and strengthening organizational culture over time.

This work also addresses important methodological challenges in using LLMs for academic research, including reproducibility and validation. By employing open-source models and implementing a rigorous, multi-step validation process that combines classification performance assessment with external empirical validation, the proposed approach is designed to

be accessible to other researchers. The successful completion of this research yields a validated methodology for measuring the four dimensions of the CVF from textual data, a comprehensive dataset of culture scores for public firms, and open-source tools to enable the extension of this approach to other cultural constructs and organizational settings.

These contributions advance our understanding of how organizational culture can be systematically measured and analyzed, reinforcing the strategic role of information systems in transforming unstructured data into meaningful insights that can inform better governance and contribute to more effective and sustainable business practices. By bridging advanced computational methods with established organizational theory, this research demonstrates the transformative potential of combining AI technologies with rigorous social science frameworks to address fundamental questions about organizational behavior and effectiveness.

Appendix A: Variable Definitions

Table A1. Variable Definitions

This table reports the definitions of the main validation variables, the segment-based CVF measures, the Fang et al. (2023) benchmark corporate-culture measures, and the control variables.

Variable	Definition
Collaborate	Ratio of ‘collaborate’ culture related earnings call Q&A segments as identified by the LLM model to total earnings call Q&A segments.
Compete	Ratio of ‘compete’ culture related earnings call Q&A segments as identified by the LLM model to total earnings call Q&A segments.
Create	Ratio of ‘create’ culture related earnings call Q&A segments as identified by the LLM model to total earnings call Q&A segments.
Control	Ratio of ‘control’ culture related earnings call Q&A segments as identified by the LLM model to total earnings call Q&A segments.
Collaborate_fang	Fang et al. (2023) benchmark ratio for the collaborative culture dimension.
Compete_fang	Fang et al. (2023) benchmark ratio for the competitive culture dimension.
Create_fang	Fang et al. (2023) benchmark ratio for the creative culture dimension.
Control_fang	Fang et al. (2023) benchmark ratio for the control culture dimension.
Emp/AT	Number of employees scaled by total assets.
SGA/Sales	Selling, general, and administrative expenses scaled by sales.
ROA	Return on assets, measured as EBIT scaled by total assets.
Market Share	Firm sales divided by total sales of all firms in the same two-digit SIC industry.
R&D/AT	Research and development expenditures scaled by total assets.
ln(1+Patent)	Natural logarithm of one plus the number of patents granted in a year; patent data are from Kogan et al. (2017).
Cash ETR	Cash effective tax rate, defined as income taxes paid scaled by pre-tax income adjusted for special items.
Restat Filing	It is a dummy variable; it is equal to ‘1’ if the firm files for a non-fraudulent restatement in the subsequent year, or ‘0’ otherwise.
ln(Market cap)	Natural logarithm of market capitalization. Market capitalization is the product of stock price and shares outstanding.
Leverage	Financial leverage, measured as the sum of long-term and short-term debt divided by total assets.
Cash/Net assets	Cash holdings, measured as the ratio of cash and short-term investments to total net assets. Net assets equal total assets minus cash and short-term investments.
Dividend yield	Dividend yield, measured as dividends per share divided by stock price.
ln(Firm age)	Natural logarithm of firm age, where firm age is measured as years since the firm first appears in Compustat.

References

- Afzali, M., & Thor, T. (2025). Corporate culture and tax planning. *Review of Quantitative Finance and Accounting*, 64(2), 861-898.
- Allee, K. D., & Deangelis, M. D. (2015). The Structure of Voluntary Disclosure Narratives: Evidence from Tone Dispersion. *Journal of Accounting Research*, 53(2), 241–274.
- Ashbaugh-Skaife, H., Collins, D. W., Kinney, W. R., & LaFond, R. (2008). The effect of SOX internal control deficiencies and their remediation on accrual quality. *The Accounting Review*, 83(1), 217–250.
- Bai, J., Boyson, N. M., Cao, Y., Liu, M., & Wan, C. (2023). Executives vs. Chatbots: Unmasking Insights through Human-AI Differences in Earnings Conference Q&A. *SSRN Electronic Journal*.
- Barth, A., & Mansouri, S. (2021). Corporate culture and banking. *Journal of Economic Behavior & Organization*, 186, 46–75.
- Beckmann, L., Beckmeyer, H., Filippou, I., Menze, S., & Zhou, G. (2024). Unusual Financial Communication - Evidence from ChatGPT, Earnings Calls, and the Stock Market. *SSRN Electronic Journal*.
- Bhatt, P., Vemprala, N., Valecha, R., Hariharan, G., & Rao, H. R. (2023). User Privacy, Surveillance and Public Health during COVID-19 – An Examination of Twitterverse. *Information Systems Frontiers*, 25(5), 1667–1682.
- Bochkay, K., Hales, J., & Chava, S. (2020). Hyperbole or reality? Investor response to extreme language in earnings conference calls. *The Accounting Review*, 95(2), 31–60.
- Burns, T., & Stalker, G. M. (1961). *The management of innovation*. Tavistock Publications.
- Cameron, K. S., & Quinn, R. E. (2006). *Diagnosing and changing organizational culture: Based on the competing values framework*. John Wiley & Sons.
- Cameron, K. S., Quinn, R. E., DeGraff, J., & Thakor, A. V. (2006). *Competing values leadership: Creating value in organizations*. Edward Elgar Publishing.
- Campello, M. (2003). Capital structure and product markets interactions: Evidence from business cycles. *Journal of Financial Economics*, 68(3), 353-378.
- Chan, K., Guo, R.-J. J., Wang, Y. A., & Yang, H.-L. (2022). Organization capital and analyst coverage. *Journal of Empirical Finance*, 69, 81-105.
- Chen, J., Xiao, S., Zhang, P., Luo, K., Lian, D., & Liu, Z. (2024). M3-Embedding: Multi-Linguality, Multi-Functionality, Multi-Granularity text embeddings through Self-Knowledge distillation. *arXiv (Cornell University)*.

- Chen, L., Zaharia, M., & Zou, J. (2023). How is ChatGPT's behavior changing over time? *arXiv.org*.
- Choi, F. Y. Y. (2000). Advances in domain independent linear text segmentation. *arXiv.org*.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.
- de Kok, T. (2025). ChatGPT for textual Analysis? How to use Generative LLMs in Accounting Research. *Management Science*, 71(9), 7888–7906.
- Demers, E., Wang, V. X., & Wu, K. (2024). Measuring Corporate Human capital disclosures: lexicon, data, code, and research opportunities. *Journal of Information Systems*, 38(2), 163–186.
- Deming, W. E. (1986). *Out of the crisis*. MIT Press.
- Denison, D. R., & Mishra, A. K. (1995). Toward a theory of organizational culture and effectiveness. *Organization Science*, 6(2), 204–223.
- Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv.org*.
- Doyle, J. T., Ge, W., & McVay, S. (2007). Accruals quality and internal control over financial reporting. *The Accounting Review*, 82(5), 1141–1170.
- Dutta, S., Kumar, A., Pant, P., Walsh, C., & Dutta, M. (2023). Using 10-K text to gauge COVID-related corporate disclosure. *PLoS ONE*, 18(3), e0283138.
- Dyreng, S. D., Hanlon, M., & Maydew, E. L. (2008). Long-run corporate tax avoidance. *The Accounting Review*, 83(1), 61–82.
- Edmans, A. (2011). Does the stock market fully value intangibles? Employee satisfaction and equity prices. *Journal of Financial Economics*, 101(3), 621–640.
- Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44(2), 350–383.
- Eisfeldt, A. L., & Papanikolaou, D. (2013). Organization capital and the cross-section of expected returns. *Journal of Finance*, 68(4), 1365–1406.
- Fang, Y., Fiordelisi, F., Hasan, I., Leung, W. S., & Wong, G. (2023). Corporate culture and firm value: Evidence from crisis. *Journal of Banking & Finance*, 146, 106710.
- Fiordelisi, F., & Ricci, O. (2014). Corporate culture and CEO turnover. *Journal of Corporate Finance*, 28, 66–82.
- Fiordelisi, F., Renneboog, L., Ricci, O., & Lopes, S. S. (2019). Creative corporate culture and innovation. *Journal of International Financial Markets, Institutions and Money*, 63, 101137.
- Frankel, R., Johnson, M., & Skinner, D. J. (1999). An empirical examination of conference calls as a voluntary disclosure medium. *Journal of Accounting Research*, 37(1), 133.

- Graham, J. R., Grennan, J., Harvey, C. R., & Rajgopal, S. (2022). Corporate culture: Evidence from the field. *Journal of Financial Economics*, 146(2), 552–593.
- Grennan, J. (2019). *A corporate culture channel: How increased shareholder governance reduces firm value (Working paper)*. Duke University.
- Guiso, L., Sapienza, P., & Zingales, L. (2015). The value of corporate culture. *Journal of Financial Economics*, 117(1), 60–76.
- Guo, D., Yang, D., Zhang, H., Song, J., Wang, P., Zhu, Q., Xu, R., Zhang, R., Ma, S., Bi, X., Zhang, X., Yu, X., Wu, Y., Wu, Z. F., Gou, Z., Shao, Z., Li, Z., Gao, Z., Liu, A., . . . Zhang, Z. (2025). DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning. *Nature*, 645(8081), 633–638.
- Hartnell, C. A., Ou, A. Y., & Kinicki, A. (2011). Organizational culture and organizational effectiveness: A meta-analytic investigation of the Competing Values Framework's theoretical suppositions. *Journal of Applied Psychology*, 96(4), 677–694.
- Hasan, T., John, K., Teng, H., & Wu, Q. (2024). Creative corporate culture and corporate tax avoidance. *The British Accounting Review*, 56(3), 101217.
- Hassan, T. A., Hollander, S., Van Lent, L., & Tahoun, A. (2019). Firm-Level Political Risk: Measurement and Effects. *The Quarterly Journal of Economics*, 134(4), 2135–2202.
- He, J., & Huang, J. (2017). Product market competition in a world of cross-ownership: Evidence from institutional blockholdings. *Review of Financial Studies*, 30(8), 2674–2718.
- Hennes, K. M., Leone, A. J., & Miller, B. P. (2008). The Importance of Distinguishing Errors from Irregularities in Restatement Research: The Case of Restatements and CEO/CFO Turnover. *The Accounting Review*, 83(6), 1487–1519.
- Henry, E., & Leone, A. J. (2015). Measuring Qualitative Information in Capital Markets Research: Comparison of alternative methodologies to measure disclosure tone. *The Accounting Review*, 91(1), 153–178.
- Hertel, T., Kaya, D., & Reichmann, D. (2024). Corporate culture and M&A deals: Using text from crowdsourced employer reviews to measure cultural differences. *Journal of Banking & Finance*, 161, 107118.
- Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A Large Language Model for Extracting Information from Financial Text*. *Contemporary Accounting Research*, 40(2), 806–841.
- Huang, J., & Chang, K. C. (2022). Towards Reasoning in large Language Models: A survey. *arXiv.org*.
- Irugalbandara, C., Mahendra, A., Daynauth, R., Arachchige, T. K., Dantanarayana, J., Flautner, K., Tang, L., Kang, Y., & Mars, J. (2023). Scaling Down to Scale Up: A Cost-Benefit Analysis of Replacing OpenAI's LLM with Open Source SLMs in Production. *arXiv.org*.
- Jha, M., Qian, J., Weber, M., & Yang, B. (2023). ChatGPT and Corporate Policies. *SSRN Electronic Journal*.

- Ji, Y., Rozenbaum, O., & Welch, K. T. (2017). Corporate Culture and Financial reporting risk: Looking through the glassdoor. *SSRN Electronic Journal*.
- Kanter, R. M. (1977). *Men and women of the corporation*. Basic Books.
- Kaur, H., Ahsaan, S. U., Alankar, B., & Chang, V. (2021). A proposed Sentiment Analysis deep learning algorithm for analyzing COVID-19 tweets. *Information Systems Frontiers*, 23(6), 1417–1429.
- Keiser, N. L., & Payne, S. C. (2019). Are employee surveys biased? Impression management as a response bias in workplace safety constructs. *Safety Science*, 118, 453–465.
- Kim, A., Muhn, M., & Nikolaev, V. (2025). Bloated disclosures: Can ChatGPT help investors process information? *arXiv (Cornell University)*.
- Koch, S., & Pasch, S. (2023). *CultureBERT: Measuring Corporate Culture With Transformer-Based Language Models* (pp. 3176–3184). IEEE Computer Society.
<https://doi.org/10.1109/bigdata59044.2023.10386765>
- Kogan, L., Papanikolaou, D., Seru, A., & Stoffman, N. (2017). Technological innovation, resource allocation, and growth. *Quarterly Journal of Economics*, 132(2), 665–712.
- Kreps, D. M. (1990). Corporate culture and economic theory. In: J. E. Alt & K. A. Shepsle (Eds.), *Perspectives on positive political economy* (pp. 90–143). Cambridge University Press.
- Kurumathur, S. K., Bhatt, P., Valecha, R., Hariharan, G., & Rao, H. R. (2025). Examination of Polarization in Social Media in Aggressor-Oriented and Victim-Oriented Discourse Following Vigilantism. *Information Systems Frontiers*, 27(4), 1393–1412.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159.
- Larcker, D. F., & Zakolyukina, A. A. (2012). Detecting deceptive discussions in conference calls. *Journal of Accounting Research*, 50(2), 495–540.
- Li, F. (2010). The information Content of Forward-Looking Statements in Corporate Filings—A naïve Bayesian Machine Learning approach. *Journal of Accounting Research*, 48(5), 1049–1102.
- Li, K., Mai, F., Shen, R., & Yan, X. (2021). Measuring corporate culture using machine learning. *Review of Financial Studies*, 34(7), 3265–3315.
- Li, K., Mai, F., Shen, R., Yang, C., & Zhang, T. (2026). Dissecting corporate culture using generative AI. *Review of Financial Studies*, 39(1), 253–296.
- Lopez-Lira, A., & Tang, Y. (2023). Can ChatGPT forecast stock price movements? return predictability and large language models. *SSRN Electronic Journal*.
- Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1), 35–65.

- Luu, H. N., Nguyen, L. T. M., Vu, K. T., & Nguyen, L. Q. T. (2023). The impact of organizational culture on bank stability. *Review of Quantitative Finance and Accounting*, 61(2), 501–533.
- March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, 2(1), 71–87.
- Mayew, W. J., & Venkatachalam, M. (2012). The power of voice: managerial affective states and future firm performance. *The Journal of Finance*, 67(1), 1–43.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv.org*.
- Nguyen, D. D., Nguyen, L., & Sila, V. (2019). Does corporate culture affect bank risk-taking? Evidence from loan-level data. *British Journal of Management*, 30(1), 106–133.
- O'Reilly, C. A., & Chatman, J. A. (1996). Culture as social control: Corporations, cults, and commitment. *Research in Organizational Behavior*, 18, 157–200.
- OpenAI. (2025, April 29). *Sycophancy in GPT-4o: what happened and what we're doing about it*.
- Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock market index using fusion of machine learning techniques. *Expert Systems With Applications*, 42(4), 2162–2172.
- Peters, R. H., & Taylor, L. A. (2017). Intangible capital and the investment-q relation. *Journal of Financial Economics*, 123(2), 251–272.
- Porter, M. E. (1980). *Competitive strategy: Techniques for analyzing industries and competitors*. Free Press.
- Porter, M. E. (1985). *Competitive advantage: Creating and sustaining superior performance*. Free Press.
- Price, S. M., Doran, J. S., Peterson, D. R., & Bliss, B. A. (2012). Earnings conference calls and stock returns: The incremental informativeness of textual tone. *Journal of Banking & Finance*, 36(4), 992–1011.
- Quinn, R. E. (1988). *Beyond rational management: Mastering the paradoxes and competing demands of high performance*. San Francisco, CA: Jossey-Bass.
- Quinn, R. E., & Rohrbaugh, J. (1983). A spatial model of effectiveness criteria: Towards a competing values approach to organizational analysis. *Management Science*, 29(3), 363–377.
- Salemi, A., Parmar, M., Goyal, P., Song, Y., Yoon, J., Zamani, H., Pfister, T., & Palangi, H. (2025). LLM-Based Multi-Agent Blackboard System for information discovery in data science. *ArXiv.org*.
- Schein, E. H. (2010). *Organizational culture and leadership (4th ed.)*. Jossey-Bass.

- Sharma, G., Chatterjee, S., Rani, V., Tatavarthy, A. D., & Bose, I. (2025). Driving Sustainability: unravelling the dynamics of user engagement on organization generated social media content on electric vehicles. *Information Systems Frontiers*, 27(6), 2431–2461.
- Shojaee-Mend, H., Mohebbati, R., Amiri, M., & Atarodi, A. (2024). Evaluating the strengths and weaknesses of large language models in answering neurophysiology questions. *Scientific Reports*, 14(1), 10785.
- Singh, J. P., Kumar, A., Rana, N. P., & Dwivedi, Y. K. (2022). Attention-Based LSTM network for Rumor Veracity Estimation of Tweets. *Information Systems Frontiers*, 24(2), 459–474.
- Sprague, Z., Ye, X., Bostrom, K., Chaudhuri, S., & Durrett, G. (2023). MuSR: Testing the Limits of Chain-of-thought with Multistep Soft Reasoning. *arXiv.org*.
- Storey, V. C., Yue, W. T., Zhao, J. L., & Lukyanenko, R. (2025). Generative Artificial intelligence: evolving technology, growing societal impact, and opportunities for information systems research. *Information Systems Frontiers*, 27(5), 2081–2102.
- Suzgun, M., Scales, N., Schärli, N., Gehrmann, S., Tay, Y., Chung, H. W., Chowdhery, A., Le, Q., V., Chi, E. H., Zhou, D., & Wei, J. (2022). Challenging BIG-Bench tasks and whether Chain-of-Thought can solve them. *arXiv.org*.
- Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7), 509–533.
- Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3), 1139–1168.
- Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., Chowdhery, A., & Zhou, D. (2022). Self-Consistency improves chain of thought reasoning in language models. *arXiv.org*.
- Wang, Z., Mao, S., Wu, W., Ge, T., Wei, F., & Ji, H. (2024). *Unleashing the Emergent Cognitive Synergy in Large Language Models: A Task-Solving Agent through Multi-Persona Self-Collaboration* (pp. 257–279). NAACL.
- Weber, M. (1947). *The theory of social and economic organization* (A. M. Henderson & T. Parsons, Trans.). Oxford University Press.
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2022). Chain-of-Thought prompting elicits reasoning in large language models. *arXiv.org*.
- Weick, K. E., & Sutcliffe, K. M. (2007). *Managing the unexpected: Resilient performance in an age of uncertainty* (2nd ed.). Jossey-Bass.
- Williamson, O. E. (1975). *Markets and hierarchies: Analysis and antitrust implications*. Free Press.

Yang, Y., Qin, Y., Fan, Y., & Zhang, Z. (2023). Unlocking the Power of Voice for Financial Risk Prediction: A Theory-Driven Deep Learning Design Approach. *MIS Quarterly*, 47(1), 63–96.

Figures

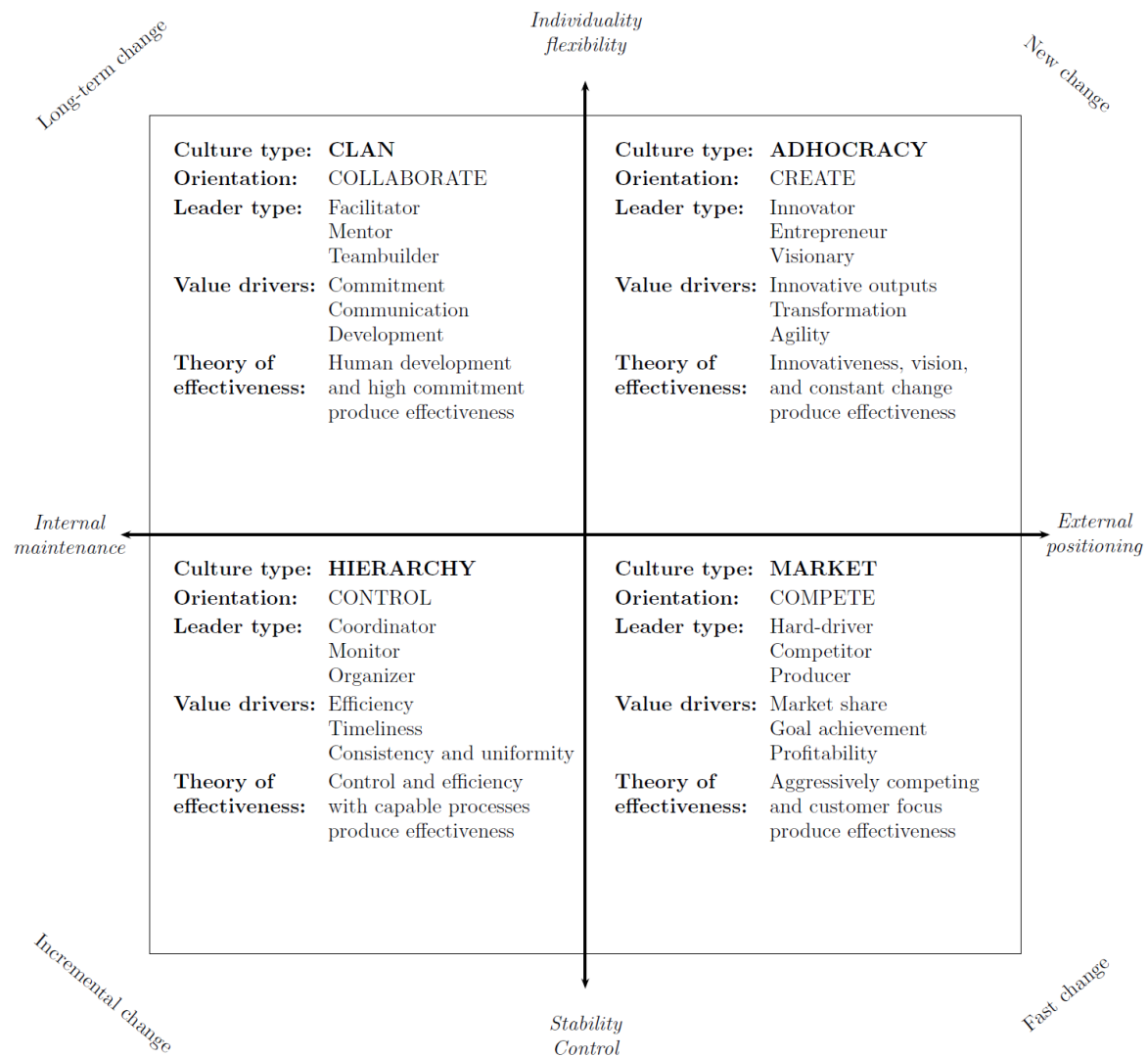


Figure 1. Competing Values Framework (CVF) (Source: Cameron et al., 2006)

CVF Dimension	Guiso et al. (2015)	Grennan (2019)	Li et al. (2021)	Graham et al. (2022)	Li et al. (2025)
Clan (Collaborate) <i>Quadrant: Internal & Flexible</i>	Communication Community Respect Teamwork	Collaboration	Respect Teamwork	Collaboration Community	Collaboration and people-focused
Adhocracy (Create) <i>Quadrant: External & Flexible</i>	Innovation	Adaptability	Innovation	Adaptability	Innovation and adaptability
Hierarchy (Control) <i>Quadrant: Internal & Stable</i>	Quality Safety	Detail-orientation	Quality	Detail-orientation	
Market (Compete) <i>Quadrant: External & Stable</i>	Hard work	Customer-orientation Results-orientation		Customer-orientation Results-orientation	Customer-oriented Performance-oriented

Figure 2. Mapping Cultural Values from Prior Literature to CVF Dimensions

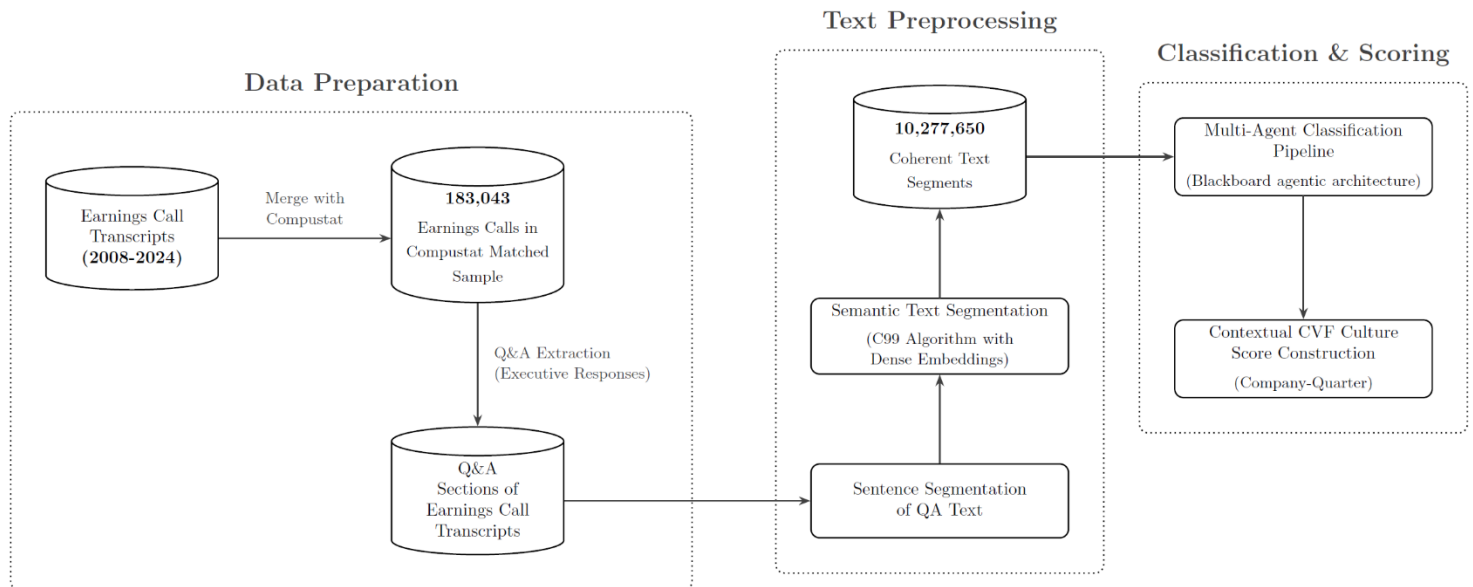


Figure 3. Process Model

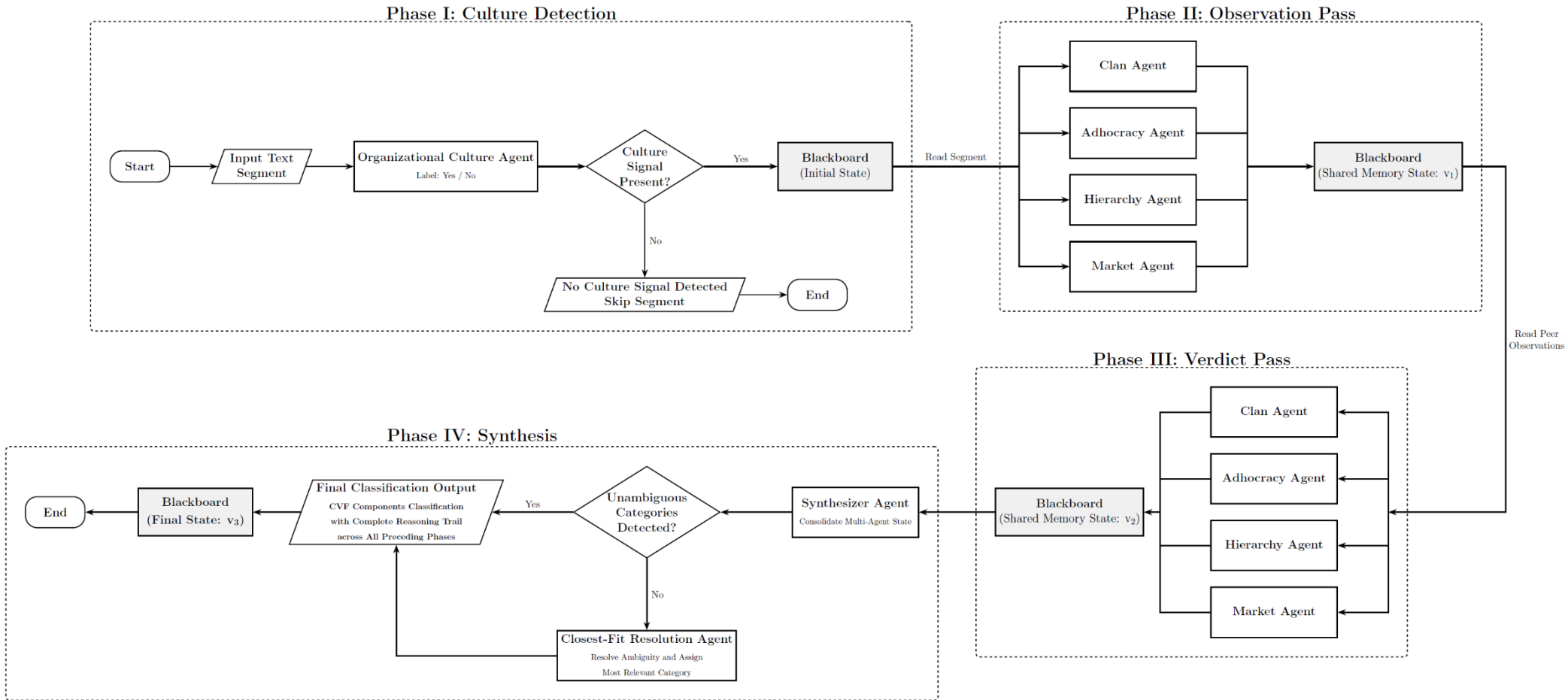


Figure 4. Multi agent organizational culture classification pipeline

Tables

Table 1: Distribution of CVF Components

Component	Number of Segments	Presence in CVF related Segments (%)	Presence in Whole Segments available in corpus (%)
Clan	211408	10.21%	2.06%
Adhocracy	576859	27.86%	5.61%
Market	1561783	75.43%	15.20%
Hierarchy	875061	42.26%	8.51%

Table 2: Classification performance of the Phase I overall CVF culture detection gate

	Accuracy	Precision (Macro)	Recall (Macro)	F1 (Macro)	F1 (Weighted)	Cohen's κ	MCC
Main Run	96.00%	96.00%	96.30%	95.99%	96.01%	0.920	0.923
Run 1	95.33%	95.29%	95.57%	95.32%	95.34%	0.907	0.909
Run 2	94.00%	93.95%	94.23%	93.99%	94.01%	0.880	0.882
Run 3	95.67%	95.58%	95.83%	95.65%	95.67%	0.913	0.914
Run 4	94.67%	94.58%	94.79%	94.65%	94.67%	0.893	0.894
Run 5	94.33%	94.27%	94.54%	94.32%	94.34%	0.887	0.888
Mean (Runs 1–5)	94.80%	94.73%	94.99%	94.79%	94.81%	0.896	0.897
SD (Runs 1–5)	–	–	–	0.62%	0.62%	0.0124	–

Table 3a: Per-class metrics for the Phase I CVF culture detection gate

	Precision	Recall	F1-Score	Support
Main Run – No	100%	92.59%	96.15%	162
Main Run – Yes	92.00%	100%	95.83%	138
Runs 1–5 Avg – No	97.66%	92.59%	95.06%	162
Runs 1–5 Avg – Yes	91.81%	97.39%	94.52%	138

Table 3b: Confusion matrix (main run)

	Predicted: No	Predicted: Yes
True: No	150	12
True: Yes	0	138

Table 4: Inter-run agreement (Cohen’s κ) – DeepSeek-R1

Subset Size (k)	N Combinations	Mean Pairwise κ	SD κ	Min κ	Max κ
2	10	0.931	0.020	0.900	0.967
3	10	0.931	0.012	0.911	0.951
4	5	0.931	0.007	0.920	0.940

Table 5: Examples of False Keyword Signals

Note: Bold and underlined terms denote words drawn from the CVF dictionary developed by Fang et al. (2023). The examples illustrate cases in which keyword presence does not correspond to meaningful CVF-related cultural content.

Sl. No.	Ticker	Publish Year	Fang et al. 2023 Keyword	Segment Text	Model Output	Model Reason
1	LOGI	2017	collab*	Let me take the second two and I'll let Vincent take first question on ASTRO. Yes, in terms of the momentum, what specifically. I would say we – it's been a while since we've seen such strong growth from some of our most profitable categories and really across the board Gaming, Video <u>Collaboration</u> , Harmony, we just across the board had very strong growth rates. I don't necessarily expect all those to continue quarter-after-quarter, but between that and the strong gross margin that gave us lot of confidence to go ahead and rise at this point. In terms of the – the timing of the raise, you asked, how would I compare this with past years, I would say yes, I mean it's a little hard to remember back to how we felt a year ago and the year before that and year before that, but I would say I feel really, really good about where we are at this point after the first quarter.	No	The text segment primarily discusses strong growth across various product categories and a decision to raise prices, reflecting financial performance and strategic planning. It does not explicitly mention cultural aspects like collaboration, innovation, competition, or control. While there is a focus on results and market performance, this aligns more with general business operations rather than a specific CVF quadrant. There are no direct references to internal processes, employee involvement, innovation, or external market competition in a cultural sense. Therefore, the evidence does not strongly support any of the four CVF categories.
2	CCJ	2024	partner*, team*	We'll sort it out. We're in discussions with them now. Our team is talking about what the production allocation is for this year and going forward and we should be able to recoup there. But don't forget, it's 10% I think of our production. So, it's important to us. We're happy to be there, we're happy to be <u>partners</u> with Kazatomprom, but in the bigger ticket Cameco Westinghouse, it's a smaller portion of our business.	No	The text segment primarily discusses production allocation and business partnerships, focusing on operational and financial aspects without indicating cultural orientations. There are no direct references to themes like employee involvement, innovation, competition, or control, which are central to the CVF quadrants. The language is about business discussions and partnerships, not organizational culture.
3	TGLS	2024	create*	So, our expectation is that everybody is probably going to be on the same boat and that's going to create inflation. And I think you're seeing that reflected on the markets today. And also, I may add that Colombia is one of the few countries that has either a balance or a deficit with the U.S. in trade. In other words, out of the 16 billion trade that we do with the U.S., 9 billion is in commodities that we ship like oil, coffee, bananas, flowers, coal.	No	The text discusses economic expectations and trade specifics with the U.S., focusing on commodities like oil and coffee. While it mentions market reactions and trade balances, these points are related to external market conditions and financial performance rather than organizational culture. There are no references to internal company practices, employee relations, management styles, or cultural values that align with any of the four CVF quadrants. The discussion is centered on external factors and trade, which doesn't provide sufficient evidence to classify the organizational culture under Clan, Adhocracy, Market, or Hierarchy.
4	SANW	2017	idea*, pioneer*	And we are just in the process of finalizing our appraisals on that. And once the appraisals are done, we can finalize the exact loan amount and close. Looking to do that here in the next 30 days or <u>ideally</u> this month. And this appraisal, is that the appraisals on the <u>Pioneer</u> assets that I know you are going to get -- Well, yes. Sorry, to clarify, this term loan is going to be collateralized by the mill in California and the mill that we bought from <u>Pioneer</u> that's located in Idaho and our R&D facilities.	No	The text segment discusses the process of finalizing appraisals and loan amounts, as well as the collateralization of assets. These statements focus on financial and operational procedures, such as appraisals, loan closings, and collateral, which are indicative of standard business operations rather than cultural orientation. There is no mention of themes related to collaboration, innovation, competition, or control within the organization. The language used is transactional and lacks evidence of shared goals, innovation, market focus, or internal processes that define the four CVF quadrants. Therefore, the text does not reflect any specific cultural quadrant under the Competing Values Framework.
5	QVCGA	2024	revenue*	As I think I mentioned in my prepared remarks, viewership across our five channels in the U.S. were down 4%, and so that just drove a level of softness throughout categories and like I said, we think it had a 1% or 2% impact on total <u>revenue</u> and then corresponding P&L. When you look at the category specifics, home <u>revenue</u> decreased 3%.	No	The text segment discusses viewership decline and its impact on revenue, which is a financial and operational update. It does not mention any cultural aspects related to collaboration, innovation, competition, or control. There are no references to employee involvement, innovation, market competition, or internal processes, making it impossible to classify under any of the CVF quadrants.

Table 6: CVF Term Classification Rates

Note: The table reports, for each of the 40 most frequent terms (top 10 keywords for each CVF component) from the CVF dictionary developed by Fang et al. (2023), the percentage of segments containing that term that were classified as “Yes.”

Fang et al. (2023) Keywords	CVF Component of Keyword	Number of Yes segments	Yes/Total segment (%)	Fang et al. (2023) Keywords	CVF Component of Keyword	Number of Yes segments	Yes/Total segment (%)
partic*	Clan	131665	33.73%	certain*	Hierarchy	187492	39.37%
people*	Clan	185395	54.39%	cost*	Hierarchy	192808	42.22%
help*	Clan	141253	44.96%	process*	Hierarchy	104938	47.66%
team*	Clan	179056	72.67%	system*	Hierarchy	78027	49.40%
partner*	Clan	104924	56.43%	efficien*	Hierarchy	85684	72.03%
capab*	Clan	70796	66.93%	detail*	Hierarchy	24550	27.88%
contribut*	Clan	32305	35.13%	inform*	Hierarchy	28651	34.45%
commit*	Clan	46287	53.38%	control*	Hierarchy	49764	64.85%
commun*	Clan	44444	53.28%	analys*	Hierarchy	19171	33.32%
engag*	Clan	40052	65.81%	productiv*	Hierarchy	37302	68.63%
market*	Market	580024	53.97%	new*	Adhocracy	332835	50.99%
growth*	Market	276458	43.97%	start*	Adhocracy	210335	38.46%
customer*	Market	329931	57.16%	chang*	Adhocracy	145118	38.06%
strong*	Market	169242	43.71%	develop*	Adhocracy	112749	53.16%
revenue*	Market	132421	34.42%	future*	Adhocracy	86624	47.30%
share*	Market	174545	54.33%	trend*	Adhocracy	53996	32.81%
result*	Market	96177	39.47%	begin*	Adhocracy	51031	35.78%
compet*	Market	195161	80.95%	init*	Adhocracy	70542	49.90%
value*	Market	131940	54.80%	creat*	Adhocracy	86709	65.02%
perform*	Market	100841	44.01%	thought*	Adhocracy	46301	35.54%

Table 7. Descriptive Statistics

This table reports descriptive statistics for the dependent variables, segment-based CVF measures, Fang et al. (2023) benchmark measures, and control variables used in the validation tests.

Variables	N	Mean	SD	P25	Median	P75
Collaborate	39206	0.026	0.029	0.007	0.019	0.037
Compete	39206	0.190	0.087	0.132	0.189	0.247
Create	39206	0.077	0.059	0.034	0.065	0.108
Control	39206	0.106	0.052	0.071	0.102	0.136
Collaborate_fang	39206	0.006	0.003	0.004	0.006	0.007
Compete_fang	39206	0.017	0.007	0.012	0.016	0.021
Create_fang	39206	0.008	0.003	0.006	0.008	0.009
Control_fang	39206	0.006	0.002	0.004	0.005	0.007
ln(Market cap)	37499	7.131	2.064	5.710	7.130	8.514
Leverage	37439	0.256	0.227	0.053	0.223	0.390
Cash/Net assets	37502	0.948	2.899	0.054	0.161	0.518
Dividend yield	37499	0.013	0.029	0.000	0.000	0.016
ln(Firm age)	37502	2.749	0.869	2.079	2.833	3.367
Emp/AT	38910	0.004	0.006	0.001	0.002	0.004
SGA/Sales	38352	0.440	0.932	0.103	0.228	0.439
ROA	39203	-0.017	0.267	-0.032	0.055	0.106
Market Share	39203	0.017	0.042	0.000	0.002	0.011
R&D/AT	39205	0.069	0.132	0.000	0.010	0.080
ln(1+Patent)	39206	0.952	1.553	0.000	0.000	1.609
Cash ETR	24681	0.249	0.317	0.094	0.203	0.304
Restat Filing	39206	0.090	0.286	0.000	0.000	0.000

Table 8. Validation of the Collaborate Dimension

This table reports validation tests for the Collaborate dimension. The dependent variables are employees scaled by total assets (Emp/AT) and SG&A scaled by sales (SGA/Sales). Columns (1)-(4) use the segment-based CVF measures, and columns (5)-(8) use the Fang et al. (2023) benchmark measures. For each measure, the first column for a dependent variable includes only the focal culture dimension, and the second includes the four CVF dimensions jointly. All specifications include the control variables shown in the table, industry fixed effects, and fiscal-year fixed effects. Firm-clustered standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Variables	(1) Emp/AT	(2) Emp/AT	(3) SGA/Sales	(4) SGA/Sales	(5) Emp/AT	(6) Emp/AT	(7) SGA/Sales	(8) SGA/Sales
Collaborate	0.006*** (0.002)	0.005** (0.002)	1.520*** (0.264)	0.724*** (0.248)				
Compete		0.001 (0.001)		-0.469*** (0.156)				
Create		-0.010*** (0.001)		1.558*** (0.197)				
Control		0.010*** (0.001)		0.745*** (0.183)				
Collaborate_fang					-0.002 (0.017)	-0.010 (0.017)	16.598*** (2.678)	16.452*** (2.675)
Compete_fang						0.025*** (0.009)		-5.833*** (1.354)
Create_fang						-0.010 (0.019)		8.716*** (2.414)
Control_fang						0.073*** (0.020)		8.360** (3.335)
ln(Market cap)	-0.000*** (0.000)	-0.000*** (0.000)	-0.057*** (0.005)	-0.060*** (0.005)	-0.000*** (0.000)	-0.000*** (0.000)	-0.056*** (0.005)	-0.054*** (0.005)
Leverage	-0.002*** (0.000)	-0.002*** (0.000)	-0.106** (0.044)	-0.074* (0.043)	-0.002*** (0.000)	-0.002*** (0.000)	-0.100** (0.044)	-0.099** (0.044)
Cash/Net assets	-0.000*** (0.000)	-0.000*** (0.000)	0.069*** (0.010)	0.069*** (0.010)	-0.000*** (0.000)	-0.000*** (0.000)	0.069*** (0.010)	0.066*** (0.010)
Dividend yield	0.003 (0.002)	0.002 (0.002)	-0.849*** (0.162)	-0.752*** (0.159)	0.003 (0.002)	0.003 (0.002)	-0.830*** (0.161)	-0.818*** (0.161)
ln(Firm age)	0.001*** (0.000)	0.001*** (0.000)	-0.108*** (0.011)	-0.100*** (0.010)	0.001*** (0.000)	0.001*** (0.000)	-0.106*** (0.010)	-0.106*** (0.011)
Constant	0.006*** (0.000)	0.005*** (0.000)	1.074*** (0.048)	0.981*** (0.049)	0.006*** (0.000)	0.005*** (0.000)	1.001*** (0.050)	0.972*** (0.055)
Obs.	37157	37157	36698	36698	37157	37157	36698	36698
Adj. R2	0.366	0.378	0.157	0.163	0.365	0.367	0.158	0.160

Table 9. Validation of the Compete Dimension

This table reports validation tests for the Compete dimension. The dependent variables are ROA and market share. Columns (1)-(4) use the segment-based CVF measures, and columns (5)-(8) use the Fang et al. (2023) benchmark measures. For each measure, the first column for a dependent variable includes only the focal culture dimension, and the second includes the four CVF dimensions jointly. All specifications include the control variables shown in the table, industry fixed effects, and fiscal-year fixed effects. Firm-clustered standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Variables	(1) ROA	(2) ROA	(3) Market Share	(4) Market Share	(5) ROA	(6) ROA	(7) Market Share	(8) Market Share
Collaborate		-0.213*** (0.071)		0.034*** (0.013)				
Compete	0.065*** (0.025)	0.318*** (0.035)	0.025*** (0.005)	0.016*** (0.005)				
Create		-0.377*** (0.045)		0.002 (0.008)				
Control		-0.241*** (0.041)		0.014* (0.007)				
Collaborate_fang						-6.420*** (0.734)		0.723*** (0.118)
Compete_fang					3.536*** (0.326)	4.194*** (0.336)	0.207*** (0.059)	0.154*** (0.059)
Create_fang						-0.686 (0.653)		0.023 (0.122)
Control_fang						-8.083*** (0.775)		0.389*** (0.141)
ln(Market cap)	0.043*** (0.001)	0.043*** (0.001)	0.007*** (0.000)	0.007*** (0.000)	0.041*** (0.001)	0.040*** (0.001)	0.007*** (0.000)	0.007*** (0.000)
Leverage	-0.003 (0.012)	-0.011 (0.012)	0.000 (0.002)	0.000 (0.002)	-0.000 (0.012)	-0.005 (0.012)	0.000 (0.002)	0.001 (0.002)
Cash/Net assets	-0.025*** (0.001)	-0.024*** (0.001)	0.001*** (0.000)	0.001*** (0.000)	-0.024*** (0.001)	-0.023*** (0.001)	0.001*** (0.000)	0.001*** (0.000)
Dividend yield	0.360*** (0.066)	0.343*** (0.065)	0.028** (0.012)	0.028** (0.012)	0.359*** (0.066)	0.346*** (0.065)	0.027** (0.012)	0.028** (0.012)
ln(Firm age)	0.033*** (0.003)	0.031*** (0.003)	0.006*** (0.001)	0.006*** (0.001)	0.033*** (0.003)	0.032*** (0.003)	0.006*** (0.001)	0.006*** (0.001)
Constant	-0.399*** (0.013)	-0.382*** (0.013)	-0.056*** (0.003)	-0.057*** (0.003)	-0.434*** (0.014)	-0.350*** (0.014)	-0.054*** (0.003)	-0.061*** (0.004)
Obs.	37435	37435	37435	37435	37435	37435	37435	37435
Adj. R2	0.390	0.395	0.493	0.493	0.396	0.405	0.491	0.494

Table 10. Validation of the Create Dimension

This table reports validation tests for the Create dimension. The dependent variables are R&D intensity (R&D/AT) and the natural logarithm of one plus annual patents (ln(1+Patent)). Columns (1)-(4) use the segment-based CVF measures, and columns (5)-(8) use the Fang et al. (2023) benchmark measures labeled with the _fang suffix. For each measure, the first column for a dependent variable includes only the focal culture dimension, and the second includes the four CVF dimensions jointly. All specifications include the control variables shown in the table, industry fixed effects, and fiscal-year fixed effects. Firm-clustered standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Variables	(1) R&D/AT	(2) R&D/AT	(3) ln(1+Patent)	(4) ln(1+Patent)	(5) R&D/AT	(6) R&D/AT	(7) ln(1+Patent)	(8) ln(1+Patent)
Collaborate		0.070** (0.033)		-1.609*** (0.412)				
Compete		-0.247*** (0.018)		-0.874*** (0.237)				
Create	0.174*** (0.021)	0.371*** (0.024)	3.349*** (0.315)	4.300*** (0.345)				
Control		-0.018 (0.019)		-0.177 (0.278)				
Collaborate_fang						3.273*** (0.405)		26.084*** (4.433)
Compete_fang						-1.710*** (0.177)		12.688*** (2.741)
Create_fang					0.455 (0.318)	0.630** (0.317)	27.087*** (5.241)	16.612*** (5.044)
Control_fang						2.075*** (0.407)		1.472 (5.082)
ln(Market cap)	-0.009*** (0.001)	-0.008*** (0.001)	0.321*** (0.012)	0.323*** (0.012)	-0.009*** (0.001)	-0.007*** (0.001)	0.328*** (0.012)	0.323*** (0.012)
Leverage	-0.015** (0.006)	-0.012* (0.006)	-0.093 (0.069)	-0.081 (0.069)	-0.019*** (0.006)	-0.018*** (0.006)	-0.160** (0.070)	-0.139** (0.070)
Cash/Net assets	0.014*** (0.001)	0.013*** (0.001)	0.021*** (0.004)	0.018*** (0.004)	0.014*** (0.001)	0.013*** (0.001)	0.019*** (0.004)	0.022*** (0.004)
Dividend yield	-0.214*** (0.027)	-0.209*** (0.026)	0.800** (0.406)	0.831** (0.406)	-0.228*** (0.027)	-0.223*** (0.027)	0.569 (0.413)	0.603 (0.409)
ln(Firm age)	-0.010*** (0.001)	-0.010*** (0.001)	0.212*** (0.022)	0.216*** (0.022)	-0.011*** (0.001)	-0.011*** (0.001)	0.195*** (0.022)	0.196*** (0.022)
Constant	0.142*** (0.006)	0.165*** (0.006)	-2.152*** (0.107)	-2.025*** (0.110)	0.152*** (0.007)	0.138*** (0.007)	-2.092*** (0.114)	-2.366*** (0.131)
Obs.	37436	37436	37436	37436	37436	37436	37436	37436
Adj. R2	0.420	0.434	0.417	0.420	0.416	0.425	0.408	0.412

Table 11. Validation of the Control Dimension

This table reports validation tests for the Control dimension. The dependent variables are the cash effective tax rate (Cash ETR) and an indicator for error-restatement disclosure (Error Filing). Columns (1)-(4) use the segment-based CVF measures, and columns (5)-(8) use the Fang et al. (2023) benchmark measures. For each measure, the first column for a dependent variable includes only the focal culture dimension, and the second includes the four CVF dimensions jointly. All specifications include the control variables shown in the table, industry fixed effects, and fiscal-year fixed effects. Firm-clustered standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Variables	(1) Cash ETR	(2) Cash ETR	(3) Restat Filing	(4) Restat Filing	(5) Cash ETR	(6) Cash ETR	(7) Restat Filing	(8) Restat Filing
Collaborate		-0.255*** (0.096)		-0.377 (0.814)				
Compete		-0.047 (0.051)		0.448 (0.408)				
Create		0.042 (0.063)		-1.296** (0.523)				
Control	0.137*** (0.053)	0.205*** (0.067)	1.639*** (0.408)	1.574*** (0.511)				
Collaborate_fang						36.545 (54.606)		-3.002 (8.115)
Compete_fang						-21.427 (19.846)		-4.281 (3.841)
Create_fang						136.768 (147.209)		3.230 (7.794)
Control_fang					13.805 (37.613)	3.708 (28.980)	7.999 (8.497)	8.595 (8.529)
ln(Market cap)	-0.003* (0.002)	-0.003* (0.002)	-0.068*** (0.012)	-0.067*** (0.012)	-0.027 (0.070)	-0.042 (0.084)	-0.067*** (0.012)	-0.065*** (0.012)
Leverage	-0.050*** (0.015)	-0.050*** (0.015)	0.489*** (0.105)	0.459*** (0.107)	-3.888 (3.507)	-3.787 (3.391)	0.491*** (0.105)	0.487*** (0.105)
Cash/Net assets	-0.016*** (0.003)	-0.016*** (0.003)	-0.066*** (0.014)	-0.064*** (0.014)	0.097 (0.152)	0.049 (0.166)	-0.070*** (0.014)	-0.072*** (0.015)
Dividend yield	0.028 (0.118)	0.025 (0.118)	-1.527* (0.799)	-1.629** (0.807)	16.900 (14.596)	17.039 (14.778)	-1.503* (0.797)	-1.512* (0.799)
ln(Firm age)	0.005 (0.004)	0.005 (0.004)	-0.052 (0.033)	-0.057* (0.032)	0.895 (0.876)	0.901 (0.883)	-0.052 (0.033)	-0.052 (0.033)
Constant	0.261*** (0.016)	0.266*** (0.016)	-2.495*** (0.475)	-2.483*** (0.471)	-1.849 (2.177)	-2.624 (3.097)	-2.367*** (0.482)	-2.330*** (0.488)
Obs.	24008	24008	37154	37154	24008	24008	37154	37154

Appendix B: Additional Analysis Material

Table B1: Benchmark performance of candidate models

In this table, BBH refers to BigBench Hard (Suzgun et al. 2022), and MuSR denotes Multistep Soft Reasoning (Sprague et al. 2023). Boldface values indicate the best performance within each benchmark. Notably, DeepSeek-R1-Distill-Qwen-14B attains the highest score on MuSR, which is the benchmark most closely aligned with our long-context, multistep classification task.

Model	BBH	MuSR
DeepSeek-R1-Distill-Qwen-14B	40.69%	28.71%
DeepSeek-R1-Distill-Qwen-32B	17.15%	16.14%
Qwen/QwQ-32B	2.87%	11.05%
Qwen2.5-14B-Instruct	48.36%	10.16%
Qwen2.5-32B-Instruct	56.49%	13.50%
Microsoft Phi-4	52.58%	23.72%
Mistral-Small-24B-Base-2501	48.54%	10.89%

Table B2: Examples of CVF Components Identified in the Earnings Call Discourse

CVF Components Present	Segment Text	Synthesizer Agent Reasoning	Full Agentic Reasoning Trail
Adhocracy	Early results are promising, but we believe that there's even more potential to refine these features to drive even stronger engagement. The team is continuing to iterate on both Spotlight Drops and Explore, aiming to make it easier for users to get better matches with less effort. By broadening the Explore tab, we believe that these new titles will focus on shared interest and intention and creating a more dynamic tailored experience that resonates, especially with younger users and women.	The text segment primarily reflects the Adhocracy culture, as indicated by the emphasis on innovation and the pursuit of new opportunities. The team's focus on iterating and refining features to create a dynamic and tailored user experience aligns with the entrepreneurial and flexible nature of Adhocracy. While the Market Agent initially considered the strategic targeting of user segments, the absence of explicit competitive language led to a 'NO' verdict. The Clan and Hierarchy cultures were also not strongly represented, as there was no mention of teamwork, employee focus, or hierarchical structures. Thus, the strongest evidence supports the presence of Adhocracy Culture in the segment.	[Clan Agent] Verdict: NO Observation: The text mentions the team is iterating on features to improve user engagement, which relates to the idea of refining processes for better outcomes. It emphasizes creating a tailored experience that resonates with specific user groups, indicating a focus on user satisfaction and personalization. The mention of shared interests and intentions suggests an effort to foster connections and common ground among users. However, there is no explicit reference to collaboration, teamwork, or employee involvement, which are key aspects of Clan Culture. The focus is more on user experience and feature development rather than organizational cohesion or employee well-being. Reasoning: The text focuses on user engagement and feature development but does not explicitly mention teamwork, employee involvement, or organizational cohesion. There are no references to employee well-being, shared goals, or collaborative environments, which are key aspects of Clan Culture. The observations from the Clan Agent highlight a lack of evidence related to employee focus or participative management, making it difficult to align the segment with the Clan Culture definition. [Adhocracy Agent] Verdict: YES Observation: The text mentions the team is continuing to iterate on features to improve user engagement, indicating a focus on refining and enhancing offerings. It references the development of new titles aimed at creating a more dynamic and tailored experience, particularly targeting younger users and women. The emphasis on innovation and the pursuit of new opportunities aligns with the Adhocracy culture's preference for flexibility and future possibilities. The mention of 'shared interest and intention' suggests a focus on differentiation and user-centric innovation. The goal of making it easier for users to get better matches with less effort reflects an entrepreneurial approach to problem-solving. Reasoning: The text mentions the team iterating on features to improve user engagement, which indicates a focus on innovation and refinement. It references developing new titles aimed at creating a dynamic and tailored experience, particularly targeting specific user groups, which aligns with the Adhocracy culture's emphasis on differentiation and future possibilities. Additionally, the mention of fostering shared interests and intentions suggests an entrepreneurial approach to meeting user needs, further supporting the presence of Adhocracy Culture. [Market Agent] Verdict: NO Observation: The text mentions 'promising' early results and the potential for 'stronger

			engagement,’ indicating a focus on improving outcomes. It references ‘refining features’ to enhance user experience, suggesting a strategic approach to achieving better performance. The goal of making it ‘easier for users to get better matches with less effort’ implies a focus on efficiency and effectiveness. The emphasis on ‘shared interest and intention’ and creating a ‘dynamic tailored experience’ highlights the importance of meeting user needs effectively. The mention of targeting ‘younger users and women’ suggests a strategic approach to market segments, aligning with competitive strategies. Reasoning: The text focuses on refining features to improve user engagement and create a tailored experience, which touches on efficiency and user satisfaction. However, there is no explicit mention of competition, market dominance, or beating rivals. While there is a strategic approach to targeting specific user segments, the language does not emphasize winning, maximizing market share, or driving productivity to outperform competitors. The absence of direct references to competitive positioning or aggressive growth targets makes it difficult to align the text with the Market Culture definition. [Hierarchy Agent] Verdict: NO Observation: The text segment discusses the refinement of features to enhance user engagement, particularly focusing on ‘Spotlight Drops’ and ‘Explore’ functionalities. It emphasizes making these features more efficient for users by reducing effort and tailoring experiences. The mention of creating a ‘more dynamic tailored experience’ suggests a focus on improving user interaction, which aligns with operational efficiency. However, there is no explicit reference to hierarchy, control, or standardized procedures. The absence of terms like ‘management layers,’ ‘rules,’ or ‘job descriptions’ indicates that Hierarchy Culture elements are not present in the text. Reasoning: The text segment discusses feature refinement and user experience improvements but does not explicitly mention hierarchy, control, or standardized procedures. There are no references to multiple layers of management, strict rules, or clearly defined job descriptions. The focus is on user engagement and dynamic experiences rather than internal processes or control mechanisms, which are key aspects of Hierarchy Culture.
Market	So – and I think that speaks to despite the price increases that we’ve taken, the attrition has remained in line with Classic Card back when the Classic Card was \$10. And so it’s quite a testament to the value that people get with the Black Card. And we’ve talked about, you’ve probably heard us talk about this Megan that when we advertise the lion’s, I keep using lion’s share, I’ll use a different phrase, but the vast majority of our marketing dollars go to support the Classic Card price at \$10 and now still a heck of a value at \$15. But when people sign up now more	The text primarily focuses on pricing strategies and customer behavior in a competitive market. The Market Agent identified elements such as price increases, strategic allocation of marketing dollars, and a focus on maximizing profitability, which align with Market Culture. The other agents noted the absence of key cultural elements like employee engagement (Clan), innovation (Adhocracy), and hierarchical structures (Hierarchy). The emphasis on customer value and market share reflects a results-driven approach characteristic of Market Culture.	[Clan Agent] Verdict: NO Observation: The text discusses the value perceived by customers in the Black Card, noting that despite price increases, attrition remains low. It mentions that the majority of marketing dollars support the Classic Card, indicating a focus on customer value. The text also highlights that a significant portion of sign-ups are for the higher-priced Black Card, suggesting customer preference. However, there is no explicit mention of employee involvement, teamwork, or organizational cohesion, which are key aspects of Clan Culture. The focus is on customer behavior and pricing strategies rather than internal company culture or employee well-being. Reasoning: The text segment focuses on customer pricing strategies and product value without mentioning employee involvement, teamwork, or organizational culture. The Clan Culture requires elements like employee engagement, shared goals, and participative management, which are absent here. Observations from all agents highlight customer behavior and marketing efforts, not internal company culture aspects.

	online, way more online than we had seen historically, six out of 10 or plus are taking the higher priced Black Card even though they probably thought they were signing up for used to be \$10 and now \$15 and they're signing up for \$24.99.		[Adhocracy Agent] Verdict: NO Observation: The text discusses price increases for the Classic Card and Black Card, noting that despite the higher cost, there is a shift in customer preference towards the Black Card. It mentions that a significant portion of sign-ups are now online, with six out of ten choosing the Black Card. The focus is on marketing strategies and customer value, particularly how the Black Card is perceived as a better deal despite its higher price. There is an emphasis on customer behavior and the effectiveness of marketing efforts in driving product adoption. The discussion centers around financial aspects and customer preferences, with no explicit mention of innovation, future-oriented strategies, or flexible structures associated with Adhocracy Culture. Reasoning: The text focuses on pricing strategies, customer behavior, and marketing efforts but does not mention innovation, R&D, or future-oriented initiatives. There is no indication of flexibility, creativity, or entrepreneurial activities. The emphasis is on maintaining customer value and market share rather than pioneering new opportunities or fostering an agile environment. [Market Agent] Verdict: YES Observation: The text discusses price increases for the Classic Card and Black Card, indicating a focus on maintaining market share and customer value. It mentions the allocation of marketing dollars to support the Classic Card price, suggesting strategic control over pricing. The emphasis on online sign-ups and the higher-priced Black Card reflects an understanding of customer behavior in a competitive market. The reference to 'lion's share' of marketing dollars highlights prioritization of certain products to drive profitability. The mention of beating competitors by offering a 'heck of a value' at increased prices underscores a focus on maximizing market share and customer satisfaction. Reasoning: The text discusses price increases and the shift in customer preference towards the higher-priced Black Card, indicating a strategic focus on maximizing profitability. The emphasis on online sign-ups and the allocation of marketing dollars to support product pricing reflects a competitive approach to maintaining market share. The mention of 'beating competitors' by offering value despite higher prices underscores a results-driven, market-focused strategy. [Hierarchy Agent] Verdict: NO Observation: The text discusses the pricing strategy for Classic Card and Black Card, focusing on how price increases have affected attrition rates. It mentions the allocation of marketing dollars to support the Classic Card price, indicating a structured approach to financial management. The emphasis on online sign-ups and the shift towards higher-priced Black Card products suggests a strategic focus on operational efficiency and customer acquisition. The mention of standardized pricing and structured marketing strategies reflects a controlled and organized approach to maintaining product value. The discussion of customer behavior in relation to pricing changes implies a systematic analysis of market responses. Reasoning: The text discusses pricing strategies and customer behavior but does not
--	---	--	--

			explicitly mention operational processes, standardization, or hierarchical structures. While there is a focus on efficiency in customer acquisition, there is no clear evidence of Hierarchy Culture elements like strict rules, standardized procedures, or a clear organizational structure.
Hierarchy	I'd say our plan kind of on a unit-by-unit basis is it's been consistent over the last few years and is still what we believe is the right balance of kind of how to develop our assets going forward. Just as a reminder, we are kind of very specific to the different areas we're developing and what the rock needs. There are some DSUs, a lot of them on the Texas side, where you need to go co complete kind of all the different benches, and that's the strategy that we execute there. As we move to New Mexico, there are some benches that need to be co completed, but others that have plenty of height separation or frac barriers that allow us to break different zones into different development packages.	The text segment primarily reflects a Hierarchy culture. The consistent plan over the years and structured approach to asset development indicate a preference for stability and control. The emphasis on standardized strategies and specific operational procedures highlights a controlled and organized method. The mention of detailed procedures and consistent application of strategies aligns with maintaining a stable operational framework. Other cultures like Clan, Adhocracy, and Market were not strongly evidenced, as the text lacked references to collaboration, innovation, competition, or market focus. The absence of these elements supports the conclusion that Hierarchy is the predominant culture present.	[Clan Agent] Verdict: NO Observation: The text discusses a consistent plan for developing assets on a unit-by-unit basis, focusing on specific areas and the needs of the 'rock.' It mentions strategies like co-completing benches in Texas and New Mexico, with attention to height separation and frac barriers. However, there are no references to themes such as staff involvement, consensus-building, morale, open feedback systems, or employee well-being. The focus is on operational strategies and geological specifics rather than organizational culture or human resource development. Reasoning: The text segment focuses on operational strategies for asset development, specifically discussing unit-by-unit plans and geological considerations. There are no mentions of staff involvement, employee development, or collaborative culture elements. The observations from all agents indicate a lack of Clan Culture themes, with the Clan Agent noting no references to teamwork, morale, or employee well-being. The emphasis is on technical and regional strategies rather than organizational culture. [Adhocracy Agent] Verdict: NO Observation: The text discusses a consistent plan over the years, focusing on the development of assets, specific areas, and strategies like co-completing benches in Texas and New Mexico. It emphasizes balance, specific needs of different regions, and the execution of strategies tailored to local conditions. There is a focus on resource allocation and operational efficiency, with mention of separating zones into different development packages. However, there are no explicit references to innovation, future possibilities, flexibility, or risk-taking, which are key aspects of Adhocracy Culture. Reasoning: The text focuses on consistent operational strategies and resource management without mentioning innovation, flexibility, or future-oriented initiatives. The Adhocracy Culture requires elements like R&D, entrepreneurial mindset, or agile approaches, which are absent here. Observations from all agents highlight a lack of these key indicators. [Market Agent] Verdict: NO Observation: The text discusses a consistent plan over the years, focusing on developing assets and balancing strategies. It mentions specific areas like Texas and New Mexico, indicating a localized approach. The emphasis is on executing strategies tailored to different regions, such as co-completing benches or using frac barriers. There is a focus on optimizing development packages for efficiency. The language suggests a methodical and controlled approach to resource management and expansion. Reasoning: The text focuses on operational strategies and geological specifics without mentioning competition, market dominance, or beating rivals. There are no references to maximizing profitability, market share, or driving productivity to satisfy customers or shareholders. The emphasis is on consistent asset development and localized execution, which aligns more with stability and control rather than a competitive, results-driven market

			culture. [Hierarchy Agent] Verdict: YES Observation: The text segment discusses a consistent plan over the years, indicating a preference for stability. It mentions specific strategies for different areas, suggesting a structured approach. The emphasis on detailed procedures for asset development reflects a focus on operational processes. The mention of standardized strategies and specific requirements for different regions highlights a controlled and organized method. The consistent application of these strategies over time aligns with the idea of maintaining a stable and reliable operational framework. Reasoning: The text segment discusses a consistent plan over the years, indicating a preference for stability. It mentions specific strategies for different areas, suggesting a structured approach. The emphasis on detailed procedures for asset development reflects a focus on operational processes. The mention of standardized strategies and specific requirements for different regions highlights a controlled and organized method. The consistent application of these strategies over time aligns with the idea of maintaining a stable and reliable operational framework.
Market, Hierarchy	So the first part is our leasing strategy when we are signing leases is to drive all of the elements of revenues. And so we are getting increased base rents and we are getting increased expense recoveries from our tenants largely on a fixed TAM basis. So that's why you're seeing the growth on the revenue side in both base minimum and in fixed TAM. On the expense side of the house, we continue to be a very operating efficiency and seeking ways to minimize as much of our expenses as possible. We've talked a little bit about the seasonality of our expense load this year relative to last year, where we expected a larger OpEx level in the second half of the year.	The text segment indicates a strong emphasis on revenue growth and operational efficiency, which aligns with both Market and Hierarchy Cultures. The Market Agent highlights the focus on maximizing financial outcomes and achieving hard targets, such as increasing base rents and expense recoveries. The Hierarchy Agent notes the importance of standardized procedures, efficiency, and cost reduction, reflecting a controlling and structured approach. While the Clan and Adhocracy Agents found no evidence of collaboration or innovation, the presence of Market and Hierarchy elements suggests a culture focused on external competition and internal control. The absence of employee-focused or innovative practices supports the exclusion of Clan and Adhocracy Cultures. The combined emphasis on financial performance and operational efficiency underscores the dual presence of Market and Hierarchy Cultures in the described strategy.	[Clan Agent] Verdict: NO Observation: The text discusses leasing strategies focused on revenue growth and operational efficiency. It mentions increasing base rents, expense recoveries, and minimizing expenses. There is no mention of employee involvement, teamwork, or shared goals, which are key aspects of Clan Culture. The emphasis is on financial performance and cost management rather than human resource development or organizational cohesion. Reasoning: The text segment focuses on leasing strategies, revenue growth, and operational efficiency. There is no mention of employee involvement, teamwork, or shared goals, which are key aspects of Clan Culture. The emphasis is on financial performance and cost management rather than human resource development or organizational cohesion. The observations from all agents, including the Clan Agent, highlight the absence of any elements related to employee well-being, collaboration, or participative management, which are essential for Clan Culture classification. [Adhocracy Agent] Verdict: NO Observation: The text discusses a leasing strategy focused on increasing revenues through base rents and expense recoveries on a fixed TAM basis. It mentions operating efficiency and minimizing expenses, particularly noting seasonality in OpEx levels. There is an emphasis on structured approaches to revenue growth and cost management, which contrasts with the dynamic and innovative nature typically associated with Adhocracy Culture. No specific references to innovation, future possibilities, or entrepreneurial activities are present in the text. Reasoning: The text focuses on leasing strategies, revenue growth, and cost management, which are more aligned with Market and Hierarchy Cultures. There is no mention of innovation, R&D, or future-oriented initiatives characteristic of Adhocracy Culture. The emphasis is on financial performance and operational efficiency rather than flexibility or

			creativity. [Market Agent] Verdict: YES Observation: The text mentions a leasing strategy focused on increasing revenues, specifically noting increased base rents and expense recoveries on a fixed TAM basis. This indicates a focus on maximizing financial outcomes. The emphasis on operating efficiency and minimizing expenses suggests a goal of improving productivity. The reference to market conditions and tenant behavior implies an external focus on the marketplace. The discussion of revenue growth and expense management aligns with the idea of achieving hard targets. The mention of tenant expense recoveries and fixed terms reflects a competitive approach to managing external relationships. Reasoning: The text discusses a leasing strategy focused on increasing revenues through base rents and expense recoveries, indicating a focus on financial outcomes. It emphasizes operating efficiency and minimizing expenses, suggesting a goal of improving productivity. The reference to market conditions and tenant behavior implies an external focus on the marketplace, aligning with the Market Culture’s emphasis on competitiveness and achieving hard targets. [Hierarchy Agent] Verdict: YES Observation: The text mentions a focus on operational processes, specifically in leasing strategy and revenue growth. It highlights the importance of efficiency and minimizing expenses, which aligns with the controlling aspects of Hierarchy Culture. The emphasis on standardized procedures, such as fixed TAM basis and operational efficiency, suggests a preference for stability and control. The mention of seeking ways to minimize expenses further indicates a focus on logistical efficiency and cost reduction. Additionally, the structured approach to analyzing both revenue and expense sides of the business reflects a hierarchical, rule-based mindset. Reasoning: The text segment discusses a focus on operational processes, specifically in leasing strategy and revenue growth. It highlights the importance of efficiency and minimizing expenses, which aligns with the controlling aspects of Hierarchy Culture. The emphasis on standardized procedures, such as fixed TAM basis and operational efficiency, suggests a preference for stability and control. Additionally, the structured approach to analyzing both revenue and expense sides of the business reflects a hierarchical, rule-based mindset.
Clan, Market, Hierarchy	So we’ll use both the brand positioning and the communication of our evolution from strength in our mix and format while also highlighting the price value that we offer and making sure that that’s also well communicated in the marketing messaging. So I’d say that’s an urgency, I’d say, underpinning all of it. I touched on a couple of the new roles. Making sure that we’ve got the	The text exhibits elements of Clan, Market, and Hierarchy cultures. Clan culture is present due to the focus on building a strong, cohesive team and collaboration with franchisees. Market culture is evident through the emphasis on brand positioning, competitive strategies, and maximizing market share. Hierarchy culture is reflected in the structured leadership roles and	[Clan Agent] Verdict: YES Observation: The text discusses the importance of building a ‘blue-ribbon team’ and the recruitment of key roles such as CMO and CDO, indicating a focus on team composition. It mentions the support received from franchisees, highlighting collaboration. The emphasis on member experience and franchisee economics suggests a people-focused approach. The term ‘blue-ribbon team’ implies a commitment to assembling a strong, cohesive group. The collaborative efforts between the organization and franchisees are noted as encouraging. Reasoning: The text emphasizes building a ‘blue-ribbon team’ and recruiting key roles, indicating a focus on team composition. It highlights collaboration with franchisees and a

	right team, and I call it the blue-ribbon team, making sure we're building our blue-ribbon team, I'm leaning in very heavily on the recruitment of the CMO and the CDO, we're making excellent progress there and have really strong slates of candidates. So that's a focus between now and year end. And you heard me talk about the execution already on the things that are affecting member experience. So getting the format optimization, floor plan optimization, right. The support that we've received from our franchisees in that endeavor has been really quite encouraging. And I think again as I mentioned in my remarks, it's about delivering that in a way that's also, that's accretive to the member experience and also favorable from a unit economic standpoint and an operating cost standpoint for our franchisees. And as it relates to accelerating growth, driving joins, enhancing, increasing membership, enhancing the member experience and being mindful of franchisee economics, those are the things that culminate in our ability to accelerate growth.	standardized operational processes. The absence of Adhocracy culture is due to a lack of innovation and future-oriented initiatives discussed in the text.	people-focused approach, aligning with Clan Culture's emphasis on teamwork and employee well-being. The mention of shared goals and cohesive group assembly further supports the presence of Clan Culture elements. [Adhocracy Agent] Verdict: NO Observation: The text discusses the importance of communication in marketing, the recruitment of high-level executives like the CMO and CDO, and the optimization of format and floor plans to enhance member experience. It also mentions the support from franchisees and the focus on unit economics and operating costs. While there is a focus on strategic planning and execution, there are no specific references to innovation, future possibilities, or a preference for flexibility and dynamism, which are key aspects of Adhocracy Culture. Reasoning: The text focuses on strategic planning, team recruitment, and operational optimizations such as format and floor plan changes. While there is a mention of innovation in recruitment and team-building, there is no explicit reference to innovation, future-oriented initiatives, or a preference for flexibility and dynamism. The emphasis is more on execution, efficiency, and structured processes, which aligns more with Market or Hierarchy cultures rather than Adhocracy. The lack of discussion on R&D, new product launches, or entrepreneurial activities further supports the conclusion that Adhocracy Culture is not present. [Market Agent] Verdict: YES Observation: The text discusses the importance of brand positioning and communication to highlight price value, which relates to external focus and market positioning. It mentions the recruitment of key roles like CMO and CDO, indicating a strategic focus on building a strong team for competitive advantage. The emphasis on optimizing format and floor plans, supported by franchisees, suggests a focus on operational efficiency and control. The goal of accelerating growth, increasing membership, and improving member experience ties to maximizing market share and profitability. The mention of favorable unit economics and operating costs for franchisees aligns with delivering results through aggressive transactions in a competitive environment. Reasoning: The text segment discusses the importance of brand positioning and communication to highlight price value, which relates to an external focus on the marketplace. The recruitment of key roles like CMO and CDO indicates a strategic focus on building a strong team for competitive advantage. The emphasis on optimizing format and floor plans, supported by franchisees, suggests a focus on operational efficiency and control. The goal of accelerating growth, increasing membership, and improving member experience ties to maximizing market share and profitability. The mention of favorable unit economics and operating costs for franchisees aligns with delivering results through aggressive transactions in a competitive environment. [Hierarchy Agent] Verdict: YES Observation: The text mentions the importance of having the right team, specifically referring to the recruitment of the CMO and CDO, indicating a focus on building a strong leadership structure. The emphasis on standardized processes is evident in the discussion of
--	---	---	---

			format optimization and floor plan optimization, which suggests a preference for operational efficiency. The mention of clear roles and responsibilities, such as the blue-ribbon team, aligns with the idea of structured job descriptions. The focus on predictable outcomes, like improving member experience and franchisee economics, reflects a desire for stability and control. The structured approach to recruitment and process optimization highlights a preference for maintaining order through defined procedures. Reasoning: The text emphasizes the importance of building a strong leadership structure through the recruitment of key roles like CMO and CDO, which aligns with the Hierarchy Culture's focus on clear roles and structured leadership. The discussion of format and floor plan optimization indicates a preference for standardized processes, a key aspect of Hierarchy Culture. Additionally, the emphasis on predictable outcomes such as improving member experience and franchisee economics reflects a desire for stability and control, further supporting the presence of Hierarchy Culture.
--	--	--	--

Table B3: Confusion matrices

Model	Run	True Negative (TN)	False Positive (FP)	False Negative (FN)	True Positive (TP)
Original Model	Main Run	150	12	0	138
DeepSeek-R1-14B	1	150	12	2	136
	2	148	14	4	134
	3	152	10	3	135
	4	151	11	5	133
	5	149	13	4	134

Table B4: Pairwise Inter-Run Agreement

Subset Size (k)	Runs Included	Average Pairwise κ	Interpretation
2	R1, R2	0.947	Almost Perfect
	R1, R3	0.953	Almost Perfect
	R1, R4	0.920	Almost Perfect
	R1, R5	0.967	Almost Perfect
	R2, R3	0.913	Almost Perfect
	R2, R4	0.933	Almost Perfect
	R2, R5	0.940	Almost Perfect
	R3, R4	0.900	Almost Perfect
	R3, R5	0.920	Almost Perfect
	R4, R5	0.913	Almost Perfect
3	R1, R2, R3	0.938	Almost Perfect
	R1, R2, R4	0.933	Almost Perfect
	R1, R2, R5	0.951	Almost Perfect
	R1, R3, R4	0.924	Almost Perfect
	R1, R3, R5	0.947	Almost Perfect
	R1, R4, R5	0.933	Almost Perfect
	R2, R3, R4	0.916	Almost Perfect
	R2, R3, R5	0.924	Almost Perfect
	R2, R4, R5	0.929	Almost Perfect
	R3, R4, R5	0.911	Almost Perfect
4	R1, R2, R3, R4	0.928	Almost Perfect
	R1, R2, R3, R5	0.940	Almost Perfect
	R1, R2, R4, R5	0.937	Almost Perfect
	R1, R3, R4, R5	0.929	Almost Perfect
	R2, R3, R4, R5	0.920	Almost Perfect

Table B5: Yes Segment % for CVF Wordlist (based on Fang et al., 2023)

Fang et al. (2023) Keyword	CVF Component	Number of Yes	Number of No	Number of Total	Yes %
partic*	Clan	131665	258713	390378	33.73%
people*	Clan	185395	155482	340877	54.39%
help*	Clan	141253	172918	314171	44.96%
team*	Clan	179056	67347	246403	72.67%
partner*	Clan	104924	81003	185927	56.43%
capab*	Clan	70796	34986	105782	66.93%
contribut*	Clan	32305	59660	91965	35.13%
commit*	Clan	46287	40432	86719	53.38%
commun*	Clan	44444	38974	83418	53.28%
engag*	Clan	40052	20804	60856	65.81%
employ*	Clan	42259	18246	60505	69.84%
life*	Clan	25772	32762	58534	44.03%
hir*	Clan	28519	13880	42399	67.26%
involv*	Clan	20272	20630	40902	49.56%
train*	Clan	24570	12263	36833	66.71%
reten*	Clan	13200	9772	22972	57.46%
collab*	Clan	14928	3397	18325	81.46%
retain*	Clan	8980	7639	16619	54.03%
responsib*	Clan	8723	6395	15118	57.70%
dialog*	Clan	7686	6866	14552	52.82%
social*	Clan	9420	5075	14495	64.99%
skill*	Clan	8963	4357	13320	67.29%
human*	Clan	6316	5668	11984	52.70%
loyal*	Clan	8871	2526	11397	77.84%
parent*	Clan	2357	4178	6535	36.07%
collectiv*	Clan	3152	2693	5845	53.93%
cooperat*	Clan	3754	2030	5784	64.90%
mutual*	Clan	2554	2943	5497	46.46%
consens*	Clan	1720	3660	5380	31.97%
competen*	Clan	3712	1510	5222	71.08%
coordin*	Clan	2790	1562	4352	64.11%
willingness*	Clan	2383	1825	4208	56.63%
empower*	Clan	1590	327	1917	82.94%
decentr*	Clan	1002	125	1127	88.91%
cohes*	Clan	528	128	656	80.49%
mentor*	Clan	458	114	572	80.07%
teamwork*	Clan	313	8	321	97.51%
facilitator*	Clan	129	103	232	55.60%
workgroup*	Clan	78	82	160	48.75%

longterm*	Clan	14	17	31	45.16%
interper*	Clan	16	8	24	66.67%
longlast*	Clan	1	1	2	50.00%
market*	Market	580024	494597	1074621	53.97%
growth*	Market	276458	352298	628756	43.97%
customer*	Market	329931	247240	577171	57.16%
strong*	Market	169242	217913	387155	43.71%
revenue*	Market	132421	252273	384694	34.42%
share*	Market	174545	146731	321276	54.33%
result*	Market	96177	147512	243689	39.47%
compet*	Market	195161	45937	241098	80.95%
value*	Market	131940	108840	240780	54.80%
perform*	Market	100841	128276	229117	44.01%
position*	Market	112562	105399	217961	51.64%
hard*	Market	78817	101757	180574	43.65%
acquis*	Market	83378	80027	163405	51.03%
target*	Market	63176	81046	144222	43.80%
profit*	Market	69357	70252	139609	49.68%
win*	Market	75199	64005	139204	54.02%
client*	Market	68433	47936	116369	58.81%
expand*	Market	69741	46352	116093	60.07%
challeng*	Market	54024	58718	112742	47.92%
fast*	Market	49515	45618	95133	52.05%
pressur*	Market	34401	56693	91094	37.76%
achiev*	Market	35745	38956	74701	47.85%
goal*	Market	39384	34329	73713	53.43%
aggress*	Market	38598	18229	56827	67.92%
acquir*	Market	27038	26232	53270	50.76%
rapid*	Market	15023	13949	28972	51.85%
speed*	Market	15030	12697	27727	54.21%
analyst*	Market	8943	15838	24781	36.09%
satisf*	Market	11114	7723	18837	59.00%
signal*	Market	7265	11385	18650	38.95%
outsourc*	Market	6207	3983	10190	60.91%
attack*	Market	4449	3320	7769	57.27%
superior*	Market	4284	2885	7169	59.76%
reputat*	Market	3602	1173	4775	75.43%
certain*	Hierarchy	187492	288784	476276	39.37%
cost*	Hierarchy	192808	263836	456644	42.22%
process*	Hierarchy	104938	115231	220169	47.66%
system*	Hierarchy	78027	79938	157965	49.40%
efficien*	Hierarchy	85684	33275	118959	72.03%
detail*	Hierarchy	24550	63507	88057	27.88%

inform*	Hierarchy	28651	54516	83167	34.45%
control*	Hierarchy	49764	26975	76739	64.85%
analys*	Hierarchy	19171	38366	57537	33.32%
productiv*	Hierarchy	37302	17051	54353	68.63%
measur*	Hierarchy	21014	27097	48111	43.68%
standard*	Hierarchy	19557	23776	43333	45.13%
conservat*	Hierarchy	16767	25413	42180	39.75%
cautio*	Hierarchy	15726	24153	39879	39.43%
discipline*	Hierarchy	30179	9405	39584	76.24%
enhance*	Hierarchy	21777	13794	35571	61.22%
outcom*	Hierarchy	13822	19296	33118	41.74%
rule*	Hierarchy	9061	17940	27001	33.56%
regular*	Hierarchy	9629	12121	21750	44.27%
method*	Hierarchy	8011	10075	18086	44.29%
procedur*	Hierarchy	7709	9412	17121	45.03%
administrat*	Hierarchy	6736	9241	15977	42.16%
solv*	Hierarchy	8559	6695	15254	56.11%
logic*	Hierarchy	4372	7009	11381	38.41%
predictab*	Hierarchy	7280	3389	10669	68.24%
document*	Hierarchy	3970	6406	10376	38.26%
fail*	Hierarchy	3539	6390	9929	35.64%
chief*	Hierarchy	3722	4256	7978	46.65%
uniform*	Hierarchy	1112	1453	2565	43.35%
boss*	Hierarchy	610	1896	2506	24.34%
bureaucr*	Hierarchy	495	194	689	71.84%
new*	Adhocracy	332835	319944	652779	50.99%
start*	Adhocracy	210335	336489	546824	38.46%
chang*	Adhocracy	145118	236181	381299	38.06%
develop*	Adhocracy	112749	99336	212085	53.16%
future*	Adhocracy	86624	96526	183150	47.30%
trend*	Adhocracy	53996	110575	164571	32.81%
begin*	Adhocracy	51031	91605	142636	35.78%
init*	Adhocracy	70542	70835	141377	49.90%
creat*	Adhocracy	86709	46653	133362	65.02%
thought*	Adhocracy	46301	83968	130269	35.54%
research*	Adhocracy	19031	52229	71260	26.71%
origin*	Adhocracy	20107	42619	62726	32.06%
innovat*	Adhocracy	53937	4605	58542	92.13%
learn*	Adhocracy	35881	22018	57899	61.97%
idea*	Adhocracy	24556	26523	51079	48.07%
imagin*	Adhocracy	15322	21111	36433	42.06%
ventur*	Adhocracy	8296	12014	20310	40.85%
fantas*	Adhocracy	10235	7685	17920	57.11%

vision*	Adhocracy	9183	3829	13012	70.57%
adapt*	Adhocracy	6828	2993	9821	69.52%
elabor*	Adhocracy	2073	5180	7253	28.58%
experim*	Adhocracy	4390	2570	6960	63.07%
envis*	Adhocracy	2314	3130	5444	42.51%
discontin*	Adhocracy	1314	3837	5151	25.51%
intellect*	Adhocracy	2381	1594	3975	59.90%
pioneer*	Adhocracy	1779	738	2517	70.68%
dream*	Adhocracy	1346	1156	2502	53.80%
entrepre*	Adhocracy	2105	241	2346	89.73%
radic*	Adhocracy	993	1096	2089	47.53%
freedom*	Adhocracy	1171	890	2061	56.82%
rd*	Adhocracy	649	867	1516	42.81%
inventive*	Adhocracy	64	25	89	71.91%
unafra*	Adhocracy	10	2	12	83.33%