



**HKU
BUSINESS
SCHOOL**
港大經管學院



**HKU Jockey Club
Enterprise Sustainability
Global Research Institute**

Hong Kong University Business School:

HKU Jockey Club Enterprise Sustainability
Global Research Institute Paper Series

Dissecting Inconsistencies in Corporate Disclosures: Evidence from Public and Private Earnings Calls

Paper No. 2026/038

Shangrong Jiang

The University of Hong Kong

Chen Lin

The University of Hong Kong

Yutong Sun

Chinese Academy of Sciences

Mingzhu Tai

The University of Hong Kong

Dissecting Inconsistencies in Corporate Disclosures: Evidence from Public and Private Earnings Calls

Shangrong Jiang, Chen Lin, Yutong Sun, Mingzhu Tai*

August 2026

Abstract

Using a novel dataset of paired public and private earnings calls, we examine inconsistencies in corporate disclosures across topic coverage, narrative content, and vocal delivery. We find that managers introduce a non-trivial proportion of new, and potentially material, topics in private meetings. They also demonstrate distinct narrative styles and vocal delivery strategies when addressing the same topics at different meetings. We find that these inconsistencies generate significant price and trading reactions around private meetings. Investors attending private meetings can make more informed trading by integrating multiple dimensions of disclosure inconsistencies rather than evaluating each dimension in isolation.

Keywords: Corporate Disclosure; Earnings Calls; Investor Behaviors; Informed Trading; Large Language Models; Topic Modelling

JEL Code: G14, G30, L22, O16

* Shangrong Jiang is from the Faculty of Business and Economics, The University of Hong Kong, Hong Kong. Chen Lin is from the Faculty of Business and Economics, The University of Hong Kong. Yutong Sun is from the State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences. Mingzhu Tai is from the Faculty of Business and Economics, The University of Hong Kong, Hong Kong. We would like to thank Xavier Giroud, Yongmiao Hong, Kai Li, Zhe Li, Roni Michaely, Lynn Wang, Shouyang Wang, and seminar participants at China International Conference in Finance, Chinese Economic Association (UK/Europe) Conference, Northern Finance Association Meetings, and Symposium on AI-Driven Mathematics and Interdisciplinary Sciences, as well as workshop participants at Central University of Finance and Economics, Columbia University, Hong Kong Polytechnic University, Renmin University, University of Chinese Academy of Sciences, University of Hong Kong, and Xiamen University for helpful comments and suggestions. This research is funded by the Research Grant Council of the Hong Kong Special Administrative Region and the National Natural Science Foundation of China.

1. Introduction

Despite firms providing a variety of public channels that give market participants broad access to corporate information, investors continue to devote substantial time to meeting privately with managers.¹ A large volume of literature shows that, relative to investors who only participate in public interactions, those with access to private meetings are more likely to make informed trading decisions (Bradley et al., 2022; Becht et al., 2026), develop a deeper understanding of firm fundamentals (Green et al., 2014; Cao et al., 2025), and better interpret firms' public disclosure (Li et al., 2023; Tseng et al., 2024). These findings suggest that private meetings may contain valuable information beyond what is available through public disclosures.

However, given the fact that most private meetings were conducted in-person and restricted to approved participants, we still know little about the actual content of these meetings.² This data constraint becomes less binding in recent years as firms and investors increasingly adopt online and hybrid (online and in-person) meeting formats, a trend accelerated by the COVID-19 pandemic and continued afterward because of their convenience and cost efficiency.³ These formats generate records that make private disclosures more observable. Leveraging these records, we employ a novel data source that provides verbatim transcripts and audio recordings of private meetings held by Chinese listed firms to examine inconsistencies in corporate disclosures across public and private meetings. Specifically, we aim to address three questions: what managers discuss in private meetings, whether the information conveyed differs from public disclosures, and how investors incorporate any such differences into their trading decisions?

We take three steps to derive the measures of disclosure inconsistency. First, we employ a novel dataset that comprises 115,357 earnings calls for Chinese listed firms. Our dataset consists of two types of earnings calls. The first type is public earnings calls held by the company in conjunction with

¹ Consultancy Ipreo reports that publicly traded firms engage in more than 100 private meetings with investors annually on average (Park & Soltes, 2018). In addition, investors globally spend more than \$2 billion a year for corporate private access (<https://www.bloomberg.com/news/articles/2017-04-25/bank-research-rules-open-door-to-match-com-for-investors-ceos?embedded-checkout=true>).

² Existing works use the summary reports of private meeting (Bowen et al., 2018; Ru et al., 2025) and meeting notes prepared by meeting participants (Becht et al., 2026) to explore the nature and consequences of private events. However, the actual content of private meetings remains largely inaccessible in the existing literature. Without access to this information, we cannot directly examine what information managers convey in private meetings and how private disclosures differ from their public counterparts.

³ During the COVID-19 pandemic, online and hybrid formats rapidly became dominant in China: more than 99% of over 600 institutional meetings conducted since 2020 were held via online conference (<https://finance.sina.com.cn/stock/s/2020-03-17/doc-iimxyqwa1028269.shtml>). This practice persisted after the pandemic: in 2024, about 90% of Chinese listed firms conducting earnings conference calls used online or hybrid formats due to their convenience and cost efficiency. (https://www.capco.org.cn/xhdt/xhyw/202408/20240815/j_2024081515593300017237088942130178.html)

an earnings release, where most investors can participate in the earnings calls with an easy registration. The second type is private meetings jointly organized by financial institutions and the invited firms to discuss the same earnings release. In contrast to public calls, participation in private calls is restricted to investors whose registrations are reviewed and approved by the hosts. We identify and pair public and private calls delivered by the same firm managers regarding the identical earnings release as the unit of observation.⁴ Our final dataset consists of 20,480 pairs of public and private calls for 976 publicly listed companies in China from the time period of 2020 to 2024.

Second, we use TopicGPT modelling framework to construct the topic taxonomy of our earnings call sample. TopicGPT is a human-centric approach to topic modeling that relies on prompting large language models to perform in-context topic generation and assignment, producing results that more closely match human-annotated ground truth topics than traditional topic modelling approaches (Pham et al., 2023; Lyonnet et al., 2025). We first use TopicGPT to generate comprehensive lists of top- and sub-topics for a sample of earnings calls. We then create a labeled training set by instructing TopicGPT to classify sentences from this subset into top-subtopic categories. We train a multilayer perceptron (MLP) on this labeled sample and apply it to the full sample. Our topic taxonomy consists of 31 top topics and the 176 associated subtopics.

Third, after obtaining the top-subtopic lists for the full sample, we construct disclosure inconsistency measures for each public-private call pair in three aspects: the topic, textual and vocal inconsistencies. For topic inconsistencies, we count the number of new topics discussed in the private call that are absent from the corresponding public call (*#NewTopic*), as well as the number of public-discussed topics that omitted in the private call (*#OmitTopic*). For the topics covered in both meetings, we first measure textual inconsistencies between public and private meetings, including differences in numerical usage ($\Delta Number$), narratives style ($\Delta Narrative$) and text-based sentiment ($\Delta Sentiment$). Utilizing the audio recordings of earnings calls, we further compute vocal inconsistencies, including differences in audio length ($\Delta Length$), listenability ($\Delta Listenability$), vocal valence ($\Delta Valence$) and arousal ($\Delta Arousal$). We examine the consistency between human annotators and our TopicGPT-based framework across alternative prompting strategies and samples. The results indicate that our framework produces disclosure inconsistency measures that closely align with those assigned by human annotators.

⁴ In Figure A1 of Appendix A1, we provide examples of pairs of public and private calls discussing the same earnings release for publicly listed firms in our call sample.

Do managers disclose differently across public and private meetings? We dissect inconsistencies in corporate disclosure using the above features. We first show that the measures of *#NewTopic* and *#OmitTopic* have a mean of 1.238 and 1.813, accounting for 15.03% and 18.76% of all topics discussed in private and public calls, respectively. These statistics suggest that managers in our sample firms disclose a non-trivial proportion of new topics in private calls while omitting a substantial proportion of topics discussed in public calls. The largest values of *#NewTopic* are observed in topics such as *Corporate Governance* and *Research and Development*. Their frequent appearance as new topics in private calls suggests that managers may potentially reserve material information for private audiences.⁵ The largest values of *#OmitTopic* are concentrated in topics such as *Political Alignment*, and *Media Engagement*, suggesting that managers are more inclined to exclude these value-irrelevant discussions from formal public communication. This asymmetric topic structure highlights the critical role of private meetings in conveying incremental information that is valuable for investors' valuation assessments and trading decisions.

Next, we explore the textual inconsistencies between public and private calls. We find that both numerical (*ΔNumber*) and narrative (*ΔNarrative*) inconsistencies are more pronounced in topics that may convey large portions of soft information, such as *Business Strategy* and *Mergers and Acquisitions*, etc. The largest sentiment gaps emerge in risk-sensitive and performance-related topics, including *Profit and Loss*, *Risk Exposure*, *Compliance and Enforcement*. We then identify whether the sentiment inconsistency is driven by managers being more positive in private (private-driven) or in public (public-driven) meetings. We find that most topics are primarily public-driven in tone, indicating that managers reallocate not only what they disclose but also how they emotionally frame information across public and private meetings.

We further decompose the vocal dimension of disclosure inconsistencies. In several topics such as *Debt and Capital Structure* and *Auditing and Internal Controls*, we find that managers speak substantially longer in public calls while the corresponding listenability is relatively lower than in private calls. These results indicate that, when discussing sensitive operational and risk-related topics in public meetings, managers may seek to satisfy extensive disclosure expectations under regulatory and market scrutiny while preserving some ambiguity to audience. We also find that inconsistencies

⁵ Unlike many developed capital markets that impose strict restrictions on selective disclosure, such as the U.K.'s parity-of-information rule and the U.S. Regulation Fair Disclosure (Reg FD), the Chinese disclosure regime does not have an equivalent regulatory framework. As a result, firms in China have greater flexibility to communicate with investors through private meetings.

in vocal valence and arousal are significantly larger than their text-based sentiment counterpart, suggesting that vocal signals may capture layers of managerial emotion and tension that are not fully reflected in textual content.

To gain a better understanding of how disclosure inconsistencies arise, we examine their determinants across earnings call-, manager- and firm-level characteristics. The results show that firm executives are more likely to disclose inconsistently between public-private calls when the earnings calls are held offline, by a single manager or co-hosted by financial institutions affiliated with the listed firms. Managers who are male, in their elder age and with longer experience in their current positions appear to generate higher levels of disclosure inconsistencies. The results further show that a better financial position of the firm is associated with fewer disclosure inconsistencies.

Do disclosure inconsistencies generate market reactions following private meetings? We examine the informativeness of disclosure inconsistencies. We find that our disclosure inconsistency measures are significantly associated with absolute abnormal returns following private earnings. For example, one-standard-deviation increases in *#NewTopic*, predict a 0.153 (*t*-stat: 4.55) increase in the three-day absolute abnormal returns post private meetings. In addition, we find that market responses to disclosure inconsistencies are driven less by single dimensions, and more by the combined interpretation of multiple communication cues. We then regress abnormal share turnover on disclosure inconsistencies and find that they appear to have a more pronounced impact on trading volume than on stock price reactions. One possible explanation is that the incremental information conveyed in private meetings updates investors' beliefs about firm fundamentals, but does not necessarily lead to a common interpretation of that information.

Do private meeting attendees trade on disclosure inconsistencies? By identifying 6,371 public-private call pairs with holdings data for private meeting attendees, we find that disclosure inconsistencies exhibit significant impacts on quarterly changes in attendees' holdings of the meeting firms. Such trading appears to be informed, as disclosure inconsistencies are positively associated with the trading gains of private meeting attendees. In addition, we find that investors attending private meetings make more informed trades by integrating multiple dimensions of disclosure inconsistency rather than evaluating each dimension in isolation. This pattern is consistent with the mosaic theory of investing (Cheynel & Levine, 2020), whereby investors collect and integrate pieces

of information to update their beliefs in a more accurate manner and thus make more informed trading decisions.

We contribute to the literature in the following aspects. First, we extend research on private interactions between firms and investors (Bowen et al., 2018; Frenkel et al., 2020; Gerken & Painter, 2023). Recent studies have begun to explore the nature of information exchanged in private meetings. For example, Ru et al. (2025) show that timely disclosure mandates for private meetings attract immediate analyst attention to peer firms that were not directly visited. Using meeting notes from 4,700 private meetings, Becht et al. (2026) find that private interactions primarily convey soft information through qualitative discussions rather than explicit disclosure of hard information, and that such meetings enhance investors' ability to process information and interpret firm fundamentals. Existing studies rely primarily on summary reports of private meetings or meeting notes from private meeting participants. Given that the actual contents of private meetings remain largely exclusive, one cannot directly examine what information managers convey in private meetings and how such information differs from public disclosures. In this paper, we complement the literature by using a novel dataset of verbatim transcripts and audio recordings from paired public and private earnings calls delivered by the same managers. We provide the first comprehensive evidence on how corporate disclosures differ across public and private meetings in terms of topic coverage, narrative content, and vocal delivery.

Second, we add to the literature examining investors' activities around private meetings (Cicero et al., 2024; Choy et al., 2025; Lee, 2026). Using corporate jet flight to identify private meetings with investors, Bushee et al. (2018) find that these private events are associated with both signed and unsigned changes in local institutional ownership. Bradley et al. (2022) further show that the subsequent trades of private meeting participants are highly informed: a long-short portfolio constructed from their trades generates a return of 202 basis points over a 12-month holding period. These results highlight that private meetings contain valuable information beyond what is available through public disclosures. However, it remains unclear whether and how investors trade on information that is incremental to public disclosures. By identifying private meeting attendees in our earnings call sample, we show that they incorporate information embedded in public-private disclosure inconsistencies into their trades. In addition, rather than evaluating each dimension of these inconsistencies in isolation, investors attending private meetings appear to integrate multiple

dimensions to construct a more complete picture of the firm. This integrated assessment enables them to update their beliefs more accurately and thus make informed trading decisions.

Furthermore, we relate to the burgeoning literature on the applications of large language models in financial markets (Eisfeldt et al., 2023; Fan et al., 2024). Recent studies assess whether LLMs can effectively extract information from unstructured text about complex concepts such as business complexity (Bernard et al., 2026), corporate culture (Li et al., 2026), and finance sentiment (Jha et al., 2026). In this paper, we examine whether LLMs can identify the disclosure inconsistencies of managerial communications across different meeting occasions. We show that LLMs assess the informational content of input texts more like a sophisticated human reader, rather than relying solely on simple heuristics or correlations.

2. Measuring Disclosure Inconsistencies

2.1 Data on Public and Private Earnings Calls

We obtain 115,357 conference calls from Comein Finance for publicly listed companies in Chinese stock markets during the period of January 2020 to December 2024. Comein Finance is a leading provider of real-time corporate conference events in China.⁶ We complement the dataset by deriving earnings call data from the 3C China corporate conference platform in the Wind database.⁷ Our dataset consists of call transcripts and audio recordings for two types of earnings calls. The first type is public earnings calls held by the company itself in conjunction with an earnings release, where investors can participate in the earnings calls with easy registration. The second type is private earnings calls jointly organized by financial institutions and listed firms to discuss the same earnings release. In contrast to public calls, participation in private calls is restricted to investors whose registrations are reviewed and approved by the hosts. The broader adoption of online and hybrid meeting formats, accelerated by the COVID-19 pandemic, makes records of these previously difficult-to-observe private meetings increasingly available. These formats remain widely used because of their convenience and cost efficiency. As a result, our sample begins in January 2020, when online private meetings became widespread to generate sufficient transcript and audio records.

By observing whether the call is publicly accessible or not, we identify 60,829 public calls and 54,528 private calls from our call dataset. In Figure A1 of Appendix A1, we provide examples of

⁶ The details can be found in <https://www.comein.cn/home/index>.

⁷ The official website of 3C conference platform can be found at <https://peacall.com/#/>

public-private call pairs discussing the same earnings release in our sample.⁸ To construct valid measures of disclosure inconsistency, we pair the public and private calls held by firms regarding the identical earnings release as the unit of observation. Each earnings call in our sample contains the presentation section delivered by firm executives (e.g., the Chief Executive Officer, the Investor Relations Officer, or the Board Secretary) and a question-and-answer (Q&A) session with call attendees. We only use the presentation session of the earnings call since the Q&A session may contain information from call attendees. We further restrict the sample to call pairs delivered by the same managers to control for the individual bias of information disclosure. The filtering procedure is summarized in Table A2 of Appendix A1. After filtering, we derive 20,480 public-private call pairs across 976 publicly listed companies in China. In our matched sample, private earnings calls are generally held after the corresponding public earnings calls, and most take place within one week of the public calls.

2.2 Topic Modelling for Earnings Calls

In this section, we construct a comprehensive topic taxonomy to capture the full scope of corporate disclosure, allowing us to analyze disclosure inconsistencies between public and private calls at the topic level. We leverage the TopicGPT modelling to perform in-context topic generation and assignment (Pham et al., 2023). TopicGPT is a human-centric topic modeling approach that employs attention mechanisms and multidimensional logic to assess the narrative content of input texts, as opposed to relying solely on simple heuristics and correlations (de Kok, 2025; Becht et al., 2026). As a result, it demonstrates zero-shot and few-shot capabilities to address problems not explicitly covered during the training process. Existing studies show that TopicGPT’s topic generation and assignment align more closely with human-annotated ground truth topics than those from LDA or other topic modelling methods (Fieberg et al., 2025; Lyonnet et al., 2025).

Specifically, our TopicGPT framework consists of two steps: First, we use GPT-4o to obtain comprehensive lists of top-subtopics from the presentation sections of the earnings calls. Second, we construct a labeled training set by prompting GPT-4o to classify sentences into these top-subtopics for a subset of earnings calls. We then train a multilayer perceptron (MLP) on the labeled sample and extend it to the full sample.

2.2.1 Topic Generation

⁸ The public and private earnings calls are typically held in succession after the earnings release (normally within one week). Refer to Figure A1 in Appendix A1 for the examples.

As a first step, we prompt GPT-4o to generate a set of topics based on input examples. Specifically, given a sentence S from the earnings call dataset and a set of example topics K , the model is instructed either to assign S to an existing topic in K or to generate a new topic that better describes S and append it to K . Initially, K contains a small set of human-written example topics (three in our prompts). For each example, the model receives a topic branch comprising a top-level topic K_t , a subtopic K_s , and the sentence S associated with the top-subtopic pair. These topics serve as few-shot demonstrations of the topic generation format. This iterative procedure encourages newly generated topics to be both distinctive and comparable in specificity to the existing ones. With these inputs, the LLM generates comprehensive top-subtopic lists that capture common themes among the provided sentences. The prompts for topic generation are shown in Table A12 of Appendix A2

Instead of running topic generation over the entire dataset, which would be computationally costly, we apply the process to a carefully constructed subsample. We collect a sample of firms based on the China Securities Regulatory Commission (CRSC) industry codes and 2 time windows into 30 strata ($15 \text{ industries} \times 2 \text{ periods}$). For each time-industry stratum, we randomly sample five earnings calls and input the corresponding transcripts into GPT-4o. The model is then instructed to generate top-subtopic lists that are encountered within the five input call transcripts at the sentence level. As a result, the subsample of 150 call transcripts ($30 \text{ strata} \times 5 \text{ documents}$) serves as our basis for generating topic lists.

We consolidate and reconcile top- and sub-level topics across strata. Specifically, GPT-4o first aggregates top-subtopic lists within each industry. We use Sentence-Transformer embeddings to identify topic pairs with high similarity (cosine similarity ≥ 0.6). We prompt the GPT-4o with two such topic pairs, instructing it to merge near-duplicate pairs where appropriate (The prompts for topic refining and merging are shown in Table A13 of Appendix A2). We then aggregate the refined top-subtopic lists across industries to construct a comprehensive topic taxonomy based on the 150 sampled earnings calls. GPT-4o is instructed to avoid industry-specific topics but rather develop industry independent top-subtopic lists.

2.2.2 Topic Assignment

As a second step, we construct a labeled training dataset using a sample of 300 documents, comprising the initial 150 transcripts and an additional stratified random sample of 150 transcripts. For each sentence in these documents, we instruct GPT-4o to classify it into one of the predefined top-level topics and corresponding subtopics, with the option to assign sentences to “Others” if no

existing subtopic fits. To capture the narrative sentiment in managerial presentations, GPT-4o also labels the sentiment of each sentence on a scale from -1 (most negative) to 1 (most positive). To address topic assignments with incorrect formatting or low quality, we incorporate a self-correction step (Shinn et al., 2023). Specifically, we implement a parser to identify hallucinated or invalid assignments (e.g., incorrect formatting, low quality, out-of-taxonomy labels). We provide the GPT-4o with the identified documents along with the error type and prompt the model to reassign a valid topic. The final output contains the assigned top- and sub-topic label (with digital codes), and a sentiment score produced by GPT-4o. The prompts for topic assignment are shown in Table A14 of Appendix A2.

Note that a document contains approximately 300 sentences on average, and hence we do not apply this labeling procedure to the full population of earnings calls. Instead, we convert each sentence into a textual vector using the OpenAI *text-embedding-3-large* model. OpenAI vector embeddings include 3,072 elements ordered by importance. We condense the dimensionality from 3,072 to 256 for computational efficiency.⁹ We train a multilayer perceptron (MLP) on our labeled dataset using 256-dimensional vector embeddings as input features. We deploy this classifier to generate topic-subtopic classifications and sentiment scores for the entire sample of sentences in the earnings calls. The overall pipeline of our topic modelling methods is provided in Figure A3 of Appendix A2.

2.3 Construction of Disclosure Inconsistency Measures

2.3.1 Topic Inconsistencies

Using the topic taxonomy derived from the TopicGPT framework, we first construct disclosure inconsistencies between public and private calls in terms of topic coverage. We construct the number of new topics (*#NewTopic*) as the number of topics discussed in the private call that do not appear in the corresponding public call. For each public-private call pair, we also calculate the share of sentences labeled by TopicGPT as non-public-call topics in the private call, scaled by the total number of sentences in the private call:

$$\%NewTopic_{i,t} = \frac{1}{S_{i,m,t}} \sum_{s=1}^{S_{i,m,t}} I[s \in K_n \setminus K_m] \quad (1)$$

⁹ OpenAI reports show that, on the Massive Text Embedding Benchmark (MTEB), the *text-embedding-3-large* model truncated to 256 dimensions continues to exhibit strong performance and even outperforms earlier-generation models with significantly higher dimensionality, i.e., the dimensionality of embeddings generated by the *text-embedding-3-large* can be reduced with limited loss of semantic information. (<https://openai.com/index/new-embedding-models-and-api-updates/>)

where $s = 0, 1 \dots S_{i,m,t}$ are the sentences in the private earnings call of firm i in quarter t . K_m and K_n present the sets of topics discussed in public and private calls, respectively. $I[\cdot]$ is an indicator function that equals one if a sentence is assigned to a topic that does not appear in the public call, and zero otherwise. In addition, we calculate the number of topics discussed in the public call but not disclosed in the private call as a measure of omitted topics, denoted by $\#OmitTopic$.

2.3.2 Textual Inconsistencies

For the topics that overlapped between public and private calls, we aggregate sentences with the same topic label to construct measures of textual inconsistency. We first compute a numerical inconsistency measure $\Delta Number$, defined as the absolute difference in numerical content (including digital numbers, percentages, dates and monetary figures) between public and private calls for the same topic, normalized by total text length.¹⁰ Next, we derive topic-level disclosure narratives by averaging the embedding vectors of sentences within each topic. We use these narratives vectors in public call $\mathbf{V}_{i,k,m,t}$ and private call $\mathbf{V}_{i,k,n,t}$ to calculate narrative inconsistency $\Delta Narrative$:

$$\Delta Narrative_{i,t} = \frac{1}{K} \sum_{k=1}^K (1 - \cos(\mathbf{V}_{i,k,m,t}, \mathbf{V}_{i,k,n,t})) \quad (2)$$

where $\cos(\mathbf{V}_{i,k,m,t}, \mathbf{V}_{i,k,n,t})$ denotes the cosine similarity between the narrative vectors in public call $\mathbf{V}_{i,k,m,t}$ and private call $\mathbf{V}_{i,k,n,t}$ of the topic k . For ease of interpretation, we define $1 - \cos(\mathbf{V}_{i,k,m,t}, \mathbf{V}_{i,k,n,t})$ as the measure of narrative inconsistency, where the public-private call pairs have the most (least) similar narratives take the lowest (highest) value.

We further use sentiment scores assessed by TopicGPT to compute the text-based sentiment inconsistency measures for each public-private call pair. Specifically, for each topic, we average sentiment scores across sentences to obtain topic-level sentiment measures. We then compute sentiment inconsistency as:

$$\Delta Sentiment_{i,t} = \frac{1}{K} \sum_{k=1}^K |\text{Sentiment}_{i,k,m,t} - \text{Sentiment}_{i,k,n,t}| \quad (3)$$

where $\Delta Sentiment_{i,t}$ denotes the sentiment inconsistency for firm i in quarter t . $\text{Sentiment}_{i,k,m,t}$ and $\text{Sentiment}_{i,k,n,t}$ are the topic-level sentiment scores of topic k in public

¹⁰ A higher value of the number used may indicate the provision of more concrete and verifiable information during a call.

and private earnings calls, respectively. In Section 4.2, we further analyze the directional component of sentiment inconsistency, i.e., whether the sentiment differences are primarily driven by managers being more positive in private (private-driven) or in public (public-driven) meetings.

2.3.3 Vocal Inconsistencies

In addition to text-based measures, we analyze inconsistencies in managers' vocal delivery across earnings calls. We obtain audio recordings of matched public and private earnings call pairs from the Comein Finance database. Using *Whisper*, a speech-to-text algorithm developed by OpenAI (Radford et al., 2023), we transcribe the audio recordings into timestamped text. We employ the *ccalign* algorithm developed by Ewertz et al. (2026) to map textual transcripts with conference call audio at the sentence level, such that we can extract topic-level vocal features and identify the inconsistencies in managers' vocal delivery across meetings.

For each call pair, we aggregate sentence-level audio recordings and the associated timestamped transcripts at the topic-level to construct measures of vocal inconsistency. We first calculate the differences in audio length ($\Delta Length$) between public and private calls for the same topic. We then follow Call et al. (2023) and construct a measure of managers' vocal listenability during earnings calls, based on the average word-level confidence scores produced by *Whisper* when transcribing speech. Specifically, we obtain *Whisper*'s confidence scores for each spoken word in an earnings call, which represent the model's posterior confidence in the correctness of each recognized token. Higher transcription confidence indicates greater listenability of managerial vocal disclosures. Vocal listenability is defined as follows:

$$Listenability_{i,t} = \frac{1}{K} \sum_{K=1}^K \sum_{w \in S} \frac{1}{N_s} Conf(w) \quad (4)$$

where $Listenability_{i,t}$ is the listenability of managerial presentation for firm i in quarter t , with $Conf(w)$ representing the confidence score of word w generated by *Whisper*, and N_s is the number of words in sentence S . A higher value of $Listenability$ reflects clearer and more intelligible vocal delivery, whereas a lower value indicates ambiguous or less intelligible speech. To quantify differences in vocal listenability between public and private settings, we calculate $\Delta Listenability$ as the absolute difference in listenability for the same topic across meetings.

We further construct vocal cognitive measures by feeding the earnings call audio into the pre-trained Wav2Vec2 model, a self-supervised learning framework of speech representations (Baik et

al., 2024). We focus on vocal emotion along two dimensions—valence and arousal (Cao et al., 2023). Valence measures the direction of the vocal feature (positive versus negative affect), while arousal captures the emotional intensity of vocal attitude (exciting versus calm).¹¹ As a result, we derive valence inconsistency $\Delta Valence$ and arousal inconsistency $\Delta Arousal$ between public and private calls using the similar formula in Equation (3).

2.4 Robustness Checks

In this section, we examine the robustness of our identification framework to variations in prompt design and sampling strategies. Specifically, we make three modifications to the benchmark settings. First, we change and expand the example topics to 5 pairs in the prompts and evaluate how alternative prompt strategies affect TopicGPT’s performance (*Prompt adjustment*). Second, it is important to consider that relying on pre-trained language models can introduce look-ahead bias (Sarkar and Vafa, 2024), as the models embed knowledge up to their training cutoff and may inadvertently incorporate future information.¹² As a precautionary check, we exploit the fact that OpenAI’s embedding model has a knowledge cutoff in September 2021, which allows us to conduct an out-of-sample evaluation. We apply the TopicGPT framework to earnings calls falling outside of the embedding model’s training sample (i.e., a subsample from the time period of 2022 to 2024). We examine whether TopicGPT generates the similar outcomes on this subsample (*Subsample*). Third, we randomize the order of documents in the topic generation and assignment stages to test whether the processing sequence affects the resulting topic classifications (*Sample shuffling*).

The evaluation is based on the premise that no algorithm can interpret the context of human conversation better than humans. We engage five Chinese finance-major doctoral students and randomly choose 200 pairs of public and private earnings calls for human assessment. Each individual independently generates topics, assigns topic labels to sentences in the earnings call transcripts, and evaluates the degree of disclosure inconsistency between public-private call pairs instructed by the similar guideline in the pipeline of our identification framework.

We first assess the robustness of our identification framework in terms of topic generation and assignment. For topic generation, we examine whether the latent topics identified by the TopicGPT

¹¹ We exclude one dimension of vocal feature named “Dominance” that reflects the speaker’s perceived sense of control and authority. The presentation session of earning calls is inherently one-directional, with executives serving as the primary information providers. As a result, speakers naturally occupy a dominant role in the interaction. The Dominance dimension therefore exhibits limited variation across observations and provides little incremental explanatory power beyond Valence and Arousal.

¹² Given that we employ generative AI for topic modelling rather than prediction task, our results are less likely to be influenced by look-ahead bias.

are consistent with those recognized by human annotators. We measure topic-level agreement using Purity, Adjusted Rand Index (ARI), and Normalized Mutual Information (NMI). These metrics capture topic consistency from the perspectives of cluster purity, pairwise agreement, and shared information, respectively (Pham et al., 2023). For topic assignment, we examine whether TopicGPT assigns individual sentences to the same topics as human annotators. We compare TopicGPT-generated topic labels with human labels and report Accuracy, Precision, Recall, and F1 scores (Li et al., 2025). The evaluation results are presented in Table A3 of Appendix A1. We find that the topics generated and assigned by TopicGPT are highly consistent with human annotations across different prompting and sampling strategies. In addition, the benchmark pipeline employed in our main analysis exhibits the strongest alignment with human judgments. For example, the benchmark framework achieves a Purity score of 0.805 in topic generation and an Accuracy score of 0.910 in topic assignment, outperforming other prompting and sampling specifications.

Next, we implement a Turing Test to determine whether LLM-generated disclosure inconsistency scores can be distinguished from human judgments (Lagakos et al., 2025). We use the human consensus score as the ground-truth assessment, which is the average disclosure inconsistency measure of the five human annotators. We compute the accuracy of each human and TopicGPT-based method relative to the human consensus score. If the TopicGPT scores fall within the range of human accuracies, we conclude that our identification framework passes the Turing test. Figure A2 reports the Turing test results for all disclosure inconsistency measures, with horizontal lines indicating the range of human reader accuracies.

The statistics in Figure A2 of Appendix A1 indicate our benchmark modelling yields the best performance with human judgments across all the measures of disclosure inconsistency. For example, when assessing the new topic in private meetings (*#NewTopic*), human annotators obtain accuracies ranging from 82.23% to 94.53%, with an average of 88.05%. The prompt adjustment strategy reaches 85.84% accuracy and passes the Turing test, but achieves the lowest accuracy among all the strategies, suggesting that simply adding more example topics is not always beneficial. Accuracy rises to 91.58% under the benchmark modelling. We find similar patterns when substituting the proxies into text and vocal inconsistency measures. Figure A2 also shows that our identification framework passes Turing tests on the subsample covering 2022-2024, beyond the LLM’s training period. In sum, these results indicate that our identification framework generates and assigns topics and produces disclosure

inconsistency measures that closely match human judgments. These features are robust to alternative prompt designs and out-of-sample evaluations.

3. Dissecting Disclosure Inconsistencies

3.1 Topic Taxonomy and Summary Statistics

In Table A4 of Appendix A1, we present the topic taxonomy generated by the TopicGPT framework, which consists of 31 top topics and 176 associated subtopics, broadly covering the range of themes disclosed by firm managers. The taxonomy includes value-relevant topics such as *Accounting Policies and Standards*, *Business Strategy*, *Mergers and Acquisitions*, and *Profit and Loss*, as well as value-irrelevant topics including *Branding and Reputation*, *Employee Welfare*, and *Corporate Culture*, etc.

In Table 1, we provide summary statistics for the disclosure inconsistency measures and other variables of interest. The measures of *#NewTopic* and *#OmitTopic* have mean values of 1.238 and 1.813, accounting for 15.03% and 18.76% of all topics discussed in private and public calls, respectively. These statistics suggest that managers in our sample firms disclose a non-trivial proportion of new topics in private calls while omitting a substantial proportion of topics discussed in public calls. The statistics of textual and vocal inconsistency measures further indicate that managers demonstrate distinct narrative styles and vocal delivery strategies when discussing the same topic across different meetings.

[Insert Table 1 Here]

In Table A5 of Appendix A1, we report the industry-level variations in disclosure inconsistency measures. Industry classification follows the sectoral standards of the China Securities Regulatory Commission (CSRC). We find that firm managers of listed firms in Scientific Research and Technical Services, Environment and Public Facility Management, and the Manufacturing sector tend to disclose higher levels of differentiated information across meetings. These industries are typically characterized by stricter information control, higher disclosure sensitivity, and more extensive informal communication channels. In contrast, firms operating in Education and Culture, Sports, and Entertainment exhibit substantially lower levels of disclosure inconsistencies across earnings calls.

3.2 Decomposition of Disclosure Inconsistencies

3.2.1 Decomposition of Topic Inconsistencies

In this section, we decompose the three measures of topic inconsistency across the topics in our taxonomy, including *#NewTopic*, *%NewTopic* and *#OmitTopic*. Figure 1 provides the distributions of these three measures across the 31 top-level topics, with each topic expressed as a share of the corresponding measure across all topics.

Panel (a) in Figure 1 shows that the emergence of new topics in private calls is concentrated in business, strategic, and operational domains. The largest shares of both *#NewTopic* and *%NewTopic* are accounted for by topics such as *Corporate Governance*, *Operation Management*, *Research and Development*. Their relatively large shares among new topics discussed in private calls suggests that managers may potentially reserve material information for private audiences, which may grant private call attendees informational advantages in valuing firms and executing trades.¹³

[Insert Figure 1 Here]

Panel (b) in Figure 1 reveals a different structure for omitted topics. The largest shares of *#OmitTopic* are accounted for by *Political Alignment*, *Social Impact*, and *Media Engagement*, etc. Their omission from private calls suggests that managers are more inclined to exclude value-irrelevant discussions from public communication. We also provide the top-and bottom-ranked subtopics in terms of topic inconsistencies in Table A6 of Appendix A1. Taken together, Figure 1 and Table A6 document a clear reallocation of discussion topics across public and private calls. It is difficult to determine whether the new information disclosed in private meetings is material or not, as fair disclosure regulations do not provide bright-line definitions of material information (Allee et al., 2022). Nevertheless, our statistics show that new topics are more closely related to firms' operations, strategies, and other value-relevant activities, than topics omitted from private meetings, making them more likely to update investors' beliefs about firm fundamentals.

3.2.2 Decomposition of Textual Inconsistencies

Next, we analyze the characteristics of textual inconsistencies developed by our identification framework. Figure 2 reports the magnitudes of numerical, narrative, and sentiment inconsistencies across 31 top-level topics. Figure 3 further provides the directional components of numerical and

¹³ In contrast to many developed capital markets where selective disclosure is subject to strict regulatory constraints (Becht et al., 2026), the Chinese disclosure environment imposes comparatively weaker restrictions on private communications between firms and investors, giving managers greater latitude to discuss potentially material information in private interactions.

sentiment inconsistencies between public and private calls, where a topic is classified as public-call driven if the corresponding measure in the public call exceeds its private-call counterpart, and as private-call driven otherwise. In addition, we provide example transcripts of textual inconsistencies between public and private earnings calls in Table A9 of Appendix A1.

[Insert Figure 2 Here]

Panel (a) of Figure 2 shows that both numerical ($\Delta Number$) and narrative ($\Delta Narrative$) inconsistencies are more pronounced in topics that may convey large portions of soft information, such as *Business Strategy*, *Mergers and Acquisitions*, *Operation Management*. In these topics, managers retain considerable flexibility in how numbers are highlighted, and how narratives are constructed between public and private calls. In panel (b), the sentiment inconsistency displays a distinct structure. The largest sentiment gaps are identified in topics that are inherently performance- and risk-related, such as *Profit and Loss*, *Risk Exposure*, *Compliance and Enforcement*, etc.

[Insert Figure 3 Here]

Figure 3 explores how these textual inconsistencies arise. Numerical content in topics such as *Sales and Revenue*, *ESG* and *Auditing and Internal Controls* is predominantly public-driven, indicating that managers provide more extensive quantitative detail in formal public disclosures to response to regulatory scrutiny. In contrast, numerical discussion of *Business Strategy*, *Operation Management*, and *Debt and Capital Structure* is largely private-driven, suggesting a more concrete and detailed discussion during private meetings. In terms of the textual sentiment, we find that most performance- and risk-related topics exhibit public-driven sentiment, whereas only a limited set of topics display private-driven sentiment. This pattern indicates that managers tend to adopt a more optimistic tone in public calls while conveying more measured tones in private meetings. This finding is consistent with the “tone management” documented in the literature (Huang et al., 2014), suggesting that managers may employ an abnormally optimistic tone in public disclosures relative to the firm’s underlying fundamentals.

Overall, Figures 2 and 3 demonstrate numerical and narrative inconsistencies are concentrated in topics with soft information that allow managerial discretion, whereas sentiment differences primarily arise in risk-sensitive and performance-related topics. We provide the top- and bottom-ranked subtopics in terms of textual inconsistencies in Table A7 of Appendix A1. These patterns reinforce the interpretation that managers strategically reallocate not only what they disclose but also how they present and emotionally frame information across public and private meetings.

3.2.3 Decomposition of Vocal Inconsistencies

In this section, we decompose the vocal dimensions of disclosure inconsistency. Figure 4 reports the magnitudes of inconsistencies in audio length, listenability, vocal valence, and arousal across topics, while Figure 5 illustrates the directions of these inconsistencies between public and private calls. We also provide the top-and bottom-ranked subtopics in terms of vocal inconsistencies in Table A8 of Appendix A1

[Insert Figure 4 Here]

The panels (a) and (b) of Figure 4 and Figure 5 reveal an informative pattern in the relationship between speech quantity and quality. In several topics, such as *Compliance and Enforcement*, *Auditing and Internal Controls*, and *Property Management*, managers speak substantially longer in public calls while the corresponding listenability is relatively lower than in private calls. In other words, although public communication in these topics contains longer speaking time, it is simultaneously less coherent, less fluent, and harder for the audience to process. By increasing vocal output while reducing informational clarity when discussing sensitive operational and risk-related topics in public meetings, managers may seek to satisfy extensive disclosure expectations under regulatory and market scrutiny while preserving some ambiguity to public audience.

The panels (c) and (d) of Figure 4 show that inconsistencies in vocal valence and arousal are significantly larger than their text-based sentiment counterparts reported in Figure 2. The largest gaps occur in topics such as *Profit and Loss*, *Risk Exposure*, *Cash Flow and Liquidity*, and *Debt and Capital Structure*. This result indicates that vocal signals capture layers of managerial emotion and tension not fully reflected in prepared textual content. One possible explanation is that, while textual sentiment can be carefully scripted and reviewed ex ante, vocal expressions are harder to control,

making voice cues a more sensitive channel for detecting latent disclosure frictions and informative signals.

[Insert Figure 5 Here]

Figure 5 provides the directional structure of these vocal inconsistencies. In the length and valence dimensions, most topics are public-call driven, indicating that managers speak longer and display more positive emotional tone in public calls. However, this expansion in speech volume is frequently accompanied by a deterioration in listenability, which for many topics is private-call driven. The statistics in Figures 4 and 5 further our understanding of managerial disclosure strategies. Compared with text-based measures, vocal features exhibit stronger inconsistencies and may capture aspects of managerial intentions, emotional regulation, and strategic obfuscation that remain unexplored in textual records.

3.3 The Determinants of Disclosure Inconsistencies

3.3.1 Research Design

In this section, we explore the determinants of disclosure inconsistencies identified between public and private calls. We expect that the disclosure inconsistencies will have different level determinants, including earnings call-, manager- and firm-level characteristics. We commence with a dataset encompassing all publicly listed companies on Chinese stock exchanges from 2020 to 2024 in the China Stock Market and Accounting Research Database (CSMAR).¹⁴ The variables of interest are defined as follows.

Call Characteristics. The first set of variables includes the earnings call characteristics. We construct dummy variables *Online* that take one if the public-private call pairs are both held online (i.e., no physical meeting location), zero otherwise, and *Team* for the call-pairs both held by firm executive teams rather than individuals. For private call co-hosted by financial institutions, we further include a dummy variable (*Affiliation*) that takes one if the co-host financial institution was the lead or co-underwriter of an initial public offering of the firm during the past five years, or for a seasoned equity offering during the past two (Malmendier and Shanthikumar, 2014).

¹⁴ CSMAR provides an extensive dataset that includes stock prices, financial statements, and details on corporate governance and ownership structure. (<https://data.csmar.com/>)

Manager Characteristics. Given that the presentation sessions of both public and private calls are delivered by the same managers, we include a range of manager-level characteristics for analysis. We identify the gender (*Gender*) and age (*Age*) of the manager. We also include the number of years the manager has held the position at the firm (*Tenure*). We further include a dummy variable *Degree*, which takes the value of one if the manager has a degree in finance or law related majors, and zero otherwise.

Firm Characteristics. We introduce a set of firms' financial and operational proxies to examine their impacts on the measures of disclosure inconsistency, including the financial leverage ratio (*Leverage*), dividend payout (*Dividend*), return on assets (*ROA*), firm age (*Firm age*) and total assets (*Asset*) for our sample firms. We introduce the percentage of the firm shares owned by institutional investors (*IO*) as the proxy of investor pressure. We also include a dummy variable that denotes whether a firm is a state-owned enterprise (*SOE*).

Our model to examine the determinants of the disclosure inconsistencies is then specified as follows:

$$\begin{aligned}
 DisclosureInc_{i,j,t} &= \beta_1 + \beta_2 CallCharacter_{i,j,t} + \beta_3 ManagerCharacter_{i,j,t} + \beta_4 FirmCharacter_{i,j,t} \\
 &+ \lambda_j + \mu_t + \varepsilon_{i,j,t}
 \end{aligned} \tag{5}$$

where $DisclosureInc_{i,j,t}$ are the measures of disclosure inconsistencies between public-private call pairs for firm i in industry sector j at quarter t , including the measures of topic inconsistencies and textual inconsistencies, and vocal inconsistencies. *CallCharacter* includes dummy variables *Online*, *Team*, and *Affiliation*. *ManagerCharacter* presents the gender, age and position tenure of the manager. We also include *Degree* to denote the managers' finance or law-related backgrounds. *FirmCharacter* consists of financial leverage ratio (*Leverage*), dividend payout (*Dividend*), return on assets (*ROA*), institutional ownership (*IO*), firm age (*Firm age*), total assets (*Asset*), and state-owned enterprise dummy (*SOE*) for our sample firms. δ_j , μ_t and $\varepsilon_{i,j,t}$ are the industry, time fixed effects, and error term.

3.3.2 Empirical Results

Table 2 reports the estimation results of Equation (5). The dependent variables are the measures of disclosure inconsistencies developed from our identification framework, including topic, textual

and vocal inconsistencies. The results provide consistent evidence that disclosure inconsistencies are associated with the characteristics of earnings calls, managers and listed firms.

[Insert Table 2 Here]

At the earnings call level, we find that *Online* and *Team* are negatively associated with the measures of disclosure inconsistency. For example, the coefficients of *Online* and *Team* are -0.164 (t -stat: -3.17) and -0.117 (t -stat: -2.51) on *#NewTopic*, -0.051 (t -stat: -2.40) and -0.046 (t -stat: -1.94) on *ΔLength*, suggesting that firm executives are less likely to disclose new topics and adjust vocal deliveries during private calls when the meetings are held online and by executive teams. In contrast, the coefficients of *Affiliation* show that the disclosure inconsistencies between public and private calls are larger when private meetings are co-hosted by institutions affiliated with the listed firms.

In terms of manager characteristics, the coefficients of *Gender* and *Age* are statistically significant across measures of topic and textual inconsistency, suggesting that managers who are male and in their elder age are more likely to provide new information and adjust wording during private earnings calls. The coefficients of *Tenure* (-0.012, t -stat: -3.14) and *Age* (-0.013, t -stat: -2.47) on *ΔArousal* are significantly negative, indicating that more experienced managers exhibit smaller variations in emotional intensity and tend to adopt a more even-tempered vocal delivery across meetings. In addition, we find that the *Degree* are significantly negatively associated with all disclosure inconsistency measures. It is possible that managers with finance or law related backgrounds are more sensitive to the regulations of information disclosure and insider trading, and thus less likely to provide incremental information during private meetings.

For firm-level characteristics, the estimation results show that firms' better financial positions are associated with lower levels of disclosure inconsistencies, as evidenced by the coefficients of *Leverage*, *Dividend* and *ROA* in Table 2. Furthermore, we find that firm executives show lower disclosure divergence across earnings calls if the listed firms have longer years of business or are state-owned enterprises. Overall, these results highlight that disclosure inconsistencies are shaped by a combination of communication environments, managerial experience, and financial incentives.

4. The Informativeness of Disclosure Inconsistencies

4.1 Abnormal Stock Returns

We first examine the impacts of disclosure inconsistencies on abnormal stock returns around private meetings. We use the absolute value of abnormal returns over the three-day windows (*ARET*) following private earnings calls as the dependent variable. *ARET* is the difference between three-day absolute market-adjusted returns and the mean three-day absolute market-adjusted returns in a benchmark period, divided by the standard deviation of mean absolute market-adjusted returns in the period (Bushee et al., 2018; Brown et al., 2018).¹⁵ The regression specification is as follows:

$$ARET_{i,j,[t,t+2]} = \alpha + \beta DisclosureInc_{i,j,t} + Controls_{i,j,t} + \lambda_j + \mu_t + \varepsilon_{i,j,t} \quad (6)$$

Where $ARET_{i,j,[t,t+2]}$ is the absolute value of abnormal returns for firm i in industry j over the post-meeting window $[0,2]$. $DisclosureInc_{i,j,t}$ are the disclosure inconsistency measures between public-private earnings calls, which consist of topic, textual, and vocal inconsistencies. We standardize all the measures of disclosure inconsistency to have zero mean and unit variance throughout the regressions such that the coefficients can be interpreted and compared across specifications. $Controls_{i,j,t}$ denote a set of firm-level control variables, including financial leverage ratio (*Leverage*), dividend payout (*Dividend*), return on assets (*ROA*), institutional ownership (*IO*), firm age (*Age*), total assets (*Asset*), and state-owned enterprise dummy (*SOE*) for our sample firms. δ_j , μ_t and $\varepsilon_{i,j,t}$ are the industry, time fixed effects, and error term.

[Insert Table 3 Here]

In Table 3, we present the effects of topic inconsistencies on three-day absolute market-adjusted returns. In Panel A of Table 3, we find that the proxies of new topics *#NewTopic* (0.153, t -stat: 4.55) and *%NewTopic* (0.147, t -stat: 4.47) are both positively associated with absolute abnormal returns following private earnings calls. However, we do not find such significant relationship between *#OmitTopic* and *ARET*. We next examine whether the informativeness of new topics varies with the manner in which they are communicated. In Panel B and C, we introduce the textual and vocal characteristics of managers' discussions of new topics in private meetings, including number usage

¹⁵ In this paper, the benchmark for both stock returns and trading volume is defined as the non-private meeting period spanning from day -45 to day -6 and from day +6 to day +45 relative to the private meeting date.

(*NumberPriv*), textual sentiment (*SentimentPriv*), listenability (*ListenabilityPriv*), valence (*ValencePriv*) and arousal (*ArousalPriv*).¹⁶ The coefficients in Panel B and C of Table 5 indicate that the impacts of new topics in private calls are more pronounced when managers provide more quantitative information and convey the information with stronger emotional intensity. Overall, the results of Table 3 suggest the new information contained in private meetings generates significant price reactions during the event window.

[Insert Table 4 Here]

Table 4 examines the impacts of textual and vocal inconsistencies on three-day absolute market-adjusted returns. Panels A and B report the standalone effects of textual and vocal inconsistency measures, respectively, while Panel C reports pairwise interactions between textual and vocal inconsistency proxies on *ARET*. In Panels A and B, we find that only $\Delta Narrative$ and $\Delta Valence$ exert significant impacts on *ARET* when examined individually. In contrast, a larger number of interaction terms are significantly associated with three-day absolute market-adjusted returns in Panel C, such as $\Delta Number \times \Delta Valence$ (0.030, *t*-stat: 2.65), $\Delta Narrative \times \Delta Listenability$ (0.062, *t*-stat: 3.19), and $\Delta Sentiment \times \Delta Arousal$ (0.015, *t*-stat: 2.50), etc. These findings suggest that the market response to public-private disclosure differences is driven less by single dimensions, and more by the combined interpretation of multiple communication cues. Investors appear to process textual and vocal signals jointly, as disclosure inconsistencies become more informative when they reinforce one another across different dimensions of communication.

4.2 Abnormal Share Turnover

Next, we examine the impacts of disclosure inconsistencies on share turnover following private meetings. We measure abnormal share turnover (*ATURN*) to capture deviations in turnover around a private meeting relative to normal periods. Specifically, we calculate *ATURN* as the difference between three-day turnover ratio and the mean three-day turnover ratio in a benchmark period, divided by the standard deviation of the three-day turnover ratios in the period. Higher values of *ATURN* indicate elevated trading activities relative to normal levels.

¹⁶ The definitions of all the textual and vocal features in private calls are identical to those in Section 2.

[Insert Table 5 Here]

In Table 5, we regress the abnormal share turnover on the measures of topic inconsistency between public and private calls. The coefficients in Panel A show that one-standard-deviation increases in *#NewTopic* and *%NewTopic* are associated with 0.383 (*t*-stat: 5.26) and 0.355 (*t*-stat: 5.12) increases in the abnormal share turnover over a three-day horizon relative to the private meetings. In Panel B and C, we incorporate the textual and vocal characteristics of managers' discussions of new topics in private meetings. We find that both textual and vocal communication characteristics moderate the relation between topic inconsistencies and abnormal share turnover. For example, the impacts of *#NewTopic* on *ATURN* increase by 27.94% ($0.107/0.383$) and 20.62% ($0.079/0.383$) for one-standard-deviation increases the number usage and vocal valence in private meetings, respectively. The results in Table 5 indicate that new topics discussed in private meetings are associated with significant market reactions in terms of trading volume.

[Insert Table 6 Here]

In Table 6, we examine the impacts of textual and vocal inconsistencies on three-day abnormal share turnover. We report the standalone effects of textual and vocal inconsistency measures in Panel A and B, and the pairwise interaction effects among inconsistency proxies in Panel C. We find that several inconsistency measures are individually associated with abnormal share turnover around private meetings, including public-private differences in numerical usage, narrative content, vocal valence and arousal. Nevertheless, their effects become substantially stronger when combined with other dimensions of disclosure inconsistency. For example, the impact of $\Delta Number$ on *ATURN* increases by 82.52% ($0.085/0.103$) when accompanied by greater valence inconsistency, as captured by the interaction term $\Delta Number \times \Delta Valence$. Consistent with the return-based evidence in Section 4.1, the interaction effects generally exhibit larger economic magnitudes than the standalone effects, indicating that the combined interpretation of multiple communication cues plays a more important role in explaining abnormal trading volume around private meetings.

In addition, taking the results in Sections 4.1 and 4.2 together, we find that the effects of disclosure inconsistencies exhibit stronger significance for abnormal turnover than for abnormal

returns. One possible explanation is that the incremental information conveyed in private meetings updates investors' beliefs about firm fundamentals, but does not necessarily lead to a common interpretation of that information. That said, such information may generate heterogeneous beliefs and disagreement among investors. In line with the literature (Banerjee & Kremer, 2010; Cookson et al., 2026), information that induces divergent interpretations can generate substantial trading responses while contributing relatively less to price discovery. As a result, public-private disclosure inconsistencies appear to have a more pronounced impact on trading volume than on stock price reactions around private meetings.

5. Do Private Meeting Attendees Trade on Disclosure Inconsistencies?

Thus far, our results show that the disclosure inconsistencies between public and private calls convey valuable information to market participants. In this section, we further examine whether private meeting attendees trade in response to disclosure inconsistencies, and whether such trading is informed.

5.1 Attendee Trading

Shenzhen Stock Exchange (SZSE), one of the major stock exchanges in China, requires listed firms to publish summary reports on private meetings with investors through its online investor relations platform, HudongYi. These disclosures contain information on meeting dates, locations, and the identities of participating investors and institutions (Bowen et al., 2018; Ru et al., 2025). We collect data on private meeting activities from the China Listed Firms' Investor Relations Database in CSMAR.¹⁷ We map these reports with our earnings call dataset to identify institutional investors attending private meetings. We further obtain quarterly institutional holdings data from CSMAR and merge them with the identified attendee records. This matching procedure yields 6,371 public-private earnings call pairs with holdings data for private meeting attendees.

We first examine the impact of disclosure inconsistencies on the trading behaviors of private meeting attendees. For each fund attending firm i 's private meeting at quarter t , we calculate the absolute value of the difference in the fund's holdings of firm i from quarter t to quarter $t + 1$, scaled by the firm's total shares outstanding. We then construct the firm-level measure of attendee

¹⁷ The official website of HudongYi is <http://irm.cninfo.com.cn>. We manually complement the data from HudongYi for private meeting summary reports missing information on meeting participants.

trading (*Attendee Trading*) by aggregating the unsigned holding changes of all investors participating in firm i 's private meeting (Bradley et al., 2022). We run the following regression specification:

$$AttendeeTrading_{i,j,t} = \alpha + \beta DisclosureInc_{i,j,t} + Controls_{i,j,t} + \lambda_j + \mu_t + \varepsilon_{i,j,t} \quad (7)$$

where $AttendeeTrading_{i,j,t}$ is the unsigned change in holdings of firm i by private meeting attendees from quarter t to quarter $t + 1$. $DisclosureInc_{i,j,t}$ are the disclosure inconsistency measures between public-private earnings calls, which consist of topic, textual, and vocal inconsistencies. $Controls_{i,j,t}$ denote a set of firm-level control variables, including financial leverage ratio (*Leverage*), dividend payout (*Dividend*), return on assets (*ROA*), institutional ownership (*IO*), firm age (*Age*), total assets (*Asset*), and state-owned enterprise dummy (*SOE*) for our sample firms. δ_j , μ_t and $\varepsilon_{i,j,t}$ are the industry, time fixed effects, and error term.

[Insert Table 7 Here]

Table 7 examines whether topic inconsistencies in private meetings are associated with subsequent trading by private meeting attendees. The coefficients in Panel A of Table 7 indicate that one-standard-deviation increases in *#NewTopic* and *%NewTopic* are associated with average 2.70% (t -stat: 3.69) and 2.53% (t -stat: 4.28) increases in the unsigned holding changes of investors participating in private meetings, respectively. Panels B and C further examine whether the way managers communicate new topics in private meetings affects attendees' trading responses. The results show that the effect of new topics is stronger when managers discuss these topics with more numerical content and stronger vocal cues. For example, the impacts of *#NewTopic* and *%NewTopic* on *Attendee Trading* increase by 7.03% (0.189/2.689) and 8.85% (0.224/2.530) for one-standard-deviation increases the vocal valence ($\Delta Valence$) in private meetings, respectively.

[Insert Table 8 Here]

In Table 8, we examine whether textual and vocal disclosure inconsistencies are associated with the subsequent trading volume of private meeting attendees. Panels A and B show that several

standalone inconsistency measures are significantly associated with attendees' unsigned holding changes. For example, public-private differences in narrative content ($\Delta Narrative$) and vocal valence ($\Delta Valence$) are associated with 0.91% (t -stat: 3.43) and 0.16% (t -stat: 2.46) increases in the total trading of private meeting attendees, respectively. In addition, these effects become stronger when accompanied by other dimensions of textual and vocal inconsistency, as evidenced by the significant interaction terms reported in Panel C of Table 8. Overall, these findings suggest that the informational content embedded in disclosure inconsistencies can be incorporated into the trading decisions of private meeting attendees.

To examine whether our results are driven by general institutional trading around the same firms rather than information acquired by private meeting attendees, we conduct a placebo test using non-attendee trading data. Specifically, for each firm-quarter, we identify institutional investors that hold shares of firm i in quarter t but do not participate in firm i 's private earning calls. We then aggregate their unsigned holding changes from quarter t to quarter $t + 1$ to construct a firm-level measure of non-attendee trading. Table A10 of Appendix A1 reports the placebo test results. The estimated coefficients are statistically insignificant across topic, textual, vocal inconsistencies, and their pairwise interactions. These results suggest that disclosure inconsistencies do not predict trading by institutional investors who did not attend the private meetings. The trading responses documented in our analysis are thus less likely to reflect general market-wide institutional rebalancing, and are more consistent with private meeting attendees incorporating information from public-private disclosure inconsistencies into their subsequent portfolio decisions.

5.2 Trading Gains

While the results in Section 5.1 suggest that private meeting attendees trade in response to the disclosure inconsistencies between public and private earnings calls, it is unclear whether these trades are informed. In this Section, we investigate whether private meeting attendees earn higher trading gains when trading on disclosure inconsistencies. *Trading Gains* is defined as the quarterly changes in institutional investors' holdings multiplied by the firm's buy-and-hold market-adjusted returns over the subsequent quarter. *Trading Gains* is positive (negative) when institutional investors trade in the same (opposite) direction as future stock returns. We then re-estimate the regression specifications in Equation (7) using *Trading Gains* as the dependent variable.

[Insert Table 9 Here]

We report the impacts of topic inconsistencies on in *Trading Gains* Table 9. The coefficients of *#NewTopic* and *%NewTopic* in Panel A are both statistically positive, indicating that investors are able to earn trading gains based on incremental information collected on private meetings. One-standard-deviation increases in *#NewTopic* and *%NewTopic* are associated with 16.7 bps (*t*-stat: 2.79) and 14.5 bps (*t*-stat: 2.36) in quarterly trading gains for private meeting attendees, respectively. In Panels B and C, we find that these trading gains are stronger when new topics are communicated with richer textual and vocal cues. For example, the interaction terms with *NumberPriv* are positive and significant for both *#NewTopic* and *%NewTopic*, with additional trading gains of 4.5 to 5.0 bps.

Table 10 reports the relation between textual and vocal disclosure inconsistencies and the trading gains of private meeting attendees. In Panel A and B, we find that only measure of narrative inconsistency can predict significant trading gains when examined individually. In contrast, a variety of interaction terms between textual and vocal inconsistencies are significantly associated with the trading gains of private meeting attendees in Panel C. For example, the money made is 2.1 bps (*t*-stat: 2.75) from $\Delta\text{Number} \times \Delta\text{Listenability}$, 6.7 bps (*t*-stat: 2.40) from $\Delta\text{Narrative} \times \Delta\text{Arousal}$, and 3.4 bps (*t*-stat: 2.77) from $\Delta\text{Sentiment} \times \Delta\text{Valence}$. That said, they appear to combine multiple dimensions of public-private disclosure inconsistencies, including differences in narrative content, numerical information, and vocal delivery, to become better informed through private meetings. This pattern is consistent with the mosaic theory of investing (Cheynel & Levine, 2020), whereby investors collect and integrate pieces of information to construct a more complete picture of the firm. By gathering together disclosure inconsistencies to form an information mosaic, investors are able to update their beliefs in a more accurate manner and thus make more informed trading decisions.

We also conduct a placebo test using the trading gains of non-attendees. We construct non-attendee trading gains in the similar format in Section 5.1. Table A11 of the Appendix A1 reports the results. The insignificant coefficients in Table A11 results suggest that disclosure inconsistencies do not predict profitable trading by investors who did not attend the private meetings. This further supports the interpretation that the trading gains are more likely driven by information acquired through private meeting attendance, rather than by general institutional trading around the same firms.

6. Conclusion

In this paper, we investigate how firm managers disclose inconsistently across public and private earnings calls and evaluate the informativeness of such disclosure inconsistencies. By leveraging a novel dataset of 20,480 paired public-private earnings calls for Chinese listed firms between 2020 and 2024, we use the TopicGPT-based identification framework to construct novel measures of disclosure inconsistencies, including topic, textual and vocal inconsistencies. Our analysis reveals that managers disclose a non-trivial proportion of new, and potentially material topics in private meetings and omit a number of topics discussed in public calls. Managers also demonstrate distinct narrative approaches and vocal delivery strategies when discussing the same topic at different meeting occasions. We explore the determinants of these disclosure inconsistencies. We find that inconsistencies are more likely to arise in offline meetings, when disclosures are delivered by a single executive, or when meetings are organized by non-state-affiliated institutions. In addition, older and male managers tend to generate higher levels of inconsistencies, while executives with finance or law-related training are more compliant with disclosure regulations.

We next examine the informativeness of disclosure inconsistencies. The inconsistency measures are significantly associated with absolute abnormal returns following private meetings. Market responses are driven less by individual dimensions than by the joint interpretation of multiple communication cues. We find that disclosure inconsistencies generate more pronounced responses in abnormal share turnover than in stock prices. This difference suggests that disclosure inconsistencies may generate heterogeneous beliefs and disagreement among investors.

We further show that disclosure inconsistencies significantly affect changes in attendees' quarterly holdings in meeting firms. These trades appear to be informed, as disclosure inconsistencies are positively associated with attendees' subsequent trading gains. Private meeting attendees make more informed trades by integrating multiple dimensions of disclosure inconsistencies rather than evaluating each dimension in isolation. That said, by gathering together disclosure inconsistencies to form an information mosaic, investors are able to update their beliefs in a more accurate manner and thus make more informed trading decisions.

Our findings carry important policy and managerial implications. We show that the informational value of private meetings arises from systematic inconsistencies between public and private disclosures in what managers discuss, how they construct their messages, and how they vocally deliver them. In essence, fair disclosure regulation in all economics prohibit selective disclosure of

material information while allowing managers to communicate non-material information privately, even when sophisticated investors may combine that information with other signals to form a material information mosaic. Private meetings can therefore improve understanding of firm fundamentals, but they may also create informational advantages for investors with selective access.

Our evidence does not establish that every disclosure inconsistency constitutes material information or that the observed inconsistencies violate fair disclosure regulations, as policymakers do not provide a bright-line definition of material information in the current stage. Nevertheless, we show that information embedded in these inconsistencies are economically consequential: they generate market reactions, shape the trades of private meeting attendees, and can be translated into trading gains. These findings raise concerns about whether fair disclosure regulations effectively maintain a level playing field, and highlight the need for clearer regulatory guidance on private disclosures and the timely public dissemination of value-relevant information disclosed in private meetings.

References

- Allee, K. D., Bushee, B. J., Kleppe, T. J., & Pierce, A. T. (2022). Did the siebel systems case limit the SEC's ability to enforce regulation fair disclosure?. *Journal of Accounting Research*, 60(4), 1235-1291.
- Baik, B., Kim, A. G., Kim, D. S., & Yoon, S. (2024). Vocal delivery quality in earnings conference calls. *Journal of Accounting and Economics*, 101763.
- Becht, M., Franks, J., & Wagner, H. F. (2026). The benefits of access: Evidence from private meetings with portfolio firms. *The Journal of Finance*.
- Banerjee, S., & Kremer, I. (2010). Disagreement and learning: Dynamic patterns of trade. *The Journal of Finance*, 65(4), 1269-1302.
- Bernard, D., Blankespoor, E., de Kok, T., & Toynbee, S. (2026). Using GPT to measure business complexity. *The Accounting Review*, 1-36.
- Bowen, R. M., Dutta, S., Tang, S., & Zhu, P. (2018). Inside the "black box" of private in-house meetings. *Review of Accounting Studies*, 23(2), 487-527.
- Bradley, D., Jame, R., & Williams, J. (2022). Non-deal roadshows, informed trading, and analyst conflicts of interest. *The Journal of Finance*, 77(1), 265-315.
- Bushee, B. J., Gerakos, J., & Lee, L. F. (2018). Corporate jets and private meetings with investors. *Journal of Accounting and Economics*, 65(2-3), 358-379.
- Call, A. C., Wang, B., Weng, L., & Wu, Q. (2023). The listenability of disclosures and firms' information environment. *Available at SSRN 4634029*.
- 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. *The Review of Financial Studies*, 36(9), 3603-3642.
- Cao, S., Gong, G., Kim, Y., Shi, H., & Wang, A. (2025). Site visits and corporate investment efficiency. *Management science*, 71(1), 635-658.
- Cheyne, E., & Levine, C. B. (2020). Public disclosures and information asymmetry: A theory of the mosaic. *The Accounting Review*, 95(1), 79-99.
- Choy, S., & Hope, O.K. (2025). Private communication between managers and financial analysts: Evidence from taxi-ride patterns in New York city. *Rotman School of Management Working Paper*, (3920680).
- Cicero, D. C., Puckett, A., Wang, A. Y., & Zhang, S. (2024). TAXI! Do mutual funds pursue and exploit information on local companies?. *Journal of Financial and Quantitative Analysis*, 59(7), 3340-3375.
- Cookson, J. A., Fos, V., & Niessner, M. (2026). Does disagreement facilitate informed trading?. *Journal of Financial and Quantitative Analysis*, 61(2), 612-639.
- de Kok, T. (2025). ChatGPT for textual analysis? How to use generative LLMs in accounting research. *Management Science*.
- Eisfeldt, A. L., Schubert, G., & Zhang, M. B. (2023). Generative AI and firm values (No. w31222). *National Bureau of Economic Research*.
- Ewertz, J., Knickrehm, C., Nienhaus, M., & Reichmann, D. (2026). Listen closely: Measuring vocal tone in corporate disclosures. *Journal of Accounting Research*, 64(1), 229-277.
- Fan, J., Liu, Q., Song, Y., & Wang, Z. (2024). Measuring misinformation in financial markets. *Available at SSRN 4922648*.
- Frenkel, S., Guttman, I., & Kremer, I. (2020). The effect of exogenous information on voluntary disclosure and market quality. *Journal of Financial Economics*, 138(1), 176-192.

- Fieberg, C., Hesse, M., Liedtke, G., & Zaremba, A. (2025). Predicting financial stability with TopicGPT: Insights from corporate and central bank communications. *Journal of Banking & Finance*, 107598.
- Gerken, W. C., & Painter, M. O. (2023). The value of differing points of view: Evidence from financial analysts' geographic diversity. *The Review of Financial Studies*, 36(2), 409-449.
- Green, T. C., Jame, R., Markov, S., & Subasi, M. (2014). Access to management and the informativeness of analyst research. *Journal of Financial Economics*, 114(2), 239-255.
- Huang, X., Teoh, S. H., & Zhang, Y. (2014). Tone management. *The Accounting Review*, 89(3), 1083-1113.
- Jha, M., Qian, J., Weber, M., & Yang, B. (2024). ChatGPT and corporate policies (No. w32161). *National Bureau of Economic Research*.
- Jha, M., Liu, H., & Manela, A. (2026). Does finance benefit society? A language embedding approach. *The Review of Financial Studies*, 39(6): 1227–1266.
- Lagakos, D., Michalopoulos, S., & Voth, H. J. (2025). American life histories (No. w33373). *National Bureau of Economic Research*.
- Lee, R.Y. (2026). Face-to-face social interactions and local informational advantage. *Journal of Financial Economics*, 180, 104280.
- Li, K., Mai, F., Shen, R., Yang, C., & Zhang, T. (2026). Dissecting corporate culture using generative AI. *The Review of Financial Studies*, 39(1), 253-296.
- Li, X., Schloetzer, J. D., Tseng, A., & Wang, H. (2023). Private meetings with management and investor uncertainty at subsequent earnings announcements. *HKUST Business School Research Paper*, (2022-071).
- Lyonnet, V., Shams, A., & Zhang, S. (2025). What Do Early-Stage Investors Ask? An LLM Analysis of Expert Calls. *Fisher College of Business Working Paper*, (2025-03), 16.
- Malmendier, U., & Shanthikumar, D. (2014). Do security analysts speak in two tongues?. *The Review of Financial Studies*, 27(5), 1287-1322.
- Park, J., & Soltes, E. F. (2018). What do investors ask managers privately?. *Available at SSRN 3087369*.
- Pham, C. M., Hoyle, A., Sun, S., Resnik, P., & Iyyer, M. (2023). TopicGPT: A prompt-based topic modeling framework. *arXiv preprint arXiv:2311.01449*.
- Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. *International conference on machine learning (pp. 28492-28518)*. PMLR.
- Ru, Y., Zheng, R., & Zou, Y. (2025). Public disclosure of private meetings: Does observing peers' information acquisition affect analysts' attention allocation?. *Journal of Accounting Research*, 63(4), 1629-1677.
- Sarkar, S. K., & Vafa, K. (2024). Lookahead bias in pretrained language models. *Available at SSRN 4754678*.
- Shinn, M., Hu, A., Turner, L., Noble, S., Preller, K. H., Ji, J. L., ... & Murray, J. D. (2023). Functional brain networks reflect spatial and temporal autocorrelation. *Nature Neuroscience*, 26(5), 867-878.
- Solomon, D., & Soltes, E. (2015). What are we meeting for? The consequences of private meetings with investors. *The Journal of Law and Economics*, 58(2), 325-355.
- Tseng, A., Schloetzer, J. D., Li, X., & Wang, H. (2024). The role of private meetings with management in reducing investor uncertainty around subsequent earnings announcements. *Available at SSRN 4965526*.

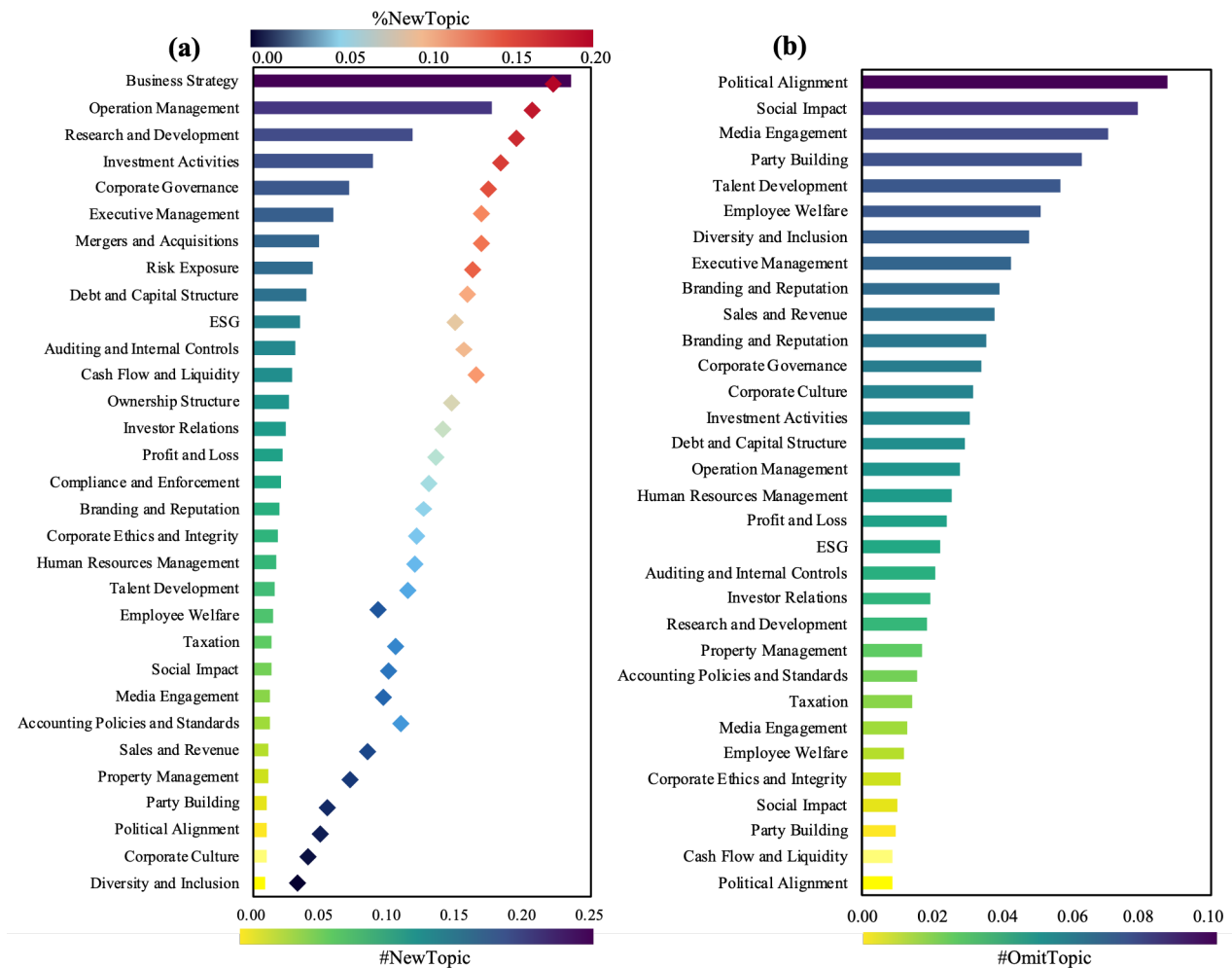


Figure 1 Decomposition of Topic Inconsistencies. This figure presents the distribution of topic inconsistency measures across topics in our taxonomy. For each measure, the value reported for a given topic represents that topic’s share of the corresponding aggregate measure across all topics. The values for each measure sum to one across topics. In Panel (a), the bar chart reports the share of new topic occurrences (*#NewTopic*) accounted for by each topic, while the dot plot reports the corresponding distribution based on *%NewTopic*. In Panel (b), the bar chart reports the share of omitted topic occurrences (*#OmitTopic*) accounted for by each topic.

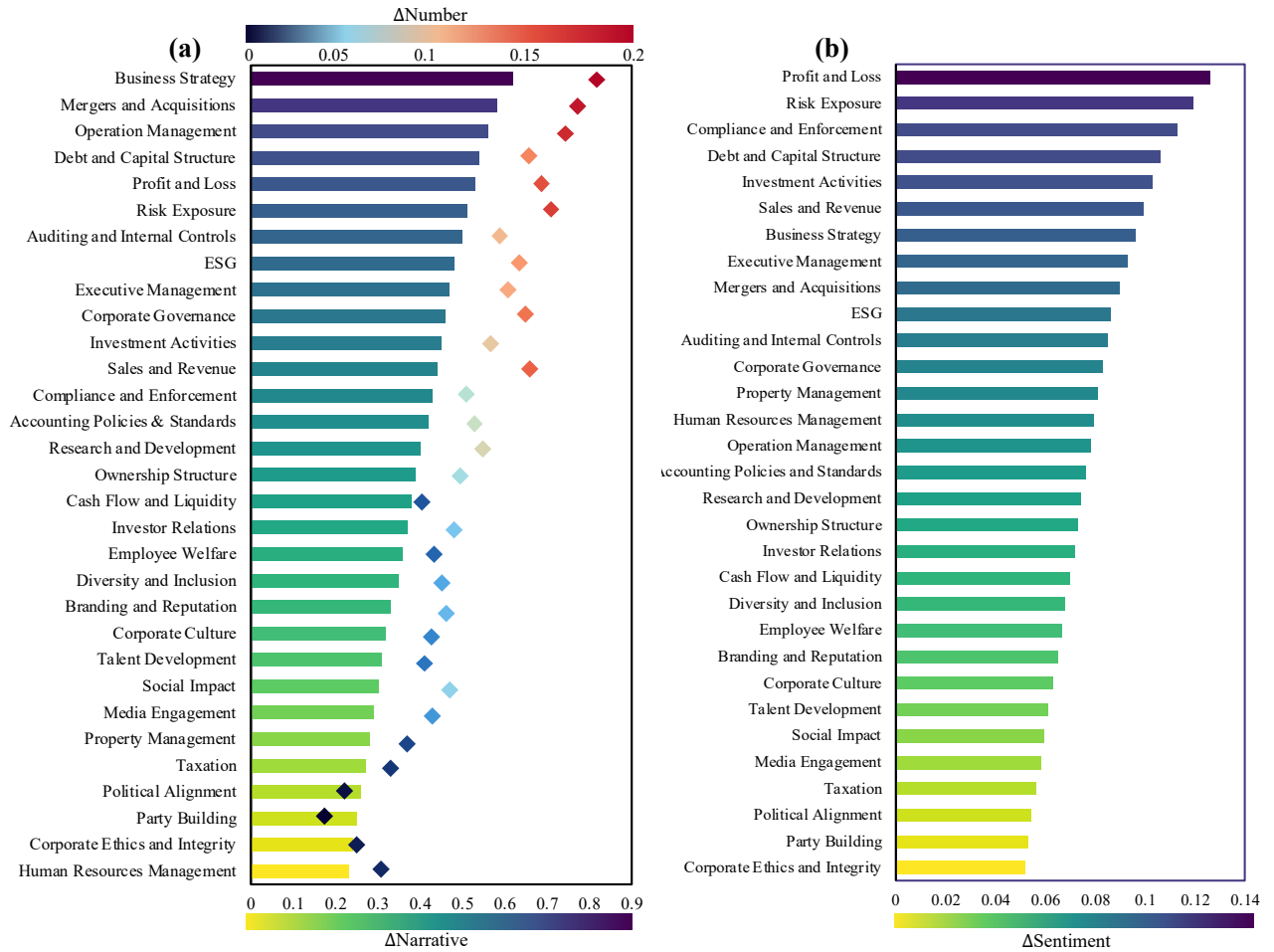


Figure 2 Decomposition of Textual Inconsistencies. This Figure provides the distributions of textual inconsistency measures across topics in our topic taxonomy. In Panel (a), the bar chart shows the narrative differences for each topic disclosed between public and private calls ($\Delta Narrative$), the dot plot reports the differences in the numerical usage for the respective topics ($\Delta Number$), while. In Panel (b), the bar chart reports the text-based sentiment inconsistency between public and private calls for each topic ($\Delta Sentiment$).

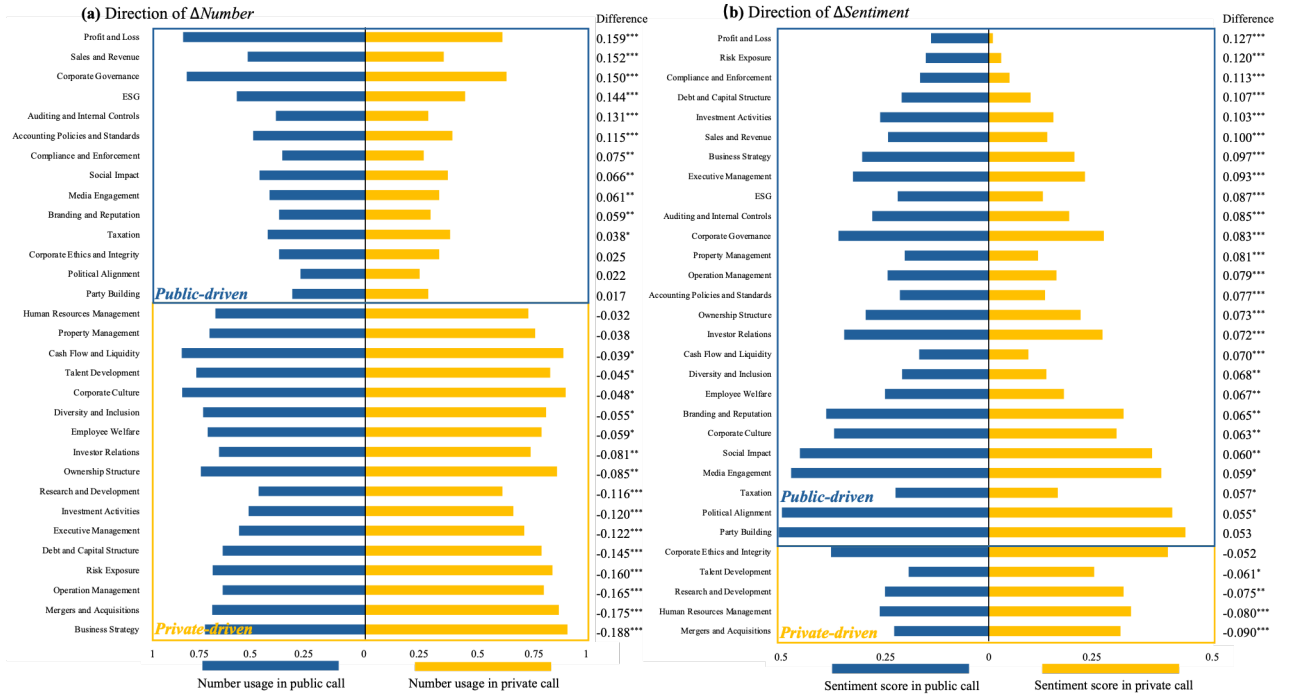


Figure 3 Direction of Textual Inconsistencies. Panels (a) and (b) illustrate the directional differences in numerical usage ($\Delta Number$) and sentiment ($\Delta Sentiment$) between public and private earnings calls, respectively. In both panels, the left (right) bar reports the disclosure statistics in the public (private) call for each topic. For a given topic, the textual inconsistency is classified as public-driven if the public disclosure statistic exceeds the private counterpart, and as private-driven otherwise. We provide the public-private differences (calculated as public minus private) and their statistical significance alongside the figures. *, **, *** are the statistical significance at the 10%, 5%, and 1% levels, respectively.



Figure 4 Decomposition of Vocal Inconsistencies. This Figure provides the distributions of vocal inconsistency measures across the topics in our taxonomy. We provide the differences in audio length ($\Delta Length$) and listenability ($\Delta Listenability$) between public and private calls at the topic level in Panel (a) and Panel (b), respectively. The decompositions of vocal valence inconsistency ($\Delta Valence$) and arousal inconsistency ($\Delta Arousal$) are reported in Panel (c) and Panel (d), respectively.

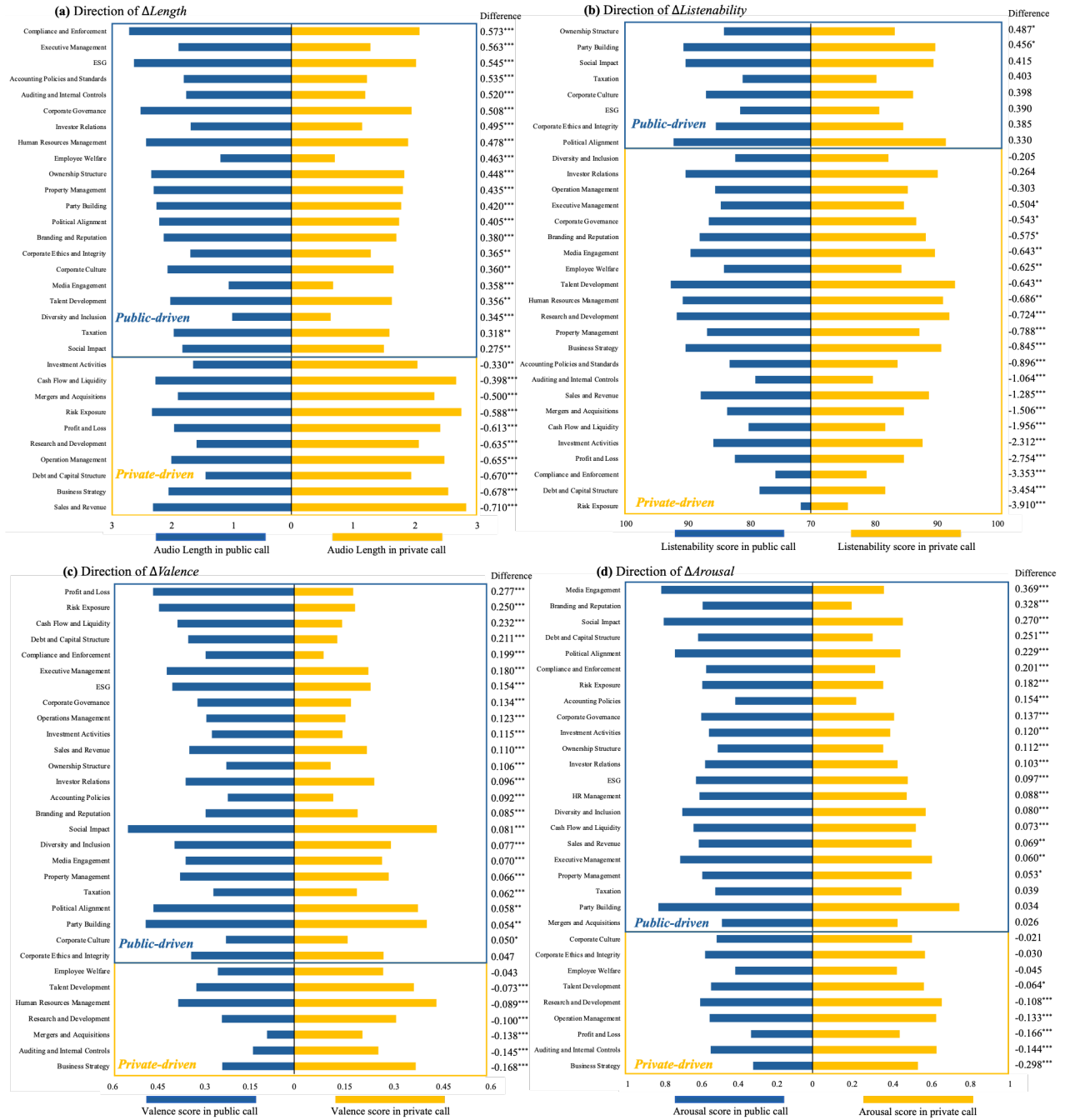


Figure 5 Direction of Vocal Inconsistencies. We provide the directional differences in audio length ($\Delta Number$), listenability ($\Delta Listenability$), vocal valence ($\Delta Valence$) and arousal ($\Delta Arousal$) between public and private earnings calls in Panels (a), (b), (c) and (d), respectively. In all panels, the left (right) bar reports the disclosure statistics in the public (private) call. For a given topic, the vocal inconsistency is classified as public-driven if the public disclosure statistic exceeds the private counterpart, and as private-driven otherwise. We provide the public-private differences (calculated as public minus private) and their statistical significance alongside the figures. *, **, *** are the statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 1 Summary Statistics

This Table reports the descriptive statistics of the disclosure inconsistency measures and the independent variables we introduced in Section 3. The detailed definitions of variables are provided in Table A1 of Appendix A1.

	Mean	Median	Std. Dev.	Q1	Q3	N
<i>Measures of Disclosure Inconsistency</i>						
#NewTopic	1.238	1	1.326	0	2	20,480
%NewTopic	0.135	0.129	0.155	0.087	0.173	20,480
#OmitTopic	1.813	1	2.092	0	3	20,480
ΔNumber	0.125	0.107	0.116	0.053	0.158	20,480
ΔNarrative	0.397	0.359	0.248	0.365	0.456	20,480
ΔSentiment	0.079	0.070	0.083	0.035	0.103	20,480
ΔLength	0.415	0.356	0.324	0.325	0.608	20,480
ΔListenability	1.024	0.677	1.646	0.369	1.284	20,480
ΔValence	0.121	0.115	0.146	0.063	0.158	20,480
ΔArousal	0.132	0.118	0.154	0.066	0.190	20,480
<i>Call Characteristics</i>						
Online	0.682	0.468	0.468	0	1	20,480
Team	0.819	0.404	0.404	0	1	20,480
Affiliation	0.091	0.087	0.055	0	0	20,480
<i>Manager Characteristics</i>						
Gender	0.112	0	0.135	0	0	20,480
Age	48	48	8.653	41	54	20,480
Tenure	3.295	3.082	2.948	0.783	5.459	20,480
Degree	0.127	0.116	0.094	0	0	20,480
<i>Firm Characteristics</i>						
IO	0.364	0.349	0.248	0	0.865	20,480
Leverage	0.520	0.501	0.193	0.389	0.680	20,480
Dividend	2.325	2.249	1.856	1.071	3.578	20,480
ROA	0.044	0.043	0.065	0.001	0.089	20,480
Firm age	18.284	17.475	9.070	11	25	20,480
Asset	19.883	19.860	1.795	18.617	21.980	20,480
SOE	0.512	0.512	0.478	0	1	20,480

Table 2 The Determinants of Disclosure Inconsistencies

In this table, we examine the determinants of disclosure inconsistencies between public-private call pairs at different levels. The dependent variables are the measures of topic inconsistencies (*#NewTopic*, *%NewTopic*, and *#OmitTopic*) in Columns (1)-(3), measures of textual inconsistencies (*ΔNarrative*, *ΔSentiment* and *ΔSentiment*) in Columns (4)-(6) and measures of vocal inconsistencies (*ΔLength*, *ΔListenability*, *ΔValence* and *ΔArousal*) in Columns (7)-(10). The variables of interest are included at the earnings call-, manager- and firm-level. Standard errors are clustered at the firm level, and *t*-statistics are reported in parentheses. *, **, *** are the statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	#NewTopic	%NewTopic	#OmitTopic	ΔNumber	ΔNarrative	ΔSentiment	ΔLength	ΔListenability	ΔValence	ΔArousal
Online	-0.164*** (-3.17)	-0.029*** (-4.64)	-0.277*** (-3.41)	-0.035** (-2.53)	-0.049*** (-2.76)	-0.011** (-2.11)	-0.051** (-2.40)	-0.160*** (-2.75)	-0.018** (-2.18)	-0.021*** (-3.17)
Team	-0.117** (-2.51)	-0.025*** (-3.15)	-0.176 (-0.92)	-0.019*** (-2.67)	-0.032** (-2.35)	-0.009** (-2.12)	-0.046* (-1.94)	-0.128*** (-2.35)	-0.010* (-1.85)	-0.012*** (-3.05)
Affiliation	0.281*** (2.86)	0.040*** (3.10)	0.389** (2.19)	0.022*** (3.23)	0.045*** (3.14)	0.003*** (3.08)	0.075** (2.48)	0.124*** (2.23)	0.006*** (3.47)	0.010*** (4.12)
Gender	-0.026** (-2.16)	-0.011 (-0.85)	-0.045** (-1.98)	-0.012 (-1.39)	-0.083** (-2.32)	-0.005** (-2.20)	-0.040 (-0.62)	-0.082* (-1.86)	-0.011** (-2.45)	-0.009* (-1.84)
Age	0.043*** (2.70)	0.009** (2.19)	0.075*** (3.10)	0.026*** (2.79)	0.094*** (2.83)	0.004** (2.42)	0.008 (1.32)	0.070 (1.27)	0.012** (2.25)	-0.013** (-2.47)
Tenure	0.074*** (3.42)	0.028*** (3.03)	0.150*** (3.74)	0.010*** (3.05)	0.056*** (3.01)	0.010** (2.17)	0.073* (1.90)	0.114*** (2.73)	0.019*** (2.80)	-0.012*** (-3.14)
Degree	-0.029*** (-2.83)	-0.024*** (-2.75)	-0.101*** (-3.85)	-0.017** (-2.50)	-0.061** (-2.18)	-0.012** (-2.29)	-0.065* (-1.87)	-0.150*** (-1.98)	-0.016*** (-3.25)	-0.008*** (-3.46)
IO	-0.139*** (-2.71)	-0.015** (-2.38)	-0.171*** (-3.75)	-0.004 (-1.17)	-0.027 (-0.87)	-0.004 (-1.05)	0.009 (0.91)	-0.073** (-2.34)	-0.008** (-2.50)	-0.014*** (-3.05)
Leverage	0.154*** (2.94)	0.016*** (3.10)	0.317*** (2.60)	0.022*** (2.99)	0.043*** (3.15)	0.005*** (3.04)	0.046*** (2.66)	0.059*** (2.02)	0.013** (2.03)	0.011** (2.40)
Dividend	0.033 (1.09)	0.008 (1.30)	0.009 (0.60)	-0.018** (-2.32)	-0.092** (-2.26)	-0.004*** (-3.21)	0.044 (0.63)	0.065 (0.96)	0.024 (1.32)	0.018 (1.29)

ROA	-0.248*** (-3.13)	-0.058*** (-4.19)	-0.205*** (-2.69)	-0.024* (-1.83)	-0.108** (-2.43)	-0.008*** (-2.82)	-0.054*** (-2.89)	-0.074* (-1.69)	-0.018*** (-3.60)	-0.019*** (-3.24)
Firm age	0.134* (1.89)	0.027 (1.28)	0.092 (0.95)	0.016** (2.09)	0.033*** (2.88)	0.004*** (2.84)	0.010 (0.77)	0.095** (2.03)	0.010** (2.11)	0.013*** (2.45)
Asset	-0.182*** (-2.44)	-0.019** (-2.51)	-0.265 (-1.31)	-0.009** (-2.30)	-0.007*** (-2.81)	-0.005** (-2.45)	-0.037*** (-2.80)	-0.075** (-2.26)	-0.013** (-1.99)	-0.015*** (-2.21)
SOE	-0.141*** (-3.71)	-0.043*** (-3.50)	-0.220*** (-2.87)	-0.016** (-2.10)	-0.020*** (-2.93)	-0.008*** (-2.79)	-0.009* (-1.71)	-0.041** (-2.40)	-0.002*** (-2.28)	-0.004*** (-2.79)
Fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	20,480	20,480	20,480	20,480	20,480	20,480	20,480	20,480	20,480	20,480
Adj. R ²	0.425	0.412	0.356	0.360	0.375	0.298	0.285	0.312	0.334	0.328

Table 3 The Informativeness of Topic Inconsistencies: Abnormal Stock Returns

In this table, we regress the absolute value of three-day abnormal returns (*ABET*) on the topic inconsistency measures for our sample firms in Panel A. We then examine the variations in the effects of *#NewTopic* and *%NewTopic* across textual and vocal characteristics in private meetings in Panel B and C, respectively. Control variables include a set of firm-level variables, including financial leverage ratio, dividend payout, return on assets, institutional ownership, firm age, total assets, and a state-owned enterprise dummy for each firm. Standard errors are clustered at the firm level, and *t*-statistics are reported in parentheses. *, **, *** are the statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Topic Inconsistencies and ABET

	(1)	(2)	(3)
#NewTopic	0.153*** (4.55)		
%NewTopic		0.147*** (4.47)	
#OmitTopic			-0.001 (-1.10)
Control variables	Yes	Yes	Yes
Fixed effect	Yes	Yes	Yes
Observation	20,480	20,480	20,480
Adj. R ²	0.196	0.197	0.194

Panel B: Variations in the Effects of #NewTopic

	(1)	(2)	(3)	(4)	(5)
#NewTopic	0.136*** (3.95)	0.152*** (4.12)	0.153*** (4.50)	0.150*** (4.23)	0.150*** (4.40)
#NewTopic*NumberPriv	0.042*** (3.36)				
#NewTopic*ListenabilityPriv		0.009 (1.60)			
#NewTopic*SentimentPriv			-0.001 (-0.85)		
#NewTopic*ValencePriv				0.013** (2.53)	
#NewTopic*ArousalPriv					0.011** (2.21)
Control variables	Yes	Yes	Yes	Yes	Yes
Fixed effect	Yes	Yes	Yes	Yes	Yes
Observation	20,480	20,480	20,480	20,480	20,480
Adj. R ²	0.199	0.196	0.196	0.197	0.197

Panel C: Variations in the Effects of %NewTopic

	(1)	(2)	(3)	(4)	(5)
%NewTopic	0.129** (3.43)	0.415*** (4.25)	0.147*** (4.30)	0.142*** (4.36)	0.142*** (4.24)
%NewTopic*NumberPriv	0.047*** (3.32)				
%NewTopic*ListenabilityPriv		0.007* (1.86)			
%NewTopic*SentimentPriv			0.002 (1.32)		
%NewTopic*ValencePriv				0.014** (2.06)	
%NewTopic*ArousalPriv					0.012* (1.74)
Control variables	Yes	Yes	Yes	Yes	Yes
Fixed effect	Yes	Yes	Yes	Yes	Yes
Observation	20,480	20,480	20,480	20,480	20,480
Adj. R ²	0.201	0.198	0.197	0.198	0.198

Table 4 The Informativeness of Textual and Vocal Inconsistencies: Abnormal Stock Returns

In this Table, we report the estimated coefficients for textual and vocal disclosure inconsistencies and their pairwise interactions on the absolute abnormal stock return (*ARET*). Panel A reports the results for textual inconsistencies, including ΔNumber , $\Delta\text{Narrative}$, and $\Delta\text{Sentiment}$. Panel B reports the results for vocal inconsistencies, including ΔLength , $\Delta\text{Listenability}$, $\Delta\text{Valence}$, and $\Delta\text{Arousal}$. Panel C reports the coefficients on pairwise interaction terms between textual inconsistency measures in rows and vocal inconsistency measures in columns. Control variables include a set of firm-level variables, including financial leverage ratio, dividend payout, return on assets, institutional ownership, firm age, total assets, and a state-owned enterprise dummy for each firm. Standard errors are clustered at the firm level, and *t*-statistics are reported in parentheses. *, **, *** are the statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Textual Inconsistencies

	(1)	(2)	(3)
ΔNumber	0.012 (1.49)		
ΔNarrative		0.049*** (2.99)	
ΔSentiment			0.018 (0.96)
Control variables	Yes	Yes	Yes
Fixed effect	Yes	Yes	Yes
Observation	20,480	20,480	20,480
Adj. R ²	0.194	0.196	0.194

Panel B: Vocal Inconsistencies

	(1)	(2)	(3)	(4)
ΔLength	0.002 (0.83)			
ΔListenability		0.003 (1.46)		
ΔValence			0.032* (1.78)	
ΔArousal				0.007 (0.87)
Control variables	Yes	Yes	Yes	Yes
Fixed effect	Yes	Yes	Yes	Yes
Observation	20,480	20,480	20,480	20,480
Adj. R ²	0.194	0.194	0.196	0.194

Panel C: Pairwise Interactions between Textual and Vocal Inconsistencies

	ΔLength	ΔListenability	ΔValence	ΔArousal
ΔNumber	0.006 (1.37)	0.024** (2.24)	0.030*** (2.65)	0.007 (0.87)
ΔNarrative	0.037** (2.17)	0.062*** (3.19)	0.064*** (3.25)	0.045*** (3.28)
ΔSentiment	0.018 (0.76)	0.023 (1.32)	0.017** (2.35)	0.015** (2.50)

Table 5 The Informativeness of Topic Inconsistencies: Abnormal Share Turnover

In this table, we regress the three-day abnormal share turnover (*ATURN*) on the topic inconsistency measures for our sample firms in Panel A. We then examine the variations in the effects of *#NewTopic* and *%NewTopic* across textual and vocal characteristics in private meetings in Panel B and C, respectively. Control variables include a set of firm-level variables, including financial leverage ratio, dividend payout, return on assets, institutional ownership, firm age, total assets, and a state-owned enterprise dummy for each firm. Standard errors are clustered at the firm level, and *t*-statistics are reported in parentheses. *, **, *** are the statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Topic Inconsistencies and ATURN					
	(1)	(2)	(3)		
#NewTopic	0.383*** (5.26)				
%NewTopic		0.355*** (5.12)			
#OmitTopic			0.057 (1.45)		
Control variables	Yes	Yes	Yes		
Fixed effect	Yes	Yes	Yes		
Observation	20,480	20,480	20,480		
Adj. R ²	0.238	0.238	0.235		
Panel B: Variations in the Effects of #NewTopic					
	(1)	(2)	(3)	(4)	(5)
#NewTopic	0.321*** (4.40)	0.380*** (5.10)	0.382*** (5.13)	0.354*** (4.75)	0.362*** (4.82)
#NewTopic*NumberPriv	0.107*** (3.46)				
#NewTopic*ListenabilityPriv		0.032* (1.85)			
#NewTopic*SentimentPriv			0.014* (1.94)		
#NewTopic*ValencePriv				0.079*** (3.11)	
#NewTopic*ArousalPriv					0.051*** (2.90)
Control variables	Yes	Yes	Yes	Yes	Yes
Fixed effect	Yes	Yes	Yes	Yes	Yes
Observation	20,480	20,480	20,480	20,480	20,480
Adj. R ²	0.245	0.240	0.238	0.240	0.239
Panel C: Variations in the Effects of %NewTopic					
	(1)	(2)	(3)	(4)	(5)
%NewTopic	0.310*** (4.19)	0.351*** (5.12)	0.355*** (5.12)	0.343*** (4.87)	0.331*** (4.85)
%NewTopic*NumberPriv	0.132*** (3.07)				
%NewTopic*ListenabilityPriv		0.025* (1.93)			
%NewTopic*SentimentPriv			0.010* (1.70)		
%NewTopic*ValencePriv				0.058*** (3.55)	
%NewTopic*ArousalPriv					0.060*** (3.47)
Control variables	Yes	Yes	Yes	Yes	Yes
Fixed effect	Yes	Yes	Yes	Yes	Yes
Observation	20,480	20,480	20,480	20,480	20,480
Adj. R ²	0.242	0.238	0.238	0.240	0.240

Table 6 The Informativeness of Textual and Vocal Inconsistencies: Abnormal Share Turnover

In this Table, we report the estimated coefficients for textual and vocal disclosure inconsistencies and their pairwise interactions on the abnormal share turnover (*ATURN*). Panel A reports the results for textual inconsistencies, including ΔNumber , $\Delta\text{Narrative}$, and $\Delta\text{Sentiment}$. Panel B reports the results for vocal inconsistencies, including ΔLength , $\Delta\text{Listenability}$, $\Delta\text{Valence}$, and $\Delta\text{Arousal}$. Panel C reports the coefficients on pairwise interaction terms between textual inconsistency measures in rows and vocal inconsistency measures in columns. Control variables include a set of firm-level variables, including financial leverage ratio, dividend payout, return on assets, institutional ownership, firm age, total assets, and a state-owned enterprise dummy for each firm. Standard errors are clustered at the firm level, and *t*-statistics are reported in parentheses. *, **, *** are the statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Textual Inconsistencies				
	(1)	(2)	(3)	
ΔNumber	0.062** (2.38)			
$\Delta\text{Narrative}$		0.103*** (2.96)		
$\Delta\text{Sentiment}$			0.017 (1.35)	
Control variables	Yes	Yes	Yes	
Fixed effect	Yes	Yes	Yes	
Observation	20,480	20,480	20,480	
Adj. R ²	0.236	0.237	0.235	
Panel B: Vocal Inconsistencies				
	(1)	(2)	(3)	(4)
ΔLength	0.024 (1.32)			
$\Delta\text{Listenability}$		0.029 (1.16)		
$\Delta\text{Valence}$			0.075** (2.48)	
$\Delta\text{Arousal}$				0.052* (1.90)
Control variables	Yes	Yes	Yes	Yes
Fixed effect	Yes	Yes	Yes	Yes
Observation	20,480	20,480	20,480	20,480
Adj. R ²	0.235	0.234	0.237	0.235
Panel C: Pairwise Interactions between Textual and Vocal Inconsistencies				
	ΔLength	$\Delta\text{Listenability}$	$\Delta\text{Valence}$	$\Delta\text{Arousal}$
ΔNumber	0.038 (1.05)	0.102*** (3.49)	0.085*** (3.24)	0.071** (2.36)
$\Delta\text{Narrative}$	0.075** (2.36)	0.117*** (3.60)	0.152*** (3.58)	0.097** (2.14)
$\Delta\text{Sentiment}$	0.028 (1.32)	0.044* (1.84)	0.079*** (3.23)	0.062** (2.19)

Table 7 Topic Inconsistencies and Attendee Trading

In this table, we regress the unsigned change in holdings (*Attendee Trading*) by private meeting attendees on the topic inconsistency measures for our sample firms in Panel A. We then examine the variations in the effects of *#NewTopic* and *%NewTopic* across textual and vocal characteristics in private meetings in Panel B and C, respectively. Control variables include a set of firm-level variables, including financial leverage ratio, dividend payout, return on assets, institutional ownership, firm age, total assets, and a state-owned enterprise dummy for each firm. Standard errors are clustered at the firm level, and *t*-statistics are reported in parentheses. *, **, *** are the statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Topic Inconsistencies and Attendee Trading					
	(1)	(2)	(3)		
#NewTopic	2.689*** (3.69)				
%NewTopic		2.530*** (4.28)			
#OmitTopic				0.063 (0.77)	
Control variables	Yes	Yes		Yes	
Fixed effect	Yes	Yes		Yes	
Observation	6,371	6,371		6,371	
Adj. R ²	0.363	0.354		0.350	
Panel B: Variations in the Effects of #NewTopic					
	(1)	(2)	(3)	(4)	(5)
#NewTopic	2.273*** (3.39)	2.688*** (3.61)	2.614*** (3.60)	2.430*** (3.23)	2.518*** (3.14)
#NewTopic*NumberPriv	0.842*** (3.35)				
#NewTopic*ListenabilityPriv		0.045 (1.39)			
#NewTopic*SentimentPriv			0.081* (1.76)		
#NewTopic*ValencePriv				0.403** (3.21)	
#NewTopic*ArousalPriv					0.189*** (3.45)
Control variables	Yes	Yes	Yes	Yes	Yes
Fixed effect	Yes	Yes	Yes	Yes	Yes
Observation	6,371	6,371	6,371	6,371	6,371
Adj. R ²	0.370	0.363	0.363	0.369	0.368
Panel C: Variations in the Effects of %NewTopic					
	(1)	(2)	(3)	(4)	(5)
%NewTopic	2.152** (3.43)	2.525*** (4.25)	2.528*** (4.20)	2.398*** (4.16)	2.427*** (4.19)
%NewTopic*NumberPriv	0.865*** (3.15)				
%NewTopic*ListenabilityPriv		0.070 (1.47)			
%NewTopic*SentimentPriv			0.092 (1.50)		
%NewTopic*ValencePriv				0.516** (2.25)	
%NewTopic*ArousalPriv					0.224*** (2.80)
Control variables	Yes	Yes	Yes	Yes	Yes
Fixed effect	Yes	Yes	Yes	Yes	Yes
Observation	6,371	6,371	6,371	6,371	6,371
Adj. R ²	0.359	0.354	0.355	0.358	0.358

Table 8 Textual and Vocal Inconsistencies and Attendee Trading

In this Table, we report the estimated coefficients for textual and vocal disclosure inconsistencies and their pairwise interactions on the attendee trading volume (*Attendee Trading*). Panel A reports the results for textual inconsistencies, including ΔNumber , $\Delta\text{Narrative}$, and $\Delta\text{Sentiment}$. Panel B reports the results for vocal inconsistencies, including ΔLength , $\Delta\text{Listenability}$, $\Delta\text{Valence}$, and $\Delta\text{Arousal}$. Panel C reports the coefficients on pairwise interaction terms between textual inconsistency measures in rows and vocal inconsistency measures in columns. Control variables include a set of firm-level variables, including financial leverage ratio, dividend payout, return on assets, institutional ownership, firm age, total assets, and a state-owned enterprise dummy for each firm. Standard errors are clustered at the firm level, and *t*-statistics are reported in parentheses. *, **, *** are the statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Textual Inconsistencies

	(1)	(2)	(3)
ΔNumber	0.529** (2.51)		
ΔNarrative		0.908*** (3.43)	
ΔSentiment			0.043 (1.06)
Control variables	Yes	Yes	Yes
Fixed effect	Yes	Yes	Yes
Observation	6,371	6,371	6,371
Adj. R ²	0.351	0.354	0.350

Panel B: Vocal Inconsistencies

	(1)	(2)	(3)	(4)
ΔLength	0.055 (1.15)			
ΔListenability		0.071 (1.27)		
ΔValence			0.161** (2.46)	
ΔArousal				0.096 (0.68)
Control variables	Yes	Yes	Yes	Yes
Fixed effect	Yes	Yes	Yes	Yes
Observation	6,371	6,371	6,371	6,371
Adj. R ²	0.350	0.350	0.352	0.350

Panel C: Pairwise Interactions between Textual and Vocal Inconsistencies

	ΔLength	ΔListenability	ΔValence	ΔArousal
ΔNumber	0.309 (0.96)	0.721** (1.98)	0.983*** (2.61)	0.685** (2.33)
ΔNarrative	0.455* (1.68)	1.153*** (2.86)	1.482*** (2.88)	1.158*** (2.71)
ΔSentiment	0.062 (0.40)	0.079 (0.71)	0.246*** (3.40)	0.113** (2.38)

Table 9 Topic Inconsistencies and Trading Gains

In this table, we regress the quarterly trading gains (*Trading Gains*) by private meeting attendees on the topic inconsistency measures for our sample firms in Panel A. We then examine the variations in the effects of *#NewTopic* and *%NewTopic* across textual and vocal characteristics in private meetings in Panel B and C, respectively. Control variables include a set of firm-level variables, including financial leverage ratio, dividend payout, return on assets, institutional ownership, firm age, total assets, and a state-owned enterprise dummy for each firm. Standard errors are clustered at the firm level, and *t*-statistics are reported in parentheses. *, **, *** are the statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Topic Inconsistencies and Trading Gains					
	(1)	(2)	(3)		
#NewTopic	0.167*** (2.79)				
%NewTopic		0.145** (2.36)			
#OmitTopic				0.018 (0.64)	
Control variables	Yes	Yes		Yes	
Fixed effect	Yes	Yes		Yes	
Observation	6,371	6,371		6,371	
Adj. R ²	0.108	0.108		0.105	
Panel B: Variations in the Effects of #NewTopic					
	(1)	(2)	(3)	(4)	(5)
#NewTopic	0.144** (2.30)	0.165*** (2.75)	0.164*** (2.78)	0.158** (2.52)	0.161*** (2.59)
#NewTopic*NumberPriv	0.045*** (2.83)				
#NewTopic*ListenabilityPriv		0.008 (1.19)			
#NewTopic*SentimentPriv			0.012 (0.94)		
#NewTopic*ValencePriv				0.028** (2.10)	
#NewTopic*ArousalPriv					0.024* (1.89)
Control variables	Yes	Yes	Yes	Yes	Yes
Fixed effect	Yes	Yes	Yes	Yes	Yes
Observation	6,371	6,371	6,371	6,371	6,371
Adj. R ²	0.114	0.108	0.109	0.112	0.111
Panel C: Variations in the Effects of %NewTopic					
	(1)	(2)	(3)	(4)	(5)
%NewTopic	0.132** (2.01)	0.144** (2.32)	0.145** (2.35)	0.139** (2.15)	0.141** (2.18)
%NewTopic*NumberPriv	0.050*** (3.15)				
%NewTopic*ListenabilityPriv		0.011 (1.41)			
%NewTopic*SentimentPriv			0.015 (0.80)		
%NewTopic*ValencePriv				0.031*** (2.74)	
%NewTopic*ArousalPriv					0.025** (2.43)
Control variables	Yes	Yes	Yes	Yes	Yes
Fixed effect	Yes	Yes	Yes	Yes	Yes
Observation	6,371	6,371	6,371	6,371	6,371
Adj. R ²	0.115	0.108	0.108	0.112	0.112

Table 10 Textual and Vocal Inconsistencies and Trading Gains

In this Table, we report the estimated coefficients for textual and vocal disclosure inconsistencies and their pairwise interactions on the trading gains (*Trading Gains*) by private meeting attendees. Panel A reports the results for textual inconsistencies, including ΔNumber , $\Delta\text{Narrative}$, and $\Delta\text{Sentiment}$. Panel B reports the results for vocal inconsistencies, including ΔLength , $\Delta\text{Listenability}$, $\Delta\text{Valence}$, and $\Delta\text{Arousal}$. Panel C reports the coefficients on pairwise interaction terms between textual inconsistency measures in rows and vocal inconsistency measures in columns. Control variables include a set of firm-level variables, including financial leverage ratio, dividend payout, return on assets, institutional ownership, firm age, total assets, and a state-owned enterprise dummy for each firm. Standard errors are clustered at the firm level, and *t*-statistics are reported in parentheses. *, **, *** are the statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Textual Inconsistencies

	(1)	(2)	(3)
ΔNumber	0.012 (0.72)		
$\Delta\text{Narrative}$		0.068** (2.04)	
$\Delta\text{Sentiment}$			0.017 (0.29)
Control variables	Yes	Yes	Yes
Fixed effect	Yes	Yes	Yes
Observation	6,371	6,371	6,371
Adj. R ²	0.105	0.108	0.105

Panel B: Vocal Inconsistencies

	(1)	(2)	(3)	(4)
ΔLength	0.003 (0.84)			
$\Delta\text{Listenability}$		0.008 (1.33)		
$\Delta\text{Valence}$			0.016 (1.07)	
$\Delta\text{Arousal}$				0.015 (0.79)
Control variables	Yes	Yes	Yes	Yes
Fixed effect	Yes	Yes	Yes	Yes
Observation	6,371	6,371	6,371	6,371
Adj. R ²	0.105	0.105	0.107	0.107

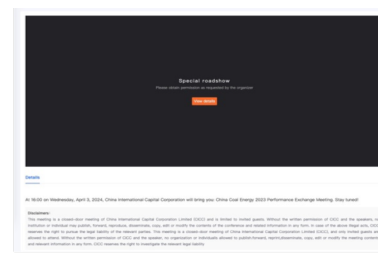
Panel C: Pairwise Interactions across Textual and Vocal Inconsistencies

	ΔLength	$\Delta\text{Listenability}$	$\Delta\text{Valence}$	$\Delta\text{Arousal}$
ΔNumber	0.019 (0.31)	0.021*** (2.75)	0.039** (2.23)	0.025 (1.09)
$\Delta\text{Narrative}$	0.034** (2.40)	0.089*** (3.01)	0.075*** (3.22)	0.067** (2.40)
$\Delta\text{Sentiment}$	0.417 (0.69)	0.010 (0.73)	0.034*** (2.77)	0.019 (1.51)

Appendix A1. Supplementary Figures and Tables



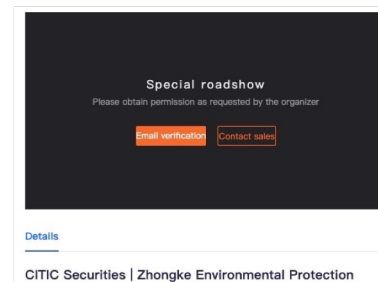
(a) Freely-access public earnings call (China Coal Energy, 601898.SH)



(b) Restricted-access private earnings call (China Coal Energy, 601898.SH)



(c) Freely-access public earnings call (Zhongke Environment, 301175.SZ)



(d) Restricted-access private earnings call (Zhongke Environment, 301175.SZ)

Figure A1 Examples of Public and Private Earnings Calls

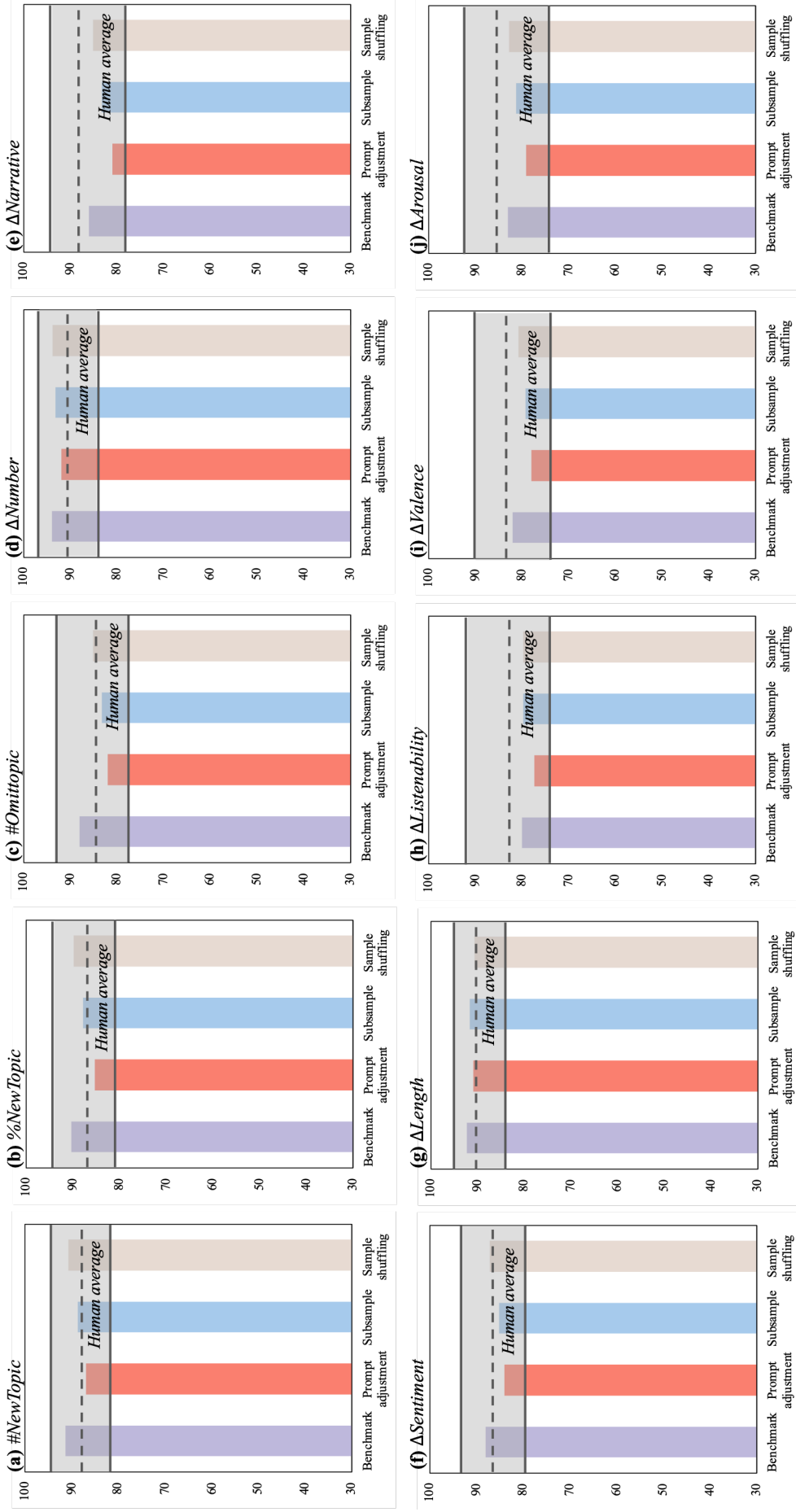


Figure A2 Turing Tests for LLMs' Disclosure Inconsistency Evaluations. This figure presents the performance of alternative prompting strategies and sampling methods in evaluating disclosure inconsistencies between public and private earnings calls. We compare the accuracy between LLM-generated scores and the assessments of five human annotators, using the human consensus score as the ground-truth scores. Agreement is measured as the accuracy relative to the human consensus scores. The solid lines represent the minimum and maximum accuracy scores of human readers, and dashed lines indicate the average accuracy across human readers. An LLM-based method is considered to pass the Turing test if its accuracy falls within the range of human annotator accuracies.

Table A1 Variable Definitions

Variable	Unit	Definition
#NewTopic	-	The number of new topics that discussed in private call are not disclosed in the public call
%NewTopic	-	The number of sentences labeled by TopicGPT as non-public-call topics in private call scaled by the total number of sentences in private call
#OmitTopic	-	The number of topics discussed in public call but not disclosed in the private call
Δ Number	-	The absolute differences of the numerical usage for the same topic disclosed between public and private calls
Δ Narrative	-	1-cosine similarity between narrative vectors in public call $V_{i,k,m,t}$ and private call $V_{i,k,n,t}$ for the same topic
Δ Sentiment	-	The absolute differences of sentiment scores for the same topic in public and private earnings calls
Δ Length	-	The absolute differences of the total audio length between public and private calls for the same topic
Δ Listenability	-	The absolute differences of the vocal listenability between public and private calls for the same topic
Δ Valence	-	The absolute differences of the vocal valence between public and private calls for the same topic
Δ Arousal	-	The absolute differences of the vocal arousal between public and private calls for the same topic
Online	-	Dummy that takes one if the public-private call pairs are both held online (no physical meeting location) otherwise zero
Teams	-	Dummy that takes one if the public-private call pairs are both held by firm executive teams otherwise zero
Affiliation	-	Dummy that takes one if the co-host financial institution was the lead or co-underwriter of an initial public offering of the firm during the past five years, or of a seasoned equity offering during the past two.
Gender	-	Dummy that takes one if the manager is a woman otherwise zero
Age	Year	The age of the firm manager
Tenure	Year	The number of years the manager has held the position at the firm
Degree	-	Dummy that takes one if the manager has a degree in finance or law related majors otherwise zero
IO	-	The percentage of the firm's shares that are owned by institutional investors such as mutual funds, pension

		funds, and insurance companies during the fiscal quarter
Leverage	-	The ratio of total debt to total assets of the firm during the fiscal quarter
Dividend	%	The amount of common dividend scaled by the net income during the fiscal quarter
ROA	-	Return on assets during the fiscal quarter
Firm age	Year	The number of years since the firm was founded
Asset	-	The natural log of total assets
SOE	-	Dummy that takes one if the firm is a state-owned enterprise otherwise zero

Table A2 Sample Construction

This Table tabulates the number of filters added to the sample for our main analysis. Given that our unit of observations is built on the public-private earnings call pairs, we filter the 115,372 earnings calls from Comein Finance database during our sample period using the following process. Filter [1] pairs of public and private calls being held by the same firm. Filter [2] identifies public-private call pairs discussing the same earnings release. Filter [3] requires the pairs of calls that are delivered by the same firm executives. Filter [4] further requires these firms to have non-missing firm-level variables, as outlined in Section 4.

	Number of call pairs	Number of call pairs dropped	Number of firms
Download earnings calls from Comein Finance Database	-	-	2,463
Filter [1]: Pairs of public and private calls being held by the same firm	36,902	-	1,395
Filter [2]: Pairs of public and private calls discussing the same earnings release	31,409	5,493	1,276
Filter [3]: Pairs of public and private calls that are delivered by the same firm executives	22,875	8,534	1,028
Filter [4]: Firm-level variables	20,480	2,395	976
Final Sample	20,480		976

Table A3 Performance Evaluation of LLM Modelling: Topic Generation and Assignment

In this Table, we assess the robustness of our identification framework in terms of topic generation and assignment. In Panel A, we use Purity, Adjusted Rand Index (ARI), and Normalized Mutual Information (NMI) to examine whether the latent topics identified by the TopicGPT are consistent with those recognized by human annotators. In Panel B, we introduce Accuracy, Precision, Recall, and F1 scores to examine whether TopicGPT assigns individual sentences to the same topics as human annotators.

<i>Panel A: Topic Generation Evaluation</i>				
	Benchmark	Prompt Adjustment	Subsample	Sample Shuffling
Purity	0.805	0.735	0.734	0.775
ARI	0.651	0.611	0.612	0.633
NMI	0.746	0.697	0.703	0.720

<i>Panel B: Topic Assignment Evaluation</i>				
	Benchmark	Prompt Adjustment	Subsample	Sample Shuffling
Accuracy	0.910	0.884	0.891	0.897
Precision	0.912	0.879	0.886	0.906
F1	0.908	0.863	0.875	0.886

Table A4 Topic Taxonomy

Code	Top topics	Code	Subtopics
01	Accounting Policies and Standards	0101	Accounting Methods
		0102	Accounting Valuation
		0103	Asset Impairment
		0104	Inventory Valuation
		0105	Revenue Recognition
		0106	Others
02	Auditing and Internal Controls	0201	Auditor Change
		0202	Audit Committee Review
		0203	Audit Committee Assessments
		0204	Audit Opinions
		0205	Internal Control Standard
		0206	Others
03	Branding and Reputation	0301	Branding Strategy
		0302	Company Mission and Slogan
		0303	Marketing Campaign
		0304	Media Engagement
		0305	Public Perception
		0306	Others
04	Business Strategy	0401	Business Model
		0402	Business Realignment
		0403	Digitalization Transformation
		0404	Diversification
		0405	Geographic Strategy
		0406	Strategic Vision
		0407	Others
05	Cash Flow and Liquidity	0501	Cash Reserve
		0502	Liquidity Risk
		0503	Short-Term Funding
		0504	Working Capital
		0505	Others
06	Compliance and Enforcement	0601	Compliance Reporting
		0602	Litigation and Legal Disputes
		0603	Penalties and Settlements
		0604	Regulatory Investigations
		0605	Others
07	Corporate Culture	0701	Code of Conduct
		0702	Core Values
		0703	Employee Training
		0704	Internal Communication
		0705	Others
08	Corporate Ethics and Integrity	0801	Anti-Corruption
		0802	Business Ethics
		0803	Compliance Culture

		0804	Ethical Leadership
		0805	Others
09	Corporate Governance	0901	Board Composition
		0902	Charter and Bylaws
		0903	Committee Structure
		0904	Director Independence
		0905	Stakeholder Representation
		0906	Others
10	Debt and Capital Structure	1001	Capital Adjustment
		1002	Debt Instruments
		1003	Debt Restructuring
		1004	Equity Issuance
		1005	Leverage Ratio
		1006	Share Buyback
		1007	Capital Structure Outlook
		1008	Others
11	Diversity and Inclusion	1101	Anti-Discrimination
		1102	Diverse Leadership Composition
		1103	Gender Equality
		1104	Inclusive Hiring
		1105	Others
12	Employee Welfare	1201	Employee Satisfaction
		1202	Labor Union
		1203	Mental Health
		1204	Work-life Balance
		1205	Others
13	ESG	1301	Carbon Emissions
		1302	Environmental Issues
		1303	ESG Rating
		1304	ESG Strategy
		1305	Governance Issues
		1306	Sustainability Issues
		1307	Others
14	Executive Management	1401	Executive Appointment and Termination
		1402	Executive Succession
		1403	Shareholder Activism
		1404	Team Restructuring
		1405	Others
15	Human Resources Management	1501	Employee Turnover and Retention
		1502	Labor Relations
		1503	Recruitment
		1504	Occupational Health and Safety
		1505	Workforce Planning
		1506	Others

16	Investment Activities	1601	Derivatives Exposure
		1602	Investment Strategy
		1603	Real Estate Investments
		1604	Securities Holdings
		1605	Trading Income
		1606	Others
17	Investor Relations	1701	Analyst Report
		1702	Earnings Call
		1703	Institutional Outreach
		1704	Roadshow Events
		1705	Others
18	Media Engagement	1801	Executive Interviews
		1802	Media Communication
		1803	Press Releases
		1804	Social Media Posting
		1805	Others
19	Mergers and Acquisitions	1901	Acquisition Strategy
		1902	Asset Divestiture
		1903	Business Separation
		1904	Merger Integration
		1905	Transaction Financing
		1906	Others
20	Operation Management	2001	Inventory Control
		2002	Procurement Optimization
		2003	Product and Service Operations
		2004	Supply Chain
		2005	Vendor and Contractor
		2006	Others
21	Ownership Structure	2101	Insider Holdings
		2102	Major Ownership Changes
		2104	Shareholder Rights
		2105	Others
22	Party Building	2201	Ideological Education
		2202	Party Activities
		2203	Party Leadership
		2204	Party Organization
		2205	Thematic Education
		2206	Others
23	Political Alignment	2301	Belt and Road Initiative
		2302	Common Prosperity
		2303	Dual Carbon Goals
		2304	Rural Revitalization
		2305	Others
24	Profit and Loss	2401	Earnings Breakdown
		2402	Earning Outlook

		2403	Profit Margin
		2404	Unexpected Loss
		2405	Others
25	Property Management	2501	Facility Upgrades
		2502	Lease Arrangements
		2503	Property Acquisition
		2504	Property Disposal
		2505	Space Utilization
		2506	Others
26	Research and Development	2601	Licensing Rights
		2602	Innovation
		2603	Intellectual Property
		2604	Patents
		2605	Technology Commercialization
		2606	Others
27	Risk Exposure	2701	Cybersecurity Risk
		2702	Covid-19
		2703	Geopolitical Risks
		2704	Macro Policy Risk
		2705	Technology Restriction
		2706	Trade war
		2707	Others
28	Sales and Revenue	2801	Competitive Pressure
		2802	Market Demand
		2803	Pricing Strategy
		2804	Revenue Composition
		2805	Sales Performance Indicators
		2806	Others
29	Social Impact	2901	Charity
		2902	Donations
		2903	Non-profit Collaborations
		2904	Social Awards
		2905	Others
30	Talent Development	3001	Career Pathing
		3002	Internship Programs
		3003	Mentorship Systems
		3004	Rotation and Mobility
		3005	Skill Training
		3006	Others
31	Taxation	3101	Tariff
		3102	Deferred Tax
		3103	Tax investigations
		3104	Tax Planning
		3105	Others

Table A6 Top and Bottom Topic Inconsistencies: Subtopic Statistics

This Table reports the subtopic statistics of topic inconsistencies between public and private meetings. For each measure of topic inconsistencies, we present the 5 top- and bottom-ranked top-subtopics across topics in our taxonomy.

#NewTopic			%NewTopic			#OmitTopic		
Top Topic	Subtopic	Mean	Top Topic	Subtopic	Mean	Top Topic	Subtopic	Mean
Top-Ranked Top-subtopics								
Corporate Governance	Board Composition	0.075	Business Strategy	Strategic Vision	0.035	Political Alignment	Others	0.089
Operation Management	Supplier Chain	0.052	Operation Management	Supplier Chain	0.030	Social Impact	Donations	0.083
Research and Development	Technology Commercialization	0.029	Research and Development	Technology Commercialization	0.026	Media Engagement	Press Releases	0.077
Investment Activities	Investment Strategy	0.021	Investment Activities	Investment Strategy	0.024	Party Building	Party Activities	0.074
Business Strategy	Strategic Vision	0.015	Corporate Governance	Board Composition	0.024	Talent Development	Skill Training	0.070
Bottom-Ranked Top-subtopics								
Diversity and Inclusion	Gender Equality	0	Diversity and Inclusion	Gender Equality	0	Cash Flow and Liquidity	Liquidity Risk	0
Corporate Culture	Others	0	Corporate Culture	Others	0	Business Strategy	Strategic Vision	0
Political Alignment	Others	0	Political Alignment	Others	0	Mergers and Acquisitions	Merger Integration	0
Party Building	Party Activities	0	Party Building	Party Activities	0	Risk Exposure	Macro Policy Risk	0
Property Management	Others	0	Property Management	Others	0	Debt and Capital Structure	Leverage Ratio	0

Table A7 Top and Bottom Textual Inconsistencies: Subtopic Statistics

This Table reports the subtopic statistics of textual inconsistencies between public and private meetings. For each measure of textual inconsistencies, we present the 5 top- and bottom-ranked top-subtopics across topics in our taxonomy.

Δ Number			Δ Narrative			Δ Sentiment		
Top Topic	Subtopic	Mean	Top Topic	Subtopic	Mean	Top Topic	Subtopic	Mean
<i>Top-ranked Top-subtopics</i>								
Business Strategy	Strategic Vision	0.228	Business Strategy	Strategic Vision	0.704	Profit and Loss	Unexpected Loss	0.133
Business Strategy	Digitalization Transformation	0.216	Business Strategy	Business Model	0.676	Risk Exposure	Macro Policy Risk	0.122
Mergers and Acquisitions	Merger Integration	0.202	Mergers and Acquisitions	Acquisition Strategy	0.671	Risk Exposure	Trade war	0.115
Mergers and Acquisitions	Acquisition Strategy	0.190	Operation Management	Supplier Chain	0.668	Profit and Loss	Earning Outlook	0.113
Operation Management	Supplier Chain	0.185	Debt and Capital Structure	Capital Structure Outlook	0.664	Compliance and Enforcement	Penalties and Settlements	0.112
<i>Bottom-ranked Top-subtopics</i>								
Party Building	Ideological education	0.025	Human Resources Management	Labor Relations	0.192	Corporate Ethics and Integrity	Others	0.012
Party Building	Party Activities	0.028	Corporate Ethics and Integrity	Business Ethics	0.204	Party Building	Party organization	0.018
Political Alignment	Common Prosperity	0.29	Corporate Ethics and Integrity	Others	0.209	Political Alignment	Common Prosperity	0.021
Corporate Ethics and Integrity	Business Ethics	0.031	Party Building	Party Activities	0.215	Taxation	Others	0.022
Human Resources Management	Labor Relations	0.032	Party Building	Ideological education	0.217	Media Engagement	Media Communication	0.022

Table A8 Top and Bottom Vocal Inconsistencies: Subtopic Statistics

This Table reports the subtopic statistics of vocal inconsistencies between public and private meetings. For each measure of vocal inconsistencies, we present the 5 top- and bottom-ranked top-subtopics across topics in our taxonomy.

Δ Length			Δ Listenability			Δ Valence			Δ Arousal		
Top Topic	Subtopic	Mean	Top Topic	Subtopic	Mean	Top Topic	Subtopic	Mean	Top Topic	Subtopic	Mean
<i>Top-Ranked Top-subtopics</i>											
Business Strategy	Strategic Vision	0.728	Risk Exposure	Macro Policy Risk	4.025	Profit and Loss	Unexpected Loss	0.286	Media Engagement	Others	0.370
Debt and Capital Structure	Debt Instruments	0.716	Risk Exposure	Trade War	3.846	Risk Exposure	Macro Policy Risk	0.278	Branding and Reputation	Branding Strategy	0.362
Operation Management	Supplier Chain	0.702	Debt and Capital Structure	Leverage Ratio	3.657	Risk Exposure	Trade war	0.264	Media Engagement	Media Communication	0.351
Research and Development	Intellectual Property	0.696	Risk Exposure	Covid-19	3.593	Profit and Loss	Earning Outlook	0.249	Business Strategy	Business Model	0.346
Profit and Loss	Earning Outlook	0.689	Compliance and Enforcement	Penalties and Settlements	3.443	Cash Flow and Liquidity	Liquidity Risk	0.245	Social Impact	Social awards	0.317
<i>Bottom-Ranked Top-subtopics</i>											
Social Impact	Charity	0.252	Diversity and Inclusion	Gender Equality	0.157	Employee Welfare	Mental Health	0.033	Corporate Culture	Core values	0.022
Social Impact	Donations	0.258	Diversity and Inclusion	Others	0.184	Corporate Ethics and Integrity	Business Ethics	0.039	Political Alignment	Common Prosperity	0.027
Taxation	Tax Planning	0.263	Investor Relations	Earnings Call	0.216	Corporate Culture	Core values	0.044	Corporate Ethics and Integrity	Business Ethics	0.029
Diversity and Inclusion	Others	0.270	Ownership Structure	Others	0.242	Party Building	Party organization	0.051	Party Building	Party organization	0.030
Media	Media	0.277	Political Alignment	Common Prosperity	0.265	Political Alignment	Common Prosperity	0.056	Political Alignment	Rural Revitalization	0.030
Engagement	Communication										

Table A9 Examples of Textual Inconsistencies

This Table provides example transcripts of textual inconsistencies between public and private earnings calls, including differences in numerical usage ($\Delta Number$), disclosure narrative ($\Delta Narrative$), and text-based sentiment ($\Delta Sentiment$). For each textual inconsistency measure, we provide the transcripts of call pairs with both high and low inconsistency measures in the respective dimension.

	Public Call Transcript	Private Call Transcript
Call pair with high $\Delta Number$	Our strategy focuses on long-term growth in intelligent solutions, with R&D investment remaining stable at around 10% of revenue. We expect overall revenue growth to stay broadly in line with the industry average over the next few years.	Internally, we plan to raise R&D intensity from 10% to 15% over the next year, while reallocating nearly 30% of capital expenditure away from traditional hardware businesses. This adjustment is expected to increase long-term revenue growth by 3–5 percentage points.
Call pair with low $\Delta Number$	The company organized more than 20 thematic party-building activities, with full participation from core management. These efforts continue to support compliant and stable operations.	We conducted over 20 party-building and ideological education sessions, achieving full participation among senior executives. This work remains a routine component of corporate governance.
Call pair with high $\Delta Narrative$	The company will steadily advance core businesses, optimize product structure, and enhance operational efficiency, while maintaining sufficient strategic flexibility in response to macroeconomic uncertainty and evolving industry conditions.	Over the next two years, the company will pivot aggressively toward intelligent solutions, exit low-margin legacy projects, and rebuild its revenue structure around software and platform-based businesses, with China as the primary testing ground before global rollout.
Call pair with Low $\Delta Narrative$	The company continues to optimize its talent structure, strengthen internal training systems, and improve organizational efficiency to support long-term development.	The company is further optimizing its talent structure, enhancing internal training mechanisms, and improving organizational efficiency to support sustainable development.
Call pair with high $\Delta Sentiment$	In the quarter, the advanced ceramics and electronic materials businesses continued to contribute the majority of revenue. Although production costs increased, management believes the overall profitability remains manageable and performance is expected to improve gradually as utilization rates recover.	Although the advanced ceramics and electronic materials businesses continued to contribute the revenue, declining prices in the dental ceramics business and delayed orders from automotive customers have materially weakened overall profit expectations for the second half of the year.
Call pair with low $\Delta Sentiment$	In the poultry industry cycle analysis, the company continues to emphasize strict compliance with regulatory requirements in breeding, epidemic prevention, and food safety across the entire value chain, including parent stock procurement, broiler production, and downstream processing.	During the same poultry cycle, the company continues to place high priority on compliance with breeding standards, disease control protocols, and food safety regulations across parent stock sourcing, broiler farming, and processing operations.

Table A10 Disclosure Inconsistencies and Non-attendee Trading

Table A10 reports placebo tests using non-attendee trading as the dependent variable. Non-attendee trading is measured as the aggregate unsigned change in holdings from quarter t to quarter $t + 1$ by institutional investors that hold shares of firm i in quarter t but do not participate in firm i 's private meeting. Panel A through C report the results for topic, textual and vocal inconsistency measures, respectively. Panel D reports the coefficients on pairwise interaction terms between textual inconsistency measures in rows and vocal inconsistency measures in columns. Control variables include firm-level characteristics. Standard errors are clustered at the firm level, and t -statistics are reported in parentheses. *, **, *** are the statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Topic Inconsistencies				
#NewTopic	0.121 (0.98)			
%NewTopic		0.116 (0.75)		
#OmitTopic			0.058 (0.24)	
Control variables	Yes	Yes	Yes	Yes
Fixed effect	Yes	Yes	Yes	Yes
Observation	6,371	6,371	6,371	6,371
Adj. R ²	0.350	0.350	0.350	0.350
Panel B: Textual Inconsistencies				
	(1)	(2)	(3)	
ΔNumber	0.051 (0.56)			
ΔNarrative		0.062 (1.01)		
ΔSentiment			0.041 (0.52)	
Control variables	Yes	Yes	Yes	Yes
Fixed effect	Yes	Yes	Yes	Yes
Observation	6,371	6,371	6,371	6,371
Adj. R ²	0.351	0.350	0.350	0.350
Panel C: Vocal Inconsistencies				
	(1)	(2)	(3)	(4)
ΔLength	0.033 (0.69)			
ΔListenability		0.038 (0.60)		
ΔValence			0.027 (0.71)	
ΔArousal				0.044 (0.92)
Control variables	Yes	Yes	Yes	Yes
Fixed effect	Yes	Yes	Yes	Yes
Observation	6,371	6,371	6,371	6,371
Adj. R ²	0.350	0.350	0.351	0.350
Panel D: Pairwise Interactions across Textual and Vocal Inconsistencies				
	ΔLength	ΔListenability	ΔValence	ΔArousal
ΔNumber	0.058 (0.31)	0.022 (0.51)	0.024 (0.68)	0.042 (0.76)
ΔNarrative	0.086 (0.53)	0.109 (1.21)	0.095 (1.05)	0.066 (0.48)
ΔSentiment	0.022 (0.33)	0.045 (0.78)	0.041 (0.54)	0.036 (1.30)

Table A11 Disclosure Inconsistencies and Non-attendee Trading Gains

Table A11 reports placebo tests using non-attendee trading gain as the dependent variable. Non-attendee trading is measured as the aggregate quarterly trading gains from quarter t to quarter $t + 1$ by institutional investors that hold shares of firm i in quarter t but do not participate in firm i 's private meeting. Panel A through C report the results for topic, textual and vocal inconsistency measures, respectively. Panel D reports the coefficients on pairwise interaction terms between textual inconsistency measures in rows and vocal inconsistency measures in columns. Control variables include firm-level characteristics. Standard errors are clustered at the firm level, and t -statistics are reported in parentheses. *, **, *** are the statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Topic Inconsistencies				
#NewTopic	0.015 (1.20)			
%NewTopic		-0.018 (-1.15)		
#OmitTopic				-0.003 (-0.41)
Control variables	Yes	Yes		Yes
Fixed effect	Yes	Yes		Yes
Observation	6,371	6,371		6,371
Adj. R ²	0.105	0.106		0.105
Panel B: Textual Inconsistencies				
	(1)	(2)		(3)
ΔNumber	0.008 (0.44)			
ΔNarrative		-0.012 (-0.78)		
ΔSentiment				-0.003 (-0.26)
Control variables	Yes	Yes		Yes
Fixed effect	Yes	Yes		Yes
Observation	6,371	6,371		6,371
Adj. R ²	0.105	0.105		0.105
Panel C: Vocal Inconsistencies				
	(1)	(2)	(3)	(4)
ΔLength	0.005 (0.20)			
ΔListenability		-0.003 (0.63)		
ΔValence			0.008 (1.17)	
ΔArousal				0.007 (0.74)
Control variables	Yes	Yes	Yes	Yes
Fixed effect	Yes	Yes	Yes	Yes
Observation	6,371	6,371	6,371	6,371
Adj. R ²	0.106	0.105	0.105	0.105
Panel D: Pairwise Interactions across Textual and Vocal Inconsistencies				
	ΔLength	ΔListenability	ΔValence	ΔArousal
ΔNumber	-0.005 (-0.50)	-0.007 (-0.83)	0.009 (0.71)	0.010 (0.62)
ΔNarrative	-0.003 (-0.48)	-0.015 (-0.82)	-0.004 (-1.31)	0.000 (1.35)
ΔSentiment	-0.001 (-0.35)	0.008 (0.95)	0.005 (1.20)	0.004 (0.88)

Appendix A2 Technical Appendix

1. The Pipeline for TopicGPT Modelling

The pipeline of TopicGPT modelling consists of two steps: First, we use GPT-4o to obtain exhaustive top-subtopic lists for the presentation section of each earnings call. Second, for each item, we create a labeled training set by instructing GPT-4o to classify sentences into top-level topics and subtopics for a sample of earnings calls. We then train a multilayer perceptron (MLP) on the labeled sample and extend it to the full sample. The pipeline of TopicGPT modelling employed in this paper is provided in Figure A3.

1.1 Topic Generation

1.1.1 Data Preprocessing

We obtain 115,357 conference calls from the Comin Finance database for publicly listed companies in Chinese stock markets during the period of January 2020 to December 2024. Comin Finance is the leading call provider that streams real-time corporate conference events and converts audio to structured text records. After filtering, we derive 20,480 pairs of public and private calls across 976 publicly listed companies in China. Each earnings call is then segmented into sentence-level units via deterministic rules using sentence-ending punctuation. The sentence-level units are used as the basic inputs for topic generation to ensure consistent granularity and reduce cross-sentence ambiguity.

1.1.2 Topic Initialization and Expansion

We initialize the topic set K with a small number of human-written example topics (three in our implementation). Each example topic includes a concise label and a one-sentence, domain-agnostic description. These topics serve purely as few-shot demonstrations of the desired output format and are not designed to cover the full topical space of the data.

Next, we collect a sample of firms based on the China Securities Regulatory Commission (CRSC) industry codes and three two-year time windows (spanning our sample) into 30 strata (15 industries $\times$ 2 periods). For each of the resulting time-industry strata, we randomly sample five earnings calls and feed these five call transcripts into GPT-4o. The initial sample of 150 call transcripts (30 strata $\times$ five documents) serves as our basis for generating topic lists.

For each sentence S drawn from a random sample of earnings calls, we prompt the GPT-4o with the topic branches that contain top-level topic K_t , subtopic K_s and instruct it to either:

1. assign S to an existing topic branch in K , or
2. generate a new topic branch that better captures the content of S .

Table A9 presents the prompts we used to instruct ChatGPT 4o for topic generations. A topic branch is defined as a top topic label coupled with a subtopic label (both with digital code). The iterative design ensures that newly generated topics match the level of granularity and abstraction of previous topics. New topics proposed by the model are appended to K , and the procedure continues until the topic set converges (i.e., no new topics are produced over multiple iterations). Running the generation step on a representative sample, rather than the entire corpus, substantially reduces computational cost while preserving topical diversity.

1.1.3 Topic Refinement

The raw topic set produced from the expansion step may contain near-duplicates and low-utility topics. We encode each topic-subtopic using Sentence-Transformer embeddings and compute pairwise cosine similarities. Topic pairs with similarity ≥ 0.6 are treated as candidate duplicates. For each such pair, we prompt the ChatGPT 4o with both topic descriptions and instruct it to determine whether the pair reflects the same conceptual theme. If so, the model produces a merged topic representation, which replaces the two originals.

Next, we aggregate the top-subtopic lists generated from each industry into a single unified set. To prevent the taxonomy from inheriting industry-specific noise, we explicitly instruct GPT-4o to avoid industry-dependent terminology and to derive subtopics that reflect cross-industry regularities in managerial disclosure. The prompts we used to instruct ChatGPT 4o for topic refining and merging are provided in Table A13. The resulting aggregation produces a comprehensive and industry-independent top-subtopic list that serves as the basis for subsequent topic assignment, which consists of 31 top topics and 150 subtopics.

1.2 Topic Assignment

1.2.1 Labeled Training Set

To construct a supervised training set for downstream classification, we begin with a sample of 300 documents consisting of the initial 150 transcripts and an additional stratified random sample of 150 transcripts. Each document is segmented into sentence-level units. For every sentence in this subset, GPT-4o is prompted to assign:

1. a top topic label from the predefined topic taxonomy:

2. a subtopic label from the corresponding subtopic list; and
3. a textual sentiment score associated with the input sentence.

If none of the existing subtopics appropriately capture the sentence content, GPT-4o is instructed to assign the sentence to an “Others” category. Table A14 provides prompt we used to instruct ChatGPT 4o for topic assignment and sentiment analysis. To mitigate hallucinated or invalid classifications (e.g., incorrect formatting, low quality, out-of-taxonomy labels), we incorporate a self-correction mechanism following Shinn et al. (2023). Specifically, we develop a rule-based parser that identifies assignment errors such as, invalid or non-existent topic codes, multi-topic combinations violating prompt constraints, and structural output inconsistencies. Next, we provide the GPT-4o with the identified documents along with the error type and prompt the model to reassign a valid topic. The corrected output is retained only if it follows the prescribed formatting and taxonomy. A sample topic assignment is:

[24] Profit and Loss
[2402] Earning Outlook
Sentiment Score: 0.15

Because a typical earnings call transcript contains approximately 300 sentences, only the 300-document subsample is processed through this LLM annotation pipeline to balance coverage and computational cost.

1.2.2 Topic Assignment for the Full Sample

In this step, we convert each sentence into a textual vector using the OpenAI *text-embedding-3-large* model. OpenAI embeddings include 3,072 elements ordered by importance. We condense the dimensionality from 3,072 to 256 for computational efficiency. To reduce computational load and mitigate overfitting, we apply dimensionality reduction to obtain 256-dimensional embeddings, which serve as input features for the classifier.

We then train a multilayer perceptron (MLP) on the labeled dataset using 256-dimensional *text-embedding-3-large* sentence embeddings as input, and the outputs are: (i) top topic code, (ii) subtopic code, and (iii) sentiment score. The loss functions of MLP are the cross-entropy for topic/subtopic classification and mean-squared error for sentiment regression. The model is trained using an 80/20 train-validation split with early stopping. Hyperparameters are tuned to optimize the F1-score for

topic accuracy and the RMSE for sentiment prediction. After training, the MLP is deployed to assign topic labels, subtopic labels, and sentiment values for all sentences in the full sample of earnings calls.

2. Measuring Vocal Characteristics of Corporate Disclosure

In this section, we provide the full technical details underlying the construction of our audio-based measures of vocal inconsistencies across public and private earnings calls. The procedure consists of three stages: (i) sentence-level alignment of transcripts and audio, (ii) extraction of listenability scores, and (iii) derivation of valence-arousal emotion measures.

2.1 Alignment of Audio Recording and Transcripts

We begin with paired public and private earnings call recordings obtained from the Comein Finance database and their corresponding human-edited transcripts. Because identifying vocal characteristics requires precise mapping between spoken audio and textual sentences, we adopt a two-step alignment framework based on WhisperX (multilingual large-v2 model) developed by OpenAI and *ccalign* package developed by Ewertz et al. (2026). The Inputs of our alignment framework include the 40,960 audio files of public and private call, the corresponding transcripts provided by Comein Finance, and the topic labels for each transcript sentence (derived from the TopicGPT modelling) .

Step 1: WhisperX-based preliminary timestamping

We run WhisperX multilingual large-v2 model on the raw audio of each earnings call. WhisperX returns word-level timestamps, generated text, and the confidence scores. These timestamps serve as the initial noisy mapping between audio recording and transcript text for public and private call pairs.

Step 2: Forced alignment with transcript sentences

We employ the *ccalign* Python package to align each transcript sentence to WhisperX output in a forced-alignment-like procedure, where the numbers and symbols in each earnings call are normalized. For each transcript sentence, *ccalign* outputs: i) beginning and ending timestamps $[t_s^{start}, t_s^{end}]$; ii) word-level alignment statistics; forced-alignment word-level timestamps. The outputs of *ccalign* algorithm are stored in a directory called *ccalign_results*.

By incorporating the two-step alignment framework, we obtain sentence-level audio-text alignments with exact timestamps for each sentence in both public and private calls. These timestamps enable the extraction of sentence-specific acoustic properties and consistent aggregation at the topic level.

2.2 Construction of Listenability

Next, we use the aligned timestamps to construct the measure of listenability. For each sentence, WhisperX provides posterior confidence scores for all recognized words, reflecting the automatic speech recognition model’s certainty in decoding the spoken content. Sentence-level listenability is defined as the average of these word-level confidence values:

$$Listenability_{i,t}(s) = \frac{1}{N_s} \sum_{w \in S} Conf(w) \quad (8)$$

where $Listenability_{i,t}$ is the sentence-level listenability of managers’ presentation for firm i in quarter t . $Conf(w)$ denotes the confidence score of word w generated by *WhisperX*, and N_s is the number of words in sentence s . Higher values of Listenability indicate clearer and more intelligible speech, whereas lower values reflect ambiguous presentation for the *WhisperX* system to recognize. To quantify topic-level inconsistency in Listenability, we aggregate sentence-level values within each topic and compute $\Delta Listenability$ as the absolute difference between the public and private call for the same topic.

2.3 Construction of Valence and Arousal Measures

To capture affective components of vocal behavior not reflected in the measure of listenability, we extract emotional features along the valence-arousal dimensions using a pre-trained Wav2Vec2 speech emotion recognition model (Wagner et al., 2023). Wav2Vec2 was trained on audiobooks, collected crowd-sourced read-out text snippets, and (noisy) telephone speech recordings. These training materials enable Wav2Vec2 to recognize vocal emotions in a more robust manner.

We feed each aligned audio segment into a pre-trained Wav2Vec2 model fine-tuned for speech emotion recognition. For each sentence-level audio segment, the model outputs continuous *valence* and *arousal* statistics. Valence measures the positive affectivity of the vocal feature, while arousal captures the strength of such an attitude (exciting versus calm). We aggregate these sentence-level scores to the topic level and compute $\Delta Valence$ and $\Delta Arousal$ as the absolute differences between public and private calls on the same topic.

Together, these audio-based measures, including speech duration (length), listenability, valence, and arousal, provide a comprehensive representation of how managers’ vocal behavior differs across meeting occasions and complement the text-based measures used in the analysis.

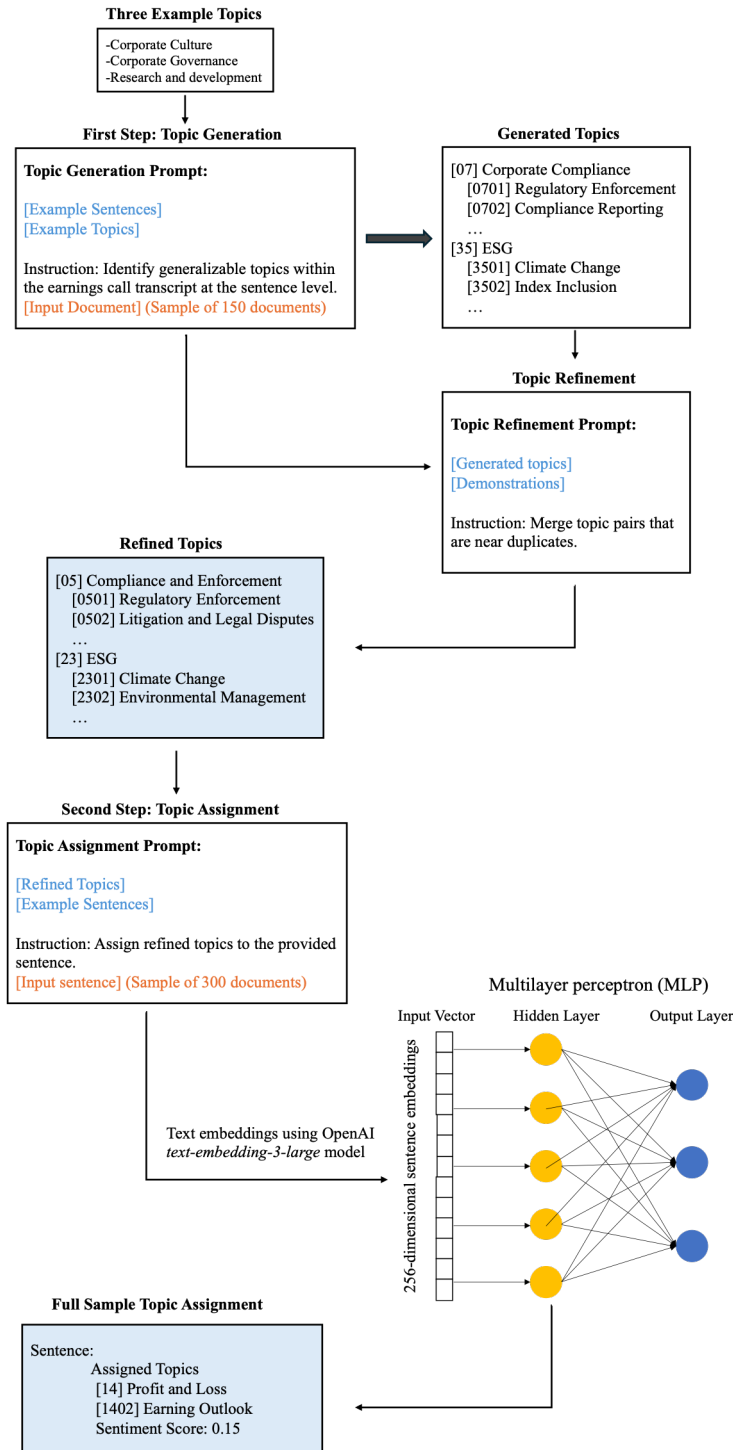


Figure A3 Pipeline of Topic Modelling using TopicGPT. In this paper, the TopicGPT modelling consists of two steps. 1) Topic Generation: Given three manually curated example topics, TopicGPT generates new topics in the 150 sample earnings call documents. The framework then refines the list by merging repeated topics and removing infrequent topics. 2) Topic Assignment: Given the generated topics, TopicGPT assigns the most relevant topic to each sentence for 300 sample documents. The labeled sentences are then converted into textual vectors using the OpenAI *text-embedding-3-large* model. The framework trains a multilayer perceptron (MLP) on the labeled dataset and deploys this classifier to generate topic-subtopic classifications and sentiment assignments for the entire sample of sentences in earnings calls.

Table A12 Prompt Template for Topic Generation

Prompt for topic generation

You will receive a sentence from an earnings call transcript dataset and a set of top-subtopics from a topic taxonomy. Your task is to identify generalizable top-subtopics within the sentence that can act as top-level topics and the associated subtopics in the taxonomy. If any relevant topics are missing from the provided set, please add them. Otherwise, output the existing top-subtopics as identified in the sentence.

[Top-subtopic List]

[01] Corporate Culture

[0101] Core Values

[02] Corporate Governance

[0201] Committee structure

[03] Research and Development

[0301] Patents

[Examples]

Example 1

Sentence:

Our company embraces a culture of integrity and innovation, committed to green development through energy conservation, emission reduction, and sustainable growth.

Your response:

[01] Corporate Culture

[0102] Core Values

Example 2

Sentence:

We have accelerated the development of the AVIDM industrial software series, released product versions on schedule, and secured eight patents, including one for parametric generation of 3D models.

Your response:

[03] Research and Development

[0301] Patents

[Instructions]

Step 1: Determine topics mentioned in the sentence.

- The topic labels must be as GENERALIZABLE as possible. They must not be industry-specific.
- The topics must reflect a SINGLE topic instead of a combination of topics.
- The new topics must have a level number, a general top-level label, and a general sub-level label.
- The topics must be broad enough to accommodate future top-subtopics.

Step 2: Perform ONE of the following operations:

1. If there are already duplicates or relevant topics in the taxonomy, output those topics and stop here.
2. If the sentence contains no topic, return "None".
3. Otherwise, add your topic as a top-subtopic. Stop here and output the added topic(s). DO NOT add any additional levels.

[Sentence]

{Sentence}

Please ONLY return the relevant or modified top-subtopics in JSON format.

[Your response]

Table A13 Prompt Template for Refining and Merging Topics

Prompt for topic refining and merging

You will receive example lists of topics that belong to the same level of a topic taxonomy. Your task is to merge topics that are paraphrases or near duplicates of one another. Return "None" if no modification is needed.

[Examples]

Example 1

Topic List:

[08] Cash Flow and Liquidity

[0805] Short-Term Funding

[08] Cash Flow and Liquidity

[0807] Commercial Papers

[08] Cash Flow and Liquidity

[0812] Internal Financing

Your response:

[04] Cash Flow and Liquidity

[0405] Short-Term Funding

Example 2

[18] Mergers and Acquisitions

[1808] Market Expansion Acquisitions

[18] Mergers and Acquisitions

[1810] Talent Acquisitions

[18] Mergers and Acquisitions

[1817] Cross-Border Acquisition

Your response:

[18] Mergers and Acquisitions

[1808] Acquisition Strategy

[Rules]

- Each line represents a topic, with a level indicator, a top-level label and sub-level label.
- Perform the following operations as many times as needed:
- Merge relevant topics into a single topic.
- Do nothing and return "None" if no modification is needed.
- When merging, the output format should contain a level indicator, the updated top- and sub-level label, followed by the original topics.

[Topic List]

{Topics}

Output the modification or "None" where appropriate. Do not output anything else.

[Your response]

Table A14 Prompt Template for Topic Assignment and Sentiment Analysis

Prompt for topic assignment and sentiment analysis

You will receive a sentence and a topic taxonomy. Assign the sentence to the most relevant top-subtopics in the taxonomy. Please also evaluate the sentiment conveyed in the given sentence (-1 denote the most negative sentiment and 1 the most positive). Then, output the top-subtopic labels and the sentiment score for the sentence. DO NOT make up new topics.

Here is the topic taxonomy:

{tree}

[Examples]

Example 1

Sentence:

The prepaid business segment is experiencing rapid revenue growth, already accounting for 15% of total hotel booking revenue, and is expected to further expand its share and serve as a key driver of overall revenue growth.

Your response:

[24] Profit and Loss

[2402] Earning Outlook

Sentiment Score: 0.15

Example 2

Sentence:

Given a more rational market and the rise of personalized consumption, we have strengthened our brand influence and enhanced market recognition through leveraging brand effects and optimizing our product portfolio.

Your response:

[03] Branding and Reputation

[0301] Branding Strategy

Sentiment Score: 0.2

[Instructions]

1. Topic labels must be presented in the provided topic hierarchy. You MUST NOT make up new topics.
2. If the assigned topic is not on the top level, you must also output the path from the top-subtopic to the assigned topic.
3. If the assigned topic does not have an appropriate subtopic, assign the sentence to the subtopic "Others".

[Sentence]

{Sentence}

[Your response]
