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

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

March 2026

Abstract

We employ a novel dataset comprising public and private earnings call transcripts for listed firms in China. We construct a comprehensive topic taxonomy to dissect disclosure inconsistencies by firm managers across public and private meetings. Our analysis reveals 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 disclosure inconsistencies are associated with meeting-, manager- and firm-level characteristics. We further examine the informativeness of disclosure inconsistencies and show that these measures are predictive of firms' future stock returns and illiquidity. Our findings highlight the managers' inconsistent disclosure behaviors across meetings and their adverse implications for a level playing field in the capital market.

Keywords: Disclosure Inconsistencies; Earnings Calls; Private Meeting; Large Language Model; 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, China. Email: srjiang@hku.hk. Chen Lin is from the Faculty of Business and Economics, The University of Hong Kong, Hong Kong, China. Email: chenlin1@hku.hk. Yutong Sun is from the State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China. Email: sunyutong@amss.ac.cn. Mingzhu Tai is from the Faculty of Business and Economics, The University of Hong Kong, Hong Kong, China. taimzh@hku.hk.

1. Introduction

Do firm managers disclose inconsistently across public and private meetings? A large volume of literature shows that investors may benefit from private communications with managers. In comparison to investors who only attend public meetings, investors who gain access to private meetings are subsequently more likely to make informed trading decisions (Solomon & Soltes, 2015; Bradley et al., 2022; Cao et al., 2025), and better interpret firms' public disclosure (Green et al., 2014; Li et al., 2023). These findings highlight the potential value of the information collected by investors during private meetings, yet we still do not understand what information is disclosed during private meetings, how this information differs from the sources of public information, and whether these differences are informative of firms' future stock returns.[1]

In this paper, we address these questions by dissecting disclosure inconsistencies by firm managers across public and private meetings. We take three steps to derive the measures of disclosure inconsistencies. First, we exposit a novel dataset that comprises 115,357 conference call transcripts for listed companies in China from the Comein Finance dataset.[2] Our transcript dataset consists of two types of earnings calls. The first type is public earnings calls held by the company in conjunction with an earnings release, where most investors can participate in the corporate 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. One can only attend the private calls once the conference registrations are reviewed and approved by the host. We identify and pair public and private calls delivered by the same firm managers regarding the identical earnings release as the unit of observation.[3] Our final dataset consists of 20,480 pairs of public and private call transcripts for 976 publicly listed companies in China from the time period of 2020 to 2025.[4]

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. Its topics and assignments

[1] The actual contents of private meeting narratives remain largely elusive in the existing literature. Without access to this information, one can only speculate about the nature and consequence of these events (Park & Soltes, 2018).

[2] Comine Finance is the leading call transcript provider that streams real-time corporate conference events and converts audio to text records in China. (https://www.comein.cn/home/index)

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

[4] Due to the outbreak of COVID-19 in China, many private meetings that were previously held off-line had to be conducted virtually. After the pandemic, virtual and hybrid (online and offline) meeting formats have continued to be widely used by firms and investors, given the convenience and cost-efficiency of online private meetings. As a result, our sample period begins in January 2020, following the COVID-19 outbreak in China.

align significantly more closely with human-annotated ground truth topics than those from traditional topic modelling approaches (Pham et al., 2023; Kim et al., 2024). We use TopicGPT to obtain exhaustive top topic and subtopic lists for the presentation section of earnings calls. Then, we create a labeled training set by instructing GPT to classify sentences into top- subtopics for a sample of call transcripts. Finally, we train a multilayer perceptron (MLP) on the labeled sample and extend it to the full sample. Our topic taxonomy finally consists of 31 top topics and the associated 150 subtopics.

Third, after obtaining the top-subtopic lists for all the call transcripts, we construct disclosure inconsistency measures for each public-private call pair in three dimensions: the topic, textual and vocal inconsistencies. For topic inconsistencies, we calculate how many new topics that discussed in the private call are not disclosed in the public call (*#NewTopic*), as well as the number of public-discussed topics that omitted in the private call (*#OmitTopic*). For the topics shared by public and private calls, we first develop the measures of textual inconsistencies, including differences in numerical usage (Δ *Number*), disclosure narratives (Δ *Narrative*) and text-based sentiment (Δ*Sentiment*). Utilizing the audio recordings of earnings calls, we further compute the measures of vocal inconsistencies, including the differences in audio length (Δ *Length*), listenability (Δ*Listenability*), vocal valence (Δ*Valence*) and arousal (Δ*Arousal*). To validate the performances of TopicGPT modelling, we evaluate the consistency between human auditors and TopicGPT across alternative prompting strategies and samples. The results show that TopicGPT generates and assigns the similar topics to earnings call transcripts relative to human auditors.

How do managers' disclosure behaviors change 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, suggesting that managers in our sample firms disclose a non-trivial proportion of new topics in private calls, and omit a number of topics discussed in public calls. The largest values of *#NewTopic* are observed in topics such as *Corporate Governance*, *Operation Management*, *Research and Development*. Their frequent appearance as new topics in private calls suggests that managers may potentially reserve material information for selective disclosure to private audiences.[5] However, the largest values of *#OmitTopic* are concentrated in topics such as *Political Alignment*, *Social Impact*, and *Media Engagement*, suggesting

[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.

that managers are more inclined to exclude these non-material discussions from formal public communication. This asymmetric topic structure highlights the critical role of private meetings in conveying 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 most pronounced in topics conveying soft information such as *Business Strategy* and *Mergers and Acquisitions*, etc. The largest sentiment gaps are identified in risk-sensitive and performance-related topics such as *Profit and Loss*, *Risk Exposure*, *Compliance and Enforcement*, etc. We then identify whether the sentiment inconsistencies are primarily driven by managers being more positive in private (private-driven) or in public (public-driven) settings. We find that most topics are strongly public-driven in tone, while only a smaller set of topics exhibits private-driven sentiment. These findings highlight firms' tone management across different meetings: adopting an optimistic tone in public calls while conveying more measured tones in 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. This divergence between how much is said and how clearly it is said suggests a disclosure strategy: managers appear to increase vocal output in public settings while reducing informational clarity, thereby creating the impression of extensive disclosure without increasing transparency. We also find that vocal valence and arousal are significantly larger than their text-based sentiment counterparts, suggesting that vocal signals may capture layers of managerial emotion and tension that are not fully reflected in prepared textual content. Given the fact that textual sentiment can be carefully scripted and reviewed ex ante, vocal expressions are harder to fully control, making voice a more sensitive channel for detecting latent disclosure frictions and informative information.

To gain a better understanding of how the disclosure inconsistencies arise, we explore determinants of the disclosure inconsistencies identified between public and private calls. We expect that the disclosure inconsistencies measures have different level determinants, including 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-host by financial institutions that are non-state-owned and affiliated with the

listed firms. Managers who are male and in their elder age and with longer experience in their current positions appear to generate higher levels of disclosure inconsistencies. In contrast, managers with accounting or law related background are more sensitive to the regulations of information disclosure and insider trading, and thus less likely to provide superior information during private meetings. The estimation results further show that a better financial position of the firm is associated with fewer disclosure inconsistencies between public and private meetings.

Do these disclosure inconsistencies grant private call attendees advantages in assessing firm value and executing stock trades? We further examine the informativeness of the disclosure inconsistency measures. We first examine the impacts of these measures on the stock returns post private meeting dates. The results indicate that the majority of disclosure inconsistency measures possess predictive power for future cumulative stock returns. For example, one-standard-deviation increases in $\Delta Narrative$, $\Delta Listenability$ and $\Delta Arousal$ increase the one-month market adjusted return by 0.198%, 0.215% and 0.389%, respectively. Among all the measures, we find that the proxies of new topic (*%NewTopic*) convey more informative signals (0.610%) to the market than their counterparts. We then examine the impact of disclosure inconsistencies on the information asymmetry among investors post the private meeting date, where we use the bid-ask spread (*BA spread*) as the proxies of information asymmetry. We find that the increases in disclosure inconsistencies between public and private earnings calls are associated with higher levels of average bid-ask spread and stock illiquidity, likely due to the adverse selection concerns across investors.

We contribute to the literature in the following aspects. First, we contribute to the literature investigating the private interactions between firms and investors (Bowen et al., 2018; Zhang et al., 2023). A stream of prior literature shows the benefits that private meetings offer to investors. Investors who attend private meetings make more informed trades (Bradley et al., 2022; Cao et al., 2025), have better interpretations of firms' public disclosure (Li et al., 2023), and actively express governance concerns (Ellis et al., 2021). Recent studies begin to explore the nature of narratives in private meetings. Ru et al. (2025) show that timely disclosure mandates of private meetings draw more immediate analyst attention to peer firms that were not directly visited. Using meeting notes of 4,700 private meetings between firms and investors, Becht et al. (2025) document that private interactions primarily convey soft information through qualitative discussions rather than explicit disclosure of hard information. They further show that private meetings improve investors'

information processing abilities, allowing them to better interpret firm fundamentals. However, given that the actual contents of private interactions are largely unavailable, we still do not understand how the private information differs from sources of public information. In this paper, we exposit a novel dataset that comprises verbatim records of public and private earnings calls delivered by the same firm managers. We directly identify the disclosure inconsistencies conveyed between the public-private call pairs. We show that these inconsistencies can be converted into capital gain in the market.

Second, our paper is related to works exploring firms' strategic disclosure behaviors (Jung et al., 2018; Bushee et al., 2018; Liu et al., 2024). Huang et al. (2014) find that firms may strategically manage their tones of the qualitative text in earnings press releases, which is incommensurate with the concurrent quantitative information. Brochet et al. (2015) show that the short-term horizon of firm disclosure is associated with managers' myopic behavior. Fabrizio and Kim (2019) find that firm managers use linguistic obfuscation in disclosure to manage the tension between the pressure for complete disclosures and the desire to project a positive image. We provide new evidence on the corporate strategic disclosure behaviors: firms can use different meeting occasions to influence investor perceptions of firms' fundamentals.

Furthermore, we add to the burgeoning literature on the applications of large language models in financial markets (Eisfeldt et al., 2023; Fan et al., 2024). The recent studies assess whether LLMs can effectively extract information from unstructured text about complex concepts such as future corporate policies (Jha et al., 2024), operations risk (Kim et al., 2023) and corporate culture (Li et al., 2025). In this paper, we examine whether LLMs can help to identify the disclosure inconsistencies of firm managers across meeting occasions. We show that LLMs assess the information contents of input texts more like a sophisticated human reader rather than relying solely on simple heuristics and correlations.

2. Measuring Disclosure Inconsistencies

2.1 Data on Public and Private Earnings Calls

We obtain 115,357 conference call transcripts from the Comin Finance database for publicly listed companies in Chinese stock markets during the period of January 2020 to January 2025. Comein Finance is the leading call transcript provider that streams real-time corporate conference events and converts audio to text records.[6] We complement the transcript dataset by deriving

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

conference transcript data from the 3C China corporate conference platform in the Wind database.[7] Our transcript dataset consists of two types of corporate conference calls. The first type is public earnings calls held by the company itself in conjunction with an earnings release, where most investors can participate in the corporate earnings calls with an 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, one can only attend the private calls once the conference registrations are reviewed and approved by the host.

By observing whether the call is publicly accessible or not, we identify 60,829 public calls and 54,528 private calls from our call transcript dataset. In Figure A1 of the Appendix A1, we provide the examples of public-private call pairs discussing the same earnings release in our sample.[8] To construct valid measures of disclosure inconsistencies, we pair the public and private call transcripts being held by firms regarding the identical earnings release as the unit of observation. Each transcript file 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 call transcript since the Q&A session may contain information from call attendees. We further restrict our sample to the pairs of calls that are delivered by the same firm managers to control the individual bias of information disclosure. The filtering procedures are illustrated in Table A2 of the Appendix A1. After filtering, we derive 20,480 pairs of public and private call transcripts across 976 publicly listed companies in China.

2.2 Topic Modelling for Earnings Calls

In this section, we construct a comprehensive topic taxonomy to capture the full scope of corporate disclosure, which enables us to dissect disclosure inconsistencies between public and private calls at the topic level. We leverage TopicGPT modelling to perform in-context topic generation and assignment (Pham et al., 2023). GPT is a human-centric approach to topic modeling that employs an attention mechanism and multidimensional logic to assess the narrative content of input texts, as opposed to relying solely on simple heuristics and correlations (de Kok, 2025). As a result, it demonstrates zero-shot and few-shot capabilities to address problems not explicitly covered

[7] See the official website of 3C conference platform can be found at https://peacall.com/#/

[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.

during the training process. The existing papers show that TopicGPT's topic generations and assignments align significantly more closely with human-annotated ground truth topics than those from LDA or other topic modelling methods (Pham et al., 2023; Li et al., 2024).

Specifically, our TopicGPT modelling framework consists of two steps: First, we use GPT-4o to obtain exhaustive topics and subtopic lists for the presentation section of the earnings call transcript. Second, we create a labeled training set by prompting GPT-4o to classify sentences into topics and subtopics for a sample of earnings call transcripts. We then train a multilayer perceptron (MLP) on the labeled sample and extend it to the full sample.

2.2.1 Topic Generation

As a first step, we prompt GPT-4o to generate a set of topics given input examples (The prompts for topic generation are shown in Table A8 of the Appendix A2). Specifically, given a sentence *S* from the earning call dataset and a set of example topics *K*, the model is instructed to either assign *S* to an existing topic in *K* or generate a new topic that better describes *S* and append it to *K*. Initially, *K* consists of a few human-written example topics (three in our prompts). For each example, we provide the model with a topic branch comprising a top-level topic K_t, a subtopic *Ks*, 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 distinctive and match the specificity seen in other topics. With these inputs, we instruct the LLM to generate top-subtopic lists that capture common themes among the provided sentences.

Instead of running topic generation over the entire dataset, which can be computationally costly, we apply the process to a carefully constructed subsample. We follow Kim et al. (2024) and collect a sample of firms based on the China Securities Regulatory Commission (CRSC) industry codes and 2 time windows into 30 strata (15 industries × 2 periods). For each of the time-industry stratum, we randomly sample five earnings calls and input the corresponding transcripts into GPT-4o. The model is then instructed to generate a list of top-subtopics that are encountered within the five input call transcripts at the sentence level. As a result, the subsample of 150 call transcripts (30 strata × 5 documents) serves as our basis for generating topic lists.

We then consolidate and reconcile top- and sub-level topics across the strata. Specifically, GPT-4o first aggregates top-subtopic lists within each industry. We use Sentence-Transformer embeddings to identify pairs of topics with high similarity (cosine similarity ≥0.5). 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 A9 of the Appendix A2). We then aggregate the refined top-subtopic lists across industries to construct a comprehensive topic taxonomy based on the 150 sampled earnings call transcripts. 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 (the initial 150 documents and an additional stratified random sample of 150 documents). 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 the sentence to "Others" if no existing subtopic fits. To capture the narrative sentiment conveyed in firms' presentations, we instruct GPT-4o to label 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. "None"/"Error"). 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 topic and sub-topic label (with digital code), and a sentiment score produced by GPT-4o (The prompts for topic assignment are shown in Table A10 of the 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 transcripts. Instead, in the third step, 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. 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 call transcript. 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*, based on how many new topics that discussed in private call are not

disclosed 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 private call scaled by the total number of sentences in private call:

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

Where $s = 0,1 \cdots S_{i,m,t}$ are sentences in private earnings call of firm i in quarter t. K_m and K_n present the sets of topics discussed in public and private earnings 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 develop the textual inconsistencies measures. We first compute a numerical inconsistency measure *ΔNumber*, defined as the absolute differences of the numerical usage (including digital numbers, percentages, dates and monetary figures, normalized by total text length) for the same topic between public and private calls.[9] Next, we derive the top-level disclosure narratives by averaging the embedding vectors of sentences within each topic. We use these narratives vectors in public call $\boldsymbol{V}_{i,k,m,t}$ and private call $\boldsymbol{V}_{i,k,n,t}$ to calculate narrative inconsistencies *ΔNarrative*:

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

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

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

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

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

Where $SentimentInc_{i,t}$ denote the sentiment inconsistency for firm i in quarter t. $Sentiment_{i,k,m,t}$ and $Sentiment_{i,k,n,t}$ are the topic-level sentiment scores of topic k in public and private earnings call, respectively. We further analyze the directional component of sentiment inconsistencies in Section 4.2, i.e., whether the sentiment differences are primarily driven by managers being more positive in private (private-driven) or in public (public-driven) settings.

2.3.3 Vocal Inconsistencies

In addition to the text-based measures, we analyze the inconsistencies embedded in the voice of the firm manager 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 then 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 the sentence-level audio recordings and the associated timestamped transcripts at topic-level to construct measures of vocal inconsistencies. We first calculate the differences in audio length (*ΔLength*) in-between public and private calls for the same topic. We then follow Call et al. (2023) and construct the listenability of managers' voices during earnings calls. It is based on the *Whisper* model's average word-level confidence scores when transcribing managers' 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 managers' vocal disclosures. The 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 managers' presentation for firm i in quarter t $Conf(w)$ denotes 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, more decisive, and more intelligible vocal delivery, whereas a lower value indicates ambiguous or harder-to-interpret speech. To capture the difference in vocal listenability between public and private settings, we calculate

ΔListenability as the absolute difference in listenability between the public and private calls for the same topic.

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 directions of the vocal feature (positive versus negative affect), while arousal captures the emotional intensity of vocal attitude (exciting versus calm).[10] As a result, we derive valence inconsistencies *ΔValence* and arousal inconsistencies *ΔArousal* in-between public and private calls using the similar formula in Equation (3).

2.4 Validations

In this section, we examine the robustness of our identification framework to changes in prompt design and sampling strategies used for topic generation and assignment. Specifically, we make three modifications to the settings of our baseline 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 be affected by look-ahead bias (Sarkar and Vafa, 2024). The models embed knowledge up to their training cutoff and may inadvertently incorporate future information.[11] 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 call transcripts falling outside of the embedding model's training sample (i.e., subsample from the time period of 2022 to 2025). We examine whether TopicGPT generate the similar outcomes on this subsample (*Subsample*). Third, we randomize the order of documents used in the topic generation and assignment stages to test whether the sequence in which TopicGPT processes inputs affects the resulting topic classifications (*Sample shuffling*).

The evaluation builds on the premise that no algorithm understands the context of a human conversation better than human beings do. We employ five Chinese finance-major doctoral students

[10] 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.

[11] 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.

and randomly choose 200 pairs of public and private earnings calls for human assessment. Each individual independently reads, listens and assesses the disclosure inconsistencies between public-private call pairs instructed by the similar guideline in the pipeline of TopicGPT modelling framework. We implement the Turing Test that examines whether we can distinguish the machine's answers from the human answers (Lagakos et al., 2025). We use the human consensus score as a benchmark, which is the average disclosure inconsistencies of the five human readers. We compute the accuracy of each human relative to the human consensus score and all the introduced GPT-based methods. If the accuracy of the TopicGPT-generated scores falls within the range of accuracies achieved by human readers, we conclude that TopicGPT passes the Turing test. Figure A2 reports the Turing test results for all disclosure inconsistency measures, where the horizontal lines indicate the range of human reader accuracies.

The statistics in Figure A2 of the Appendix A1 indicate our baseline modelling yields the best performance with human coding across all the measures of disclosure inconsistencies. For example, when assessing the new topic discussed in a private call (*#NewTopic*), human readers obtain the accuracy from a minimum of 82.23% to a maximum of 97.53%, and an average of 89.05%. 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 helpful. Accuracy rises to 91.58% with the benchmark modelling. We find similar patterns when substituting the proxies into text and vocal inconsistencies. The statistics in Figure A2 also show that our identification framework passes Turing tests on the subsample from the period of 2022 to 2025. In sum, the results indicate that our identification framework is robust to alternative prompt settings and subsamples out of the LLM's training period.

3. Dissecting Disclosure Inconsistencies

3.1 Topic Taxonomy and Summary Statistics

In Table A3 of the Appendix A1, we present the topic taxonomy generated by the TopicGPT framework, which consists of 31 top topics and the associated 150 subtopics and broadly covers the range of themes disclosed by firms. The taxonomy includes material information topics such as *Accounting Policies and Standards*, *Business Strategy*, *Mergers and Acquisitions*, and *Profit and Loss*, as well as non-material information topics including *Branding and Reputation*, *Employee Welfare*, and *Corporate Culture*, etc.

In Table 1, we provide summary statistics for the disclosure inconsistencies measures and other variables of interest. The statistics in Table 1 show that the measures of *#NewTopic* and *#OmitTopic* have a mean of 1.238 and 1.813 respectively, suggesting that managers in our sample firms disclose a non-trivial proportion of new topics while omitting a number of topics discussed in public calls. The statistics of textual and vocal inconsistency measures further indicate that firm managers demonstrate distinct narrative styles and vocal delivery strategies when discussing the same topic at different meeting occasions.

[Insert Table 1 Here]

In Table A4, 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 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 Topic Inconsistencies

In this section, we decompose the three measures of topic inconsistencies 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.

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 values of both *#NewTopic* and *%NewTopic* are observed in topics such as *Corporate Governance*, *Operation Management*, *Research and Development*. Their frequent appearance as new topics in private calls suggests that managers systematically may potentially reserve material information for selective disclosure to private audiences. The disclosure behaviors may constitute a violation of fair disclosure regulations and grant private call attendees advantages in assessing firm value and executing stock trades.[12]

[12] In contrast to many developed capital markets where selective disclosure is subject to strict regulatory constraints, 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.

[Insert Figure 1 Here]

Panel (b) in Figure 1 reveals a different structure for omitted topics. The highest values of *#OmitTopic* are concentrated in *Political Alignment*, *Social Impact*, and *Media Engagement*, etc. Their omission from private calls suggests that managers are more inclined to exclude non-material discussions from formal public communication. We also provide the top-and bottom-ranked subtopics in term of topic inconsistencies in Table A4 of the Appendix A1. Taken together, Figure 1 and Table A5 document a clear reallocation of discussion topics across public and private calls. Private calls are primarily used to introduce material and forward-looking information and exclude non-material and non-financial topics. 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.

3.3 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 topics. Figure 3 further illustrates the directions of the numerical and sentiment inconsistencies between public and private calls, where a topic is classified as public-call driven if the disclosure measure in the public call exceeds its private-call counterpart, and as private-call driven otherwise.

[Insert Figure 2 Here]

Panel (a) of Figure 2 shows that both numerical (*ΔNumber*) and narrative (*ΔNarrative*) inconsistencies are most pronounced in topics conveying soft information such as *Business Strategy*, *Mergers and Acquisitions*, *Operation Management*, etc. Although their underlying information is constrained by accounting and regulatory standards, managers retain considerable flexibility in how numbers are highlighted and how narratives are constructed between public and private calls.[13] In panel (b), the sentiment inconsistencies display a distinct structure. The largest sentiment gaps are

[13] We provide example transcripts of textual inconsistencies between public and private earnings calls in Table A8 of the Appendix A1.

identified in topics such as *Profit and Loss*, *Risk Exposure*, *Compliance and Enforcement*, etc. These topics are inherently performance- and risk-related, and therefore particularly prone to tone management. Managers appear to adjust sentiment when discussing firm performance and operations uncertainty in response to the audience and disclosure context.

[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. By contrast, numerical discussion of *Business Strategy*, *Operation Management*, and *Debt and Capital Structure* is largely private-driven, suggesting a more concrete manner during private meetings. A similar asymmetry appears in textual sentiment. 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.

Overall, Figures 2 and 3 demonstrate numerical and narrative inconsistencies concentrate in topics containing soft information with managerial discretion, whereas sentiment differences arise primarily in risk-sensitive and performance-related topics. We provide the top-and bottom-ranked subtopics in term of textual inconsistencies in Table A6 of the 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 Decomposition of Vocal Inconsistencies

In this section, we decompose the vocal dimension of disclosure inconsistencies. Figure 4 reports the magnitudes of inconsistencies in audio length, listenability, vocal valence, and arousal across topics, while Figure 5 further illustrates the directions of these inconsistencies between public and private calls. We also provide the top-and bottom-ranked subtopics in term of vocal inconsistencies in Table A7 of the Appendix A1

[Insert Figure 4 Here]

The panel (a) and (b) of Figure 4 and Figure 5 reveal an informative pattern in the relationship between speech quantity and speech 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. This divergence between how much is said and how clearly it is said suggests a disclosure strategy: managers appear to increase vocal output in public settings while reducing informational clarity, thereby creating the impression of extensive disclosure without increasing clarity. Such a quantity-clarity substitution strategy is consistent with strategic obfuscation incentives when discussing sensitive operational and risk-related topics under regulatory and market scrutiny.

The panel (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 are concentrated 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 that are not fully reflected in prepared textual content. A possible explanation is that while textual sentiment can be carefully scripted and reviewed ex ante, vocal expressions are harder to fully control, making voice a more sensitive channel for detecting latent disclosure frictions and informative information.

[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 intention, emotional regulation, and strategic obfuscation that remain unexplored in textual records.

4. The Determents of Disclosure Inconsistencies

4.1 Empirical Strategies

In this section, we explore determinants of the 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 2025 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) otherwise zero, and *Team* for the call-pairs both held by firm executive teams rather than individuals. For the private earnings call co-hosted by financial institutions, we identify whether the co-host financial institution is a state-owned financial institution (*SOFI*) or not. 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 of a seasoned equity offering during the past two (Malmendier and Shanthikumar, 2014).

Manager Characteristics. Given that the presentation sessions of both public and private calls are delivered by the same managers, we include a variety 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 two dummy variables *Finance degree* and *Law degree*, which take the value of one if the manager has a degree in finance or law related majors, otherwise zero.

Firm Characteristics. We introduce a set of firms' financial and operational proxies to examine their impacts on the measures of disclosure inconsistencies, 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's shares that are owned by

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

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*) in China.

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

$$DisclosureInc_{i,j,t} = \beta_1 + \beta_2 CallCharater_{i,j,t} + \beta_3 ManagerCharater_{i,j,t} + \beta_3 FirmCharater_{i,j,t} + \lambda_j + \mu_t + \varepsilon_{i,j,t} \quad (5)$$

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. $CallCharacter$ include dummy variables *Online*, *Team*, *SOFI*, and *Affiliation*. $ManagerCharacter$ presents gender, age and the position tenure of the manager. We also include *Finance degree* and *Law degree* to denote the managers' finance or law related backgrounds. $FirmCharater$ 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.

4.2 Determents of Topic Inconsistencies

Table 2 reports the estimation results of Equation (5). The dependent variables are the three topic inconsistency measures, including *#NewTopic*, *%NewTopic* and *#OmitTopic*. The results provide consistent evidence that topic inconsistencies are systematically associated with the characteristics relating to the earnings calls, managers and listed firms.

[Insert Table 2 Here]

In column (1), we find that the *Online* (-0.164, *t*-stat: -3.17) and *Team* (-0.117, *t*-stat: -2.51) are negatively correlated with *#Newtopic*, suggesting that firm executives are less likely to disclose new topics during private calls when the earnings calls are held online or by executive teams. The coefficients of *SOFI* and *Affiliation* show that the topics overlap between public and private calls is lower when private meetings are co-hosted by non-state-owned financial institutions or by institutions

affiliated with the listed firms. We find similar patterns when we substitute the dependent variables with *%NewTopic* and *#OmitTopic* in columns (2) and (3).

In terms of manager characteristics, the coefficients of *Gender* and *Age* are statistically significant across all measures of topic inconsistencies, suggesting that managers who are male and in their elder age are more likely to provide new information during private earnings calls. In addition, managers with longer experience in their current positions appear to generate higher levels of topic inconsistencies. We find that *Finance degree* and *Law degree* are significantly negatively associated with all topic inconsistency variables. It is possible that managers with finance or law related background 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 fewer topic inconsistencies, as evidenced by the coefficients of *Cash*, *Leverage* and *ROA* in Table 2. Furthermore, we find that firm executives are less likely to discuss new topics during private calls if the listed firms have longer years of business or are state-owned enterprises.

4.3 Determents of Textual Inconsistencies

We examine the determinants of textual inconsistencies in Table 3, including differences in numerical usage (*ΔNumber*), narrative content (*ΔNarrative*), and text sentiment (*ΔSentiment*) between public and private earnings calls. At the earnings-call level, both *Online* and *Team* are negatively associated with all three textual measures, indicating that online communication and team meetings constrain managers' ability to adjust wording and sentiment across earnings calls. In contrast, *Affiliation* is positively associated with these measures, suggesting that closer relationships between firms and financial institutions facilitate more pronounced textual differentiation.

[Insert Table 3 Here]

In terms of manager characteristics, we find that the coefficients of *Age* and *Tenure* are consistently positive across all measures of textual inconsistencies. For example, a one-year increase in a manager's age is associated with a 0.026 (*t*-stat: 2.79) increase in *ΔNumber* and a 0.094 (*t*-stat: 2.83) increase in *ΔNarrative*. We find that executives with finance or law backgrounds exhibit significantly lower levels of textual inconsistencies. In addition, a set of firm-level characteristics is

significantly associated with textual inconsistencies: Firms with higher return on assets (*ROA*) and larger size (*Asset*) exhibit significantly lower textual inconsistencies, whereas more leveraged firms show greater textual inconsistencies.

4.3 Determents of Vocal Inconsistencies

Table 4 reports the determinants of vocal inconsistencies, including audio length (*ΔLength*), listenability (*ΔListenability*), vocal valence (*ΔValence*), and arousal (*ΔArousal*) between public and private earnings calls. The coefficients in Table 4 show online and team meetings are negatively associated with the measures of vocal inconsistencies, while *Affiliation* positively predicts vocal inconsistencies, indicating that closer ties between firms and financial institutions also facilitate vocal delivery divergences in-between public and private calls.

[Insert Table 4 Here]

Manager attributes also exhibit pronounced effects on vocal inconsistencies. For example, *Tenure* significantly raises *ΔListenability* (0.114, *t*-stat: 2.73) and *ΔValence* (0.019, *t*-stat: 2.80), indicating that experienced managers adjust not only what they say but how they sound across meetings. In contrast, the coefficients of *Tenure* and *Age* on *ΔArousal* are significantly negative, suggesting that more experienced managers exhibit smaller variations in emotional intensity and tend to adopt a more even-tempered vocal delivery across meetings. Managers with finance or law backgrounds consistently display lower vocal inconsistencies across all four measures. Firm fundamentals reinforce these patterns. More leveraged firms exhibit significantly greater vocal inconsistencies across all dimensions, while firms with higher return on assets and larger size show lower vocal divergence. Overall, these results highlight that vocal inconsistencies are shaped by a combination of communication environments, managerial experience, and financial incentives.

5. The Informativeness of Disclosure Inconsistencies

In this section, we examine the informativeness of the disclosure inconsistencies measures. Specifically, we examine the impacts of disclosure inconsistencies on the stock returns and information asymmetry post private meeting dates.

5.1 Stork Returns

We first provide evidence on whether disclosure inconsistencies between public and private calls allow investors to execute profitable trades. To capture the dynamic return response to private meetings, we estimate a series of regressions using cumulative abnormal returns (CAR) over multiple post-meeting windows. Specifically, the dependent variable is the CAR over the time window [0,1] following the private meeting, and we subsequently extend the event window forward in time up to [29,30] to examine the evolution of market reactions. If disclosure inconsistencies enable investors to buy (sell) prior to positive (negative) information being revealed in price, then future trading gains will be associated with the timing of private calls. The regression specification is as follows:

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

Where $CAR_{i,j,[t+n,t+m]}$ is the cumulative market-adjusted return for firm i in industry j over the post-meeting window $[t+n, t+m]$. We begin with the time window [0, 1] and incrementally expand the horizon in a sequence of regressions until [29, 30]. $DisclosureInc_{i,j,t}$ are the disclosure inconsistencies measures between public-private earnings calls, which consists of topic, textual, and vocal inconsistencies. We standardize the all the measures of disclosure inconsistencies 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 asset (*ROA*), institutional ownership (*IO*), firm age (*Age*), total asset (*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 Figure 6 Here]

Figure 6 presents the regressions of cumulative market-adjusted return on measures of disclosure inconsistencies from [0,1] to [29,30] post-meeting horizons. The coefficients in Figure 6 indicate a clear trend of increasing returns following the private earnings call in cross-sectional regressions over a 30-day trading.[15] In terms of the economic significance, we find that a number of disclosure inconsistency measures demonstrate significant impacts on the cumulative market-adjusted return.

[15] Notice that the coefficients of our disclosure inconsistency measures are predominantly positive. One possible explanation is the absence of a short-selling mechanism in China's capital market. As a result, private information revealed in such meetings is more likely to be incorporated into prices through buying pressure rather than short selling.

For example, one-standard-deviation increases in Δ*Narrative*, Δ*Listenability* and Δ*Arousal* increase the market adjusted return by 0.198% (*t*-stat: 3.16), 0.215% (*t*-stat: 3.07) and 0.389% (*t*-stat: 3.30). Given that private meetings generally are held in relatively small scale and not as important of public information event, these coefficients likely represent a reasonable economic magnitude for potential trading gains.

[Insert Table 5 Here]

In Table 5, we present the effects of disclosure inconsistencies on CAR over a one-month horizon. Among all the measures of disclosure inconsistencies, we find that the proxies of new topics convey more informative signals (0.610, *t*-stat: 3.47) to the market than their counterparts, which may presents the profitability of illegal insider trading based on material non-public information. That said, the information content of new topics in private meetings can be directly leveraged by private meeting attendees to execute profitable insider trading.

We also observe that sentiment-based inconsistency measures, including *ΔSentiment*, *ΔValence* and *ΔArousal*, exhibit a more gradual information diffusion pattern. As shown in Figure 6, these affective measures are weakly informative immediately after the private meeting, but their informativeness increases steadily over time. These delayed effects likely reflect the fact that emotional cues are less directly linked to fundamental information and thus require longer processing by private meeting attendees, as investors gradually interpret and incorporate these affective signals into their valuation and trading decisions.

5.2 Information Asymmetry

Next, we examine the impacts of disclosure inconsistencies on the information asymmetry among investors post the private meeting date. We use bid-ask spread (*BA spread*) as the proxies of information asymmetry. Bid-ask spread is an estimator from daily high and low prices following Corwin and Schultz (2012). For each firm, we compute the bid-ask spread daily and run the similar regression specification in Equation (6) by substituting the dependent variables as *BA spread* over the post-meeting period from [0, 1] and [29, 30]. We multiply the coefficients by 100 times for easier interpretation.

[Insert Figure 7 Here]

We report the results in Figure 7. We find that the majority of disclosure inconsistencies exert positive impacts on bid-ask spread. For example, one-standard-deviation increases in *#Newtopic*, Δ*Narrative*, Δ*Listenability* lead to 10.374% (*t*-stat: 3.78), 3.294% (*t*-stat: 2.85) and 4.215% (*t*-stat: 2.28) increase in the bid-ask spread over one-month horizon, respectively. The evidence in Figure 7 shows that increases in disclosure inconsistencies between public and private earnings calls are associated with higher levels of average bid-ask spread and stock illiquidity, likely due to the increased risks associated with information asymmetry across investors.

[Insert Table 6 Here]

In Table 6, we report the effects of disclosure inconsistencies on bid-ask spread, where the coefficients indicate significant relationships between bid-ask spread and various disclosure inconsistency measures. Among all the measures, we observe that *%NewTopic* exerts the largest economic magnitude on one-month (10.521%, *t*-stat: 3.54) average bid-ask spread. Because the topic inconsistencies may capture the unlawful disclosure of material non-public information contained in private meetings, its effect on bid-ask spread reflects heightened information asymmetry that deters uninformed investors from trading due to adverse selection concerns. Overall, our results show that the disclosure inconsistencies we identified between public and private calls amplify information asymmetry and may contribute to an uneven playing field in the market.

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 transcripts for Chinese listed firms between 2020 and 2025, we use TopicGPT 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 material topics in private meetings and omit a number of non-material 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 accounting or law-related training are more compliant with disclosure regulations.

We further document that disclosure inconsistencies are informative for subsequent stock returns and illiquidity. Specifically, a number of our inconsistency measures are positively associated with cumulative abnormal returns and bid-ask spread in the post private meeting days. In particular, the measures of new topics discussed in private calls contain more incremental information than their counterparts, which allows investors to execute profitable trades. Together, these findings reveal that disclosure inconsistencies in private meetings not only shape investor behavior but also contribute to a more uneven informational environment in financial markets.

Our findings carry policy and managerial implications. For policymakers, our evidence raises questions about informational fairness and selective disclosure. While private meetings can facilitate investor engagement, they also risk amplifying advantages for investors with privileged access, undermining the principle of a level playing field. Enhancing oversight of private communications or providing clearer guidelines on permissible disclosure content may help reduce these disparities. For firms, our findings imply that inconsistent narratives across public and private settings may have unintended market consequences. Establishing more standardized internal disclosure protocols and aligning communication strategies across channels could help reduce information frictions and avoid potential regulatory or reputational risks. For investor, our results suggest that disclosure inconsistencies across communication channels are economically consequential and may affect price formation and market liquidity in the capital market.

Acknowledgement

The authors acknowledge financial support from the Research Grant Council of the Hong Kong SAR, China [Project T35-710/20-R]. The authors declare that they have no financial relationships or other conflicts of interest related to this research.

References

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. (2025). The benefits of access: Evidence from private meetings with portfolio firms. The Journal of Finance.

Bernard, D., Blankespoor, E., de Kok, T., & Toynbee, S. (2023). A modular measure of business complexity. *Available at SSRN 4480309*.

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.

Brochet, F., Loumioti, M., & Serafeim, G. (2015). Speaking of the short-term: Disclosure horizon and managerial myopia. *Review of Accounting Studies*, 20(3), 1122-1163.

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.S., Gong, G., Kim, Y., Shi, H., & Wang, A. (2025). Site visits and corporate investment efficiency. *Management science*, 71(1), 635-658.

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*.

Ellis, J., Gerken, W., & Jame, R. (2022). The meeting before the meeting: Governance roadshows and mutual fund voting. *Available at SSRN 3966579*.

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.

Fabrizio, K. R., & Kim, E. H. (2019). Reluctant disclosure and transparency: Evidence from environmental disclosures. *Organization Science*, *30*(6), 1207-1231.

Fan, J., Liu, Q., Song, Y., & Wang, Z. (2024). Measuring misinformation in financial markets. *Available at SSRN 4922648*.

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*.

Jung, M. J., Naughton, J. P., Tahoun, A., & Wang, C. (2018). Do firms strategically disseminate? Evidence from corporate use of social media. *The Accounting Review*, *93*(4), 225-252.

Kim, A., Muhn, M., & Nikolaev, V. (2023). From transcripts to insights: Uncovering corporate risks using generative ai. *arXiv preprint arXiv:2310.17721*.

Kim, A., Muhn, M., Nikolaev, V. V., & Zhang, Y. (2024). Learning Fundamentals from Text. Chicago Booth Accounting Research Center Research Paper, Fama-Miller Working Paper.

Lagakos, D., Michalopoulos, S., & Voth, H. J. (2025). American Life Histories (No. w33373). *National Bureau of Economic Research*.
Li, K., Mai, F., Shen, R., Yang, C., & Zhang, T. (2024). Dissecting corporate culture using generative AI–Insights from analyst reports. *Available at SSRN 4558295*.
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).
Liu, Y., Madsen, J., & Zhou, F. S. (2024). Strategic (Inconsistent) Disclosures and Sophisticated Investors: Evidence from Hedge Funds. Journal of Accounting Research.
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, July). Robust speech recognition via large-scale weak supervision. *In 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.
Zhang, R. X. (2023). Do managers learn from institutional investors through direct interactions?. *Journal of Accounting and Economics*, 75(2-3), 101554.

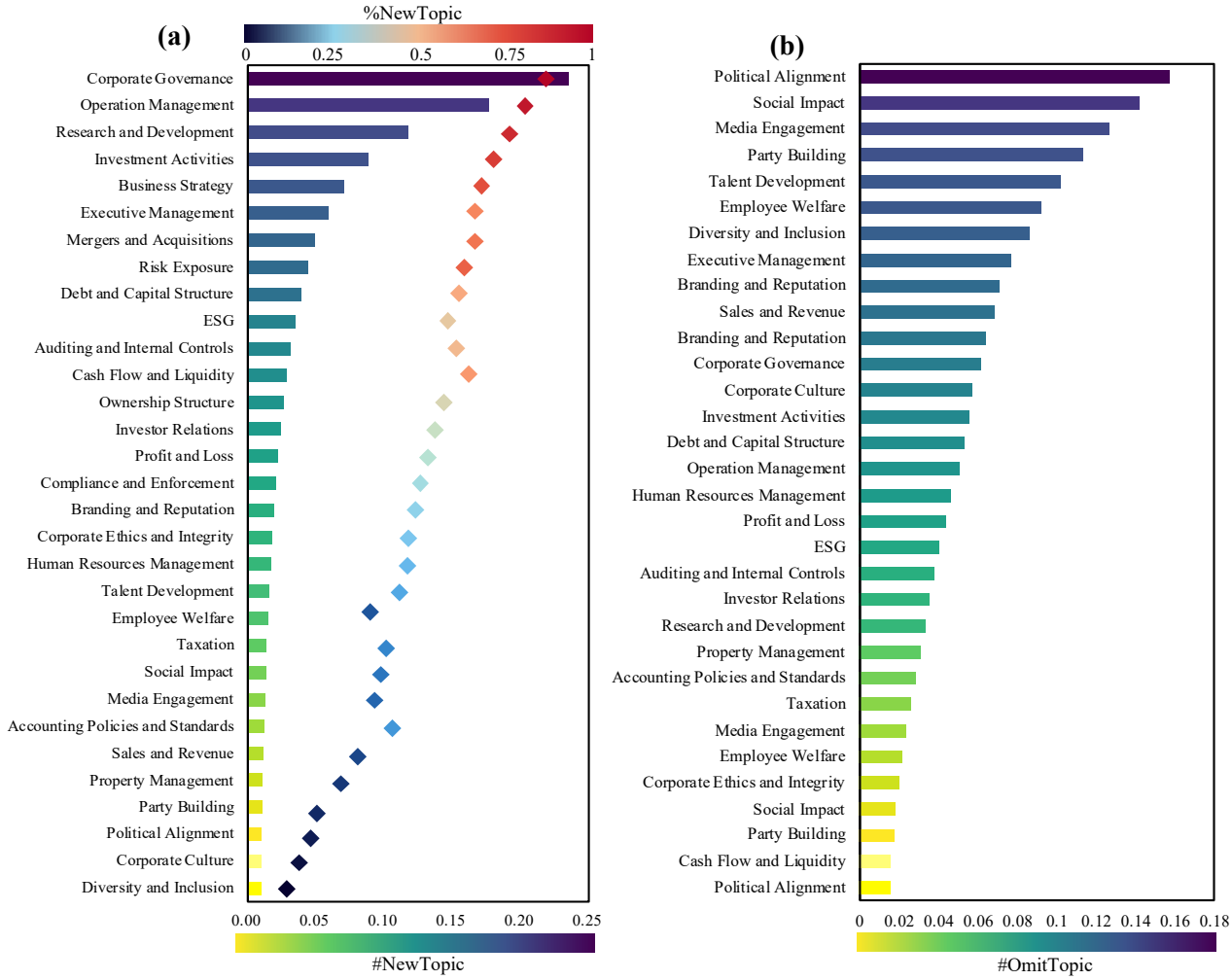


Figure 1. Decomposition of Topic Inconsistencies. This Figure provides the distributions of topic inconsistency measures across topics in our taxonomy. In Panel **(a)**, the bar chart reports the number of new topics discussed in private earnings calls (*#NewTopic*), while the dot plot shows the proportion of these new topics relative to all topics discussed in the corresponding private calls (*%NewTopic*). In Panel **(b)**, the bar chart reports the number of topics that are discussed in public calls but omitted in private calls (*#OmitTopic*).

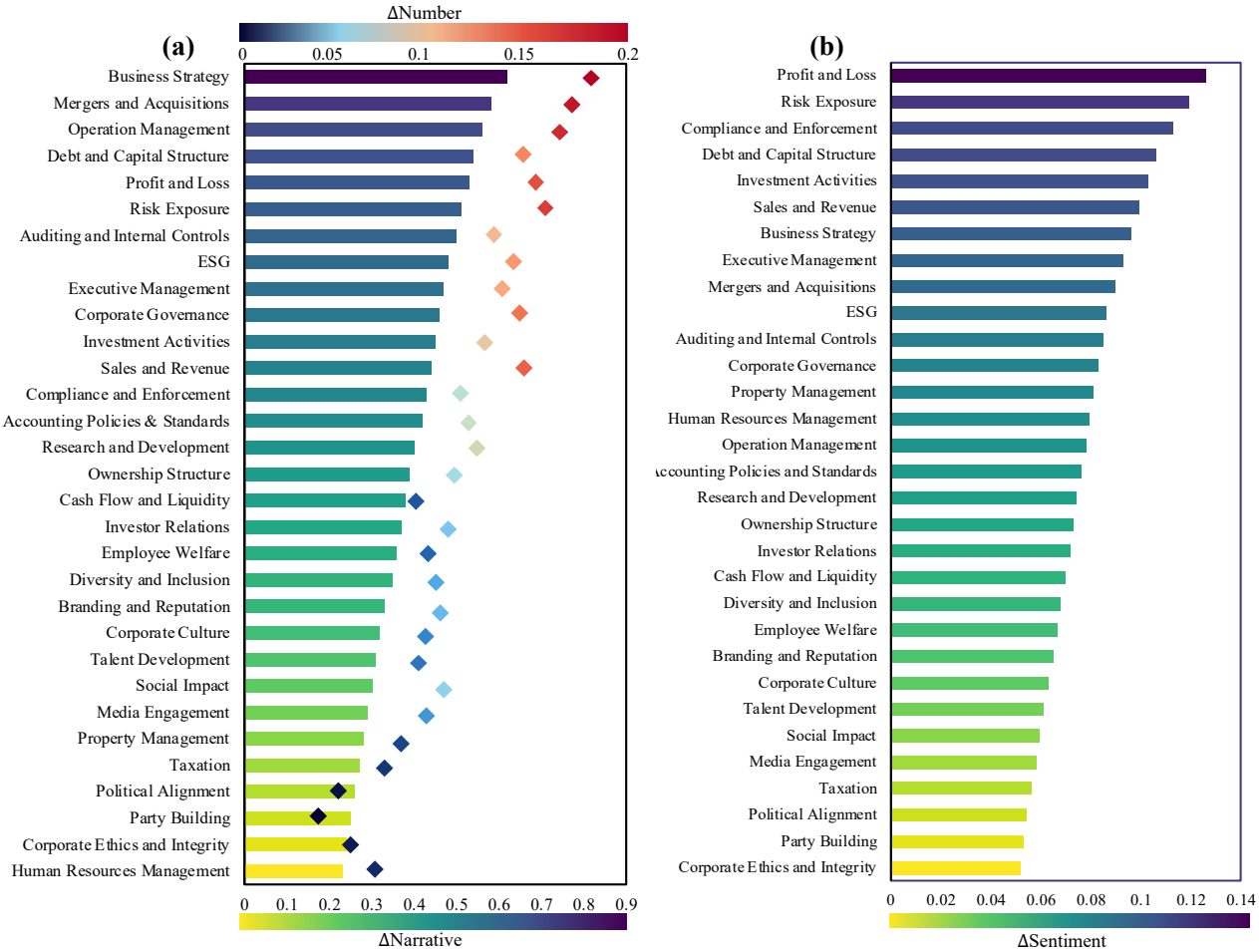


Figure 2. Decomposition of Textual Inconsistencies. This Figure provides the distributions of textual inconsistency measures across topics in our topic taxonomy. In Panel **(a)**, the bar chart reports the differences in the numerical usage for the respective topics disclosed between public and private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

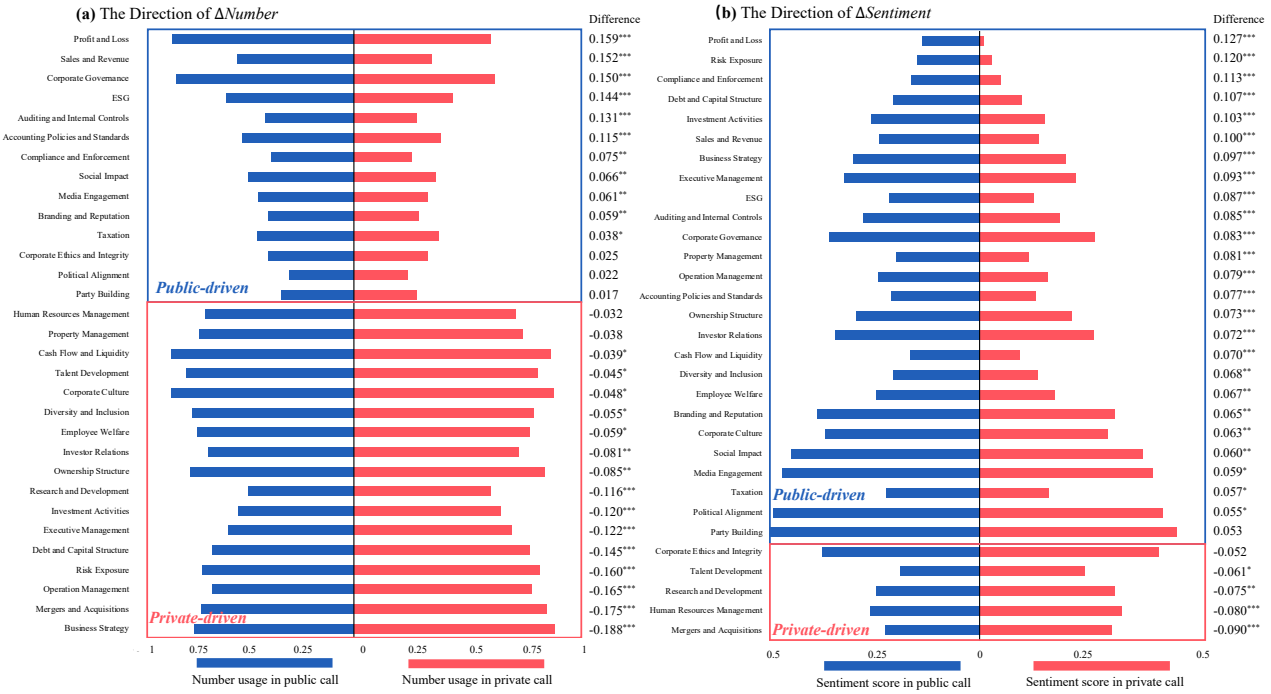


Figure 3. Direction of Textual Inconsistencies. Panels **(a)** and **(b)** illustrate the directional differences in numerical usage (*ΔNumber*) and sentiment (*ΔSentiment*) between public and private earnings calls, respectively. In both panels, the left (right) bar reports the disclosure statistics in the public (private) call for each topic. For a given topic, the 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 (Δ*Length*) and listenability (Δ*Listenability*) between public and private calls at topic level in Panel **(a)** and Panel **(b)**, respectively. The decomposition of vocal valence inconsistencies (Δ*Valence*) and arousal inconsistencies (Δ*Arousal*) are reported in Panel (**c**) and Panel (**d**), respectively.

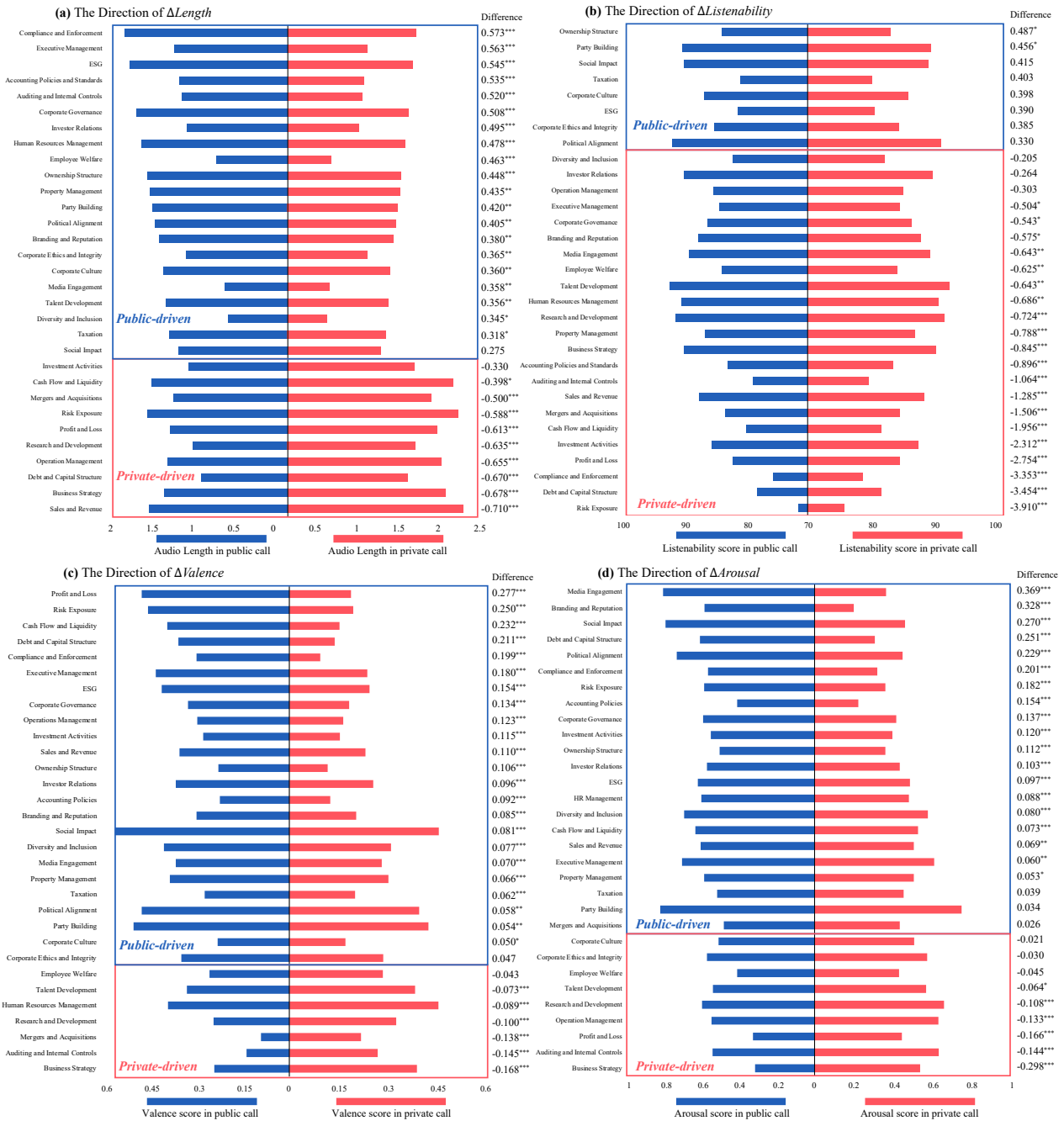


Figure 5. Direction of Vocal Inconsistencies. We provide the directional differences in audio length (*ΔNumber*), listenability (*ΔListenability*), vocal valence (*ΔValence*) and arousal (*ΔArousal*) between public and private earnings calls in Panels **(a)**, **(b)**, **(c)** and **(d)** , respectively. In all panels, the left (right) bar reports the disclosure statistics in the public (private) call. For a given topic, the 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.

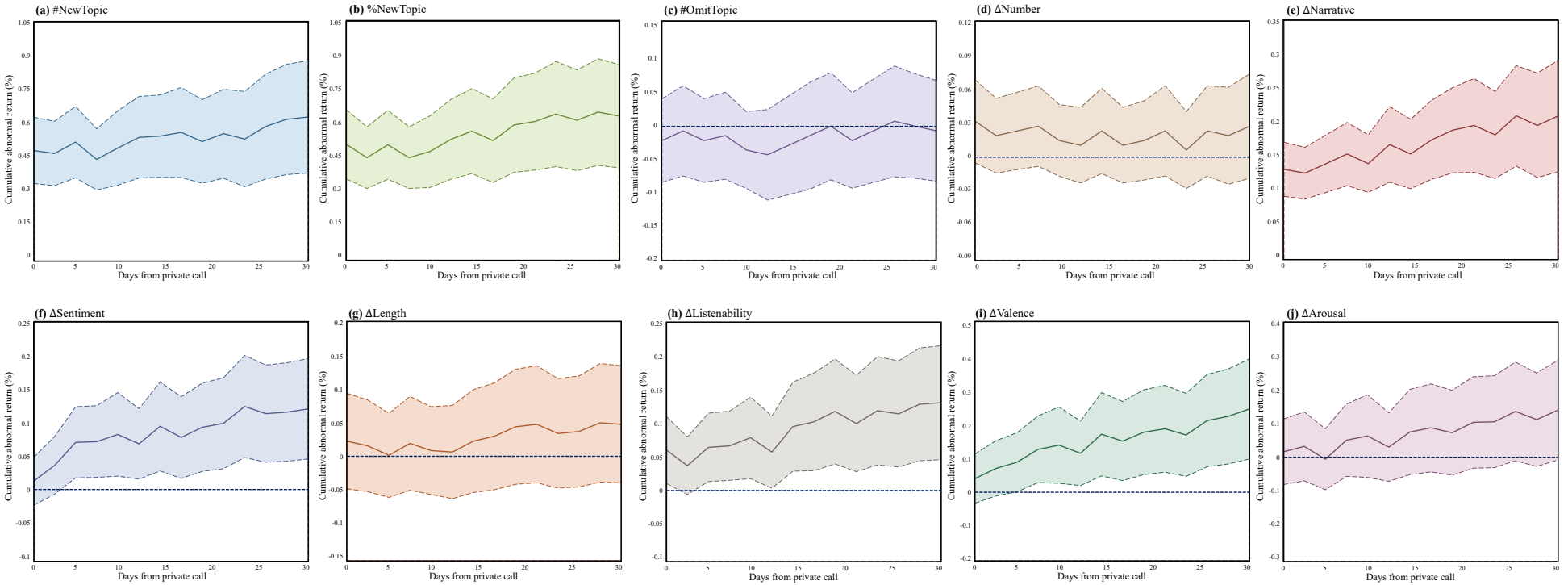


Figure 6. Disclosure Inconsistencies and Stock Returns. In this Figure, we present the coefficients from a series of regressions of cumulative market-adjusted returns (CAR) on the disclosure inconsistencies measures, including #NewTopic **(a)**, %NewTopic **(b)**, #OmitTopic **(c)**, ΔNumber **(d)**, ΔNarrative **(e)**, ΔSentiment **(f)**, ΔLength **(g)**, ΔListenability **(h)**, ΔValence **(i)**, ΔArousal **(j)**. In each Panel, the solid line indicates the estimated coefficients β (Specification (5)) from CAR[0,1] to CAR[29,30] while dashed lines show 95% confidence intervals.

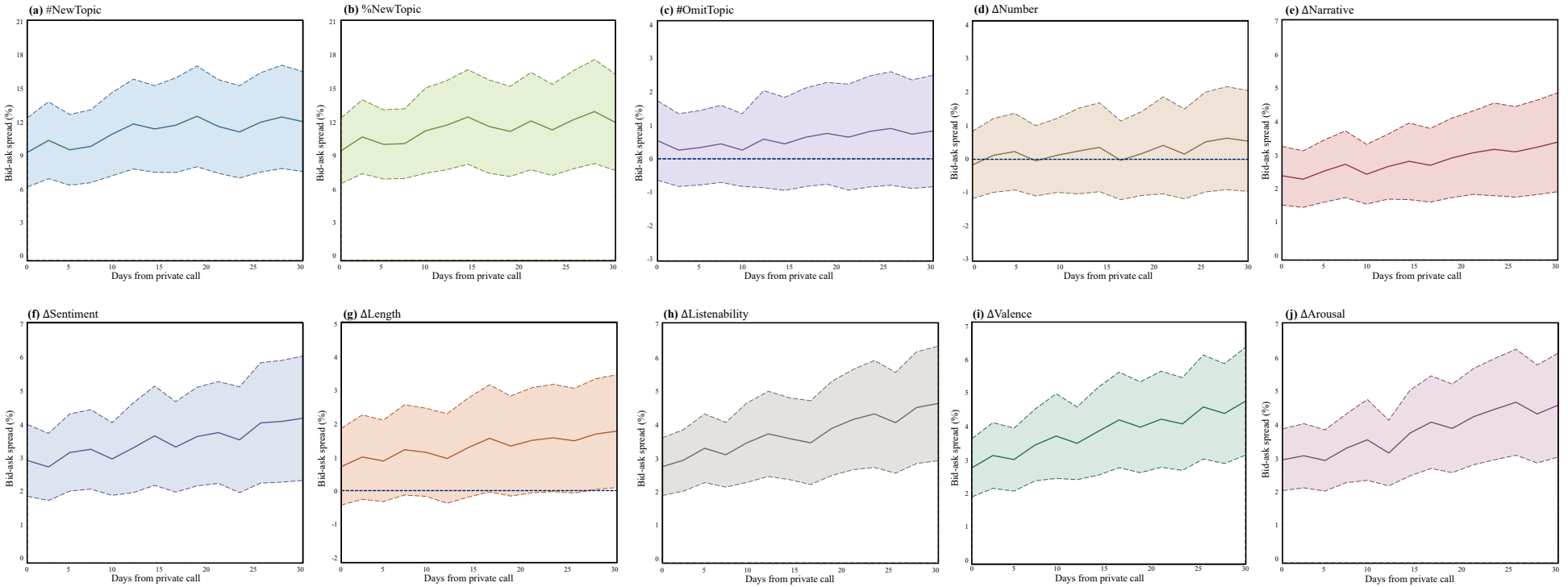


Figure 7. Disclosure Inconsistencies and Information Asymmetry. In this Figure, we present the coefficients from a series of regressions of bid-ask spread on the disclosure inconsistencies measures, including #NewTopic **(a)**, %NewTopic **(b)**, #OmitTopic **(c)**, ΔNumber **(d)**, ΔNarrative **(e)**, ΔSentiment **(f)**, ΔLength **(g)**, ΔListenability **(h)**, ΔValence **(i)**, ΔArousal **(j)**. In each Panel, the solid line indicates the estimated coefficients β (Specification (5)) from Bid-ask spread [0,1] to Bid-ask spread [29,30] while dashed lines show 95% confidence intervals.

Table 1 Summary Statistics

This Table reports the descriptive statistics of the disclosure inconsistencies measures and the independent variables we introduced in Section 4. The detailed definitions of variables are provided in Table A1 of the Appendix A1.

	Mean	Median	Std. Dev.	Q1	Q3	N
Measures of Disclosure Inconsistency						
#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
Call Characteristics						
Online	0.682	0.468	0.468	0	1	20,480
Team	0.819	0.404	0.404	0	1	20,480
SOFI	0.685	0.630	0.372	0	1	20,480
Affiliation	0.091	0.087	0.055	0	0	20,480
Manager Characteristics						
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
Finance degree	0.107	0.096	0.084	0	0	20,480
Law degree	0.063	0.058	0.058	0	0	20,480
Firm Characteristics						
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 Determents of Topic Inconsistencies

In this table, we examine the determinants of topic inconsistencies between public-private call pairs at different levels. *#NewTopic*, *%NewTopic* and *#OmitTopic* are three measures of topic inconsistencies. The variables of interest are included at earnings call-, manager- and firm-level. Regressions are run at the firm-quarter 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)
	#NewTopic	%NewTopic	#OmitTopic
Online	-0.164***	-0.029***	-0.277***
	(-3.17)	(-4.64)	(-3.41)
Team	-0.117**	-0.025***	-0.176
	(-2.51)	(-3.15)	(-0.92)
SOFI	-0.093***	-0.027***	-0.185***
	(-3.10)	(-2.79)	(-2.93)
Affiliation	0.281***	0.040***	0.389**
	(2.86)	(3.10)	(2.19)
Gender	-0.026**	-0.011	-0.045**
	(-2.16)	(-0.85)	(-1.98)
Age	0.043***	0.009**	0.075***
	(2.70)	(2.19)	(3.10)
Tenure	0.074***	0.028**	0.150***
	(3.42)	(3.03)	(3.74)
Finance degree	-0.029***	-0.024***	-0.101***
	(-2.83)	(-2.75)	(-3.85)
Law degree	-0.138***	-0.020***	-0.215***
	(-3.65)	(-2.87)	(-3.50)
IO	-0.139***	-0.015**	-0.171***
	(-2.71)	(-2.38)	(-3.75)
Leverage	0.154***	0.016***	0.317***
	(2.94)	(3.10)	(2.60)
Dividend	0.033	0.008	0.009
	(1.09)	(1.30)	(0.60)
ROA	-0.248***	-0.058***	-0.205***
	(-3.13)	(-4.19)	(-2.69)
Firm age	0.134*	0.027	0.092
	(1.89)	(1.28)	(0.95)
Asset	-0.182**	-0.019**	-0.265
	(-2.44)	(-2.45)	(-1.31)
SOE	-0.141***	-0.043***	-0.220***
	(-3.71)	(-3.50)	(-2.87)
Fixed effect	Yes	Yes	Yes
Observation	20,480	20,480	20,480
Adj. R^2	0.325	0.412	0.356

Table 3 The Determents of Textual Inconsistencies

In this table, we examine the determinants of textual inconsistencies between public-private call pairs. Δ*Number*, Δ*Narrative* and Δ*Sentiment* are three measures of textual inconsistencies. The variables of interest are included at earnings call-, manager- and firm-level. Regressions are run at the firm-quarter 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)
	ΔNumber	ΔNarrative	ΔSentiment
Online	-0.035**	-0.049***	-0.011**
	(-2.53)	(-2.76)	(-2.11)
Team	-0.019***	-0.032**	-0.009**
	(-2.67)	(-2.35)	(-2.12)
SOFI	-0.006	-0.013	-0.001
	(-1.31)	(-1.33)	(-1.23)
Affiliation	0.022***	0.045***	0.003***
	(3.23)	(3.14)	(3.08)
Gender	-0.012	-0.083**	-0.005**
	(-1.39)	(-2.32)	(-2.20)
Age	0.026***	0.094***	0.004**
	(2.79)	(2.83)	(2.42)
Tenure	0.010***	0.056***	0.010**
	(3.05)	(3.01)	(2.17)
Finance degree	-0.017**	-0.061**	-0.012**
	(-2.50)	(-2.18)	(-2.29)
Law degree	-0.022***	-0.046***	-0.014***
	(-3.51)	(-3.10)	(-3.27)
IO	-0.004	-0.027	-0.004
	(-1.17)	(-0.87)	(-1.05)
Leverage	0.022***	0.043***	0.005***
	(2.99)	(3.15)	(3.04)
Dividend	-0.018**	-0.092**	-0.004***
	(-2.32)	(-2.26)	(-3.21)
ROA	-0.024*	-0.108**	-0.008***
	(-1.83)	(-2.43)	(-2.82)
Firm age	0.016**	0.033***	0.004***
	(2.09)	(2.88)	(2.84)
Asset	-0.009**	-0.007***	-0.005**
	(-2.44)	(-2.81)	(-2.45)
SOE	-0.016**	-0.020***	-0.008***
	(-2.10)	(-2.93)	(-2.79)
Fixed effect	Yes	Yes	Yes
Observation	20,480	20,480	20,480
Adj. R^2	0.360	0.375	0.298

Table 4 The Determents of Vocal Inconsistencies

In this table, we examine the determinants of vocal inconsistencies between public-private call pairs. Δ*Length*, Δ*Listenability*, Δ*Valence* and Δ*Arousal* are four measures of vocal inconsistencies. The variables of interest are included at earnings call-, manager- and firm-level. Regressions are run at the firm-quarter 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)
	ΔLength	ΔListenability	ΔValence	ΔArousal
Online	-0.051**	-0.160***	-0.018**	-0.021***
	(-2.40)	(-2.75)	(-2.18)	(-3.17)
Team	-0.046*	-0.128**	-0.010*	-0.012***
	(-1.94)	(-2.35)	(-1.85)	(-3.05)
SOFI	0.005	-0.109*	-0.010	-0.015
	(0.46)	(-1.83)	(-1.16)	(-1.08)
Affiliation	0.075**	0.124**	0.006***	0.010***
	(2.48)	(2.23)	(3.47)	(4.12)
Gender	-0.040	-0.082*	-0.011**	-0.009*
	(-0.62)	(-1.86)	(-2.45)	(-1.84)
Age	0.008	0.070	0.012**	-0.013**
	(1.32)	(1.27)	(2.25)	(-2.47)
Tenure	0.073*	0.114***	0.019***	-0.012***
	(1.90)	(2.73)	(2.80)	(-3.14)
Finance degree	-0.065*	-0.150**	-0.016***	-0.008***
	(-1.87)	(-1.98)	(-3.25)	(-3.46)
Law degree	-0.055**	-0.144***	-0.012***	-0.010***
	(-2.47)	(-2.94)	(-3.77)	(-4.01)
IO	0.009	-0.073**	-0.008**	-0.014***
	(0.91)	(-2.34)	(-2.50)	(-3.05)
Leverage	0.046***	0.059**	0.013**	0.011**
	(2.66)	(2.02)	(2.03)	(2.40)
Dividend	0.044	0.065	0.024	0.018
	(0.63)	(0.96)	(1.32)	(1.29)
ROA	-0.054***	-0.074*	-0.018***	-0.019***
	(-2.89)	(-1.69)	(-3.60)	(-3.24)
Firm age	0.010	0.095**	0.010**	0.013**
	(0.77)	(2.03)	(2.11)	(2.45)
Asset	-0.037***	-0.075**	-0.013**	-0.015**
	(-2.80)	(-2.26)	(-1.99)	(-2.21)
SOE	-0.009*	-0.041**	-0.002**	-0.004***
	(-1.71)	(-2.40)	(-2.28)	(-2.79)
Fixed effect	Yes	Yes	Yes	Yes
Observation	20,480	20,480	20,480	20,480
Adj. R^2	0.285	0.312	0.334	0.328

Table 5 The Effects of Disclosure Inconsistencies on Stock Returns

In this table, we regress the cumulative market-adjusted returns on the disclosure inconsistency measures for our sample firms. CAR [0, 30] are the cumulative market-adjusted return during the window of one month to the private meeting. We report the coefficients of all the disclosure inconsistency measures on CAR [0, 30] in this Table. 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. Regressions are run at the firm-quarter 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.

	CAR [0, 30]						CAR [0, 30]				
	(1)	(2)	(3)	(4)	(5)		(6)	(7)	(8)	(9)	(10)
#NewTopic	0.591***					ΔSentiment	0.121**				
	(3.55)						(2.02)				
%NewTopic		0.610***				ΔLength		0.004*			
		(3.47)						(1.76)			
#OmitTopic			-0.006			ΔListenability			0.135***		
			(-1.10)						(3.07)		
ΔNumber				0.056*		ΔValence				0.242***	
				(1.75)						(2.84)	
ΔNarrative					0.198***	ΔArousal					0.339**
					(3.16)						(2.30)
Control variables	Yes	Yes	Yes	Yes	Yes	Control variables	Yes	Yes	Yes	Yes	Yes
Fixed effect	Yes	Yes	Yes	Yes	Yes	Fixed effect	Yes	Yes	Yes	Yes	Yes
Observation	20,480	20,480	20,480	20,480	20,480	Observation	20,480	20,480	20,480	20,480	20,480
Adj. R^2	0.196	0.197	0.194	0.196	0.196	Adj. R^2	0.196	0.194	0.196	0.196	0.197

Table 6 The Effects of Disclosure Inconsistencies on Information Asymmetry

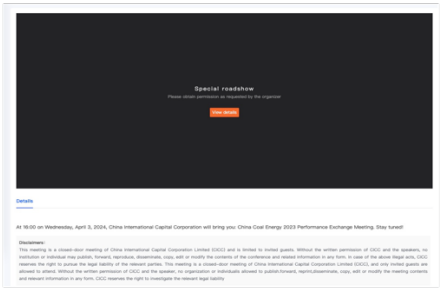
In this table, we regress the bid-ask spread on the disclosure inconsistency measures for our sample firms. Bid-ask spread [0, 30] is the average bid-ask spread during the window of one month to the private meeting. We report the coefficients of all the disclosure inconsistency measures on Bid-ask spread [0, 30] in this Table. 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. Regressions are run at the firm-quarter 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.

	Bid-ask spread [0, 30]						Bid-ask spread [0, 30]				
	(1)	(2)	(3)	(4)	(5)		(6)	(7)	(8)	(9)	(10)
#NewTopic	10.374***					ΔSentiment	3.686***				
	(3.78)						(3.24)				
%NewTopic		10.521***				ΔLength		1.067			
		(3.54)						(1.16)			
#OmitTopic			0.496			ΔListenability			4.215**		
			(0.22)						(2.28)		
ΔNumber				0.232		ΔValence				4.340***	
				(0.95)						(3.40)	
ΔNarrative					3.294***	ΔArousal					4.173***
					(2.85)						(3.06)
Control variables	Yes	Yes	Yes	Yes	Yes	Control variables	Yes	Yes	Yes	Yes	Yes
Fixed effect	Yes	Yes	Yes	Yes	Yes	Fixed effect	Yes	Yes	Yes	Yes	Yes
Observation	20,480	20,480	20,480	20,480	20,480	Observation	20,480	20,480	20,480	20,480	20,480
Adj. R^2	0.430	0.429	0.426	0.425	0.429	Adj. R^2	0.429	0.428	0.429	0.430	0.429

Appendix A1. Supplementary Figures and Tables



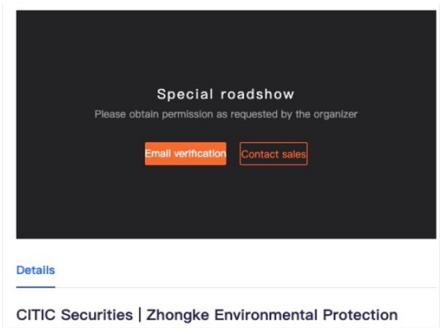
(a) Freely-access earnings call (China Coal Energy, 601898.SH)



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



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



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

Figure A1 Examples of Public and Private Earnings Calls

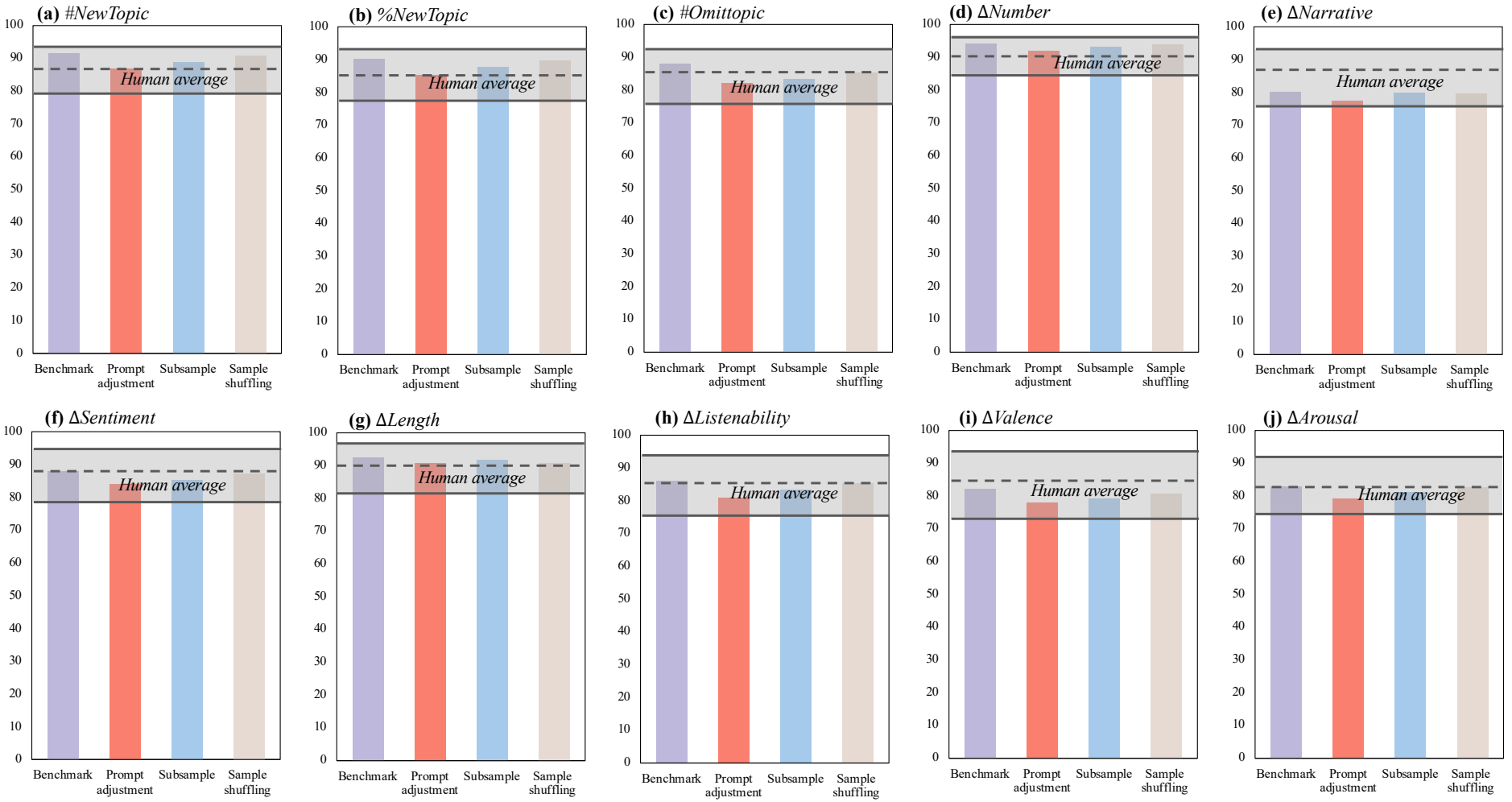


Figure A2. Turing Tests for Alternative Identification Frameworks. This Figure applies the Turing tests for alternative prompting strategies and sampling methods in assessing disclosure inconsistencies between public-private calls. We compare the accuracy of the disclosure inconsistencies measures for 200 randomly selected public-private call pairs by 5 human readers and alternative prompting strategies and sampling methods, where the benchmark is the human consensus scores. Accuracy is 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.

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
ΔNumber	-	The absolute differences of the numerical usage for the same topic disclosed between public and private calls
ΔNarrative	-	1-cosine similarity between narrative vectors in public call $\boldsymbol{V}_{i,k,m,t}$ and private call $\boldsymbol{V}_{i,k,n,t}$ for the same topic
ΔSentiment	-	The absolute differences of sentiment scores for the same topic in public and private earnings calls
ΔLength	-	The absolute differences of the total audio length in-between public and private calls for the same topic
ΔListenability	-	The absolute differences of the vocal disclosure listenability in-between public and private calls for the same topic
ΔValence	-	The absolute differences of the vocal disclosure valence in-between public and private calls for the same topic
ΔArousal	-	The absolute differences of the vocal disclosure arousal in-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
SOFI	-	Dummy that takes one if the co-host financial institution in private call is state-owned financial institution 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
Finance degree	-	Dummy that takes one if the manager has a degree in finance related majors otherwise zero
Law degree	-	Dummy that takes one if the manager has a degree in 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 earning call pairs, we filter the 115,372 earning call transcripts from Comein Finance database during our sample period using the following process. Filter [1] pairs of public and private call transcripts 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 4.3.

	Number of transcript pairs	Number of transcripts pairs dropped	Number of firms
Download transcripts from Comein Finance Database	-	-	2,463
Filter [1]: Pairs of public and private call transcripts 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 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 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 A4 Industrial Variations of Disclosure Inconsistencies

Industry	N	#NewTopic	%NewTopic	#OmitTopic	ΔNumber	ΔNarrative	ΔSentiment	ΔLength	ΔListenability	ΔValence	ΔArousal
Agriculture	583	1.226	0.134	1.808	0.079	0.303	0.040	0.192	0.780	0.069	0.076
Communication Services	441	0.809	0.089	2.195	0.077	0.324	0.074	0.357	0.938	0.129	0.110
Construction	1042	1.119	0.116	1.818	0.120	0.350	0.082	0.357	1.092	0.117	0.143
Culture, Sports and Entertainment	685	1.024	0.103	1.887	0.094	0.388	0.072	0.313	0.903	0.106	0.104
Education	665	1.440	0.148	1.951	0.144	0.453	0.076	0.306	0.901	0.122	0.112
Environment and Public Facility	1991	1.500	0.150	2.439	0.179	0.525	0.097	0.465	1.318	0.131	0.161
Health Care and Social Work	1741	1.512	0.154	2.370	0.197	0.481	0.113	0.491	1.308	0.183	0.210
Leasing and Commercial Services	1267	1.238	0.126	2.014	0.156	0.506	0.128	0.526	1.409	0.148	0.239
Manufacturing	2771	1.740	0.189	2.539	0.104	0.406	0.063	0.292	0.899	0.084	0.110
Mining	989	0.750	0.079	1.866	0.081	0.314	0.077	0.380	1.007	0.104	0.149
Others	1222	1.286	0.136	1.733	0.092	0.368	0.058	0.251	0.778	0.094	0.082
Real Estate	1954	1.167	0.121	1.924	0.117	0.334	0.078	0.367	1.115	0.134	0.115
Residential and Other Services	816	1.100	0.117	2.020	0.118	0.384	0.070	0.341	0.950	0.091	0.095
Scientific Research and Technical Services	1704	1.855	0.204	2.555	0.158	0.479	0.124	0.617	1.505	0.177	0.201
Transportation and Postal Services	750	0.881	0.091	1.802	0.082	0.321	0.068	0.295	0.873	0.105	0.127
Utilities	793	1.052	0.113	1.707	0.146	0.367	0.087	0.377	1.008	0.137	0.142
Wholesale and Retail	1064	0.910	0.099	1.829	0.119	0.388	0.063	0.282	0.831	0.091	0.114

Table A5 Top and Bottom Topic Inconsistencies: Sub-topic Level Statistics

This Table reports the sub-topic level 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 A6 Top and Bottom Textual Inconsistencies: Sub-topic Level Statistics

This Table reports the sub-topic level 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.

ΔNumber			ΔNarrative			ΔSentiment		
Top Topic	Subtopic	Mean	Top Topic	Subtopic	Mean	Top Topic	Subtopic	Mean
Top-ranked Top-subtopics								
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
Bottom-ranked Top-subtopics								
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 A7 Top and Bottom Vocal Inconsistencies: Sub-topic Level Statistics

This Table reports the sub-topic level 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.

ΔLength			ΔListenability			ΔValence			ΔArousal		
Top Topic	Subtopic	Mean	Top Topic	Subtopic	Mean	Top Topic	Subtopic	Mean	Top Topic	Subtopic	Mean
Top-Ranked Top-subtopics											
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
Bottom-Ranked Top-subtopics											
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 Engagement	Media Communication	0.277	Political Alignment	Common Prosperity	0.265	Political Alignment	Common Prosperity	0.056	Political Alignment	Rural Revitalization	0.030

Table A8. Demonstration of Textual Inconsistencies

This Table provides example transcripts of textual inconsistencies between public and private earnings calls, including differences in numerical usage (*ΔNumber*), disclosure narrative (*ΔNarrative*), and text-based sentiment (*ΔSentiment*). For each textual inconsistency measure, we provide the transcripts of call pairs with both higher and lower inconsistency measures in the respective dimension.

	Public Call Transcript	Private Call Transcript
Call pair with high* Δ*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* Δ*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* Δ*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* Δ*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* Δ*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* Δ*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.

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 topics and 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 sentence into 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 call transcripts from the Comin Finance database for publicly listed companies in Chinese stock markets during the period of January 2020 to January 2025. Comein Finance is the leading call transcript 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 call transcripts across 976 publicly listed companies in China. Each transcript 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 follow Kim et al. (2024) and 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 ×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 ×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 topic K_t, subtopic Ks 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 A10 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 A11. 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 A12 provide 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 identifies assignment errors such as, invalid or non-existent topic codes ("None", "Error"), 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:

[14] Profit and Loss

[1402] 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 the third 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) 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 call transcripts.

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. (2025). 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 $\left[t_s^{start},\ t_s^{end}\right]$; 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 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:

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

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 articulation and rapid pacing for the WhisperX system to recognize. To quantify topic-level inconsistency in Listenability, we aggregate sentence-level values within each topic and compute ***ΔListenability*** 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 recognize vocal emotions in a more robust manner.

As a result, 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*. 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 *ΔValence* and *ΔArousal* as the absolute differences between public and private calls on the same topic.

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

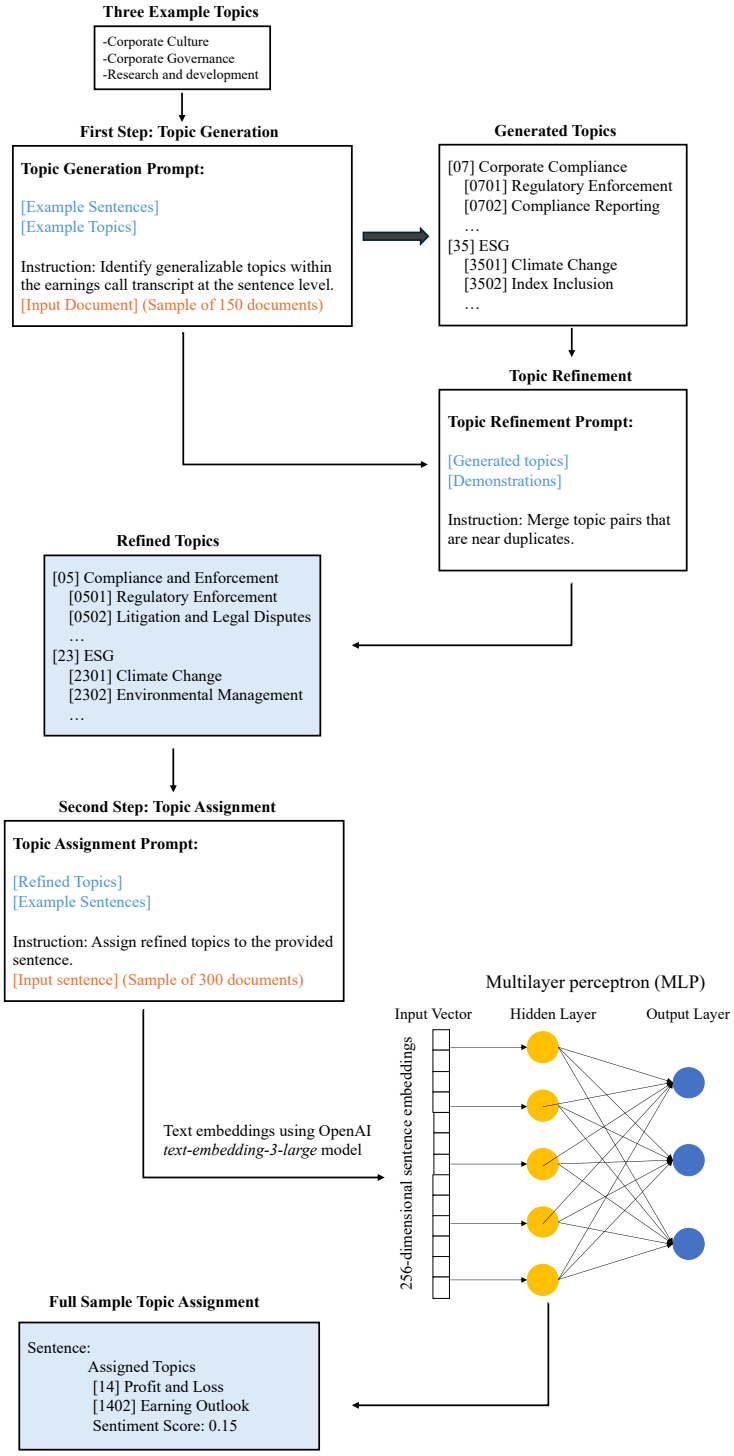


Figure A3. Pipeline of Topic Modelling using TopicGPT. In this paper, the Topic GPT modelling consists of two steps. 1) Topic Generation: Given three manually curated example topics, TopicGPT generates new topics in the 100 sample earning call documents. The framework then refines the list by merging 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 A8 Prompt Template for Topic Generations

Prompt for topic generations

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 A9 Prompt Template for Refining and Merging Topics

Prompt for topic refining and merging

You will receive a list 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 A10 Prompt Template for Topic Assignment and Sentiment Analysis

Prompt for topic assignment and sentiment analysis

You will receive a document 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:
[14] Profit and Loss
[1402] 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:
[20] Branding and Reputation
[2001] 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]