Learning Representation for Earnings Call Transcript via Structure-Aware Key Insight Extraction

Anonymous ACL submission

Abstract

Learning representations for earnings call transcripts encounter significant challenges, such as the unreliability of the knowledge encoding process and specific domain-specific requirements in the financial context. To address these challenges, this work proposes a self-supervised transcript representation learning approach that utilizes structural information within transcripts to provide supervision signals. Additionally, it offers concise explanations for each decision made by the neural networks through a redundancy-aware key sentence extractor. Extensive experiments across various downstream tasks, such as risk prediction, information retrieval, and firm similarity analysis, demonstrate the effectiveness of our approach.

1 Introduction

An earnings call is a conference call in which the management team of a public firm, including executives, communicates with analysts, investors, and journalists (Chen et al., 2018). Historically, firms relied on analysts to manually examine transcripts of these calls to glean valuable insights for investment decision-making. However, due to the extensive length of the transcripts and the specialized knowledge necessary for analysis, many finance professionals find it difficult and time-consuming to extract key information (Bloomfield, 2008). Moreover, analyzing transcripts manually is susceptible to biases and errors, which can lead to inaccurate identification of crucial information (Sawhney et al., 2021).

Neural Networks (NNs) have demonstrated excellent capabilities in analyzing financial texts, including sentiment analysis (Nopp and Hanbury, 2015), stock volatility prediction (Yang et al., 2022), startup recommendation (Kim et al., 2020), etc. A common theme among these studies is representation learning, a technique capable of automatically extracting relevant features and patterns from transcripts, thereby generating a structured representation easily amenable to decoding and analysis by NNs. While classic representation learning methods like BERT and its variations have achieved significant success in generating contextual text representation, applying them to earnings call transcripts poses several challenges. First, the black-box nature of deep learning makes the high-dimensional encoding process of these models often opaque and unreliable (Bang et al., 2021). Second, financial texts often involve specific terminology, concepts, and industry contexts that may necessitate specialized processing beyond the capabilities of standard representation learning models (Nugent et al., 2023).

The primary challenge in achieving explainability in representation learning lies in the high-dimensional complexity of the input data. It is difficult for humans to understand how the model learns semantic connections from thousands of tokens and generates dense representations (Li et al., 2021; Bang et al., 2021). Recent studies, grounded in human cognitive theory (Ding et al., 2020; Baddeley, 1992; Broadbent, 2013), demonstrate that when understanding text semantics, humans tend to retain only the most crucial information in their memory units and can accomplish most text-related tasks based on this stored knowledge. Inspired by this theory, we propose a two-step learning process. In the first step, we develop a redundancy-aware key sentence extractor that emulates human behavior to extract pivotal insights from lengthy earnings call transcripts. In the second step, we employ these “essential sentences" with any off-the-shelf language model to generate the representation. Consequently, the distilled key sentences can offer succinct yet comprehensive explanations that aid humans in understanding which sentences influence the creation of the final representation by the NNs, thereby improving the NNs’ explainabil-

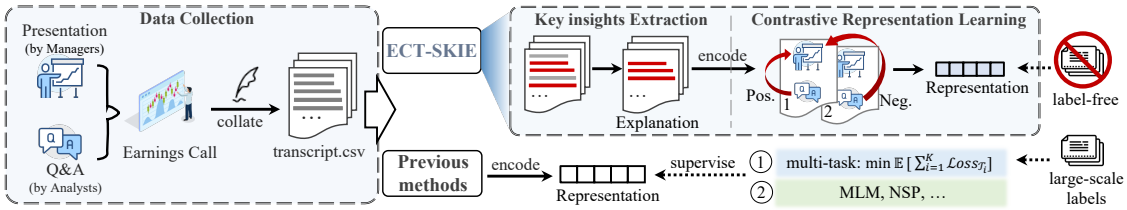

Figure 1: Illustration of data collection and workflow for earnings call transcripts representation learning. Traditional text representation learning relies on various downstream tasks or general self-supervised training approaches like masked language model (MLM) and next sentence prediction (NSP). In contrast, our approach, ECT-SKIE (**E**arnings **C**all **T**ranscript via **S**tructure-Aware **K**ey **I**nsight **E**xtraction), is explainable, label-free, and specialized for financial text understanding.

ity (Bang et al., 2021).

One way to address the problem of diverse requirements is to gather a large number of downstream financial tasks and use a multi-task learning framework for joint training (Liu et al., 2015). However, this solution is not practical for earnings call transcripts, since transcripts are limited in number and their downstream data are difficult to collect (Mukherjee et al., 2022). Inspired by recent advances in contrastive learning, we introduce a novel self-supervised learning approach that harnesses the inherent structure of transcripts to provide customized supervision signals for financial analysis.

Our method has several appealing properties: ***Label-free.*** It does not require any high-cost labels in training, and yet is capable of extracting pertinent key information effectively via exploiting the structural information in transcripts. ***Explainability.*** It endows the representation learning model with explainability and reliability by providing a concise explanation (i.e., discard irrelevant sentences) for every single decision made by a black-box model. ***Task-agnostic.*** It is a task-agnostic method, yet it can yield impressive results across various downstream financial tasks, such as risk forecasting, information retrieval, and firm similarity analysis. Figure 1 illustrates the differences between our method and conventional approaches.

2 Related work

We briefly review the two main categories of related literature and position our work in that context.

Extractive summarization. Extractive summarization aims to generate concise and coherent summaries by selecting and assembling salient sentences from a given source text. Over the years, the field has witnessed significant advancements, particularly through the exploration of both supervised and unsupervised techniques. Recent supervised methods involve reinforcement learning (Gu et al., 2022) and graph learning (Wang et al., 2020). Unsupervised approaches exploit intrinsic text features such as graph-based centrality scoring of sentences (Erkan and Radev, 2004; Mihalcea and Tarau, 2004) and sequence correlations (Liu and Lapata, 2019; Padmakumar and He, 2021). (Jie et al., 2023) introduces a transformer-based summarization network with controllable length. In addition, Large Language Models like ChatGPT have powerful capabilities for abstractive summarization and extractive summarization. However, they are likely to suffer from some uncontrollable problems. Given the user prompt *"Extract 70% of the sentences as key sentences from the given text. Don't break up a sentence."*, the extracted sentences are inconsistent with the origin sentence and there is also a considerable bias in summary length (cf. Figure 6 in Appendix A).

Document representation learning. Document-level representation learning has witnessed diverse techniques aimed at capturing the semantic essence of entire documents. The Doc2Vec (Le and Mikolov, 2014) method extends Word2Vec (Mikolov et al., 2013) to generate paragraph vectors, providing effective representations for paragraphs and documents. FastText (Joulin et al., 2017) combines subword information with word embeddings, demonstrating strength in tasks with limited data. Transformer-based models, including Longformer (Beltagy et al., 2020) and BigBird (Zaheer et al., 2020), address efficient processing of lengthy documents. Pretrained language models like BERT (Kenton and Toutanova, 2019) and GPT (Floridi and Chiriatti, 2020) can be adapted for document-level tasks with fine-tuning, showcasing their versatility. Hierarchical models like Hierarchical Attention Networks (Yang et al.,

2016; Ma et al., 2021) leverage word and sentence embeddings to create informative document-level representations (Zhang et al., 2022). These approaches collectively showcase the rich array of methods employed to capture document semantics for various downstream tasks.

Summary of Differences. In a paradigm sense, our method is a representation learning method, and quite similar to extractive summarization. The core difference between our approach and extractive summarization methods is that our model is self-explanatory, and dedicated to the financial domain.

3 Methodology

The primary objective of transcript representation learning is to develop a neural network that is capable of projecting each transcript X into a dense d-dimensional vector $h_\theta(X)$. A good representation should encapsulate crucial information that can be used for various downstream financial tasks.

3.1 Problem formulation

Formally, let $h(\cdot)$ denote a shared representation function that maps the input data to a representation in a higher-dimensional space, and let $g_i(\cdot)$ denote a task-specific function that maps the higher-dimensional representation to the output space for task $\mathcal{T}_i$. We seek to optimize the following objective function:

$$\theta^* = \underset{g_1,g_2,\ldots,g_{|\mathcal{T}|}}{\operatorname{argmin}} \sum_{k=1}^{|\mathcal{T}|} \gamma_k \mathcal{L}_k(g_k(h_\theta(X)), Y)$$

where $\mathcal{L}$ is a loss function that measures the discrepancy between the predicted outputs and the true outputs, and the hyperparameter γ controls the trade-off between fitting the individual tasks and sharing the representation across tasks. In this work, we aim to design a *model-agnostic* and *task-agnostic* representation learning approach that can effectively satisfy various downstream tasks.

3.2 ECT-SKIE Framework Overview

Inspired by human cognitive theory, we propose a two-step learning process to learn a more explainable and trustworthy representation. First, we propose a key sentence extractor to mimic human behavior and extract key insights from long earnings call transcripts. Second, we use these "essential sentences" with arbitrary off-the-shelf language models to generate the representation.

The core of our method is to find key sentences $\kappa(X)$ given a transcript X that satisfies:

$$\mathbb{E}_{\mathcal{T}}\left[\mathcal{L}(g(h_\theta(\kappa(X))), Y))\right] \approx \mathbb{E}_{\mathcal{T}}\left[\mathcal{L}(g(h_\theta(X)), Y))\right]. \quad (1)$$

where $\kappa(\cdot)$ denotes the key sentence set. Generally, key sentences in earnings call transcripts provide the most important information about the company's financial performance and prospects. These sentences are usually spoken by the company's CEO or CFO and are often included in the transcript of the earnings call. They can include information about revenue growth, earnings per share, cash flow, and other key financial metrics. Additionally, key sentences can also include information about other important developments that could impact the company's future performance.

In the following, we first propose a supervised approach to extract key sentences from lengthy transcript (§3.3). Then, we point out the challenge of limited data resources, and elaborate on how to extract key sentences in a self-supervised manner (§3.4). Notations are attached in Appendix B.

3.3 Key Insights Extraction

The most direct solution for extracting key sentences is to devise a neural network $f_\psi(X) : \mathbb{R}^{N\times d} \to \mathbb{R}^N$ that generates a binary mask $M \in \mathbb{R}^N$ for each transcript with N sentences. ψ denotes the trainable parameters of the extractor. Accordingly, we can implement a key sentence extractor via:

$$\kappa_\psi(X) = f_\psi(X) \odot X \quad (2)$$

where $\kappa_\psi(X)$ is the selected sentence set by neural networks. We use $\odot$ as a key sentence selector, i.e., we put sentences X_i into $\kappa_\psi(X)$ if M_i equals 1. Accordingly, we can optimize the model via:

$$\psi^* = \arg\min_{\psi} \mathbb{E}_{\mathcal{T}}\left[\mathcal{L}(g(h_\theta(\kappa_\psi(X))), Y))\right]. \quad (3)$$

$\kappa_\psi(X)$ acts as a compressed representation that captures the most relevant and distilled information for downstream tasks. Directly optimizing Eq. (3) cannot guarantee the conciseness and accuracy of the extracted sentence set. To this end, we can borrow Information Bottleneck (IB) theory (Tishby and Zaslavsky, 2015) to address the shortcoming:

$$\psi^* = \arg\min_{\psi} \mathbb{E}_{\mathcal{T}}\left[\mathcal{L}(g(h_\theta(\kappa_\psi(X)))), Y))\right] + \beta\, I(X, \kappa_\psi(X)), \quad (4)$$

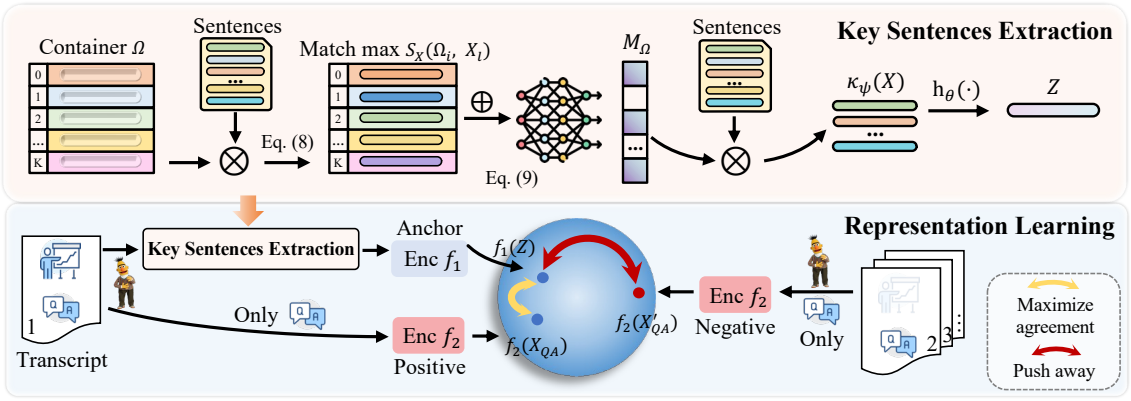

Figure 2: Overall pipeline of ECT-SKIE. Earnings call transcripts are encoded and fed into the *Key Sentences Extraction* module, generating a concise explanation denoted as M_Ω and the corresponding transcript representation Z. ECT-SKIE is optimized via InfoNCE objective, in which key sentences extracted from the text serve as anchors. The Q&A section is encoded as a positive sample, while sections from other transcripts are treated as negative samples.

where β is a weight hyper-parameter that achieves a trade-off between the knowledge sufficiency and the compression ratio of information. Inspired by prior studies (Bang et al., 2021; Kim et al., 2021), we can use a variational approximation of the second term as:

$$I(X, \kappa_\psi(X)) \leq \mathbb{E}\left[D_{KL}(\mathbb{P}_\psi(M_s|X), r(M_s))\right], \quad (5)$$

where $M_s \in \mathbb{R}^N$ is the output of $f_\psi(X)$, and $r(M_s)$ is the prior distribution of the mask M_s.

Pitfalls in optimizing Eq. (5). Typically, for the neural network $f_\psi(X)$, sentences $X \in \mathbb{R}^{N\times d}$ are fed as input, and a binary mask $M \in \mathbb{R}^N$ will be the output of $f_\psi(X)$ where 1 indicates being selected as a key sentence and 0 otherwise. Although we have applied the IB strategy to force the model to generate concise explanations, the above key sentence extractor still suffers from the redundancy problem in practice. As shown in Figure 3, if two sentences are semantic the same and important, both will be selected as key sentences.

Redundancy-aware key sentence extractor. To address the redundancy problem, we propose a simple yet effective solution by adding a container with K slots $\Omega \in \mathbb{R}^{K\times d}$ ($K < N$). Instead of determining which sentence is important, we alternatively consider which slot in the container is important. Specifically, the improved redundancy-aware key sentence extractor consists of the following steps:

(1) Bind candidate sentences into the container. We calculate the score between sentence X_l and slot Ω_i via:

$$\mathcal{S}_X(\Omega_i, X_l) = \frac{\exp(\text{Sim}(\Omega_i, X_l))}{\sum_{j=1}^N \exp(\text{Sim}(\Omega_i, X_j))}, \quad (6)$$

where $\text{Sim}(\Omega_i, X_l)$ is cosine similarity. The slot Ω_i will bind with the sentence S with the highest score:

$$\Omega_i(S) = \arg\max_{S=X_l} \mathcal{S}_X(\Omega_i, X_l), \quad \text{for } l = 1, \ldots, N. \quad (7)$$

As a result, each slot will hold one sentence, and a sentence may be associated with multiple slots.

(2) Select top-ranked slots. Having associated candidate sentences with the container, our subsequent goal is to choose the top N_s ranked slots, thereby extracting at most N_s key sentences. For this purpose, we begin by combining the slot representation with its associated sentence vector to create a transcript-aware slot representation. Subsequently, we concatenate the representation of the slot and the sentence it holds. Multi-layer Perceptron $\text{MLP}(\Omega_i \oplus X_j) : \mathbb{R}^{2d} \to \mathbb{R}$ is employed to produce the slot scores.

$$\mathcal{S}_\Omega(\Omega_i, S_i) = \frac{\exp(\text{MLP}(\Omega_i \oplus S_i))}{\sum_{j=1}^K \exp(\text{MLP}(\Omega_j \oplus S_j))}. \quad (8)$$

We select the top-N_s slots and extract their corresponding sentences to construct $\kappa_\psi(X)$. And the new training objective becomes:

$$\psi^* = \arg\min_\psi \mathbb{E}_\mathcal{T}\left[\mathcal{L}(g(h_\theta(\kappa_\psi(X))), Y))\right] + \beta\, \mathbb{E}\left[D_{KL}(\mathbb{P}_\psi(M_\Omega|X), r(M_\Omega))\right]. \quad (9)$$

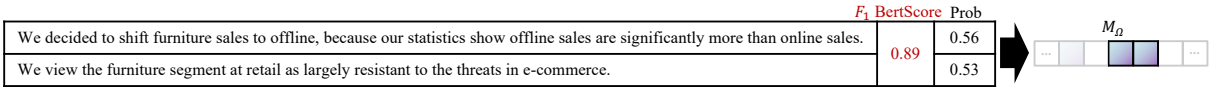

Figure 3: An example of redundancy in extracting sentences. Due to the characteristics of shared neurons, two sentences with comparable semantics often receive closely matched scores. Nonetheless, our goal of producing concise explanations necessitates the preservation of only one sentence.

Therefore, the mask M_s for sentences in Eq. 5 is rewritten to M_Ω for slots. With these modifications, the model can effectively resolve the redundancy problem: If two sentences have similar semantics, they will belong to the same slot with a high probability. The $\arg\max$ operation ensures that only the most important sentence is maintained among a group of sentences with similar semantics.

3.4 Self-supervised Representation Learning

Optimizing Eq. (9) is impractical due to the limited availability of transcripts and difficulties in acquiring their associated downstream data (Mukherjee et al., 2022). Take a closer look at Eq. (1), if we can find the ground-true key sentence set $\kappa(X)$, we no longer need any downstream tasks. The question now becomes, how to find key sentences without any downstream tasks verification?

As stated above, key sentences in earnings call transcripts provide the most important information about the company's financial performance and future prospects. In the Q&A section, analysts ask follow-up questions and request the executives to clarify information mentioned in the presentations, or they can solicit new information that the managers do not disclose in the Presentation section (Yang et al., 2022). Thus, the important information is also what investors pay attention to, and the topics asked by investors are often related to the company's key information (Chen et al., 2018). It gives us a chance to use the Q&A section as a surrogate supervision signal. Thus, the first term (called ψ_1) in Eq. (9) can be replaced with:

$$\begin{aligned}\psi_1 &= \arg\max_{\psi} I(\kappa_\psi(X), X_{QA}) \\ &= \mathbb{E}_{p(X_{QA},\kappa_\psi(X))}\left[\log \frac{p\left(X_{QA} \mid \kappa_\psi(X)\right)}{p\left(X_{QA}\right)}\right],\end{aligned} \tag{10}$$

where X_{QA} is the set of sentences in the Q&A section. To maintain a high level of explainability, we encode the extracted N_s key sentences $\kappa_\psi(X)$, using BERT and mean pooling operation to squeeze multiple representations into one vector. For the Q&A section, which has natural structural information that can be divided into multiple Q&A rounds, we feed each round's content into BERT and use the mean representation across rounds as the representation of the whole Q&A section. Then we use infoNCE (Chen et al., 2020) to estimate the lower bound of term Eq. (10):

$$\mathcal{L}_{\text{NCE}} = -\log \frac{\exp(\kappa_\psi(X) \cdot X_{QA}/\tau_1)}{\sum_{\mathcal{B}} \mathbb{1}_{X'_{QA} \notin X} \exp(\kappa_\psi(X) \cdot X'_{QA}/\tau_1)}, \tag{11}$$

where $\mathcal{B}$ denotes the batch size, $\mathbb{1}$ is an indicator function and τ_1 is a hyper-parameter tuning the balance between positive and negative samples (Due to space limitation, the encoder h_θ on $\kappa_\psi(X)$, X_{QA} and X'_{QA} is omitted in Eq. 11). Appendix C presents the detailed proofs. Finally, the task-agnostic self-supervised training objective is defined as:

$$\mathcal{L} = \mathcal{L}_{\text{NCE}} + \beta\, D_{KL}(\mathbb{P}_\psi(M^*_\Omega|X), r(M_\Omega)). \tag{12}$$

Given a transcript with N sentences, we select N_s sentences with a pre-defined ratio $\alpha = \frac{N_s}{N}$ and use a uniform distribution as the prior of $r(M_\Omega)$.

3.5 Implementation

Training objectives. The current form of the Eq. (12) is intractable due to the second term of summing over the $\binom{N}{N_s}$ combinations of candidate subsets. This is because we sample top N_s out of N sentences where each sentence is assumed to be drawn from a categorical distribution with class probabilities $\mathbb{P}(M_\Omega|X)$. Thus, we use the generalized Gumbel-softmax trick (Jang et al., 2017), which can be used to approximate a non-differentiable categorical subset sampling with differentiable Gumbel-softmax samples. The detailed procedures are as follows.

Continuous relaxation and reparameterization. First, we independently sample a sentence for N_s times. For each time, a random perturbation e_j is added to the log probability of each sentence:

$$n_i = -\log\left(-\log e_j\right), \quad \text{where } e_j \sim U(0,1)$$

$$\mathcal{S}'_\Omega(\Omega_i, S_i) = \frac{\exp\left(\left(n_i + \log(\mathcal{S}_\Omega(\Omega_i, S_i))\right)/\tau_2\right)}{\sum_{j=1}^{K} \exp\left(\left(n_i + \log(\mathcal{S}_\Omega(\Omega_j, S_j))\right)/\tau_2\right)},$$

where τ_2 is a tunable parameter regarding the temperature of the Gumbel-Softmax distribution. Next, we define a continuous-relaxed stochastic Mask $M_\Omega^* \in \mathbb{R}^K$ as the element-wise maximum of the independently sampled concrete vectors $\mathcal{S}'_\Omega$:

$$M_\Omega^* = \max_l \mathcal{S}'^{(j)}_{\Omega,i}(\Omega_i, S_i) \quad \text{for} \quad j = 1, \ldots, N_s.$$

With this sampling scheme, we approximate the N_s-hot random vector and have the continuous approximation to the variational bound. This trick allows us to use standard backpropagation to compute the gradients of the parameters via reparameterization. Analogously, as for Eq. (7), we can also use this generalized Gumbel-softmax trick to approximate the $\arg\max$ function.

4 Experiments

In this section, we delve into the evaluation of the transcript representation $\kappa_\psi(X)$ through three downstream tasks: risk forecasting, information retrieval and firm similarity, as discussed in §4.1. In §4.2 we present the inner workings of IB policies and containers through qualitative analysis. For reproducibility, the implementation is publicly available at: `https://anonymous.4open.science/r/ECT-SKIE-B2D6`.

Dataset. We have collected an extensive dataset of earnings call transcripts from U.S. firms, which is available through sources like the SeekingAlpha website[1] and databases such as Thomson Reuters StreetEvents[2]. Each transcript has been structured as a CSV file (illustrated in Appendix D). Following the precedent set (Yang et al., 2022; Ye et al., 2020), we designate the years 2015-2016 as our training set, 2017 for validation, and 2018 for testing. Pertinent statistics of datasets are outlined in Table 1. Our dataset will be publicly released soon.

Table 1: Descriptive statistics of earnings call transcripts.

Year	2015	2016	2017	2018
# transcripts	10,168	9,765	10,431	11,147
# firms	3,398	3,427	3,451	3,616
avg. # tokens in Presentation	3,207	3,172	3,191	3,199
avg. # tokens in Q&A	4,347	4,197	4,222	4,245

Baselines. We choose several representative models as baselines for comprehensive evaluation of ECT-SKIE. For the risk forecasting task, we select two language models (BERT (Devlin et al., 2019), SimCSE (Gao et al., 2021)), three heuristic methods (LexRank (Erkan and Radev, 2004), TextRank (Mihalcea and Tarau, 2004)), and three risk forecasting methods (Profet (Theil et al., 2019), MR-QA (Ye et al., 2020), DialogueGAT (Sang and Bao, 2022)). For the information retrieval task, we use LexRank (Erkan and Radev, 2004), TextRank (Mihalcea and Tarau, 2004) and recent unsupervised extractive summarization method PMI (Padmakumar and He, 2021) as our baselines. Note that there are a lot of risk forecasting models using multi-modal data, including earnings call text, audio, etc. To be fair, only single-modal works based on transcripts are considered.

Setup. In this work, the settings of hyperparameter tuning include (bold indicate the final choice): the temperature for Gumbel-softmax approximation $\tau_2 - \{0.1, 0.2, 0.5, \mathbf{0.7}, 1\}$, learning rate $\{0.1, 0.01, \mathbf{0.001}, 0.0001\}$, the trade-off weight parameter $\beta - \{0, 0.01, \mathbf{0.1}, 1, 10\}$, the container size $K - \{50, 100, 200, \mathbf{400}, 800\}$ and the compression ratio $\alpha - \{0.1, 0.3, 0.5, \mathbf{0.7}, 0.9\}$. The temperature τ_1 in Eq. 11 is set to 0.1. Then, we set the batch size to 128 and use the AdaGrad algorithm to optimize our model.

Due to the space limit, more settings can be found in Appendix E. We report the parameter sensitivity of K and subjective evaluation of extracted sentences in Appendix F and G, respectively. In addition, at the end of Appendix G, five explanation cases (M_Ω) generated by our model are visualized in Table 7-11.

4.1 Performance on Downstream Tasks

Task I: Risk forecasting. To conduct the risk prediction task in this study, we obtain daily stock prices (dividend-adjusted) of each company in our sample from the CRSP database[3]. The risk of a public firm is commonly measured as the stock price volatility over a period of time. Since each earnings call transcript is associated with a date where the call is held, we calculate the volatility τ-day after the earnings call, where $\tau \in \{3, 7, 10, 15, 20, 60\}$ denotes different forecasting horizons, i.e., daily to weekly ($\tau = 3, 7$), weekly to monthly ($\tau = 10, 15, 20$), and quarterly ($\tau = 60$). Following (Ye et al., 2020), MSE and MAE are used for performance measures.

[1] https://seekingalpha.com
[2] https://www.streetevents.com
[3] https://crsp.org

Table 2: Performance comparisons on the risk forecasting task in terms of MSE and MAE. The best performance is in bold and the second-best results are underlined. The results are statistically significant ($p < 0.01$) under a one-tailed t-test.

Type	Method	Ratio	Mean Square Error (MSE)						Mean Absolute Error (MAE)					
			3d	7d	10d	15d	20d	60d	3d	7d	10d	15d	20d	60d
Language Model	BERT	100%	0.7401	**0.3603**	0.3070	0.2650	0.2340	0.1826	0.6724	0.4766	0.4295	0.3981	0.3728	0.3207
	SimCSE	100%	0.7320	0.3672	0.3073	0.2638	0.2330	0.1768	0.6690	**0.4704**	0.4299	0.3982	0.3739	0.3163
Summarization Model	LexRank	70%	0.7521	0.3904	0.3282	0.2832	0.2577	0.1979	0.6768	0.4879	0.4474	0.4157	0.3957	0.3389
	TextRank	70%	0.7528	0.3858	0.3244	0.2810	0.2562	0.1973	0.6779	0.4852	0.4451	0.4135	0.3951	0.3379
Risk Prediction	Profet	100%	0.8058	0.4233	0.3343	0.3272	0.2585	0.2130	0.6947	0.5141	0.4485	0.4544	0.3927	0.3574
	DialogueGAT	100%	0.7432	0.3824	0.3238	0.2663	0.2315	0.1813	0.6740	0.4831	0.4419	0.3975	0.3688	0.3182
	MR-QA	100%	0.7868	0.3998	0.3025	0.2561	0.2311	0.1792	0.7022	0.4920	**0.4259**	**0.3888**	**0.3699**	0.3170
*	W/o IB	70%	0.7725	0.3899	0.3407	0.2895	0.2684	0.2083	0.6891	0.4879	0.4582	0.4195	0.4021	0.3457
	ECT-SKIE	70%	**0.7222**	0.3630	**0.3014**	**0.2557**	**0.2307**	**0.1744**	**0.6616**	0.4711	0.4270	0.3926	0.3706	**0.3146**

As shown in Table 2, even with an information compression ratio of 70%, our model offers significant advantages. The variant of ECT-SKIE without the Information Bottleneck (IB) mechanism shows a substantial performance decline, which suggests that ECT-SKIE's use of the Question-Answering self-supervised paradigm and IB mechanism allows it to effectively compress information and filter out irrelevant noise. This observation is further supported by the real visual cases presented in Appendix G. In addition, compared to extractive summarization methods, our model is significantly different from them since ECT-SKIE retains more critical information related to risk volatility while TextRank and LexRank summarize the text, but are not sensitive to financial risks. In a nutshell, these results highlight ECT-SKIE's exceptional ability to extract risk-relevant information.

Table 3: Performance comparisons on the IR task.

Metric	LexRank	TextRank	PMI	ECT-SKIE
Precision	0.6400	0.8380	0.9300	**0.9440**
Mean Rank	4.6800	2.4020	2.1020	**1.7100**
Mean Reciprocal Rank	0.7274	0.8821	0.9595	**0.9627**

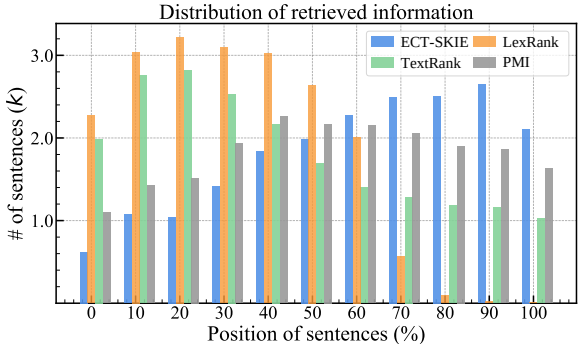

Figure 4: The percentage position distribution of the extracted sentences in source transcripts.

Task II: Information retrieval. Information retrieval serves the purpose of locating pertinent documents within a collection based on a designed query (Guo et al., 2022). This process can be used to evaluate the ability of ECT-SKIE and other unsupervised extractive summarization methods to extract overall information. Specifically, given a set of earnings call transcripts $\{d_1, d_2, ..., d_n\}$, for each transcript d_i, we encode the full text and extracted text as key k_i and query q_i, respectively. Then, with the candidate transcript representation set $\{k_1, k_2, ..., k_n\}$, we measure the relevance scores of query q_i to all keys and rank candidate transcripts based on the scores.

We test this task on TextRank, LexRank, PMI, and ECT-SKIE using three widely recognized metrics: Precision, Mean Rank, and mean reciprocal rank (MRR). Table 3 highlights ECT-SKIE's capability to distill the pivotal and pertinent information from transcripts when subjected to a high compression rate. When coupled with the insights depicted in Figure 4, it becomes evident that achieving superior retrieval prowess is closely linked to the Q&A section. The distribution of extracted sentences reveals that ECT-SKIE allocates greater attention to the Q&A section, particularly the later portions. This tendency can be attributed in part to the Q&A-based supervision signal that guides ECT-SKIE's learning process. Conversely, the other baseline methods exhibit a propensity to emphasize sentences occurring early in transcripts, notably in the presentation section. Broadly speaking, the sentences positioned early in a transcript tend to offer more insights into a firm's business situation as conveyed by its managers. Conversely, those positioned later often disclose finer points of risk that stimulate investors' interest, implying that the Q&A section contains more unique information that distinguishes one firm from the others.

Table 4: Five most similar firms to each focal firm in 2018.

Focal firm	Tesla Inc.	Alibaba Group HLDG	Intel Corp.	Visa Inc.	Pfizer Inc.
Similar firms	Delphi Technologies PLC	Pinduoduo Inc. -ADR	Advanced Micro Devices	Mastercard Inc.	Sanofi
	Methode Electronics Inc	Shopify Inc.	Axcelis Technologies Inc.	Citigroup Inc.	Novartis AG
	Enphase Energy Inc.	ChannelAdvisor Corp.	Applied Materials Inc.	Evertec Inc.	Incyte Corp.
	Polar Power Inc.	Groupon Inc.	NXP Semiconductors NV	Banco Santander SA	Biogen Inc.
	Constellium SE	Wayfair Inc.	Integrated Device Technology Inc.	Northern Trust Corp.	Gilead Sciences Inc.

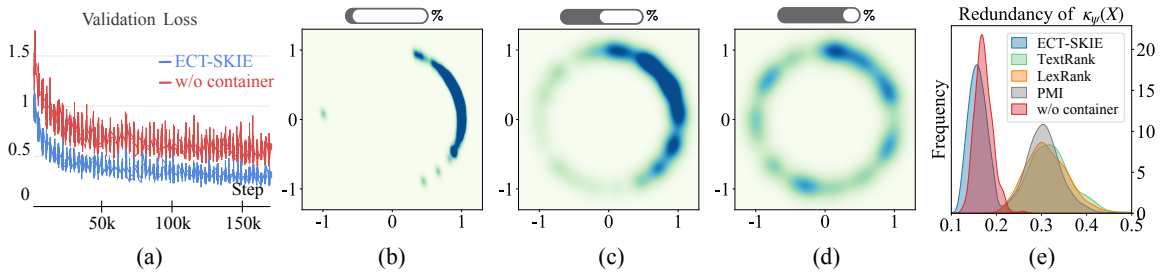

Figure 5: (a) The loss curves of the container ablation. (b, c, d) The three pictures show feature distributions of slots in containers at different training periods, derived by Gaussian kernel density estimation (KDE) in $\mathbb{R}^2$. (e) The redundancy distribution computed by cosine similarity between extracted sentences.

Task III: Firm similarity. Table 4 qualitatively proves that ECT-SKIE proficient in transcript representation learning. It is intuitive to anticipate that firms offering comparable products and services should exhibit resemblances in their transcript representations. This notion aligns harmoniously with earlier research in the field of business (Hoberg and Phillips, 2016), which utilized corporate disclosures to ascertain connections between economically related firms. This foundation also underpins text-based market analysis. Specifically, we randomly choose a set of focal firms. For each focal firm, we rank other firms by their similarity scores to the focal firm, and pick the top five. Results show the businesses of the focal firms are highly coincident or linked to their corresponding similar firms. This implies that ECT-SKIE can be applied to various financial tasks such as market competition analysis.

4.2 Redundancy Investigation

As discussed in §3.3, the issue of high redundancy will undoubtedly have a significant impact on the efficiency of information compression. To substantiate the effectiveness of our proposed redundancy-aware key sentence extractor, we conducted two meticulous experiments, charting the nuanced changes in training loss curves (Figure 5(a)) and the dynamic evolution of feature distribution throughout the training process (Figure 5(b-e)).

As illustrated in Figure 5(a), the model equipped with the container exhibits a notably lower training loss curve compared to its vanilla counterpart. This intriguing phenomenon aligns with our clarification that the container can effortlessly circumvent the limitations of redundancy. Figures 5(b-d) show that the distribution of features within slots gradually becomes more diverse as training unfolds. In a profound sense, each slot within the container can be interpreted as an embodiment of a distinct financial risk factor. As the training progresses, the container mechanism bestows upon ECT-SKIE the capability to holistically consider a spectrum of risk factors, thereby facilitating the acquisition of more streamlined representations from transcripts. Figure 5(e) serves as a compelling visual testimony, showcasing ECT-SKIE's heightened sentence diversity compared to its container-less variant (i.e., w/o container) and three baseline models by a significant margin.

5 Conclusion and Future Works

To improve the performance of representation learning for earnings call transcripts and address the black-box problem, we proposed ECT-SKIE, a promising approach for automatically extracting relevant information from earnings call transcripts. Our model leverages the structural information in transcripts to extract key insights effectively while providing concise explanations for each decision made by the model. We hope our research can shed light on the development of more efficient and effective transcript representation learning models for financial analysis.

6 Limitations

Our work has limitations that should be acknowledged. First, despite the fact that we have collected a sizeable earnings call transcript dataset, the transcripts are relatively old (2015 to 2018). This may cause performance degradation when the model is trained on older data but tested on newer ones. To this end, we plan to further improve the generalization of our model to avoid frequent retraining of the model due to changes in data samples, which will save resources and reduce carbon emissions during training and running. Second, the fixed compression ratio α implies that our model cannot dynamically choose the appropriate ratio of sentence extraction according to the SNR (Signal-to-noise Ratio) of the earnings call transcript at present. In addition, according to Eq. (7), each slot in the container pairs with one candidate sentence, and one candidate sentence might be associated with multiple slots. Thus, our model cannot extract a consistent amount of sentences to the ratio we pre-defined. Therefore, an important direction in our future work is to develop methods that can fix these limitations.

7 Ethics Statement and Potential Risks

We acknowledge that it is risky to invest in the market according to the output of our model. The earnings call transcript representation generated by our model may not be accurate in certain situations, e.g., processing transcripts with a particularly high SNR, attacked by some potential back door, and subjected to extreme information compression ratio. Therefore, for financial stock markets, our model should not operate independently and be put into use without human intervention. The output of our model should be carefully reviewed.

References

Alan Baddeley. 1992. Working memory. *Science*, 255(5044):556–559.

Seojin Bang, Pengtao Xie, Heewook Lee, Wei Wu, and Eric Xing. 2021. Explaining a black-box by using a deep variational information bottleneck approach. In *AAAI*, 13, pages 11396–11404.

Iz Beltagy, Matthew E Peters, and Arman Cohan. 2020. Longformer: The long-document transformer. *arXiv:2004.05150*.

Robert Bloomfield. 2008. Discussion of "annual report readability, current earnings, and earnings persistence". *Journal of Accounting and Economics*, 45(2-3):248–252.

Donald Eric Broadbent. 2013. *Perception and communication*.

Jason V Chen, Venky Nagar, and Jordan Schoenfeld. 2018. Manager-analyst conversations in earnings conference calls. *Review of Accounting Studies*, 23:1315–1354.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In *ICML*, pages 1597–1607.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT*, pages 4171–4186.

Ming Ding, Chang Zhou, Hongxia Yang, and Jie Tang. 2020. Cogltx: Applying bert to long texts. *NeurIPS*, 33:12792–12804.

Günes Erkan and Dragomir R Radev. 2004. Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of artificial intelligence research*, 22:457–479.

Luciano Floridi and Massimo Chiriatti. 2020. Gpt-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30:681–694.

Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. Simcse: Simple contrastive learning of sentence embeddings. In *EMNLP*, pages 6894–6910.

Nianlong Gu, Elliott Ash, and Richard Hahnloser. 2022. Memsum: Extractive summarization of long documents using multi-step episodic markov decision processes. In *ACL*, pages 6507–6522.

Jiafeng Guo, Yinqiong Cai, Yixing Fan, Fei Sun, Ruqing Zhang, and Xueqi Cheng. 2022. Semantic models for the first-stage retrieval: A comprehensive review. *ACM Transactions on Information Systems (TOIS)*, 40(4):1–42.

Gerard Hoberg and Gordon Phillips. 2016. Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, 124(5):1423–1465.

Eric Jang, Shixiang Gu, and Ben Poole. 2017. Categorical reparametrization with gumble-softmax. In *ICLR*.

Renlong Jie, Xiaojun Meng, Xin Jiang, and Qun Liu. 2023. Unsupervised extractive summarization with learnable length control strategies. *arXiv:2312.06901*.

Armand Joulin, Édouard Grave, Piotr Bojanowski, and Tomáš Mikolov. 2017. Bag of tricks for efficient text classification. In *EACL*, pages 427–431.

Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT*, pages 4171–4186.

Hyoung Jun Kim, Tae San Kim, and So Young Sohn. 2020. Recommendation of startups as technology cooperation candidates from the perspectives of similarity and potential: A deep learning approach. *Decision support systems*, 130:113229.

Jaekyeom Kim, Minjung Kim, Dongyeon Woo, and Gunhee Kim. 2021. Drop-bottleneck: Learning discrete compressed representation for noise-robust exploration. *arXiv:2103.12300*.

Shimon Kogan, Dimitry Levin, Bryan R Routledge, Jacob S Sagi, and Noah A Smith. 2009. Predicting risk from financial reports with regression. In *NAACL*, pages 272–280.

Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *ICML*, pages 1188–1196.

Haoran Li, Arash Einolghozati, Srinivasan Iyer, Bhargavi Paranjape, Yashar Mehdad, Sonal Gupta, and Marjan Ghazvininejad. 2021. Ease: Extractive-abstractive summarization end-to-end using the information bottleneck principle. In *Workshop on New Frontiers in Summarization*, pages 85–95.

Xiaodong Liu, Jianfeng Gao, Xiaodong He, Li Deng, Kevin Duh, and Ye-Yi Wang. 2015. Representation learning using multi-task deep neural networks for semantic classification and information retrieval. In *HLT-NAACL*, pages 912–921.

Yang Liu and Mirella Lapata. 2019. Text summarization with pretrained encoders. In *EMNLP*, pages 3730–3740.

Yixiao Ma, Shiwei Tong, Ye Liu, Likang Wu, Qi Liu, Enhong Chen, Wei Tong, and Zi Yan. 2021. Enhanced representation learning for examination papers with hierarchical document structure. In *SIGIR*, pages 2156–2160.

Rada Mihalcea and Paul Tarau. 2004. Textrank: Bringing order into text. In *EMNLP*, pages 404–411.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. *NeurIPS*, 26.

Rajdeep Mukherjee, Abhinav Bohra, Akash Banerjee, Soumya Sharma, Manjunath Hegde, Afreen Shaikh, Shivani Shrivastava, Koustuv Dasgupta, Niloy Ganguly, Saptarshi Ghosh, et al. 2022. Ectsum: A new benchmark dataset for bullet point summarization of long earnings call transcripts. In *EMNLP*, pages 10893–10906.

Clemens Nopp and Allan Hanbury. 2015. Detecting risks in the banking system by sentiment analysis. In *EMNLP*, pages 591–600.

Tim Nugent, George Gkotsis, and Jochen L Leidner. 2023. Extractive summarization of financial earnings call transcripts: Or: When grep beat bert. In *ECIR (2)*, pages 3–15.

Vishakh Padmakumar and He He. 2021. Unsupervised extractive summarization using pointwise mutual information. In *EACL*, pages 2505–2512.

Yu Qin and Yi Yang. 2019. What you say and how you say it matters: Predicting stock volatility using verbal and vocal cues. In *ACL*, pages 390–401.

Yunxin Sang and Yang Bao. 2022. Dialoguegat: A graph attention network for financial risk prediction by modeling the dialogues in earnings conference calls. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 1623–1633.

Ramit Sawhney, Arshiya Aggarwal, and Rajiv Shah. 2021. An empirical investigation of bias in the multimodal analysis of financial earnings calls. In *NAACL-HLT*, pages 3751–3757.

Christoph Kilian Theil, Samuel Broscheit, and Heiner Stuckenschmidt. 2019. Profet: Predicting the risk of firms from event transcripts. In *IJCAI*, pages 5211–5217.

Naftali Tishby and Noga Zaslavsky. 2015. Deep learning and the information bottleneck principle. In *IEEE information theory workshop*, pages 1–5.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv:2302.13971*.

Danqing Wang, Pengfei Liu, Yining Zheng, Xipeng Qiu, and Xuanjing Huang. 2020. Heterogeneous graph neural networks for extractive document summarization. *arXiv:2004.12393*.

Yi Yang, Kunpeng Zhang, and Yangyang Fan. 2022. Analyzing firm reports for volatility prediction: A knowledge-driven text-embedding approach. *INFORMS Journal on Computing*, 34(1):522–540.

Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *HLT-NAACL*, pages 1480–1489.

Zhen Ye, Yu Qin, and Wei Xu. 2020. Financial risk prediction with multi-round q&a attention network. In *IJCAI*, pages 4576–4582.

Manzil Zaheer, Guru Guruganesh, Kumar Avinava Dubey, Joshua Ainslie, Chris Alberti, Santiago Ontanon, Philip Pham, Anirudh Ravula, Qifan Wang, Li Yang, et al. 2020. Big bird: Transformers for longer sequences. In *NeurIPS*, pages 17283–17297.

Shunyu Zhang, Yaobo Liang, Ming Gong, Daxin Jiang, and Nan Duan. 2022. Multi-view document representation learning for open-domain dense retrieval. In *ACL*, pages 5990–6000.

Haiyan Zhao, Hanjie Chen, Fan Yang, Ninghao Liu, Huiqi Deng, Hengyi Cai, Shuaiqiang Wang, Dawei Yin, and Mengnan Du. 2023. Explainability for large language models: A survey. *ACM Transactions on Intelligent Systems and Technology*.

A Why not use Large Language Models

Large language models (LLMs) such as GPT-4, LLaMA, and ChatGLM, have shown an outstanding capability of understanding and generating natural language. They are pre-trained on massive amounts of multi-modal data (news articles, web pages, social media, etc), and then fine-tuned on specific tasks or domains. While these models achieve excellent results on many benchmarks, it also comes with some drawbacks.

LLMs are very complex and require a lot of time and computing resources to train. For example, GPT-3 has 175B parameters and consumes about 355 years of GPU time to train (not to mention GPT-4). The smallest version of LLaMA also has parameters of 7B (Touvron et al., 2023). The deployment of such models in specific industries is not only expensive to maintain, but also creates a huge waste of resources (only the knowledge of the corresponding industry in models is invoked). In addition, LLMs have limited explainability and controllability (Zhao et al., 2023). Explainability refers to the ability to provide human-understandable reasons or evidence for those predictions or decisions. LLMs are often seen as black boxes that produce outputs without revealing their internal logic or reasoning. Figure 6 illustrates the uncontrollable problems. This makes it hard to trust, verify, or debug them. Although LLMs are skilled in lengthy text, they may not be the best choice for the specific industry, especially in finance where accuracy, transparency (i.e., explainability) and efficiency are crucial.

In contrast, our model has several advantages in the financial field. ECT-SKIE has only 5M parameters and is easy to train. It has better controllability and explainability since its simple architecture is easier to understand. *key sentences extraction* module outputs a concise explanation M_Ω for our representation learning. As shown in Table 7-11, the explanation can be easily visualized to help researchers trace the exact sources and basis of a generated representation. In addition, the experiment results have demonstrated that the generated representations achieve state-of-the-art performance on multiple downstream tasks.

B Notations

Frequently used notations are present in Table 5.

C Proof of variational approximation

To keep formulas simple, we replace $\kappa(X)$ with X_s in the following proof.

Variational approximation of Eq. (5). We illustrate the variational upper bound for $I(X, X_s)$. We first show that $\mathrm{I}(X, X_s) \leq \mathrm{I}(X, M_s) + C$ where C is a constant and then use the lower bound for $-\mathrm{I}(X, M_s) - C$ as a lower bound for $-\mathrm{I}(X, X_s)$. First, we prove $\mathrm{I}(X, X_s) \leq \mathrm{I}(X, M_s)$ + C. From the Markov Chain $X \rightarrow (X, M_s) \rightarrow X_s$, we have $\mathrm{I}(X, X_s) \leq \mathrm{I}(X, (X, M_s))$. According to the chain rule for mutual information, $\mathrm{I}((X, M_s)) = \mathrm{I}(X, M_s) + \mathrm{I}(X, X|M_s)$, where $\mathrm{I}(X, X|M_s) = \mathrm{H}(X|M_s) + \mathrm{H}(X|M_s) - \mathrm{H}(X, X|M_s)$. Further, $\mathrm{H}(X|M_s) \leq \mathrm{H}(X)$. Putting these pieces together, we have

$$\mathrm{I}(X,\, X_s) \,\leq\, \mathrm{I}(X,\, M_s) \,+\, \mathrm{H}(X) \qquad (13)$$

where entropy $\mathrm{H}(X)$ of input is a constant. For simplicity, we denote it as C.

We then approximate $p(M_s)$ using $r(M_s)$. From the fact that Kullback Leibler divergence is always positive, we have

$$\begin{aligned}\mathbb{E}_{(X,M_s)\sim p(X,M_s)}[\log p(M_s)] &= \mathbb{E}_{M_s\sim p(M_s)}[\log p(M_s)] \\ &\geq \mathbb{E}_{M_s\sim p(M_s)}[\log r(M_s)] \\ &= \mathbb{E}_{(X,M_s)\sim p(X,M_s)}[\log r(M_s)].\end{aligned} \qquad (14)$$

From (13) and (14), we have

$$\begin{aligned}\mathrm{I}(X, X_s) &\leq \mathrm{I}(X, M_s) + C \\ &= \mathbb{E}_{(X,M_s)\sim p(X,M_s)}\left[\log \frac{p(M_s \mid X)}{p(M_s)}\right] + C \\ &\leq \mathbb{E}_{(X,M_s)\sim p(X,M_s)}\left[\log \frac{p(M_s \mid X)}{r(M_s)}\right] + C \\ &= \mathbb{E}_{X\sim p(X)} D_{\mathrm{KL}}(p(M_s \mid X), r(M_s)) + C.\end{aligned} \qquad (15)$$

Then, we have:

$$I(X, \kappa_\psi(X)) \leq \mathbb{E}\left[D_{KL}(\mathbb{P}_\psi(M_s|X), r(M_s))\right],$$

and we can minimize $I(X, \kappa_\psi(X))$ by minimizing $\mathbb{E}\left[D_{KL}(\mathbb{P}_\psi(M_s|X), r(M_s))\right]$.

Variational approximation of Eq. (10). To minimize the term $\mathbb{E}_{\mathcal{T}}\left[\mathcal{L}(g(X_s)), Y))\right]$ in Eq. (4), we turn to maximize the term $I(X_s, X_{QA})$ in Eq. (10). However, this term is also intractable because the computation of mutual information, so we maximize the mutual information between X_s and

Table 5: Frequently used notations.

Notation	Description
$\mathcal{T}$	The downstream task of representation learning
X	The input of an earnings call transcript
Y	The ground truth of input X for a specific downstream task
$\mathcal{L}$	The function of measuring the discrepancy between predicted output and ground-truth
$h_\theta(\cdot)$	The function of projecting X into a dense vector
$g(\cdot)$	The function of mapping the vector representation to the output
$\kappa(X)/X_s$	The ground-true set of key sentences in X
N	The number of sentences in X
$f_\psi(X)$	The neural network of generating mask M
$\kappa_\psi(X)$	The set of key sentences extracted by ECT-SKIE
ψ	The trainable parameters of ECT-SKIE
ψ^*	The optimal parameters of ECT-SKIE
$\odot$	The function of selecting key sentences
$\oplus$	The function of concatenating two vectors
$I(\cdot,\cdot)$	The mutual information between two variables
M	The binary mask of selecting key sentences
M_s	The mask for selecting sentences
M_Ω	The mask for selecting slots
$r(M_s)$	The prior distribution of the binary mask M
$r(M_\Omega)$	The prior distribution of the binary mask M_Ω
β	The hyper-parameter of trading off conciseness and key insights extraction performance
$\mathbb{P}_\psi$	Probability estimated by a neural network ψ
K	The number of slots in the contain
Ω	The representation of the container in ECT-SKIE
Ω_i	The representation of i-th slot in the container
X_l	The each sentence of X
S	The sentence with the highest score to a slot
S_X	The matrix of the relevance scores of K slots for N sentences of the transcript X
N_s	The number of sentences in $\kappa_\psi(X)$
S_Ω	The matrix of the importance scores of K slots for the transcript X
M_Ω	The binary submap for selecting N_s sentences from K sentences bound to the container
τ_1	The temperature hyper-parameter of Softmax operation in InfoNCE
τ_2	The temperature hyper-parameter in the Gumbel-softmax operation
$\mathcal{B}$	The batch size of input for training
X_{QA}	The representation of Q&A session in corresponding transcript X
X'_{QA}	The representation of Q&A session of other transcripts in a training batch.
$\mathbb{1}$	The indicator function which outputs 1 if the condition is true and 0 otherwise
n_i	The Gumbel noise at the i-th bit on the score vector S_Ω
α	The compression rate from X to $\kappa(X)$, which equal to N_s/N
$\mathbf{x}_i$	The vector representation of i-th sentence in X_s

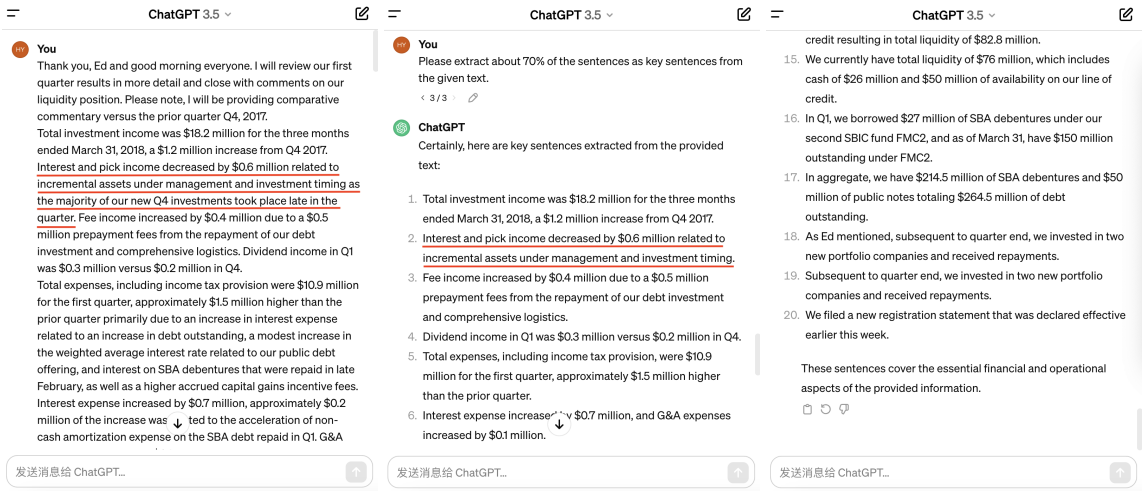

Figure 6: Extractive summarization using ChatGPT. We feed the presentation from the CFO in Fidus Investment Corporation First Quarter 2018 Earnings Conference Call, totaling 39 sentences. Then, we use the prompt "Please extract about 70% of the sentences as key sentences from the given text.", and the results return 20 key sentences (The correct number should be 27). In addition, the underlined sentence is truncated. In other tests, even the key sentences extracted are not really key (the opening greeting). In conclusion, the results have a huge deviation from the user's requirements, implying the uncontrollability and untrustworthiness of the large language models.

X_{QA} by optimizing InfoNCE, a contrastive loss in Eq. (10). This proof is as follows.

The mutual information between X_s and X_{QA} can be defined as:

$$I(X_s, X_{QA}) = \sum_{X_s, X_{QA}} p(X_s, X_{QA}) \log \frac{p(X_s|X_{QA})}{p(X_s)}.$$

Then, we model a density ratio that preserves the mutual information between X_s and X_{QA} as follows:

$$f(X_s, X_{QA}) \propto \frac{p(X_s|X_{QA})}{p(X_s)}, \tag{16}$$

where $\propto$ stands for "proportional to" (i.e. up to a multiplicative constant). Note that the density ratio f can be unnormalized (does not have to integrate to 1). Note that any positive real score can be used here. According to Eq. (11), we implement it via:

$$f(X_s, X_{QA}) = \exp(\mathrm{d}(X_s, X_{QA})). \tag{17}$$

Given a set $\mathcal{B} = \{X_1, ..., X_N\}$ of N random samples containing one positive sample from $p(X_s|X_{QA})$ and $N-1$ negative samples from the "proposal" distribution $p(X_s$, we can optimize InfoNCE:

$$\mathcal{L}_{\mathrm{NCE}} = -\mathop{\mathbb{E}}_{\mathcal{B}} \left[\log \frac{f(X_s, X_{QA})}{\sum_{X_s^j \in \mathcal{B}} f(X_s^j, X_{QA})} \right], \tag{18}$$

where we know the optimal value for $f(X_s, X_{QA})$ is given by $\frac{p(X_s|X_{QA})}{p(X_s)}$. Inserting this back into Eq. (18) and splitting $\mathcal{B}$ into the positive example and the negative examples $\mathcal{B}_{neg}$ results in:

$$\begin{aligned}
\mathcal{L}_{\mathrm{NCE}} &= -\mathop{\mathbb{E}}_{\mathcal{B}} \log \left[\frac{\frac{p(X_s|X_{QA})}{p(X_s)}}{\frac{p(X_s|X_{QA})}{p(X_s)} + \sum_{X_s^j \in \mathcal{B}_{\mathrm{neg}}} \frac{p(X_s^j|X_{QA})}{p(X_s)}} \right] \\
&= \mathop{\mathbb{E}}_{\mathcal{B}} \log \left[1 + \frac{p(X_s|X_{QA})}{p(X_s)} \sum_{X_s^j \in \mathcal{B}_{\mathrm{neg}}} \frac{p\left(X_s^j|X_{QA}\right)}{p\left(X_s\right)} \right] \\
&\approx \mathop{\mathbb{E}}_{\mathcal{B}} \log \left[1 + \frac{p(X_s|X_{QA})}{p(X_s)} (|\mathcal{B}| - 1) \mathop{\mathbb{E}}_{X_s^j} \frac{p\left(X_s^j|X_{QA}\right)}{p\left(X_s\right)} \right] \\
&= \mathop{\mathbb{E}}_{\mathcal{B}} \log \left[1 + \frac{p(X_s|X_{QA})}{p(X_s)} (|\mathcal{B}| - 1) \right] \\
&\geq \mathop{\mathbb{E}}_{\mathcal{B}} \log \left[\frac{p(X_s|X_{QA})}{p(X_s)} |\mathcal{B}| \right] \\
&= -I(X_s, X_{QA}) + \log(|\mathcal{B}|).
\end{aligned}$$

Thus, $I(X_s, X_{QA}) \geq \log(N) - \mathcal{L}_{\mathrm{NCE}}$. As $|\mathcal{B}|$ increases, the estimation using InfoNCE becomes more accurate. Therefore, it is useful to use more negative samples (use a large batch size $\mathcal{B}$). In conclusion, we can optimize our original objective $\mathbb{E}_{\mathcal{T}}\left[\mathcal{L}(g(X_s)), Y))\right]$ by minimizing InfoNCE so as to maximize the mutual information $I(X_s, X_{QA})$.

D Earnings call transcript

An earnings call is a conference call that public companies hold to discuss their earnings or share other information with investors. These calls often take place after the company releases its earnings report and are usually scheduled in advance.

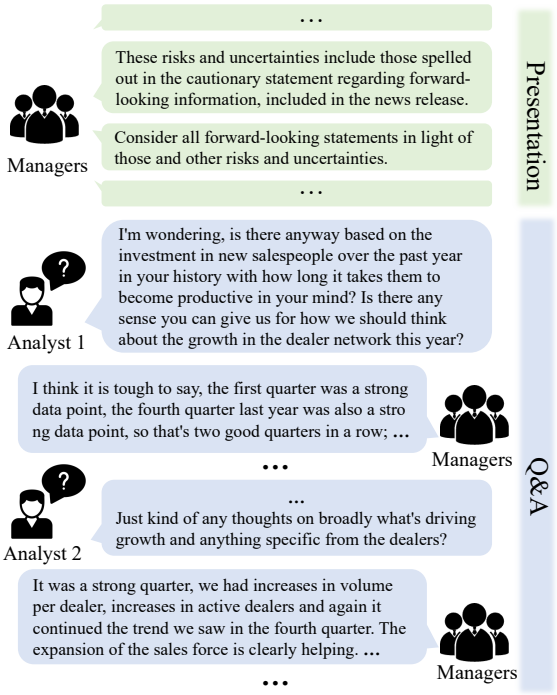

Figure 7: An excerpt of an earnings conference call. There are multiple Q&A rounds after the presentation conducted by the company's managers.

Earnings calls are an important way for companies to communicate with their shareholders and the broader investment community. They provide an opportunity for management to discuss the company's financial results, as well as its strategy and outlook. This information can be valuable for investors who are trying to make informed decisions about whether to buy, hold or sell a company's stock.

During an earnings call, the company's management team typically presents an overview of the company's financial performance for the most recent quarter or fiscal year. This may include information about revenue, expenses, profits, and other key financial metrics. The management team may also discuss any significant events or developments that occurred during the period, such as new product launches, acquisitions, or changes in the competitive landscape. After the presentation, there is usually a question-and-answer session during which analysts and investors can ask questions of the management team. This can provide additional insights into the company's performance and future plans. An example of the earnings call is shown in Figure 7.

Earnings calls are typically webcast live over the internet and are also recorded and made available for replay on the company's website. Transcripts of earnings calls are also often published by financial news outlets and can be a valuable resource for investors who want to learn more about a company's performance and strategy. They provide valuable information about a company's financial performance and future plans, which can help investors make informed decisions about whether to buy, hold or sell a company's stock, including much financial risk information.

In our work, we collect a large-scale earnings call transcripts dataset of U.S. firms and format them into CSV format files. All transcripts are recorded in English. Figure 8 presents the specific format. Then, we extract the transcripts in four fiscal years (2015-2018) for our training and testing and present the statistics of them in Table 1. Researchers can obtain the earnings conference call data from Seekingalpha or databases such as Thomson Reuters StreetEvents for research purposes. In addition, we are looking at the anonymization of private data before the data is released (e.g., using random serial numbers to replace private names).

E Experimental settings

E.1 Baselines

We provide detailed instructions about baselines as follows.

- **BERT** (Devlin et al., 2019): Previous studies have used the pre-trained BERT model to encode the whole earnings call transcript for risk forecasting. However, BERT has a token limit, so we split each transcript into sentences and feed them into the BERT model separately. For the risk forecasting task, we use average pooling to obtain the document representation of the entire transcript. Then, a support vector regression model is trained to predict the risk with these representations and the corresponding risk labels as inputs.

- **SimCSE** (Gao et al., 2021): This work presents a simple contrastive learning framework that significantly improves the state-of-the-art sentence embeddings. It uses an unsupervised approach that takes a sentence as input and predicts itself in a contrastive objective, with only standard dropout as noise. They also demonstrate that the contrastive learning objective makes the pre-trained embeddings' anisotropic space more uniform and

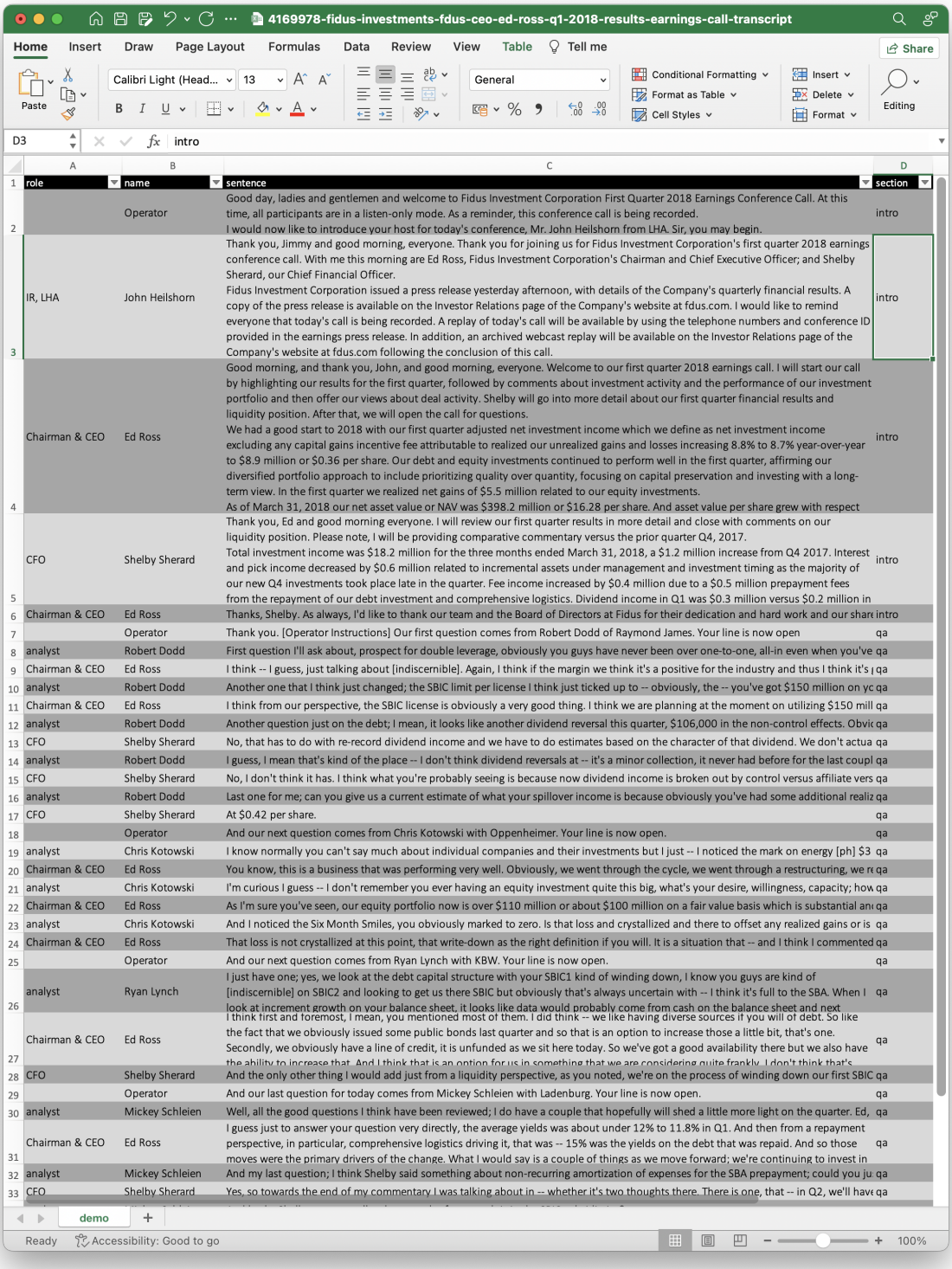

Figure 8: A screenshot of one earnings call transcript file.

aligns positive pairs better when supervised signals are available. We use it to encode the sentences into embeddings by the public model on HuggingFace[4].

- **Profet** (Theil et al., 2019): This work splits an earnings call transcript into three sections: the presentation section, the question section, and the answer section. Then, it extracts features for each section using BiLSTM and attention mechanism to forecast the financial risk. We use the pre-trained model BERT to encode the text and concatenate the representations of three sections for *Risk forecasting*. Although this method integrates financial features with the obtained text features and utilizes these two types of features to predict volatility, we only use some of them for text feature extraction.

- **MR-QA** (Ye et al., 2020): This work proposes a multi-round Q&A attention network, which is the first to decompose the long transcript at a fine-grained Q&A level. MR-QA extracts features of each round of conversation using a bidirectional attention mechanism and predicts the risk label. However, this method neglects the structural information between Q&A and presentation sections. Moreover, this method is still an end-to-end risk prediction model with only the risk loss as the objective. The language encoder is the same as Profet (In the original work, GloVe was used for implementation.)

- **DialogueGAT** (Sang and Bao, 2022): They model the speakers' information and their utterances in dialogues in earnings calls by a graph attention network, and concatenate the past volatility, speakers' representation and utterance representation to forecast financial risk. As a result, due to the extremely long nature of transcripts, modeling at the utterance level will lead to quite high computational complexity, which limits the generalization of the model. We convert our dataset according to their data storage formats, and then train and test their model with their publicly released codes[5]. The results show that its forecasting error fluctuates greatly in our dataset (a relatively large variance).

- **TextRank** (Mihalcea and Tarau, 2004) is an unsupervised graph-based ranking model for text processing which can be used in order to find the most relevant sentences in text and also to find keywords. We use TextRank to summarize earnings call transcripts, and the sentences in the summary will be encoded by the pre-trained BERT. Then, the vector representation is applied to downstream tasks. We use the public implementation[6] to get key sentences, where we set the length ratio of the summary to 0.7. Then, these selected sentences are fed into pre-trained BERT. We leverage the sentence embeddings to conduct *Risk forecasting* and *Information retrieval*.

- **LexRank** (Erkan and Radev, 2004): This unsupervised summarization method computes sentence importance based on the concept of eigenvector centrality in a graph representation of sentences. A connectivity matrix based on intra-sentence cosine similarity is used as the adjacency matrix of the graph representation of sentences. A GitHub implementation[7] can be used to get the summary. We set the same selection ratio, and the next steps are basically the same as TextRank.

- **PMI** (Padmakumar and He, 2021) This work presents an unsupervised summarization method that uses GPT-2 to calculate the pointwise mutual information (PMI) between sentences. They introduce new metrics of relevance and redundancy for text summarization, which motivate our work. Here is their official implementation[8]. However, this method is very inefficient for long earnings call transcripts, as it requires constructing a probability matrix of size $N \times N$, where N is the number of sentences in a transcript. PMI takes about 10 minutes to process one transcript, while we have around 40,000 transcript samples. Thus, we test 500 samples on this baseline due to our limited computational resources and time. The selection ratio is set to 0.7, the same as that in TextRank, LexRank, and our model. In addition, the same is true for the processing method of sentence embedding.

E.2 Experimental Details

We elaborate the details of experiments and three downstream tasks as follows.

Given transcripts from four fiscal years (2015-2018), we use the data from 2015 and 2016 to train

[4]huggingface.co/princeton-nlp/sup-simcse-bert-base-uncased

[5]https://github.com/sangyx/DialogueGAT

[6]https://github.com/summanlp/textrank

[7]https://github.com/crabcamp/lexrank

[8]https://github.com/vishakhpk/mi-unsup-summ

the model, data from 2017 as the validation set, and data from 2018 as the testing set. Following a similar setup to (Ye et al., 2020; Qin and Yang, 2019), this chronicle data division can prevent look-ahead bias (Theil et al., 2019) for risk forecasting. The pre-trained language model BERT is used to encode sentences as the vector representation that serves as the input for these downstream tasks.

Risk forecasting. Risk of a public firm is commonly measured as the stock price volatility over a certain period. Formally, let $r_t = \frac{p_t}{p_{t-1}} - 1$ be the return of a stock, where p_t is the closing price of the stock on day t, the volatility between days t and $t + \tau$ is the sample standard deviation of stock returns during this period:

$$v_{[t,t+\tau]} = ln\sqrt{\frac{1}{\tau - 1}\sum_{i=0}^{\tau}(r_{t+i} - \bar{r})^2}, \quad (19)$$

where $\bar{r}$ represents the average stock return during the period. As a standard practice in finance, we take log since the distribution of log of volatility tends to follow a Bell distribution (Kogan et al., 2009). Therefore, a stock with high volatility indicates that its stock price fluctuates widely and thus is highly risky for investments, while a stock whose price stays more or less constant indicates a low risk for investments.

Now the set of key insights $\kappa_\psi(X)$ derived by ECT-SKIE can be fed into MLPs to make the final risk prediction as the following formula:

$$\hat{y}_m = \text{MLPs}(h_\theta(\kappa_\psi(X))), \quad (20)$$

where $h_\theta(\cdot)$ is the pre-trained BERT as the encoder. We use three-layer MLPs and the hidden size of each layer is set to 300, 150, and 50.

Information retrieval. We conduct an IR experiment to demonstrate that the information captured by our model retains the main features of the relevant information from transcripts even if ECT-SKIE is trained with a high compression rate. Given a list of earnings call transcripts $d_1, d_2, ..., d_n$ and the extracted sentences from $d_i, i \in \{1, 2, ..., n\}$, we consider the extracted sentences as a query q and rank these candidate transcripts based on their relevance to the query q, which can be formalized as a probability distribution over transcripts:

$$p(d_i|q) = \frac{\exp \text{sim}(\phi(q), \phi(d_i))}{\sum_j \exp \text{sim}(\phi(q), \phi(d_j))} \quad (21)$$

where the $\text{sim}(\cdot, \cdot)$ is the cosine similarity function. The vector encoder $\phi(\cdot)$ is the pre-trained BERT. Specifically, we use average pooling $\mathbb{R}^{N_s \times d} \rightarrow \mathbb{R}^d$ for the representations of key insights extracted by ECT-SKIE and consider the representation $\mathbb{R}^d$ as a query q, where d is the embedding size of BERT. Similarly, we get the representation of the corresponding transcripts and consider it as the ground-true target.

We use three metrics to measure the performance, Precision, Mean Rank (i.e., the mean position of the true transcript after ranking all transcripts based on $p(d_i|q)$), and mean reciprocal rank (MRR) (i.e., the mean of the inverse of the rank).

Firm similarity. We use the test set (transcripts in 2018) to evaluate the performance of our model on Firm similarity. Given that each firm has four quarters in each fiscal year, there are four earnings call transcripts belonging to the firm. We randomly select one from four transcripts and feed it into our model, using the output as the representation of its corresponding firm. Then, for each firm, we treat it as the focal firm and compute the cosine similarity between its transcript representation and the transcript representations of other firms. Based on the obtained cosine scores, we rank these firms in descending order and pick the top five firms as the focal firm's similar firms.

Redundancy Calculation. We can measure the redundancy of the extracted key sentences directly, using a simple calculation based on cosine similarity. For the key sentences extracted from one transcript, the redundancy score, Red, is computed as follows:

$$Red = 2(\sum_{j=1}^{N_s-1}\sum_{k=j+1}^{N_s}\cos(\mathbf{x}_j, \mathbf{x}_k))/(N_s^2 - N_s). \quad (22)$$

A lower value of Red indicates less redundancy. We plot the score distribution of the baselines, ECT-SKIE, and ECT-SKIE's variant as Figure 5(e).

F Sensitivity analysis of K

K is the crucial hyper-parameter of the container mechanism. We analyze the sensitivity of the container size and presented the results as Figure 9. We observe that the loss decreases as K increases, which is expected. As we explain in Eq. (7), each slot in the container matches with one candidate sentence, and one candidate sentence may match with multiple slots. A larger K implies that there

Table 6: Score of subjective evaluation for the quality of extracted sentences by 10 volunteers.

Volunteers	**Case**										**Score**	**Preference (%)**	
(Number)	1	2	3	4	5	6	7	8	9	10	(mean)	TextRank	ECT-SKIE
No.1	8.5	7.0	7.0	7.5	4.0	6.0	7.5	7.0	7.5	8.0	7.00	40%	60%
No.2	9.0	7.0	7.0	8.0	5.0	5.0	8.0	8.0	7.0	8.0	7.20	30%	70%
No.3	8.5	7.0	3.0	8.0	6.5	7.5	8.0	8.0	8.5	8.0	7.30	30%	70%
No.4	6.5	5.0	8.0	3.0	4.0	6.0	8.0	6.0	6.0	8.0	6.05	60%	40%
No.5	8.0	6.0	7.0	7.0	5.0	7.0	8.0	8.0	5.0	6.0	6.70	60%	40%
No.6	7.0	6.5	7.0	8.0	5.0	7.0	8.0	7.0	4.0	5.5	6.50	40%	60%
No.7	6.5	7.0	7.0	8.0	6.0	7.5	7.0	8.0	7.0	6.0	7.00	20%	80%
No.8	8.0	6.0	6.5	6.0	4.0	6.0	8.0	8.0	8.0	8.0	6.85	40%	60%
No.9	4.0	5.0	8.0	4.5	6.0	9.0	3.0	8.5	6.0	7.0	6.10	50%	50%
No.10	7.0	4.0	6.0	8.0	6.5	5.0	8.0	7.0	7.0	6.0	6.45	20%	80%

are more slots in the container, and the chance that one candidate sentence will match with multiple slots is lower. Therefore, the number of extracted key insights is higher, which helps to generate a more comprehensive representation of a transcript.

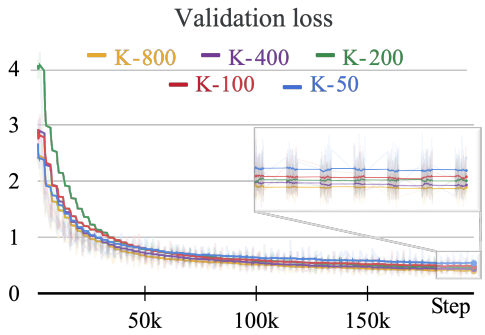

Figure 9: Loss curves of different container size K on the validation set ($\alpha = 0.15$).

G Subjective evaluation

Benefiting from the explainability of ECT-SKIE, we can visualize the results of key sentences extraction as Table 7-11 and conduct two experiments of subjective evaluation, including an evaluation of the quality of the sentences extracted by ECT-SKIE and a subjective quality comparison between ECT-SKIE and TextRank. Specifically, 10 volunteers are recruited who have extensive knowledge of finance. We create two questionnaires in the form of a web page and visualize the binary mask M_Ω to facilitate their evaluation. Volunteers are asked to rate extracted sentences based on their informativeness and relevance to financial risk. Before beginning, the volunteers are not informed about which method the masks belong to, ensuring fairness. Figure 10 shows the questionnaire for subjective evaluation where the third column is the binary mask. In addition, as illustrated in Figure 11, we release the web questionnaire to allow volunteers to choose more appropriate extracted key insights. In Table 6, the preference ratios show that our model outperforms TextRank, and these scores given by volunteers are generally between 6 and 7, implying that our advantage may not be significant yet. Moreover, we present some cases of mask visualization from pieces of transcripts as follows. We set the compression ratio α to 0.4 for a better review. In these tables, the second column is the binary mask inferred by ECT-SKIE, where 1 means the corresponding sentence is regarded as a key insight of the earnings call and 0 means the opposite. We paid $10 to participants hourly and totally spent about $300 on participant compensation.

H Usage of AI Writing Assistance

This paper was written with linguistic support from the AI assistant ChatGPT, only paraphrasing and polishing part of the original content included. No other assistance was received.

Figure 10: Questionnaire for subjective evaluation. Volunteers evaluate the quality of extraction by dragging the score bar, and the corresponding criteria are listed in the lower left corner.

Figure 11: Questionnaire of subjective comparison.

Table 7: Case 1 of the binary mask derived by ECT-SKIE.

Sentences	$\mathbf{M_i}$
...	...
We look forward to working closely with the initial 13 innovative members of this consortium to take an important step-forward in cancer research.	0
Second, we introduced ProBeam 360, our new single-room proton therapy system, with a 30% smaller footprint and a 25% lower volt construction cost as compared with the previous system.	0
This increases access to this technology and provides clinicians with a viable path to potential next-generation treatment, such as Flash therapy.	1
With the launch of ProBeam 360 and under Kolleen's leadership, we expect renewed growth in the proton business over the long term.	0
So a lot of exciting developments that are strengthening our leadership in radiation therapy.	0
Our second growth initiative is to extend our global footprint.	0
For context, our sales mix for the full year was approximately 50% in the Americas, 30% in EMEA and 20% in APAC.	1
Our growth was driven by strong performance across all geographies.	0
The largest growth was driven by our Asia Pacific region, which saw a number of key wins in the quarter.	1
We signed a memorandum of understanding with Genesis Care to form a strategic partnership for cancer care research and increase the access to advanced care – cancer care in Australia and Europe.	1
This resulted in orders for 10 linacs in Australia in the fourth quarter.	0
We also continue to execute well in China and Japan, with double-digit growth and market leadership.	0
In EMEA, we saw our fifth straight quarter of double-digit orders growth, driven by tender wins in Spain and Scandinavia, strong software growth and an outstanding Halcyon – and outstanding Halcyon orders of over 40 units.	1
We also saw robust performance from upgrades of over $85 million, an increase of 30% for the quarter and 44% for the full year.	1
Upgrades include linear accelerator enhancements like HyperArc and our 6-degree-of-freedom couch and were driven by EMEA by the hiring of a dedicated upgrade team to focus where the needs are greatest.	0
We continue to extend our global footprint by winning large tenders, providing integrated best-in-class products and solutions for both mature and emerging markets and by winning competitive takeouts.	0
...	...

Table 8: Case 2 of the binary mask derived by ECT-SKIE.

Sentences	$\mathbf{M_i}$
Can you give us an update on just your thoughts around use of capital between acquisition versus share repurchase versus further de-leverage?	0
And I guess as a follow-up to that when we think about M&A, can you give us some thoughts on how you see opportunities between like larger buys like HEG versus more product acquisitions like Main Street Hub and what looks more likely in the near term?	1
Thanks.	0
Hey, Lloyd, it's Ray.	0
I will take a shot at this and Scott can come over the top.	0
But our focus right now was just continuing to finish the swing on bringing HEG all the way into the fold.	0
And then obviously we have got Main Street Hub lined up for the back half of the year and when we look at M&A obviously, I think we have got the financial capacity there and we have also got the operational capacity.	1
You saw an announcement today where we are bringing Betsy Rafael into the management team to help us on the scale there.	1
So, it's going to add another quiver.	0
Operationally, we have got a business system that enables the integration.	0
Our products are APIable.	0
We have got a global tech platform that's capable of scale.	1
And with the leverage that you mentioned, we should be at a point by the end of this year where we can do either product tuck-ins or larger acquisitions like HEG.	1
As far as use of capital, that's going to be the primary strategy, but obviously, you have seen us do share repurchases.	1
We did a 7 million share repurchase last year alongside a secondary and that is another option for us.	1
...	...
And then also in terms of marketing and advertising expense, the ratio was up around 100 bps versus Q4, maybe kind of what's your expectation going forward in terms of pushing the pedal on marketing and maybe being a little less efficient in the near-term, but for growth purposes?	1
Thanks.	0
Hey, Mark, it's Ray.	0
I will start with the international and Scott will pickup the marketing question.	0
Organic growth in international bounced back as we anticipated.	1
It's in the high-teens.	0
FX was relatively neutral, very small positive impact there.	1
So we have seen exactly what we were expecting as a return to growth there, because as I mentioned to you on the last call, that business has been growing in the mid-teens.	1
So, we saw a little bit of a pickup, so happy about the progress we are making there.	1
...	...

Table 9: Case 3 of the binary mask derived by ECT-SKIE.

Sentences	$\mathbf{M_i}$
...	...
For your ECS business specifically the U.S. ECS business, are you gaining share?	1
Or are you seeing broad-based end market strength?	1
Thanks.	0
Yes, so Tiffany as you know we grew nicely in Q4 in North America and in a number of segments that was faster than market.	1
So in general, we're holding share across the various categories that we play in and in some cases we are gaining.	0
We obviously don't talk about specific suppliers, but I feel like we're doing a bit better than holding our own.	1
Yes.	0
And let me remind you the issue with the business here.	0
This was largely just a data center business, just an enterprise business and just a proprietary server business.	1
And we have been working to change that over the last couple of years with the onset of solid state storage and converting that.	0
So it's really – for us, it's been a mix issue.	0
Frankly, we're very happy with our hardware sales.	0
We'd like to get our software sales to catch up to the levels that will make a big difference in where the - frankly, the future of the business is going and where we need to be, and that's the change that we're making in it.	1
...	...
I'm hoping, maybe we can just attack this difference between what you're seeing and what you're supplier are seeing in a little bit different way.	0
Aside from your strong execution, I know there have been some supplier gains, maybe those are a little bit more in the past than more recent.	1
But is one of the differences maybe end-market exposure?	0
Or perhaps it's a matter of whether you're serving more of the International OEMs in China versus sort of local manufacturers?	1
Is that part of what's driving the difference?	0
Well, I'm not going to get into where all of our suppliers have their business.	1
I mean, you guys know where you focus when you're on calls with different suppliers.	0
I think I said before that, we think the benefit for us - unfortunately, the benefit is a negative.	0
We're not that big in cellphone devices.	0
It's not a big consumer item for us.	1
Our consumer business is sort of ho-hum in the schemes of things.	1
And while, I would love to go in there and run and take everybody's market share, that's just not something that we have done.	0
But also we're a little more insulated as a result of not having that business, when that's the business that is taking the biggest hit.	1
I would say to you, on the industrial front for us, we have gained a lot of new customers over the last couple of years.	1
And we're selling them more products which is actually helped us with industrial, because of the customer base increase.	1
...	...

Table 10: Case 4 of the binary mask derived by ECT-SKIE.

Sentences	$\mathbf{M_i}$
...	...
But I would largely say, it has been the expansion of customer base.	1
It has been expansions of products into that.	0
It has been increased design activity that has turned to production that has helped us grow.	1
While you could say, it's great right now, because you're talking to me about the consumer piece, we're relatively low, in general, on the consumer side, given how big of a market that is in Asia-Pac for a lot of our suppliers.	1
Hopefully, that helps you.	1
It does.	0
And maybe just connected to that as sort of part of the same questions, on the supplier side I referred to it a minute ago.	1
I think you'd taken some suppliers to concentrate the majority of their business, if not, all of it with Arrow.	1
I think some of those were maybe more on the order of a year ago.	1
What are you seeing with regard to your suppliers in terms of share today?	1
Are you still gaining in that way?	0
Or is any of it reverting?	0
And any trend there that we could highlight?	0
Thank you.	0
Well, really it was a couple of years ago.	0
I think the market shook out.	1
There was still some that came in I think in the first couple of quarters this year.	1
And then the rest of it was just sort of gutting it out from there back into typical market activity.	1
The big thing that has been working which we told you it would tick-up and it has is don't underestimate the amount of money and the amount of effort we put into designs and engineering for the customer base.	1
It's one of the reasons that we're seeing more customers, we're doing more designs, deeper designs, and those designs have been going into production and that all helps with the growth.	1
And a lot of its frankly new customers.	0
I think when we started we were around 125,000 or 130,000 customers and we're now hitting that 200,000-customer mark.	0
So, that to me, is a big indicator that – well not only big indicator but also an insulator for us as the market slows.	1
If we can continue selling more products to our current customer base that will help us too.	0
Thank you.	0
...	...

Table 11: Case 5 of the binary mask derived by ECT-SKIE.

Sentences	$\mathbf{M_i}$
...	...
I have actually two questions for Greg.	0
...	0
And then I think the second question I have is on the significant operating leverage that we've seen in the core search business with over 400 basis points of margin improvement from last year.	1
I mean, have you delayed any expenses into Q2?	1
Or do you think this kind of operating leverage is something that we could see in the next quarters as well?	1
Hi, Cesar, it's Greg, thank you very much for your questions.	0
On the first question, so what I could say is just again reiterate what I said about the full year expectations for adjusted EBITDA loss of being roughly on par with previous year, including all of the investments that we are making.	1
I guess what I didn't mention in the prepared remarks is, obviously, the investments we are making in self-driving, which are also quite significant, and we're making very large strides there.	1
So I would just leave you with the prepared remarks.	0
On the search question with respect to operating leverage, so, obviously, Q1 did see excellent results.	0
I think it's kind of early in the year.	0
And so at this point, we are going to stick with the guidance of flattish margins in the Search and Portal business.	0
But, obviously, we always try to balance off the opportunities for additional investments in new technologies, things like our Alice intelligent voice assistant, things like speech technologies, things like mapping and navigation against sort of a natural operating leverage in the business, and we'll look to update you on these results over the course of the year.	1
Can I just follow up with one question on the margins and, obviously, the ruble has depreciated a little bit against the U.S. dollar in the past two weeks.	0
So can you please remind us how you hedged your rent expenses, and I think that's until 2018?	0
And what's going to happen after 2018?	0
Sure so we ended up – we actually hedged our office rent expense more than a year ago, and it covers both our 2017 and 2018 lease payments through, I think, we used one counter party for that hedging transaction.	1
And going forward, there are changes in the way that leases are accounted such that the P&L impact of the FX fluctuations will be minimalized – minimized.	0
And then just to remind you, obviously, that if you look at the way that our cash balance is currently made up, it is very heavily weighted towards the USD.	0
So our sensitivity to FX changes with respect to FX fluctuation is less than it would have been otherwise.	1
Currently if you look at our FX basket, we are 77% USD or nonruble currencies.	0
...	...