Transactions on Computational and Scientific Methods | Vo. 4, No. 7, 2024
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Financial Report Summarization via Structure-Aware Modeling and Information Fidelity Constraints

Manqing Wang1*, Po-Lun Peng[2], Jianan Chen[2]
[1]Trine University, Phoenix, USA
[2]Westcliff University, Irvine, USA
*Corresponding author: mwang244@my.trine.edu

Abstract: This study addresses the challenges of complex structural hierarchies, strong cross-section dependencies, and frequent factual drift in long document summarization of financial reports. It proposes a unified framework that integrates structural constraints with information fidelity modeling. Financial reports are divided into sections or paragraph-level structural units and encoded in a structure-aware manner. Structural position representations and structure-aware attention mechanisms are introduced so that decoding performs content aggregation and organization under global ordering and boundary constraints. In addition, source consistency and structural consistency constraints are incorporated into the training objective. These constraints guide the generated content to remain traceably aligned with source evidence and reduce the risk of numerical and statement deviations. To evaluate reliability in high-risk financial text scenarios, the study constructs a comprehensive evaluation framework that includes both summary quality and fidelity metrics. Systematic comparisons and sensitivity analyses are conducted on an open financial report summarization dataset. Model stability is further examined from the perspectives of hyperparameters, environmental perturbations, and data composition. The analysis focuses on the impact of boundaries of structural position embedding dimension, long document sample proportion, numerical noise injection intensity, and learning rate on supportability and numerical consistency. Experimental results show that the proposed framework outperforms multiple baseline methods on ROUGE-L, BERTScore, NumMatch, and Supported Rate. It also maintains more robust evidence alignment under various perturbation conditions. These results verify the effectiveness and necessity of combining structural constraints with information fidelity modeling for financial report summarization.

Keywords: Structural constraints; information fidelity; financial report summary; evidence alignment

1. Introduction

With the continuous deepening of digitalization and intelligent technologies, financial reports have become the core medium for corporate information disclosure and external communication. The scale of text and the complexity of structure keep increasing[1]. Annual reports, periodic reports, and accompanying notes usually contain large volumes of highly specialized financial information with cross-sectional dependencies and strict logical constraints. The cost of reading and understanding is therefore rising. For investors, regulators, auditors, and corporate managers, accurately identifying key information, risk signals, and underlying business conditions within a limited time has become an urgent practical demand. Under this context, applying intelligent text summarization techniques to financial reports is regarded as an important approach to improve information accessibility and decision-making efficiency.

In contrast to general news or social media texts, financial report texts exhibit fundamentally different generation requirements and risk characteristics. Financial narratives rely heavily on structural constraints.

Clear logical orders and semantic dependencies exist across sections. Indicators, accounts, and descriptive statements must satisfy strict consistency relationships. In addition, numerical values, causal explanations, and policy-related expressions in financial reports are highly sensitive. Any omission, ambiguity, or factual deviation may lead to misleading interpretations. Such errors can further result in serious compliance and risk management consequences. Therefore, financial report summarization is not merely a compression task. It is a modeling problem that requires strict preservation of structure and semantic accuracy[2].

Although automatic summarization methods have achieved progress on general corpora, they still show clear limitations in financial report scenarios. On the one hand, most approaches lack explicit modeling of inherent structural constraints. This often leads to disrupted section logic or missing contextual dependencies during generation. On the other hand, existing models tend to emphasize linguistic fluency while weakening factual consistency control[3]. As a result, generated summaries may deviate from the source text in numerical values, conclusions, or causal relationships. In the financial domain, such information fidelity issues are particularly critical. Summaries are frequently used directly for decision support and risk assessment. Their reliability is therefore of much higher importance.

From an application perspective, structural constraints and information fidelity are closely related to the practical usability of generated summaries. High-quality structure-constrained summaries can preserve the original logical framework of financial reports while highlighting core operating conditions, financial changes, and risk signals. This reduces information filtering costs and improves analytical efficiency. Meanwhile, a controllable information fidelity mechanism helps mitigate semantic drift in automated processing. It also enhances user trust in intelligent system outputs. This provides a reliable foundation for intelligent auditing, investment analysis, and financial risk early warning[4].

Considering these factors, research on structure-constrained text summarization and information fidelity modeling for financial reports is of significant importance. This direction promotes the extension of summarization techniques from general language generation to high-risk and high-constraint domains. It also advances the capability of natural language processing methods in professional text understanding. Furthermore, systematically incorporating structural constraints and fidelity perspectives enables the establishment of technical paradigms that better align with real business needs. This supports the intelligent upgrading of financial information disclosure, analysis, and regulation. It also provides essential theoretical and methodological foundations for building trustworthy and controllable financial intelligent systems.

2. Related work

Existing studies in text summarization research mainly focus on extractive and abstractive approaches. Early methods emphasized selecting representative sentences from the source text. They relied on importance scoring or positional heuristics to achieve information compression[5]. These approaches show certain advantages in information fidelity. However, they often fail to produce coherent and highly abstract summaries. With the development of deep learning, abstractive summarization has gradually become dominant. End-to-end models generate natural language summaries directly. They significantly improve fluency and semantic integration. Nevertheless, these methods primarily emphasize readability and overall coverage. They provide limited modeling of implicit structural relations and logical constraints in the source text. In complex long document scenarios, this often leads to structural misalignment or omission of critical information[6].

To address the summarization of long documents and professional texts, some studies introduce hierarchical and paragraph-level modeling strategies. These methods attempt to simulate multi-level document structures to alleviate excessive context span. They improve the model's ability to perceive section-level information to some extent. However, most of them still treat document structure as an implicit feature. Explicit representation and controllable mechanisms for structural constraints are largely missing. In highly structured financial reports, different sections carry clear and irreplaceable functional roles. Structural

dependencies affect not only information completeness but also the correctness of semantic interpretation. As a result, existing long document summarization methods remain insufficient for meeting the strict requirements of structural consistency and logical rigor in financial text scenarios.

Research related to information fidelity mainly explores factual consistency detection, generation constraints, and post-generation verification. One line of work evaluates semantic or factual alignment between generated text and the source to identify potential inconsistencies. Another line introduces constraint signals during generation to reduce the probability of hallucinated content. These studies have shown effectiveness in general text generation tasks. However, challenges remain in financial report summarization. Information fidelity in financial texts goes beyond factual consistency. It also involves numerical accuracy, causal relations, conditional statements, and cross-sectional references. Simple fact matching or consistency scoring is therefore insufficient to fully capture real risks[7].

In recent years, natural language processing research on financial and accounting texts has increased. It covers information extraction, risk identification, text classification, and question answering. However, systematic studies on financial report summarization remain limited. Existing work often focuses on readability or information coverage. A unified modeling perspective that jointly considers structural constraints and information fidelity during generation is still lacking. Overall, current methods do not fully integrate the structural characteristics and high-risk nature of financial texts. This gap highlights clear research opportunities and theoretical extensions for structure-constrained text summarization and information fidelity modeling in financial reports.

3. Proposed Framework

This paper addresses the task of financial report text summarization, constructing a unified generative framework that simultaneously and explicitly models document structure constraints and information fidelity. Let the original financial report be represented as a sequence of multiple structural units:

$$D = \{S_1, S_2, \ldots, S_N\}$$

Each S_i corresponds to a chapter or a semantically complete paragraph unit. First, each structural unit is encoded to obtain its corresponding structure-aware representation:

$$h_i = f_{enc}(S_i)$$

And chapter order and hierarchy information are introduced through a structural position mapping function:

$$\tilde{h}_i = h_i + p_i$$

Where p_i is the structural position embedding, which is used to explicitly constrain the model's perception of the document's logical order and structural boundaries, thereby avoiding disruption of the original report's overall organizational form during the generation process. This paper also presents the overall model architecture, as shown in Figure 1.

In the summary generation stage, the model performs conditional decoding based on a structure-aware representation to generate a summary sequence:

$$Y = \{y_1, \ldots, y_n\}$$

The conditional probability of each generated word is defined as:

$$P(y_t / y_{<t} / D) = Softmax(g(s_t))$$

Decoding state s_t depends on both historically generated content and the global structural context:

$$s_t = f_{dec}(y_{t-1}, \sum_{i=1}^{N} \alpha_i, \tilde{h}_i)$$

The structural attention weight α_i is calculated using a normalization mechanism:

$$\alpha_i = \frac{exp(s_{t-1}^T h_i)}{\sum_{j=1}^{N} exp(s_{t-1}^T \tilde{h}_j)}$$

This ensures that information references to different structural units remain globally consistent and subject to order constraints during the abstract generation process.

To enhance information fidelity, this paper introduces explicit source text consistency modeling during the generation process. Let the overall representation of the abstract in the latent space be:

$$z_y = \frac{1}{T} \sum_{t=1}^{T} s_t$$

The overall aggregation of the original text structure is as follows:

$$z_D = \frac{1}{T} \sum_{t=1}^{N} h_i$$

Constraining the semantic consistency between the generated summary and the original text by minimizing the distance between them:

$$L_{cons} = //z_y - z_D//_2$$

This constraint, at the global level, inhibits the generated content from deviating from the original financial semantic distribution, which helps reduce the risk of factual deviation and information omission.

Regarding the overall optimization objective, the model jointly models the language generation objective with structural constraints and information fidelity constraints. The basic generation loss is defined as:

$$L_{gen} = -\sum_{t=1}^{T} logP(y_t / y_{<t}, D)$$

And introduce structural consistency regularization terms.

$$L_{struct} = \sum_{t=1}^{T} \sum_{i=1}^{N} \alpha_i \bullet \Delta(i)$$

Where A represents the penalty weight for structural units deviating from the expected logical order. The final optimization objective is expressed as:

$$L = L_{gen} + \lambda_1 L_{cons} + \lambda_2 L_{struct}$$

By using multi-objective joint constraints, the model is guided to simultaneously meet the requirements of language readability, structural consistency, and information fidelity when generating summaries, thereby better adapting to the high-constraint, high-risk text generation scenario of financial reports.

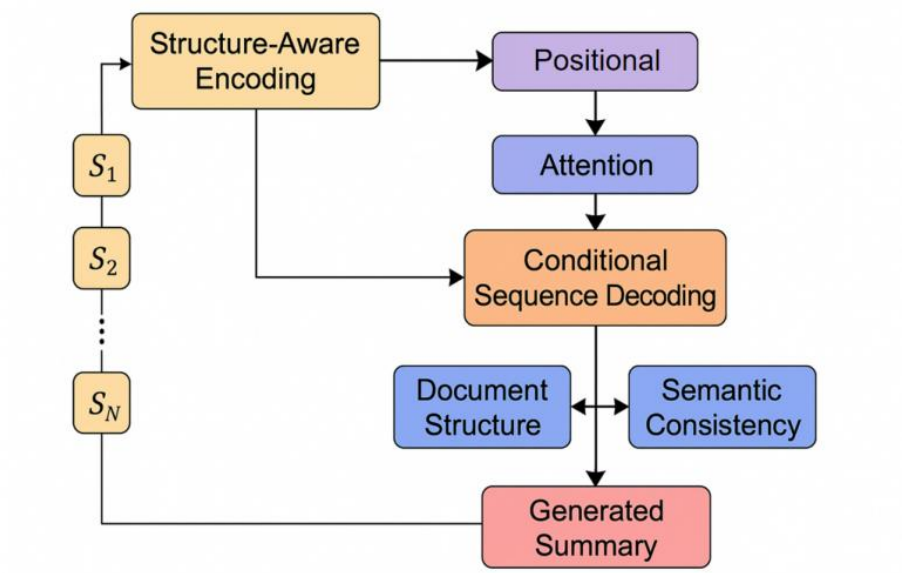

Figure 1. Overall model architecture

4. Experimental Analysis

4.1 Dataset

This study uses the open-source FINDSum dataset for financial report summarization as the primary data resource. The dataset targets realistic scenarios that combine long financial narratives with multiple tables. The samples are collected from corporate annual reports. They naturally exhibit section-based narratives, cross-paragraph dependencies, and extensive numerical tables with financial indicator descriptions. These characteristics make FINDSum suitable for studying high-risk generation tasks such as structure-constrained summarization and information fidelity control.

In terms of data scale and coverage, FINDSum is built on 21,125 annual reports from 3,794 companies. The dataset is divided into two subsets that correspond to the most critical and error-prone narrative regions in financial reports. These include results of operations and liquidity-related narratives. Each sample contains long report text together with multiple tables. A corresponding target summary is also provided. This forms a standard supervised summarization setting with structured inputs and summary outputs.

From the perspective of task alignment, the section organization and multi-source inputs in FINDSum make it well-suited for structure unit-level modeling. Reports can be segmented into sequences of sections, paragraphs, or subsections. Tables can be treated as structural evidence associated with adjacent narratives. This design supports modeling of cross-unit consistency, numerical description alignment, and factual fidelity in generated summaries. In addition, FINDSum is released under an open license and is publicly accessible. This satisfies requirements for reproducibility and public verification in academic research.

4.2 Experimental Results

This article first presents the results of the comparative experiments, as shown in Table 1.

A global comparison shows that the proposed method achieves the best performance across all four metrics. The advantage is not limited to ROUGE-L and BERTScore, which mainly reflect readability and semantic similarity. It is more pronounced on NumMatch and Supported Rate, which directly measure information fidelity. For financial report summarization, evaluation is not only about stylistic similarity. The critical issue is whether numbers, entities, and conclusions remain faithful to the original evidence. The simultaneous improvement on fidelity-oriented metrics therefore provides stronger support for the claim of effective information fidelity modeling.

In terms of summary quality, the ROUGE-L score of Ours reaches 34.1. This clearly exceeds the strongest baseline Hyfit at 32.2. The BERTScore also increases from 0.882 to 0.896. These results indicate that introducing structural constraints does not sacrifice overall abstraction capability. Instead, it improves the aggregation and reexpression of core semantics in long documents. Financial reports are typically lengthy and hierarchically structured, with very high information density. When a model can explicitly perceive structural units and aggregate them in an ordered manner, it is more likely to preserve key operating conclusions and important disclosures. This advantage is then reflected in ROUGE and semantic similarity metrics.

Table 1. Comparative experimental results

Method	ROUGE-L	BERTScore	NumMatch	Supported Rate
ROUGE-SEM[8]	27.8	0.851	0.66	0.71
FineSurE[9]	29.4	0.862	0.70	0.76
V2xum-llm[10]	30.1	0.868	0.72	0.78
UniSumEva[11]	28.6	0.857	0.69	0.74
Tofueval[12]	31.0	0.874	0.75	0.82
Hyfit[13]	32.2	0.882	0.78	0.83
Ours	**34.1**	**0.896**	**0.84**	**0.92**

Improvements in information fidelity are more interpretable. Ours achieves a NumMatch of 0.84, which is 0.06 higher than Hyfit at 0.78. The Supported Rate reaches 0.92, which is 0.09 higher than Hyfit at 0.83. These gains are substantially larger than those observed for summary quality metrics. This suggests that the proposed method more effectively suppresses the most common and critical error sources in financial summaries. These include numerical misstatements, unit or definition mismatches, and generalized claims without explicit textual evidence. In financial applications, this shift from readability gains to verifiable consistency is crucial. It indicates that the summaries are more suitable for high-risk use cases such as investment analysis, auditing, and compliance.

A closer comparison with baseline methods further highlights this pattern. Some approaches improve ROUGE-L or BERTScore, but show limited gains in NumMatch and Supported Rate. This implies that stronger language generation or semantic matching alone is insufficient to resolve factual consistency issues in financial summarization. The proposed method leads on all four metrics. The margin is especially clear for numerical matching and evidence support. This suggests that structural constraints and consistency modeling impose effective boundary control during generation. The model is guided to organize summaries around traceable evidence rather than relying on linguistic priors for abstract rewriting. This outcome aligns well with the central theme of the study. It also provides empirical support for emphasizing trustworthy generation and controllable outputs.

The structural position embedding dimension directly affects the model's ability to encode the chapter order, hierarchical boundaries, and cross-paragraph dependencies in financial reports. Too small an embedding dimension may limit the expressive capacity of structural signals, while too large an embedding dimension may introduce redundancy and increase training instability. To characterize the boundary of this hyperparameter's effect on summarization quality and information fidelity, a single-index sensitivity evaluation needs to be performed while keeping other configurations fixed. The experimental results are shown in Figure 2.

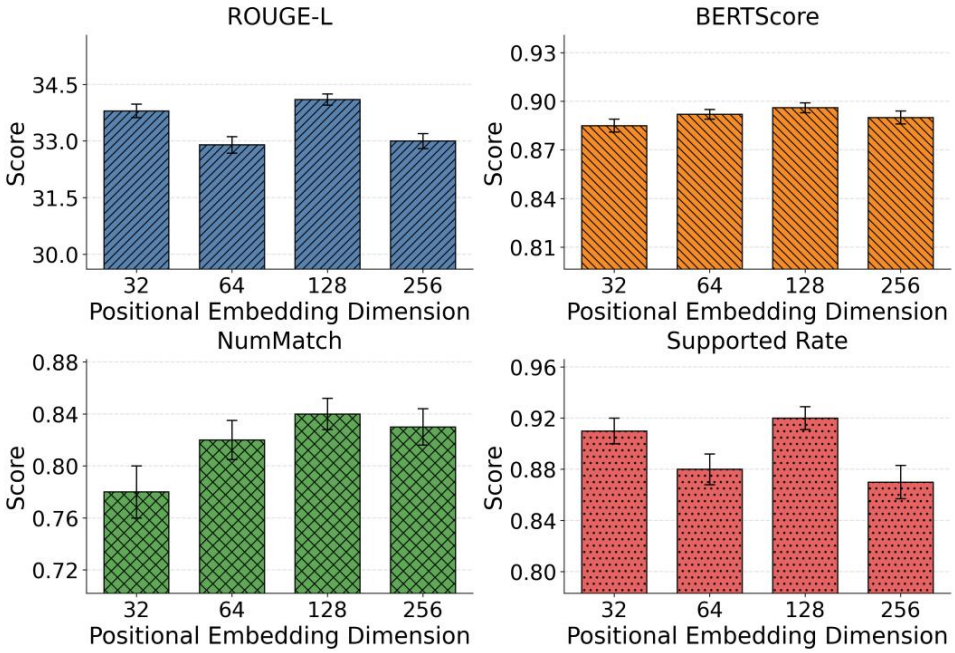

Figure 2. The impact of structural position embedding dimension on experimental results

The figure illustrates the impact of structural position embedding dimension on summary quality and information fidelity. Four subplots report the same model under different dimensions using ROUGE-L, BERTScore, NumMatch, and Supported Rate. Overall, a larger dimension is not always better. There exists an appropriate capacity range to encode structural signals such as section hierarchy and ordering boundaries in financial reports. For structure-constrained summarization, this hyperparameter acts like a bandwidth for structural information. Insufficient bandwidth weakens structural control. Excessive bandwidth introduces redundancy and amplifies noise. This degrades generation stability and verifiability.

At the level of summary quality, ROUGE-L increases clearly as the dimension grows from 32 to 128. It then declines at 256. This indicates that a moderate dimension better supports organizing key information and forming a document-level structure closer to the reference summary. Changes in BERTScore are smaller. The curve shows a mild increase followed by stabilization. This suggests that semantic similarity is less sensitive to dimensional changes than ROUGE-L. This matches the nature of financial report summaries. The semantic gist is easier to preserve. Correct placement and coherent coverage under structural constraints depend more on structural encoding capacity.

At the level of information fidelity, the trends of NumMatch and Supported Rate provide stronger evidence for the study focus. NumMatch increases gradually from 32 to 128 and then decreases slightly at 256. This indicates that richer structural signals guide the model to extract and restate key numbers from correct structural units. Numerical mismatches and definition drift are reduced. Supported Rate peaks at 128 and is clearly lower at 64 and 256. This shows that structural constraints affect not only expression form but also evidence alignment reliability. An appropriate dimension enables the decoder to rely more consistently on traceable evidence. This increases the proportion of supported statements.

Considering all four metrics together, 128 dimensions provide a better balance between quality and fidelity. This setting yields higher ROUGE-L, competitive BERTScore, and the highest NumMatch and Supported Rate. A dimension of 32 underrepresents structural information. Structural control is insufficient. Coverage and fidelity suffer. A dimension of 256 likely overrepresents structure or introduces redundancy. Attention becomes more dispersed. The evidence focus weakens. Numerical matching and support decline. For financial report summarization, these results emphasize the need for proper capacity allocation in structural

constraint modeling. With an appropriate setting, information fidelity can be enforced as a hard constraint without sacrificing readability.

The proportion of long document samples alters the cross-paragraph dependencies and structural boundary complexity encountered by the model during training, thus affecting the learning difficulty and generalization performance of structurally constrained summarization. For long, strongly structured texts such as financial reports, changes in the proportion of long documents directly impact the model's ability to characterize chapter-level information organization, cross-chapter consistency, and evidence alignment. To isolate the influence of this data factor, it is necessary to keep the rest of the training configuration constant, adjust only the proportion of long document samples, and observe the sensitivity response of each indicator. The experimental results are shown in Figure 3.

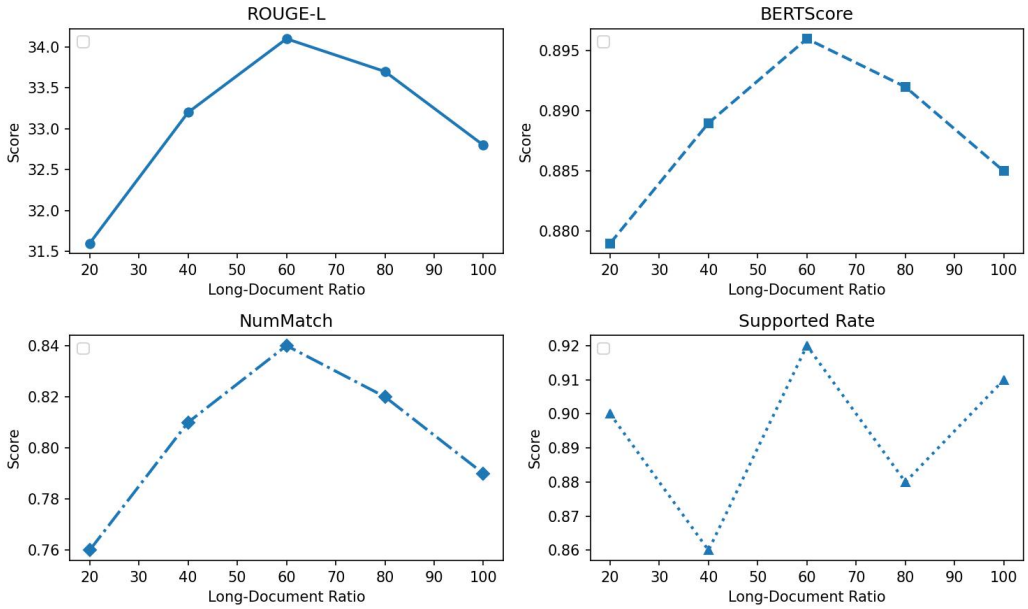

Figure 3. The impact of the proportion of long document samples on experimental results

This figure depicts the single-factor response of the model to changes in the proportion of long document samples in the training set. The responses are measured on summary quality and information fidelity metrics. Financial reports are inherently long and strongly structured texts. Adjusting the long document ratio effectively changes the density of cross-paragraph dependencies and the complexity of structural boundaries exposed to the model. The impact, therefore, appears across content coverage, semantic proximity, and evidence alignment. All other training settings are kept fixed. Only the long document proportion is varied. As a result, the observed curves can be interpreted as effects of data distribution structure rather than confounding factors from hyperparameters or training instability.

Regarding summary quality, ROUGE-L increases clearly as the long document ratio rises from 20 percent to 60 percent. It then shows a slight decline at higher ratios. This forms a typical inverted U-shaped pattern. The result indicates that a moderate increase in long document samples helps the model learn a more complete document-level organization. Coverage of key paragraphs becomes more sufficient. This is particularly suitable for cross-sectional information aggregation in financial reports. When the proportion becomes too high, training difficulty increases. The model may adopt more conservative extraction or focus on local content. Global coverage and structured compression weaken slightly. This is reflected as a decline in ROUGE-L.

The variation of BERTScore is more moderate. The curve shows a mild increase followed by a small decrease. This suggests that semantic similarity is less sensitive to changes in long document proportion than ROUGE-L. This aligns with the characteristics of financial summarization. Under different data

compositions, the model can usually preserve similar semantic meaning. The more challenging part is optimal information selection and organization under a long context. This mainly affects structure and coverage-oriented metrics. The observation also implies that semantic similarity alone is insufficient to assess summary reliability. Fidelity metrics are still required to verify whether generated content is supported by source evidence.

At the level of information fidelity, NumMatch also follows an increase then decrease pattern. Higher values appear around a moderately long document ratio. This indicates that appropriate exposure to long documents strengthens the integration of numerical context. Numerical restatements become better aligned with the source. Supported Rate shows a non-monotonic trend. A clear drop appears around 40 percent. Recovery is observed at 60 percent and 100 percent. This suggests that evidence support is influenced not only by document length. It is also affected by structural diversity and the difficulty of aggregating cross-paragraph evidence. For the theme of this study, the results highlight that structure-constrained and fidelity-oriented modeling is highly sensitive to data distribution. A reasonable long document proportion can maintain a more stable balance between readability and verifiable consistency. It also suggests that future financial report settings may benefit from finer-grained structural sampling or difficulty stratification to achieve more stable fidelity outcomes.

Numerical noise injection can perturb sensitive elements such as key amounts and proportions in financial texts, thereby testing the robustness boundaries of the model in terms of evidence alignment and statement supportability. The experimental results are shown in Figure 4.

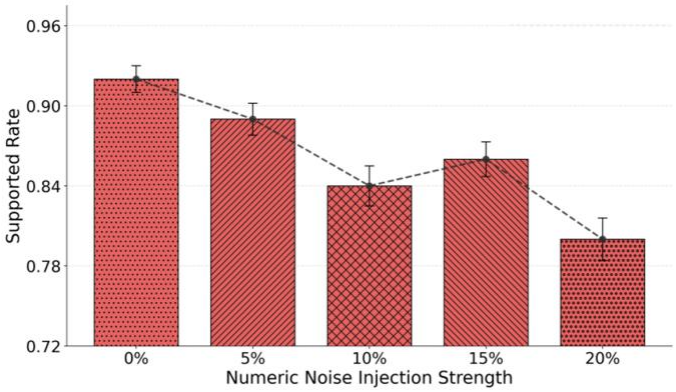

Figure 4. Environmental sensitivity analysis of numerical noise injection intensity on the supported rate

This figure shows the response curve of the Supported Rate under different levels of numerical noise injection. It reflects the robustness of evidence alignment in financial text summarization when numerical values are perturbed. In financial report summaries, the main sources of risk often lie in the restatement and interpretation of amounts, ratios, and period-over-period changes. Even small numerical deviations can determine whether a statement remains supported by the source text. In this experiment, the Supported Rate therefore serves as a direct proxy for information fidelity and verifiability. It measures whether structure-constrained summaries can preserve traceable evidence chains under numerical uncertainty.

From the overall trend, the Supported Rate decreases clearly as noise intensity increases from 0 percent to 10 percent. This indicates strong vulnerability to perturbations in numerically sensitive fields. Numerical noise disrupts the binding between values and their contextual explanations in the source. The model then struggles to select the correct evidence segments during generation. Some summary sentences lose explicit support as a result. For this task, the observation confirms that financial summarization is not only a language generation problem. It is also an evidence selection and fact verification problem. Once noise is introduced, evidence consistency degrades rapidly.

A partial recovery appears at 15 percent, followed by another decline at 20 percent. This non-monotonic pattern shows that model degradation is not linear with respect to noise strength. A plausible explanation is that at moderate noise levels, the model adopts more conservative expressions. It reduces reliance on precise numbers. This temporarily buffers the supported rate. When noise increases further, even conservative strategies fail. Mismatch between evidence and statements becomes more frequent. Supported Rate then drops again. For financial report summarization, this suggests that the model may adjust its generation strategy under numerical uncertainty. However, such an adjustment does not guarantee reliable evidence support at high noise levels.

In relation to the study focus, these results emphasize that information fidelity modeling must explicitly address evidence consistency and robustness under numerical perturbations. Otherwise, even strong structural constraints can be undermined by numerical noise. In financial applications, users care more about traceability and verifiability than surface fluency. The sensitivity of the Supported Rate, therefore, reveals a critical boundary. The model's ability to align numerical fields largely determines its trustworthiness in high-risk scenarios. Future work can enhance robustness through explicit alignment of numerical entities, normalization of units and definitions, and pre-generation evidence verification. This would improve stable and supported outputs under noisy conditions.

Finally, this paper presents the impact of the learning rate on the experimental results, which are shown in Table 2.

Table 2. The impact of the learning rate on experimental results

Learning Rate	ROUGE-L	BERTScore	NumMatch	Supported Rate
5e-4	31.9	0.880	0.76	0.86
4e-4	32.8	0.886	0.79	0.88
3e-4	33.4	0.890	0.81	0.89
2e-4	33.8	0.893	0.83	0.91
1e-4	34.1	0.896	0.84	0.92

The learning rate sensitivity results indicate that as the learning rate is gradually reduced from 5e-4 to 1e-4, all four metrics show a consistent upward trend. This demonstrates that this hyperparameter has a direct impact on convergence quality and generation stability. For structure-constrained summarization of financial reports, the learning rate affects more than optimization speed. It also influences the balance between structural signals and fidelity constraints during training, as well as gradient stability. The consistent improvement can therefore be interpreted as the model learning section level organization and evidence alignment more effectively under smoother parameter updates.

From the perspective of summary quality, ROUGE-L increases from 31.9 to 34.1, and BERTScore rises from 0.880 to 0.896. This indicates that a smaller learning rate is more favorable for improving coverage and semantic proximity in long documents. Financial report texts have high information density and strong cross-paragraph dependencies. Structure-constrained modeling requires the model to gradually form stable representations of structural boundaries and global key points. When the learning rate is too large, parameter updates become overly aggressive. Attention related to structural information may fluctuate. This weakens the aggregation and coherent expression of key sections in the generated summary.

In terms of information fidelity, NumMatch improves from 0.76 to 0.84, and Supported Rate increases from 0.86 to 0.92. These gains are highly task-relevant. The main risk in financial summarization lies in whether numerical restatements and conclusions are supported by source evidence. A smaller learning rate yields a

more stable optimization process. This allows the model to learn precise bindings between numerical values and their contexts. It also reduces the tendency to rely on linguistic priors for generalized rewriting. The continuous increase in the Supported Rate indicates that generated statements are more robust to evidence verification. This aligns to produce verifiable outputs under information fidelity modeling.

Overall, a learning rate of 1e-4 achieves the best performance across all four metrics. This suggests that a more conservative learning rate setting can balance readability and reliability. The benefit is especially clear for fidelity-related metrics. This conclusion directly supports the theme of the study. Structural constraints and consistency losses are sensitive to training stability. Only with sufficiently smooth gradient updates can the model develop robust boundaries for structural and evidence alignment. In deployment and transfer scenarios, the learning rate can therefore be treated as a key control factor for trustworthy generation. It can help maintain stable Supported Rate outputs under varying data noise or distribution shifts.

5. Conclusion

This work focuses on financial reports as long documents with a strong structure and a high risk of information disclosure. It systematically studies structure-constrained text summarization and information fidelity modeling. Section hierarchy, ordering boundaries, and cross-paragraph dependencies are explicitly incorporated into the generation process. Verifiable consistency signals are also integrated into the training objective. The study argues that summary quality should not be judged only by linguistic similarity. Evidence support and numerical consistency should serve as core constraints. In this way, clearer trust boundaries are established for abstractive summarization. From task formulation, modeling framework, and evaluation criteria, this study shifts financial summarization from general readability-oriented objectives toward reliability-oriented goals suitable for compliance and auditing contexts. It provides a technical path that better aligns with real-world high-risk applications.

From a practical perspective, summarization methods that jointly emphasize structural constraints and information fidelity can substantially reduce reading and information filtering costs for financial reports. They also improve accessibility and interpretability of key disclosures. In applications such as investment research, corporate risk management, internal control, and regulatory analysis, such methods help transform massive disclosure texts into more actionable decision inputs. This is especially valuable in scenarios with multiple entities, multiple reporting periods, and dispersed cross-sectional evidence. Core signals such as operating changes, cash flow pressure, risk warnings, and accounting policy adjustments can be extracted more consistently. The introduction of fidelity metrics such as support rate and numerical matching further enables audit-style inspection and traceable verification. This reduces the risk of misuse or mistrust of automated summaries in high-risk business settings. As a result, intelligent financial analysis tools become more deployable in real operational pipelines.

The findings also reveal several critical boundary conditions for financial summarization. Factors such as the proportion of long documents, numerical perturbations, and training hyperparameters have significant effects on the stability of evidence alignment. This indicates that robustness should be treated as an objective as important as accuracy when building trustworthy generation systems for finance. Structural constraints improve information organization and section-level consistency. However, without explicit constraints and validation for numerically sensitive elements, summaries may still lose support under high noise or complex contexts. For practical deployment, model outputs should therefore be integrated with evidence chain management, traceable referencing, and risk grading strategies. When uncertainty increases, the system should narrow its output scope or trigger stricter validation procedures. This is necessary to meet audit and compliance requirements for verifiability and responsibility boundaries.

These boundary conditions are also consistent with recent studies on financial intelligence and trustworthy system execution. Explainable cross-market causal discovery with financial knowledge graphs shows that relation-aware causal structures can make complex market signals more interpretable for downstream

analysis [14]. Cross-period financial statement anomaly detection further indicates that contrastive representation learning can capture subtle temporal deviations in accounting disclosures [15]. Microstructure-aware risk detection in high-frequency trading highlights the value of embedding and sequence modeling for recognizing high-risk financial behavior under dense temporal observations [16]. Trustworthy LLM-agent-oriented ETL orchestration also supports the need for neural execution verification when financial analysis outputs depend on structured data pipelines and automated reasoning procedures [17]. Together, these perspectives suggest that future financial summarization systems should integrate structural generation with causal evidence, temporal risk modeling, and verifiable execution control.

Looking ahead, future research can further extend structural constraints from the section level to finer-grained disclosure templates and financial semantic units. This would strengthen cross-section information fusion and consistency across reporting standards. It would also support more specialized summary forms, such as liquidity-focused, operational risk-focused, or accounting policy-oriented summaries and comparative summaries. In addition, information fidelity modeling can move from passive evaluation to active control. Stronger evidence retrieval and verification components, numerical and unit normalization mechanisms, and uncertainty-aware adaptive generation strategies can be introduced. These improvements would help maintain stable, supported outputs across different noise conditions and disclosure styles. As generative technologies continue to penetrate financial, auditing, and regulatory domains, the joint modeling of structural constraints and information fidelity will remain a fundamental pillar for trustworthy financial intelligence. It is expected to further enhance efficiency and quality in intelligent investment analysis, automated compliance review, and corporate risk early warning.

References

[1] Zhang, Haopeng, Philip S. Yu, and Jiawei Zhang. "A systematic survey of text summarization: From statistical methods to large language models." arXiv preprint arXiv:2406.11289 (2024).

[2] Shakil, Hassan, Ahmad Farooq, and Jugal Kalita. "Abstractive text summarization: State of the art, challenges, and improvements." Neurocomputing 603 (2024): 128255.

[3] Zhang, Yang, et al. "A comprehensive survey on process-oriented automatic text summarization with exploration of llm-based methods." arXiv preprint arXiv:2403.02901 (2024).

[4] Wibawa, Aji Prasetya, and Fachrul Kurniawan. "A survey of text summarization: Techniques, evaluation and challenges." Natural Language Processing Journal 7 (2024): 100070.

[5] Raza, Asif, et al. "Abstractive text summarization for Urdu language." Journal of Computing & Biomedical Informatics 7.02 (2024).

[6] Van Veen, Dave, et al. "Adapted large language models can outperform medical experts in clinical text summarization." Nature medicine 30.4 (2024): 1134-1142.

[7] Liu, Wenjun, et al. "Automatic text summarization method based on improved TextRank algorithm and K-means clustering." Knowledge-Based Systems 287 (2024): 111447.

[8] Zhang, Ming, et al. "ROUGE-SEM: Better evaluation of summarization using ROUGE combined with semantics." Expert Systems with Applications 237 (2024): 121364.

[9] Song, Hwanjun, et al. "FineSurE: Fine-grained summarization evaluation using LLMs." arXiv preprint arXiv:2407.00908 (2024).

[10] Argaw, Dawit Mureja, et al. "Scaling Up Video Summarization Pretraining with Large Language Models." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2024: 8332-8341.

[11] Lee, Yuho, et al. "UniSumEval: Towards unified, fine-grained, multi-dimensional summarization evaluation for llms." arXiv preprint arXiv:2409.19898 (2024).

[12] Tang, Liyan, et al. "Tofueval: Evaluating hallucinations of llms on topic-focused dialogue summarization." Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers). 2024.

[13] Zhao, Shu, et al. "Hyfit: hybrid fine-tuning with diverse sampling for abstractive summarization." IEEE Transactions on Big Data (2024).

[14] Chen, L., and Liu, S. "Explainable Cross-Market Causal Discovery via Financial Knowledge Graphs." (2024).

[15] Peng, Z. "Cross-Period Financial Statement Anomaly Detection via Contrastive Representation Learning." (2024).

[16] Su, Z. "Microstructure-Aware Risk Detection in High-Frequency Trading Using Embedding and Sequence Modeling." (2024).

[17] Wang, W., and Cheng, J. "Trustworthy LLM-Agent-Oriented Enterprise ETL Orchestration via Neural Execution Verification." (2024).