ECTSum: A New Benchmark Dataset For Bullet Point Summarization of Long Earnings Call Transcripts

Rajdeep Mukherjee[1], Abhinav Bohra[1], Akash Banerjee[1], Soumya Sharma[1], Manjunath Hegde[2], Afreen Shaikh[2], Shivani Shrivastava[2], Koustuv Dasgupta[2], Niloy Ganguly[1,3], Saptarshi Ghosh[1], Pawan Goyal[1]

[1]Indian Institute of Technology Kharagpur, India, [2]Goldman Sachs Data Science and Machine Learning Group, India
[3]Leibniz University of Hannover, Germany

Objectives

- To instigate research in **Financial Document Summarization**, which remains largely unexplored due to the unavailability of suitable datasets.
- To create a scalable summarization dataset in the financial domain with **minimal to zero dependency on human annotations**.

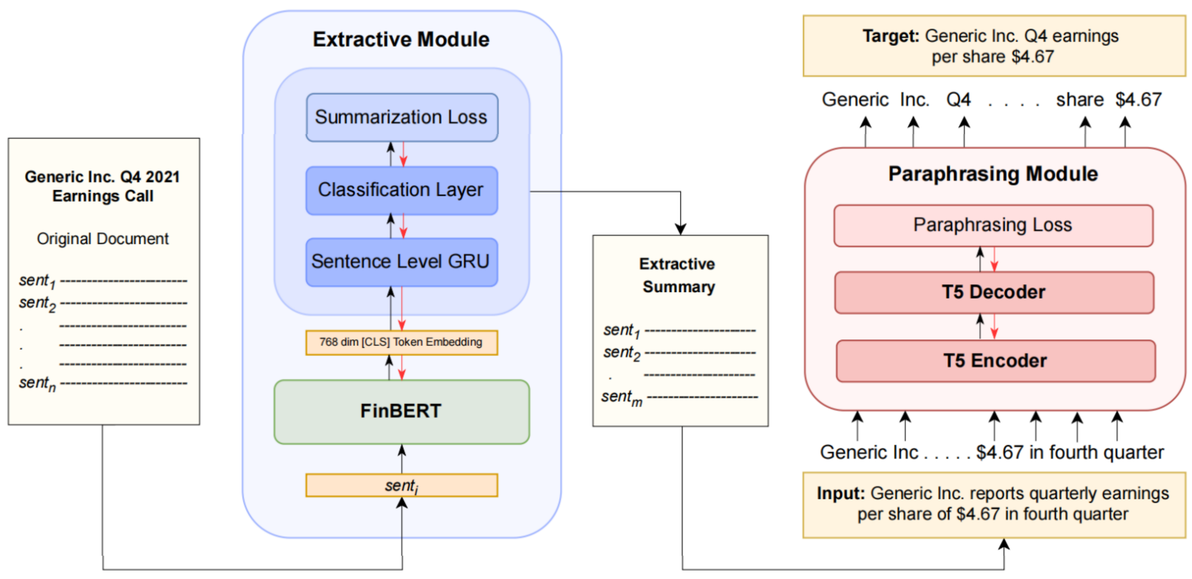
Our Contributions

- We present **ECTSum**, a **highly scalable**, and the first long document summarization dataset in the financial domain.
- Documents are **free-form lengthy transcripts** of company earnings calls (*Earnings Call Transcripts*), collected from **The Motley Fool**.
- Target summaries are a set of **telegram-style bullet points** obtained from corresponding **Reuters** articles that cover the calls.
- We benchmark the performance of a wide range of summarizers, especially long document summarizers, on ECTSum, against automatic metrics.
- We propose **ECT-BPS**, a simple-yet-effective solution for the task of *bullet point summarization* of long earning call transcripts (**ECT**s).

Key Dataset Properties

- Documents (ECTs) are unstructured; salient content evenly distributed.
- Average length of documents (earnings call transcripts): 2.9K words
- Average length of target summaries: 50 words.
- Document-to-summary *Compression Ratio* score of **103.67**.
- Train-Val-Test split ratio: 70:10:20

ECT-BPS - Our Proposed ECT Summarizer



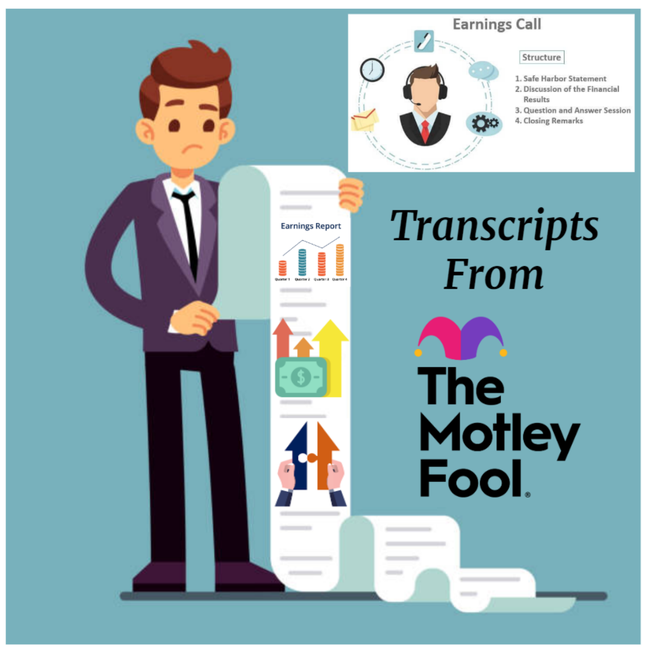
Results

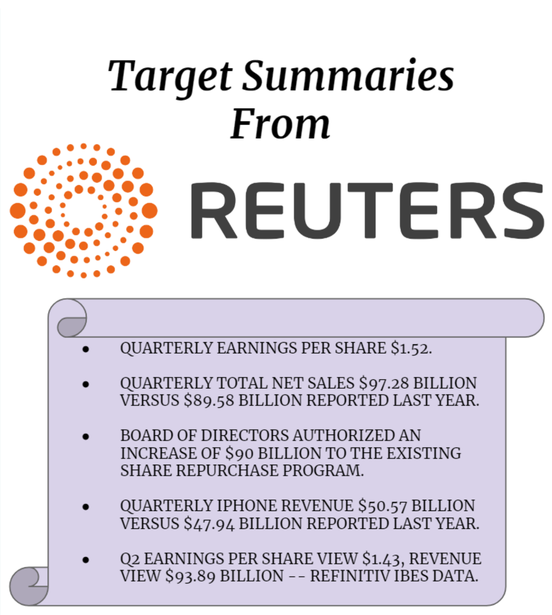
Model	ROUGE-1	ROUGE-2	ROUGE-L	BERTScore	Num-Prec.	SummaC
BigBird [1]	0.344	0.252	0.400	0.716	0.844	0.452
LongT5 [2]	0.438	0.267	0.471	0.732	0.812	0.516
LED [3]	0.450	0.271	0.498	0.737	0.679	0.439
ECT-BPS	**0.467**	**0.307**	**0.514**	**0.764**	**0.916**	**0.518**

Table 1:Comparing the performance of ECT-BPS with long document summarizers against automatic evaluation metrics. **ROUGE** and **BERTScore** evaluate the content quality, whereas **Num-Prec.** and **SummaC** evaluate the factual consistency of model-generated summaries.

References

[1] Zaheer et. al; Big Bird: Transformers for Longer Sequences; **NeurIPS 2022**
[2] Guo et. al; LongT5: Efficient Text-To-Text Transformer for Long Sequences; **NAACL 2022**
[3] Beltagy et. al; Longformer: The Long-Document Transformer; **ArXiv 2020**





ECTSum - Our Proposed Dataset

Dataset	# Docs.	Coverage	Density	Compression Ratio	# Tokens Doc.	# Tokens Summary
Arxiv/PubMed	346,187	0.87	3.94	31.17	5179.22	257.44
BillSum	23,455	-	4.12	13.64	1813.0	207.7
BigPatent	1,341,362	0.86	2.38	36.84	3629.04	116.67
GovReport	19,466	-	7.60	19.01	9409.4	553.4
BookSum Chapters	12,293	0.78	1.69	15.97	5101.88	505.32
ECTSum	2,425	0.85	2.43	**103.67**	2916.44	49.23

Table 2:Comparing the statistics of ECTSum with existing long document summarization datasets. **Covergae** and **Density** quantify the extent to which a summary is derivative of the source text. ECTSum has the **highest** document-to-summary **compression ratio** among all the datasets.

Key Takeaways

- Given the form and content of ECTs and telegram-style target summaries, ECTSum is an **extremely challenging** summarization dataset.
- **Highly extendable** dataset, with **no manual annotation** involved; documents and reference summaries collected from public domain.
- **ECT-BPS comprehensively outperforms** strong baselines.
- Ensuring **factual consistency** of model-generated summaries is crucial in the finance domain. Only evaluating content quality is not enough.

Evaluation by Financial Experts

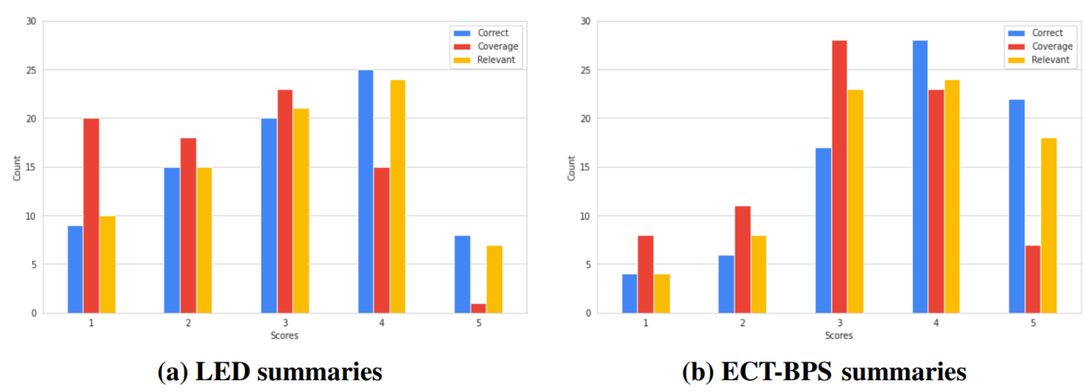


Figure 1:Histogram distribution of human evaluation scores assigned to model-generated summaries

Contact Information

Paper Details