

Email address:

lilliann.rowan@fmr.com; indraneel.biswas@fmr.com; michael.threlfall@fmr.com

Motivation, Background and Overview:

For a more targeted approach, we propose that the team concentrate on evaluative metrics for **abstractive summarizations** of financial documents produced by large language models (LLMs).

Quantitative measures are needed to identify the LLM or prompt that generates the best summary.

What are the characteristics of a good LLM generated summary? First, it needs to contain the important or relevant details needed by the user. This can be measured by comparing reference summaries (i.e., summaries written by subject matter experts) and LLM generated summaries. Second, a good summary should not contain hallucinations. A good summary should only contain entities found in the original document. In addition, the relationships between these entities should be consistent with the relationships in the reference document. And lastly, the summary needs to be linguistically fluent (i.e., grammatically correct) and coherent.

Project Description:

The project involves the creation of an evaluation framework to quantitatively measure the quality of abstractive summarization models. The team will need to:

1. Generate summaries using different LLMs or prompts.
2. Use several evaluation metrics (e.g., GLUE, ROUGE, METEOR, etc.) to pick the best model or prompt based on the above definition of a “good” summary.
3. Recommend metrics, methods, or **post LLM** models to use as part of an evaluation framework.

Data:

The team will be provided with links to a document and a corresponding reference summary written by a subject matter expert. Here is one example:

URL: [RER-1171836](#)

SME Summary:

FINRA found the firm failed to reasonably supervise for potentially manipulative trading from 2013 through 2019 and SageTrader also failed to establish and implement AML policies and procedures reasonably expected to detect and cause the reporting of suspicious activity and violated FINRA Rules 3310(a) and 2010. As a result, SageTrader consented to a censure and a \$100,000 fine to FINRA. Additionally, SageTrader also had to pay a total fine of \$775,000 which was paid to FINRA and eight exchanges, and the firm was also required to review and revise its supervisory system for detecting potentially manipulative trading by the firm’s customers.

Expected Outputs:

1. Report with the proposed evaluation framework for abstractive summarizations. The framework needs to measure all four characteristics of a good summary described in the “Motivation Background Overview” section. It needs to describe:
 - a. The different evaluation metrics explored, its limitations, the rationale for using the metric, and recommended use.
 - b. Any post-models or methods implemented.
 - c. Final recommended evaluation framework.
2. Executive summary
3. Annotated code

Stretch Goal(s):

1. Create a network graph that shows the relationships between the different entities (e.g., companies, persons, FINRA rules) mentioned in all documents.