ECE1786 Project Final Report

EcoInsight Detector

Yizhen Lu - 1002287660
Yang Song - 1002467984

Word count: 1800
Penalty:0

Permission:

Yizhen Lu

permission to post video: yes

permission to post final report: yes

permission to post source code: yes

Viola Song

permission to post video: yes

permission to post final report: yes

permission to post source code: no

Introduction

Earnings call transcripts offer critical insights for investment decisions, covering information like financial performance, strategies, and potentially predictive indicators for companies or industries. However, analyzing this extensive data manually is impractical due to its sheer volume. The goal of the project is to leverage the GPT4 model to help investors analyzing earnings call transcripts. It's a Class-2 project centered on using advanced GPT4 models to summarize these transcripts and then derive insights from the summaries. The primary focus is evaluating the model's ability in extracting relevant and time-sensitive insights, aiding the understanding of significant events, emerging trends, and patterns from the transcripts.

Background & Related Work

In the field of applying advanced language models like GPT for text analysis and summarization, there are some related works that stand out.

Width.ai leverages GPT-4 for opinions summarization, which recognize and condense key opinions into concise summaries[1]. This approach demonstrates the capability of GPT to extract key information from complex text. Another related work is that the MLQ.ai fine-tuned GPT-3 for earnings call assistant[2] . The customization of GPT-3 helps in understanding financial jargon and earnings call structures for creating more accurate summaries. This work demonstrates the ability of fine-tuning large language models to cater to specific domains like financial reporting. [2]

These works illustrate the growing trend of implementing large language models for advanced text summarization in specific fields. They provide a base for our project into summarizing earnings call transcripts and extract insights using GPT-4.

Data and Data Processing

The earnings call transcripts data were extracted from the Financial Modeling Prep API, focusing on major companies within the Technology, Financial Services, and Healthcare sectors listed on NASDAQ and NYSE from Q1 2020 to Q4 2022. The obtained transcripts were then structured as CSV files with key details as shown below:

symbol	quarter	year	date	content
BAC	1	2021	2021-04-15 18:04:07	Operator: Good day, everyone, and welcome to t...
BAC	2	2021	2021-07-14 14:28:06	Operator: Good day everyone and welcome to the...
BAC	3	2021	2021-10-14 15:26:05	Operator: Stand by, your program is about to b...
BAC	4	2021	2022-01-19 14:30:26	Operator: Good day, everyone, and welcome to t...
BAC	1	2022	2022-04-18 10:24:04	Operator: Good day, everyone, and welcome to t...
BAC	2	2022	2022-07-18 12:11:04	Operator: Good day, everyone, and welcome to t...
BAC	3	2022	2022-10-17 14:59:05	Operator: Good day, everyone, and welcome to t...
BAC	4	2022	2023-01-13 13:50:25	Operator: Good day, everyone, and welcome to t...

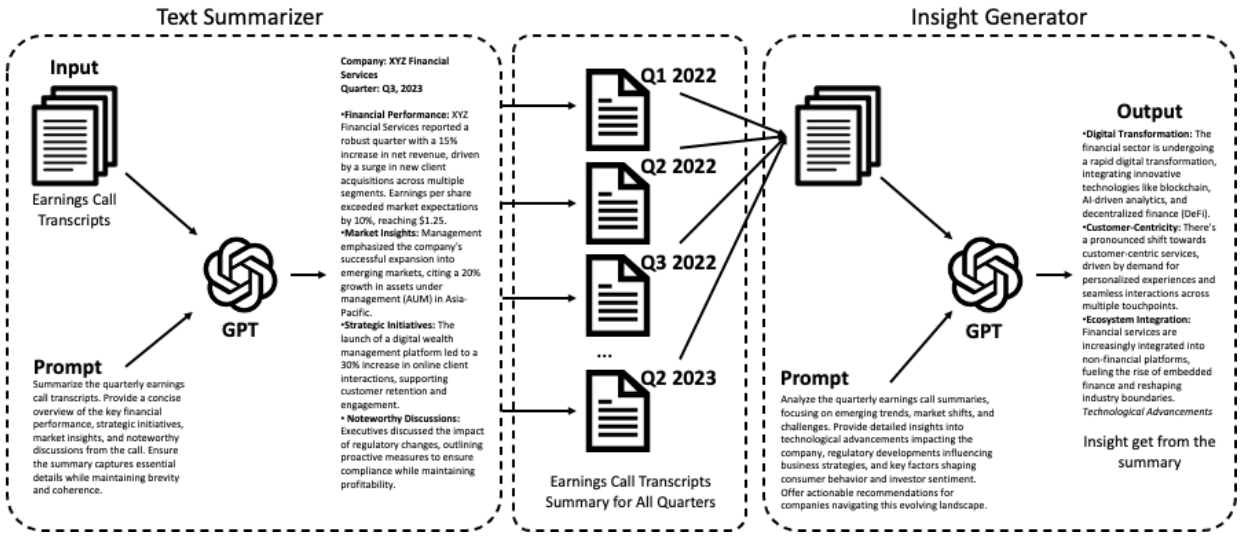
For data preprocessing, all conversations attributed to the operator were removed, retaining only discussions from the management team. Further refinement involved eliminating extraneous elements like human names, gratitude-related words, and irrelevant phrases. The goal is to keep crucial information while using GPT-4 cost-effectively.

Summaries for each transcript were created with the GPT-4 API and aggregated in JSON format. Each item in the JSON has two keys, the call data and the summary of that call. The choice of JSON allows the input provided to the insight generator to maintain the chronological context, better than just simply aggregating all summaries into one paragraph. This approach empowers GPT to generate insights that relate to the company's performance across time, uncover potential trends and detect evolving patterns, which is a crucial part of our result evaluation.


```
    {
        "date": "2021-04-15 18:04:07",
        "summary": "Financial Performance: Net income was $8.1 billion, up from $5.5 billion in the previous quarter. This increase was
driven by strong revenue growth and a significant reserve release of $2.7 billion.\n\tBalance Sheet and Liquidity: The balance sheet
expanded to $2.97 trillion, with deposits increasing by $89 billion and loans decreasing by $25 billion. The liquidity portfolio
exceeded $1 trillion.\n\tSegment Performance: Both the Global Markets and Global Wealth Management segments showed strong performance,
while Consumer and Global Banking segments are recovering from previous challenges.\n\tAsset Quality: Asset quality remained strong
with low net charge-offs of $833 million. The bank released $2.7 billion in credit reserves due to an improved economic outlook.
\n\tDigital Engagement: 70% of consumer households used the bank's digital platform, indicating a strong focus on digital engagement
and innovation.\n\tCapital Management: The CET1 ratio was maintained at 11.8%, reflecting a strong capital position.\n\tOperational
Costs: Quarterly expenses were $15.5 billion, influenced by seasonal factors and one-off charges.\nOverall, Bank of America reported a
solid quarter with growth in net income, a strong balance sheet, and continued investment in digital capabilities, while navigating
economic uncertainties and managing operational costs."
    },
    {
        "date": "2021-07-14 14:28:06",
        "summary": "Net Income: The bank reported $9.2 billion in after-tax net income, equivalent to $1.03 per diluted share.
\n\tReserve Release and Tax Adjustment: A notable $2.2 billion credit reserve release was reported due to improving asset quality.
Additionally, a $2 billion positive income tax adjustment was made following the increase in the UK corporate income tax rate.
\n\tRevenue and Profitability: Revenue declined by 6% from Q1, primarily due to lower sales and trading results. However, consumer and
wealth management revenues were strong. The return on tangible common equity stood at 20%, and return on assets was at 123 basis
points.\n\tBalance Sheet and Capital: The bank's total assets increased by $60 billion, reaching over $3 trillion. The CET1 ratio was
200 basis points above the 9.5% minimum requirement, translating into a $31 billion capital cushion.\n\tLoan Growth and Economic
Outlook: There was modest loan growth across various sectors. The bank anticipates increased loan activity as the economy recovers.
Digital engagement continued to grow with record user numbers.\n\tFuture Strategy: Bank of America expressed optimism about the
economic recovery and customer activity. Plans to resume share buybacks were discussed, reflecting the bank's strong earnings and
excess capital position.\nOverall, Bank of America's Q2 results demonstrated robust asset quality, controlled expenses, and strategic
positioning for continued growth in a recovering economy."
    },
    {
        "date": "2021-10-14 15:26:05",
        "summary": "Net Income: The bank reported a strong quarter with $7.7 billion in net income, up significantly from the previous
year, marking over $25 billion in earnings for the first nine months of the year.\n\tGrowth and Operating Leverage: The bank saw a
return to pre-pandemic growth trends, with organic growth across all business lines. Operating leverage was notable with a 12% year-
over-year revenue increase and flat expenses.\n\tLoan Growth: There was solid loan growth across various sectors, with broad-based
growth in both consumer and commercial loans. Total loans increased by $16 billion quarter-over-quarter, excluding PPP loans.\n\tNet
Interest Income (NII): NII showed significant improvement due to deposit growth, loan growth, and efficient interest rate risk
management. The bank expects modest NII growth from Q3 to Q4.\n\tDigital Banking: The bank highlighted robust digital engagement, with
nearly 41 million active digital platform users. Digital sales increased by 27% year-over-year.\n\tCapital Return: The bank returned
$12 billion to shareholders through dividends and share buybacks, underlining its strong earnings and excess capital position.
\n\tCommunity Investments: Bank of America has funded nearly $400 million of its $1.25 billion commitment to community support,
including investments in minority-focused funds and equity investments in MDIs and CDFIs.\n\tFuture Outlook: The bank expects continued
deposit and loan growth aligned with economic growth. It anticipates robust improvement in NII for the full year 2022 compared to 2021,
assuming the current forward interest rate curve.\nThe bank's leadership team expressed optimism about the economic recovery and
customer activity, with a focus on continuing to deliver strong results while investing in the community and returning capital to
shareholders."
    },
```

Architecture and Software

The architecture comprises two components: text summarizer and insight generator, both utilizing GPT-4.



Firstly, the text summarizer takes an individual transcript as input and utilizes the prompt to extract concise summary. This prompt guides the summarizer to distill details related to financial performance, strategies, and discussions from the earnings call. The summarizer outputs a coherent piece of text summarizing the input transcript.

This summarizer is applied to transcripts across all quarters, and the output summaries along with the call dates are then compiled into a JSON.

Next, the insight generator takes the summaries(JSON) and employs a prompt designed to extract insights. This prompt focuses on identifying emerging trends, challenges faced, technological advancements, and factors affecting consumer behavior and investor sentiment, etc.

The summarizer and insight generator work in sequence: the summarizers‘ output(JSON) is used by the insight generator. Continuous improvement and metric evaluation will refine the prompts used in each stage to achieve the desired quality.

Comparison Method

<u>Text Summarization</u>

We generated multiple summaries using specific prompts and compared them against the analysis reports written by analysts from websites like RBI and Yahoo Finance, etc. BERTScore, a metric assessing semantic similarity via BERT embeddings, is used to

quantitatively evaluate the summaries against these articles, ensuring a comprehensive evaluation of our model's performance in capturing key information.

Insight Generation

The 3 financial analysts designed a rubric to rate the insights from 4 different aspects as shown in the table below. Then define patterns from the feedback, for instance, if certain types of insights are consistently rated low, determine what they have in common and refine the prompt. Additionally, we labeled the insights as 1(good) or 0(bad) according to the feedback and then prompted GPT for evaluation. This classification facilitated the fine-tuning of our model and helped us do better on the prompt engineering.

Our process combined human judgment with GPT4 to make our system robust on detailed and insightful financial analysis.

Rubric for Rating the Insights

Aspects	Details	Score
Accuracy	Evaluate if the insight accurately reflects information from the summaries	0/1
Relevance	Evaluate if the insight is relevant to the company's financial performance, strategic initiatives, market insights and future prospects	0/1
Depth of Analysis	Does it go beyond surface-level information? (Trend over time, historical data, etc.)	0/1
Actionability	Analyze if the insight provides actionable information for audiences	0/1
Overall	Sum of the scores above	0-4

Quantitative Results

Quantitative results for Text Summarization

For each quarter, we focused on obtaining 1-3 financial analysis reports from financial websites. The team also leveraged GPT-4 to assist in searching for financial reports online. To evaluate the effectiveness of the summarizer, the metric 'BERT score' is calculated which can evaluate the semantic similarity between two pieces of text. For each quarter, we compared our generated summary with each financial report, we then computed the average BERT score for that quarter, providing a quantitative measurement of the quality of the summaries. We repeated the process across different quarters, subsequently, we averaged these scores to obtain one comprehensive BERT score for the entire company. The team achieved a notable average BERT score of 0.7934, indicating a high level of semantic similarity and thus, an effective summarization system. The specific results are in the appendix A.

Quantitative Results for insight generator

Three financial analysts helped on evaluating the two companies' insights by rating the generated insights according to the rubric. However, the workload was too heavy for the three people to rate the entire 9 companies. Therefore, the team worked with financial analysts to design the prompt and asked GPT-4 to assess the insights. The financial analysts also checked the GPT-4's thought processes to ensure the generated results from GPT-4 were accurate and aligned with the expectations of the analysts. After several rounds of review and adjustments, the analysts approved the thought processes and reasoning provided by GPT-4. The rubric and reasoning that compares between financial analysts and GPT-4 are provided in the Appendix B & C. The team eventually reached a 3.8 out of 4 overall, indicating the success of the insight generator and the results are shown in Appendix D.

Qualitative Results

Good Insight Examples:

- *The company consistently reports high growth in service revenue, including Apple Music, iCloud, and Apple TV+. The services division has become a focal point in the business strategy, with the company continually investing in digital services.*
- *Net income figures showed recovery from approximately $5 billion in Q3 2020 to $8.1 billion by Q1 2021, indicating a steady improvement in financial performance.*
- *The pandemic accelerated a shift towards online and mobile platforms, with the bank often noting the growth of digital sales. This trend persisted even as pandemic-related restrictions eased, indicating a permanent shift in consumer behavior towards digital services.*

Not Good Insight Examples:

- *Collaborations and strategic acquisitions can bolster market position and drive innovation.*
- *Clear communication with stakeholders about challenges, strategies, and outlook, particularly during uncertain times, can foster trust.*
- *Building flexible and resilient supply chains is critical to mitigating disruptions from global events and shortages*

The first three insights are good because they are specific, data-driven, and closely tied to the companies they describe. They provide concrete figures, such as revenue growth and net income recovery, or highlight significant trends, like the shift towards digital services in banking. These insights go beyond surface-level information, offering a clear analysis of trends over time and strategic shifts within the companies. This level of detail not only enhances their relevance but also makes them actionable, providing valuable information for decision-making in investment and business strategy.

In contrast, the latter three insights are not good because of their lack of specificity and direct connection to particular data or company-specific trends. They read more like general business principles rather than insights derived from the analysis. While they touch on important themes like collaboration, stakeholder communication, and supply chain resilience, they do not provide concrete examples or context from specific companies. This generality limits their depth of analysis and makes them less actionable for specific business or investment strategies.

Discussion and Learnings

In the project retrospective, we observed that our model's results were intuitively aligned with the global economics context, underpinning its effectiveness in capturing key industry trends. Examples include identifying growth in digital services as a significant trend for a technology company and pointing out that the shift to digital banking services was a direct response to the pandemic, marking a fundamental change in consumer behavior.

A surprising aspect of the model was its ability to accurately represent numerical trends, such as increase, steady improvement, and significant drops, etc.. Moreover, the model generated time-specific insights, successfully linking key findings to their corresponding dates of the transcripts, which provide a chronological narrative of events and trends. This level of precision in temporal analysis is noteworthy, adding a valuable dimension to our understanding and use of the GPT4 model.

However, most insights were fact-based and contextual, with limited predictive elements. Future improvements include implementing predictive prompt engineering for GPT4 to forecast trends. Broadening data sources to encompass market metrics, competitor analysis, and extensive industry reports will also enrich our analysis. These steps aim to enhance our model's current trend comprehension and enable more accurate future predictions.

Individual Contributions

Yizhen Lu
1. Collect data from Financial Modelling Prep API 2. Preprocess the data earnings call transcript data 3. Build the prompt for text summarizer 4. Use the chosen prompt to run the text summarizer for company AAPL, BAC, AZN,JNJ,PFE,NVDA 5. Use the chosen prompt to run the insight generator for company MSFT, WFC, JPM 6. Find 3 financial analysts to build and evaluation metrics for insight generator and manually evaluate the insights generated for AAPL and JNJ
Viola Song
1. Use the chosen prompt to run the text summarizer for company MSFT, WFC, JPM 2. Use aggregated summaries to build the prompt for insight generator 3. Use the chosen prompt to run the insight generator for company AAPL, BAC, AZN,JNJ,PFE,NVDA 4. Find the related financial analysis report for each transcript and evaluate the text summarizer results 5. Evaluate the Insight generators result using GPT4

6. Compare the GPT4 evaluation with the manual evaluation for AAPL and JNJ

Reference:

[1] Bhaskar, A., Fabbri, A.R. and Durrett, G. (2023) Prompted opinion summarization with GPT-3.5, arXiv.org. Available at: https://arxiv.org/abs/2211.15914 (Accessed: 11 December 2023).

[2] Foy, P. (2023) Fine-tuning GPT-3: Building an earnings call assistant, MLQ.ai. Available at: https://www.mlq.ai/fine-tuning-gpt-3-earnings-call-assistant/ (Accessed: 11 December 2023).

Appendix:

Text Summarizer Prompt:

Summarize the quarterly earnings call transcripts. Provide a concise overview of the key financial performance, strategic initiatives, market insights, and noteworthy discussions from the call. Ensure the summary captures essential details while maintaining accuracy, brevity and coherence.

Insight Generator Prompt:

Analyze the quarterly earnings call summaries, focusing on emerging trends, market shifts, and challenges. Provide detailed insights into technological advancements impacting the company,regulatory developments influencing business strategies, and key factors shaping consumer behavior and investor sentiment. Offer actionable recommendations for companies navigating this evolving landscape.

Insight evaluation Prompt:

We are doing an insight generator task. First, we take the company's earnings call transcripts and do the summarization task to a bunch of summaries with call date. This summarizer is applied to transcripts across all quarters, and the output summaries along with the call dates are then compiled into a JSON. And then the insight generator takes the summaries(JSON) and employs a prompt designed to extract insights. This prompt focuses on identifying emerging trends, challenges faced, technological advancements, and factors affecting consumer behavior and investor sentiment, etc. There will be one insight for one company but it provides insight across a couple of years. Here are the evaluation metrics of the insights. Aspects

Details
Score
Accuracy: Evaluate if the insight accurately reflects information from the internet, you can search up online for the information and compare with the insights to see if it is correct.
0/1
Relevance: Evaluate if the insight is relevant to the company's financial performance, strategic initiatives, market insights and future prospects
0/1
Depth of Analysis: Does it go beyond surface-level information? (Trend over time, historical data, etc.)
0/1
Actionability: Analyze if the insight provides actionable information for audiences
0/1
Overall: Sum of the scores above
0-4
Please return the score first, and then explain the reasons for giving the scores. For accuracy, please check the internet information rate score. Please follow the format 'overall score - score for aspects and reasoning

Appendix:

A. Bert Score for summaries

Company	Avg. Bert Score
AAPL	0.8034
BAC	0.798
JNJ	0.7811
PFE	0.7629
WFC	0.7978
JPM	0.8092
AZN	0.7892

NVDA	0.7851
MSFT	0.8141
overall	0.79342222

B. BAC Insights Evaluation by financial analysts

Aspects	Analyst 1	Reasoning
Accuracy	1	Nice! The insights match the key points from the earnings calls, like the COVID-19 impact and the push for digital banking.
Relevance	1	Totally relevant!
Depth of Analysis	1	Gave a solid, in-depth look at where the bank's heading.
Actionability	1	Super practical advice, with clear-cut recommendations.
Overall	4	

Aspects	Analyst 2	Reasoning
Accuracy	1	Correct
Relevance	1	Hits the nail on the head with what's crucial for BAC

Depth of Analysis	1	It mentions trends and history
Actionability	1	The recommendations are on point
Overall	4	

Aspects	Analyst 3	Reasoning
Accuracy	1	Right on target!
Relevance	1	relevant
Depth of Analysis	0	It's a decent overview, but it's not digging deep enough. I want more analysis on the long-term implications of these trends.
Actionability	1	The advice is sound
Overall	3	

C. BAC Insights Evaluation by GPT-4

Aspects	GPT-4	Reasoning
Accuracy	1	The insights align well with the known data and trends in Bank of America's financial reports, accurately reflecting key aspects like COVID-19 impact and digital transformation.
Relevance	1	The insights are relevant to Bank of America's financial and strategic context, touching on crucial

		areas like net interest income, loan growth, and digital banking.
Depth of Analysis	1	The analysis effectively delves into the depth of Bank of America's financial dynamics, offering a comprehensive view of trends and market shifts over time.
Actionability	1	The insights provide actionable recommendations that are practical and directly applicable to Bank of America's strategic decision-making.
Overall	4	Overall, the analysis offers a well-rounded, accurate, relevant, and in-depth view of Bank of America's financial state, providing actionable insights for future strategies.

D. Insights Evaluation by GPT-4 for all Companies

Company	Score
AAPL	4
BAC	4
JNJ	4
PFE	3
WFC	4
JPM	4
AZN	3

NVDA	4
MSFT	4
overall	3.8