



AI-Powered Slides Generation and Chat for Financial Reports

Yongwon Cho

Liping Yin

Rongxiang Zhi

Yaohai Zhou

Table of Contents

Table of Contents.....	2
Abstract.....	3
Problem Definition.....	4
Overview.....	4
Significance.....	4
Assumption.....	5
Literature Review.....	6
Methodology.....	7
Slides Generation.....	7
Chat.....	8
AI Topic Selection.....	10
Web Client.....	10
Analysis.....	11
Data Samples.....	11
Demo.....	12
Performance.....	18
Conclusion.....	25
Finding.....	25
Recommendation.....	26
Reference.....	27

Abstract

This project revolves around the development of a software application tailored for finance professionals, particularly CFOs and their assistants. The software aims to streamline the conversion of a company's 10-K financial report into a PowerPoint presentation using automation and AI. The presentation is enhanced with visually engaging charts, succinct summaries, and a chat feature for efficient information retrieval. This project holds significance as it empowers finance experts by automating presentations and using AI to enhance data understanding, leading to increased productivity. However, the current scope is restricted to one company because of time and capacity constraints, resulting in practical limitations for broader commercial utilization.

We comprehensively evaluate the performance of the software across various dimensions. We validate the accuracy of generated charts through meticulous manual calculations and automated implementations using matplotlib and python-pptx. Summarization quality is consistently achieved with a 100% success rate in over 100 trials. Our AI topic selection demonstrates an 80% success rate, highlighting challenges in generating relevant and distinct topics. The chat feature is evaluated through manual assessment, yielding a 90% success rate in responding to diverse queries. The software's overall reliability is confirmed with a 100% execution success rate in 100 instances.

The project acknowledges limitations such as the confined scope and challenges in AI decision-making and creativity. Recommendations include potential expansion to 10-Q reports, refining topic selection strategies, exploring AI-driven chart generation, and addressing ethical concerns related to AI-generated content quality.

Problem Definition

Overview

The main goal of this project is to develop a software application tailored for finance professionals, including CFOs and their assistants. This tool aims to streamline the process of transforming a company's 10-K financial report into a PowerPoint presentation. By harnessing the power of automation and AI, our software will enhance the presentation by integrating visually engaging components such as charts and succinct summaries. To enhance user experience, we have included a chat feature that accompanies the generated PowerPoint. This feature enables users to efficiently locate and extract information by asking a variety of questions related to the document. This capability can be particularly beneficial if users want to incorporate additional slides of their choosing.

Significance

Our project has the potential to bring significant advantages to its intended users, specifically targeting company CFOs and their assistants. These professionals are not only responsible for preparing financial reports but also face the challenge of effectively presenting information to company boards and the public. Our software aims to empower them by boosting their efficiency in presentations. By automating the process of creating slides, we free up their time and resources, enabling them to focus more on strategic decision-making instead of spending valuable hours manually crafting slides. The inclusion of visual elements like charts and text summaries ensures that the key aspects of the financial report are communicated succinctly and informatively, aiding in audience comprehension and engagement.

In terms of the chat feature, it could be seen as an additional capability that employs AI to help users locate and learn information more effectively from financial documents. More importantly, this feature assists users in extracting pertinent content on various topics of their choosing. This ability to add extra slides on top of the auto-generated ones ensures the software remains valuable even when users require additional customization and flexibility.

Moreover, recent advancements in AI Natural Language Processing (NLP) technology have brought about a revolutionary change in how information is structured and summarized. Our project takes advantage of these developments, making it possible to extract valuable information and generate insightful text based on the contents of financial reports. By harnessing AI, we can furnish users with accurate and pertinent summaries that capture the core of the data, facilitating efficient communication of crucial financial information. This technological breakthrough empowers users to navigate vast amounts of data effortlessly, saving time and energy while upholding the coherence of the generated content.

Assumption

Delimitation

In order to ensure practicality given our time constraints and development capacity, we have opted to narrow down the scope of this project and concentrate solely on one company, namely Apple Inc., as opposed to accommodating various companies. This decision was prompted by the realization that the formats of 10-K filings differ across companies, even though they adhere to general legal guidelines. By confining our focus to a specific company, we can effectively mitigate the intricacies and difficulties associated with automated parsing and analysis of 10-K filings. This approach ensures the successful delivery of a Minimum Viable Product (MVP) within the given timeframe, while still showcasing the overall value of our efforts.

However, the chosen scope also brings forth a notable limitation in terms of its practicality. Consequently, the software currently functions predominantly as an internal tool for the designated company, rather than a commercially feasible product capable of accommodating a diverse range of clients without extensive code customization. Expanding its applicability would necessitate more advanced AI capabilities for information extraction, exceeding our current technological capabilities.

Limitation

The primary limitation of our endeavor is that, although we have managed to produce a functional product, we lack established benchmarks for controlling or assessing AI-generated content, both in the produced slides and chat responses. Regrettably, we have not identified any useful tools, libraries, or existing research that can aid us in quantifying the quality of AI-generated content. Consequently, we depend on manual examination and assessment to determine the quality as we progress with our analysis. While we are confident that manual evaluation is adequate for our present scope, it does raise concerns that need to be addressed should we decide to expand our scope, for instance, from an internal corporate tool to a commercial product.

Literature Review

ChartGPT[1] is an AI tool designed for constructing charts through text input, aligning with some aspects of our project's objectives. However, given our deliberate choice to manually select topics/sections, utilizing prompts for chart generation lacks significance, as we can directly implement hardcoded methods. Nevertheless, it holds potential value for future developments, especially if we choose to expand our scope beyond manual selections.

The SEC API[2] is a Python package allowing you to search the entire SEC filings corpus and access over 650 terabytes of data. It offers a more convenient approach to extract sections from 10-K filings, eliminating the need for direct PDF parsing.

For our text summarization and interactive chat functionalities, we rely on GPT (Generative Pre-trained Transformer) models, specifically ChatGPT[3]. ChatGPT is a variant of GPT that has been fine-tuned for conversational interactions. It is a state-of-the-art language model developed by OpenAI, capable of generating coherent and contextually relevant text based on the input provided.

ChatPDF[4] serves as a commercial benchmark for evaluating the performance of our conversational AI system. It can extract information or answer questions from large PDF files like manuals, essays, books.

LlamaIndex[5] serves as a robust data framework tailored for applications powered by Language Models (LLMs) to efficiently process, organize, and retrieve private or domain-specific information. Within our project context, llama_index functions as the core component responsible for document indexing, which in turn drives the implementation of our chat functionality.

Python-pptx[6] is a Python library for creating and updating PowerPoint (.pptx) files. Our project used it to automate production of slides based on financial reports.

Methodology

Unless expressly indicated differently, all mentions of "AI," "GPT," "ChatGPT," or similar terms in later parts of the report refer to the selected model, specifically gpt-3.5-turbo, developed by OpenAI. This decision was driven by its extensive capabilities and cost-efficiency. While our trials might have revealed slight improvements when utilizing GPT-4 or other models across various AI-related sectors, the significantly increased costs associated with their usage did not justify the benefits attained. As a result, this report will not cover outcomes arising from the implementation of these models.

Slides Generation

Section Extraction

In the course of our investigation, we utilized the SEC API to access and retrieve specific portions of SEC submissions and related data using a standardized approach. More precisely, the Extractor API [\[7\]](#) aided in extracting text segments from the 10-K report. This API delivered the extracted content in a refined and standardized format, either as plain text or HTML. By providing the API with the filing's URL, the section designation (e.g., Item 1A), and the desired data format (e.g., raw text), we obtained the extracted content. In our project, we archived these extracted sections as text files. This approach greatly facilitated easy reading, storage, and access to pertinent information, enabling our team to refer back to specific sections effortlessly.

We initially attempted to employ AI to extract appropriate sections. However, we eventually decided to manually select most of the slides. This decision was influenced by the recognition that AI might encounter difficulties in precisely identifying pivotal sections of financial reports, given its absence of human contextual understanding and specialized expertise in this domain. Opting for the manual selection of critical sections could provide a clearer representation for finance professionals within the company. It's worth noting that, in the latter stages of the project, we ceased using the SEC API when experimenting with [AI-driven topic selection and information generation](#). This was due to the need for a more adaptable and automated information retrieval approach from the document.

Chart Generation

We began our data exploration using Pandas for data manipulation, paired with matplotlib for data representation. However, when we transitioned to the presentation aspect, we recognized the limitations of matplotlib's static graphics and saw a growing demand for more interactive visualizations. This realization led us to adopt the python-pptx library, which allowed for the creation of native PowerPoint graphing. We aimed to elevate the visual quality of our slides, and the built-in graphs in PowerPoint catered to this by giving a more polished look apt for business

settings. Additionally, they could be customized post-generation, making it easier to modify according to feedback or changing needs.

Text Summarization

This pertains to the assignment in which AI was utilized to succinctly summarize text segments from predetermined sections. It's important to note that this task did not entail employing AI to scan the entire document for relevant information, in contrast to the subsequent [AI Topic Selection](#) phase. Instead, for this specific task, the AI was given access to a single text snippet provided each time the API was invoked. In the creation of each slide, we employed a prompt that directed the AI to summarize a given text, which was then submitted to the API alongside the previously extracted text. The prompt was designed to include supplementary instructions, such as requesting the AI to format its response using bullet points rather than paragraphs.

The AI's reply consisted of a text string containing four bullet points summarizing the given text. Subsequently, this output underwent a reconfiguration process, which encompassed actions like division and exclusion of specific symbols. Due to the diverse nature of the generated responses, the code was designed to adapt to various formatting styles. For example, each bullet point could initiate with a number, period, or hyphen. Moreover, automated exception and retry mechanisms were incorporated to handle instances where difficulties arose during the reformatting of the response. The end result of this sequence was a Python list comprising four strings, each representing a bullet point intended for inclusion on the slide.

PowerPoint Assembly

Python-pptx^[6] is a library for the dynamic creation and manipulation of PowerPoint (.pptx) files using Python code. This library provides an interface for working with various elements of PowerPoint presentations, including slides, shapes, text, images, charts, and more. One of the advantages of python-pptx for this project was that it facilitated the creation of native PowerPoint charts.

In our code, we also implemented features that provided users with customization options. We introduced fields for users to customize the title and name on the slides. Moreover, we offered three templates: plain, green, and yellow. This variety ensured that users could choose a style that aligned with their content's mood and their audience's preferences. Additional customization functionalities could be incorporated in the future to enhance the comprehensiveness of the product, if deemed necessary.

Chat

The chat functionality enables users to inquire about any aspect of the document. While one might assume that the process entailed merely forwarding the entire document text alongside the question to the AI, the reality was quite different. The foremost constraint stemmed from an

imposed maximum input length for the API, rendering the transmission of the complete document infeasible. Even if this were possible, the cost-effectiveness of such an approach was questionable, given that the API cost correlated with input length. As a result, it became imperative to devise a method for discerning contextually relevant information pertaining to the user's query and exclusively transmitting this segment from the document to the AI. To fulfill this objective, we integrated the llama_index data framework into our implementation.

Indexing

In order to expedite the retrieval of contextually relevant information aligned with the user's query, it was imperative to construct an indexing database based on the input document. Given the document's content, our code initially divided it into smaller, approximately evenly-sized chunks. The sizing of each chunk was fine-tuned, as it significantly impacted the effectiveness of information retrieval. While opting for larger chunk sizes typically conserved space and time, it compromised the capacity to discern intricate details. This prompted our decision to settle on a relatively compact chunk size, as we anticipated the user's queries would often pertain to precise tasks like locating specific numbers.

Subsequently, each of these chunks underwent transformation into fixed-dimensional embeddings using GPT, with the anticipation that these embeddings would encapsulate the semantic essence of their respective chunks. The resultant output of this process materialized as a locally cached database, encompassing a data structure storing all the embeddings for subsequent querying purposes.

Querying

When a user posed a query concerning the document, that particular question was subjected to the aforementioned GPT-based transformation to generate an embedding. This embedding then functioned as the basis for querying the established database. The leading embeddings from this database, those most closely aligned with the input embedding in terms of relevance, were extracted and assembled into a consolidated collection of contextual information. This amalgamation of context information was then dispatched to the AI, accompanying the original question. It's worth noting that the quantity of embeddings that could be transmitted was restricted by the maximum input length permissible for the AI.

To enhance response quality, supplementary directives were incorporated, including instructions to exclude external knowledge. Additionally, we implemented auto-retry mechanisms to counteract potential timeouts and failures. With the aim of optimizing user experience, our code facilitated the streaming of responses from the AI. Consequently, answers were presented to the user as they were generated, eliminating the need for extended waiting periods and providing immediate access to partial responses. The AI's response served as the answer to be presented to the user in response to their question.

AI Topic Selection

The utilization of AI Topic Selection serves as an experimental component based on the [chat function](#). Following the completion of the document indexing process, we tasked AI with envisioning a scenario in which it possessed the latest 10-K report. Its objective was to propose topic titles for a board presentation grounded in this report. The requirement was to generate two concise topic titles, each consisting of no more than five words. Importantly, these topics were to maintain close relevance to the company and be frequently addressed within the 10-K document. The emphasis was on topics beyond basic financial analysis, potentially covering market regions, global events, specific products, and more. Furthermore, the topics were expected to exhibit notable differences and contrasts in comparison to the previous year.

Subsequently, we requested AI to furnish details about each chosen topic. We presented the AI with a scenario wherein a presentation was being formulated for the company's board using the 10-K report, centered around a chosen topic. The AI's task involved extracting pertinent details from the indexing database to generate three succinct bullet points for each topic. These bullet points were to highlight distinct aspects tied to the company's revenues or losses. Each point was expected to encompass 30-40 words and provide accurate statistical insights and document-based examples.

Web Client

Once all the aforementioned features were developed, our aim was to encapsulate the project within a web interface. For this purpose, we leveraged the Streamlit package in Python. This choice was driven by its seamless integration with our existing codebase, as well as its provision of visually appealing and streamlined interface components.

Analysis

Data Samples

In our software testing procedure, we utilized 10-K documents from Apple to conduct analysis. These selected documents pertain to the years 2020, 2021, and 2022. However, documents that stem from 2019 or an earlier period were intentionally left out of the testing process. The reason behind this exclusion was primarily the variations in their formatting, which would complicate the implementation process if incorporated. Additionally, these older documents were considered to have limited value for users due to their outdated content.

Our anticipation was that the current software would function seamlessly for future documents, granted the document format stayed the same. Nevertheless, any changes made to the document format might inadvertently cause discrepancies in how sections are labeled across various submissions, giving rise to problems. This scenario highlights the limitations of using specialized APIs, like the SEC API, to precisely extract important information. If there's a desire to augment the software's capabilities, addressing this limitation becomes vital. This would entail revising the implementation strategy to accommodate potential alterations to the section naming conventions within the SEC API, or even considering an entirely different and more adaptable approach to information extraction through AI.

Demo

Web Client

MEng Capstone

Slides

Year

2022

10-K URL

<https://www.sec.gov/Archives/edgar/data/320193/000032019322000108/aapl-20220924.htm>

Load Cache

Title

Test Title

Name

Test Name

Theme

green

Generate

Figure: Web Client Interface

The user is empowered to input their desired year and the URL of the 10-K document within the web client interface. They maintain the flexibility to customize the slide's title, the speaker's name, and the chosen theme. Once all the vital settings are in place, a single click activates the slide generation process.



Figure: Title Page Example

Total Net Sales

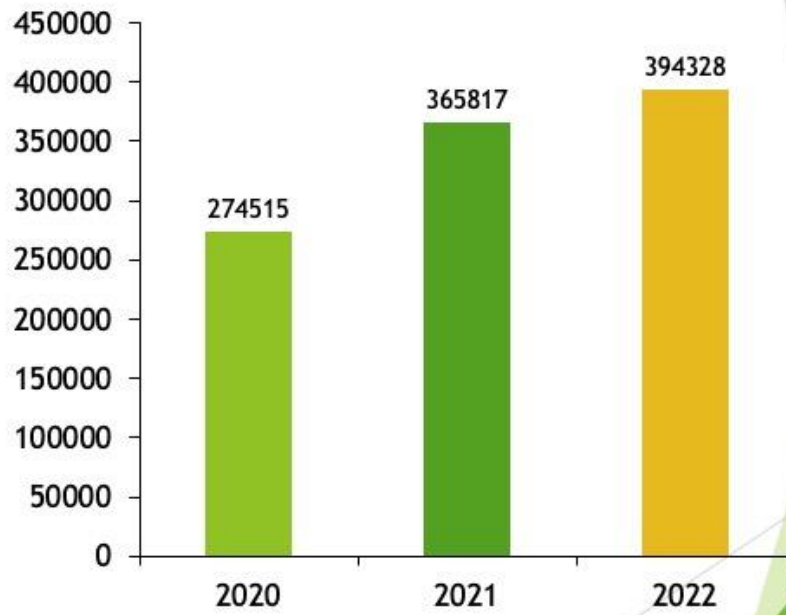


Figure: Bar Chart Example

Sales by Product

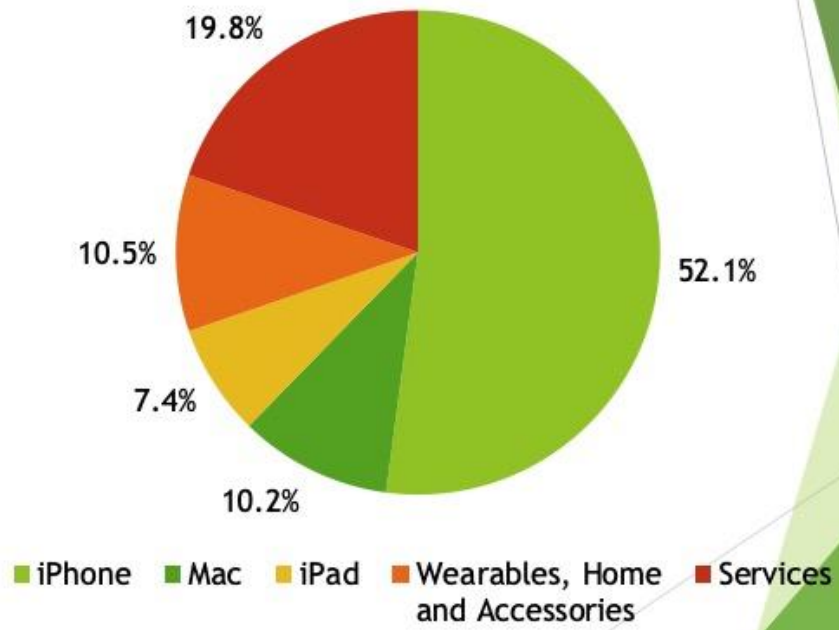


Figure: Pie Chart Example


Fiscal Year Highlights


- ▶ Total net sales increased by \$28.5 billion in fiscal year 2022, driven by higher sales of iPhone, Services, and Mac
- ▶ Significant product announcements in fiscal year 2022 included new MacBook Pro models, AirPods, iPhone SE, Mac Studio, and iPad Air
- ▶ The company raised its Program authorization to \$405 billion and increased the quarterly dividend to \$0.23 per share
- ▶ The COVID-19 pandemic had a significant impact, causing disruptions in global financial markets and supply shortages

Figure: Text Summarization Example


Chat


Chat

 Hi, I'm your AI assistant. Any question on the 10-K document?

 What is the sales of iPhone in 2022?

 The sales of iPhone in 2022 is \$205,489 million.

 How has COVID-19 affected the company? Answer in bullet points.

- 
- The COVID-19 pandemic has adversely affected the company's business, results of operations, financial condition, and stock price.
 - Governments and businesses have taken unprecedented measures in response to the pandemic, including travel and business operation restrictions, temporary closures of businesses, and quarantine and shelter-in-place orders.
 - The pandemic has significantly curtailed global economic activity and caused volatility and disruption in global financial markets.
 - The company's component suppliers and manufacturing and logistical service providers have experienced disruptions, resulting in supply shortages that affected sales worldwide.
 - Public safety measures can also adversely impact consumer demand for the company's products and services in affected areas.
 - The extent to which the pandemic may impact the company's operational and financial performance remains uncertain and depends on various factors outside the company's control, such as the timing, extent, trajectory, and duration of the pandemic, the emergence of new variants, the development, availability, distribution, and effectiveness of vaccines and treatments, the imposition of protective public safety measures, and the impact of the pandemic on the global economy and demand for consumer products and services.

Send a message



Figure: Chat Interface

Performance

Chart Correctness

To ensure the accuracy of our charts, a multi-step verification process was implemented. Initially, calculations were manually executed. This acted as a benchmark against which computer-generated charts could be compared. To augment the reliability of our results, two distinct coding methodologies were employed to generate charts: matplotlib and the python-pptx library. This was intentionally designed such that three separate individuals from our team managed each of the three segments: manual calculations, matplotlib charting, and python-pptx graphing.

This decentralization minimized the potential for personal errors, further buttressing the integrity of the outcome. Upon completion, a rigorous comparison of the three outputs was conducted. Finding them to be in alignment, we could confidently ascertain the consistency and correctness of our charts. This rigorous approach gives us profound confidence in the 100% accuracy of our implementation.

Summarization Quality

This pertains to the [Text Summarization](#) section outlined in the Methodology. To evaluate the quality of the summarized text, we conducted over 100 trials and scrutinized the produced content, confirming the absence of factual inaccuracies or extraneous material. The outcome achieved a perfect success rate of 100%. The primary factor contributing to this result was the nature of this segment, which exclusively entailed summarizing a predetermined text section. Consequently, the level of ambiguity or variability in summarizing identical text was exceedingly limited. This stood in stark contrast to the [AI Topic Selection](#) part, wherein all input contents were determined by AI before undergoing summarization.

AI Topic Selection Quality

We executed the code 40 times and examined both the generated topics and their corresponding content. Ultimately, we determined that approximately 32 instances, accounting for 80%, were not only pertinent but also carried significance. On the other hand, 8 cases fell short: among these, 4 contained placeholders for numbers such as X%, \$X, and [specific percentage increase]; 3 possessed content unrelated to the given company; and 1 exhibited an abnormal topic title.

Concerning content unrelated to the specified company, we attributed this issue to our implementation approach. Initially, we segmented the document into chunks while building the index. As a result, when interacting with AI, it only gained insight into specific details without a holistic understanding of the entire document. Consequently, its ability to produce satisfactory topics was compromised. This limitation also extended to the generated bullet points, which were often one-sided and detached from the comprehensive report. In instances where a topic had

infrequent occurrences within the report, AI's generated content tended to be generalized rather than offering specific insights about the company. This scarcity of company-specific information impeded the generation of substantial content, leading to the provision of more general facts about the topic itself.

We considered the presence of number placeholders in AI's responses as a manifestation of its lack of precision. Throughout the indexing process, AI occasionally overlooked critical details, particularly numerical statistics, resulting in the generation of ambiguous terminology to convey such information. We also had concerns that if the temperature parameter for the AI was set too high, it could exacerbate the problem by pushing the AI to prioritize excessive creativity over adhering to factual information derived from the context.

Topic 1: Expansion into Wearables Market

- ▶ In 2022, the company faced intense competition in the wearables market, with competitors offering lower-priced products and larger customer bases. This resulted in a decrease in product margins and potential revenue losses
- ▶ The company's ability to maintain its functional and design advantages in wearables is crucial for its financial success. However, competitors imitating the company's product features and applications pose a significant challenge, potentially impacting revenue growth
- ▶ While the company is focused on expanding its market opportunities in wearables, it must be cautious about competitors collaborating to offer integrated solutions that are more competitive than its current offerings. This could potentially lead to a loss of market share and revenue

Figure: Example of Good Topic Selection

Topic 1: Expansion into Wearables Market

- ▶ The company's expansion into the wearables market is a strategic move to tap into a growing market segment. According to the 10-K document, the wearables market has shown significant growth in recent years, with global shipments reaching X million units in 2021
- ▶ The company faces intense competition in the wearables market from companies with established hardware, software, and service offerings. It is crucial for Apple to differentiate its wearables products by leveraging its design and functional advantages to maintain market share and compete effectively
- ▶ The company's ability to continually improve its wearables products and services is critical to its success in the market. As competitors imitate the company's product features and applications, it is essential for Apple to invest in innovation and maintain its competitive edge to attract and retain customers

Figure: Example 1 of Inappropriate Topic Selection, "X" Used as Placeholder

Topic 2: Risk Factors and Legal Proceedings

- ▶ The Company's revenue is at risk due to the potential impact of the COVID-19 pandemic. The forward-looking statements in the 10-K document highlight the uncertainty surrounding future business operations and results
- ▶ The Company's financial condition and stock price can be materially and adversely affected by various factors, including those described in the risk factors section of the 10-K document. Past financial performance may not be a reliable indicator of future performance
- ▶ Investors should not rely on historical trends to anticipate future results or trends. The risk factors section emphasizes that other factors, in addition to those discussed, can affect the Company's results of operations and financial condition

Figure: Example 2 of Inappropriate Topic Selection, Unrelated or Too General

"Innovative"

- ▶ Total net sales for fiscal year 2022 increased by 8% or \$28.5 billion compared to 2021, driven primarily by higher net sales of iPhone, Services, and Mac products
- ▶ The weakness in foreign currencies relative to the U.S. dollar had an unfavorable year-over-year impact on all Products and Services net sales during 2022
- ▶ The company continuously introduces new product, service, and software offerings throughout the year, contributing to revenue growth and innovation

Figure: Example 3 of Inappropriate Topic Selection, Abnormal Topic Title

Overall Reliability

To assess the overall reliability of the PowerPoint generation feature, we executed our software over 100 instances, encompassing different years and template selections. The integration of auto-retry mechanisms across all AI-involved components significantly contributed to these outcomes, affirming a flawless execution rate of 100% for our software, ensuring the generation of the requested PowerPoint files without encountering crashes or any other errors.

Chat Quality

In order to assess the quality of the chat responses, we crafted a set of 70 test queries. These queries were thoughtfully designed to encompass a diverse array of question types, including numerical versus textual, varying scales, and even creative prompts. This comprehensive approach was undertaken to ensure an equitable appraisal of the overall effectiveness of our implementation. A selection of sample queries is provided as follows:

Question Type	Sample Question
Numerical	What is the value of iPhone sales in 2022?
Text, Small Scale	Why does sales in Greater China increase this year?
Text, Large Scale	How has COVID-19 affected the company?
Creative	Assume I am the CFO assistant. Write a 200-word internal email for me to the board of the company to inform them that the latest 10-K is released and what major updates they should expect. Use detailed examples.

Each response underwent manual scrutiny to ensure alignment with the specific question, factual accuracy in accordance with the document, and inclusion of a reasonable and pertinent amount of information drawn from one or more appropriate sections within the document. The findings indicate that 90% of the responses successfully met all of these criteria, surpassing our initial target (>60%).

The primary cause of unsuccessful responses stemmed from the inability to locate the pertinent information necessary to formulate an answer. This was particularly pronounced for numerical questions, where the inability to retrieve even a single numerical value resulted in overall failure. One major contributing factor to this difficulty was that certain details were so minute that they became indiscernible during the embedding transformation process. Two instances of such unsuccessful responses are provided below:

Sample Question	Response	Failure Reason
What is the value of vendor non-trade receivables in 2022?	The value of vendor non-trade receivables in 2022 is not provided in the given context information.	Inability to locate relevant info
What is the value of Cash and cash equivalents in 2022?	The value of cash and cash equivalents in 2022 is \$18,546 million.	Factually wrong / Inability to locate relevant info

To enhance comprehension of the outcome, we employed ChatPDF[4], a commercial product that enables users to pose questions based on any PDF document. Using this product as a benchmark, we achieved a success rate of 83% on the identical set of test queries. Our superior performance could be attributed to the meticulous fine-tuning of our parameters, including optimizing the chunk size for a specific document type (10-K financial reports).

Furthermore, we subjected our implementation to evaluation using an automated sanity check tool[8] provided by llama_index. This tool automatically generated 100 questions, or potentially

more if deemed necessary, by scanning the entire document. It then assessed the relevance of the responses to the questions and the contextual information selected based on those inquiries. The outcome revealed a success rate of 97%. It is important to note that this evaluation was not intended to serve as a definitive gauge of response quality, as it granted a passing score based on a minimal threshold of relevance, without delving into the actual accuracy of the responses. Rather, this exercise provided us with a certain level of confidence that our implementation is poised to effectively handle a broad spectrum of queries stemming from various sections of the document, including those that may not have been anticipated.

Conclusion

Finding

We successfully accomplish our defined scope. Our implemented features are operational, and we hold the belief that they offer a measure of benefit to users. However, our satisfaction with the scope delimitations is not particularly high. While we deliberately chose a narrower scope and a simpler implementation approach, this decision inadvertently constrains the applicability and utility of our solution. The constraints become evident in the following ways:

1. **Limited Applicability and Usefulness:** While we deliberately limited our focus to one company for straightforward implementation, this approach inherently confines the potential uses and real-world benefits of our solution. Expanding the scope to offer broader support for various companies demands a more sophisticated approach to information retrieval.
2. **Challenges in AI Decision Making:** The capabilities of current Language Model (LLM) technologies, like GPT, fall short when it comes to making decisive choices. This is most evident in the context of AI topic selection, where LLMs struggle to discern suitable topics, potentially due to limited exposure to finance-related materials. Even with efforts to enhance AI topic selection, difficulties persist in locating and extracting structural information—such as data tables—necessary for creating content like charts. Integrating AI-generated content with code-based processes presents inherent unpredictability and challenges.
3. **Dilemma of AI Creativity:** We face the conundrum of whether it's rational for our target users to leverage AI for creative tasks when they possess pre-existing domain knowledge. This is especially pertinent in finance-related tasks where expertise is essential. The uncertainty surrounding the potential worth of AI-generated creativity arises in contrast to decisions driven by human knowledge, particularly when we lack ample confidence in its quality, as indicated earlier.

In summary, our integrated functionalities effectively serve their designated objectives; however, the limitations introduce obstacles that may surpass the capacities of currently available AI solutions in the market. The difficulties highlight the intricate implementation of AI in sectors like finance, where individuals pay attention to specifics like accurate formatting and data extraction, all the while depending on experience-driven, comprehensive, and heuristic decision-making. It's plausible that a fusion of AI capabilities and human expertise remains indispensable for such specialized undertakings, at least for the time being.

Recommendation

1. Expanding the scope of supported filings to encompass 10-Q reports. While the inclusion of 10-Q filings could enhance the software's capabilities, our current decision is to confine the scope to 10-K filings. This choice is due to the greater consistency in format across samples and the richer textual data available within 10-K filings.
2. Enabling support for 10-K filings from diverse companies. However, it demands sophisticated document parsing techniques, given the variations in their formats. The intricacies of these formats prohibit the utilization of predetermined methods for section extraction. Instead, a more adaptable approach is required to dynamically extract relevant sections from these filings.
3. Improving the quality of AI topic selection. One possible strategy is to generate a short but comprehensive document summary initially as a pool of keywords and then employ AI to discern key topics. This method could potentially yield more targeted topic selections from evenly distributed areas of the document.
4. Employing AI to determine useful chart types. At present, the inclusion of charts along with their associated data sources is dependent on manual choices. This shift presents a comparable dilemma to that of topic selection. Furthermore, the unpredictability of the output format from current GPT-based AIs presents challenges in ensuring consistent data formatting for chart generation through code. Alternative libraries that employ natural-language prompts for chart generation, similar to ChartGPT^[1], could provide a solution, but this necessitates further exploration.
5. Ethical considerations, particularly in guaranteeing and quantifying the quality of AI-generated content. Presently, manual ratings serve this purpose. However, if the product transitions from an internal tool to a commercial solution encompassing multiple companies, ensuring the quality of AI-generated content becomes a paramount concern.

Reference

1. Whoiskatrin. (2023). Chart-gpt. <https://github.com/whoiskatrin/chart-gpt>
2. SEC API (2023). SEC API - A SEC.Gov EDGAR Filings Query & Real-Time Stream API. SEC EDGAR Filings API. <https://pypi.org/project/sec-api/>
3. OpenAI. (2023). ChatGPT (Aug 3 version) [Large language model]. <https://chat.openai.com/chat>
4. ChatPDF (2023). ChatPDF. <https://www.chatpdf.com/>
5. Liu, J. (2022). LlamaIndex. https://github.com/jerryliu/llama_index. doi:10.5281/zenodo.1234
6. Scanny. (2021). Python-pptx. <https://python-pptx.readthedocs.io/en/latest/>
7. SEC API (2023). SEC API - A SEC.Gov EDGAR Filings Query & Real-Time Stream API. Extractor API. <https://sec-api.io/docs/sec-filings-item-extraction-api>
8. Liu, J. (2022). Usage Pattern. LlamaIndex 0.8.3. https://gpt-index.readthedocs.io/en/latest/core_modules/supporting_modules/evaluation/usage_pattern.html