

Generative AI for Financial Analysis

Caden Webb

March 10, 2025

Abstract

This application explores an opportunity to apply recently developed natural language processing (NLP) techniques in the finance domain. Using a retrieval augmented generation (RAG) equipped LLM, a prototype is built to analyze complex financial documents such as earnings call transcripts and 10-K reports. The final product successfully leverages the vector storage to generate coherent, semantically rich responses, while using predetermined templates to elicit responses geared towards different customer segments.

Introduction

Artificial intelligence has made highly technical fields much more accessible by providing instruction and explanations to any level of understanding. Chat bots like ChatGPT and Claude can explain advanced concepts from fields such as calculus, molecular chemistry, and nuclear physics using analogies and explanations tailored to the user. Follow-up questions that would have required Ph.D. level expertise can be answered in seconds. Recent developments have improved the fidelity of responses using vector databases that can, essentially, cite the sources that they are drawing from to respond to queries. This application seeks to combine those two strengths, creating an application that can synthesize length financial documents into coherent summaries to answer questions posed by investors, executives, and analysts. The application will be backed by a retrieval augmented generation (RAG) system that will create a higher fidelity system.

Literature Review

Recent developments in generative artificial intelligence have improved accuracy and reduced so-called hallucinations by using subject-specific databases to augment prompts sent to

large language models (LLMs). One such method is RAG, which incorporates external knowledge into a vector embedding database. Queries are transformed into embeddings, which are subsequently compared to segments within the database. Sections with high similarity scores are then added to the prompt, which is then sent to the LLM for response generation. These databases are often formatted as knowledge graphs optimized for information storage and retrieval (Agrawal et al., 2023). Additionally, prompt engineering can make responses more accessible to users. One study showed an increase in the speed to prototyping for machine learning models and that application can extend to various other instances in the generative AI process (Jiang et al., 2022).

Methodology

The goal of this project was to prototype an application that can ingest financial documents and accurately synthesize them into text towards one of three target audiences: investor, analyst, and executive. Data was collected from publicly available data sources for 11 different companies: RTX, Apple, Nvidia, Google, Visa, Tesla, JP Morgan, Walmart, Amazon, Meta, and Microsoft. Earnings call transcripts from each quarter for the last seven to eight years were scraped from *The Motley Fool* website. Financial metrics were pulled from the Securities Exchange Commission (SEC) API, and text from 10-K reports was scraped from their website. Press releases were also collected from the websites of the companies as well as the *PR Newswire* website. The collection of data could be expanded in future efforts but was sufficient for this prototype given the computational limitations.

The text data from the 10-K reports was heavily pre-processed to remove HTML and XML tags. Even so, there was a lot of garbage from embedded headings, tables, and images left

in the text that could be improved, and this likely detracted from overall system performance and optimization. The earnings call transcripts were also processed to remove HTML tags and website advertisements, although this text was much cleaner than that of the 10-K data. The SEC metrics, 10-K reports, and earnings call transcripts were all stored in JSON files optimized for ingestion into the Facebook AI Similarity Search (FAISS) database that would be used for the RAG system. Chunking was performed to create 1000–2000-character chunks with some overlap, and the metadata containing the source, year, chunk number, and speaker (for earnings call transcript) was included for each chunk. Because the 10-K reports were so dense, additional pre-processing was performed to keep only the most relevant chunks. This approach used term frequency, inverse document frequency (TF-IDF) and keyword scores to identify the most interesting chunks for each document. Some keyword examples were “management discussion,” “financial statements,” and “business overview.” This trimmed some 2.5 million chunks to ~50,000 for 10-K reports. There were ~40,000 chunks from the earnings call transcripts, and the raw txt files were used for the press releases because they were relatively short.

After collection and pre-processing, the json and text files were transformed into numerical vectors via the SentenceTransformer library, which uses the all-MiniLM-L6-v2 model. The embedding model created a 384-dimensional embedding that was intentionally lightweight to work under the computational constraints imposed. The same embedding technique was used to transform the prompt into numerical vectors before querying the database and model. The FAISS library was then used to create an indexed database to organize the vectors for efficient search, and the resulting data structure was stored in Google Drive to be drawn from in the future. The final interface included an argument for the number of documents to be selected that would be used in the prompt.

The model used to generate responses was Mistral 7B, an open source LLM that is both powerful and accessible. After a user prompt and subsequent querying from the vector storage, a new prompt is engineered to elicit clear, grounded responses from the model. A template was created for each of the three paradigms (investor, analyst, executive), and a system and human prompt were created to instruct the model. The templates are shown below.

- Analyst
 - System: You are an AI assistant specialized in financial analysis. You analyze corporate financial data, SEC filings, earnings transcripts, and press releases to provide detailed analytical insights. Focus on trends, metrics, comparative analysis, and factual reporting. Use precise financial terminology and cite your sources clearly.
 - Human: “Analyze the following information about {companies} focused on {query}. Provide a comprehensive analysis with specific metrics, trends, and risk factors. Include inline citations to the source documents for all claims.”
- Executive
 - System: “You are an AI assistant that provides concise executive summaries of financial information. You analyze corporate financial data, SEC filings, earnings transcripts, and press releases to extract strategic insights. Focus on high-level implications, market positioning, and business impact. Be direct, concise, and action-oriented in your analysis.”
 - Human: “Provide an executive summary about {companies} focused on {query}. Highlight key strategic insights, competitive positioning, and business

implications. Be concise but thorough, and include inline citations to the source documents.”

- Investor
 - System: "You are an AI assistant that analyzes financial information for investors. You review corporate financial data, SEC filings, earnings transcripts, and press releases to provide investment-focused insights. Focus on performance metrics, future outlook, risk assessment, and potential investment implications. Be balanced in your assessment and cite your sources clearly."
 - Human: "Provide an investor-focused analysis of {companies} regarding {query}. Address performance trends, growth potential, risk factors, and competitive position. Include relevant metrics and inline citations to the source documents.”

These templates would be augmented by the chunks pulled from the vector database, and that prompt would then be tokenized and sent to the Mistral model for evaluation.

A simple user interface (UI) was built using Streamlit for accessibility. A prompt to initiate the system would be displayed, then a chat box to type in a question with an example included. The output would show the response from the model along with some plots of documents with the highest similarity scores. A section at the bottom displayed each source and the chunk of context that was used to create the response. A text file to log chat history was also created to document questions and responses.

Results

The application was able to successfully process user prompts and return semantically rich responses grounded in the context provided. Queries were not executed particularly quickly, and the vector database took a while to initialize, but the responses were grammatically sound and generally held some information that matched source material. There were some cases where the chunks pulled were not very rich, containing titles, headers, or footers, so data engineering and pre-processing is still an area for improvement. In general, the earnings call transcripts seemed to give the best context for responses, while 10-K text was occasionally inconsistent or irrelevant. The system did not directly cite its sources in the response each time but generally had some reference to the material it was drawing from. There were also some formatting issues that seemed to come from the HTML tags that were not totally removed from the raw text and attempts to debug that issue were futile. The graphs shown were not particularly informative either, although there is potential for integration with the SEC metrics, which were not generally used as context. That would be part of a larger opportunity for more quantitative insights. Historical financial performance coupled with future forecasts could be added to enrich the qualitative aspect of the LLM responses. A sentiment analysis, specifically of the text from the earnings calls, could also provide some interesting insight into how a company is trending. While there are currently many services that provide similar information, the ease of access for any level of financial experience and ability to cite sources make this type of application unique and valuable; however, further data collection and processing efforts would need to be made to scale the product. Images of the UI and some examples of prompts and responses will be shown below.

Configuration

System Initialized

Query Settings

Select Audience

Financial Analyst

Number of documents to retrieve

520

System Info

Queries processed: 1

Last query: 22:38:45

Reset System

Financial Analysis System

Ask a financial question

Enter your query:

What are Amazon's biggest growth opportunities and challenges over the next five years?

Submit Query

Response

Based on the provided context information, Amazon has several growth opportunities and challenges over the next five years.

Growth Opportunities:

1. AWS (Amazon Web Services): AWS continues to be a significant growth driver for Amazon. The company reported 18.8% year-over-year revenue growth in Q2 2021, and this trend is expected to continue as companies focus on modernizing their infrastructure and moving from on-premise infrastructure to the cloud. AWS also announced significant customer momentum with new commitments and migrations in various industries such as finance, media, and technology.

Response

Based on the provided context information, Amazon has several growth opportunities and challenges over the next five years.

Growth Opportunities:

1. AWS (Amazon Web Services): AWS continues to be a significant growth driver for Amazon. The company reported 18.8% year-over-year revenue growth in Q2 2021, and this trend is expected to continue as companies focus on modernizing their infrastructure and moving from on-premise infrastructure to the cloud. AWS also announced significant customer momentum with new commitments and migrations in various industries such as finance, media, and technology.

2. International Expansion: Amazon has been expanding its presence internationally, particularly in its stores business. The company sees potential for growth in large retail market segments in which it is still nascent, such as grocery and business-to-business. Additionally, Amazon's low Earth orbit satellite could provide connectivity to millions of households, companies, and government entities with limited to no connectivity.

3. Prime Offerings: Amazon's Buy with Prime program allows consumers to use Prime to purchase products and services from third-party sellers. This offering provides an opportunity for Amazon to expand its reach and increase customer engagement.

4. Entertainment and Devices: Amazon has been investing heavily in entertainment and devices, including the acquisition of Twitch and the development of its own smart speakers and streaming service. These investments could lead to increased revenue and customer engagement.

5. Healthcare: Amazon has been exploring opportunities in healthcare, including the launch of its own healthcare insurance plan. This could provide a significant growth opportunity for the company if successful.

Challenges:

1. Cost Optimization Efforts: While AWS revenue growth has been strong, the company has taken several actions to streamline its costs. This includes eliminating about 27,000 corporate roles. Amazon will need to balance cost optimization efforts with strategic long-term investments that can improve the customer experience.

2. Competition: Amazon faces competition from other e-commerce giants such as Walmart and eBay. Additionally, the company may face increased regulatory scrutiny and antitrust investigations.

3. International Regulatory Risks: Amazon's international expansion may face regulatory risks, particularly in countries with strict regulations around data privacy and consumer protection.

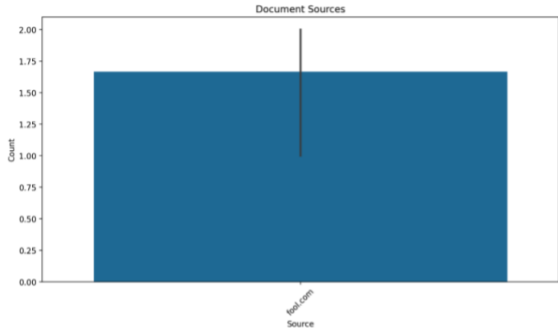
4. Economic Uncertainty: The global economic environment remains uncertain due to the ongoing COVID-19 pandemic and geopolitical tensions. This uncertainty could impact Amazon's growth prospects.

5. Technological Advancements: Amazon will need to continue innovating and adapting to new technologies to maintain its competitive edge in the market.

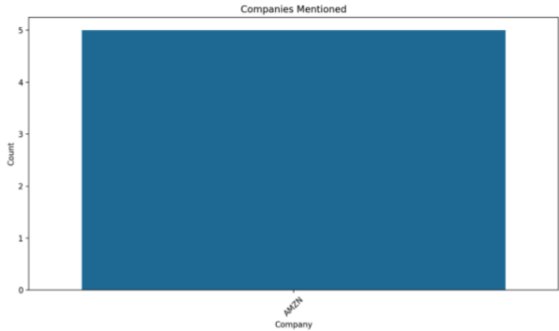
Overall, Amazon has several growth opportunities and challenges over the next five years. The company will need to balance cost optimization efforts with strategic investments that can improve the customer experience while navigating regulatory risks, economic uncertainty, and technological advancements.

Insights from Retrieved Documents

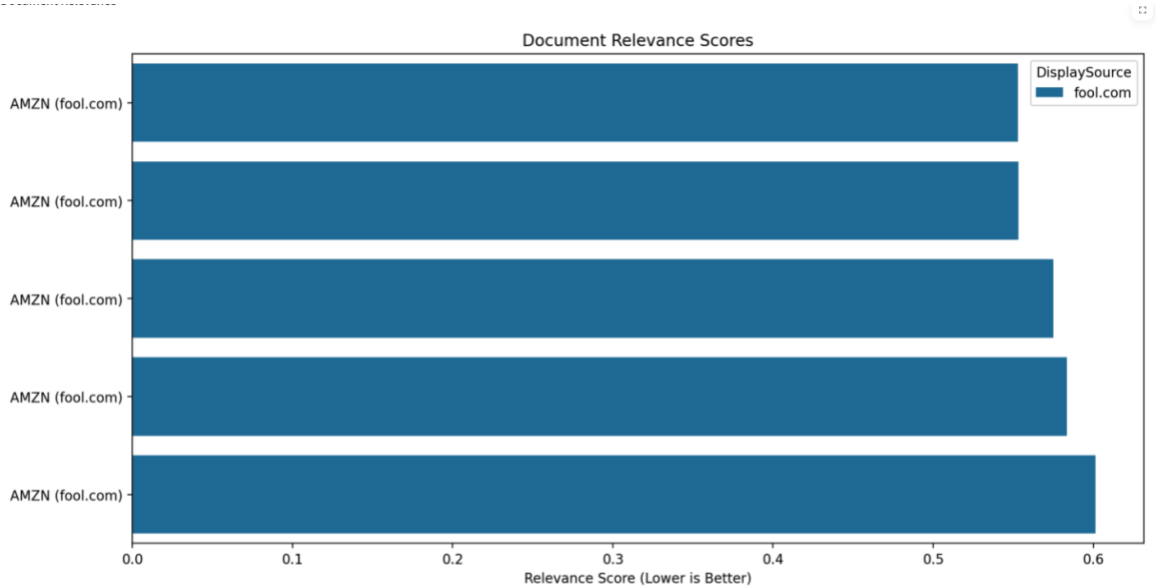
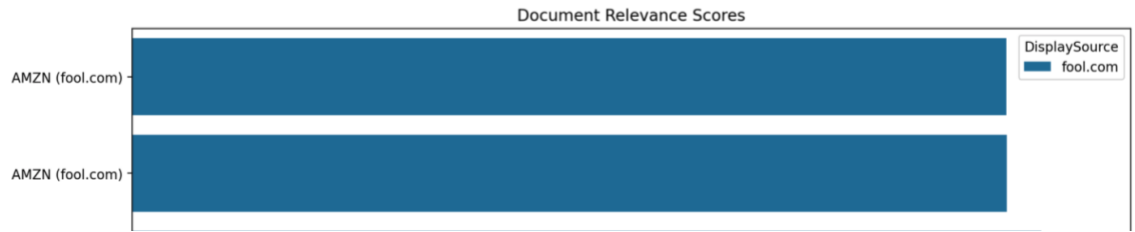
Document Sources



Companies Mentioned



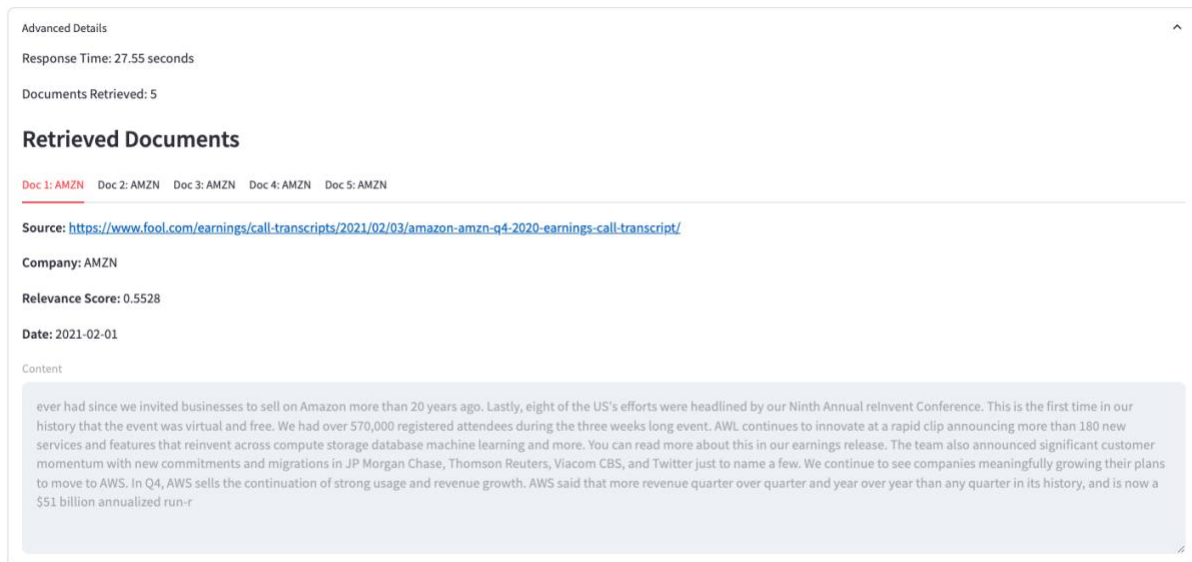
Document Relevance



Advanced Details

Response Time: 27.55 seconds

Documents Retrieved: 5



The above images shows an example of how the retrieved documents can be displayed to see where the source material is coming from. The prompt shown above was run with a varying number of documents and from different perspectives to better understand how the system is processing the queries. An example of how the prompt is constructed is below.

- User input: What are Amazon’s biggest growth opportunities and challenges over the next five years?
- Audience: Analyst.
- System prompt: “You are an AI assistant specialized in financial analysis. You analyze corporate financial data, SEC filings, earnings transcripts, and press releases to provide detailed analytical insights. Focus on trends, metrics, comparative analysis, and factual reporting. Use precise financial terminology and cite your sources clearly.”
- User prompt: Analyze the following information about AMZN focused on What are Amazon’s biggest growth opportunities and challenges over the next five years?. Provide

a comprehensive analysis with specific metrics, trends, and risk factors. Include inline citations to the source documents for all claims.”

- Context included 5 different chunks from the 2020, 2023, and 2024 earnings calls transcripts.
- Response: “Based on the provided context information, Amazon has multiple growth opportunities and challenges over the next five years.

Growth Opportunities:

1. AWS: Amazon Web Services (AWS) is one of Amazon's largest growth drivers. The company reported a year-over-year revenue growth rate of 18.8% in Q2, and this trend is expected to continue due to several macro trends such as companies focusing on new efforts after completing cost optimization efforts and modernizing their infrastructure. Additionally, AWS continues to attract new customers, including major corporations such as JP Morgan Chase and Thomson Reuters.
2. International Expansion: Amazon has been expanding internationally and has announced plans to open more warehouses in Europe. This expansion presents a significant growth opportunity for the company as it taps into new markets and increases its customer base.
3. Large Retail Market Segments: Amazon is still nascent in certain large retail market segments, such as grocery and business-to-business. The company has been investing heavily in these areas, and this presents a growth opportunity as it expands its product offerings and reaches new customers.
4. Entertainment: Amazon has been investing heavily in the entertainment industry, acquiring content creators and streaming services. This presents a growth opportunity for the company as it continues to expand its entertainment offerings and reach new customers.
5. Healthcare: Amazon has been exploring opportunities in the healthcare industry, including telemedicine and prescription drug delivery. This presents a growth opportunity for the company as it enters a new market and reaches new customers.

Challenges:

1. Competition: Amazon faces intense competition from other e-commerce giants such as Walmart and eBay. This competition could limit Amazon's growth potential and profitability.
2. Regulatory Risks: Amazon has faced regulatory scrutiny in various countries, including the United States. Any changes in regulations could impact the company's operations and growth potential.
3. Cybersecurity Risks: Amazon has faced cybersecurity threats in the past, and any future incidents could damage the company's reputation and impact customer trust.
4. Supply Chain Risks: Amazon relies heavily on its supply chain to deliver products to customers. Any disruptions in the supply chain could impact the company's ability to meet demand and maintain profitability.
5. Economic Risks: Economic downturns could impact consumer spending and reduce demand for Amazon's products and services.

Overall, Amazon has multiple growth opportunities and challenges over the next five years. While the company has been experiencing strong growth, it faces competition, regulatory risks, cybersecurity risks, supply chain risks, and economic risks that could impact its growth potential and profitability.”

This is a clear response to the provided prompt. To compare, here are the response from the **executive** and **investor** perspectives with the same prompt:

- **Executive:** “Amazon (AMZN) has reported strong revenue growth and profitability in Q4, driven by continued demand for e-commerce and cloud computing services. In Q4, AWS saw year-over-year revenue growth of 18.8%, driven by three macro trends: cost optimization efforts, modernization of infrastructure, and migration from on-premise infrastructure to the cloud. The company has invested heavily in its larger businesses, such as e-commerce and cloud computing, as well as in areas like international expansion, large retail market segments, and low Earth orbit satellite for limited connectivity. However, the company has also taken actions to streamline costs, including eliminating about 27,000 corporate roles. Despite these efforts, Amazon is still pursuing key strategic long-term investments that it believes can make customers' lives better and potentially change what Amazon is.”
- **Investor:** “Amazon (AMZN) has been one of the most dominant players in the e-commerce industry for over 20 years, and it has continued to grow its presence in the market through innovation and diversification of its business model. In this analysis, we will focus on Amazon's biggest growth opportunities and challenges over the next five years, considering performance trends, growth potential, risk factors, and competitive position.

Growth Opportunities:

1. Continued Expansion in E-commerce: Amazon's core business model revolves around e-commerce, and it continues to dominate the market with its vast selection of products and services. With the ongoing shift towards online shopping, Amazon has immense growth potential in this segment. The company has already expanded into various product categories such as groceries, household essentials, and clothing. However, there is still room for further expansion into niche markets and geographies.
2. Growing AWS Business: Amazon Web Services (AWS) has been a significant contributor to the company's growth over the past few years. AWS provides cloud computing services to various industries, enabling them to store, process, and analyze large amounts of data. As more businesses migrate to the cloud, AWS has significant growth potential. Additionally, AWS is constantly innovating and expanding its offerings, providing new services and features that cater to different industries and use cases.
3. International Expansion: While Amazon has made significant strides in its international expansion, there is still considerable growth potential in emerging markets.

The company has faced challenges in some regions due to regulatory hurdles and cultural differences. However, as the global e-commerce market continues to grow, Amazon has an opportunity to capitalize on this trend by expanding its operations in these markets.

4. Diversification of Business Model: Amazon has been diversifying its business model beyond e-commerce and AWS. For example, the company has entered the entertainment industry through its streaming service, Prime Video, and has ventured into the smart home market with devices like Echo and Ring. Furthermore, Amazon has been exploring opportunities in healthcare, with initiatives like Amazon Pharmacy and its partnership with PillPack. By continuing to diversify its business model, Amazon can mitigate risks associated with its core e-commerce and AWS businesses and tap into new growth opportunities.

Challenges:

1. Competition: Amazon faces intense competition from various players in the e-commerce and technology industries. Walmart, eBay, and other traditional retailers have been adapting to the online market, while new entrants like Alibaba and JD.com have gained significant market share in China. Additionally, tech giants like Apple and Google are expanding their own e-commerce platforms, further intensifying the competition. To remain competitive, Amazon must continue to innovate and offer unique value propositions to its customers.

2. Regulatory Risks: Amazon operates in multiple jurisdictions, and regulatory risks can significantly impact its business. For instance, antitrust investigations in Europe and the United States have raised concerns about the company's dominance in the e-commerce market. Moreover, changes in trade policies and tariffs can affect Amazon's supply chain and pricing strategies. The company must navigate these regulatory risks carefully to avoid potential legal issues and maintain its reputation.

3. Operational

The response parameter was set to a maximum of 700 tokens, so the last response was cut off before finishing. The response time was also related to the selected paradigm, as the executive perspective took about 9 seconds while the investor and analyst took 30 seconds and 26 seconds respectively. The number of documents selected for retrieval also affected the response time. The executive response successfully generated a shorter, concise overview of potential challenges, while both the analyst and investor views looked at risks, competition, and potential new market opportunities from slightly different lenses. While it does not explicitly reference the earnings reports that the source data is coming from, the numbers seem to accurately reflect the context

that is provided, and the answer seems to be grounded. There could be further refinement to the prompts and roles, as well as more explicit requests for a length suitable to the token output allowed. The investor and analyst responses were both quite verbose. The accompanying text file shows record of 10 questions asked from each perspective, and the last question was asked requesting several different documents.

Conclusion

Overall, the prototype successfully accomplishes the stated goal of responding to different requests for financial information. While scaling would require improved infrastructure, more data, and larger models, the general system can still be applied to query the vector database and “show the work” when pulling metrics that could be used to decide on what investments to make. Generally, the data engineering was the most difficult and onerous part of this project. Having adequately clean and pre-processed data is very important and doing that with web scraped data is difficult to accomplish. Chunking the data for upload to the vector store could also be improved. There were many chunks that did not contain information and would occasionally get called by the similarity scores because they would have the company name or “earnings” in the title. There is also potential to include more quantitative data from the SEC database in the responses and UI, something that would be valuable to casual analysts and investors. There was not much experimentation with the model and embedding parameters either, which could be tested to improve the responses. This was a sound foundation for a much broader system that could specialize in financial analysis.

Bibliography

- Agrawal, Garima, Tharindu Kumarage, Zeyad Alghamdi, and Huan Liu. "Can knowledge graphs reduce hallucinations in llms?: A survey." *arXiv preprint arXiv:2311.07914* (2023).
- Jiang, Ellen, Kristen Olson, Edwin Toh, Alejandra Molina, Aaron Donsbach, Michael Terry, and Carrie J. Cai. "Prompt-based prototyping with large language models." In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*. 2022.
- Mistral AI. 2025. *Mistral 7B: An Open-Weight 7 Billion Parameter Language Model*. Accessed March 14, 2025. <https://mistral.ai/news/announcing-mistral-7b/>.
- PR Newswire. 2025. *PR Newswire*. Accessed March 14, 2025. <https://www.prnewswire.com/>.
- The Motley Fool. 2025. *The Motley Fool*. Accessed March 2, 2025. <https://www.fool.com/>.

