Leveraging AI and Vector Databases for User-Driven Document Analysis and Interactive Chart Transformations

1 Introduction

Extracting and interpreting information from documents containing various charts often demands considerable time and relies significantly on a reader's prior knowledge and background. The state-of-the-art paper by [3] highlights the critical need for chart questions and answer (CQA) systems capable of synthesizing text and visuals for comprehensive insights, and provides problem space as shown in Figure 1.

Various research is being done that attempts to fill this gap through the integration of LLM and techniques to synthesize texts and/or chart data that enable users to interact at various levels and depths. Recent papers include methodologies for extracting chart data from diverse sources such as embedded SVGs in PDFs, screenshots, and user uploaded images [6, 7, 2]. Also, so and so worked on application design that integrates various APIs to ensure the right fine-tuned LLM is used for proper interpretation and insights drawn from those user-uploaded chart data However, we see some gaps in terms of the reliability of information that mostly comes from the LLM's general knowledge.

Our project builds upon contributions of these researchers (and more) and attempts to add value by integrating a vector store database, embedding implementation for semantic search, and ensuring reliability. This aspect represents only a fraction of our overall contribution.

In this experiment, we make the following contributions:

- Chart Transformation: Allow user to request chart transformation to not only enhance understandings but also facilitate specific analytic needs.

- User-Agent: We develop an application that offers a dynamic interface for user interaction with text and charts in user-uploaded documents, enabling insightful queries and analysis.

- Modification: Through conversational user commands, enable precise modification of document content, including chart transformation and text rewriting.

- LLM and database: Integration of large language models (LLMs) with vector database to ensure precise data retrieval from the document's own data, prompting reliability. [Scalability]

2 One-sentence description

Utilizing large language models and query techniques with vector databases, this project offers a dynamic interface where users can interactively query and customize text and charts from uploaded documents, ensuring authoritative insights are directly derived and enhanced from the document's own data.

3 Project Type

Application development, Visualization, interactivity, productivity

4 Audience

> *Who is the audience for this project? How does it meet their needs? What happens if their needs remain unmet?*

This project's audience is any individual or professional who wants to understand visualizations more deeply in the context of the surrounding text. Processing uploaded data through a large language model allows users to ask questions of existing visualizations and transform them into different representations that may yield hidden trends and results.

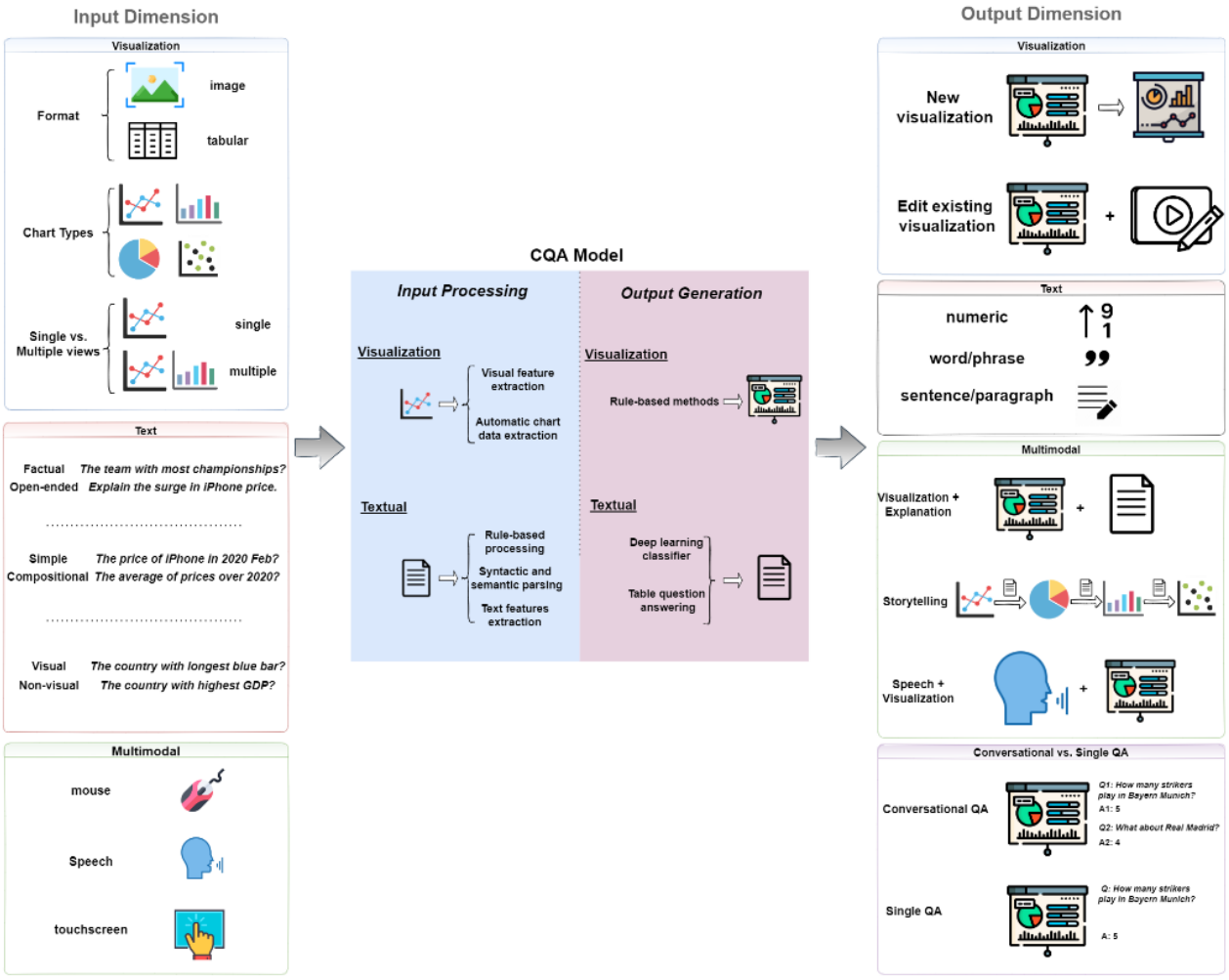

Figure 2: *An overview of the problem space of chart question answering, covering key categories of input and output dimensions* . A particular CQA problem setup may involve one or more categories of *input* and one or more categories of *output* dimensions.

Figure 1: Overview of problem space

5 Approach

The application will be designed using existing and emerging techniques relating to LLM, databases, embedding, and information retrieval techniques to offer an interactive visualization and information analysis tool.

5.1 Dataset

This application do not require a large dataset. We intent to apply existing techniques and libraries for information retrieval, processing, and storage.

- **PDFs**: retrieval techniques based on [6]
- **Create our own ??**: Generate articles with context and use of chart (bar, boxplot, line chart, etc) [2]?
- **Online articles (Not planned yet)**

5.2 Details

What is your approach?

Figure 2 shows a concept diagram of our for prototyping efforts.

5.3 Evidence for Success

Why do you think it will work?

Leveraging existing resources for prototyping and building up an existing pipeline already laid on a similar AI4VIS project.

6 Best-case Impact Statement

In the best-case scenario, what would be the impact statement (conclusion statement) for this project?

This project harnesses the capabilities of Large Language Models (LLMs) alongside current and evolving techniques for data extraction and interaction, focusing on both textual and visual elements in documents. It empowers users to explore, understand, and personalize text and graphics, thereby unlocking deeper insights.

Moreover, the framework sets the foundation for broader applications, including integration with productivity tools, cross-linguistic functionalities, and beyond.

7 Major Milestones

- **Prototyping:** Focused on ensuring the functionality of data extraction, storage, sequencing, and integration with LLMs, specifically targeting text and chart data using established techniques and libraries.
- **Conversational Chain and Prompt Engineering:** Progressing from prototyping, the aim is to develop a context-aware agent proficient in maintaining conversational history and context.
- **UI and UX Design:** Initial efforts have produced a straightforward UI layout, showcasing the document on the left and interactions with the AI agent and its responses on the right, with ongoing enhancements to both design and functionality as shown in figure 3
- **Finalization and Deployment:** Complete the development cycle by finalizing the product and deploying it for user access.
- **User Study Preparation:** Post-deployment, prepare the application for extensive user testing to gather feedback and insights.

Figure 3 shows a concept diagram of our interactive user interface.

8 Obstacles

8.1 Major obstacles

- Technical challenges with the integration of various libraries, database integration, and achieving desired output

8.2 Minor obstacles

- Expanding from simple text based document and chart types to a wide array of documents and chart types.

9 Resources Needed

What additional resources do you need to complete this project?

- OPENAI credits

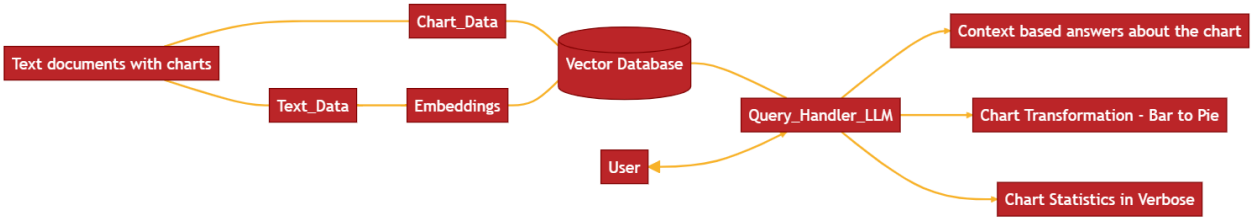

Figure 2: High-level overview of pipeline

10 Related Work

List 5 major publications that are most relevant to this project, and how they are related (sample citation).

Paper that described ongoing questions on answering questions about charts → Hoque et al.'s work [3] is a state-of-the-art report on open research questions relating to answering user questions about charts. One major open question is how answers can combine both text and chart visualizations to help users better understand key points about their data.

Papers that studied combination of visual chat and image input to generate multimodal response → The work of Chen et al. [1] shows how LLaVA-interactive combines pre-built AI models for visual chat, image segmentation, and image generation/editing, without the need for additional model training. The system supports multi-turn dialogues with users, taking multimodal inputs and generating multimodal responses. It goes beyond language prompts by enabling visual prompts to align human intents in the interaction.

Han et al.'s work [2] described how datasets and corresponding visualizations can be created by prompting GPT-4 with proper parameters, bounds, and chart types. These datasets and visualizations can be analyzed using ChartLlama, a fine-tuned LLaVA-1.5 model that can answer Q&A questions about the provided data.

Masson et al.'s work [5] introduces Charagraphs to enhance reader engagement with numeric data in texts by providing dynamic, interactive charts and annotations for in-situ visualization and manipulation. This innovation not only aids in data interpretation and insight discovery but also paves the way for applying these visualizations to physical documents via OCR and augmented reality.

Ko et al.'s work [4] proposes the VL2NL (Vega-lite to Natural Language) framework to facilitate the creation of diverse natural language datasets for data visualization, significantly advancing the development of Natural Language Interfaces (NLIs) by synthesizing chart semantics and generating accurate captions. This innovation not only enhances accessibility for individuals with visual impairments but also promises to improve NLI research through its ability to generate a wide range of utterances and its potential for generalization across various types of NL datasets.

Peng et al.'s work [6] demonstrated how domain-specific APIs can be used to interpret chart input in the proper context. They explain how graphs can be analyzed with natural language interaction and how these interactions permit changes to the underlying labels and data of a chart.

The work of Wang et al. [7] implements how LLM can be designed specifically for financial chart analysis using FinViz-GPT. It surpasses other multimodal language models in tasks such as chart description, answering finance-related questions, and forecasting market trends.

These studies connect how visualizations can be interpreted by LLMs to provide additional context to the user. However, they do not elaborate on how surrounding text can be utilized to provide further context to these visualizations. Our project aims to build on user queries to an LLM to combine text from uploaded documents to derive additional context for provided visualizations.

11 Define Success

What is the minimum amount of work necessary for this work be publishable?

For this work to be publishable, we need:

- A fully functional prototype with preliminary testing results is the immediate goal.

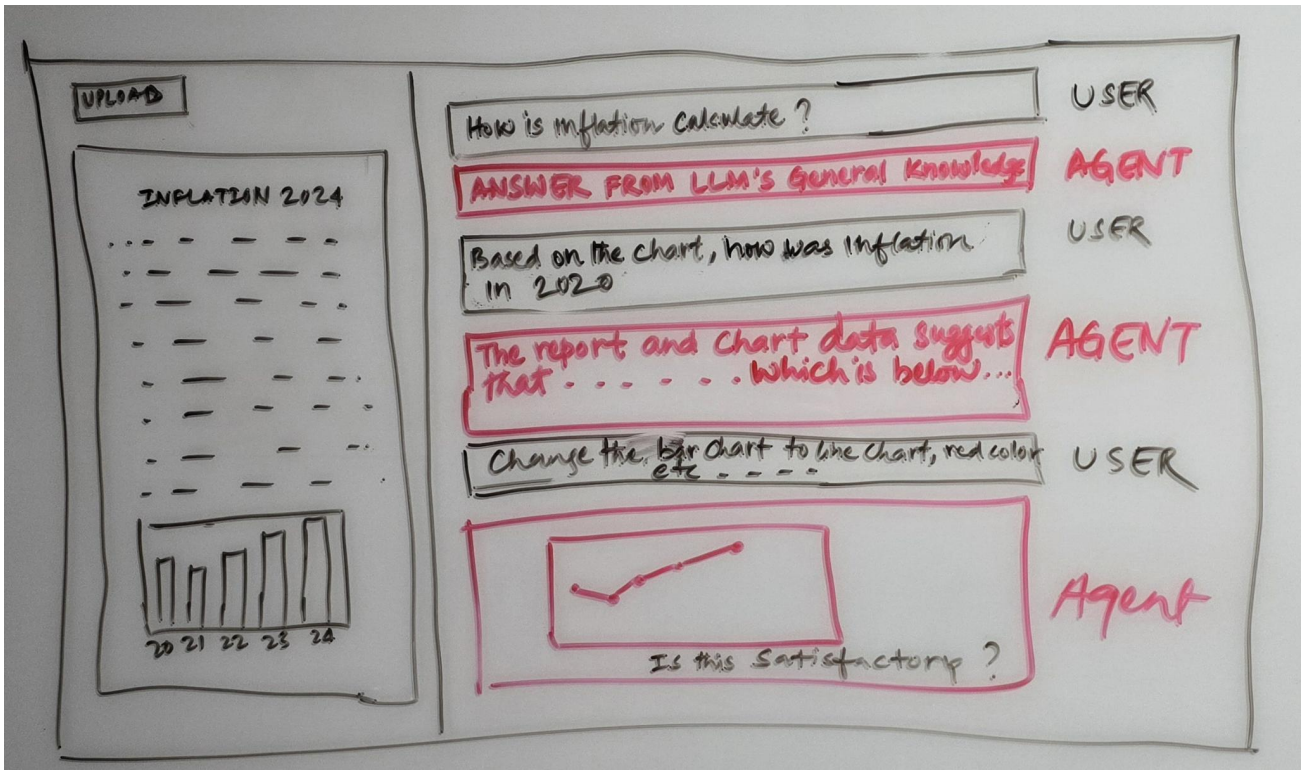

Figure 3: Concept UI for prototype

References

[1] W.-G. Chen, I. Spiridonova, J. Yang, J. Gao, and C. Li. Llava-interactive: An all-in-one demo for image chat, segmentation, generation and editing. *arXiv preprint arXiv:2311.00571*, 2023.

[2] Y. Han, C. Zhang, X. Chen, X. Yang, Z. Wang, G. Yu, B. Fu, and H. Zhang. Chartllama: A multimodal llm for chart understanding and generation. *arXiv preprint arXiv:2311.16483*, 2023.

[3] E. Hoque, P. Kavehzadeh, and A. Masry. Chart question answering: State of the art and future directions. In *Computer Graphics Forum*, volume 41, pages 555–572. Wiley Online Library, 2022.

[4] H.-K. Ko, H. Jeon, G. Park, D. H. Kim, N. W. Kim, J. Kim, and J. Seo. Natural language dataset generation framework for visualizations powered by large language models, 2023.

[5] D. Masson, S. Malacria, G. Casiez, and D. Vogel. Charagraph: Interactive generation of charts for realtime annotation of data-rich paragraphs. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–18, 2023.

[6] Y. Peng, S. Lin, Q. Chen, L. Xu, X. Ren, Y. Li, and J. Xu. Chatgraph: Chat with your graphs. *arXiv preprint arXiv:2401.12672*, 2024.

[7] Z. Wang, Y. Li, J. Wu, J. Soon, and X. Zhang. Finvis-gpt: A multimodal large language model for financial chart analysis. *arXiv preprint arXiv:2308.01430*, 2023.