LLM-Powered Visualizations and Narratives from Natural Language Queries

Eduardo Roger S. Nascimento[1]

[1]Departamento de Informática – Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)
Caixa Postal 38.097 – 22.453-900 – Rio de Janeiro – RJ – Brazil

Abstract. *The gap between non-technical users and tabular data limits effective access and data-driven decision-making. This work presents an agent-based workflow that integrates Text-to-VIS and data storytelling using Large Language Models (LLMs). Given a natural language query and a dataset, the system generates Python code to produce charts and provides narrative insights to support interpretation. Using GPT-4o, we demonstrate the approach through three case studies on a movie dataset. Results show that LLMs can not only generate appropriate visualizations but also assist users in understanding patterns and key findings, bridging the gap between data visualization and storytelling.*

1. Introduction

Data visualization plays a vital role in enabling users to understand complex data through graphical representations. Traditionally, creating effective visualizations required proficiency in data analysis and programming. Recently, the advent of Large Language Models (LLMs) such as ChatGPT[1] and the exploration of prompt engineering techniques have opened new avenues for developing more effective and user-friendly natural language interfaces (NLI) for data interaction [Zhang 2024]. With their ability to leverage pre-training on vast amounts of text data, LLMs have shown remarkable success in a wide range of natural language processing tasks, including Text-to-Vis tasks. Text-to-VIS systems allow users to generate data visualizations though charts or plots by simply describing their intent in natural language (NL). These systems leverage LLMs to translate NL queries into declarative visualization languages, such as Python[2] with Matplotlib [Hunter 2007], Vega-Lite [Satyanarayan et al. 2023] and ECharts [Li et al. 2018].

While Text-to-VIS systems focus on generating visual representations, they often lack the narrative element that helps users interpret and communicate the insights derived from data. This is where data storytelling becomes essential. More than just displaying data clearly and efficiently, data storytelling is a methodology for conveying key insights to a specific audience, aimed at informing decisions and guiding action. It involves structuring the analytical message as a narrative that enhances the communication and dissemination of findings within and across groups [Feijó 2019]. Recent research and applications have explored the potential of LLMs not only to generate charts but also to

[1] http://openai.com/pt-BR/chatgpt/overview/
[2] https://www.python.org/

act as narrators, this way, producing descriptive text that explains the visualizations, answers specific analytical questions, or even mimics expert commentary [Danqing et al. 2021; Ma et al. 2023; Sun et al. 2023].

In this work, we explore the integration of Text-to-VIS and data storytelling using LLMs in a unified workflow. We present a workflow that combines chart generation with automated narrative generation, powered by GPT-4o, to support intuitive, insightful, and engaging data exploration. Through a series of case studies, we demonstrate how LLMs can be used not only to construct relevant visualizations but also to generate compelling explanations tailored to the user's analytical goals.

2. Related Work

Text-to-VIS task has also gained attention with the rise of LLMs. Early work, such as Data2Vis [Dibia and Demiralp 2019], trained models to generate Vega-Lite visualizations from structured data and textual input. Tools like Vanna AI [Vanna 2024], PandasAI [PandasAI 2024] and LIDA [Dibia 2023], offer streamlined pipelines that transform user queries into data retrieval operations and basic visualizations. Chat2VIS [Maddigan and Susnjak 2023] applies prompt engineering to define prompts capable of understanding user queries and generating python code of the corresponding visualizations. Other strategies have also been introduced more recently. The framework Prompt4vis [Li et al. 2025] leverages LLMs and in-context learning to enhance the generation of data visualizations from natural language.

In parallel with advances in Text-to-VIS, several works have focused on integrating data storytelling into visualization workflows to enhance interpretability and engagement. “Calliope” [Danqing et al. 2021] presents an automated system that transforms spreadsheet data into visual data stories by extracting relevant facts, generating charts, and composing narrative sequences. “Erato” [Sun et al. 2023] enables human-machine collaboration by letting users define keyframes of a story, while the system interpolates intermediate steps to create smooth transitions between them. “XInsight” [Ma et al. 2023] adopts a causality-driven approach to explain data insights, combining exploratory data analysis with causal reasoning to produce semantically rich narratives. These tools move beyond static visualizations, aiming to structure data-driven insights as coherent, interpretable stories that support better decision-making.

3. Methodology

The architecture of system is implemented using LangChain[3] and LangGraph[4] , which enable the creation of modular and agentic workflows based on LLMs. As illustrated in Figure 1, the system is composed of four primary components: *Clarify Question, Human Feedback, Text-to-VIS*, *and Data Storyteller*. Each component is modeled as a node within a directed workflow graph, where transitions are governed by predefined paths but conditionally triggered based on the output of LLM decisions.

Clarify Question The system first evaluates the clarity of the user’s natural language query using an LLM, following the LLM-as-a-Judge paradigm [Galileo 2025]. If the query is ambiguous or underspecified, a clarification step is triggered. Otherwise, the workflow proceeds directly to the Text-to-VIS component.

---

[3] https://www.langchain.com/
[4] https://www.langchain.com/langgraph

Human Feedback: When clarification is needed, the system prompts the user for additional input to resolve ambiguities. This step ensures accurate query interpretation when LLM inference alone is insufficient.

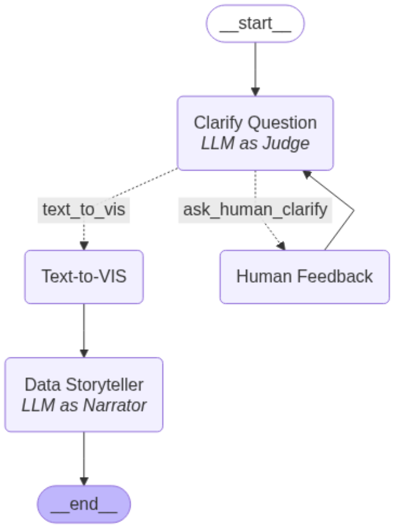

Figure 1. Workflow of the proposed system

Text-to-VIS: In this step, the visualization strategy is executed, producing a complete Python code snippet responsible for rendering a specific chart (e.g., bar, line, scatter). Vanna AI [Vanna.AI 2024] was a strategy used to generate the data visualization.

Data Storyteller: In the final stage of the workflow, the generated visualization is passed to the *Data Storyteller module*. Here, the LLM assumes the role of narrator, transforming the visualization into a natural language narrative. LLM was guided using information adapted from [Feijó 2019], enabling it to contextualize and communicate key patterns, trends, and insights in a way that is accessible and meaningful to end users.

To evaluate our approach, we conducted three case studies using the GPT-4o model provided by OpenAI [OpenAI 2024], configured with a temperature of 0 to ensure deterministic responses. We used the *Movies*[5] dataset as the sole data source for all case studies. This dataset contains metadata on movies released between 1996 and 2010, including financial information (e.g., gross income and budget), ratings, genres, and release year.

4. Results

4.1. Case Study 1: *"number of movies released by year"*

The first case explores a question that aim to analyze how LLMs handle time-based data requests. The model generated a bar chart (see Figure 2) that clearly shows the annual number of movie releases from 1996 to 2010. Regarding the narrative generated by LLM, it provided a title ("*The Evolution of Movie Releases Over Time*"), contextualized the trend, and identified key insights: *Growth* (1996–2000), *Stability* (2001–2007), and *Sharp Decline* (2008–2010). It framed the decline as a conflict ("*A Noticeable Decline in Recent*

---

[5] https://github.com/nl4dv/nl4dv/blob/master/examples/assets/data/movies-w-year.csv

Years”) and proposed causes (e.g., economic crisis, technological shifts, industry saturation).

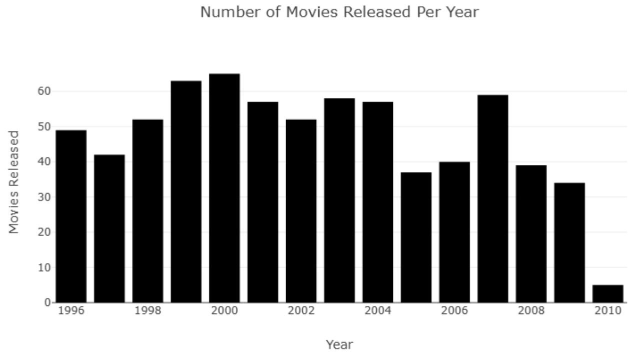

Figure 2. Case Study 1.

4.2. Case Study 2: *“Which movie genres have the highest average gross income?”*

The second case investigates how LLMs generate visual summaries from categorical and aggregated numerical data, focusing on average gross income grouped by movie genre. *Text-to-VIS Tool* produced a bar chart ranking genres and autonomously sorted the chart by average gross income (without explicit prompting) as shown in Figure 3, revealing Adventure as the top-performing genre. Its narrative, titled “*The Power of Genre in Box Office Success*”, framed the revenue disparity as a conflict ("*The Uneven Playing Field of Genres*") and highlighted key insights: Adventure’s dominance (blending action/fantasy), strong performance of Action/Musicals, and niche genres’ struggles. Notably, the LLM dedicated a paragraph to hypothesize why Adventure reigns (citing broad demographic appeal and high production value) demonstrating its ability to generate interpretive analysis.

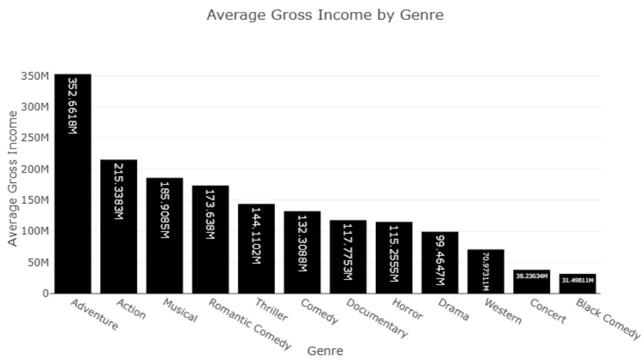

Figure 3. Case Study 2.

4.3. Case Study 3: *“Is there a relationship between movie budget and gross income?”*

The third case analyzed budget-gross correlation and began with *Clarify Question* asking more information: “*Do you want to analyze the correlation for all movies or by specific genre/period?*”. When we prompted for specificity (“*by genre: adventure*”), the *Text-to-Vis* generated a scatter plot (see Figure 4) showing a positive but non-linear relationship. *Data Storyteller* emphasized that while higher budgets generally yield higher grosses, exceptions exist (low-budget hits and high-budget flops), concluding that captivating storytelling ultimately outweighs pure financial investment.

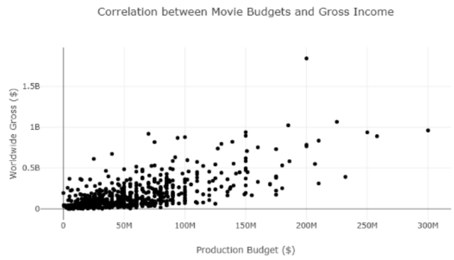

Figure 3. Case Study 3.

All prompts, LLM outputs, generated charts, and full narratives from these experiments are available in a public GitHub repository.[6]

5. Discussion

Our experiments show that LLMs can generate both accurate visualizations and meaningful narratives that support user interpretation. The charts matched the queries and data, while the narratives followed a coherent arc: providing context, highlighting trends, and suggesting explanations.

The LLM-based question clarity check was effective in prompting clarifications only when needed. Still, the quality of the narratives depends on prompt design and the model's interpretation, which may lead to generic insights. Overall, the workflow helps bridge the gap between visualization and understanding, and shows potential for building more intuitive tools for non-technical users.

6. Conclusion

This work presented a workflow that integrates Text-to-VIS and data storytelling using LLMs. Through case studies on Movies dataset, we demonstrated the system's ability to transform natural language queries into relevant charts, and provide insightful narrative explanations with minimal user intervention. As future work, we plan to explore multiple Text-to-VIS strategies and incorporate an agent-based evaluation module to select the best visualization among alternatives.

We also aim to assess both the quality of the generated story and its visualization, with a particular focus on personalization. One key challenge is how to leverage user intent to guide the storytelling process in a meaningful and adaptive way.

References

Danqing Shi, Xinyue Xu, Fuling Sun, Yang Shi, and Nan Cao (2021) "Calliope: Automatic visual data story generation from a spreadsheet", IEEE Trans. Vis. Comput. Graph., 27(2):453–463.

Dibia, V. (2023) "Lida: A tool for automatic generation of grammar-agnostic visualizations and infographics using large language models". Available at: https://arxiv.org/abs/2303.02927

[6] https://github.com/dudursn/llm_based_text2vis_and_storytelling

Dibia, V. and Demiralp, Ç. (2019) “Data2vis: Automatic generation of data visualizations using sequence-to-sequence recurrent neural networks”, IEEE Computer Graphics and Applications, 39(5):33–46.

Dyer, S., Martin, J. and Zulauf, J. (1995) “Motion Capture White Paper”. Available at: http://reality.sgi.com/employees/jam_sb/mocap/MoCapWP_v2.0.html, December.

Feijó, B. (2019) “Data Storytelling – Initial Reflections”, Technical Note DS01/2019, Department of Informatics, PUC-Rio, November. Available at: https://www-di.inf.puc-rio.br/~bfeijo/cursosjogos/metaverso/Nota_Tecnica_DS001_2019_DS_primeiras_reflexoes.pdf

Galileo AI (2025) “Mastering LLM as a Judge”, Galileo’s Evaluation Intelligence Platform. Available at: https://galileo.ai/mastering-llm-as-a-judge

Hunter, J. D. (2007) “Matplotlib: A 2D graphics environment”, Computing in Science & Engineering, 9(3):90–95.

Li, D., Mei, H., Shen, Y., Su, S., Zhang, W., Wang, J., Zu, M., and Chen, W. (2018) “ECharts: A Declarative Framework for Rapid Construction of Web-Based Visualization”, Visual Informatics, 2(2):136–146.

Li, S., Chen, X., Song, Y., Song, Y., Zhang, C. J., Hao, F., and Chen, L. (2025) “Prompt4vis: Prompting large language models with example mining for tabular data visualization”, The VLDB Journal, 34(4):1–26.

Maddigan, P. and Susnjak, T. (2023) “Chat2VIS: Generating Data Visualizations via Natural Language Using ChatGPT, Codex and GPT-3 Large Language Models”, IEEE Access, 11:45181–45193. doi:10.1109/ACCESS.2023.3274199.

Ma, P., Ding, R., Wang, S., Han, S., and Zhang, D. (2023) “XInsight: Explainable data analysis through the lens of causality”, Proc. ACM Manag. Data, 1(2):156:1–156:27.

OpenAI (2024) “GPT-4o (Generative Pre-trained Transformer 4 Omni)”, OpenAI, released 13 May 2024. Available at: https://openai.com/index/hello-gpt-4o

PandasAI (2024) “Pandasai: Interact with pandas dataframes using llms”. Available at: https://github.com/gventuri/pandas-ai

Satyanarayan, A., Moritz, D., Wongsuphasawat, K., and Heer, J. (2017) “Vega-Lite: A Grammar of Interactive Graphics”, IEEE Transactions on Visualization and Computer Graphics, 23(1):341–350.

Sun, M., Cai, L., Cui, W., Wu, Y., Shi, Y., and Cao, N. (2023) “Erato: Cooperative data story editing via fact interpolation”, IEEE Trans. Vis. Comput. Graph., 29(1):983–993.

Vanna.AI Inc. (2024) “Vanna.AI: A Python library for AI-powered SQL generation”. Available at: https://github.com/vanna-ai/vanna

Zhang, W., Wang, Y., Song, Y., Wei, V. J., Tian, Y., Qi, Y., Chan, J. H., Wong, R. C.-W., and Yang, H. (2024) “Natural Language Interfaces for Tabular Data Querying and Visualization: A Survey”, IEEE Transactions on Knowledge and Data Engineering, 36(11):6699–6718.