LLM-Powered Visualizations and Narratives from Natural Language Queries

Eduardo Roger S. Nascimento[1,*,†], Bruno Feijó[1,†], Antonio L. Furtado[1,†] and Marco A. Casanova[1,†]

[1]*Departamento de Informática – Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Caixa Postal 38.097 – 22.451-900, Rio de Janeiro, Brazil*

Abstract

The gap between non-technical users and tabular data limits effective access and data-driven decision-making. To address this challenge, this work presents an agent-based workflow integrating text to visualization (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, the approach is illustrated through three proof-of-concept demonstrations on a movie dataset. The results show that LLMs can generate appropriate visualizations and assist users in understanding patterns and key findings, bridging the gap between data visualization and storytelling.

Keywords

Text-to-VIS, Data Visualization, Data Storytelling, LLMs, Agents, LangGraph

1. Introduction

Data visualization is vital 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 [1]. 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 through 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 [2], Vega-Lite [3], and ECharts [4].

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 to inform decisions and guide action. It involves structuring the analytical message as a narrative that enhances the communication and dissemination of findings within and across groups [5]. Recent research and applications have explored the potential of LLMs not only to generate charts but also to act as narrators, producing descriptive text that explains the visualizations, answers specific analytical questions, or even mimics expert commentary [6, 7, 8].

ER2025: Companion Proceedings of the 44th International Conference on Conceptual Modeling: Industrial Track, ER Forum, 8th SCME, Doctoral Consortium, Tutorials, Project Exhibitions, Posters and Demos, October 20-23, 2025, Poitiers, France

rogerrsn@tecgraf.puc-rio.br (E. R. S. Nascimento); bruno@inf.puc-rio.br (B. Feijó); furtado@inf.puc-rio.br (A. L. Furtado); casanova@inf.puc-rio.br (M. A. Casanova)
https://wwww.inf.puc-rio.br/~casanova/ (M. A. Casanova)
0009-0005-3391-7813 (E. R. S. Nascimento); 0000-0003-4441-2632 (B. Feijó); 0000-0003-3710-624X (A. L. Furtado); 0000-0003-0765-9636 (M. A. Casanova)

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

This work explores the integration of Text-to-VIS and data storytelling using LLMs in a unified workflow. It presents 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 exploratory demonstrations, the work illustrates how LLMs can be used to construct relevant visualizations and generate compelling explanations tailored to the user's analytical goals.

2. Related Work

Text-to-VIS is the object of growing attention with the rise of LLMs. Early work, such as Data2Vis [9], trained models to generate Vega-Lite visualizations from structured data and textual input. Tools like Vanna AI [10], PandasAI [11], and LIDA [9] offer streamlined pipelines that transform user queries into data retrieval operations and basic visualizations. Chat2VIS [12] applies prompt engineering to define prompts capable of understanding user queries and generating Python code of the corresponding visualizations. Recent strategies include the framework Prompt4vis [13], which 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" [6] presents an automated system that transforms spreadsheet data into visual data stories by extracting relevant facts, generating charts, and composing narrative sequences. "Erato" [8] 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" [7] 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 to structure data-driven insights as coherent, interpretable stories that support better decision-making.

3. Methodology

The 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 predefined paths govern transitions but are conditionally triggered based on the output of LLM decisions.

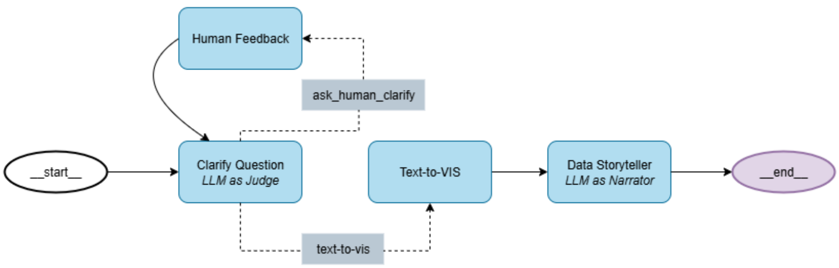


Figure 1: Workflow of the proposed system.

- **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 [14]. The workflow proceeds directly to the Text-to-VIS component if the query is sufficiently clear. Otherwise, a clarification step is triggered if the query is ambiguous or underspecified. In this case, the system prompts the user for additional information to disambiguate or enrich the original query. Once the user provides

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

the necessary feedback, the system integrates the new details into the query and generates the appropriate visualization.

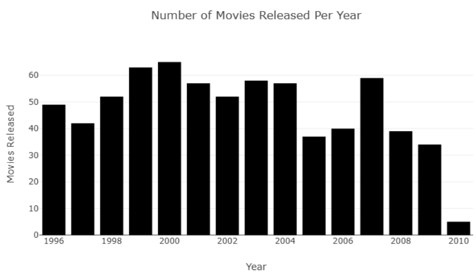
- **Human Feedback**: Prompted by the system, the user provides additional input to resolve ambiguities. This step ensures accurate query interpretation when LLM inference alone is insufficient.
- **Text-to-VIS**: In this step, a visualization strategy is executed, producing a complete Python code snippet responsible for rendering a specific chart (e.g., bar, line, scatter). Vanna AI [10] 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. The LLM is guided using information adapted from [5], enabling it to contextualize and communicate key patterns, trends, and insights in a way that is accessible and meaningful to end users. The user's question and the visualization metadata (chart type, data used, x and y labels) are passed in the prompt.

The evaluation of the proposed system is based on three proof-of-concept demonstrations, using the GPT-4o model provided by OpenAI [15], configured with a temperature of 0 to ensure deterministic responses. The evaluation used the Movies[5] dataset as the sole data source for all demonstrations. This dataset contains data on movies released between 1996 and 2010, including financial information, such as gross income and budget, ratings, genres, and release year.

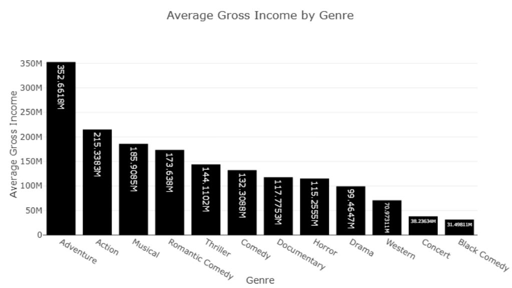
4. Results

4.1. Demonstration 1: "Number of movies released by year"

The first demonstration explores a question that analyzes how LLMs handle time-based data requests. The model generated a bar chart (see Figure 2(a)) showing the annual number of movie releases from 1996 to 2010. The narrative generated by the LLM 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 Years"*) and proposed causes, such as economic crisis, technological shifts, and industry saturation.



(a) Result from Demonstration 1.



(b) Result from Demonstration 2.

Figure 2: Charts generated by the LLM-based system for Demonstrations 1 and 2.

4.2. Demonstration 2: "Which movie genres have the highest average gross income?"

The second case illustrative example explores how LLMs generate visual summaries from categorical and aggregated numerical data, focusing on average gross income grouped by movie genre. The *Text-to-VIS Tool* produced a bar chart ranking genres and autonomously sorted the chart by average gross

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

income (without explicit prompting) as shown in Figure 2(b), 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 genre reigns Supreme*, citing broad demographic appeal and high production value. Thus, this demonstrates the model's capability to generate interpretive analysis.

4.3. Demonstration 3: "Is there a relationship between movie budget and gross income"

The third demonstration analyzed budget-gross correlation and began with *Clarify Question*, asking for more information: *"Do you want to analyze the correlation for all movies or by specific genre/period?"*. When prompted specifically for adventure genre, the *Text-to-Vis* tool was called and generated a scatter plot (see Figure 3) showing a positive but non-linear relationship. The *Data Storyteller* emphasized that, while higher budgets generally yield higher grosses, exceptions exist, such as low-budget hits and high-budget flops. Finally, it concluded that captivating storytelling ultimately outweighs pure financial investment.

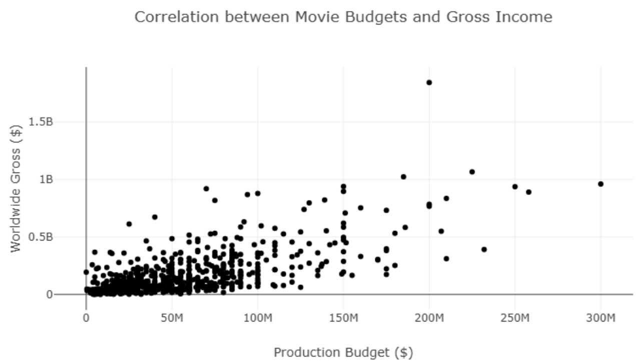


Figure 3: LLM-generated chart for Demonstration 3.

4.4. Discussion

The explorations suggest that LLMs can generate accurate visualizations and meaningful narratives supporting 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.

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

5. Conclusion

This work presented a workflow that integrates Text-to-VIS and data storytelling using LLMs. Proof-of-concept demonstrations based on the Movies dataset demonstrated the system's ability to transform natural language queries into relevant charts and provide insightful narrative explanations with minimal

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

user intervention. Future work will explore multiple Text-to-VIS strategies and incorporate an agent-based evaluation module to select the best visualization alternative. It is also planned to assess the quality of the generated story and its visualization, with a particular focus on personalization. One key challenge is leveraging user intent to guide the storytelling process in a meaningful and adaptive way.

Declaration on Generative AI

During the preparation of this work, the authors used GPT-4o and Grammarly for grammar and spelling checks, and take full responsibility for the final content.

References

[1] W. Zhang, Y. S. Yifei Wang, V. J. Wei, Y. Tian, Y. Qi, J. H. Chan, R. C.-W. Wong, H. Yang, Natural language interfaces for tabular data querying and visualization: A survey, IEEE Transactions on Knowledge and Data Engineering 36 (2024) 6699–6718.

[2] J. D. Hunter, Matplotlib: A 2d graphics environment, Computing in Science & Engineering 9 (2007) 90–95.

[3] A. Satyanarayan, D. Moritz, K. Wongsuphasawat, J. Heer, Vega-lite: A grammar of interactive graphics, IEEE Transactions on Visualization and Computer Graphics 23 (2017) 341–350.

[4] D. Li, H. Mei, S. S. Yi Shen, W. Zhang, J. Wang, M. Zu, W. Chen, Echarts: A declarative framework for rapid construction of web-based visualization, Visual Informatics 2 (2018) 136–146.

[5] B. Feijó, Data Storytelling – Initial Reflections, Technical Report DS01/2019, Department of Informatics, PUC-Rio, 2019. https://www-di.inf.puc-rio.br/~bfeijo/cursosjogos/metaverso/Nota_Tecnica_DS001_2019_DS_primeiras_reflexoes.pdf.

[6] D. Shi, X. Xu, F. Sun, Y. Shi, N. Cao, Calliope: Automatic visual data story generation from a spreadsheet, IEEE Transactions on Visualization and Computer Graphics 27 (2021) 453–463.

[7] P. Ma, R. Ding, S. Wang, S. Han, D. Zhang, Xinsight: Explainable data analysis through the lens of causality, Proceedings of the ACM on Management of Data 1 (2023) 156:1–156:27.

[8] M. Sun, L. Cai, W. Cui, Y. Wu, Y. Shi, N. Cao, Erato: Cooperative data story editing via fact interpolation, IEEE Transactions on Visualization and Computer Graphics 29 (2023) 983–993.

[9] V. Dibia, Çagatay Demiralp, Data2vis: Automatic generation of data visualizations using sequence-to-sequence recurrent neural networks, IEEE Computer Graphics and Applications 39 (2019) 33–46.

[10] Vanna.AI, Vanna.ai: A python library for ai-powered sql generation, https://github.com/vanna-ai/vanna, 2024.

[11] PandasAI, Pandasai: Interact with pandas dataframes using llms, https://github.com/gventuri/pandas-ai, 2024.

[12] P. Maddigan, T. Susnjak, Chat2vis: Generating data visualizations via natural language using chatgpt, codex and gpt-3 large language models, IEEE Access 11 (2023) 45181–45193. doi:10.1109/ACCESS.2023.3274199.

[13] S. Li, X. Chen, Y. Song, Y. Song, C. Zhang, Prompt4vis: Prompting large language models with example mining for tabular data visualization, The VLDB Journal 34 (2025) 1–26.

[14] Galileo AI, Mastering llm as a judge, https://galileo.ai/mastering-llm-as-a-judge, 2025. Accessed 2025.

[15] OpenAI, Gpt-4o (generative pre-trained transformer 4 omni), https://openai.com/index/hello-gpt-4o, 2024. Released 13 May 2024.

A. Online Resources

All prompts, LLM outputs, generated charts, and full narratives from these experiments are available via

- GitHub at https://github.com/dudursn/llm_based_text2vis_and_storytelling