**Abstraction**

Human language has long presented significant challenges for the field of natural language processing (NLP) in terms of understanding and generating natural language. In the field of Natural Language Text-Generated Data Visualization (NL2VIS), there are often challenges in providing non-experts with the ability to create accurate and informative data visualizations through existing solutions. Among the significant implications of this problem are the growing need for intuitive tools that allow individuals without technical expertise in graphical analysis to interpret and visualize data. There remains, however, a long way to go in developing NL2VIS systems that are both highly accurate and highly customizable due to the nuances of natural language and the complexity of graphical representation of data. By integrating advanced pre-trained large language models (LLMs) such as ChatGPT (4o Mini, 4o, o1), NL4DV, ncNET, and YOLOPandas into the existing state-of-the-art NL2VIS solution, Chat2VIS, we improve the current state-of-the-art NL2VIS solution. Through a series of case studies, we compare and evaluate the performance differences of these models. Furthermore, we contribute to the field of chart customization by improving the user experience, thereby making the system more practical and accessible to a broader audience.

**Introduction**

Natural language processing (NLP) has long grappled with the complexities of human language, particularly in the domains of natural language understanding (NLU) and natural language generation (NLG) [1]. The ability of machines to accurately comprehend and generate language poses significant challenges, owing to the inherent ambiguity, variability, and context-dependence of human communication. However, the advent of large-scale language models like GPT-3 and, more recently, ChatGPT by OpenAI has marked a significant milestone in overcoming these barriers. These models have demonstrated remarkable capabilities in various NLP tasks, including conversational agents, text generation, and summarization [2].

One emerging application of NLP is Natural Language to Data Visualization (NL2VIS), where natural language queries are transformed into meaningful data visualizations. This approach is particularly beneficial for users who lack expertise in graphical data analysis but need quick and accurate visual representations of complex datasets [3]. Despite its potential, most current NL2VIS systems rely on statistical parsers, limiting them to handling only simple queries. Although advanced deep learning models have excelled in other NLP tasks, they are not widely applied to NL2VIS, primarily due to the lack of large-scale, high-quality benchmarks that are essential for supporting these systems [4]. This gap poses a significant challenge in the development of more flexible and powerful NL2VIS systems.

In recent years, researchers have begun exploring the integration of large language models (LLMs) into NL2VIS to address these limitations. These efforts aim to enhance the complexity of processable natural language queries and expand the flexibility of visual outputs. For instance, Chen et al. conducted an empirical evaluation of GPT-3.5 and GPT-4 in data visualization tasks within Harvard University's CS171 Data Visualization course [5], while Vázquez developed a test set to analyze the performance of ChatGPT-3 and ChatGPT-4 in generating charts [6].

With the continuous advancement of LLM and NLP technologies, powerful visualization tools like ChatGPT (4o Mini, 4o, o1), NL4DV, ncNET, and YOLOPandas have emerged. This article evaluates the performance differences between these models using a series of datasets and integrates them into the Chat2VIS application developed by Maddigan et al. [3], providing users with more sophisticated options for generating visualizations. Furthermore, we introduce a chart customization feature to enhance the system’s practicality and flexibility, making it more adaptable to diverse user needs.

**References**

[1] Gokul Yenduri *et al.*, “GPT (Generative Pre-trained Transformer) – A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions,” *IEEE access*, pp. 1–1, Jan. 2024, doi: https://doi.org/10.1109/access.2024.3389497.

‌[2] P. P. Ray, “ChatGPT: a Comprehensive Review on background, applications, Key challenges, bias, ethics, Limitations and Future Scope,” Internet of Things and Cyber-Physical Systems, vol. 3, no. 1, pp. 121–154, Apr. 2023, doi: https://doi.org/10.1016/j.iotcps.2023.04.003.

‌[3] P. Maddigan and T. Susnjak, “Chat2VIS: Generating Data Visualisations via Natural Language using ChatGPT, Codex and GPT-3 Large Language Models,” IEEE Access, pp. 1–1, 2023, doi: https://doi.org/10.1109/access.2023.3274199.

‌[4] Y. Luo, J. Tang, and G. Li, “nvBench: A Large-Scale Synthesized Dataset for Cross-Domain Natural Language to Visualization Task,” 2015. Accessed: Sep. 24, 2024. [Online]. Available: https://arxiv.org/pdf/2112.12926

‌[5] Z. Chen et al., “Beyond Generating Code: Evaluating GPT on a Data Visualization Course,” arXiv (Cornell University), Oct. 2023, doi: https://doi.org/10.1109/eduvis60792.2023.00009.

‌[6] P. -P. Vázquez, "Are LLMs ready for Visualization?," *2024 IEEE 17th Pacific Visualization Conference (PacificVis)*, Tokyo, Japan, 2024, pp. 343-352, doi: 10.1109/PacificVis60374.2024.00049.