VIZARD: Improving Visual Data Literacy with Large Language Models

Rubab Zahra Sarfraz
rubabzsarfraz@gmail.com
BridgeLinx Technologies
Lahore, Pakistan

Samar Haider
samarh@seas.upenn.edu
University of Pennsylvania
Philadelphia, USA

ABSTRACT

Data visualizations are commonplace in both our professional and personal lives. From workplace dashboards to our health charts to spending trackers—we interact with them almost every day. However, despite their crucial role in communicating information to us, many people still struggle to effectively use these tools and draw meaningful insights from them. This issue is particularly acute in developing countries where language barriers and limited technology skills present additional challenges for data visualization literacy. In this paper, we present VIZARD: a dashboard companion that uses large language models to analyze data visualizations for users and explain their elements in their language of choice, as well as providing insights and recommendations based on the trends observed according to the user's industry and job role. We pair this with a novel framework for evaluating visualization literacy which uses procedurally generated questions that are tailored to participants' interests and current visualization literacy level. We make VIZARD open source to encourage more research in this direction.

VLDB Workshop Reference Format:
Rubab Zahra Sarfraz and Samar Haider. VIZARD: Improving Visual Data Literacy with Large Language Models. VLDB 2024 Workshop: International Workshop on Big Data Visual Exploration and Analytics (BigVis 2024).

1 INTRODUCTION

Graphical visualizations are crucial in presenting information within the data science life cycle. Data professionals typically create dashboards for business users, but if these users lack the skills to interpret them, significant information loss occurs. To aid comprehension, data analysts often add explanatory texts, but this method is outdated with evolving data applications. Additionally, data professionals may lack domain-specific knowledge, which further limits the value extracted from the data.

Recent advancements in the visual reasoning [19] and multilingual capabilities [7] of LLMs can help change this. Our research shows how state-of-the-art LLMs can be used to make dashboards more comprehensible for all business users, enhancing visualization literacy and promoting effective data utilization in organizations.

In this paper, we present VIZARD: an intervention framework for improving users' data visualization literacy (DVL) with a personalized dashboard companion for interpreting visualizations and a DVL evaluation framework that measures their progress. We make all code, resources, and examples of our framework's usage available to the community[1].

Our contributions are as follows:

(1) We introduce VIZARD, a fully customizable LLM-powered dashboard companion that analyzes and explains graphical visualizations to users of varying DVL levels.
(2) We propose a novel visualization literacy evaluation framework which procedurally generates plots and questions that can be tailored to various domains, levels of complexity, and languages.
(3) We conduct a user survey to understand the challenges faced by people in a developing country who work with data and ask for their feedback on VIZARD.

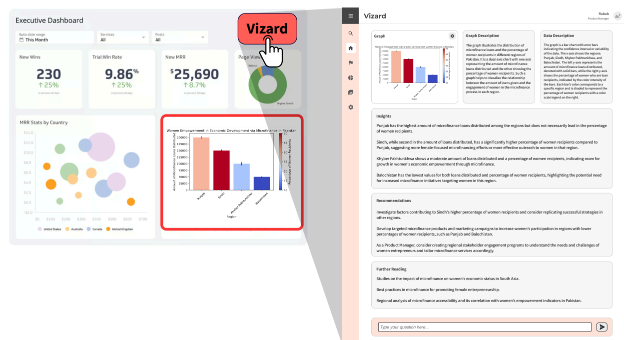


Figure 1: An illustration of VIZARD's interface showing its integration with an existing dashboard software.

2 RELATED WORK

Data Literacy: Data literacy is defined as the ability to understand, interpret, and critically evaluate data for effective decision-making [14, 18] and has been shown to be vital in people's education, research, and professional growth [6, 15]. Our work specifically focuses on a subset of data literacy called data visualization literacy (DVL) or visual data literacy (VDL), which refers to the ability to interpret and extract information from data visualizations [1, 9]. Research has shown that people with high graph literacy are able to draw better insights from data, which further underscores the importance of this skill [12]. When it comes to evaluating a user's DVL, multiple studies have tackled this problem and proposed various

Proceedings of the VLDB Endowment. ISSN 2150-8097.

[1] https://github.com/rubabzs/vizard

solutions [3, 4, 8, 13]. In this paper, we draw inspiration from these frameworks and aim to make our system similarly user-friendly.

Personalized Dashboards: In order to teach graph comprehension to more people, educational materials must be made more engaging to retain individuals' interest. The best way to do so is to tailor it to people's own use cases so they can more clearly see its applicability in their lives and work. Various different approaches have been presented for this [17], like introducing a semantic language for dashboard creation [11], or a Google Chrome extension which captures graphs and uses machine learning classifiers to identify their type for the user [12].

Multimodal and Multilingual LLMs: The ability to understand and reason about visual data has broadened LLMs' applicability in this domain [19]. They have been successfully utilized to automatically caption scientific figures [5, 16], leading to the creation of dedicated datasets for training models in this task [2]. Their multilingual capability [7] also allows them to understand and switch seamlessly between multiple languages [20], which increases their utility in multilingual settings [10]. We make use of this ability in the Urdu version of VIZARD.

3 VIZARD: SYSTEM OVERVIEW

In this section, we describe our framework for evaluating and improving visualization literacy levels within controlled environments, such as private or public organizations. VIZARD is composed of two main components: a DVL evaluation framework and a dashboard companion tool which work together as shown in Figure 2. Both components currently build on top of GPT-4o, which excels in both its multilingual and multimodal reasoning capabilities, but can be adapted to use any multi-modal LLM as a backend.

Initially, the organization's primary data custodian[2] configures VIZARD according to the business' domain and sets the parameters for the visualization literacy evaluation quiz for other members so that the questions and graph explanations are tailored appropriately. The intended business users then set their job role and preferred language and take the quiz to assess their current DVL level. The results of this step are stored in a shared database which is later used to generate personalized educational content. The data users regularly re-take the quiz to evaluate their DVL level over time to measure the impact of our framework on their skills and adapt its content accordingly. Table 1 lists the five DVL levels and their corresponding user abilities, with each level assuming knowledge of the preceding one.

In the following sections, we provide technical details on the two components of our framework.

3.1 Data Visualization Literacy Evaluation

We develop a novel framework to procedurally generate plots and accompanying quizzes at varying levels of complexity. Our pipeline has four stages:

Idea Generation: We prompt the LLM to generate ideas for graphical plots on a given topic e.g., *"women empowerment in economic development of a country"*. The model generates a list of plot ideas of varying complexity levels, each with accompanying

[2]We define data custodian to be the person who is responsible for advocating for and generating value from data in an organization.

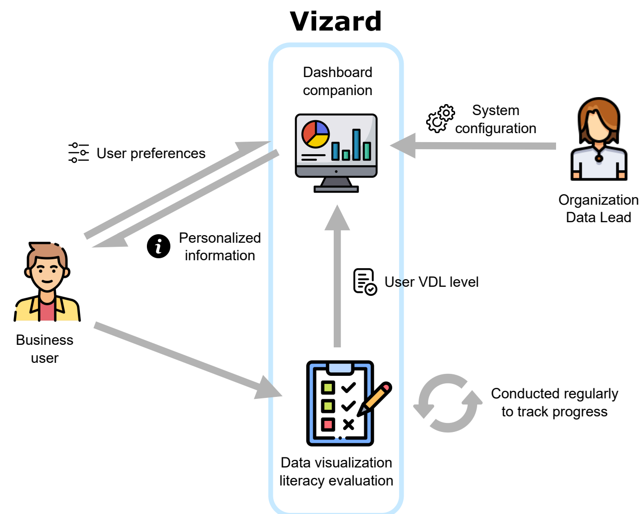


Figure 2: An overview of VIZARD. The organization's chief data officer (or equivalent) configures the system according to the organization's industry and domain while the user sets their language preferences and takes the DVL evaluation. All inputs are then used by the dashboard to customize its output for each user.

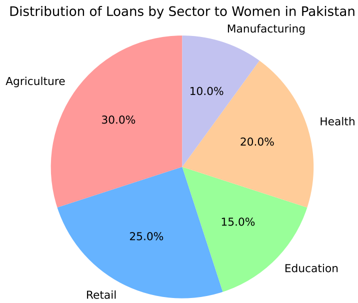
Table 1: Data Visualization Literacy Levels.

Level	Definition
1	The user does not know anything about data visualizations such as graphs and charts.
2	The user knows about different kinds of visualizations but finds it difficult to build relationships between the variables given in them.
3	The user can build relationships between two variables but finds it difficult to build relationship between more variables.
4	The user can build relationships between more than two variables but finds it difficult to apply their domain knowledge to the visualizations.
5	The user can apply domain knowledge and derive insights and recommendations from the visualizations.

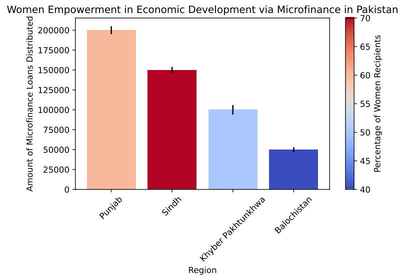
`instructions` and a `schema` in a predefined JSON format which describe the plot elements, like `"chart_type"`, `"variables"`, and `"color"`, etc.

Plot Generation: Using the `instructions` and `schema` from the previous step, the LLM generates Python code to create the plot. If the model does not have access to the Internet or actual data about the domain, it produces synthetic data that reflects real-world trends. The code is then run locally to visualize the plot, and Figure 3 shows an example of this.

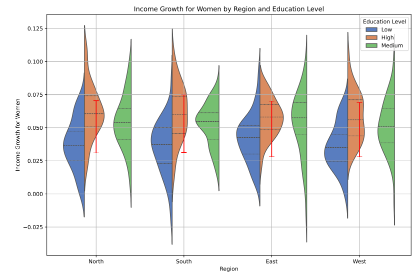
Question Generation: We provide the LLM with the generated plot and ask it to create multiple-choice questions of varying difficulty levels which focus on graph elements, data trends, and



(a) Level 1.



(b) Level 2.



(c) Level 3.

Figure 3: Example graphs produced by GPT-4o based on our specifications, one at each level of complexity. Observe how the number of variables as well as the number of layers of information increases as we go from left to right, making each higher level plot is more challenging for users to understand than the last.

Q1. What does the color of the bars in the bar chart indicate?

A) The total amount of microfinance loans distributed in each region.
B) The error or variability in the distribution of loans.
C) The population size of each region.
✔ D) The percentage of women recipients of microfinance loans in each region.

Q2. Based on the bar chart depicting the distribution of microfinance loans across various regions in Pakistan and highlighting the percentage of women recipients along with variability, what insight can most accurately be drawn?

A) Punjab has the lowest percentage of women recipients of microfinance loans.
B) Sindh shows the least variability in the distribution of loans among women.
✔ C) Sindh has the highest percentage of women recipients of microfinance loans.
D) Khyber Pakhtunkhwa and Balochistan distribute an equal number of loans to women.

Figure 4: Example questions accompanying the bar chart in Figure 3b. Observe the difference in difficulty as Q2 relies on more in-depth knowledge of the chart as well as the domain.

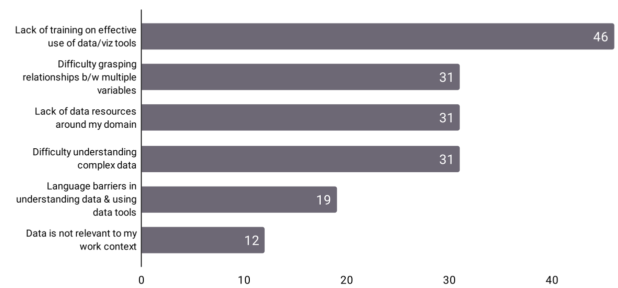


Figure 5: Common challenges faced when working with data. Lack of training on relevant tools, difficulty in understanding variable relationships and lack of domain-oriented resources came out to be on top in our survey (n=75).

insights from the plot. The proportion of questions of a certain difficulty that user's can get right decides their DVL level. Two example questions generated for the bar chart in Figure 3b are shown in Figure 4 with the correct options marked by the LLM as well.

DVL Level Assignment: Users take the evaluation quiz before they begin using Vizard as a dashboard companion. Our evaluation framework is adaptive and presents users with questions of increasing difficulty if they answer the initial questions correctly, or decreasing difficulty otherwise. This continues until we converge to a level where users get roughly half the questions right, at which point they are assigned the corresponding DVL level and the dashboard companion generates educational content accordingly.

3.2 Dashboard Companion

Once a user's visualization literacy level has been determined, Vizard is able to proceed with its core function of generating personalized educational content for dashboards.

Educational Content Generation: Vizard generates education content for a graph by providing the image to the LLM along with a prompt that contains comprehensive instructions for how it should be described. The prompt is parametrized with the user's preferences and DVL level to tailor the content accordingly. Input parameters include `"language"`, `"dvl_level"`, `"role"`, and `"domain"`. The prompt instructs the LLM to produce educational content with the following five sections, each focused on explaining a different aspect of the graph:

(1) **Graph description:** Explains the type of graph and its characteristics, tailored to the user's DVL level.
(2) **Data description:** Connects the data with the chart, explaining elements like labels, axes, colors, legends, data points, and error bars.
(3) **Insights:** Analyzes trends in the graph and customizing them according to the user's job function and domain knowledge. This section is crucial for all users as it teaches beginners how to interpret graphs and validates insights for advanced users.
(4) **Recommendations:** Provides actionable steps based on the insights, also tailored to the user's role and DVL level.
(5) **Further readings:** Recommends resources for deeper understanding of the data being analyzed.

An example of these outputs are shown in Table 2. Due to space constraints, we only include three sections. Further examples as well as the full content generation prompt can be found in our GitHub repository.

Table 2: Examples of Vizard's explanatory outputs at DVL level 3 (see Table 1) for the violin plot in Figure 3c.

Graph Description	Insights	Recommendations
• The graph depicts Income Growth for Women by Region and Education Level, represented through a multi-colored violin plot with regions classified as North, South, East, and West. • The data points within each violin plot spread differently, signaling variance and data distribution for women with different levels of education in these regions.	• Income growth for women varies significantly across different regions. • Women with a high level of education appear to have, on average, higher income growth compared to those with medium and low education levels across all regions. • The Southern region shows a narrower income growth distribution for women with high education, suggesting less variability in income outcomes for this subgroup. • The spread of income growth for women with low education in the Western region is larger than in other regions, indicating higher variability in economic outcomes for this group. • Median income growth for women with medium education appears comparable across all regions, possibly suggesting that other factors are influencing income growth for this group besides regional location and education level.	• Implement targeted policies that focus on regions with lower income growth, especially for women with low levels of education, such as in the North and East. • Increase access to higher education for women as there's a correlation between higher education levels and increased income growth. • Study the factors contributing to the lower variability of income growth in the Southern region for highly educated women to replicate successful strategies in other regions. • Conduct further research on the Western region to understand the factors behind the large spread of income growth for women with low education and address the underlying issues. • Consider uniform interventions to support women with medium education levels as their income growth does not significantly vary with the region, indicating other common barriers to economic empowerment

Prompt Engineering & Evaluation: While developing the content generation prompt, we experimented with including varying levels of detail on the user as well as refining the phrasing of the instructions to improve the content produced. We evaluated the LLM's output according to a rubric informed by what users of different DVL levels would expect from such a system and tested it out at different literacy levels as well as in both languages, English and Urdu.

User Feedback: During the development of Vizard, we conducted a survey targeted at individuals working in occupations requiring some level of data analysis in their day-to-day role (n=75). Figure 5 lists the major challenges they faced. After being shown a mock up of Vizard's interface and a description of its functionality, over 90% of respondents stated that such a tool would definitely be valuable for them and will help them improve their understanding of data.

4 CONCLUSION AND FUTURE WORK

Visualization literacy is an increasingly indispensable skill to have and being able to understand and effectively use data enables people to make better decisions and become more productive members of society. In this paper, we presented Vizard, a novel framework for adaptive visualization literacy education and evaluation which can serve as a customizable dashboard for explaining complex graphs to users at different DVL levels in their preferred language. Although our user survey provided us with positive feedback, we aim to expand on this in future work by conducting a focus group to improve Vizard's user experience and performing a long-term study to measure its effect on users' literacy levels. In the future, we would like to introduce the functionality to integrate the output from the other DVL evaluation frameworks as well. We will also evaluate additional LLMs for this task and conduct rigorous experiments to quantify Vizard's ability to adapt to various audiences and types of visualizations. We open-source all code and resources created in this work and hope that Vizard can help equip everyone with the skills needed to interpret visual data and improve our understanding of data as a society.

REFERENCES

[1] Katy Börner, Adam Maltese, Russell Nelson Balliet, and Joe Heimlich. 2016. Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization* 15, 3 (2016), 198–213.

[2] Sanjana Shivani Chintalapati, Jonathan Bragg, and Lucy Lu Wang. 2022. A dataset of alt texts from HCI publications: Analyses and uses towards producing more descriptive alt texts of data visualizations in scientific papers. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–12.

[3] Yuan Cui, W Ge Lily, Yiren Ding, Fumeng Yang, Lane Harrison, and Matthew Kay. 2023. Adaptive assessment of visualization literacy. *IEEE Transactions on Visualization and Computer Graphics* (2023).

[4] Lily W Ge, Maryam Hedayati, Yuan Cui, Yiren Ding, Karen Bonilla, Alark Joshi, Alvitta Ottley, Benjamin Bach, Bum Chul Kwon, David N Rapp, et al. 2024. Toward a More Comprehensive Understanding of Visualization Literacy. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–7.

[5] Ting-Yao Hsu, C Lee Giles, and Ting-Hao'Kenneth' Huang. 2021. SciCap: Generating captions for scientific figures. *arXiv preprint arXiv:2110.11624* (2021).

[6] Tibor Koltay. 2017. Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science* 49, 1 (2017), 3–14.

[7] Viet Dac Lai, Nghia Trung Ngo, Amir Pouran Ben Veyseh, Hieu Man, Franck Dernoncourt, Trung Bui, and Thien Huu Nguyen. 2023. Chatgpt beyond english: Towards a comprehensive evaluation of large language models in multilingual learning. *arXiv preprint arXiv:2304.05613* (2023).

[8] Sukwon Lee, Sung-Hee Kim, and Bum Chul Kwon. 2016. Vlat: Development of a visualization literacy assessment test. *IEEE transactions on visualization and computer graphics* 23, 1 (2016), 551–560.

[9] Sukwon Lee, Sung-Hee Kim, and Bum Chul Kwon. 2017. VLAT: Development of a Visualization Literacy Assessment Test. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 551–560. https://doi.org/10.1109/TVCG.2016.2598920

[10] Junyi Li, Tianyi Tang, Wayne Xin Zhao, Jian-Yun Nie, and Ji-Rong Wen. 2022. Pretrained language models for text generation: A survey. *arXiv preprint arXiv:2201.05273* (2022).

[11] Stephen MacNeil, Parth Patel, and Benjamin E Smolin. 2022. Expert goggles: Detecting and annotating visualizations using a machine learning classifier. In *Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–3.

[12] Yasmina Okan, Mirta Galesic, and Rocio Garcia-Retamero. 2016. How people with low and high graph literacy process health graphs: Evidence from eye-tracking. *Journal of Behavioral Decision Making* 29, 2-3 (2016), 271–294.

[13] Saugat Pandey and Alvitta Ottley. 2023. Mini-VLAT: A Short and Effective Measure of Visualization Literacy. In *Computer Graphics Forum*, Vol. 42. Wiley Online Library, 1–11.

[14] Luci Pangrazio and Julian Sefton-Green. 2020. The social utility of 'data literacy'. *Learning, Media and Technology* 45, 2 (2020), 208–220.

[15] Judy Robertson and E Kay M Tisdall. 2020. The importance of consulting children and young people about data literacy. *Journal of Media Literacy Education* 12, 3 (2020), 58–74.

[16] Andrea Spreafico and Giuseppe Carenini. 2020. Neural data-driven captioning of time-series line charts. In *Proceedings of the International Conference on Advanced Visual Interfaces*. 1–5.

[17] Andrea Vázquez-Ingelmo, Francisco José García-Peñalvo, and Roberto Therón. 2019. Tailored information dashboards: A systematic mapping of the literature. In *Proceedings of the XX international conference on human computer interaction*. 1–8.

[18] Annika Wolff, John Moore, Zdenek Zdrahal, Martin Hlosta, and Jakub Kuzilek. 2016. Data literacy for learning analytics. In *Proceedings of the sixth international conference on learning analytics & knowledge*. 500–501.

[19] Zhengyuan Yang, Linjie Li, Kevin Lin, Jianfeng Wang, Chung-Ching Lin, Zicheng Liu, and Lijuan Wang. 2023. The dawn of lmms: Preliminary explorations with gpt-4v (ision). *arXiv preprint arXiv:2309.17421* 9, 1 (2023), 1.

[20] Yan Zeng, Wangchunshu Zhou, Ao Luo, Ziming Cheng, and Xinsong Zhang. 2022. Cross-view language modeling: Towards unified cross-lingual cross-modal pre-training. *arXiv preprint arXiv:2206.00621* (2022).