https://doi.org/10.2352/EI.2024.36.1.VDA-360


# Evaluating the Recommendations of LLMs to Teach a Visualization Technique using Bloom's Taxonomy

***Alark Joshi*** [1]***, Chandana Srinivas*** [1]***, Elif E. Firat*** [2]***, Robert S. Laramee*** [3]
***University of San Francisco*** [1]***, Cukurova University*** [2]***, University of Nottingham*** [3]

## Abstract

*Large Language Models (LLMs) have demonstrated a huge impact on education and literacy in recent years. We evaluated the recommendations provided by two popular LLMs (OpenAI's ChatGPT and Google's Bard) to educate novices on the topic of Parallel Coordinate Plots (PCPs) using Bloom's taxonomy. We present the results of a human-expert evaluation of the recommendations provided by both the LLMs with experts from the visualization literacy field. Based on the analysis of the expert evaluation, we found that while both the LLMs provided some relevant and practical recommendations, some of the recommendations were either too difficult for novices or were in the wrong cognitive process (according to Bloom's taxonomy). In some cases, the hallucinations led to recommendations that were completely inapplicable to Parallel Coordinate Plots literacy.*

## Introduction

Large Language Models (LLMs) have been receiving a lot of attention [19, 25] in higher education due to their potential to provide solutions to tests and assignments for students [34, 24, 10, 35]. The Visualization Literacy field has also been considering novel ways to teach individuals how to read and interpret unfamiliar charts [5, 1, 29, 33, 13, 28]. Identifying the "best" way to teach individuals about new/unfamiliar visualization techniques continues to be an open research problem. Increasingly, visualization techniques such a treemaps, bubble charts, and Parallel Coordinate Plots (PCPs) are making their way into news media, fitness apps, and our every day lives [30, 31, 26].

In this paper, we analyzed the results from four visualization literacy experts who reviewed the recommendations given by two popular LLMs (OpenAI's ChatGPT-3.5 and Google's Bard 2.0.0). The experts evaluated the LLMs recommendations based on their prior experience of teaching students the topic of PCPs using the Bloom's taxonomy cognitive learning framework [3]. Based on our analysis of the scores provided by the experts, we found that while many of the LLM recommendations were useful and potentially relevant, there were some cases where the recommendations were completely inappropriate for PCP literacy. There were some cases where the recommendations were appropriate for PCP literacy, but not in the appropriate cognitive stage of Bloom's taxonomy. The experts also stated that some of the recommendations were too difficult for novices.

Figure 2 shows an overview of our process that included obtaining recommendations from LLMs, soliciting recommendations from experts, analyzing the scores, and presenting the findings.

Here are the contributions of the paper:

- We identified the recommendations provided by popular LLMs for visualization literacy using Bloom's taxonomy.
- The recommendations provided by the LLMs were evaluated by experts in the visualization literacy field with experience using Bloom's taxonomy.
- The expert evaluations suggest that the recommendations provided by the LLMs, while mostly useful, are not always applicable and some may have significant errors.

### *Background - Bloom's Taxonomy*

Bloom's taxonomy [3, 4] was proposed by Benjamin S. Bloom in 1956 as a cognitive framework for instruction and assessment of learning outcomes. It contains six cognitive processes that build on each other and enable learners to obtain and demonstrate proficiency as they traverse through the cognitive processes. Figure 1 shows the six cognitive processes starting from the Remember category at the bottom of the pyramid. We provide a quick overview of Bloom's taxonomy,

1. **Remember** facts related to the concept. Being able to remember and retrieve basic characteristics about a given concept is an essential first step. As Kratwohl [21] states, learning frequently focuses *only* on this step.
2. **Understand** - This process relates to the ability of a learner to demonstrate their ability to explain, classify, describe their understanding of a given concept.
3. **Apply** - In this stage, the learner has an opportunity to apply their acquired knowledge through hands-on learning to solve a problem or execute a set of steps.
4. **Analyze** - This process requires learners to apply their metacognitive processes incorporating their learned concept by comparing, distinguishing, and drawing connections between prior concepts and the given topic.
5. **Evaluate** - As the learner increases proficiency, their ability to critique or appraise a certain situation is crucial. This engages higher order metacognition and requires learners to justify their decision(s).
6. **Create** - In the final step, learners are required to synthesize their acquired knowledge and demonstrate proficiency by creating or formulating a new or original piece of work.

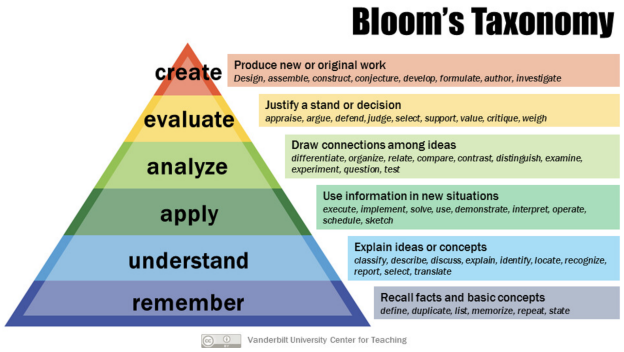


***Figure 1.*** *The Six Cognitive Processes in Bloom's Taxonomy. The Cognitive Processes start from the introductory tasks such as recalling facts and remembering concepts at the bottom of the pyramid to producing original work based on a deep understanding of a given focus concept. Image credits: Center for Teaching, Vanderbilt University [2].*

## Related Work

### *LLMs and Data Visualization*

Researchers in the data visualization community have experimented with ChatGPT in the context of graph layout algorithms [12]. Chen et al. [9] found that the Generative Pre-trained Transformer (GPT) model in GPT-4 performed better than GPT-3.5 to complete assignments in their data visualization course. Additionally, the teaching assistants could distinguish between human-generated work and GPT-generated work around 70% of the time.

### *Data Visualization Literacy*

Visualization Literacy is a popular topic in the data visualization research field [5, 6]. Firat et al. [15, 16] have recently provided survey papers detailing the state-of-the-art and research challenges related to visualization literacy. There have been many efforts to increase visualization literacy of the students in a classroom at the K-12 level [1, 11], higher education [22, 17], and general audiences [7, 33, 32, 28].

Assessing visualization literacy has been a hot topic as well with the Visualization Literacy Assessment Test from Lee et al. [23] followed by the recently introduced mini-VLAT [27]. Boy et al. [8] developed an assessment tool based on item-response theory. Recently, Ge et al. [18] introduced CALVI to assess a viewer's ability to critically evaluate charts (especially one that are created to mislead viewers).

Parallel coordinates plots are used in news media and scientific reports to visualize multidimensional data. To increase the ability of general audiences, Firat et al. [14] developed and evaluated an interactive tool to teach participants about PCPs. Peng et al. [28] created six modules to teach students about PCPs following the cognitive processes in Bloom's taxonomy.

## Methodology

Given the popularity of Large Language models (LLMs) in higher education [34, 24], we wanted to evaluate their ability to teach novices about a visualization technique. We chose Parallel Coordinates Plots due to the fact that they are still not as popular as Treemaps or Box plots, but well established in the data visualization research field.

We gave the following prompt to OpenAI's ChatGPT (GPT-3.5) and Google's Bard (2.0.0):

**"Using Bloom's taxonomy, generate prompts for a novice to learn about Parallel Coordinate Plots."**

Multiple "draft" and *regenerated responses* for recommendations received from ChatGPT and Bard were collected. The various drafts/regenerated responses were almost identical to the original (with minor word variations) and so we used the original set of recommendations from both the LLMs.

These recommendations were then rated by a team of visualization literacy experts that comprised of three professors, who conduct research in the field of Data Visualization and Visualization Literacy and one student actively conducting research on visualization literacy. Each of the recommendations were rated by each expert using a 7-point Likert Scale (1 - Strongly Disagree to 7-Strongly Agree) on the following metrics:

- Is this recommendation made by the LLM appropriate / relevant for PCPs?
- Is this recommendation made by the LLM appropriate for that cognitive process in the Bloom's taxonomy?

Additionally, we asked the experts to comment on the appropriateness of the recommendations for PCP literacy. For each cognitive process in Bloom's taxonomy, we asked the experts: *Are these questions appropriate for PCP literacy? Are they appropriate for this learning module in Bloom's taxonomy?*

The overall results of the ratings of the four experts and a comparison of the LLMs is presented in the Results Section along with an analysis of the qualitative feedback received from the experts. Figure 2 shows a schematic of our process.

We now present the recommendations made by ChatGPT and Bard to teach novices about PCPs using Bloom's taxonomy.

### *Google Bard's recommendations*

Here are the recommendations provided by Bard to teach a novice about PCPs using Bloom's taxonomy.

- **Remember / Recognize**

(Q1.1) *What is a parallel coordinates chart?*

(Q1.2) *What are the advantages and disadvantages of PCPs?*

(Q1.3) *How do you create a parallel coordinates chart?*

- **Understand**

(Q2.1) *Explain how PCPs can be used to compare different sets of data.*

(Q2.2) *Give an example of how you would use a parallel coordinates chart to make a decision.*

(Q2.3) *Discuss the different ways that PCPs can be customized.*

- **Apply**

(Q3.1) *Create a parallel coordinates chart to show the distribution of test scores in your class.*

(Q3.2) *Use a parallel coordinates chart to compare the heights of boys and girls in your school.*

(Q3.3) *Interpret the results of a parallel coordinates chart to make a recommendation.*

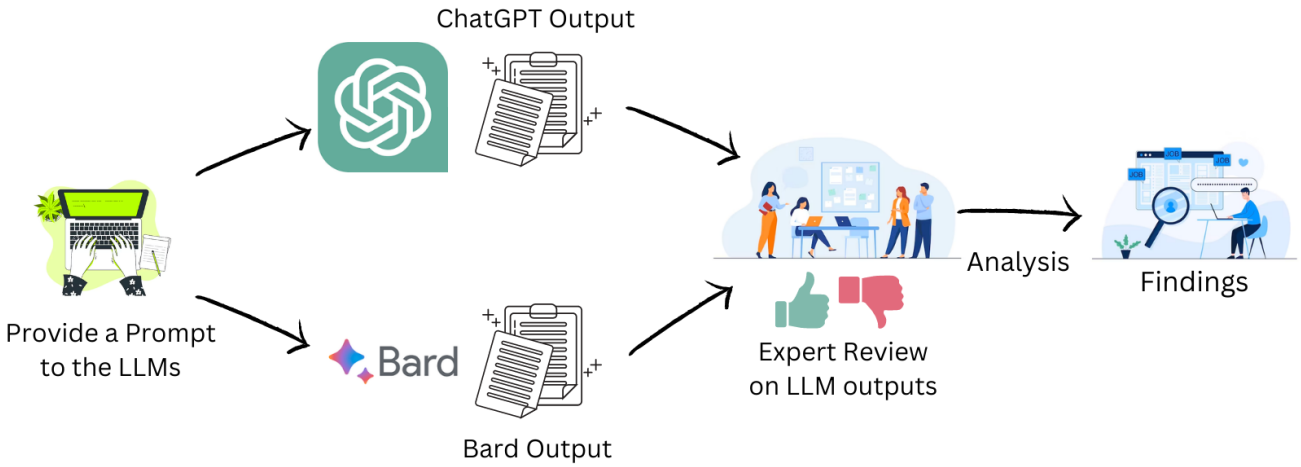


***Figure 2.*** *Methodology Overview - We provided the two Large Language Models (LLMs), OpenAI's ChatGPT and Google's Bard, with a prompt to teach a novice about PCPs using Bloom's Taxonomy. We then had four experts review the recommendations on appropriateness for PCPs and appropriateness for the specific cognitive process in Bloom's Taxonomy. We analyzed the expert reviews and present our findings.*

- **Analyze**

(Q4.1) *Identify the outliers in a set of data using a parallel coordinates chart.*

(Q4.2) *Compare the medians of two different sets of data using a parallel coordinates chart.*

(Q4.3) *Calculate the interquartile range of a set of data using a parallel coordinates chart.*

- **Create / Synthesis**

(Q5.1) *Create a new type of parallel coordinates chart that is better suited for a specific type of data.*

(Q5.2) *Develop a new method for interpreting PCPs.*

(Q5.3) *Write a report on the use of PCPs in data analysis.*

### *OpenAI ChatGPT's recommendations*

Here are the recommendations provided by ChatGPT to teach a novice about PCPs using Bloom's taxonomy.

- **Remember/Recognize**

(Q7.1) *Define what a parallel coordinates chart is and its purpose in data visualization.*

(Q7.2) *List the key components of a parallel coordinates chart.*

(Q7.3) *Explain how PCPs differ from other types of plots, such as scatter plots or bar graphs.*

- **Understand**

(Q8.1) *Summarize the steps involved in creating a parallel coordinates chart.*

(Q8.2) *Describe the types of data that can be effectively represented using PCPs.*

(Q8.3) *Explain the meaning of the different axes and lines in a parallel coordinates chart.*

- **Apply**

(Q9.1) *Given a dataset with multiple variables, create a parallel coordinates chart to visualize the relationships and patterns within the data.*

(Q9.2) *Compare and contrast PCPs of different datasets to identify similarities and differences in their patterns.*

(Q9.3) *Use a parallel coordinates chart to identify outliers or anomalies in a dataset.*

- **Analyze**

(Q10.1) *Analyze a parallel coordinates chart and identify trends, patterns, or relationships between variables.*

(Q10.2) *Evaluate the effectiveness of using a parallel coordinates chart to represent a specific dataset compared to other types of plots.*

(Q10.3) *Compare multiple PCPs to identify commonalities or differences in the relationships between variables.*

- **Evaluate / Critique**

(Q11.1) *Assess the strengths and limitations of using PCPs as a tool for data visualization.*

(Q11.2) *Critique a parallel coordinates chart and suggest improvements or alternative ways to present the data.*

(Q11.3) *Judge the appropriateness of using a parallel coordinates chart in a specific data analysis scenario, considering the nature of the data and the research question.*

- **Create / Synthesize**

(Q12.1) *Create a customized parallel coordinates chart with additional features or annotations to enhance the visual representation and interpretation of the data.*

(Q12.2) *Devise a strategy to handle missing or incomplete data when constructing a parallel coordinates chart.*

(Q12.3) *Design a set of guidelines or best practices for creating informative and visually appealing PCPs.*

## Results

In this section, we provide the results of the analysis of the expert reviews of the recommendations provided by each LLM on the following two metrics:

- Is this recommendation made by the LLM appropriate / relevant for PCPs?
- Is this recommendation made by the LLM appropriate for that cognitive process in the Bloom's taxonomy?

### *Remember*

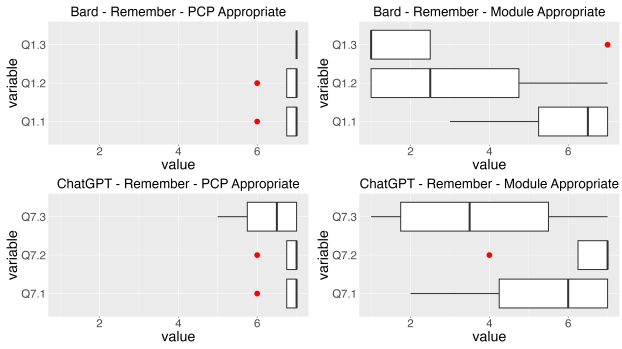


***Figure 3.*** *Remember analysis - The top row shows the results of the scores received for Bard's recommendation, while the bottom row shows the scores for ChatGPT's recommendations. While both the LLMs recommended tasks appropriate for PCP literacy, the right column shows that experts did not think that most of the recommendations were appropriate for the Remember stage of Bloom's taxonomy.*

Figure 3 shows the analysis of the scores of Bard's recommendations in the top row (Q1.1, Q1.2, Q1.3) and the analysis for ChatGPT's recommendation in the bottom row (Q7.1, Q7.2, and Q7.3). The left column shows the rating for whether the recommendation was appropriate for PCPs and the right column shows whether the recommendation was appropriate for the current module / cognitive process in Bloom's taxonomy. The top left image shows that the recommendations made by Bard for the Remember module/cognitive process are quite good for PCPs with a median of 7 for all three recommendations, whereas the top right image shows that experts did not rate the questions as appropriate for the Remember / Recognize stage of Bloom's taxonomy. Q1.2 requires a novice to state the advantages and disadvantages of PCPs. While the recommendation is appropriate for PCPs, it is too early in the learning process to ask a novice to compare and contrast. This question would be better suited for the Evaluate or Create/Synthesize stage of Bloom's taxonomy. Similarly, Q1.3 asks novices *How do you create a parallel coordinates chart?*. Given that there is a specific *Create* stage in Bloom's taxonomy, this question received a median score of 1 (strongly disagree) on whether it is appropriately placed in the Remember stage of Bloom's taxonomy.

The bottom row of Figure 3 shows the analysis of the scores of ChatGPT's recommendations. The left figure shows that all the three recommendations from ChatGPT had a high median score of 6.5 or higher. The right figure shows some disagreement among the expert reviewers for Q7.3. In Q7.3, the recommendation is that we ask a novice to *Explain how PCPs differ from other types of plots, such as scatter plots or bar graphs.* While this is a good recommendation for PCP literacy, the expert reviewer scores indicate that it is not the appropriate stage in Bloom's taxonomy, especially since the Analyze stage requires learners to compare and contrast the concept being learned (PCPs, in this case) with other similar concepts (other types of plots).

Based on analyzing the qualitative feedback of the experts, the common theme was that the recommendations from Google Bard were considered appropriate for PCP literacy, but some of them were not appropriate for the Remember module. The recommendations from ChatGPT were considered more appropriate for PCP literacy and for the Remember module, although one expert commented that "Q7.3 is too advanced for the Remember module of Bloom's taxonomy." It may be too early to ask a student to explain how PCPs differ from other types of plots. This is also reflected in the ratings as shown in Figure 3.

### *Understand*

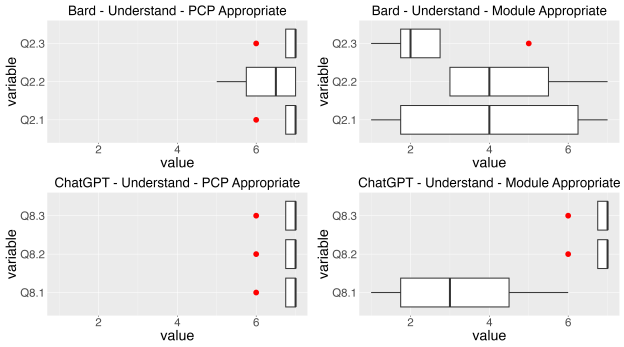


***Figure 4.*** *Understand analysis - Similar to the Remember stage, the questions recommended by the LLMs are appropriate for PCP literacy, but many are not appropriate for the Understand stage of Bloom's taxonomy. Recommendations such as Q8.2 and Q8.3 (bottom row) scored high on both the metrics, though.*

Figure 4 shows the analysis of the expert reviewers' scores for the Understand stage of Bloom's taxonomy for the LLMs. The top left plot shows that the recommendations from Bard were rated as highly appropriate for teaching PCP. The plot on the top right shows that the experts rated the questions quite low, implying that they were not appropriate for the *Understand* stage of Bloom's taxonomy. For example, Q2.3 asks learners to *Discuss the different ways that PCPs can be customized.* This is too early in the learning process and it may not be something that learners can do at this stage. Similarly, Q2.1 and Q2.2 also received low scores for asking learners to (Q2.1) *Explain how PCPs can be used to compare different sets of data* and (Q2.2) *Give an example of how you would use a parallel coordinates chart to make a decision*.

On the other hand, the recommendations by ChatGPT seem to have received high scores based on the bottom row of charts in Figure 4. The bottom left figure shows that all three recommendations were appropriate for teaching PCPs. The bottom right figure shows that Q8.2 and Q8.3 were appropriate for the *Understand* stage, but Q8.1 was not considered appropriate for this stage. This is probably because Q8.1 asks learners to *Summarize the steps in-*

*volved in 'creating' a parallel coordinates chart*. This question may be more appropriate for the *Apply* or the *Create / Synthesize* stage of the learning process.

We analyzed the qualitative feedback from the experts for this module and found that Bard's recommendation were appropriate for PCP literacy, but they were misplaced in the Understand module, where the goal is to ensure that a student can interpret and understand the chart accurately. ChatGPT's recommendations were found to be appropriate by all experts for PCP literacy, but 2/4 experts were concerned about "Q8.2 being too advanced for new learners." Q8.2 asks students to describe the types of data that can be effectively represented using PCPs.

## Apply

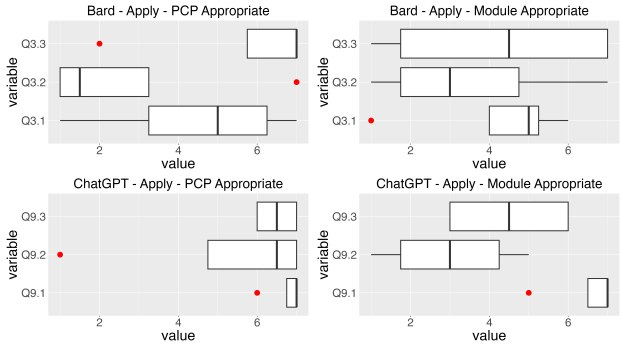


***Figure 5.*** *Apply analysis - The recommendations from Bard (top row) received much lower scores than the recommendations from ChatGPT (bottom row). The recommendations for both LLMs were not considered particularly appropriate for the Apply stage (right column).*

Figure 5 shows the results of the analysis of the reviewers' scores of the recommendations of the LLMs for the Apply stage. The top left plot shows that two of the three recommendations made by Bard received low scores for teaching about PCPs. Q3.2 *Use a parallel coordinates chart to compare the heights of boys and girls in your school* received a very low score and Q3.1 *Create a parallel coordinates chart to show the distribution of test scores in your class* also received a median score of 5. The top right plot shows that that recommendations were not appropriate for the *Apply* stage either with median scores of 5, 3, and 4.5 for the questions Q3.1, Q3.2, and Q3.3 respectively.

The bottom row of Figure 5 shows the scores received for ChatGPT. While the bottom left plot looks encouraging, with high scores for the recommendations being appropriate for teaching PCPs, the bottom right plot implies that some of the recommendations may not be appropriate for the *Apply* stage. While Q9.1 *Given a dataset with multiple variables, create a parallel coordinates chart to visualize the relationships and patterns within the data* is appropriate for PCPs and for the *Apply* stage, the (Q9.2) *Compare and contrast PCPs of different datasets to identify similarities and differences in their patterns* and (Q9.3) *Use a parallel coordinates chart to identify outliers or anomalies in a dataset* received lower scores for their appropriateness for the *Apply* stage.

The qualitative analysis of the experts' feedback indicates that Bard's suggestions were found to be somewhat related to PCP literacy, but questions such as Q3.2 were found to be "strange and inappropriate." All the experts thought that question Q3.1 was a good fit for this 'Apply' learning module, whereas Q3.3 was more appropriate for the previous 'Understand' module. Regarding the recommendations from ChatGPT, all four experts agreed that the "Although all questions are appropriate for PCP literacy," some of them (such as Q9.2) were more appropriate for the 'Analyze' module. This is also reflected in the rating shown in Figure 5.

## Analyze

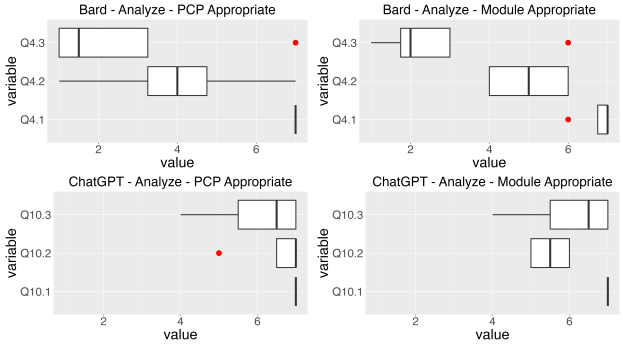


***Figure 6.*** *Analyze comparison - Two of the three recommendations provided by Bard for this stage received low scores for both the metrics (top row). The recommendations from ChatGPT were deemed more appropriate for PCP and received good scores for both the metrics.*

Figure 6 shows the analysis for the scores received for the LLM recommendations for the *Analyze* stage. The top row shows the scores for the recommendations received from Bard. Q4.1 *Identify the outliers in a set of data using a parallel coordinates chart* received a high score for being appropriate for PCP and for the *Analyze* stage. The other two recommendations for Q4.2 *Compare the medians of two different sets of data using a parallel coordinates chart* and Q4.3 *Calculate the interquartile range of a set of data using a parallel coordinates chart* received lower scores due to their low relevance to PCPs and not being appropriate for the stage in the taxonomy. Q4.3 is a particularly bad recommendation since it asks learners to calculate the interquartile range, a task that is not at all suitable for PCPs.

The bottom row shows the score for the recommendation received from ChatGPT. The high median scores for the both the bottom plots implies that the recommendations were applicable to PCP literacy and appropriate for the *Analyze* module. Q10.1 *Analyze a parallel coordinates chart and identify trends, patterns, or relationships between variables* scored a median of 7 in both the charts as it was deemed highly appropriate for teaching PCP and in the correct stage of the taxonomy.

The qualitative analysis provides more insight into the quality of the recommendations. Bard's recommendation were found to be particularly bad for PCP literacy and for this module. Q4.2 and 4.3 are not all relevant to PCP literacy and are an example of hallucinations when using an LLM [20]. The recommendations of ChatGPT for this module were relevant to PCP literacy and one of the experts said that while these tasks may be "difficult/vague" for an online, empirical study, "They might be appropriate for a traditional classroom setting."

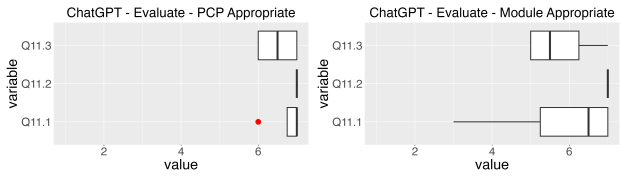


***Figure 7.*** *Evaluate analysis - There were no recommendations provided by Bard for the Evaluate stage and so there is no analysis for Bard. The recommendations provided by ChatGPT received high scores and were considered appropriate for PCP literacy and appropriate for the Evaluate stage.*

### *Evaluate*

Figure 7 shows the results of the analysis of the scores for the recommendations from ChatGPT. *Bard did not provide any recommendations for this stage.* This is unexpected, as Bloom's taxonomy (see Figure 1) contains the *Evaluate* stage as an integral step of the learning process.

ChatGPT generated three recommendations that scored high in terms of whether they were appropriate to teach learners about PCP and they were appropriate for the *Evaluate* stage as well. They included questions that required learners to *Assess the strengths and limitations of PCPs*, *Critique a PCP and suggest improvements*, and *Determine whether a PCP was the appropriate chart for a specific data analysis scenario.*

The recommendations of ChatGPT were found to be relevant and appropriate for PCP literacy by all the experts. One expert commented that "They require a learner to demonstrate that they can evaluate the appropriateness of using a PCP and what one gains from using that as compared to another visualization technique," whereas another expert said "they might be appropriate for a coursework/homework or an in-class exercise."

### *Create*

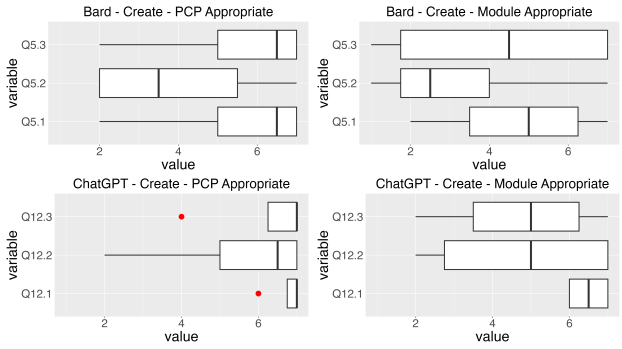


***Figure 8.*** *Create analysis - Two of the recommendations provided by Bard received high scores for being appropriate for PCP literacy (top left). All the recommendations from ChatGPT received high scores for PCP literacy (bottom left). Other than Q12.1 (bottom right), all the other five recommendations from Bard and ChatGPT received low scores for the Create stage of the taxonomy.*

Figure 8 shows the analysis of the scores for the LLM recommendations for the *Create* stage on Bloom's taxonomy. The top row shows the scores for the recommendations from Bard. While Q5.1 and Q5.3 received high scores for being PCP appropriate questions (as shown in the top left plot), they received low scores on being appropriate for the *Create* stage. Q5.2 *Develop a new method for interpreting PCPs* gets a low score on both counts due to the lack of relevance to PCPs and the stage in the taxonomy.

The bottom row shows the scores for ChatGPT's recommendations. All three recommendations received a high score for being appropriate for PCPs. They included questions requiring learners to Create a PCP, devise a strategy to handle missing data, and designing a set of guidelines to create visually appealing PCPs. The three recommendations from ChatGPT scored high on the scale of whether the recommendations were appropriate for the *Create* stage.

Most of the recommendations from Bard were found by all the experts to be irrelevant and not as appropriate for PCP literacy. One of the experts commented that the last recommendation, "Q5.3 - Write a report on the use of PCPs in data analysis" was " too open-ended and long. I don't think writing reports are appropriate in this case." Another expert did not completely agree and said, "the last question that asks the learner to write a report is relevant to PCP literacy." Examining the feedback for ChatGPT's recommendations, while the experts all agreed that the questions were appropriate for PCP literacy and for the Create learning module, 3/4 experts thought that the questions were "too advanced and would be a good question for a very advanced user or a researcher." The majority of the experts agreed that while the questions are appropriate for PCP literacy, they may be too difficult for a student learning about PCP for the first time.

## Discussion

Based on our analysis of the scores provided by visualization literacy experts, we observe that majority of the recommendations from the LLMs are applicable to PCP literacy. While many of them do not fall in the appropriate cognitive stage of Bloom's taxonomy, they could be used in conjunction with an expert in the visualization literacy, as a starting point.

Overall, based on the expert reviewers, we can conclude that the **recommendations from OpenAI ChatGPT's are contextual and applicable for PCP literacy**, whereas the recommendations from Google's Bard were comparatively less relevant (hallucinations [20]) to Parallel Coordinates Literacy. For example, for the Apply recommendation, Bard recommended "Use a parallel coordinates chart to compare the heights of boys and girls in your school." This recommendation is not appropriate and may be a better recommendation for a box plot. Similarly, Bard recommended, "Create a parallel coordinates chart to show the distribution of test scores in your class." This recommendation is also not appropriate for visualizing data using a parallel coordinates plot.

## Conclusion and Future Work

Recommendations of two popular LLMs (ChatGPT and Bard) to teach novices about PCPs using Bloom's taxonomy were evaluated by visualization literacy experts. The experts evaluated every recommendation on two metrics related to their appropriateness for PCP literacy and appropriateness for one of the six stages in Bloom's taxonomy.

Based on the analysis of the scores of the expert reviewers, we found that while there were some useful recommendations for teaching novices about parallel coordinates, there were some limitations in terms of the appropriate cognitive process where that recommendation belonged. In some cases, the recommendations

provided were not at all appropriate for PCP literacy. We also found that both the LLMs made recommendations at times that were too difficult or advanced for a novice learning about PCPs for the first time. We recommend that a human expert working with a LLM may lead to a practicable set of recommendations for visualization literacy.

In the future, we plan to perform some machine-based validation, where we will test the ability of each LLM to validate their own recommendations and the recommendations provided by other LLMs.

## References


[1] Basak Alper, Nathalie Henry Riche, Fanny Chevalier, Jeremy Boy, and Metin Sezgin. Visualization literacy at elementary school. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 5485–5497, 2017.

[2] Patricia Armstrong. Bloom's taxonomy. *Vanderbilt University Center for Teaching*, 2010.

[3] Benjamin S Bloom, Max D Engelhart, Edward J Furst, Walquer H Hill, and David R Krathwohl. Taxonomy of educational objectives: the classification of educational goals: handbook i: cognitive domain. Technical report, New York, US: D. Mckay, 1956.

[4] BS Bloom. Taxonomy of educational objectives: The classification of educational goals. *Cognitive domain*, 1956.

[5] Katy Borner. *Atlas of knowledge: anyone can map*. MIT Press, 2015.

[6] Katy Börner, Andreas Bueckle, and Michael Ginda. Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proceedings of the National Academy of Sciences*, 116(6):1857–1864, 2019.

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

[8] Jeremy Boy, Ronald A Rensink, Enrico Bertini, and Jean-Daniel Fekete. A principled way of assessing visualization literacy. *IEEE transactions on visualization and computer graphics*, 20(12):1963–1972, 2014.

[9] Zhutian Chen, Chenyang Zhang, Qianwen Wang, Jakob Troidl, Simon Warchol, Johanna Beyer, Nils Gehlenborg, and Hanspeter Pfister. Beyond generating code: Evaluating gpt on a data visualization course. *arXiv preprint arXiv:2306.02914*, 2023.

[10] Kunming Cheng, Qiang Guo, Yongbin He, Yanqiu Lu, Shuqin Gu, and Haiyang Wu. Exploring the potential of gpt-4 in biomedical engineering: the dawn of a new era. *Annals of Biomedical Engineering*, pages 1–9, 2023.

[11] Fanny Chevalier, Nathalie Henry Riche, Basak Alper, Catherine Plaisant, Jeremy Boy, and Niklas Elmqvist. Observations and reflections on visualization literacy in elementary school. *IEEE computer graphics and applications*, 38(3):21–29, 2018.

[12] Sara Di Bartolomeo, Giorgio Severi, Victor Schetinger, and Cody Dunne. Ask and you shall receive (a graph drawing): Testing chatgpt's potential to apply graph layout algorithms. *arXiv preprint arXiv:2303.08819*, 2023.

[13] E. E. Firat, A Denisova, and R S. Laramee. Treemap literacy: A classroom-based investigation. In *Eurographics Proceedings*, 2020.

[14] Elif E Firat, Alena Denisova, Max L Wilson, and Robert S Laramee. P-lite: A study of parallel coordinate plot literacy. *Visual Informatics*, 6(3):81–99, 2022.

[15] Elif E Firat, Alark Joshi, and Robert S Laramee. Interactive visualization literacy: The state-of-the-art. *Information Visualization*, 21(3):285–310, 2022.

[16] Elif E Firat, Alark Joshi, and Robert S Laramee. Vislite: Visualization literacy and evaluation. *IEEE Computer Graphics and Applications*, 42(3):99–107, 2022.

[17] Elif E Firat, Colm Lang, Bhumika Srinivas, Ilena Peng, Robert S Laramee, and Alark Joshi. A constructivism-based approach to treemap literacy in the classroom. In *COMPUTER GRAPHICS forum*, volume 42 (2), 2023.

[18] Lily W Ge, Yuan Cui, and Matthew Kay. Calvi: Critical thinking assessment for literacy in visualizations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–18, 2023.

[19] Google. Bard. `https://bard.google.com`, 2023. [Online; accessed May-04-2023].

[20] Nuno M Guerreiro, Duarte Alves, Jonas Waldendorf, Barry Haddow, Alexandra Birch, Pierre Colombo, and André FT Martins. Hallucinations in large multilingual translation models. *arXiv preprint arXiv:2303.16104*, 2023.

[21] David R Krathwohl. A revision of bloom's taxonomy: An overview. *Theory into practice*, 41(4):212–218, 2002.

[22] Andrey Krekhov, Michael Michalski, and Jens Krüger. Integrating visualization literacy into computer graphics education using the example of dear data. *arXiv preprint arXiv:1907.04730*, 2019.

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

[24] Brady D Lund and Ting Wang. Chatting about chatgpt: how may ai and gpt impact academia and libraries? *Library Hi Tech News*, 40(3):26–29, 2023.

[25] OpenAI. Chatgpt: Optimizing language models for dialogue. `https://openai.com/blog/chatgpt/`, 2022. [Online; accessed May-04-2023].

[26] Bianca Pallaro and Alicia Parlapiano. Visualizing the 13.6 billion in u.s. spending on ukraine. `https://www.nytimes.com/interactive/2022/03/18/upshot/ukraine-aid-details.html`, 2022. [Online; accessed 23-April-2022].

[27] Saugat Pandey and Alvitta Ottley. Mini-vlat: A short and effective measure of visualization literacy. *arXiv preprint arXiv:2304.07905*, 2023.

[28] Ilena Peng, Elif E Firat, Robert S Laramee, and Alark Joshi. Evaluating bloom's taxonomy-based learning modules for parallel coordinates literacy. In *COMPUTER GRAPHICS forum*, volume 41, 2022.

[29] Puripant Ruchikachorn and Klaus Mueller. Learning visualizations by analogy: Promoting visual literacy through visualization morphing. *IEEE transactions on visualization and computer graphics*, 21(9):1028–1044, 2015.

[30] Margot Sanger-Katz and Alicia Parlapiano. The democrats

have a lot of cutting to do. https://www.nytimes.com/bill.html, 2021. [Online; accessed 23-April-2022].

[31] Harry Stevens. Majority of women ages 15 to 44 would face new post-roe abortion limits. https://www.washingtonpost.com/nation/2022/05/04/abortion-numbers-us-roe-opinion/, 2022. [Online; accessed 23-April-2022].

[32] Christina Stoiber, Davide Ceneda, Markus Wagner, Victor Schetinger, Theresia Gschwandtner, Marc Streit, Silvia Miksch, and Wolfgang Aigner. Perspectives of visualization onboarding and guidance in va. *Visual Informatics*, 6(1):68–83, 2022.

[33] Christina Stoiber, Conny Walchshofer, Florian Grassinger, Holger Stitz, Marc Streit, and Wolfgang Aigner. Design and comparative evaluation of visualization onboarding methods. In *Proceedings of the 14th International Symposium on Visual Information Communication and Interaction*, pages 1–5, 2021.

[34] Miriam Sullivan, Andrew Kelly, and Paul McLaughlan. Chatgpt in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning and Teaching*, 2023.

[35] Hao Yu. Reflection on whether chat gpt should be banned by academia from the perspective of education and teaching. *Frontiers in Psychology*, 14:1181712, 2023.