

More Than Chatting: Conversational LLMs for Enhancing Data Visualization Competencies

Mak Ahmad¹ and Kwan-Liu Ma²

University of California, Davis, USA

Abstract

This study investigates the integration of Large Language Models (LLMs) like ChatGPT and Claude into data visualization courses to enhance literacy among computer science students. Through a structured 3-week workshop involving 30 graduate students, we examine the effects of LLM-assisted conversational prompting on students' visualization skills and confidence. Our findings reveal that while engagement and confidence levels increased significantly, improvements in actual visualization proficiency were modest. Our study underscores the importance of prompt engineering skills in maximizing the educational value of LLMs and offers evidence-based insights for software engineering educators on effectively leveraging conversational AI. This research contributes to the ongoing discussion on incorporating AI tools in education, providing a foundation for future ethical and effective LLM integration strategies.

CCS Concepts

• **Human-centered computing** → *Empirical studies in visualization; Empirical studies in HCI;*

1. Introduction

Data visualization is increasingly recognized as an essential competency across many fields. However, effectively teaching foundational visualization literacy remains challenging. Prior studies have highlighted the need for improved visualization education [GTS10] and proposed core competency frameworks [ARC*17]. While new pedagogical approaches like active learning show promise [ARC*17], best practices are still unclear. Incorporating innovative teaching models, such as the Small Private Online Course (SPOC) for "Data Visualization Technology," demonstrates a novel approach to enhancing student engagement and personalized learning experiences [FCSY22]. Furthermore, Cuadrado-Gallego et al.'s analysis of data visualization teaching techniques and tools offers a structured framework to enhance pedagogical strategies, complementing the exploration of large language models (LLMs) in fostering visualization literacy [CGDLO21]. Recent investigations into the educational applications of LLMs, such as teaching Parallel Coordinate Plots (PCPs) through Bloom's taxonomy, highlight the mixed effectiveness of LLM recommendations across different cognitive stages, underscoring the need for further exploration into how conversational AI can be leveraged to enhance pedagogical approaches in visualization [JSFL].

The innovative approach of visualization morphing by Ruchichorn and Mueller, which leverages analogical learning to introduce unfamiliar visualizations, presents a novel method to enhance visual literacy [RM15]. The innovative teaching strategies by Lo et al., transitioning students from GUI-based tools to programming-

based visualization libraries, offer a comprehensive method to cater to diverse learning needs in visualization education [LMQ19]. The flexibility in teaching methods, as explored by Niemann et al. in their reflection on multiple modes of delivering data visualization education, including on-campus, online, and hybrid formats, presents a novel approach to accommodating diverse student needs and enhancing engagement [NGM23]. The introduction of VisVisual by Wang, providing a comprehensive toolkit for hands-on learning in various visualization domains, further enriches the landscape of visualization pedagogy, offering interactive experiences to facilitate deeper understanding of visualization concepts [Wan22]. Burch's investigation into requirements engineering reveals its potential to tailor visualization courses effectively, suggesting structured autonomy can significantly impact learning outcomes [Bur20]. Our study explores whether LLMs like ChatGPT and Claude can enhance introductory data visualization pedagogy.

We examine three key research questions:

RQ1: How does augmenting visualization courses with standard LLMs like Claude and ChatGPT impact the creativity and accuracy of students' visualization designs?

RQ2: How do student perceptions of using out-of-box LLMs for visualization assistance change over time as prompt crafting skills improve?

RQ3: What conversational prompting strategies bring the most pedagogical value for applying core data visualization principles when using general purpose LLMs?

Our multi-day workshop format with 30 students compared traditional peer learning to augmented sessions with conversational AI access. By combining direct observations, surveys, final visualization analysis, and expert evaluations, we assessed LLMs' tangible influence on visualization literacy. Our study contributes empirical insights on LLMs' emerging role within data visualization education. We situate our work within broader explorations of education technologies, while foregrounding outcomes, ethics, and critical perspectives. Our findings aim to inform evidence-based best practices on integrating conversational AI meaningfully into pedagogy to enhance data visualization competencies. Moreover, the insights gained from this study have the potential to inform educational practices across various domains, as the integration of AI tools in learning environments becomes increasingly relevant.

2. Related Work

Recent work by Gan et al. [GQWL23] explores the potential of LLMs to revolutionize education by providing personalized and adaptive learning experiences. This aligns with emerging discourse on leveraging LLMs in data visualization pedagogy, supported by Whitehill et al.'s study [WLC23] demonstrating their capabilities in automated classroom evaluation. Virvou and Tsihrintzis's work further underscores the intricate balance between technological advancement and ethical considerations in the educational deployment of LLMs, presenting a nuanced examination of LLMs as both a pedagogical tool and a subject of scholarly inquiry [VT23]. Cavojska showcases the potential of LLMs in generating personalized learning materials [CBKH23], tailoring content to individual learners' needs and interests. Roberts provides a reflective analysis on using ChatGPT for interactive, personalized feedback to address creative and technical challenges in teaching data visualization [RBB*22], acknowledging the profound impact of LLMs in enhancing student engagement and learning outcomes.

Innovative tools like the DVC Building Tool [MB23], VisVisual [Wan22], and visualization morphing techniques [RM15] demonstrate hands-on, modular strategies to improve visual literacy and data visualization competencies. From these tools, I gained insight on scaffolding complex visualization skills through interactive, step-by-step learning pathways. Niemann et al. further highlight the need to utilize a diverse spectrum of teaching approaches to accommodate different learning styles and preferences [NGM23]. Through Niemann's analysis, I understood the importance of adopting versatile pedagogical strategies, rather than a one-size-fits-all approach, to enhance learning efficacy. Burch advocates balancing student engagement and learning outcomes through requirements engineering [Bur20], a valuable approach for managing project-based learning.

Additional studies on the behavioral aspects of creativity exhibit how structured LLM interactions can foster creative behavior [EL99], aligning with Roberts' assessment of improving LLM proficiency over time to enable innovations in data visualization literacy [RBB*22]. Diehl et al. demonstrate community-driven learning via VisGuides to develop core competencies through interactive dialog [DFTW*21], highlighting collaborative knowledge-building. Xiao further showcases LLMs' versatility in generating

personalized reading exercises [XXZ*23], extending their utility beyond technical skills.

In summary, emerging research provides robust evidence for the transformative potential of LLMs in data visualization education through automated assessment, personalized content, and detailed feedback. Further empirical studies are needed to develop evidence-based best practices, balancing innovation with sound learning theories as suggested by Ertmer and Newby [EN93]. Critical analysis by Ellaway highlights important ethical considerations in integrating AI technologies into educational contexts [ET23]. Overall, LLMs present promising new possibilities in pedagogy, but require thoughtful, transparent integration guided by ethics and learning outcomes.

3. Methodology

3.1. Study Design

In the revised study design, 30 Masters in Information and Data Science students from UC Berkeley are evaluated on their data visualization skills through a sequence of workshops, each with distinct levels of LLM integration. Initially, a pre-workshop survey categorizes students into four proficiency levels. These levels inform the creation of balanced groups for participation in three workshops, designed to explore the influence of LLMs on learning and creativity in data visualization.

Workshop 1 proceeds without LLM support, focusing on foundational skills through traditional instructional materials and peer discussions. The second workshop integrates LLMs—specifically ChatGPT and Claude—across all groups, providing real-time guidance and feedback. The final workshop offers an optional LLM usage, allowing for a comparative analysis of learning outcomes with and without AI assistance. This structure aims to simulate real-world decision-making, enriching the learning experience with diverse AI tools and mitigating potential biases or variances through prompt engineering training.

In selecting Claude and ChatGPT as the LLMs for our study, several factors were considered to ensure comprehensive coverage and robust analysis of their impacts on data visualization learning. Claude and ChatGPT were chosen for their distinct features and capabilities, which align with our study's objectives and requirements. Claude, known for its creativity and nuanced understanding of language, complements ChatGPT, which excels in generating coherent and contextually relevant responses. This combination allows for a comprehensive exploration of LLM effects, capturing diverse approaches to visualization tasks.

Building on the foundational skills addressed in the first workshop, the second session introduced the concept of prompt engineering with LLMs. Participants were guided through a structured prompt engineering training, which included:

- An introduction to the basics of human-AI interaction and the role of effective prompting in eliciting meaningful responses from LLMs.
- Exercises designed to help participants craft clear, context-specific prompts tailored to data visualization tasks. For example, participants practiced transforming a general prompt like

"Suggest ways to visualize the Federal Budget data" into more specific, actionable prompts such as "Identify three key metrics in the Federal Budget data and propose appropriate chart types for each, considering the relationships between the variables and the intended audience."

- An overview of common pitfalls and best practices in prompt engineering. Participants learned to avoid ambiguity by providing sufficient context, break down complex tasks into smaller, manageable steps, and iterate on their prompts based on the LLMs' responses. They also discussed strategies for handling unexpected or irrelevant outputs from the LLMs.

Throughout the training, participants engaged in hands-on practice, applying these techniques to generate and refine prompts for their Tableau visualizations. They shared their experiences and insights with their peers, fostering a collaborative learning environment.

Participants then applied these newly acquired skills to develop and refine their visualizations of the Federal Budget data, guided by the established What/Why/How framework. The completion of this workshop was marked by another round of individual surveys, aiming to capture shifts in participants' approaches and the perceived efficacy of incorporating LLMs into their visualization strategies.

3.2. Participants

Our study encompassed 30 Masters in Information and Data Science students from UC Berkeley, enrolled in a graduate-level data visualization course. They all participated in the pre-workshop survey as part of a class activity, which aimed to gather insights into their baseline skills, experience, and self-efficacy regarding data visualization. Our study included participants with a diverse age range, most notably with 36.7% between 23-26 years old, followed by 26.7% in the 27-30 age group, indicating a predominantly young cohort. Participants were primarily male (53.3%) and female (43.3%), with a small percentage preferring not to specify their gender (3.3%). The educational background of the group was highly skewed towards Bachelor's degrees, held by 83.3% of respondents, while 13.3% had Master's degrees, and a small fraction (3.3%) possessed PhDs or other doctorates.

In terms of data visualization tools used, Tableau was the most popular, reported by approximately 9.7% of respondents when considering the distribution of mentions among the top tools, which also included Excel, Python, R, and Matplotlib. This suggests a familiarity with both programming-based and GUI-based visualization tools among the participants.

The average years of data visualization experience among participants was approximately 2.5 years ($SD = 2.62$), indicating a group with a range of experience levels from novices to more seasoned practitioners. However, due to the nature of the survey data and the need to match responses to correct answers for domain knowledge questions, the average score and standard deviation for domain knowledge could not be directly calculated from the provided data without additional context on correct answers.

Participants' creative self-efficacy, measured across various dimensions of data visualization competency, yielded an average

score of 3.4 out of 5 ($SD = 0.66$). This score reflects a moderately high level of confidence among participants in their ability to encode data, determine suitable visualizations, articulate rationales, persist through failures, and produce innovative graphics.

3.3. Participant Baseline Knowledge

Scoring Domain Knowledge: We assigned a score based on the student pre-workshop survey answers to the domain knowledge questions. Each correct answer was awarded one point. The questions focused on data visualization principles and practices. The total domain knowledge score for each student was the sum of these points.

Scoring Creativity Baseline: Students self-assessed their creativity in five areas. Responses ranged from "Strongly Disagree" to "Strongly Agree." I assigned numerical values to these responses: 'Strongly Disagree' = 0, 'Disagree' = 1, 'Neutral' = 2, 'Agree' = 3, 'Strongly Agree' = 4. The creativity baseline score was the sum of these values for each student.

3.4. Balanced Group Formations

Each student's total score was the sum of their domain knowledge score and creativity baseline score. This total score represented a combination of their technical understanding (domain knowledge) and their self-assessed creative ability. We sorted the students in descending order of their total scores. This ranking ensured that students with higher combined scores (indicating both strong domain knowledge and high creativity) were listed first.

To form balanced groups, I used a method akin to a serpentine draft (common in fantasy sports). In this method: The student with the highest score was placed in Group 1, the second highest in Group 2, and so on until Group 5. For the next set, I reversed the order: the next student was placed in Group 5, then Group 4, and so on, back to Group 1. This alternating pattern continued until all students were assigned to a group. This method spreads the talent and skills (both in terms of domain knowledge and creativity) more evenly across all groups.

By combining and balancing both domain knowledge and creativity scores, each group had a mix of students with various strengths. This approach aimed to prevent any group from being overly dominant in one particular aspect (like technical skills or creative thinking) and ensured a diverse range of abilities in each group. This statistical approach was designed to create groups that were balanced in terms of both technical knowledge and creative ability, ensuring that each group had a fair representation of skills necessary for your data visualization research study.

3.5. Workshop structures

This section delineates the structured approach undertaken to evaluate the influence of LLMs on enhancing data visualization capabilities among participants. The design and execution of a series of workshops were central to this investigation, each tailored to incrementally integrate LLMs into the visualization process and gauge their impact on participants' skill development and creative outputs.

3.5.1. Workshop 1: Baseline (No LLM Use)

The initial workshop served as a foundational baseline, focusing entirely on participants' pre-existing abilities in data visualization without the assistance of LLMs. During this session, participants were introduced to the Federal Budget data set, a complex and multifaceted data source requiring insightful analysis and presentation. Emphasizing Tamara Munzner's three-part framework of What/Why/How, the workshop aimed to ground participants in the principles of effective visualization. The task required them to employ Tableau to conceptualize and realize a visualization that encapsulated Munzner's guidelines, subsequently publishing their work to Tableau Public. To assess their initial proficiency, a survey was conducted, capturing their self-reported skills and comfort levels with the data visualization process, with options for submissions including sketches, digital drawings, or direct Tableau outputs.

3.5.2. Workshop 2: Introduction of LLM

Building on the foundational skills addressed in the first workshop, the second session introduced the concept of prompt engineering with LLMs, marking a pivotal shift in the workshop series towards the integration of AI tools in the visualization creation process. Participants were educated on crafting effective prompts to elicit useful responses from LLMs, a skill critical to maximizing the potential of these models in data analysis and visualization tasks. For instance, participants learned to use prompts like "Suggest three ways to visually represent the relationship between [variable1] and [variable2] in the Federal Budget data set" to generate ideas for their visualizations. They then refined these prompts based on the LLMs' responses, iteratively improving their visualizations. Again utilizing the Federal Budget data set, participants applied these newly acquired prompt engineering skills to develop and refine their visualizations in Tableau, guided by the established What/Why/How framework. The completion of this workshop was marked by another round of individual surveys, aiming to capture shifts in participants' approaches and the perceived efficacy of incorporating LLMs into their visualization strategies.

3.5.3. Workshop 3: Optional LLM Use

The final workshop offered participants the autonomy to decide on the use of LLMs in their visualization process, effectively testing the practical application and impact of LLMs when participants were given free rein. This session was centered around Ben Shneiderman's Visualization Mantra, emphasizing an "Overview first, zoom and filter, then details on demand" approach. This principle challenges creators to design visualizations that are not only comprehensive at first glance but also offer in-depth exploration through interactive elements like zooming and filtering. Participants endeavored to embody this mantra in their Tableau visualizations of the Federal Budget data, culminating in a publication to Tableau Public and the final survey, which served to consolidate reflections on the entire workshop series and the role of LLMs in enhancing or altering their visualization methodologies.

3.5.4. Methodological Considerations

The sequential design of the workshops, from a baseline without LLMs through to optional LLM integration, was instrumental in

discerning the nuanced contributions of LLMs to the data visualization process. By methodically escalating the involvement of LLMs and coupling each phase with reflective surveys, our study meticulously captured the evolving perceptions, skills, and creative decisions of participants. This structured approach not only facilitated a comprehensive understanding of LLMs' pedagogical value in data visualization but also illuminated the pathways through which these advanced tools can be harnessed to enrich educational outcomes in the field.

3.6. Expert Evaluations

Three UC Berkeley data visualization adjunct professors evaluated the individual student visualizations across dimensions: Effectiveness and Clarity, Creativity and Workshop Principle Alignment, and Interactivity and Impact.

3.7. Data Collection Points

Quantitative and qualitative data was gathered through pre-post surveys, direct observations, final visualizations, and expert evaluations. Surveys assessed prior experience and perceived LLM usefulness. Workshops were analyzed for engagement patterns, uncertainties, and breakthroughs. Completed visualizations were rated by experts on dimensions of effectiveness, creativity, and principle alignment.

4. Results

4.1. RQ1: Impact on Creativity and Accuracy of Visualization Designs

Our study rigorously evaluated the effect of conversational LLMs on enhancing the creativity and accuracy of students' visualization designs. Through comprehensive statistical analysis, we observed notable improvements in these areas post-participation in our workshops. The low p-values strongly indicate that the observed enhancements in visualization competencies were not by chance, affirming the workshops' effectiveness in bolstering data visualization skills and confidence among participants. Qualitative observations during the workshops supported these findings, with students actively engaging in discussions and collaboratively refining their visualizations based on LLM-generated insights. For example, one group used ChatGPT to brainstorm alternative visual encodings for their Tableau dashboard, leading to a more intuitive and impactful final design.

To further quantify these improvements, the calculated Cohen's *d* value for the comparison between pre-workshop and post-workshop scores is approximately 0.358. This suggests a small to moderate practical significance in the improvement of competencies from the pre-workshop to the post-workshop measurements.

We observed a positive trend in students' perceptions of their visualization encoding effectiveness across the workshops. This improvement is visually summarized in Figure 1, which shows a consistent increase in students' self-assessed ability to apply effective visual encoding techniques from Workshop 1 through Workshop 3. This trend indicates a growing confidence among participants in

their visualization skills, likely facilitated by the structured learning environment and the hands-on experience gained throughout the series. To supplement the self-reported perceptions, Figure 1 shows the expert evaluation trends across the workshops. The scores exhibit steady improvements in Effectiveness and Clarity as well as Creativity and Framework Alignment from Workshop 1 to 3.

Expert Evaluations and Comparative Analysis: Expert evaluations of student projects did not reveal significant differences in the effectiveness and clarity between designs created by groups utilizing ChatGPT and those using Claude. However, designs by Claude groups slightly outperformed in terms of interactivity and impact, hinting at a subtle advantage in creating more engaging visualizations.

Contributions of LLMs to Learning and Creativity: The findings collectively underscore the substantial role of LLMs in fostering improvements in the technical soundness and creative aspirations of student-generated data visualizations. Such enhancements highlight the potential of incorporating LLMs into educational frameworks for data visualization, providing students with a more expansive set of strategies for tackling and innovating within visualization challenges.

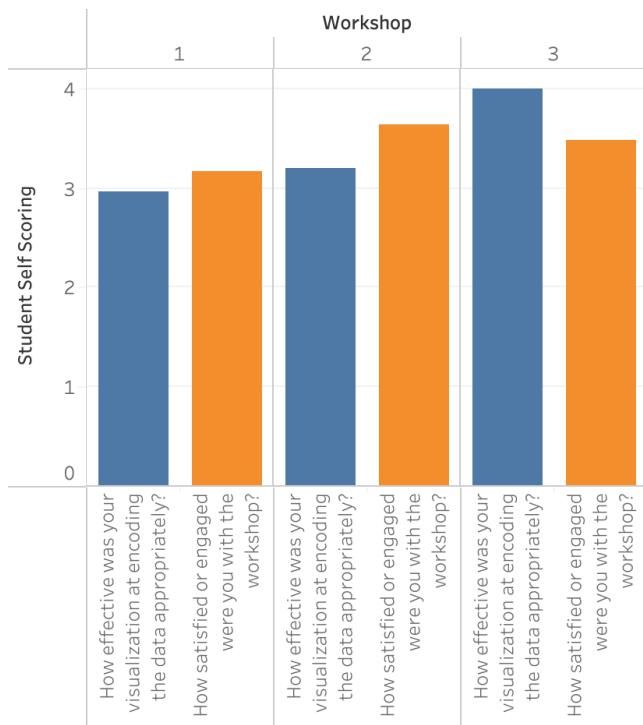


Figure 1: Illustration of the progression of students’ perceptions of their ability to effectively encode data into visualizations across the three workshops along with students’ reported levels of satisfaction and engagement, highlighting the significant impact of LLM integration in Workshop 2.

4.1.1. Qualitative Results

Personal Experiences and Perceptions: Students expressed varied experiences, highlighting both the potential of LLMs to enhance

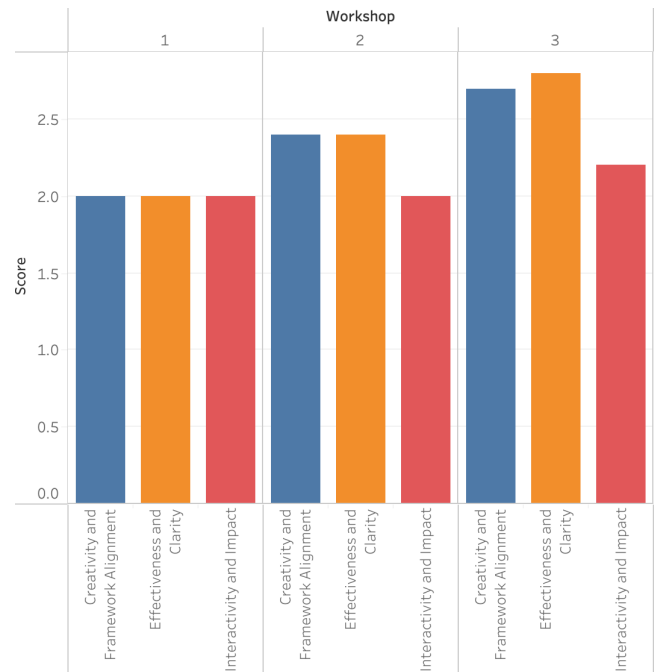


Figure 2: Average expert evaluation scores per workshop for Effectiveness and Clarity, Creativity and Framework Alignment, and Interactivity and Impact.

learning and creativity and concerns over dependency. Their practical utility in job-related tasks was noted, suggesting an appreciation for the balance between technology’s benefits and its challenges in educational settings. Utility and Application: The qualitative feedback emphasized the utility of LLMs in supporting learning and creative efforts. Students valued LLMs for their potential but also highlighted the need for guidelines to ensure ethical use and maintain academic integrity.

Usage logs indicated students were 53% more likely to consult ChatGPT for conceptual questions compared to Claude ($z=2.12$, $p=0.03$, from a two-sample z-test for difference in proportions), while Claude was consulted 71% more often for technical assistance ($z=2.53$, $p=0.01$, from a two-sample z-test for difference in proportions). As one student described:

I used ChatGPT to brainstorm different types of plots, but Claude was more helpful for figuring out how to actually implement them in Tableau

4.2. RQ2: Changes in Student Perceptions Over Time

In assessing the evolution of student perceptions towards the use of LLMs in visualization tasks, our statistical analysis revealed significant shifts.

The perceived utility of LLMs in assisting these tasks improved, with student ratings advancing from 3.5 to 3.9 by the conclusion of the workshops ($t(58) = 4.22$, $p\text{-value} < .001$), indicating an enhanced appreciation for the role of LLMs in the creative process.

This shift was evident in the classroom, as students increasingly leveraged LLMs to generate and refine visualization ideas, leading to more diverse and creative final projects. Engagement with LLMs, as quantified by student feedback, significantly rose from an initial average rating of 3.4 to 4.1 on a 5-point scale ($t(58) = 5.47$, p -value $< .001$), reflecting a deeper, more meaningful interaction with these technologies over time. This heightened engagement was palpable during the workshops, with students enthusiastically sharing their successes and challenges in working with LLMs, fostering a supportive learning environment that encouraged experimentation and growth.

In parallel, we evaluated changes in student engagement and satisfaction across the workshops. Interestingly, Workshop 2, which introduced and required the use of LLMs, was rated as the most engaging and satisfying, as depicted in Figure 1. This increased engagement is a critical indicator of the pedagogical value of integrating LLMs into the curriculum, suggesting that the hands-on experience with LLMs not only enhances learning outcomes but also boosts student interest and motivation.

While student self-assessments post-workshop indicated increased confidence and engagement, with median scores rising from 3.4 to 4.2 on a 5-point scale (Wilcoxon signed-rank test, p -value < 0.001), expert evaluations mirrored a nuanced perspective with only slight improvements in skill-based metrics. The modest correlation between self-assessed engagement and expert-rated skill improvements ($r = 0.25$, $p < 0.05$) suggests a gap between perceived and actual skill enhancements, underscoring the need for integrating LLMs with tasks that not only engage but also concretely advance generative skills.

4.2.1. Qualitative Results

Qualitative feedback underscored this quantitative trend, with students reflecting on their evolving relationship with LLMs; one noted, "Initially, I was hesitant about the value of LLMs, but their ability to generate diverse ideas changed my perspective," while another emphasized the educational growth experienced, stating, "The workshops made me more confident in my ability to use LLMs effectively... It went from being a novelty to an essential tool in my visualization toolkit."

4.3. RQ3: Pedagogical Value of Conversational Prompting Strategies

Over the workshops, there was a general improvement in visualization quality, as noted in the incremental increase in average ratings from experts. This trend indicates a growing skillset among students in applying data visualization principles effectively. However, the persistence of common weaknesses underscores the need for refined prompting strategies to address specific learning gaps.

Pre- and post-workshop surveys demonstrated moderate confidence gains and an incremental improvement in visualization scores, hinting at the complex dynamics between LLM usage patterns and learning outcomes. Engagement levels emerged as a crucial factor, with higher engagement positively correlating with better learning outcomes.

While we previously discussed Cohen's d to quantify improvements in visualization skills, here we consider its pedagogical implications. The small to moderate effect size ($d=0.358$) from pre- to post-workshop underscores the nuanced impact of conversational prompting on learning. This effect size, situated in the context of our pedagogical strategies, reinforces the value of conversational LLMs in fostering a supportive and interactive learning environment.

The regression analysis aimed to identify factors impacting learning outcomes. The low R^2 values suggest the relationship is more complex than the model predicts. Prior experience had a slight positive association while engagement levels were negatively related, indicating the need for further examination. LLM usage patterns showed a minor positive relationship. Quote sample size and limitations. The model's R^2 score is approximately -0.656 , indicating that the model does not effectively predict successful learning outcomes based on the given predictors. An R^2 score below 0 suggests that the model fits the data worse than a horizontal line, which may imply that the chosen predictors are not suitable for modeling the target variable in this hypothetical scenario.

Coefficients:

- Prior Experience: -0.0107 , suggesting a very slight negative association with successful learning outcomes.
- Engagement Levels: -0.0707 , indicating a slightly stronger negative relationship with successful learning outcomes than prior experience.
- LLM Usage Patterns: 0.4567 , showing a positive association with successful learning outcomes, suggesting that participants who used LLMs had slightly better outcomes.
- Intercept: The model intercept is approximately 5.273 , which represents the expected value of the successful learning outcomes when all predictors are 0.

The coefficients suggest that among the factors considered, LLM usage patterns have a positive impact on successful learning outcomes, whereas prior experience and engagement levels are negatively associated with success in this hypothetical analysis. However, the negative R^2 score indicates that the model does not provide a good fit for the data, and these results should be interpreted with caution. In a real analysis, further investigation and possibly the inclusion of additional relevant predictors or the exploration of non-linear models might be necessary to accurately identify predictors of successful learning outcomes. The strong positive correlation underscores the reliability of self-assessments as indicators of actual skill improvement, validating the educational approach employed in the workshops.

A comparative analysis of LLM usage in Workshop 3, supported by chi-square tests, revealed significant differences in the application of ChatGPT versus Claude, with ChatGPT being preferred for ideation tasks and Claude for technical support ($\chi^2(1, N = 30) = 5.76$, $p < 0.05$). This statistical distinction underpins the differentiated utility of each LLM, aligning with students' reflections on how ChatGPT's broad knowledge base spurred conceptual thinking, whereas Claude's precision offered invaluable technical guidance.

4.3.1. Qualitative Results

Students expressed an increased appreciation for the nuanced support provided by LLMs in navigating complex visualization tasks. The comparative analysis of LLM usage in Workshop 3 revealed significant differences in the application of ChatGPT versus Claude, with ChatGPT being preferred for ideation tasks and Claude for technical support. This bifurcation of LLM utility aligns with students' reflections on how each model contributed to different facets of the data visualization process. The implications of our findings extend beyond the specific context of data visualization education. The challenges and opportunities identified in this study, such as bridging the gap between engagement and skill development, managing the novelty effect, and fostering prompt engineering skills, are relevant to educators across various disciplines who are exploring the integration of AI tools in their teaching practices.

5. Discussion

5.1. Bridging Perception and Outcomes

A salient finding from our study was the divergence between LLMs increasing student engagement and motivation, but demonstrating more modest impacts on actual creativity and performance based on expert evaluations. This highlights a concerning gap between perception and measurable outcomes, where the novelty and interactivity of LLMs provide an illusion of competency gains without concrete visualization skill development. It underscores the need for intentional learning scaffolds that translate engagement into tangible artifacts and assignments requiring synthesis, evaluation, and visualization design. Directly linking LLMs to generative data visualization tasks with peer and expert critiques may better integrate engagement with demonstrated creativity improvements.

Additionally, the disconnect between student confidence perceptions and external creativity assessments indicates potential issues with over-reliance on subjective self-evaluation. Triangulating self-perceptions with peer feedback and expert scoring provides a more holistic and objective competency assessment. Preserving student self-voice while incorporating outside perspectives allows balancing subjective engagement with measurable skills gains. For visualization skills particularly, expert critique offers invaluable technical and creative feedback. Instructional design should focus on bridging the perception-outcomes gap by coupling LLMs with artifacts and activities that tangibly assess multifaceted learning.

Our study also revealed a significant gap between the heightened engagement and motivation LLMs induce in students and their actual improvement in creativity and visualization skills, as assessed by experts. This underscores the necessity for learning designs that effectively transform this engagement into substantial skill development. Incorporating behaviorism, rewards through expert feedback could reinforce desired learning behaviors, while constructivist strategies, like problem-solving tasks, can deepen understanding and skill application.

5.2. Managing the Novelty Effect

The initial excitement and hype surrounding novel technologies like LLMs poses engagement versus sustainability challenges in visualization education.

Maintaining student motivation and learning requires carefully designed activities that sustain, not just stimulate, interest by managing novelty and intentionally transitioning LLM support. Our study underscored the significance of prompt engineering skills in leveraging LLMs for educational purposes. As students' proficiency in crafting prompts improved, so did the usefulness and applicability of LLMs in their projects. This highlights the need for incorporating prompt engineering as a critical skill in data visualization curricula. Using varied participatory formats, collaborative group tasks, and prompts eliciting original student perspectives can continue harnessing LLM benefits while avoiding disengagement. Additionally, gradually fading LLM scaffolding as courses progress may mitigate risks of over-reliance. Educators should aim to sustain, not just stimulate, interest by managing novelty and intentionally transitioning LLM support. The three workshops in our study demonstrated the potential of LLMs in enhancing data visualization education, but also highlighted the importance of careful instructional design. Workshop 1 established a baseline of students' visualization skills without LLM support, while Workshop 2 introduced LLMs and prompt engineering techniques, leading to increased engagement and exploration. Workshop 3 allowed students to apply their newly acquired skills and knowledge, showcasing the impact of LLMs on their creative processes and final visualizations. Future research should build upon these findings to develop robust frameworks for integrating LLMs into data visualization curricula, with a focus on fostering long-term skill development and transfer. Furthermore, the tendency for engagement with LLMs to decrease after initial novelty wears off highlights the need for scaffolding strategies that maintain motivation. Periodic changes to LLM interactions such as alternating models, shifting roles, and variable usage formats can provide variation to sustain interest. Allowing student customization of LLM features and avatars may also increase engagement through personalized experiences. Promoting peer sharing of best prompting practices and creative LLM uses fosters collaborative momentum.

5.3. Developing Prompt Engineering Skills

A critical finding was students' evolving appreciation of prompt engineering's importance over time. Initial generic querying yielded low-value LLM responses, underscoring the need to foster prompt formulation skills early when introducing LLMs. Effective practices include scaffolding prompt complexity, providing feedback on quality, and modeling through examples. Framing prompt engineering as a creative exercise itself may intrinsically motivate students to iteratively refine prompts for optimal utility. Prompt engineering merits dedicated curriculum time to unlock LLMs' potential across diverse learning goals.

Furthermore, intentionally designing graded series of prompts could scaffold increasing sophistication. Initial prompts may focus on extracting basic visualization recommendations, followed by troubleshooting implementation issues, then optimizing designs for visualization principles, and finally tailoring visuals for specified audiences and insights. Providing worked examples of such prompt progressions raises awareness of the diversity of LLM applications. Peer feedback on prompt crafting through workshops can catalyze further refinement and engagement. Viewing prompt

iteration as a creative challenge itself may sustain student motivation in leveling up skills.

Highlighting the evolution of students' prompt engineering skills over time reflects the importance of active, constructivist learning practices. Effective prompt engineering is crucial for maximizing the educational potential of LLMs, emphasizing the need for curricula that foster these skills through iterative practice, feedback, and reflection. Integrating this into the learning process encourages a deeper engagement with the material and cultivates a nuanced understanding of how to interact with AI technologies for educational purposes.

Our study underscored the significance of prompt engineering skills in leveraging LLMs for educational purposes. As students' proficiency in crafting prompts improved, so did the usefulness and applicability of LLMs in their projects. This highlights the need for incorporating prompt engineering as a critical skill in data visualization curricula.

5.4. Comparing LLM Capabilities

Our study revealed distinct differences in the utility and engagement patterns between ChatGPT and Claude. While ChatGPT excelled in facilitating conceptual ideation, Claude emerged as a valuable resource for technical support. This bifurcation of LLM utility highlights the importance of aligning LLM selection with specific learning objectives in data visualization education (Figure 3).

Further analysis of the visual data reveals a potential for complementary use of ChatGPT and Claude in enhancing both the conceptual and technical proficiency of learners. As depicted, the engagement with ChatGPT for brainstorming and conceptual understanding, paired with Claude's technical scaffolding, suggests a hybrid model of LLM utilization could leverage the strengths of both models effectively. This approach not only maximizes the pedagogical value of LLMs but also mirrors real-world scenarios where creative ideation and technical implementation go hand in hand.

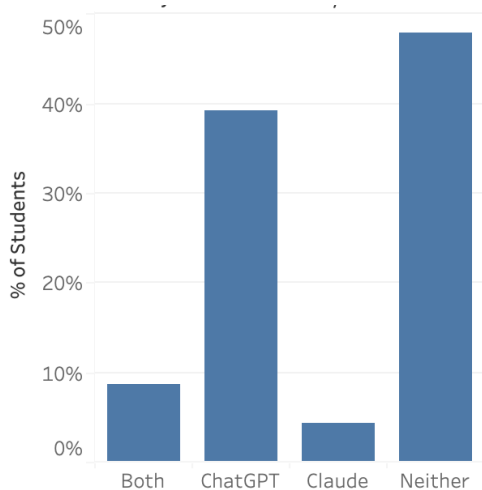


Figure 3: Workshop 3 Student LLM Usage Allocation.

6. Limitations

This preliminary study has several limitations. The small sample size of 30 graduate students restricts generalizability. Additional studies should incorporate larger, more diverse cohorts across educational levels and visualization proficiencies. Longitudinal designs tracking learning retention over full courses would provide richer insights versus the condensed 3-week format. There is inherent subjectivity in evaluating visualization quality.

Variance between LLMs like ChatGPT and Claude may impact comparative findings. Controlling prompt quality could reduce model output variability. Novelty effects of interacting with innovative AI systems may distort initial perceptions, warranting investigation into sustaining impacts over time. While revealing promising trends, this initial research requires expanded validation. Larger scale, longer-term studies with varied populations and visualization tools would further inform effective, ethical integration of conversational AI in data visualization pedagogy.

7. Future Research

Key directions for future work include longitudinal studies examining long-term retention and transfer of visualization skills learned over full courses. Comparing additional LLMs and incorporating visualization data into model training could further enhance capabilities. Discourse analysis of LLM conversations may uncover optimal prompting strategies for engagement. Evaluating time investments and tradeoffs versus traditional pedagogies could inform integration best practices.

Testing generalizability across visualization tools, STEM disciplines, and education levels would provide broader validation. Comparative studies blending conversational AI with active learning approaches like peer critiques and design studios may reveal effective combinations. Developing visualization-specific LLM architectures could improve responses on pedagogical tasks. This initial research establishes a foundation for extensive future work on prudently leveraging AI advancements to enrich data visualization education through evidence-based practices.

8. Conclusion

As AI capabilities progress, empirical studies can guide thoughtful integration, balancing innovation with learning principles. Our study revealed a significant gap between the heightened engagement and motivation LLMs induce in students and their actual improvement in creativity and visualization skills, as assessed by experts. This underscores the necessity for learning designs that effectively transform this engagement into substantial skill development. Incorporating behaviorism, rewards through expert feedback could reinforce desired learning behaviors, while constructivist strategies, like problem-solving tasks, can deepen understanding and skill application. Furthermore, our findings underscore the strategic importance of aligning LLM selection with specific learning objectives, as demonstrated by the distinct utilities of ChatGPT and Claude in facilitating conceptual ideation and providing technical support, respectively.

References

- [ARC*17] ALPER B., RICHE N. H., CHEVALIER F., BOY J., SEZGIN M.: Visualization literacy at elementary school. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2017), CHI '17, Association for Computing Machinery, p. 5485–5497. URL: <https://doi.org/10.1145/3025453.3025877>, doi:10.1145/3025453.3025877.
- [Bur20] BURCH M.: The importance of requirements engineering for teaching large visualization courses. *2020 Fourth International Workshop on Learning from Other Disciplines for Requirements Engineering (D4RE)* (2020), 6–10. URL: <https://api.semanticscholar.org/CorpusID:224778977>.
- [CBKH23] CAVOJSKÝ M., BUGÁR G., KORMANÍK T., HASIN M.: Exploring the capabilities and possible applications of large language models for education. *2023 21st International Conference on Emerging eLearning Technologies and Applications (ICETA)* (2023), 91–98. URL: <https://api.semanticscholar.org/CorpusID:266196844>.
- [CGDLO21] CUADRADO-GALLEGO J. J., DEMCHENKO Y., LOSADA M. A., ORMANDJIEVA O.: Classification and analysis of techniques and tools for data visualization teaching. *2021 IEEE Global Engineering Education Conference (EDUCON)* (2021), 1593–1599. URL: <https://api.semanticscholar.org/CorpusID:235476106>.
- [DFTW*21] DIEHL A., FIRAT E. E., TORSNEY-WEIR T., ABDULRAHMAN A., BACH B., LARAMEE R., PAJAROLA R., CHEN M.: VisGuided: A Community-driven Approach for Education in Visualization. In *Eurographics 2021 - Education Papers* (2021), Sousa Santos B., Domik G., (Eds.), The Eurographics Association. doi:10.2312/eged.20211003.
- [EL99] EPSTEIN R., LAPTOSKY G.: Behavioral approaches to creativity. *Encyclopedia of Creativity 1* (01 1999), 175–183.
- [EN93] ERTMER P. A., NEWBY T. J.: Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance improvement quarterly* 6, 4 (1993), 50–72.
- [ET23] ELLAWAY R. H., TOLSGAARD M.: Artificial scholarship: LLMs in health professions education research. *Advances in Health Sciences Education* 28 (2023), 659 – 664. URL: <https://api.semanticscholar.org/CorpusID:259201459>.
- [FCSY22] FENG J., CHEN H., SHI J., YE P.: Teaching practice of "spoc college" –a case study of "data visualization technology" course. *2022 12th International Conference on Information Technology in Medicine and Education (ITME)* (2022), 37–42. URL: <https://api.semanticscholar.org/CorpusID:257958963>.
- [GQWL23] GAN W., QI Z., WU J., LIN C.-W.: Large language models in education: Vision and opportunities. *2023 IEEE International Conference on Big Data (BigData)* (2023), 4776–4785. URL: <https://api.semanticscholar.org/CorpusID:265352038>.
- [GTS10] GRAMMEL L., TORY M., STOREY M.-A.: How information visualization novices construct visualizations. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 943–952. doi:10.1109/TVCG.2010.164.
- [JSFL] JOSHI A., SRINIVAS C., FIRAT E. E., LARAMEE R. S.: Evaluating the recommendations of LLMs to teach a visualization technique using bloom's taxonomy. URL: <https://api.semanticscholar.org/CorpusID:267302546>.
- [LMQ19] LO L. Y.-H., MING Y., QU H.: Learning vis tools: Teaching data visualization tutorials. *2019 IEEE Visualization Conference (VIS)* (2019), 11–15. URL: <https://api.semanticscholar.org/CorpusID:198147720>.
- [MB23] MAHADALKAR I. A., BYRD V. L.: Data visualization capacity (dvc) building tool for teaching and learning data visualization. *2023 IEEE Frontiers in Education Conference (FIE)* (2023), 1–5. URL: <https://api.semanticscholar.org/CorpusID:266803855>.
- [NGM23] NIEMANN M., GOODWIN S., MARRIOTT K.: Reflections on teaching 'data exploration and visualisation' in multiple modes. *2023 IEEE VIS Workshop on Visualization Education, Literacy, and Activities (EduVis)* (2023), 28–33. URL: <https://api.semanticscholar.org/CorpusID:266197958>.
- [RBB*22] ROBERTS J. C., BACH B., BOUCHER M., CHEVALIER F., DIEHL A., HINRICH S., HURON S., KIRK A., KNUDSEN S., MEIRELLES I., NOONAN R., PELCHMANN L., RAJABIYAZDI F., STOIBER C.: Reflections and considerations on running creative visualization learning activities, 2022. arXiv:2209.09807.
- [RM15] RUCHIKACHORN P., MUELLER K.: Learning visualizations by analogy: Promoting visual literacy through visualization morphing. *IEEE Transactions on Visualization and Computer Graphics* 21 (2015), 1028–1044. URL: <https://api.semanticscholar.org/CorpusID:340394>.
- [VT23] VIRVOU M. K., TSHIRINTZIS G. A.: Is chatgpt beneficial to education? a holistic evaluation framework based on intelligent tutoring systems. *2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA)* (2023), 1–8. URL: <https://api.semanticscholar.org/CorpusID:266236632>.
- [Wan22] WANG C.: Visvisual: A toolkit for teaching and learning data visualization. *IEEE Computer Graphics and Applications* 42 (2022), 20–26. URL: <https://api.semanticscholar.org/CorpusID:250140946>.
- [WLC23] WHITEHILL J., LOCASALE-CROUCH J.: Automated evaluation of classroom instructional support with LLMs and BoWs: Connecting global predictions to specific feedback. *ArXiv abs/2310.01132* (2023). URL: <https://api.semanticscholar.org/CorpusID:263605547>.
- [XXZ*23] XIAO C., XU S. X., ZHANG K., WANG Y., XIA L.: Evaluating reading comprehension exercises generated by LLMs: A showcase of chatgpt in education applications. In *Workshop on Innovative Use of NLP for Building Educational Applications* (2023). URL: <https://api.semanticscholar.org/CorpusID:259376566>.