StuGPTViz: A Visual Analytics Approach to Understand Student-ChatGPT Interactions

Category: Research

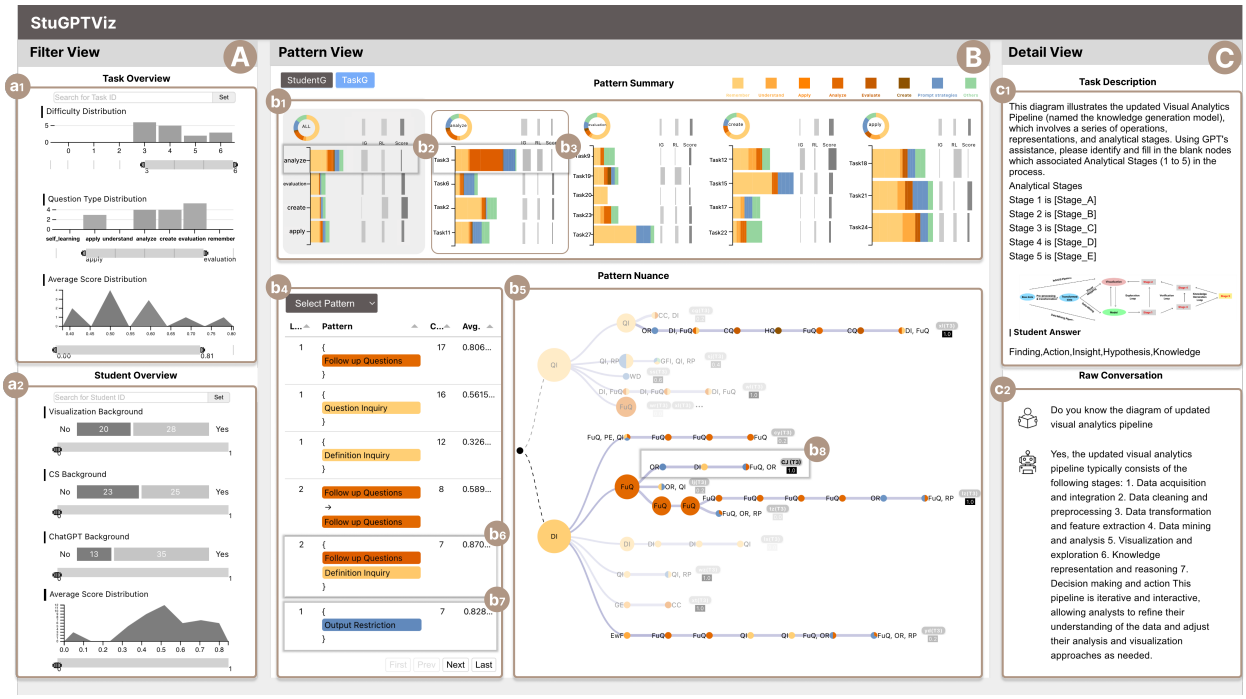

Fig. 1: System interface of StuGPTViz. The **Filter View** (A) offers an overview and enables filtering of tasks and students through the Task Overview (a1) and Student Overview (a2). The **Pattern View** (B) displays the macro-level summary of conversation characteristics in the Pattern Summary (b1), and micro-level interaction patterns along with their evolution in the Pattern Nuance, represented by the Pattern Mining Tabl (b4) and the Interaction Tree (b5). The **Detail View** (C) presents task descriptions (c1), and the raw conversation data between students and ChatGPT (c2).

Abstract—The integration of Large Language Models (LLMs), especially ChatGPT, into education is poised to revolutionize students' learning experiences by introducing innovative conversational learning methodologies. To empower students to fully leverage the capabilities of ChatGPT in educational scenarios, understanding students' interaction patterns with ChatGPT is crucial for instructors. However, this endeavor is challenging due to the absence of datasets focused on student-ChatGPT conversations and the complexities in identifying and analyzing the evolutional interaction patterns within conversations. To address these challenges, we collected conversational data from 48 students interacting with ChatGPT in a master's level data visualization course over one semester. We then developed a coding scheme, grounded in the literature on cognitive levels and thematic analysis, to categorize students' interaction patterns with ChatGPT. Furthermore, we present a visual analytics system, StuGPTViz, that tracks and compares temporal patterns in student prompts and the quality of ChatGPT's responses at multiple scales, revealing significant pedagogical insights for instructors. We validated the system's effectiveness through expert interviews with six data visualization instructors and three case studies. The results confirmed StuGPTViz's capacity to enhance educators' insights into the pedagogical value of ChatGPT. We also discussed the potential research opportunities of applying visual analytics in education and developing AI-driven personalized learning solutions.

Index Terms—Visual Analytics for Education, Student-ChatGPT Interaction

1 Introduction

Groundbreaking developments in generative AI, particularly through Large Language Models (LLMs) applications such as ChatGPT, have introduced unprecedented opportunities in educational methodologies [12, 35]. These tools not only expedite students' information searches but also assist instructors in refining classroom activities and delivering personalized guidance [13, 47, 57]. However, as the integration of LLMs into educational scenarios is still nascent, it is imperative for instructors to carefully plan and assess how students utilize LLMs in their learning activities [53] to harness the full potential of LLMs and enhance the student learning experience [34, 67]. A fundamental step in this process is to gain a comprehensive understanding of student interactions with LLMs, thereby acquiring key pedagogical insights such as students' cognitive levels, learning attitudes, and mastery of knowledge [41, 45].

Nevertheless, efforts to provide instructors with these pedagogical insights are still in the initial stages. Existing research on LLMs in education primarily focuses on educational ethics and potential application scenarios, such as using LLMs as automated tools for evaluating student essays [10, 38, 39, 55, 59]. To our knowledge, in-depth studies on understanding the students' interaction with LLMs for learning tasks are scarce and face two significant challenges. First, there are no pub-

licly available datasets dedicated to capturing students' conversations with LLMs. The existing datasets are primarily composed of everyday conversations with general users. While some include conversations about learning tasks like programming or solving math problems, they are few and lack guaranteed quality because the learning scenarios involving LLM use are not carefully crafted and evaluated by instructors. Moreover, these conversations typically follow a "one question, one answer" format, as their primary purpose is to assess LLMs' problem-solving capabilities. These deficiencies highlight the urgent need for data collection of learning-centered conversations with well-structured tasks [16, 17, 78].

Second, understanding how students interact with LLMs for pedagogical insights through conversation data presents significant challenges. One major difficulty is that instructors are eager to understand the extent of higher-order thinking (e.g., independent thinking) students engage in when using these advanced AI tools [21, 71]. To comprehend this higher-order thinking, it is essential to measure students' cognitive levels [11]. However, accurately interpreting these cognitive levels based on students' inquiries to LLMs has never been explored. Additionally, assessing students' proficiency in utilizing LLMs poses another challenge, which involves evaluating the various LLMs' responses and observing how students adjust their prompts in response. Moreover, tracking the progression of these interactions introduces an added layer of complexity [27, 68]. Visual analysis is a potential way. However, current research has often overlooked these challenges. While there are many visualizations works on conversation analysis studies focusing on topic progression or sentiment analysis [22, 36, 46], such studies do not adequately capture and visualize the cognitive levels reflected in the evolving interactions, falling short of meeting instructors' needs.

To address these challenges, we selected ChatGPT, notable for being one of the most prevalent LLM applications [77], to gather conversation data. In addition, we see a potential to introduce LLMs for visualization education to address diverse student backgrounds and manage varied learning activities such as concept comprehension, visualization literacy, and design [52]. In collaboration with experienced course instructors, We devised and integrated an in-class exercise module into a graduate-level data visualization course at the local university, allowing students to interact with ChatGPT freely. Our approach yielded a significant collection of high-quality student-ChatGPT conversation data from well-crafted learning tasks. Through comprehensive thematic analysis and an extensive literature review [14, 65, 73, 74], we developed a coding scheme that categorizes the diverse cognitive levels [44] and several metrics to evaluate the quality of ChatGPT responses [26, 48]. To further capture the evolving strategies students recurrently employ when interacting with ChatGPT, we analyzed various sequences and sets of the codes, defining these as "interaction patterns" which became the focal point of our analysis. Building on this foundation and design requirements derived from course instructors and experts, we introduce a pioneering visual analytics system for instructors to explore intricate interaction patterns and derive actionable pedagogical insights from student-ChatGPT conversation data. In particular, a customized tree visualization is designed to present the evolution and compare the characteristics of students' interaction patterns. To summarize, our key contributions are as follows:

- We introduced ChatGPT to a real data visualization course, collected student-ChatGPT conversation data and developed a coding scheme for an in-depth analysis of interaction patterns.
- We designed a visual analytics system, StuGPTViz, to help instructors discover insights into students' cognitive levels and proficiency when using ChatGPT.
- Through three case studies and expert interviews, we demonstrated the effectiveness of our coding scheme and system in enhancing educational activities such as problem-solving guidance, personalized feedback, and exercise design.

Overall, we present a design study that constitutes an initial yet crucial step toward analyzing student interaction patterns with ChatGPT, advancing the application of visual analytics in AI-driven education.

2 Related Work

In this section, we discuss the relevant research, including LLMs in education and visualization education, visualization for AI-enhanced education, and visual analytics for conversational data.

2.1 LLMs in Education and Visualization Education

The integration of LLMs such as ChatGPT into educational settings has sparked a diverse range of discussions, with plenty of initial works centered on ethical considerations regarding their use in learning environments [38, 55]. Increasingly, the academic community recognizes the transformative potential LLMs hold for education, advocating for their adoption to revolutionize learning and teaching methodologies [12]. Despite potential resistance from some educators, students inevitably turn to LLMs for assistance with coursework [35]. Therefore, researchers and instructors have explored LLMs for various applications, including serving as teaching assistants for writing and coding, generating adaptive exercises [66], supporting personalized question-answering sessions [10], and facilitating innovative learning modes like "learn by teaching", where LLM plays the role as "learner" and students assume the role of the teacher to teach the AI [62].

However, there is a notable gap in understanding how students strategize their use of ChatGPT for educational purposes. Existing literature predominantly focuses on the capabilities and applications of LLMs without delving into student interaction strategies, leaving educators without the necessary insights to fully leverage these tools in enhancing learning experiences [1, 29].

Simultaneously, the field of visualization education is gaining traction, not only within the visualization community but also more broadly [8]. The challenges of teaching data visualization range from addressing diverse student backgrounds to managing varied learning activities such as concept comprehension, visualization literacy, and design evaluation—are substantial [52]. ChatGPT's potential to support these educational challenges opens avenues to investigate how students use ChatGPT across different visualization learning tasks, particularly relevant to our project's focus [3]. A comprehensive analysis of this research direction still remains unexplored [20]. This gap presents a unique opportunity for our work to contribute to the field by offering insights into student strategies in employing ChatGPT within visualization learning contexts, thereby advancing the understanding and application of LLMs in educational settings.

2.2 Visualization for AI-Enhanced Education

The integration of visualization in AI-enhanced education is dedicated to leveraging visual analytics for learning analysis, such as interpreting complex data and AI algorithms to improve educational outcomes [23]. While learning analysis harnesses data to refine and enhance learning processes, visualization techniques render these insights accessible and actionable for educators and students [18, 58]. Despite notable advancements in each domain, the integration of visual analytics specifically tailored to learning analysis within AI-enhanced educational environments remains underexplored.

Existing research underscores the value of visual analytics in presenting student performance metrics, engagement levels, and learning behaviors, thus enriching our understanding of educational dynamics [4, 5]. Within this context, the subfield of open learner models exemplifies the potential of visual explanations, akin to Explainable AI (XAI), in demystifying AI-generated outputs, offering learners and educators transparent and trustworthy insights [15, 25]. Additionally, other works have employed visual analytics to examine students' interaction data with intelligent agents, using students' log data such as hint requests to delve into their problem-solving processes [75]. Recently, with the advent of potent LLM-based conversational agents, the intricacy of interactions between students and AI has reached a new height. Goals once considered unrealistic, such as in-depth analysis of students' cognitive levels and thought processes, are now achievable [9, 40, 72]. To our knowledge, the dedicated exploration of visual analytics to analyze and elucidate student interactions with advanced AI tools, such as ChatGPT, is just beginning. In response to this gap, our work proposes a novel visual analytics system based on the students-ChatGPT

conversation data we collected. The system is designed to identify and unravel the intricate nuances of student-ChatGPT interactions. It equips educators with profound insights into how LLMs can be utilized to customize and elevate students' learning experiences.

2.3 Visual Analytics for Conversational Data

The exploration of visual analytics for conversational text data within the visualization community has encompassed a wide range of applications, from sentiment analysis and topic modeling to mapping conversation flows and interactions within user groups [30, 32, 36, 46]. For instance, T-Cal and IneqDetect [31, 54] are centered on analyzing collaboration group conversations, extracting keywords, and estimating sentiments among group members to reflect collaboration effectiveness. Meanwhile, efforts such as ThreadReconstructor, MultiConVis, and VisOHC [6, 24, 37] probe into the structure and core topics of online forum discussions. However, these initiatives mainly focus on summarizing the dynamics of multi-party conversations without delving into the intricacies of one-on-one dialogues.

Another significant gap in current methodologies is their constrained ability to uncover the depth of evolving cognitive levels in educational dialogues between students and AI tools like ChatGPT. Visualization efforts for one-on-one medical conversations, like ConVIScope and Discursis [7, 49], focus on charting the development of patient-doctor dialogues. However, their visualizations reflect the change of sentiment and topic over time, falling short in showcasing and comparing how one party (e.g., students) adjust their answers (e.g., prompts) in response to another party (e.g., various LLMs' replies). Similarly, research aimed at analyzing educational dialogues often emphasizes engagement and comprehension, lacking a detailed visual analysis of the depth of thinking, learning strategies, or intentions revealed through these interactions [8, 50, 76].

These shortcomings highlight the necessity for a novel visual analytics framework designed to tackle the specific challenges posed by educational dialogues with LLMs [3, 33]. Consequently, our work introduces an interaction pattern tree visualization focused on meticulously displaying and comparing students' interaction patterns derived from their conversations with ChatGPT.

3 Background and Data Collection

This section outlines the background of our data visualization course, the collaborative efforts with our expert team, the in-class exercises, and the data collection considerations and procedures we adopted.

3.1 Data Collection Background and Considerations

At the invitation of our experts (two course instructors, E1&E2), we set out to incorporate ChatGPT into the curriculum of a postgraduate data visualization course for computer science majors in the first semester of 2024. This course, meeting once a week for three hours, attracted 55 registrants. Over the past four months, we have collaborated closely with E1, E2, and three teaching assistants (TA1-TA3) at our university. E1 is a professor with over 15 years of experience designing and teaching data visualization courses. E2 is a lecturer with three years of teaching experience and served as the primary instructor for this course. TA1, TA2, and TA3, are senior teaching assistants who have supported the professors in designing the in-class exercises, homework, and exams for the data visualization course for at least two semesters. All experts have used ChatGPT extensively over the last two years. In preparation, we conducted a three-hour meeting to outline how ChatGPT would be integrated into the course and how data would be collected. To maximize the benefits of ChatGPT for students, we decided to introduce an in-class exercise section. This would allow students to apply what they learned by interacting with ChatGPT and create an ideal setting for collecting diverse data on their learning interactions. During our discussion, two key considerations emerged:

C1: Diverse Learning Tasks for Comprehensive Data Collection. To capture the full spectrum of student interactions with ChatGPT, exercises should be designed across various course aspects using diverse learning tasks (e.g., visualization understanding & design). These learning tasks should encompass different cognitive levels as outlined in Bloom's Taxonomy [44] to ensure that the collected data fully reflects students' diverse engagement with ChatGPT. While these tasks are exemplified using a data visualization course, learning tasks for any course can be similarly designed to align with Bloom's Taxonomy [19].

C2: Natural Learning Scenarios for Unbiased Data Collection. To accurately reflect genuine students' behaviors, it was vital to integrate ChatGPT into the learning process as an optional tool rather than a course mandate. This approach was intended to gather authentic interaction data, showcasing real student needs and preferences in using ChatGPT for learning.

Informed by these considerations, we included a 40-minute open ChatGPT session at the end of each class (9 classes in total) to promote active student engagement with ChatGPT through various tasks. Additionally, we carried out a pilot study with about 30 HCI and data visualization Ph.D. students to evaluate the exercise section's design and procedures. Their conversation data helped us ascertain the appropriateness of our data collection approach. The data collection process was also approved by the university's ethics committee (IRB approval). The subsequent sections will provide an overview of our task designs for the in-class exercises and the data collection procedures.

3.2 Tasks Summary and In-class Exercise Procedure

Collaborating with instructors E1 and E2, we developed 27 exercise tasks spread across seven distinct types according to the levels of cognitive learning in the Revised Bloom's Taxonomy (**C1**, Fig. 2):

Task Type & Count	Task Brief	Cognitive Level
Concept Remember (2)	Multiple Choices questions for basic concept remembering	Remember (L1)
Concept Understanding (3)	Multiple Choices questions for deeper concept understanding	Understand (L2)
Concept Application (3)	Short questions for concept application	Apply (L3)
Visualization Analysis (4)	Open-ended analysis questions (e.g., encoding usage, color scheme)	Analyze (L4)
Visualization Evaluation (5)	Evaluate the given visualization design	Evaluate (L5)
Visualization Design (4)	Design visualization with the given data	Create (L6)
Self Learning (6)	Self exploration of key concepts	Others

Fig. 2: The summary of task type, count, cognitive level, and a brief description. Sample tasks are provided in the supplementary (A).

To begin the data collection, we distributed a questionnaire to the 55 enrolled students to gauge their willingness to participate and to gather their background information. 48 students consented to participate, providing details on their undergraduate majors, data visualization expertise, programming skills, and prior experience with ChatGPT. In addition, we hosted a 40-minute introductory session on ChatGPT for beginners, which included instructions on how to export and upload the conversation files to the Canvas system[1], the web-based learning management system used by our university.

Each in-class exercise session, conducted during the last 40 minutes of the lecture, consisted of a self-learning segment with ChatGPT (10 mins), a task completion segment with ChatGPT (25 mins), and a conversation log upload phase (5 mins). Students initially were required to spend 10 minutes asking ChatGPT questions to learn the lecture's key terms. This was followed by a 25-minute segment where students completed given tasks (Fig. 2) with ChatGPT's assistance and concluded with a five-minute interval for uploading their conversations to Canvas. We encouraged students to interact with ChatGPT in a way that felt natural to them. We suggested that those who were already confident with the course material could skip the self-exploration phase (**C2**) if they found it to be redundant.

[1] https://www.instructure.com/canvas

3.3 Dataset Brief

The dataset we collected involved 48 students' conversation data with ChatGPT during the in-class exercise session of data visualization course over the entire spring 2024 semester. It consists of 744 unique conversations with 2507 turns after filtering out the empty conversations and those unrelated to the learning tasks. To the best of our knowledge, it is the only existing dataset specifically for student-ChatGPT conversations under strictly-defined learning activities. The collected data is in a structured format, including metadata such as student names (anonymized), task ID and task types, and conversation content. Each conversation is logged in sequential order, capturing both student prompts and ChatGPT responses. Sample data are provided in the supplementary materials. Additionally, we collected students demographic information, including their background in computer science, data visualization, and ChatGPT usage experience.

4 Design Requirement and Data Processing

This section presents the design requirements for the visual analytics system identified through discussions with our experts (E1 and E2), the procedure for coding student prompts, and the methodology for processing ChatGPT responses.

4.1 Visualization Design Requirements

We summarized experts' requirements for analyzing student-ChatGPT conversation data as follows:

R1: Overview of students and tasks data. Experts highlighted the necessity of an overview of both students and the tasks, including the distribution of students' background information (e.g., knowledge in visualization) and tasks' characteristics (e.g., types). Instructors can then select specific students or tasks for deeper analysis of student-ChatGPT conversations.

R2: Summarizing macro-level conversation characteristics. Before detailed analysis, experts require a comprehensive summary of the student-ChatGPT conversation, especially the cognitive levels of students' prompts and the qualities of ChatGPT responses based on selected tasks and students. This summary should span multiple user-selected viewpoints, covering broad categories such as various student groups and types of tasks, since instructors are always interested in the differences in behavior and performance among different student groups, such as those with and without a CS background. Additionally, instructors often have limited time and want to quickly see the differences among groups to prioritize their focus.

R3: Identifying micro-level interaction patterns. Experts require a structured method to detect and summarize students' micro-level interaction patterns (i.e., recurring methods and strategies students employ when interacting with ChatGPT), along with essential metrics such as learning outcomes and pattern frequency. Emphasizing this information is crucial for instructors to identify which patterns are more effective for learning and merit deep exploration. They can further develop actionable insights on how students engage with ChatGPT throughout their learning journey.

R4: Tracing interaction pattern evolution. Educators necessitate a method to trace the development of students' interaction patterns, emphasizing both the shared and unique sequences within the context of task-solving. This requirement involves visualizing how students' interaction patterns evolve in response to tasks, reflecting variations in cognitive engagement and problem-solving approaches. Such visualization should facilitate a deeper understanding of varied student approaches to learning tasks.

R5: Evaluating interaction pattern performance. Experts wanted to evaluate interaction patterns of interest effectively. This involves identifying students who utilize the pattern and assessing relevant metrics such as their learning outcomes and ChatGPT's response quality. Such comprehensive analysis helps gauge the effectiveness of different interaction patterns, enabling instructors to provide targeted feedback and identify exemplary patterns as recommended learning strategies.

R6: Examining detailed raw data. Experts expressed a desire to access the raw data, which includes original in-class activities, students' answers or responses, and students' conversation logs with ChatGPT. These details can be used to justify their analysis results and provide straightforward examples for students to master effective interaction with ChatGPT.

4.2 Students' Prompts Coding

The open coding process for students' prompts is based on thematic analysis methodology [69] and enriched by a literature review to identify prompt patterns [65, 73, 74]. Following expert requirements (**R2**), we categorized codes into two types: learning-related codes reflecting students' cognitive levels regarding course materials, and ChatGPT-related codes denoting students' comprehension and proficiency in using ChatGPT as a supplementary tool.

Firstly, we reviewed literature analyzing general user prompt patterns and intents [65, 73, 74], identifying 16 general prompt patterns as the initial ChatGPT-related codes. While these codes could represent students' proficiency with ChatGPT, they did not capture their cognitive levels. Therefore, we invited three visualization researchers, each with over four years of experience in data visualization education and thematic analysis, to independently conduct open coding of students' prompts. They refined the ChatGPT-related codes and developed learning-related codes that encapsulate the intent and thought processes behind each student's prompt. Our researchers engaged in repeated cross-checking and refinement until a consensus was reached, ultimately establishing 27 codes, including 15 learning-related and 12 ChatGPT-related codes.

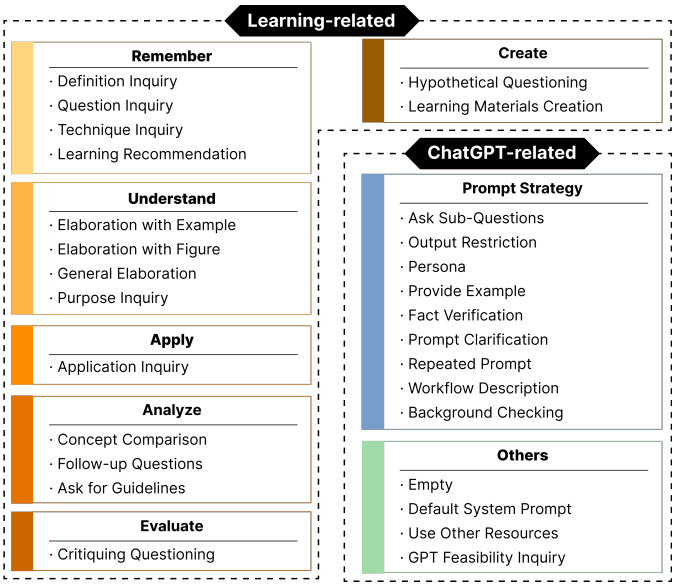

Fig. 3: The code schema with revised bloom taxonomy [44] classification.

After finalizing the code space, we consulted with our experts (TA1-TA3) to further refine our codes. At TA3's suggestion, we used the revised Bloom's taxonomy to categorize the 15 learning-related codes, enhancing our understanding of students' learning intentions and mapping their cognitive processes [44]. This taxonomy describes six stages of cognitive learning: remember, understand, apply, analyze, evaluate, and create, each representing an advancement in cognitive level. The authors independently categorized the learning-related codes, then discussed and refined any inconsistencies until agreement was reached. The final code schema is summarized in Fig. 3. To ensure coding quality and consistency, the authors coded 30 conversations from the pilot experiment with the finalized labels and calculated the Inter-Rater Reliability (IRR) scores, resulting in an IRR of 0.84, validating the coding process's reliability. Finally, the authors coded all students' prompts. For prompts embodying multiple learning intents and strategies, we applied multiple codes. Additionally, we used the ordered sequence of codes and unordered sets of codes from each student's conversation data to illustrate the "interaction patterns" identified via expert requirements (**R3**). For each coded conversation, we mined all possible ordered code

Fig. 4: The **Pattern Summary** section of the *Pattern View* summarizes both between-group and within-group interaction patterns based on the students and tasks selected by the user. Users can click on each grey bar to sort the students according to the selected metric.

sequences (e.g., [Definition Inquiry, Follow up Question, ...]) and unordered code sets (e.g., {Definition Inquiry, Application Inquiry, ...}) for further analysis.

4.3 Processing ChatGPT's Responses

To enhance our understanding of students' interactions with ChatGPT, we analyze and evaluate ChatGPT's response quality based on expert suggestions (TA1). Through our literature review, common metrics for evaluating ChatGPT's responses include response relevance, length, and correctness [28, 42, 43], all recognized as important by our experts. We employed the Ragas response relevance package [26], a renowned framework for evaluating large language models, to assign a numerical score to the relevance of each "user prompt - LLM response" pair. Additionally, we measured response length and, in collaboration with experts E2 and TA2, assessed response correctness by categorizing them into "basically correct" (score 1), "partially correct" (score 0.5), and "basically wrong" (score 0), respectively. Through further discussion, experts (E1 & E2) expressed interest in evaluating the amount of accurate information a student can acquire from each turn of interaction with ChatGPT. Consequently, we combine the two metrics, response relevance score and response correctness, to develop the new metric "information gain" inspired by the literature [51]. The metric's formula is shown below, which primarily leverages the KL-divergence principle [70], calculates the amount of new information provided by the latest ChatGPT response compared to the existing set of responses:

$$\mathrm{IG}(P,Q) = \sum_i P(i) \log \left(\frac{P(i)}{Q(i)} \right) \times R \times C,$$

In this formula, IG represents the information gain of the incoming response P under the cumulative response set Q. $P(i)$ is calculated as the frequency of word i in the current response divided by the total number of words in that response. $Q(i)$, on the other hand, is the cumulative frequency of word i up to the current response, divided by the total number of words in all responses up to that point. The variables R and C represent the numeric relevance score and correctness score, respectively. This computation quantifies the new information provided by ChatGPT's latest response compared to the existing knowledge base.

5 Visualization

Based on the design requirements identified in Sec. 4.1 and the data collected in Sec. 3, we developed a visual analytics system, StuGPTViz (Fig. 1), aimed at enabling instructors to analyze student interactions with ChatGPT effectively.

5.1 System Overview

StuGPTViz is intricately designed to facilitate a multi-level analysis of students' interactions with ChatGPT from the perspective of both tasks and students. It supports instructors in selecting specific tasks and students as focal points of analysis, thereby accommodating diverse analytical interests. Moreover, it enables a "gradually deepening" analysis process, allowing users to acquire both a broad overview and detailed insights into the interactions between students and ChatGPT. Specifically, StuGPTViz is structured into three main components:

Initial selection: Instructors begin by selecting particular tasks or students of interest through the *Task Overview* and *Student Overview* in the **Filter View** (Fig. 1-a1, a2). This selection offers an overview of students and tasks data **(R1)**, which also triggers updates in the *Pattern Summary* of the *Pattern View* (Fig. 1-b1).

Gradually deepening exploration: Starting from the *Pattern Summary* in the **Pattern View** (Fig. 1-b1), instructors can gain a summary of macro-level conversation characteristics of the selected students and tasks, covering both groups and individuals **(R2)**. To delve deeper, instructors can further analyze the micro-level student-ChatGPT interaction patterns via *Pattern Nuance*. While the *Pattern Mining Table* (Fig. 1-b4) enable instructors to identify the pattern summary together with significant metrics like learning outcome **(R3)**, the *Interaction Tree* traced the pattern evolution of each student **(R4)**. Moreover, the interplay between these two visualizations enables instructors to explore and assess patterns of interest effectively **(R5)**.

Detailed inspection: Finally, the *Task Description* and *Raw Conversation* from the **Detailed View** (Fig. 1-c1, c2) provides access to the original tasks, students' responses, and their raw conversation logs with ChatGPT. This component allows instructors to examine task specifications, student prompts, and ChatGPT's replies, leveraging their expertise to interpret the data comprehensively **(R6)**.

5.2 Filter View

The *Filter View* (Fig. 1-A) serves as the gateway to analysis, presenting the distribution of various background metrics and enabling instructors to filter the students and tasks of interest (**R1**). The *Task Overview* (Fig. 1-a1) displays information about learning tasks and features a search box for quickly finding specific tasks by ID. The distribution of task difficulties and types, determined by experts (E1 & E2), is depicted through two bar charts, while an area chart illustrates the distribution of students' normalized average scores (x-axis) across tasks. Instructors can select tasks by clicking on the bars or adjusting sliders, with each metric chart dynamically updating in response to real-time user selections. Moreover, the *Student Overview* (Fig. 1-a2) provides insights into students' backgrounds. A search box allows quick location of students by aliases, and an area chart at the bottom visualizes the distribution of students' average scores. Positioned between them, three-segmented bar charts represent the distribution of students' prior experience in data visualization and computer science and their familiarity with ChatGPT, derived from a background survey conducted during the first class. By freely filtering each metric, instructors can effortlessly isolate students and tasks of interest (**R1**), then proceed to the *Pattern View* (Fig. 1-B).

5.3 Pattern View

After identifying tasks and students of interest, instructors can delve into detailed analysis using the *Pattern View* (Fig. 1-B). Comprising the *Pattern Summary* and *Pattern Nuance*, the *Pattern View* enables a gradually

deepening examination of interaction patterns, offering insights from macro-level summaries to micro-level details **(R2, R3)**. Throughout the *Pattern View*, we consistently apply color scheme (Fig. 1, top-right corner) to represent the code categories defined in Fig. 3. Specifically, we use a sequential color scheme, transitioning from light yellow to dark brown, to denote increasing cognitive levels (from "remember" to "create") as demonstrated in students' prompts. Concurrently, a distinct color scheme (blue for effective prompt strategies from literature and green for others) signifies students' proficiency in using ChatGPT.

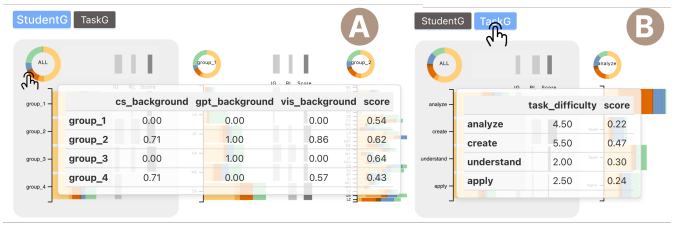

Fig. 5: (A) The between group-level background comparison. By default, students are grouped by their background. (B) Under the "Task-Grouping" mode, tasks are grouped by types.

5.3.1 Pattern Summary

The *Pattern Summary* is introduced with a control button at the top-left corner (Fig. 4) for selecting the grouping mode. A comprehensive macro-level pattern summary **(R2)** under the chosen mode is displayed below. The grey card component presents a between-group level summary (Fig. 4-A), adjacent to which are the within-group level summary cards (Fig. 4-B). To begin, instructors can view the summary donuts chart (Fig. 4-a1), which shows the distribution of all students' prompt categories as defined in Sec. 4.2. For example, here, a predominantly light yellow section suggests that a majority of student prompts are at the "Remember" cognitive stage, while minor blue and major green segments indicate a limited use of literature-supported effective prompt strategies. This donuts chart design aims to clearly display the percentage of each cognitive level in students' prompts while optimizing space usage. Additionally, to effectively display and compare the overall quality of ChatGPT responses and students' learning outcomes, we selected two light grey bars and a dark grey bar (Fig. 4-a1) to represent the three metrics due to their simplicity and effectiveness [56]. The details of metrics are introduced in Sec. 4.3.

Instructors can easily check each group's background information by hovering over the summary donut chart (Fig. 5-A) and perform between-group comparisons of cognitive stage distribution and ChatGPT response quality using stacked bar charts and accompanying grey bar charts (Fig. 4-a2). Moreover, by switching to "TaskG" mode (Fig. 5-B), instructors can change the grouping from student background to task types, shifting the analysis towards task-specific student performance and interaction summaries. This flexible approach enables an efficient multi-viewpoint summary of macro-level students-ChatGPT conversations (**R2**). After identifying a specific group of interest, instructors can click on a stacked bar in the summary card (Fig. 4-a2), which highlights the selected group card (Fig. 4-b1) for deeper, within-group analysis. Instructors can sort students by various metrics by clicking on the metric summary bars in the first row, as shown in the example where students are sorted by their scores in ascending order (Fig. 4-b1). To move on, instructors can click on any stacked bar or the donuts chart (Fig. 4-b1) to investigate micro-level details in *Pattern Mining Table* (Fig. 6-A) and the *Interaction Tree* (Fig. 7-A).

5.3.2 Pattern Mining Table

The *Pattern Mining Table* (Fig. 1-b4, Fig. 6-A) catalogs micro-level nuanced interaction patterns mined from the tasks and students selected by users (**R3**). Each interaction pattern is associated with specific metrics: length ("L"), frequency ("C"), and average score ("Avg."). Aligned with the previous definition Sec. 4.2, interaction patterns are delineated as either a set of codes (denoted by curly braces {}, Fig. 6-B)

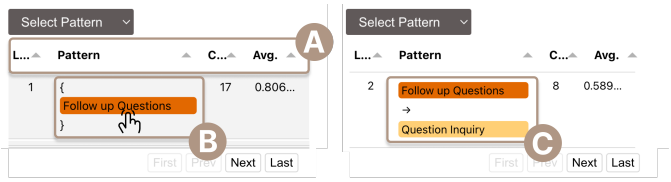

Fig. 6: The **Pattern Mining Table** within *Pattern Nuance*. (A) The table headers include pattern length ("L"), interaction pattern ("Pattern"), pattern frequency ("C"), and average score ("Avg."). (B) Example of the "unordered code set" type pattern. Users can click the pattern row to highlight the students utilizing this pattern in (Fig. 7-A). (C) Example of the "ordered code list" type pattern.

or an ordered list of codes (indicated by an arrow → in the "Pattern" column, Fig. 6-C), with each code's background color reflecting its category. Instructors can sort these patterns by any metric, aiding in the identification of prevalent or effective patterns.

5.3.3 Interaction Tree

The *Interaction Tree* employs a decision-tree format design to trace the evolution of interaction patterns (**R4**) for individual students under each task (Fig. 1, Fig. 7-A). Each path within the *Interaction Tree* represents a student's detailed interaction process with ChatGPT under a specific task, beginning from a common root where nodes symbolize students' prompts, and adjacent solid links depict ChatGPT's responses. This format enables the aggregation of similar interaction paths, highlighting variations and commonalities in student interaction patterns. Consistent with the above settings, colors in the node coding for the category of the prompt and abbreviations inside or beside the node detailing the code content. When prompts encompass multiple codes, nodes incorporate pie-chart coloring to represent each code visually (Fig. 7-a1). To aggregate prompts with the same code contents, we use node size to indicate the number of students at the same round in their conversation. On the other hand, the solid links between nodes convey ChatGPT's response characteristics, including "Response Token Length" (RL) and "Information Gain" (IG). Each ChatGPT response's "Information Gain" is represented through variations in the horizontal length of the link, as it is the major information instructors (E1 & E2) want to notice, while link width and opacity double encode "Response Token Length". This visual encoding offers insights into the value and quality of ChatGPT's replies. Furthermore, the end of each path features one grey tag representing the student's alias and task's ID (Fig. 7-a2), together with another tag below encoding the student's performance on this task. Here, we utilized numerical values and color intensity to denote scores, thereby linking interaction patterns directly to learning outcomes. For instance, the highlighted student gained the full mark (Fig. 7-a2) after various interactions with ChatGPT. Through these meticulously designed components, instructors are not only equipped to perform a granular analysis of student interactions but also able to correlate them with educational outcomes, facilitating a comprehensive understanding of students' learning behaviors and the effectiveness of their interactions with ChatGPT (**R5**).

To further facilitate in-depth evaluation of each identified interaction pattern, we introduced an interplay between the *Pattern Mining Table* and the *Interaction Tree*. By clicking on each row, which corresponds to a specific interaction pattern (Fig. 6-B), the corresponding students who engage in this pattern will be highlighted in the *Interaction Tree* (Fig. 7-A). This enables instructors to assess relevant metrics such as learning outcomes and the quality of ChatGPT's responses for each identified pattern or individual student (**R5**).

Design alternative. During the design process, some alternatives were raised and discussed with our experts. For instance, we proposed to use a Sankey diagram [60] together with a stacked bar chart design to represent the students' different prompt choices at each step. Specifically, each stacked bar represent the distribution of different categories of codes in each conversation turn (Fig. 7-b1). The Sankey flow represents the selected student's interaction pattern, ending with a

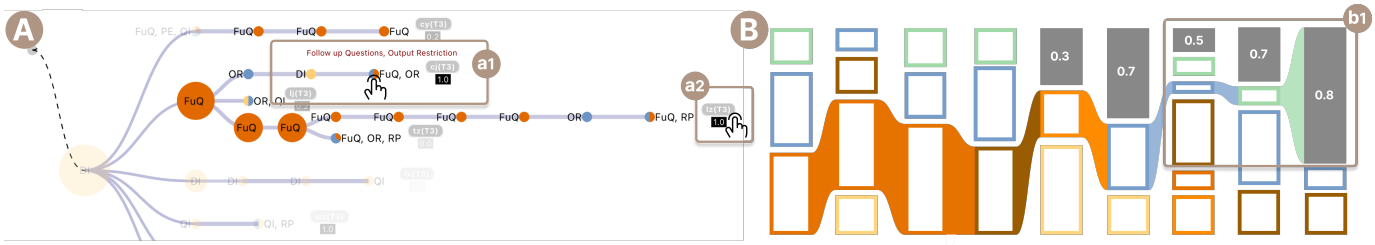

Fig. 7: (A) The **Interaction Tree** visualization within *Pattern Nuance* traces the detailed evolution of each student's interaction pattern for specific tasks. The students' paths utilizing {"Follow up Questions"} pattern are highlighted. (B) An alternative design featuring a Sankey-stacked bar chart for tracking pattern evolution.

grey bar with numerical values representing the student's score in this task. Although this design was clear to show the percentage of students' diverse interaction choices with ChatGPT at each conversation turn, the experts (E1 &E2) prioritized our Interaction Tree visualization since the quality of each ChatGPT's response was lacking and could be hard to add to the Sankey stacked bar chart easily. Meanwhile, they preferred a nuanced comparison between different students' interaction patterns, which is also a limitation of the Sankey-form diagram.

5.4 Detail View

Instructors can select the student's alias and task's ID tag at the end of each path in the *Interaction Tree* (Fig. 7-a2) to examine the specifics of each task and students' responses in the *Task Description* (Fig. 1-c1), and explore students' raw conversation with ChatGPT in the *Raw Conversation* (Fig. 1-c2). This functionality not only validates the analytical findings but also equips instructors with concrete examples and references for crafting feedback to students or refining task designs, serving as the end of our whole analysis workflow (**R6**).

6 Evaluation

This section delves into the assessment of StuGPTViz. We present the result of a student questionnaire to validate the settings of our open-ChatGPT in-class exercises. Then, we demonstrate the effectiveness of the Information Gain (IG) metric we introduced in 4.3. Subsequently, we showcase the system's capacity to facilitate the identification and analysis of student interactions with ChatGPT through three case studies. Additionally, we discuss the collective feedback from interviews with six domain experts (**E1-E6**). The insights from these experts evaluated the StuGPTViz's effectiveness and impact.

6.1 Questionnaire Feedback

We administered a mid-semester voluntary questionnaire consisting of multiple-choice and short-answer questions. The multiple-choice questions were designed to gauge students' experiences and attitudes towards using ChatGPT to learn data visualization and complete the designed in-class exercises. The short-answer questions aimed to collect students' general feedback, including their level of trust in and feelings about using ChatGPT for learning data visualization and other subjects. The detailed statistics of the questionnaire are provided in the supplementary (B). To summarize, the results indicated a strong positive reception: more than 90% students reported enjoying using ChatGPT in their learning process and expressed a willingness to utilize it extensively in our data visualization course. These findings affirmed the rationality behind our course material design and ensured the quality of the data collected for our study. Some other insights from the short-answer questions are discussed in the following Sec. 7.

6.2 Metric Evaluation

We evaluated the effectiveness of our Information Gain (IG) metric by sampling 10% of student-ChatGPT conversations. Two experts (E1 & E2) manually labeled the data into three categories: "low information gain" (score 0), "average information gain" (score 0.5), and "high information gain" (score 1). A score of 0 was assigned to responses containing mostly incorrect information, a score of 1 to responses providing rich and accurate new information, and a score of 0.5 to responses that were partially inaccurate or partially redundant with previous. To measure the correlation between IG metric and experts' labeling, we calculated Pearson correlation [63], Spearman correlation [64] and Kendall Rank correlation [2] between them. The results are 0.609 ($p = 0.00$), 0.621 ($p = 0.00$) and 0.497 ($p = 0.00$), respectively. All the results show there is a moderate to strong and significant positive correlation between the IG metric and experts' judgment of ChatGPT's response quality. The sample of labeled data and metric results is provided in the supplementary materials. Although effective, the IG metric is a simplified measure that primarily considers word frequency, designed to provide an initial assessment. In the future, we plan to use more advanced third-party ChatGPT response quality evaluation methods to enhance the metric accuracy.

6.3 Case Study

We engaged experts (E1-E6) to evaluate StuGPTViz independently. We introduced the background, visual designs, and a brief workflow demo to them and yielded three case studies that underscore the system's utility in analyzing student interactions with ChatGPT. Experts E1 and E2 are the course instructors we collaborated with, and E3-E6 are newly invited experts, including three assistant professors and one lecturer from three different universities, all with expertise in data visualization.

6.3.1 Case 1: Enhancing Students' ChatGPT Utilization

In the first case study, E1 and E3 leveraged StuGPTViz to derive instructional strategies to optimize students' use of ChatGPT for learning and addressing challenging tasks.

Initial Overview and Task Filtering: The investigation began with an overview of tasks and student backgrounds. To focus on challenging tasks, the experts used the *Task Overview* to filter out tasks with difficulty scores below 3, those categorized under "self-learning" and "remember," and those with average scores exceeding 0.8 (Fig. 1(**a1**)). In the *Student Overview*, all students were retained to ensure a comprehensive analysis of diverse interaction patterns (Fig. 1(**a2**)).

Identifying Challenges and Specific Tasks: To review task summary patterns, the experts switched to the "Task-Grouping" mode in the *Pattern View* (Fig. 1 **b1**) and identified "analyze" tasks as particularly challenging, evidenced by longer stacked bars indicating higher cognitive engagement (Fig. 1 **b2**). Although the thick grey bar representing "Information Gain" suggested students acquired significant information from ChatGPT, the overall learning outcomes for these tasks were suboptimal (Fig. 1 **b2**). Consequently, the experts focused on the "analyze" task group, identifying "Task 3" as notably difficult due to the extensive cognitive processing it required(Fig. 1 **b3**).

Analysis of Interaction Patterns: The examination of "Task 3" through the pattern mining table and Interaction Tree (Fig. 1-b4, b5) revealed nuanced dynamics of student-ChatGPT interactions. Initially, sorting by average score revealed infrequent and overly specific patterns. To uncover broadly applicable patterns, experts shifted to sorting by frequency ("Count"), identifying a recurring and effective pattern: a combination of "Definition Inquiry" and "Follow up Questions" (Fig. 1-b6), notable for its prevalence and high average score exceeding 0.8.

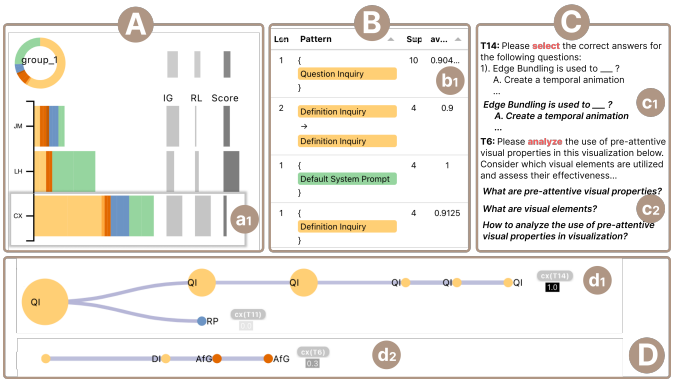

Fig. 8: (A) The interaction pattern summary within Group 1. (B) The pattern table, sorting all interaction patterns mined from students with alias "cs". (C) The descriptions for T14 & T6 and the student's key prompts. (D) The pattern paths of student "cs" for T14 and T6.

Experts agreed this simple yet effective combination was crucial for foundational understanding and deeper exploration of ChatGPT's responses. To observe students using this pattern, experts selected it in the table, highlighting corresponding students in the Interaction Tree (Fig. 1-b8, Fig. 7-B). A standout instance involved a student, "CJ", who used this pattern to achieve a full score with minimal conversation (Fig. 7-b2). By clicking the student-task ID tag (Fig. 7-b2) and reviewing the raw conversations (Fig. 1-c3) along the tree branch, experts saw "CJ" begin with a "Definition Inquiry", asking if ChatGPT understood the pipeline concept [61]. After confirmation, "CJ" asked a "Follow up Question" for further exploration. Noticing that ChatGPT did not provide answers from the multiple choices listed in the task description, "Output Restrictions" were strategically implemented. Despite these restrictions, ChatGPT's responses remained flawed, leading "CJ" to revert to "Definition Inquiry" for double verification before progressing. Once a correct response was secured from ChatGPT, "CJ" advanced through the remaining sections, receiving partially correct responses from ChatGPT but ultimately providing completely correct answers independently. Experts concluded that "Output Restrictions" in prompts maximize ChatGPT's utility for precise information retrieval. Revisiting the pattern mining table (Fig. 1-b4), they noted "Output Restrictions" as a significant pattern (Fig. 1-b7) due to its high frequency and average score. Thus, the experts finally identified the combination of "Definition Inquiry" and "Follow up Questions" with "Output Restrictions" as a potent strategy for engaging with ChatGPT effectively. The raw conversation data of "CJ" is provided in the supplementary (C). This evaluation also highlighted ChatGPT's limitations in addressing abstract questions, advising against outright reliance on its answers but encouraging a focus on its reasoning and factual accuracy. Ultimately, the experts planned to incorporate these insights into feedback for their students. .

This case study showcases StuGPTViz's capability to not only unearth effective student interaction patterns with ChatGPT but also distill these insights into actionable strategies for educators.

6.3.2 Case 2: Provide Personalized Feedback

Experts E3 and E4 aimed to provide personalized feedback to students lagging behind, focusing on those with average scores below 0.5 in the *Student Overview*. They identified "Group 1" from the *Pattern View* donuts chart as the weakest, with student "cx" showing extensive ChatGPT interaction but the lowest scores (Fig. 8-a1).

To better understand "cx's" interactions, the experts analyzed the *Pattern Table* (Fig. 8-B) sorted by usage frequency ("Count"), revealing "cx" predominantly engaged in basic "Question Inquiry" and "Definition Inquiry" and rarely modified default settings (Fig. 8-b1). This pattern indicated that "cx" primarily operated at initial cognitive levels, showing a dependency on ChatGPT's capabilities rather than engaging in self-driven learning or critical thinking.

However, further analysis showed a distinct contrast in "cx's" engagement with various tasks. In task T14, consisting of straightforward multiple-choice questions, "cx" achieved full marks by directly copying questions into ChatGPT (Fig. 8-c1, d1). While effective for securing marks, this approach had limited educational value as it bypassed the learning process. In contrast, despite the lower score in T6, "cx" demonstrated a significant effort to grasp the underlying concepts and principles of analysis, as evidenced by the detailed prompts samples (Fig. 8-c2) and the orange nodes in the interaction path (Fig. 8-d1).

The experts identified that T6 was an exercise in "visualization analysis and evaluation" and speculated that "cx" had a keen interest in tasks requiring analysis and evaluation rather than just recalling basic concepts. Consequently, the experts proposed a shift in "cx's" educational strategy, advocating for an increased focus on tasks that emphasize analysis and evaluation and moving away from tasks that merely require regurgitating information.

This case demonstrates the power of StuGPTViz to aid instructors in providing in-depth personalized feedback to students.

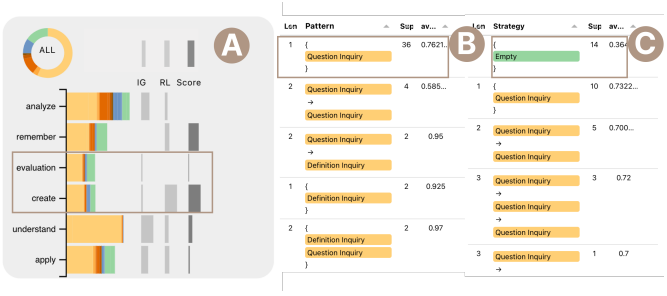

Fig. 9: (A) Summary of students' interaction patterns for each type of task. (B) The pattern table, sorting all interaction patterns mined from students under the "evaluate" type task. (C) The pattern table, sorting all interaction patterns mined from students under the "create" type task.

6.3.3 Case 3: Refine Course Material Design

Experts E2 and E4 employed StuGPTViz to assess the alignment of existing in-class exercises with educational objectives, especially considering unrestricted access to ChatGPT. Their initial step involved excluding "self-learning" tasks from the analysis within the *Task Overview*.

Progressing to the *Pattern View*, they aimed to compare cognitive engagement levels across various task types. The experts found that "evaluate" and "create" tasks elicited notably lower levels of advanced cognitive engagement compared to "analyze" and "remember" tasks (Fig. 9-A). They further examined the "evaluate" tasks and identified a predominant pattern of "Question Inquiry" (Fig. 9-B), indicating students often relayed task descriptions to ChatGPT for answers. A detailed check of the "evaluate" task descriptions in the *Detailed View* revealed this was particularly evident in tasks where students were explicitly instructed to analyze specific visual properties. The direct provision of analysis criteria encouraged a straightforward query-response dynamic with ChatGPT, bypassing deeper cognitive processing.

Based on these insights, E4 recommended a shift in exercise design towards more open-ended questions that prompt students to independently determine relevant metrics or principles before engaging in analysis. This approach aims to ensure alignment with educational objectives considering unrestricted access to ChatGPT. On the other hand, the experts investigated "create" type tasks and found that the analysis results and task description indicated a lack of clear direction for students in formulating queries to ChatGPT, often resulting in incomplete ("Empty") or superficial ("Question Inquiry", mainly refer to question copy & paste) interactions (Fig. 9-C). To address this, it was suggested that instructors could guide students toward requesting alternative design options from ChatGPT, including the rationale behind each. The experts concurred that prompting students to assess these options and incorporate their critical reasoning can greatly enhance the learning experience.

This case illustrates the effectiveness of StuGPTViz in supporting instructors with the redesign of course materials, particularly in integrating ChatGPT, to guarantee the achievement of educational objectives.

6.4 Expert Interview

Following the case studies, one-on-one interviews were conducted with six experts, each lasting around 80 minutes with a $80 compensation.

System Workflow. The workflow of StuGPTViz was praised by all experts for its clarity and functionality. Experts noted its ease of use, allowing for a streamlined narrowing of analysis scope to achieve insights across multiple levels. Whether focusing on groups of students, individuals, or tasks, the system's design facilitated seamless navigation without added complexity. E5, an assistant professor specializing in teaching visualization to business students, highlighted, *"The workflow's logical progression and the interconnection of each view were particularly impressive, enabling a diverse analytical focus through a unified procedure."* E3, E4, and E6 emphasized the importance of understanding students' learning outcomes. They appreciated the system's capability to filter, and highlight scores at various analytical stages.

Visual Design and Interactions. Experts agreed that the visual design and interactive elements of StuGPTViz are clear and user-friendly, significantly enhancing the analytical process. The use of stacked bar charts was particularly commended for facilitating easy understanding and comparison of cognitive levels across students. The color coding, distinguishing between learning-related and ChatGPT usage-related codes, was found intuitive. E2 highlighted the clarity provided by the visual design: *"The ability to discern students' overall cognitive level at a glance is highly appreciated."* The Interaction Tree visualization emerged as a favorite for its detailed representation of different patterns and the entire interaction journey, including ChatGPT responses and learning outcomes. E4, an assistant professor, praised the decision-tree format for showcasing diverse student strategies and ChatGPT's varied response quality. However, concerns were raised about the scalability of this visualization, especially for large classes over 100 students, suggesting a need for refining the summary of popular interaction patterns and detailed comparison of individual paths.

Suggestions. Experts provided several actionable suggestions for enhancing StuGPTViz. E5 proposed adding a summary report panel to capture screenshots and annotate findings directly within the system, facilitating a comprehensive and customizable analysis experience. E6 recommended more flexible options for grouping tasks and students, suggesting a user-defined grouping mechanism to enable richer cross-correlations between different cohorts and tasks, visualized through a matrix-form panel. Despite the potential for increased complexity, E6 believed this feature could unveil deeper insights. Additionally, E1 and E4 suggested integrating the coding of students' conversations with ChatGPT directly into the workflow, similar to a "grading the assignment" process. This would streamline the evaluation process and enrich instructors' understanding of student interactions with ChatGPT.

7 Discussion

This section discusses the significance and insights towards ChatGPT for data visualization education, and the generalizability and scalability of the proposed visual analytics system.

ChatGPT for Data Visualization Education The introduction of ChatGPT into educational ecosystems marks a pivotal moment and necessitates a nuanced understanding of technological integration in learning. Our investigation into the patterns and strategies of student engagement with ChatGPT is critical, providing insights that guide instructors in facilitating the effective use of AI. Our study revealed several key findings from student questionnaires, student-ChatGPT conversation data, and discussions with course instructors. First, students reported that ChatGPT excels in summarizing key concepts, providing quick access to vast information, and offering tailored Q&A sessions. Instructors agreed that these functionalities are particularly beneficial for students with limited backgrounds, such as those who changed their major in graduate school, allowing them to keep pace with coursework without hindering class progress. Additionally, over 90% of students expressed satisfaction with ChatGPT's ability to handle data visualization queries, indicating a strong positive perception of its utility in data visualization education. However, instructors pointed out that recognizing ChatGPT's limitations in interpreting and processing visual data compared to textual information is crucial. For instance, through the collected student-ChatGPT conversations, instructors identified that some students resorted to asking ChatGPT for URLs of existing sample visualizations (e.g., demos of parallel coordinates) to obtain better figure quality after receiving a low-quality response. These findings inspire instructors to develop innovative approaches, such as guiding ChatGPT to provide descriptions or links to reference visualizations, for better educational content. Instructors also emphasized that the demand-driven nature of data visualization requires students to employ precise and strategic questioning techniques. This underscores the importance of teaching students how to prompt effectively to maximize ChatGPT's capabilities. Furthermore, as we navigate the ChatGPT-enhanced educational paradigm, our work highlighted the need to focus on cultivating students' higher-order cognitive skills, such as critical thinking and evaluative judgment. For instance, instructors suggested encouraging students to ask ChatGPT for alternative designs with underlying rationales, enabling them to critically assess options and fostering a collaborative learning dynamic. Approaches like this redefine the role of ChatGPT in education—from a universal solver to a pedagogical partner—and emphasize the importance of critical engagement and decision-making skills in the data visualization domain.

Generalizability & Scalability The application of StuGPTViz, while rooted in the context of data visualization education, unveils broader implications for ChatGPT-assisted learning environments. It offers a robust framework for analyzing student interactions with ChatGPT across a variety of courses. StuGPTViz's core workflow, which includes filtering tasks and students of interest, analyzing cognitive levels through prompts, assessing prompt engineering skills, evaluating ChatGPT's response quality, and identifying interaction patterns, encapsulates the universal aspects of leveraging ChatGPT in education. This approach enriches our understanding of effective AI integration into pedagogy and opens avenues for examining ethical considerations regarding ChatGPT's involvement in teaching practices. Regarding scalability, StuGPTViz demonstrates competence in managing classes of 48 students through the interaction tree visualization techniques equipped with simple pruning for clarity. This capability ensures the StuGPTViz's efficacy in distilling actionable insights from complex datasets, catering to the needs of regular-sized classes. However, scalability challenges arise as class sizes expand beyond this scope. For larger cohorts (e.g., exceeding 100 students) the incorporation of edge bundling techniques emerges as a potential refinement to enhance pattern visualization and comparison. Additionally, expanding the pattern mining table to include interaction patterns from different tasks or courses would further enhance its scalability and provide additional benefits. This adaptation will form a component of our further efforts, aiming to ensure that StuGPTViz remains effective in in diverse educational settings, advancing the goal of inclusive AI-enhanced education.

8 Conclusion

This study introduces StuGPTViz, a visual analytics system for instructors to analyze student-ChatGPT interactions. In particular, we collected student-ChatGPT conversations in a graduate-level data visualization course and developed a comprehensive coding scheme to categorize students' prompts from cognitive levels and ChatGPT's response qualities. We then build StuGPTViz to visualize student-ChatGPT interaction patterns and support multi-level, multi-perspective analysis. StuGPTViz empowers instructors with deep insights into students' cognitive processes, their reliance on ChatGPT, and their ability to use it effectively, highlighting areas for pedagogical intervention to promote higher-order thinking. The system's effectiveness, validated through expert interviews and case studies, confirms its potential to impact student-ChatGPT conversation analysis and visualization education. As we look to the future, StuGPTViz sets the stage for broader research into the application of visual analytics in education and the development of AI-enhanced personalized learning experiences.

REFERENCES

[1] R. Abdelghani, Y.-H. Wang, X. Yuan, T. Wang, P. Lucas, H. Sauzéon, and P.-Y. Oudeyer. Gpt-3-driven pedagogical agents to train children's curious question-asking skills. *International Journal of Artificial Intelligence in Education*, pp. 1–36, 2023. doi: 10.1007/s40593-023-00340-7 2

[2] H. Abdi. The kendall rank correlation coefficient. *Encyclopedia of measurement and statistics*, 2:508–510, 2007. 7

[3] M. A. AlAfnan, S. Dishari, M. Jovic, and K. Lomidze. Chatgpt as an educational tool: Opportunities, challenges, and recommendations for communication, business writing, and composition courses. *Journal of Artificial Intelligence and Technology*, 3(2):60–68, 2023. doi: 10.37965/jait.2023.0184 2, 3

[4] S. S. Alhadad. Visualizing data to support judgement, inference, and decision making in learning analytics: Insights from cognitive psychology and visualization science. *Journal of Learning Analytics*, 5(2):60–85, 2018. doi: 10.18608/jla.2018.52.5 2

[5] N. R. Aljohani, A. Daud, R. A. Abbasi, J. S. Alowibdi, M. Basheri, and M. A. Aslam. An integrated framework for course adapted student learning analytics dashboard. *Computers in Human Behavior*, 92:679–690, 2019. doi: 10.1016/j.chb.2018.03.035 2

[6] D. Angus, A. Smith, and J. Wiles. Conceptual recurrence plots: Revealing patterns in human discourse. *IEEE Transactions on Visualization and Computer Graphics*, 18(6):988–997, 2012. doi: 10.1109/TVCG.2011.100 3

[7] D. Angus, B. Watson, A. Smith, C. Gallois, and J. Wiles. Visualising conversation structure across time: Insights into effective doctor-patient consultations. *PloS one*, 7(6):e38014, 2012. doi: 10.1371/journal.pone.0038014 3

[8] B. Bach, M. Keck, F. Rajabiyazdi, T. Losev, I. Meirelles, J. Dykes, R. S. Laramee, M. AlKadi, C. Stoiber, S. Huron, et al. Challenges and opportunities in data visualization education: A call to action. *IEEE Transactions on visualization and computer graphics*, 2023. doi: 10.1109/TVCG.2023.3327378 2, 3

[9] L. Bai, X. Liu, and J. Su. Chatgpt: The cognitive effects on learning and memory. *Brain-X*, 1(3):e30, 2023. doi: 10.1002/brx2.30 2

[10] D. Baidoo-Anu and L. O. Ansah. Education in the era of generative artificial intelligence (ai): Understanding the potential benefits of chatgpt in promoting teaching and learning. *Journal of AI*, 7(1):52–62, 2023. doi: 10.61969/jai.1337500 1, 2

[11] A. L. Ball and B. L. Garton. Modeling higher order thinking: The alignment between objectives, classroom discourse, and assessment. *Journal of Agricultural Education*, 46(2):58–69, 2005. doi: 10.5032/jae.2005.02058 2

[12] C. A. Bonfield, M. Salter, A. Longmuir, M. Benson, and C. Adachi. Transformation or evolution?: Education 4.0, teaching and learning in the digital age. *Higher education pedagogies*, 5(1):223–246, 2020. doi: 10.1080/23752696.2020.1816847 1, 2

[13] E. Bonner, R. Lege, and E. Frazier. Large language model-based artificial intelligence in the language classroom: Practical ideas for teaching. 2023, 02 2023. doi: 10.56297/BKAM1691/WIEO1749 1

[14] V. Braun and V. Clarke. *Thematic analysis.* American Psychological Association, 2012. doi: 10.1080/17439760.2016.1262613 2

[15] S. Bull and J. Kay. Open learner models. In *Advances in intelligent tutoring systems*, pp. 301–322. Springer, 2010. doi: 10.1007/978-3-642-14363-2_15 2

[16] S. Chaudhary. Code alpaca: An instruction-following llama model for code generation. https://github.com/sahil280114/codealpaca, 2023. 2

[17] W. Chen, M. Yin, M. Ku, P. Lu, Y. Wan, X. Ma, J. Xu, X. Wang, and T. Xia. Theoremqa: A theorem-driven question answering dataset. *arXiv preprint arXiv:2305.12524*, 2023. 2

[18] L. Corrin, G. Kennedy, P. De Barba, A. Bakharia, L. Lockyer, D. Gasevic, D. Williams, S. Dawson, and S. Copeland. Loop: A learning analytics tool to provide teachers with useful data visualisations. In *Proceedings of the 32nd Annual Conference of the Australasian Society for Computers in Learning and Tertiary Education (ASCILITE 2015)*, pp. 409–413. Ascilite, 2015. 2

[19] A. Crowe, C. Dirks, and M. P. Wenderoth. Biology in bloom: implementing bloom's taxonomy to enhance student learning in biology. *CBE—Life Sciences Education*, 7(4):368–381, 2008. doi: 10.1187/cbe.08-05-0024 3

[20] Y. Cui, W. G. Lily, Y. Ding, F. Yang, L. Harrison, and M. Kay. Adaptive assessment of visualization literacy. *IEEE Transactions on Visualization and Computer Graphics*, 2023. doi: 10.1109/TVCG.2023.3327165 2

[21] A. Darvishi, H. Khosravi, S. Sadiq, D. Gašević, and G. Siemens. Impact of ai assistance on student agency. *Computers & Education*, 210:104967, 2024. doi: 10.1016/j.compedu.2023.104967 2

[22] W. Dou, L. Yu, X. Wang, Z. Ma, and W. Ribarsky. Hierarchicaltopics: Visually exploring large text collections using topic hierarchies. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2002–2011, 2013. doi: 10.1109/TVCG.2013.162 2

[23] E. Duval. Attention please! learning analytics for visualization and recommendation. In *Proceedings of the 1st international conference on learning analytics and knowledge*, pp. 9–17, 2011. doi: 10.1145/2090116.2090118 2

[24] M. El-Assady, R. Sevastjanova, D. Keim, and C. Collins. Threadreconstructor: Modeling reply-chains to untangle conversational text through visual analytics. vol. 37, pp. 351–365, 2018. doi: 10.1111/cgf.13425 3

[25] C. D. Epp and S. Bull. Uncertainty representation in visualizations of learning analytics for learners: Current approaches and opportunities. *IEEE Transactions on Learning Technologies*, 8(3):242–260, 2015. doi: 10.1109/TLT.2015.2411604 2

[26] S. Es, J. James, L. Espinosa-Anke, and S. Schockaert. Ragas: Automated evaluation of retrieval augmented generation. *arXiv preprint arXiv:2309.15217*, 2023. doi: 10.48550/arXiv.2309.15217 2, 5

[27] L. I. D. Faruk, R. Rohan, U. Ninrutsirikun, and D. Pal. University students' acceptance and usage of generative ai (chatgpt) from a psycho-technical perspective. In *Proceedings of the 13th International Conference on Advances in Information Technology*, pp. 1–8, 2023. doi: 10.1145/3628454.3629552 2

[28] S. Fergus, M. Botha, and M. Ostovar. Evaluating academic answers generated using chatgpt. *Journal of Chemical Education*, 100(4):1672–1675, 2023. doi: 10.1021/acs.jchemed.3c00087 5

[29] M. Firat. How chat gpt can transform autodidactic experiences and open education. *Department of Distance Education, Open Education Faculty, Anadolu Unive*, 2023. doi: 10.31219/osf.io/9ge8m 2

[30] S. Fu, Y. Wang, Y. Yang, Q. Bi, F. Guo, and H. Qu. Visforum: A visual analysis system for exploring user groups in online forums. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 8(1):1–21, 2018. doi: 10.1145/3162075 3

[31] S. Fu, J. Zhao, H. F. Cheng, H. Zhu, and J. Marlow. T-cal: Understanding team conversational data with calendar-based visualization. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, 13 pages, p. 1–13. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3174074 3

[32] S. Fu, J. Zhao, W. Cui, and H. Qu. Visual analysis of mooc forums with iforum. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):201–210, 2017. doi: 10.1109/TVCG.2016.2598444 3

[33] K. Fuchs. Exploring the opportunities and challenges of nlp models in higher education: is chat gpt a blessing or a curse? In *Frontiers in Education*, vol. 8, p. 1166682. Frontiers, 2023. doi: 10.3389/feduc.2023.1166682 3

[34] W. Gan, Z. Qi, J. Wu, and J. C.-W. Lin. Large language models in education: Vision and opportunities. In *2023 IEEE International Conference on Big Data (BigData)*, pp. 4776–4785, 2023. doi: 10.1109/BigData59044.2023.10386291 1

[35] S. Grassini. Shaping the future of education: Exploring the potential and consequences of ai and chatgpt in educational settings. *Education Sciences*, 13(7), 2023. doi: 10.3390/educsci13070692 1, 2

[36] M. Hao, C. Rohrdantz, H. Janetzko, U. Dayal, D. A. Keim, L.-E. Haug, and M.-C. Hsu. Visual sentiment analysis on twitter data streams. In *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 277–278, 2011. doi: 10.1109/VAST.2011.6102472 2, 3

[37] E. Hoque and G. Carenini. Multiconvis: A visual text analytics system for exploring a collection of online conversations. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, IUI '16, 12 pages, p. 96–107. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2856767.2856782 3

[38] J. J. Huallpa et al. Exploring the ethical considerations of using chat gpt in university education. *Periodicals of Engineering and Natural Sciences*, 11(4):105–115, 2023. doi: 10.21533/pen.v11i4.3770 1, 2

[39] M. Imran and N. Almusharraf. Analyzing the role of chatgpt as a writing assistant at higher education level: A systematic review of the literature. *Contemporary Educational Technology*, 15(4):ep464, 2023. doi: 10.30935/cedtech/13605 1

[40] M. Javaid, A. Haleem, R. P. Singh, S. Khan, and I. H. Khan. Unlocking

the opportunities through chatgpt tool towards ameliorating the education system. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, 3(2):100115, 2023. doi: 10.1016/j.tbench.2023.100115 2

[41] J. Jeon and S. Lee. Large language models in education: A focus on the complementary relationship between human teachers and chatgpt. *Education and Information Technologies*, pp. 1–20, 2023. doi: 10.1007/s10639-023-11834-1 1

[42] D. Johnson, R. Goodman, J. Patrinely, C. Stone, E. Zimmerman, R. Donald, S. Chang, S. Berkowitz, A. Finn, E. Jahangir, et al. Assessing the accuracy and reliability of ai-generated medical responses: an evaluation of the chat-gpt model. *Research square*, 2023. doi: 10.21203/rs.3.rs-2566942/v1 5

[43] J. Kocoń, I. Cichecki, O. Kaszyca, M. Kochanek, D. Szydło, J. Baran, J. Bielaniewicz, M. Gruza, A. Janz, K. Kanclerz, et al. Chatgpt: Jack of all trades, master of none. *Information Fusion*, p. 101861, 2023. doi: 10.1016/j.inffus.2023.101861 5

[44] D. Krathwohl. A revision of bloom's taxonomy: An overview. *Theory Into Practice - THEORY PRACT*, 41:212–218, 11 2002. doi: 10.1207/s15430421tip4104_2 2, 3, 4

[45] H. Kumar, I. Musabirov, M. Reza, J. Shi, A. Kuzminykh, J. J. Williams, and M. Liut. Impact of guidance and interaction strategies for llm use on learner performance and perception. *arXiv preprint arXiv:2310.13712*, 2023. doi: 10.48550/arXiv.2310.13712 1

[46] B. C. Kwon, S.-H. Kim, S. Lee, J. Choo, J. Huh, and J. S. Yi. Visohc: Designing visual analytics for online health communities. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):71–80, 2016. doi: 10.1109/TVCG.2015.2467555 2, 3

[47] K. Lan and N.-S. Chen. Teachers' agency in the era of llm and generative ai: Designing pedagogical ai agents. *Educational Technology and Society*, 27:i–xviii, 01 2024. doi: 10.30191/ETS.202401_27(1).PP01 1

[48] C. Lee and G. G. Lee. Information gain and divergence-based feature selection for machine learning-based text categorization. *Information processing & management*, 42(1):155–165, 2006. doi: 10.1016/j.ipm.2004.08.006 2

[49] R. Li, E. Hoque, G. Carenini, R. Lester, and R. Chau. Conviscope: Visual analytics for exploring patient conversations. In *2021 IEEE Visualization Conference (VIS)*, pp. 151–155. IEEE, 2021. doi: 10.1109/VIS49827.2021.9623269 3

[50] K. Littleton and C. Howe. *Educational dialogues: Understanding and promoting productive interaction*. Routledge, 2010. doi: 10.1080/02601370.2011.636237 3

[51] C. Liu and H.-Y. Shum. Kullback-leibler boosting. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, vol. 1, pp. I–I, 2003. doi: 10.1109/CVPR.2003.1211407 5

[52] L. Y.-H. Lo, Y. Ming, and H. Qu. Learning vis tools: Teaching data visualization tutorials. In *2019 IEEE Visualization Conference (VIS)*, pp. 11–15. IEEE, 2019. doi: 10.1109/VISUAL.2019.8933751 2

[53] J. Lodge, K. Thompson, and L. Corrin. Mapping out a research agenda for generative artificial intelligence in tertiary education. *Australasian Journal of Educational Technology*, 39:1–8, 05 2023. doi: 10.14742/ajet.8695 1

[54] S. MacNeil, K. Kiefer, B. Thompson, D. Takle, and C. Latulipe. Ineqdetect: A visual analytics system to detect conversational inequality and support reflection during active learning. In *Proceedings of the ACM Conference on Global Computing Education*, CompEd '19, 7 pages, p. 85–91. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3300115.3309528 3

[55] D. Mhlanga. Open ai in education, the responsible and ethical use of chatgpt towards lifelong learning. pp. 387–409, 2023. doi: 10.1007/978-3-031-37776-1_17 1, 2

[56] T. Munzner. *Visualization analysis and design*. CRC press, 2014. 6

[57] V. Owan, K. Abang, D. Idika, and B. Bassey. Exploring the potential of artificial intelligence tools in educational measurement and assessment. *Eurasia Journal of Mathematics, Science and Technology Education*, 19:Article ID em2307, 06 2023. doi: 10.29333/ejmste/13428 1

[58] Y. Park and I.-H. Jo. Development of the learning analytics dashboard to support students' learning performance. *JUCS - Journal of Universal Computer Science*, 21(1):110–133, 2015. doi: 10.3217/jucs-021-01-0110 2

[59] M. M. Rahman and Y. Watanobe. Chatgpt for education and research: Opportunities, threats, and strategies. *Applied Sciences*, 13(9), 2023. doi: 10.3390/app13095783 1

[60] P. Riehmann, M. Hanfler, and B. Froehlich. Interactive sankey diagrams. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pp. 233–240, 2005. doi: 10.1109/INFVIS.2005.1532152 6

[61] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. Ellis, and D. A. Keim. Knowledge generation model for visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1604–1613, 2014. doi: 10.1109/TVCG.2014.2346481 8

[62] R. Schmucker, M. Xia, A. Azaria, and T. Mitchell. Ruffle&riley: Towards the automated induction of conversational tutoring systems. *arXiv preprint arXiv:2310.01420*, 2023. doi: 10.48550/arXiv.2310.01420 2

[63] P. Sedgwick. Pearson's correlation coefficient. *Bmj*, 345, 2012. 7

[64] P. Sedgwick. Spearman's rank correlation coefficient. *Bmj*, 349, 2014. 7

[65] S. Shankar, H. Li, P. Asawa, M. Hulsebos, Y. Lin, J. Zamfirescu-Pereira, H. Chase, W. Fu-Hinthorn, A. G. Parameswaran, and E. Wu. Spade: Synthesizing assertions for large language model pipelines. *arXiv preprint arXiv:2401.03038*, 2024. doi: 10.48550/arXiv.2401.03038 2, 4

[66] R. Sheng, L. Yang, H. Li, Y. Luo, Z. Xu, Z. Zhou, D. Gotz, and H. Qu. Knowledge compass: A question answering system guiding students with follow-up question recommendations. In *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23 Adjunct, article no. 65, 4 pages. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3586182.3615785 2

[67] E. Tajik and F. Tajik. A comprehensive examination of the potential application of chat gpt in higher education institutions. 04 2023. doi: 10.2139/ssrn.4699304 1

[68] E. Tajik and F. Tajik. A comprehensive examination of the potential application of chat gpt in higher education institutions. *TechRxiv. Preprint*, pp. 1–10, 2023. doi: 10.36227/techrxiv.22589497 2

[69] M. Vaismoradi, H. Turunen, and T. Bondas. Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing & Health Sciences*, 15(3):398–405, 2013. doi: 10.1111/nhs.12048 4

[70] T. van Erven and P. Harremos. Rényi divergence and kullback-leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, 2014. doi: 10.1109/TIT.2014.2320500 5

[71] A. R. Vargas-Murillo, I. N. M. de la Asuncion, F. de Jesús Guevara-Soto, et al. Challenges and opportunities of ai-assisted learning: A systematic literature review on the impact of chatgpt usage in higher education. *International Journal of Learning, Teaching and Educational Research*, 22(7):122–135, 2023. doi: 10.26803/ijlter.22.7.7 2

[72] Q. Wang, K. Saha, E. Gregori, D. Joyner, and A. Goel. Towards mutual theory of mind in human-ai interaction: How language reflects what students perceive about a virtual teaching assistant. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, article no. 384, 14 pages. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445645 2

[73] J. White, Q. Fu, S. Hays, M. Sandborn, C. Olea, H. Gilbert, A. Elnashar, J. Spencer-Smith, and D. C. Schmidt. A prompt pattern catalog to enhance prompt engineering with chatgpt. *arXiv preprint arXiv:2302.11382*, 2023. doi: 10.48550/arXiv.2302.11382 2, 4

[74] J. White, S. Hays, Q. Fu, J. Spencer-Smith, and D. C. Schmidt. Chatgpt prompt patterns for improving code quality, refactoring, requirements elicitation, and software design. *arXiv preprint arXiv:2303.07839*, 2023. doi: 10.48550/arXiv.2303.07839 2, 4

[75] M. Xia, X. Zhao, D. Sun, Y. Huang, J. Sewall, and V. Aleven. Involving teachers in the data-driven improvement of intelligent tutors: A prototyping study. In *Artificial Intelligence in Education: 24th International Conference, AIED 2023, Tokyo, Japan, July 3–7, 2023, Proceedings*, 13 pages, p. 340–352. Springer-Verlag, Berlin, Heidelberg, 2023. doi: 10.1007/978-3-031-36272-9_28 2

[76] B. R. A. N. Ximena Zúñiga and T. D. Sevig. Intergroup dialogues: An educational model for cultivating engagement across differences. *Equity & Excellence in Education*, 35(1):7–17, 2002. doi: 10.1080/713845248 3

[77] J. Yang, H. Jin, R. Tang, X. Han, Q. Feng, H. Jiang, S. Zhong, B. Yin, and X. Hu. Harnessing the power of llms in practice: A survey on chatgpt and beyond. *ACM Trans. Knowl. Discov. Data*, feb 2024. Just Accepted. doi: 10.1145/3649506 2

[78] L. Zheng, W.-L. Chiang, Y. Sheng, T. Li, S. Zhuang, Z. Wu, Y. Zhuang, Z. Li, Z. Lin, E. Xing, et al. Lmsys-chat-1m: A large-scale real-world llm conversation dataset. *arXiv preprint arXiv:2309.11998*, 2023. doi: 10.48550/arXiv.2309.11998 2