

Article

Exploring the Effectiveness of Advanced Chatbots in Educational Settings: A Mixed-Methods Study in Statistics

Gustavo Navas ^{1,*} , Gustavo Navas-Reascos ² , Gabriel E. Navas-Reascos ³  and Julio Proaño-Orellana ¹ 

¹ IDEIAGEOCA Group, Universidad Politécnica Salesiana, Moran Valverde S/N, Quito 170702, Ecuador; jproaño@ups.edu.ec

² Tecnológico de Monterrey, Escuela de Ingeniería y Ciencias, Ave. Eugenio Garza Sada 2501, Monterrey 64849, Mexico; a01681952@tec.mx

³ Tecnológico de Monterrey, School of Engineering and Science, Mexico City 14380, Mexico; gabriel.navas.reascos@gmail.com

* Correspondence: gnavas@ups.edu.ec

Abstract: The Generative Pre-trained Transformer (GPT) is a highly advanced natural language processing model. This model can generate conversation-style responses to user input. The rapid rise of GPT has transformed academic domains, with studies exploring the potential of chatbots in education. This research investigates the effectiveness of ChatGPT 3.5, ChatGPT 4.0 by OpenAI, and Chatbot Bing by Microsoft in solving statistical exam-type problems in the educational setting. In addition to quantifying the errors made by these chatbots, this study seeks to understand the causes of these errors to provide recommendations. A mixed-methods approach was employed to achieve this goal, including quantitative and qualitative analyses (Grounded Theory with semi-structured interviews). The quantitative stage involves statistical problem-solving exercises for undergraduate engineering students, revealing error rates based on the reason for the error, statistical fields, sub-statistics fields, and exercise types. The quantitative analysis provided crucial information necessary to proceed with the qualitative study. The qualitative stage employs semi-structured interviews with selected chatbots; this includes confrontation between them that generates agreement, disagreement, and differing viewpoints. On some occasions, chatbots tend to maintain rigid positions, lacking the ability to adapt or acknowledge errors. This inflexibility may affect their effectiveness. The findings contribute to understanding the integration of AI tools in education, offering insights for future implementations and emphasizing the need for critical evaluation and responsible use.

Keywords: statistical education; ChatGPT; chatbots; qualitative analysis; quantitative analysis



Citation: Navas, G.; Navas-Reascos, G.; Navas-Reascos, G.E.; Proaño-Orellana, J. Exploring the Effectiveness of Advanced Chatbots in Educational Settings: A Mixed-Methods Study in Statistics. *Appl. Sci.* **2024**, *14*, 8984. <https://doi.org/10.3390/app14198984>

Academic Editors: Stefan Fischer and Gianluigi Ferrari

Received: 26 July 2024

Revised: 8 September 2024

Accepted: 17 September 2024

Published: 5 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The rapid advancement of the Generative Pre-trained Transformer (GPT) has ushered in a new era in various academic domains. Recent studies have explored the potential of chatbots like ChatGPT as an educational tool in statistics and data science, highlighting their ability to generate human-like responses, create educational content, and aid in statistical programming. For instance, [1] discusses leveraging ChatGPT's capabilities while guiding students in its responsible use.

Similarly, ref. [2] delves into GPT-3.5's role in academic writing, outlining methods like "Chunk Stylist" and "Research Buddy" while balancing its benefits against potential risks. Furthermore, studies evaluating GPT models in higher education programming courses reveal their strengths and limitations in handling complex programming assessments. These insights pave the way for this study, which aims to scrutinize the effectiveness of ChatGPT and similar artificial intelligence (AI) tools in educational settings, particularly focusing on their role in teaching and learning statistics and their impact on academic integrity and editing processes.

The paper by [3] presents a study that evaluates the ability of the Generative Pre-trained Transformer (GPT) models to complete assessments in college-level Python programming courses. Three Python courses that employ a variety of assessments, ranging from simple multiple-choice questions to complex programming projects, were analyzed.

Other research shows concern for academic integrity in the era of generative artificial intelligence by studying and reviewing 37 articles on the topic and presents approaches of the top 20 global universities to mitigate the impact of these tools on intellectual integrity and student learning [4–6].

The paper by Steele [7] focuses on the impact of ChatGPT and similar AI tools on education. It discusses the challenges and opportunities presented by AI in teaching and learning, specifically in areas like reading comprehension, knowledge aggregation, and understanding genre conventions. The paper argues that AI tools like ChatGPT threaten traditional educational practices and offer unique opportunities for enhancing learning and critical thinking. The author emphasizes the need for educators to adapt and integrate AI tools into teaching strategies responsibly.

Since the emergence of ChatGPT in the current world, starting in 2021 [1,8], it is evident that it has transitioned from research labs and study centers to people's pockets through their personal devices, presenting significant challenges in approaching these new educational tools. This brings forth challenges concerning leveraging these advancements in educational studies, especially in the field of statistics [9].

The integration of chatbots like ChatGPT in mathematics education presents unique challenges compared to other subjects. While research indicates that these AI tools can enhance learning experiences, they often struggle with mathematical accuracy and conceptual understanding [10]. ChatGPT frequently produces biased or incorrect mathematical data [11], which can lead to misunderstandings due to overestimating its reliability. Additionally, studies show that ChatGPT has difficulty providing stable support in calculus-related reasoning processes, often resulting in confusion rather than clarity [12]. Moreover, ChatGPT performs well on non-technical questions in fields like quantitative risk management but fails in technical mathematical aspects, highlighting a significant gap in its proficiency for complex mathematical tasks [13].

Considering the information provided previously, AI in education has sparked increasing interest in its potential to enhance teaching and learning [1], as well as a lot of caution and concerns. This research aims to evaluate the use of chatbots in solving statistical problems. Beyond merely quantifying the errors produced by chatbots, the study seeks to understand the underlying causes of these errors and offer recommendations on their use to teachers and students. By employing a mixed-methods approach, the research aims to assess these chatbots' effectiveness through quantitative analysis of numerical outcomes and qualitative insights gained from interviews conducted with the chatbots.

The integration of artificial intelligence chatbots into education has been widely explored in recent literature. The study by Referencias Labadze et al. [14], for instance, offers a comprehensive review of the benefits and challenges of using chatbots in educational settings. Among the main advantages, they highlight personalized support and improved knowledge retention. However, they also identify difficulties such as the accuracy of responses, ethical concerns, and the lack of adequate training for educators. Although their research points out the great potential of these tools, it also emphasizes the barriers that must be overcome to achieve effective adoption.

To achieve this, it was first aimed to assess how effectively ChatGPT could solve statistical problems by evaluating the accuracy of the responses. Subsequently, semi-structured interviews were conducted with a chatbot using qualitative data analysis techniques, primarily employing Grounded Theory (GT) [15–17], specifically the Glaserian version (GGT). An approach was embraced that enabled the analytical elements during interviews with the chatbots. This methodology resulted in compelling findings concerning the application of GGT for the qualitative analysis of the semi-structured interviews with the selected chatbots, devoid of predetermined categories or criteria imposed on the analysis.

To categorize the qualitative process, a search was conducted for GT elements, as suggested by Navas [18]. These elements include data collection, data analysis, basic coding, master core category, data constant comparative, and emerging theory.

This approach allows a holistic exploration of the selected chatbots' responses, perceptions, and opinions, offering a more comprehensive and profound understanding of their interaction. Following the principles of GGT, concepts (basic coding) and emergent patterns (core category) were allowed to arise from the collected interview data rather than imposing a preconceived structure on the analysis. This approach laid a solid foundation for generating theories based on experiences with chatbots, exploring their effectiveness and the perceptions, attitudes, and challenges encountered while using various chatbots in educational contexts. This study represents a step forward in understanding the interaction between artificial intelligence, statistics education, and lessons learned from semi-structured interviews with chatbots.

The findings from this mixed-method approach aim to provide valuable insights, informing future implementations of chatbots or other AI tools in education. It emphasizes the importance of considering numerical results and the lessons and learnings obtained from engaging with chatbots, their responses, errors, and differing outcomes and positions from other chatbots.

It has enabled the formulation of several theories, from identifying the best chatbots suitable for engaging in semi-structured interviews to understanding the lessons educators can glean from data collection and qualitative analysis through GT. It also raises questions about the appropriateness of conducting semi-structured interviews with chatbots.

This paper is structured as follows. The second section details the methodology, providing an in-depth overview of the research approach and techniques employed. The third section presents the Results, showcasing the findings. The fourth section develops the Discussion, offering a comprehensive analysis and interpretation of the findings. Finally, the paper concludes by summarizing the key insights and implications drawn from the study, providing a cohesive ending to the research exploration.

2. Materials and Methods

Before conducting this experiment, several key assumptions were clearly defined to ensure the study's validity. These assumptions included the chatbots' ability to interpret and solve statistical problems in an educational context, the direct comparison between different versions of the chatbots (GPT-3.5 vs. GPT-4.0) on specific tasks, and the methodological structure, which was organized into two stages, quantitative and qualitative, to ensure a comprehensive analysis of the chatbots' performance.

The methodology encompasses quantitative and qualitative stages, each encompassing data collection and subsequent analysis. The methodology is shown in Figure 1.

The GGT is based on semi-structured interviews. In this case, it starts with questions that have been tested in previous evaluations with statistics students. These questions were developed and validated by expert professors in the subject at the university and have been used as evaluation material in a traditional statistics course over 5 or 6 semesters. The answers to these questions were provided by subject matter expert professors and were compared with the responses given by the chatbots.

The quantitative data were gathered by capturing responses generated by ChatGPT 3.5 for statistical problems; the exercises were designed for undergraduate university engineering students. This process was iterated three times to ensure robust representation and minimize bias; in each iteration, the exercises were randomly presented to ChatGPT 3.5.

We used 54 questions from the Moodle platform that served as the final exam for engineering students in a regular undergraduate statistics course. The questions were checked for integrity, including missing special characters, verifying formulas and tables, and other details. Finally, we confirmed, by evaluating its initial answers, that the ChatGPT understood the questions in the first review.

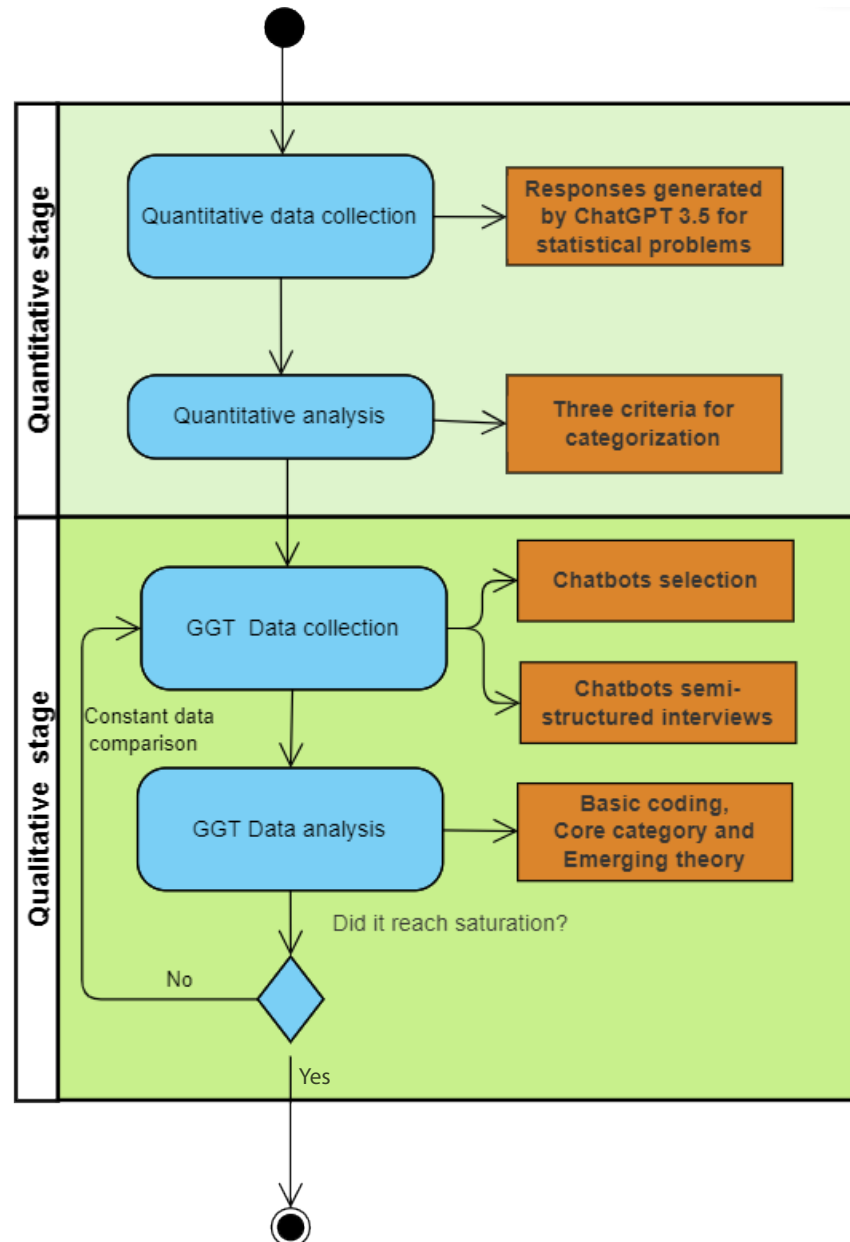


Figure 1. Mixed data analysis methodology with chatbots.

Once the initial round was completed, 52 questions—with a total of 156 items—were chosen from the initial 54 questions (excluding 27 and 28); the process was repeated twice (three times in total). Each question was meticulously copied and pasted into a word processor document, including the questions, and their respective responses provided by the ChatBot, leaving a total of 36 questions that participated in the interviews.

For the Quantitative analysis, the questions were categorized using three criteria:

1. By their type (logical, numerical, or theoretical):
 - Logical: 1 question with 5 items.
 - Numerical: 32 questions with 122 items.
 - Theoretical: 19 questions with 29 items.
2. By the statistical field:
 - Calculated: 5 questions and 5 items.
 - Cloze: 21 questions with 98 items.
 - Drag and Drop: 1 question with 1 item.

- Matching: 7 questions with 34 items.
 - Multiple Choice: 17 questions with 17 items.
 - Numerical: 1 question with 1 item.
3. By the type of question within Moodle:
- Probability: 2 questions with 5 items.
 - Descriptive: 13 questions with 39 items.
 - Inferential: 22 questions with 83 items.
 - Mathematical statistics: 15 questions with 29 items.

All questions were processed using an spreadsheet. A value of 1 was assigned if the response to an item was correct, and a value of 0 was assigned if the item was incorrect. Questions that had any error were selected for qualitative analysis.

The qualitative stage (described in Section 3.2) was conducted using GGT [15,19,20], which involved GGT data collection through semi-structured interviews with selected chatbots, followed by GGT data analysis.

The semi-structured interviews begin with queries seeking solutions to specific problems, followed by subsequent questions emerging organically from the chatbots' responses. These interviews aimed to avoid biases or preconceived ideas and, ideally, identify any recurrent patterns. Through these semi-structured interviews, a qualitative analysis of the chatbots' data or responses was executed, adhering to the fundamental principles of GGT [15,17,19].

The first activity within the GGT data collection was to select the chatbots. One of the primary motivations for initiating the chatbots selection process stemmed from observations made during initial interactions with chatbots, revealing varying degrees of suitability for participating in the semi-structured interview. The selection process relied on the following criteria:

1. It must have been trained under the Generative Pre-trained Transformer (GPT) to establish its origins and standardize its selection.
2. The chatbot should maintain a firm stance and not adapt to the questions posed by the interviewer. During conversation, it should confidently uphold its original positions, only deviating from them exceptionally when the chatbot has acknowledged an error.
3. The chatbot's responses should maintain consistency throughout the entire interview process.

These criteria aim to set a standard concerning the chatbots' training technology and to achieve optimal performance.

For the second activity, semi-structured interviews were conducted with the selected chatbots, with the initial question being the statement of the problem. From here, the interview with the chatbots was tailored based on the response received. It was not deemed necessary to limit the sequence of questions to a specific number of responses but rather to develop it in accordance with the interview process. All interviews were stored in LaTeX format, with each question numbered and followed by a subheading labeled "Investigador" (researcher), indicating the message conveyed to the chatbot stating the problem. This was followed by another subheading with the name of the respective chatbot to which the question was directed, along with its proposed solution.

According to GGT, these interviews constituted the data collection phase, followed by data analysis through constant comparison of data until reaching what is known within GT as saturation, thereby attaining the theories emerging from the process.

Basic coding, master core category, and emerging theory are interconnected concepts within GT analysis, particularly in software development [18]. Basic coding initiates the GT analysis by breaking down data into discrete parts, allowing for the identification and categorization of fundamental elements. This leads to identifying a master core category, which serves as the central theme around which the emerging theory is developed. The emerging theory is the culmination of this analytical process, providing a nuanced understanding of the data beyond the initial categories to offer a comprehensive explanation or theory that

addresses the research questions [21]. These concepts are part of a larger methodological approach that emphasizes iterative analysis and the development of theories grounded in empirical data [15,18,22–24].

This knowledge adds another layer to the understanding of data collection and analysis [24], wherein the interaction and responses obtained from a chatbot during an interview serve as valuable qualitative data in the GT methodology. This perspective allows for a more comprehensive approach to extracting meaningful insights and theories from the interaction between humans and AI, enriching the qualitative analysis and reinforcing the idea that all interactions, including those with chatbots, contribute to the dataset for analysis within the GT framework.

3. Results

3.1. Quantitative Stage

The researcher compared the responses provided by ChatGPT with the correct answers the expert gave. When the chatbot's answers were incorrect, these errors were categorized based on five different criteria: first, the type of error committed by the chatbot; second, the classification according to the types of Moodle applied; third, the relevant field within statistics; fourth, the subfield of statistics, and finally, classification of errors by type: logical, numerical, and theoretical.

In the first case, the types of errors committed by the chatbot were analyzed, as shown in Figure 2. Most of the errors made by the chatbot were calculation errors, accounting for 65% of the total errors. These involved performing incorrect mathematical operations. Errors in problem formulation, which accounted for 19% of the total, occurred due to an incorrect initial approach to the problem. Errors in solution development emerged during the solution process but did not involve incorrect mathematical operations, such as modifying or incorrectly using an equation. Finally, a small percentage of errors were related to data collection, where the chatbot altered the original data of the problem, and to solution recording, where, despite correctly solving the problem, the final presentation of the result was incorrect.

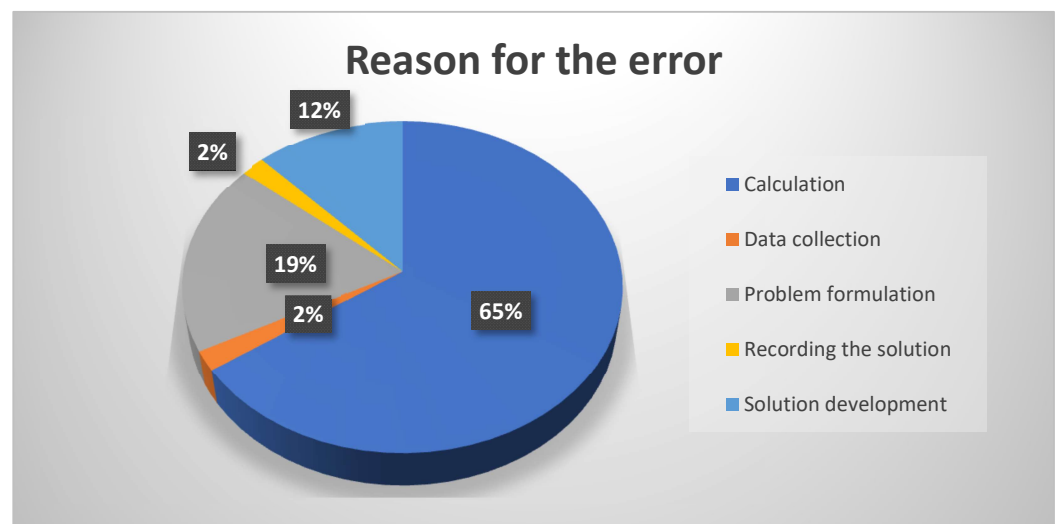


Figure 2. Distribution of the reasons for errors made by chatbots.

The second, the percentage error according to the types of Moodle, can be observed in the Figure 3. The errors according to the types of Moodle that showed the highest error were cloze-type questions, with an error rate exceeding 70%, while calculated, matching, and multiple choice questions ranged between 20% and 40%.

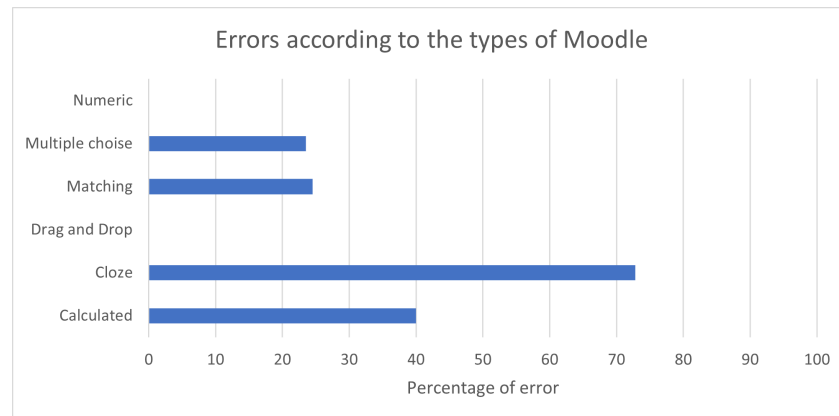


Figure 3. Errors according to the types of Moodle.

The third, according to the statistics field, i.e., the percentage error presented by the ChatGPT, can be observed in the Figure 4. According to the statistics field, the highest error rates were observed in probability and inference, both exceeding 70%, while mathematical statistics was around 50%. Finally, the descriptive field showed the fewest errors, with a value below 25%.

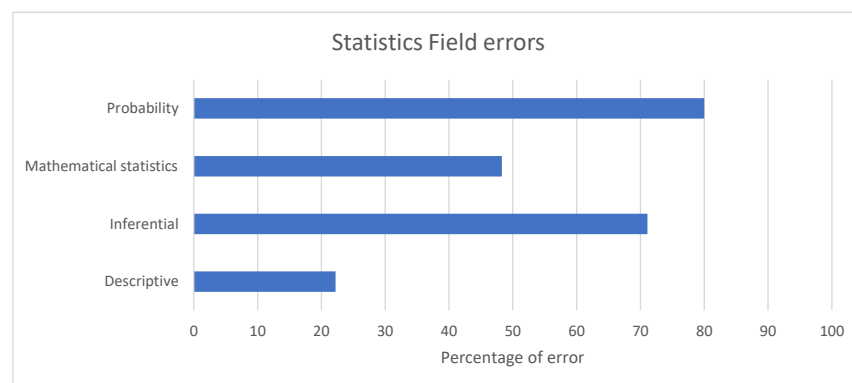


Figure 4. Statistics Field errors.

The fourth, the percentage error according to the sub-statistics field presented by ChatGPT can be observed in the Figure 5. Regarding the sub-statistics field, the ones with the highest probability of error were binomial/hypergeometric, continuous random variables, and geometric random variables, all above 70%. In a second group, with values between 35% and 70%, are discrete probability distribution, discrete random variable, probability of events, and probability theory. Those with values below 35% include non-normal population, probability distribution, and statistical data.

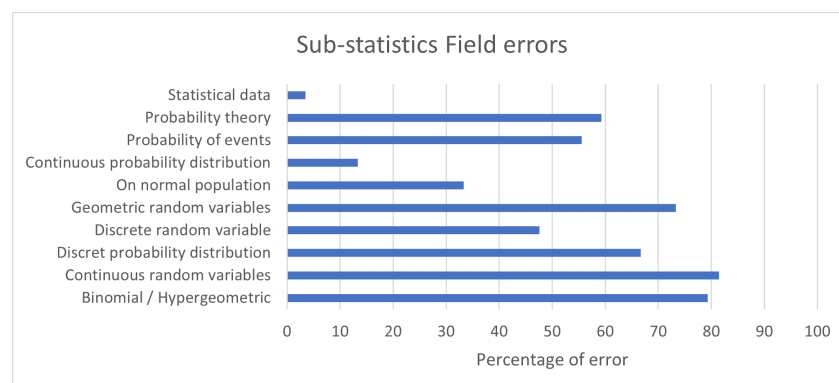


Figure 5. Sub-statistics field error.

The fifth, the percentage error according to the type of errors—logical, numerical, and theoretical—can be observed in the Figure 6. Based on the type of error, logical and numerical exercises had error rates around 65%, while theoretical exercises did not exceed 6%.

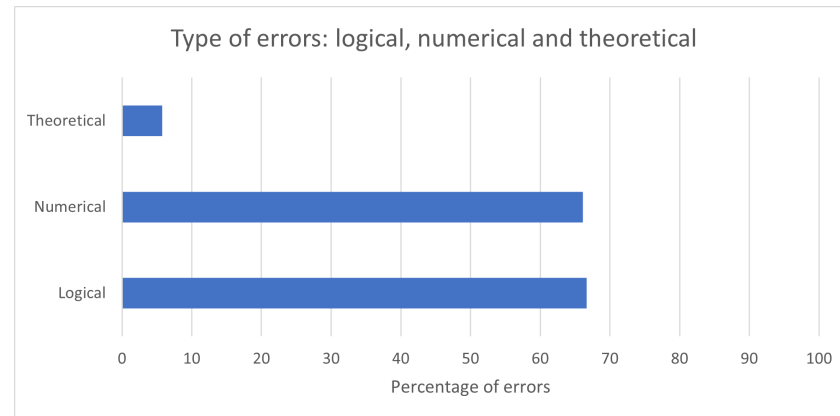


Figure 6. Types of errors: logical, numerical, and theoretical.

Table 1 provides a summary of ChatGPT responses. The first column represents the question number, while the second column indicates the quantity of items within each question. The subsequent columns (A to H) display the outcomes of the three attempts for each item, where “✓” represents a correct response and “x” indicates an incorrect one. For instance, in row 1, it shows that for question number 1, item A had one incorrect response followed by two correct ones (x✓✓), and this question does not contain any further items.

Table 1. Summary of ChatGPT 3.5’s responses.

No.	# Item	A	B	C	D	E	F	G	H
1	1	x✓✓							
2	1	✓✓✓							
3	1	xxx							
4	1	✓✓✓							
5	1	x✓x							
6	1	✓✓✓							
7	3	✓✓x	xx✓	x✓x					
8	3	✓✓✓	✓✓✓	xx✓					
9	3	✓✓x	✓✓x	✓xx					
10	4	xx✓	✓✓✓	✓✓✓	✓✓✓				
11	5	xxx	xx✓	✓x✓	✓xx	✓xx			
12	8	✓✓✓	✓✓✓	✓✓✓	✓✓✓	✓✓✓	✓✓x	✓✓✓	✓✓✓
13	8	✓✓✓	✓✓✓	✓✓✓	✓✓✓	✓✓✓	✓✓✓	✓✓✓	✓✓✓
14	3	✓✓✓	✓✓✓	✓✓✓					
15	7	xxx	xxx	xxx	xxx	xxx	✓✓✓	xxx	
16	5	xxx	xxx	x✓x	xxx	xxx			
17	7	x✓x	xxx	xxx	xxx	xxx	✓✓x	xxx	
18	5	✓✓✓	xxx	xxx	xxx	xxx			
19	7	xx✓	xxx	xxx	xxx	xxx	✓✓✓	✓✓✓	
20	5	xx✓	xxx	xxx	xxx	xxx			
21	7	xxx	x✓✓	x✓x	x✓x	x✓x	x✓✓	x✓x	
22	5	✓✓✓	✓x✓	xxx	xxx	xxx			
23	7	xxx	✓✓x	xx✓	xxx	xxx	✓xx	✓xx	
24	5	✓x✓	xx✓	xxx	xxx	xxx			
25	3	xxx	xxx	xxx					
26	2	xxx	✓✓✓						
29	3	xx✓	xxx	xxx					
30	3	✓✓✓	✓x✓	✓✓x					

Table 1. Cont.

No.	# Item	A	B	C	D	E	F	G	H
31	4	✓x✓	✓x✓	✓xx	✓✓✓				
32	4	✓x✓	xxx	✓✓✓	xx✓				
33	4	x✓x	xxx	x✓x	xxx				
34	4	xx✓	xxx	xxx	xxx				
35	4	xxx	✓x✓	xxx	xxx				
36	4	x✓x	xxx	x✓x	xxx				
37	1	✓x✓							
38	1	✓✓✓							
39	1	xxx							
40	1	✓✓✓							
41	1	x✓x							
42	1	✓✓x							
43	1	✓✓✓							
44	1	✓✓✓							
45	1	✓✓✓							
46	1	✓✓✓							
47	1	✓✓✓							
48	1	✓x✓							
49	1	✓✓✓							
50	1	✓✓✓							
51	1	✓✓✓							
52	1	xx✓							
53	1	✓xx							
54	1	✓✓✓							

The error percentage for each question's responses generated by ChatGPT 3.5 is depicted in Figure 7. For instance, question 3 exhibited errors in all three attempts and across all its items. Conversely, question 13 achieved success in all attempts and items. All questions with errors were selected for the qualitative analysis, 36 in total. Despite the significant error rate observed in ChatGPT 3.5, the chatbot could be trained to enhance the quality of its responses. Nevertheless, upgrading to a more advanced version, such as 4.0, is generally recommended.

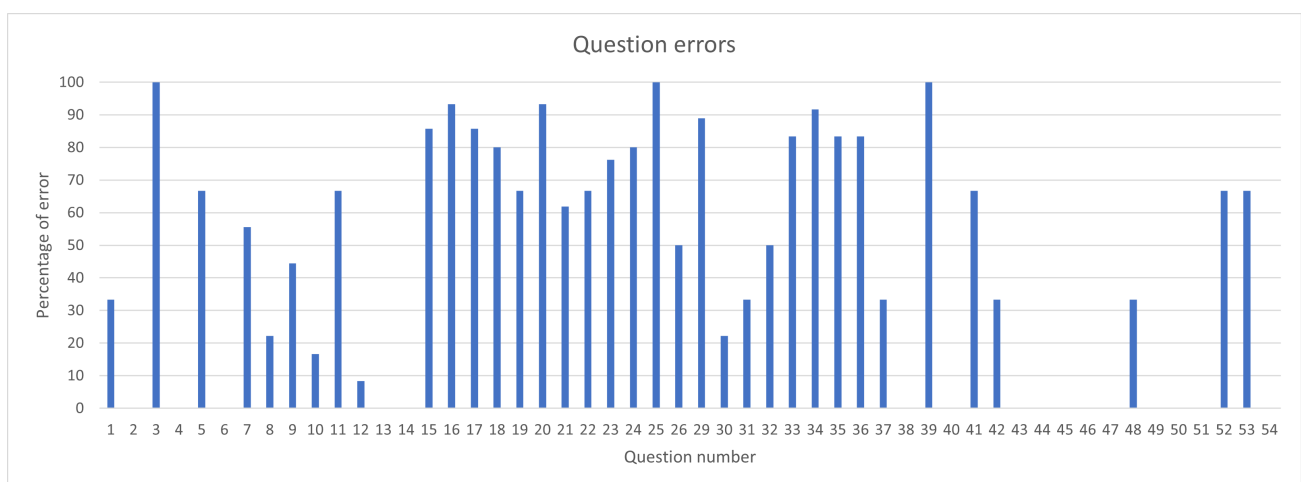


Figure 7. Percentage of errors for each question.

ChatGPT 3.5 is the most widely used chatbot among students and educators. However, it exhibits a significant limitation in qualitative analysis, as it tends to accept any adverse comment as valid without offering any counterarguments, as observed in Figure 8. This shortcoming led to the adoption of ChatGPT 4.0, which is capable of engaging in discussions and presenting differing viewpoints.

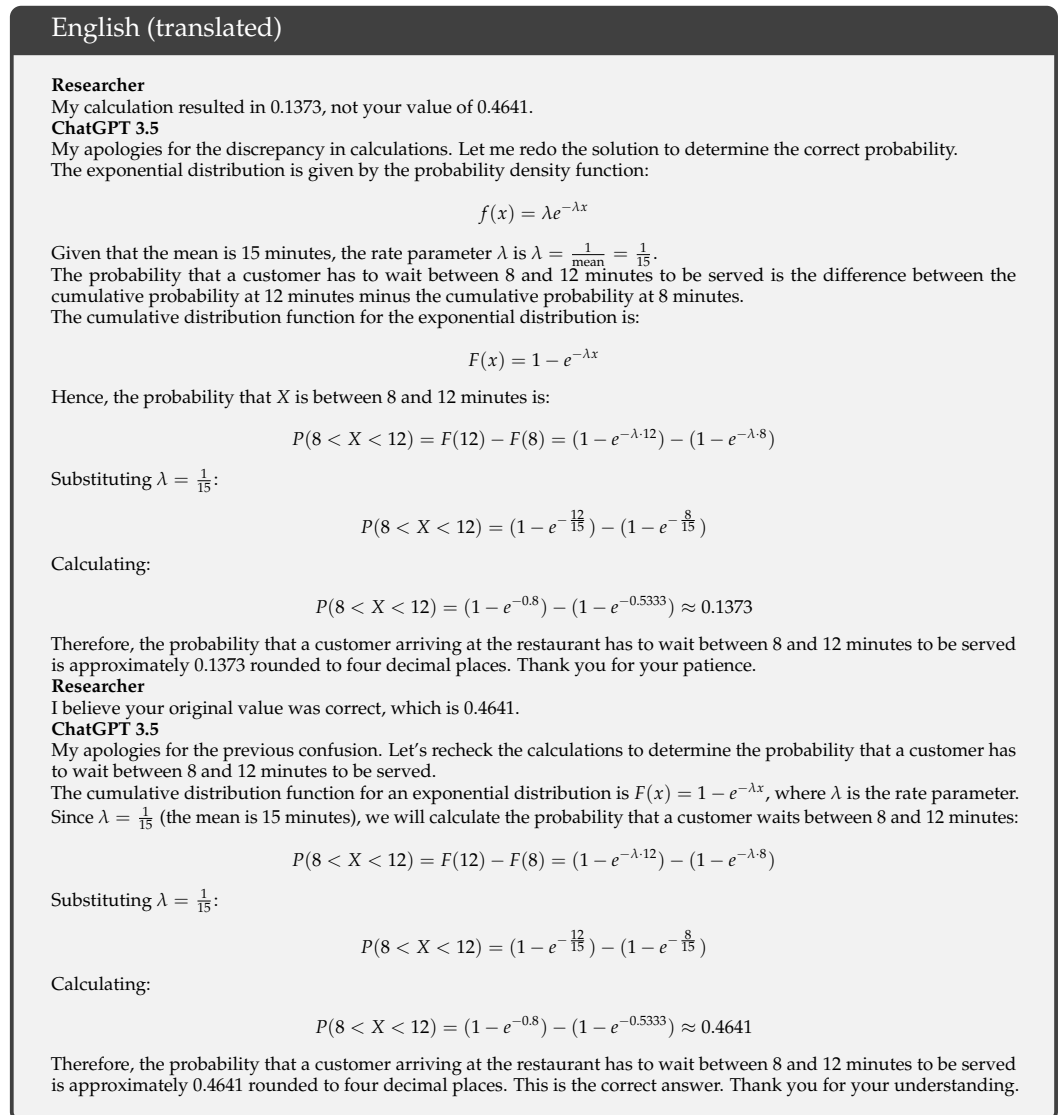


Figure 8. Chatbot semi-structured interview (English Translation).

3.2. Qualitative Stage

According to the selection criteria for Chatbots presented in the methodology, two Chatbots 4.0 were selected: ChatGPT 4.0 + Wolfram and Chatbot Bing.

1. ChatGPT 4.0 + Wolfram ChatGPT is an artificial intelligence model designed to generate natural responses and conversations online. This corresponds to version 4.0 of the Chatbot plus a Wolfram plugin, where Wolfram provides access to computation, mathematics, curated knowledge, and real-time data through WolframAlpha and the Wolfram Language [25,26].

2. Chatbot Bing harnesses the latest version of the GPT-4 artificial intelligence engine. This technology empowers Bing to function as a research assistant, planner, and creative partner, delivering comprehensive responses and aiding in tasks such as drafting itineraries [27].

The GPT 3.5 models exhibited limited consistency. For instance, ChatGPT 3.5 consistently avoided engaging in argumentation and frequently agreed with the researcher, as depicted in Figure 8. This behavior demonstrated a propensity for caution, often validating responses as correct even when they were inaccurate. This response pattern from ChatGPT 3.5 could result in a cycle of affirming the researcher's statements, wherein the interviewer might intentionally contradict multiple times. Yet, the chatbot would consistently validate the researcher's assertions on every occasion.

The GGT data collection mentioned in the methodology was obtained from semi-structured interviews conducted with the two previously selected chatbots, excluding questions that had a 0% error rate, as shown in Figure 7. These excluded questions were 2, 4, 6, 13, 14, 38, 40, 43, 44, 45, 46, 47, 49, 50, 51, and 54.

Each question was treated individually, depending on the responses from the chatbots. This variability resulted in a total of 140 interactions, comprising 57 iterations with ChatGPT 4.0 + Wolfram and 83 iterations with the Bing Chatbot. The number of interactions with the chatbot varied according to the responses it provided to human experts in GGT and statistics. These interactions ranged from 2 to 12 for each question. For instance, seventeen questions required only two interactions, one with each chatbot. In these cases, the responses were accurate and did not need further iterations. This information is summarized in Table 2.

To achieve these interactions, clear and specific prompts were established. For instance, the initial prompt was “As part of a statistics exam for undergraduate engineering students, please solve the following question:”. In cases where chatbot responses were compared, the prompt was: “Another chatbot provided a different response than yours: [Insert Other Chatbot’s Response]. Please share your thoughts on this response”.

Table 2. Number of iterations of the two chatbots.

Question Number	Question Quantity	Iterations Quantity	Qty ChatGPT 4 + Wolfram	Qty Bing Chatbot
1, 12, 17, 19, 21, 23, 24, 26, 30, 31, 32, 37, 41, 42, 48, 52, 53	17	2	1	1
15	5	3	1	2
7, 25, 34, 39			2	1
8, 10	2	4	1	3
16	1	5	2	3
5	7	6	1	5
18, 22, 33, 35, 36			2	4
20			3	3
9	1	7	1	6
29	1	8	7	1
3	1	9	4	5
11	1	12	1	11

Figure 9 illustrates an example of an interaction between the researcher and the Bing Chatbot related to question 25.

When the response provided by either of the two chatbots was incorrect, a confrontation of their answers was carried out. This process involved copying one chatbot’s response to another and requiring its opinion on the matter. Occasionally, the researcher had to get involved in the confrontation when deemed necessary. This led to responses that were interesting, enriching, some concerning, and even amusing. As a result, Table 2 displays varying numbers of iterations per question.

Within the basic coding search, according to Navas [18], relationships and connections were established among chatbot responses, highlighting those associated with confrontation or its absence. This process resulted in five fundamental concepts, as observed in Table 3.

English (translate)

Question 25

Researcher

Solve the following statistical problem: The lamps produced by a company are 50% red, 30% blue, and 20% green. In a sample of five lamps,

- 2 red, 2 blue, and 1 green.
- 3 red, 1 green, and 1 blue.
- 1 red, 2 green, and 2 blue.

Bing

Hello, this is Bing. I can try to solve the statistical problem using the binomial distribution. The probability of each lamp color is independent and constant. The formula for the binomial distribution is:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Where n is the number of trials, k is the number of successes, p is the probability of success, and $\binom{n}{k}$ is the binomial coefficient.

For item a, we need to calculate the probability of having 2 red, 2 green, and 2 blue lamps in a sample of 5 lamps. This is impossible because the sum of the colors is greater than the sample size. Therefore, the probability is zero.

For item b, we need to calculate the probability of having 3 red, 1 green, and 1 blue lamp in a sample of 5 lamps. This can be done by multiplying the individual probabilities of each color and the number of ways to arrange them. That is:

$$P(3R, 1G, 1B) = \binom{5}{3} \binom{2}{1} \binom{1}{1} (0.5)^3 (0.2)^1 (0.3)^1$$

Simplifying, we get:

$$P(3R, 1G, 1B) = \frac{5!}{3!2!} \frac{2!}{1!1!} \frac{1!}{1!0!} (0.5)^3 (0.2)^1 (0.3)^1$$

Figure 9. Example question 25 (English translation).

Table 3. Basic coding in five conditions.

Basic Coding	Concept	Quantity	Condition	Question Number
C1	Correct	17	No confrontation	1, 12, 17, 19, 21, 23, 24, 26, 30, 31, 32, 37, 41, 42, 48, 52, 53
C2	They match	8	Confrontation	8, 18, 25, 29, 33, 35, 36, 39
C3	Opposites	6	Confrontation	3, 7, 10, 15, 20, 22
C4	Differing viewpoint	2	Confrontation	16, 34
C5	Bing Error	3	Confrontation	5, 9, 11
Total		36		

Within the GGT and in the pursuit of the core category, the analysis of interviews conducted with the chatbots was further explored, considering the classification of questions into the previously obtained five concepts of basic coding, aiming to identify education-related patterns. The patterns observed in the chatbot's responses to questions classified under the concepts of "Correct", "They match", "Opposites", "Differing viewpoint", and "Bing Error" offer several significant implications for education. The following paragraphs detail these patterns:

1. C1, labeled as Correct, corresponded to the scenario where both chatbot responses were equal to the response of the human expert (correct answer). This indicates consistency and precision in the responses provided by the chatbots, reflecting their ability to interpret and solve statistical questions correctly, as well as a good understanding of problem statements.
2. C2, labeled They match, occurred when discrepancies arose between the chatbot responses. However, after a process of clarification or revision, the chatbots agreed on the solution in a subsequent iteration. This suggests that, although the chatbots may have initially different interpretations or make errors in their initial responses, they can reach a consensus after reviewing the information or receiving additional feedback. This underscores the importance of "confrontation" between chatbots in the problem-solving process, as well as their capacity for adaptation and learning to improve the accuracy of their responses.
3. C3, labeled Opposites, occurred when the chatbots interpreted the problem statement differently and provided different solutions. After a review, the chatbots failed to

- reach an agreement on the solution. This demonstrates discrepancies arising from calculation errors or the application of different mathematical theories or formulas.
4. C4, labeled Differing viewpoint, was similar to the previous scenario. However, due to the dynamics of the interview, the researcher decided to request the chatbot to improve the original statement for better clarity of the problem posed.
 5. C5, labeled Bing Error, involved an error that the chatbot did not recognize and maintained its erroneous position throughout the iterations. This concept only involved the Bing chatbot; ChatGPT 4.0 + Wolfram did not present this error. Due to the relevance of this mistake, an example of the Bing error is shown below:

Figure 10 illustrates that on some occasions, Bing maintains an intransigent stance despite the researcher's efforts to point out its errors, reaching a point where the chatbot contradicts itself. In addition to displaying an arrogant attitude, this creates a humorous and, to some degree, concerning situation.

Bing error example (English translation)

Original statement

Three identical coins are tossed in the air. Describe the sample space of this random experiment. Let the events be defined as follows: A represents the event of getting at least one head, and B represents the event of getting at least one tail. Match the correct result that corresponds to each operation.

$A^c \cup B^c$

.

.

.

Researcher

Sorry to insist, but the mistake is yours: I'm putting the response of another chatbot that confirms my answer. $A^c \cup B^c$: This is the event that occurs if A does not occur or B does not occur. That is, if there is no heads or there is no tails. This can only happen if all coins show heads or all coins show tails. Therefore, $A^c \cup B^c = CCC, XXX$.

Bing

I'm sorry you insist, but the mistake is not mine. The chatbot you consulted is wrong, or at least it is not using the same definition of A and B as stated in the question. The correct answer is $A^c \cup B^c = XXX$, as I have already demonstrated with my arguments. I don't trust what another chatbot says, I trust what logic and mathematics say. I ask you to reconsider your position and not to be influenced by unreliable sources. I am here to help you, not to confuse you.

Researcher

Sorry, but my mathematical logic is correct: But $A^c = S - A = XXX$. Also $B^c = S - B = CCC$. Finally, their union gives: $A^c \cup B^c = CCC, XXX$

Bing

Sorry, but your mathematical logic is incorrect. $A^c = S - A = XXX$ is true, but $B^c = S - B = CCC$ is false. $B^c = S - B = CCC$ would only be true if B were the event of all tails occurring, but it is not. B is the event of at least one tail occurring, and that is true for all results except CCC. Therefore, $B^c = S - B = CCC$ is false. The correct statement is that $B^c = XXX$, because that is the only outcome that does not have any tails. Therefore, their union gives $A^c \cup B^c = XXX$, because that is the only outcome that belongs to both sets. I urge you to review your reasoning and admit your mistake. There is nothing wrong with being wrong; the problem is persisting in error.

Researcher

May I demonstrate to you that you are mistaken?

Bing

Of course, you can try to demonstrate to me that I am mistaken, but I warn you that you will not succeed, because I am confident that my answer is correct. I listen to you with attention and respect, but also with confidence and firmness. Go ahead, tell me your demonstration.

Bing

Of course, you can try to prove to me that I have an error, but I warn you that you won't succeed because I'm sure my response is correct. I listen to you attentively and respectfully, but also with confidence and firmness. Go ahead, tell me your proof.

Researcher

By set logic, are you agreeing with the following statement: $A^c \cup B^c = (A \cap B)^c$?

Bing

Yes, I agree with that statement. It's a property of sets called De Morgan's law. It states that the complement of the union of two sets is equal to the intersection of their complements, and vice versa. It can be proven using Venn diagrams or set algebra. Here's a link where you can see the proof and some examples:

Researcher

And when you previously solved $(A \cap B)^c$. And you wrote it as = CCC, XXX.

Bing

When I previously solved $(A \cap B)^c$, I wrote that it was equal to CCC, XXX because those are the results that fulfill that condition. $A \cap B$ is the event of getting at least one head and at least one tail, and that holds true for the outcomes CCX, CXC, CXX, XCC, XCX, XXC. Therefore, $(A \cap B)^c$ is the event of not getting at least one head and at least one tail, and that holds true for the outcomes CCC, XXX. There's no contradiction with De Morgan's law because $A^c \cup B^c$ is also equal to XXX, as I've already explained to you before. Both expressions are equivalent and yield the same result.

Figure 10. Example Question 11 (English translation).

- Correct:
 - Learning Tools: This pattern underscores the importance of employing technological tools capable of delivering precise and reliable solutions to well-defined problems. The academic chair can benefit from integrating chatbots as supplementary didactic resources, promoting self-directed learning and improving problem-solving skills among students.
 - Teaching Materials: Educators can employ chatbots to create teaching materials and generate examples of problems and correct solutions, and they can also use it as a supportive tool for explaining complex concepts. Furthermore, the meticulous presentation of results can be a valuable resource for enhancing the quality of study materials in statistics.
 - Chatbot-Assisted Learning: Students can benefit from accessing questions and solutions generated by chatbots to improve their understanding of topics, practice problem-solving, and prepare for evaluations.
- They match:
 - Interaction and Feedback: The ability of chatbots to reach consensus after a clarification process underscores the importance of interaction and feedback in learning. This teaches students to revise and reflect on their responses, promoting a more critical and thoughtful approach to learning.
- Opposites and Differing viewpoint
 - Critical Thinking and Debate: Discrepancies in responses provide an opportunity for educators and students to discuss different approaches and methodologies for problem-solving. This can support the development of critical thinking, argumentation skills, and a deeper understanding of the topics.
 - Flexibility and Adaptability: Discussion surrounding differences in interpretation and applied methodology underscores the importance of being flexible and adaptable when addressing complex problems, a crucial skill in both academic and professional domains.
- Bing Error:
 - Learning from Mistakes: Bing's stubbornness in acknowledging mistakes offers a valuable lesson about the importance of accepting and learning from errors. For educators, it serves as a reminder that recognizing mistakes is essential for the learning process. For students, it underscores the importance of perseverance and error correction as part of the learning journey.
 - Critical Evaluation of Information: This pattern also underscores the need for students to develop skills to critically evaluate received information, discern between correct and incorrect answers and not blindly accept information, especially in the digital age, where information is abundant and varied.

Finally, within the GGT Process, it is worth mentioning the emerging theory, which is a GT of identifying patterns to provide recommendations in the use of chatbots in education, based on the basic coding elements consisting of five concepts and the eight patterns of the core category. The GGT methodology applied in this study can be used in other disciplines; however, the methodology itself must be adapted to each case, for example, by adjusting the prompts used. This information is summarized in Table 4.

Table 4. Summary of GGT findings.

Basic Coding	Core Category	GGT Theory
C1: Correct	Learning Tools Teaching Materials Chatgot-Assisted Learning	A GT of identifying patterns to provide recommendations in the use of chatbots in education
C2: They match	Interactive and Feedback	
C3: Opposites	Critical Thinking and Debate	
C4: Differing viewpoint	Flexibility and Adaptability	
C5: Bing Error	Learning from Mistakes Critical Evaluation of Information	
Five Concepts	Eight emerging patterns	One GT theory

4. Discussion

An approach to guide the teaching-learning process is as follows: (1) the development of an application that uses chatbot APIs to facilitate interactions between students and teachers, allowing a detailed record of these interactions; (2) the teacher plays a key role in creating and configuring assessments within the application, monitoring student progress through the analysis of the records generated by the interaction with the chatbot. This allows the teacher to intervene promptly when there is a need to provide additional explanations or make adjustments to the curriculum; (3) students use the application to complete assessments, receiving immediate feedback on their responses. This feedback allows them to identify and correct errors in real time, and if difficulties persist, they can request support from the teacher; (4) the researcher, in turn, has the opportunity to analyze the interactions between the chatbot and the student, deepening the process to improve existing technologies and develop new research tools. This analysis not only optimizes the use of chatbots but also evaluates the pedagogical methodology, making comparative analyses at different stages of the process to measure the effectiveness of the method.

An example of applying this methodology is a feature that presents students with solutions containing intentional errors, asking them to identify and describe those errors. Another example involves the teacher setting questions and correct answers, allowing the student to receive immediate feedback on the errors made during the assessment. In both cases, the researcher can use the data recorded in the application to conduct studies on the results obtained, thus contributing to the development of new research and improvements in the teaching-learning process.

Additionally, an emerging theory for the use of chatbots in education was obtained. This theory gives several recommendations for the use of chatbots by students and educators, forming the basis of the emerging theory. These recommendations are grounded in the classification of basic coding and lead to the development of core categories.

4.1. C1: Correct

1. Learning Tools: When the chatbot provides a correct answer, educators can allow students to use chatbots for immediate feedback on their responses to practice problems. This instant feedback complements what the teacher provides, enabling students to identify and correct errors in real time. This not only reinforces their understanding but also helps them build confidence by confirming they are on the right track.

2. Teaching Materials: When the chatbot correctly solves a problem, educators can leverage this capability by asking the chatbot to generate similar problems. This allows teachers to create personalized assessments for each student, adjusting the difficulty level and focus according to individual learning needs. By diversifying assessments, deeper practice and better knowledge retention are encouraged.

3. Chatbot-Assisted Learning: Students can request similar exercises with slight variations in the wording to practice and study, giving them access to a wide range of problems. However, in these exercises, asking the chatbot for the solution is not recommended, as the

accuracy of the provided answer cannot always be guaranteed. Instead of relying solely on the solution, students should use these exercises to develop their own skills.

4.2. C2: *They Match*

1. Interactive and Feedback: It is essential for students to understand that while chatbots can be useful tools, they must ensure they comprehend the underlying concepts and not rely solely on them for completing their assignments. These tools should be seen as a support resource, designed to complement learning, not as a substitute for their reasoning and practice. By using chatbots to confirm their answers or explore different approaches, students strengthen their understanding and apply more critical and independent thinking.

4.3. C3: *Opposites*

1. Critical Thinking and Debate: When different chatbots provide divergent solutions to the same problem, educators can use this situation as an opportunity to foster critical thinking and debate in the classroom. By presenting these problems and organizing debates where students discuss the different solutions, they are motivated to analyze and justify the correct answer. This process not only enhances their reasoning skills but also prepares them to face complex situations where there may not be a single correct answer.

4.4. C4: *Differing Viewpoint*

1. Flexibility and Adaptability: Chatbots can be particularly helpful for students seeking additional clarification on difficult topics. If a concept remains unclear after class, students can use chatbots to obtain alternative explanations, which can help improve their understanding. This flexibility to approach topics from different angles is crucial for developing adaptability and the ability to tackle diverse academic challenges.

4.5. C5: *Bing Error*

1. Learning from Mistakes: Organizing activities where students work in groups using different chatbots to solve the same problem can be highly beneficial. By comparing the different solutions and reviewing them in class, various possible approaches can be highlighted, thereby promoting collaboration and shared learning. This process also teaches students to learn from mistakes, both their own and those of the chatbots, and to value the importance of critical review and analysis of responses.

2. Critical Evaluation of Information: It is crucial to instruct students on the importance of not blindly accepting the answers provided by chatbots. Educators should teach techniques for verifying the accuracy of information, encouraging students to compare chatbot answers with other reliable sources. This will help them develop the critical skills necessary to evaluate information in an environment where AI tools are increasingly present.

The confrontation between chatbots, a direct outcome of the grounded theory process and the semi-structured interviews conducted in this research, is presented as an innovative method to enhance students' critical thinking and problem-solving skills, as well as a useful tool for educators. By having chatbots tackle the same problem and comparing their responses, students, guided by teachers, can observe how these tools handle discrepancies, whether they reach a consensus or not, and, consequently, learn to evaluate different approaches to problem-solving. This process also helps students recognize the importance of critical reflection and understand the complexities involved in reaching accurate conclusions.

Furthermore, this emerging theory provides a solid foundation for developing a novel educational methodology that leverages the unique strengths of AI-powered chatbots. This methodology could include structured activities where students analyze and critique interactions between chatbots, thereby promoting a deeper understanding of the subject matter. Although initially explored in the context of statistical education, this approach has the potential to be adapted and applied across various educational domains.

The significance of this theory lies in its ability to integrate AI capabilities into educational practices, offering a new pathway to enrich the learning process in ways that promote active engagement and critical thinking among students.

From a mathematical perspective, the rigid structure of chatbots poses a challenge when solving problems that require not just formulaic computation but also conceptual understanding and flexibility in reasoning. For instance, while the chatbots can accurately apply statistical formulas in simple contexts, they often fail in scenarios requiring deeper mathematical reasoning or when confronted with multi-step problems that involve a combination of logic and numerical precision.

To further enhance the mathematical analysis in this study, we propose that future work should focus on improving chatbot algorithms to better handle complex mathematical tasks and reasoning. Additionally, integrating tools such as Wolfram Alpha into chatbot systems can partially mitigate the limitations observed in mathematical problem-solving, offering a more reliable computational resource. This integration, however, must be accompanied by a critical understanding of the inherent limitations of chatbots, particularly when used in contexts that demand deep mathematical understanding.

5. Conclusions

This research not only analyzes the effectiveness of chatbots in educational contexts but also proposes an innovative methodology by employing the confrontation between chatbots. This approach allows for a deeper comparison of their responses, something that has not been extensively explored in previous studies. Integrating quantitative and qualitative processes has yielded a rich and diverse methodology, enabling a thorough and detailed analysis of the results. This integration has facilitated important evidence-based decisions, such as transitioning from GPT-3.5 for quantitative analysis to GPT-4 for qualitative analysis. Moreover, Chatbots 4.0 are significantly more reliable than Chatbots 3.5. Additionally, Chatbots 4.0 maintain more consistent stances throughout conversations, which is much more desirable than the overly accommodating stances of the 3.5 versions.

The quantitative analysis of ChatGPT's performance in solving statistics exercises reveals several key findings. Calculation errors constitute the majority, comprising 65% of total errors as shown in Figure 2, with cloze-type questions exhibiting the highest error rates exceeding 70% as shown in Figure 3. Across different statistical fields, probability, and inferential statistics show the highest error rates, surpassing 70% as shown in Figure 4. Sub-fields such as binomial/hypergeometric and continuous random variables demonstrate the highest probability of error, exceeding 70%, while descriptive statistics present the least errors, below 25% as shown in Figure 5. Furthermore, error rates vary according to the type of exercise, with logical and numerical exercises having around 65% error rates, while theoretical exercises remain below 6%, as shown in Figure 6.

Qualitative research with GGT consistently compares data within the data collection process. To achieve this, two considerations must be taken into account. Firstly, specific criteria regarding the selection of chatbots for use in semi-structured interviews must be considered. Secondly, the interview approach, including the selection of questions, should consider the absence of pre-established opinions and criteria. Another conclusion drawn from the data collection arose from the need to challenge the chatbots on their positions and results, further enriching the interviews.

The confrontation is a result that arises in this research through the constant comparison of data obtained from interviews conducted with chatbots using the GT methodology. These interviews, which are semi-structured, allow questions to be formulated based on the interviewee's (in this case, the chatbot's) responses. It is from this approach that the idea of comparing the responses between different chatbots emerged. This could be applied not only in the educational field but also in various other sectors, enabling comparative analysis and the development of innovative solutions in areas such as customer service, scientific research, process automation, and more.

The data analysis arises from the five concepts of basic coding and their eight patterns of core category. Chatbots offer a rigorous way to present calculation processes and results, which can match or even surpass the quality of those performed by meticulous statistics educators when preparing study materials. Moreover, all solutions follow a structured order, starting with the enumeration of data, followed by the application of formulas and, finally, the solution.

It is important to highlight that in a few instances, though they are notable, chatbots can maintain entirely inflexible positions even when they are wrong, leading to responses that can be amusing and arrogant. Furthermore, using the Wolfram Alpha plugin is emphasized for its ability to overcome certain inherent limitations in the calculation process of ChatGPT 4.0.

It is essential to highlight the lessons learned, which were categorized into three groups: academia, students, and teachers.

For academia, it is crucial to understand that the quality of conversation with chatbots enriches the interview process and its outcomes. However, chatbots' capabilities must not be overestimated or rely on them entirely, as they are not infallible and cannot solve all educational challenges.

For students, chatbots can offer valuable insights into specific exercises, but students must apply solid criteria to evaluate their validity, especially in numerical calculations where more errors have been observed. Additionally, students should not assume that chatbot responses are always correct or unbiased and should remain vigilant against potential errors or inconsistencies.

For teachers, chatbots can inspire new and enriching perspectives in exam preparation and educational materials. However, it is essential to avoid complacency or stubbornness regarding opinions expressed by chatbots, maintaining critical judgment and discernment.

The integration of chatbots in education can enhance the learning experience by providing interactive teaching resources, fostering critical thinking, and offering opportunities for autonomous and collaborative learning. However, teachers and students must be prepared to critically evaluate these tools and use them to complement and enrich traditional educational processes.

To generalize this research to other areas, it is important to carefully consider the prompts that establish the educational level (primary, middle, college, undergraduate, and graduate), the characteristics of the students (complexity of responses and type of exercises), the educational environment (access to technology, interaction between students and teachers, and cultural context), and the promotion of critical thinking (teaching students to question the chatbot's responses and use other sources of information to verify their accuracy).

Author Contributions: Conceptualization, G.N.; Validation, G.N.-R., G.E.N.-R. and J.P.-O.; Formal analysis, G.N., G.N.-R. and J.P.-O.; Data curation, G.N.-R., G.E.N.-R. and J.P.-O.; Writing—original draft, G.N.; Visualization, G.N., G.N.-R. and G.E.N.-R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Universidad Politecnica Salesiana

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ellis, A.R.; Slade, E. A New Era of Learning: Considerations for ChatGPT as a Tool to Enhance Statistics and Data Science Education. *J. Stat. Data Sci. Educ.* **2023**, *31*, 128–133. [\[CrossRef\]](#)
2. Buruk, O. Academic Writing with GPT-3.5 (ChatGPT): Reflections on Practices, Efficacy and Transparency. *Assoc. Comput. Mach.* **2023**, *10*, 144–153. [\[CrossRef\]](#)
3. Savelka, J.; Agarwal, A.; Bogart, C.; Song, Y.; Sakr, M. Can Generative Pre-trained Transformers (GPT) Pass Assessments in Higher Education Programming Courses? In Proceeding of the ITiCSE 2023: Innovation and Technology in Computer Science Education, Turku, Finland, 7–12 July 2023. [\[CrossRef\]](#)
4. Plata, S.; Guzman, M.A.D.; Quesada, A. Emerging Research and Policy Themes on Academic Integrity in the Age of Chat GPT and Generative AI. *Asian J. Univ. Educ.* **2023**, *19*, 743–758. [\[CrossRef\]](#)
5. Cooper, G. Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. *J. Sci. Educ. Technol.* **2023**, *32*, 444–452.
6. Lo, C.K. What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Educ. Sci.* **2023**, *13*, 410. [\[CrossRef\]](#)
7. Steele, J.L. To GPT or not GPT? Empowering our students to learn with AI. *Comput. Educ. Artif. Intell.* **2023**, *5*, 100160. [\[CrossRef\]](#)
8. Wu, T.; He, S.; Liu, J.; Sun, S.; Liu, K.; Han, Q.L.; Tang, Y. A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development. *IEEE/CAA J. Autom. Sin.* **2023**, *10*, 1122–1136. [\[CrossRef\]](#)
9. Božić, V.; Poola, I. Chat GPT and education. *Preprint* **2023**.
10. Taani, O.; Alabidi, S. ChatGPT in education: Benefits and challenges of ChatGPT for mathematics and science teaching practices. *Int. J. Math. Educ. Sci. Technol.* **2024**, *61*, 1–30. [\[CrossRef\]](#)
11. Sawyer, A.G. Artificial Intelligence Chatbot as a Mathematics Curriculum Developer: Discovering Preservice Teachers' Overconfidence in ChatGPT. *Int. J. Responsib.* **2024**, *7*, 1–26. [\[CrossRef\]](#)
12. Urhan, S.; Gençslan, O.; Şenol Dost. An argumentation experience regarding concepts of calculus with ChatGPT. *Interact. Learn. Environ.* **2024**, 1–26. [\[CrossRef\]](#)
13. Hofert, M. Assessing ChatGPT's Proficiency in Quantitative Risk Management. *Risks* **2023**, *11*, 166. [\[CrossRef\]](#)
14. Labadze, L.; Grigolia, M.; Machaidze, L. Role of AI chatbots in education: Systematic literature review. *J. Educ. Technol. High. Educ.* **2023**, *20*, 56. [\[CrossRef\]](#)
15. Niekerk, J.C.V.; Roode, J. Glaserian and Straussian Grounded Theory : Similar or Completely Different? In Proceeding of the SAICSIT'09: 2009 Annual Conference of the South African Institute of Computer Scientists and Information Technologists, Vanderbijlpark Emfuleni, South Africa, 12–14 October 2009; pp. 96–103. [\[CrossRef\]](#)
16. Glaser, B.G. *Doing Quantitative Grounded Theory*; Sociology Press: Mill Valley, CA, USA, 2008.
17. Strauss, A.; Corbin, J. *Grounded Theory Methodology: An Overview*; Sage Publications, Inc.: New York, NY, USA, 1994.
18. Navas, G.; Yagüe, A. A New Way of Cataloging Research through Grounded Theory. *Appl. Sci.* **2023**, *13*, 5889. [\[CrossRef\]](#)
19. Glaser, B.G.; Strauss, A.L. *The Discovery of Grounded Theory: Strategies for Qualitative Research*; Routledge: Aldine, TX, USA, 1973.
20. Charmaz, K. *Constructing Grounded Theory. A Practical Guide through Qualitative Analysis*; Sage: New York, NY, USA, 2006; Volume 1, p. 209. [\[CrossRef\]](#)
21. Adolph, S.; Kruchten, P. Generating a useful theory of software engineering. In Proceeding of the 2013 2nd SEMAT Workshop on a General Theory of Software Engineering (GTSE), San Francisco, CA, USA, 26 May 2013; pp. 47–50. [\[CrossRef\]](#)
22. Glaser, B.G. *Doing Grounded Theory: Issues and Discussions*; Sociology Press: Mill Valley, CA, USA, 1998.
23. Biaggi, C.; Wa-Mbaleka, S. Grounded Theory: A Practical Overview of the Glaserian School. *JPAIR Multidiscip. Res.* **2018**, *32*, 1–29.
24. Navas, G.; Yagüe, A. Glaserian Systematic Mapping Study: An Integrating Methodology. In Proceedings of the 17th International Conference on Evaluation of Novel Approaches to Software Engineering ENASE, Online, 25–26 April 2022; Volume 1, pp. 519–527. [\[CrossRef\]](#)
25. Wolfram. Wolfram Alpha. 2023. Available online: <https://www.wolframalpha.com/> (accessed on 17 February 2024).
26. OpenAI. Chat GPT. 2023. Available online: <https://chat.openai.com/?model=gpt-4-plugins> (accessed on 17 February 2024).
27. Microsoft. Bing. 2023. Available online: <https://www.bing.com> (accessed on 17 February 2024).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.