Preprints.org

Design and Evaluation of a Research-Guided Chatbot for Novice Researchers

V. Sithira Vadivel [*] , Shaleeza Sohail , Xin Gu

Article

Design and Evaluation of a Research-Guided Chatbot for Novice Researchers

V. Sithira Vadivel [1],*, Shaleeza Sohail [1] and Xin Gu [2]

[1] University of Newcastle

[2] Kings Own Institute 3

* Correspondence: vsithira.vadivel@newcastle.edu.au; Tel.: +65 96890302

Abstract: (1) Background: Evaluating popular educational chatbots (EC) shows that existing EC impacts users' ability to synthesize and take ownership of their work, potentially leading to misinformation. This study aims to design and evaluate a research-guided chatbot that improves research synthesis. (2) Method: One hundred sixty-five graduate students evaluated the chatbot. The qualitative data were analysed using natural language processing for sentiment analysis and emotion detection. (3) Results: The findings show that "interesting" and "enjoyable" were frequent words used in user feedback. Sentiment analysis findings show that 47% suggested users are satisfied with the chatbot's recommendation potential; 46.3% found the guided chatbot to be the most effective in simplifying research methodology concepts; 48.4% identified the chatbot as the most effective tool for conceptualizing research questions and hypotheses and excellent in fostering ownership of their ideas; 85.3% rated the chatbot as fostering creativity; and 81.0% rated the chatbot positively for its ability to minimize misinformation. Two themes were identified in the thematic analysis: attitudes and improvements to future research-guided chatbots, emphasizing simplicity. (4) Conclusions: The strong preference for the guided chatbot underlines the potential for future improvements in EC with the flexibility to gain ownership of work with a guided approach.

Keywords: Artificial Intelligence; AI for education; guided chatbot foundation model; ownership of ideas; minimize misinformation

1. Introduction

Many existing educational chatbots and tools focus on delivering content and assessing knowledge. There is a significant opportunity for chatbots to foster creativity, critical thinking, and ownership of learning. A chatbot designed to promote users' ideas could potentially create a more engaging and empowering experience.

The idea of this project is to design a guided chatbot for novice researchers that supports them in starting their research planning while allowing them to take control and ownership of all the ideas expressed in their research. The chatbot guides them by offering examples for them to consider the best options for their research plan. The designed and developed chatbot was evaluated by 98 tertiary students regarding their experience and sentiments about owning their ideas and the guidance provided by the chatbot while conducting research.

2. Materials and Methods

A chatbot, also known as a conversational agent, is computer software capable of taking natural language as input, processing it, and providing an output in real time. Several educational chatbots, such as StudyBuddy, Ada, and Socratic, assist students in academic planning by providing feedback and individualizing learning support [1–4] Martinez et al., 2023; Kabiljo et al., 2020; Konecki et al., 2023; Duelen et al., 2024). Other chatbots like ZenoChat, Duolingo, Bard, ChatGPT, and Quillbot enhance writing skills to provide high-quality documents.

GPT, for example, launched by OpenAI in 2018, is an architecture that employs large pre-trained neural networks known as Large Language Models (LLMs). These models generate human-like text and grasp input data intricacies, facilitating coherent and contextually relevant responses. GenAI excels in various applications, including machine translation, question answering (QA), Named Entity Recognition (NER), data cleansing and preprocessing, text summarization, and sentiment analysis [5]. In education, ChatGPT and LLMs can enhance content generation, personalize learning experiences, and boost student engagement, provided that fact-checking, bias mitigation, and guidelines are implemented [6]. However, concerns about academic integrity and plagiarism are raised by [7,8], highlighting the need to foster critical thinking, problem-solving, and communication skills.

Research evaluating popular educational chatbots shows that existing educational chatbots (EC) impact users' ability to synthesize and take ownership of their work, potentially leading to misinformation [9]. It is important to ground hypotheses in existing evidence and structured methodology. Good hypotheses should challenge existing ideas thoughtfully rather than destructively, ensuring they are backed by prior research and can be tested effectively. Proper planning and statistical power are crucial for validating these hypotheses in scientific research [10]. Novice researchers, such as student researchers, may struggle in the research planning process, and guidance is required in the planning phase as they navigate the research process. The use of Gen AI among researchers has been extensive in recent years [11]. The use of Gen AI in research could lead to blind research, which has become common among novice and student researchers. Students introduced to research planning must be guided. However, the design of current Gen AI tools, such as ChatGPT, could amplify misinformation and disinformation. This not only increases plagiarism issues but also fosters a misguided understanding of the different planning phases of research. Novice researchers must receive guidance and examples to inspire ideas and suggest the planning process. A systematic approach to guiding student researchers was the aim of this project, with an emphasis on ownership of their research work and stimulating their thought process.

The problem of the extensive use of ChatGPT or any such Gen AI could tarnish the institution's reputation if it is not handled properly. Academics will not tolerate the possibility of using such tools or an increase in academic misconduct.

In this study, we examine the evolution of these tools in the academic field. We designed a chatbot that guides students in planning their research and taking ownership of their work. We evaluate the prototype with sentiment analysis to determine the users' emotional state when guided by a chatbot that supports researchers' thought processes.

In this paper, we discuss the evaluation process of the research-guided chatbot. Students are asked to plan their research on a topic related to teaching and learning using the research-guided chatbot. The chatbot suggests examples to stimulate ideas and proposes sample hypotheses and research questions, identifying the research type as qualitative or quantitative and the sample size. Students are guided in the next research phase by receiving education on the research methodology process. The guided chatbot proposes examples for formulating hypotheses and research questions. It suggests a sample testable hypothesis, explains how it will be tested, and encourages students to write their hypotheses similarly. The students were given the same tasks to complete the activities for their chosen research topic. The tasks will be repeated if the chatbot finds them incomplete and encourages the students with more recommendations until an optimal hypothesis statement is written. This way, it encourages a thorough understanding of the subject matter by the student rather than solely relying on the output generated by the GenAI. The activities and tasks were checked for completeness, and the type of questions submitted by the students were validated by the guided chatbot. A report will be generated to summarize the completeness of the research phases. Respondents' experiences using the guided chatbot were assessed based on ownership and ease of use.

The tasks of the guided chatbot activity include the following ten steps: 1) Choosing a research topic; 2) Identifying the problem area; 3) Developing research questions; 4) Defining independent

and dependent variables; 5) Determining actions for the identified variables; 6) Formulating hypotheses; 7) Determining the type of experiment (qualitative or quantitative); 8) Deciding on the sampling size; 9) Selecting a data collection technique (e.g., survey); 10) Validating the research methodology and outcome.

The following image shows the screen capture of the research-guided chatbot explaining the tasks in the activity to plan a research methodology.
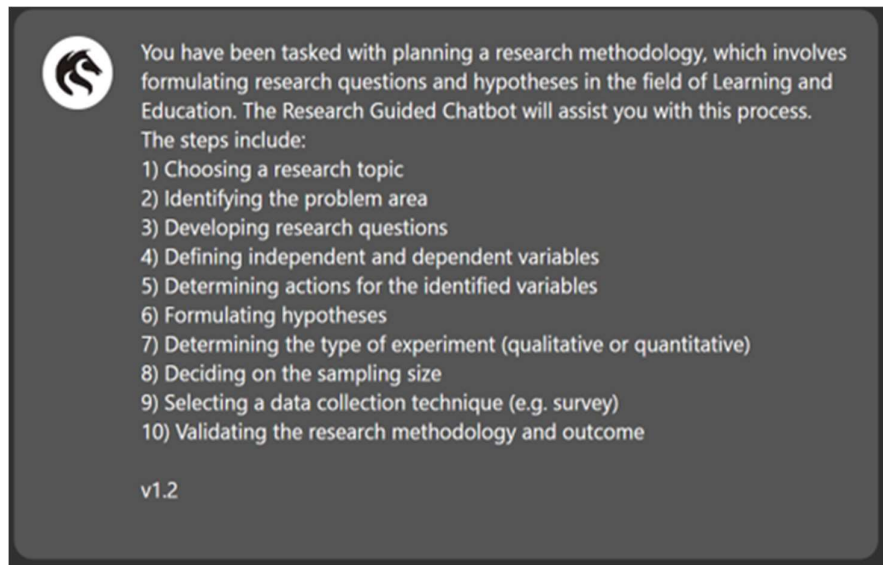

Figure 1. The tasks listed by the Research-guided chatbot.

The guided options with examples for forming a research question are shown in the screenshot below. The chatbot suggests keywords related to the chosen topic and provides examples to guide users in formulating their research questions.
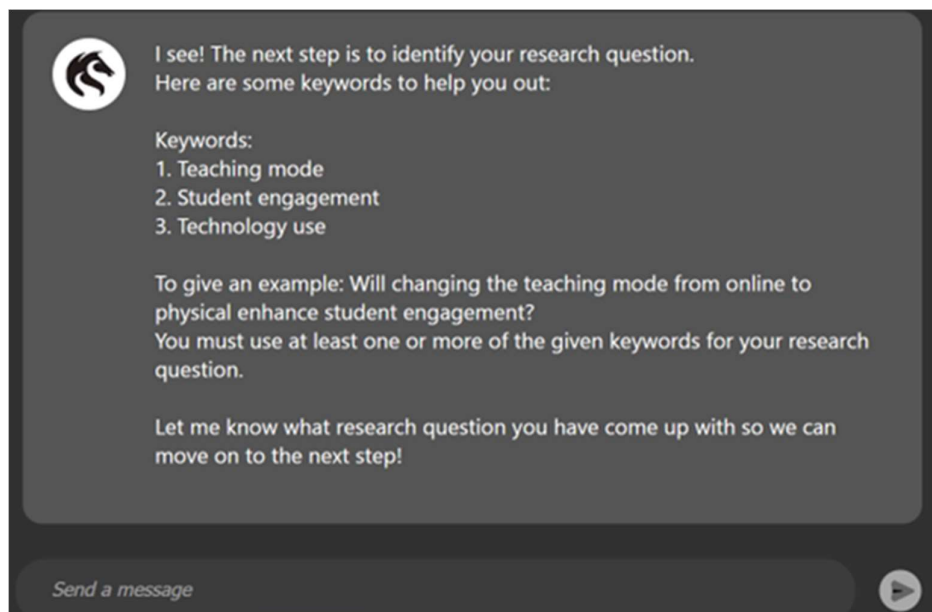

Figure 2. The research-guided chatbot keyword-guided options for research questions.

The chatbot guides users in formulating their hypotheses using independent and dependent variables. The guided options, along with examples for forming a hypothesis and an explanation of independent and dependent variables, are shown in the following screenshot.
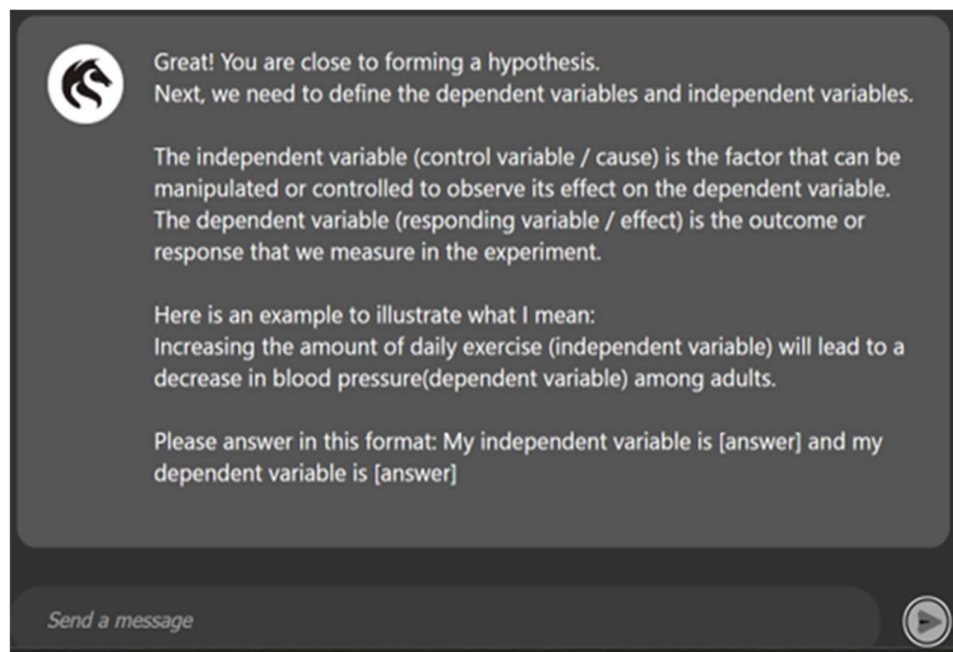

Great! You are close to forming a hypothesis.
Next, we need to define the dependent variables and independent variables.

The independent variable (control variable / cause) is the factor that can be manipulated or controlled to observe its effect on the dependent variable. The dependent variable (responding variable / effect) is the outcome or response that we measure in the experiment.

Here is an example to illustrate what I mean:
Increasing the amount of daily exercise (independent variable) will lead to a decrease in blood pressure(dependent variable) among adults.

Please answer in this format: My independent variable is [answer] and my dependent variable is [answer]

Send a message

Figure 3. The research-guided chatbot guided options with independent and dependent variables for formulating the hypothesis.

3. Results and Discussion

A pilot test was conducted with 20 undergraduate students from the Bachelor of Information Technology at the University. The pilot test encountered several issues as the undergraduate students struggled to comprehend certain research planning phases, such as formulating the hypothesis independently. Given this limitation, the criteria for selecting respondents were based on their educational background. Only graduates were eligible to participate in the study. Approximately 250 graduate respondents were invited to partake in the actual evaluation, with only 165 accepting the invitation. The 165 respondents are from Singapore and Australia, and they must be graduates or postgraduates. This selection criterion is mandatory as the prototype activity focuses on research planning. The first phase involves respondents evaluating the research-guided chatbot, while the second phase requires them to complete a survey regarding their experience with the guided chatbot. The estimated time to complete the research-guided activity is 20 to 25 minutes. Generally, all respondents completed the activity within 17 to 19 minutes.

The survey questions were designed with four sections. Section A of the survey is about respondents' demographics; section B is to obtain information on preferences for effective concepts; section C is to find experience with methods and ownership of generated output; and section D covers qualitative questions about guided chatbot experiences. The objective of the survey is to obtain the respondents' sentiments using the guided chatbot compared to a general design of GenAI.

User feedback provides a significant source of information to understand user experience, which will be used to improve the chatbot's design. Two types of feedback were provided by the users:

- General feedback was given by the users after using the chatbot.
- The questionnaire, as discussed in section A, was completed by the users.

In total, 78 relevant text responses were collected from users about their experience using the chatbot. To quantify this qualitative data, two types of natural language processing (NLP) analyses

are performed: sentiment analysis and emotion detection. These methods were utilized to analyze user sentiment trends and gain insights into the emotional responses provided by users.

The word cloud shown in Figure 4 illustrates the significance of the words in user comments in terms of word frequency to visualize themes that dominate user feedback. This graphical representation of word frequency, indicated by word size, shows that "interesting" and "enjoyable" were the most frequently used words in user feedback.

Figure 4. User comments on a word cloud.

The Valence Aware Dictionary and Sentiment Reasoner (VADER) [12] algorithm was employed for sentiment analysis to categorize user comments as positive, negative, or neutral. VADER is suitable for analyzing short text (reviews, comments) as it is attuned to the use of special characters and capital letters in comments and reviews. A compound score is calculated to identify sentiment polarity, where thresholds of +0.05 and -0.05 differentiate the three sentiments. Figure 5 shows the sentiment for 78 user comments, where around 60 comments are positive, 10 comments are negative, and 8 comments are neutral. In total, the positive or neutral comments are more than 88% (70 out of 78).
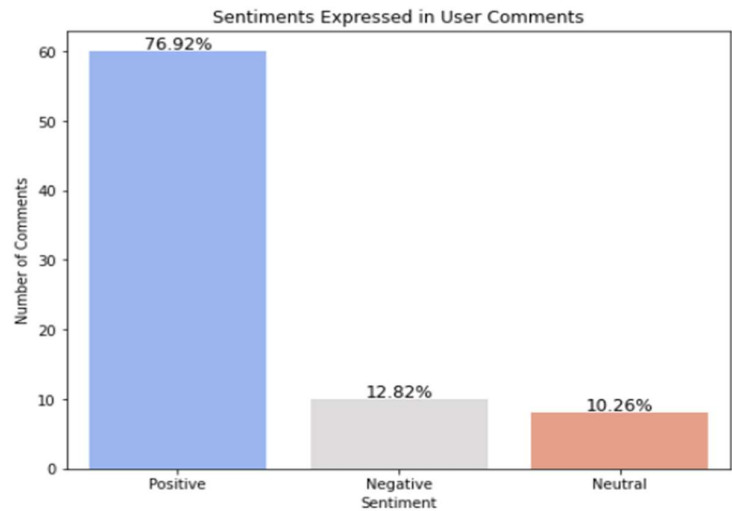
Figure 5. User comments sentiment analysis.

For further understanding of user experience, emotion analysis was conducted for the user comments. A pre-trained emotion model, the CardiffNLP Twitter Emotion Model [13], is utilized to detect emotions like joy, sadness, anger, fear, and optimism. For each user comment, scores are calculated for expressing these emotions, and the emotion with the highest score is chosen. Figure 6 shows that 56 comments are emotionally classified as "optimism," whereas less than 10 comments are classified as "joy," "anger," and "sadness." In total, about 82% of the comments (64 out of 78) are classified with a positive emotion ("optimism" and "joy").
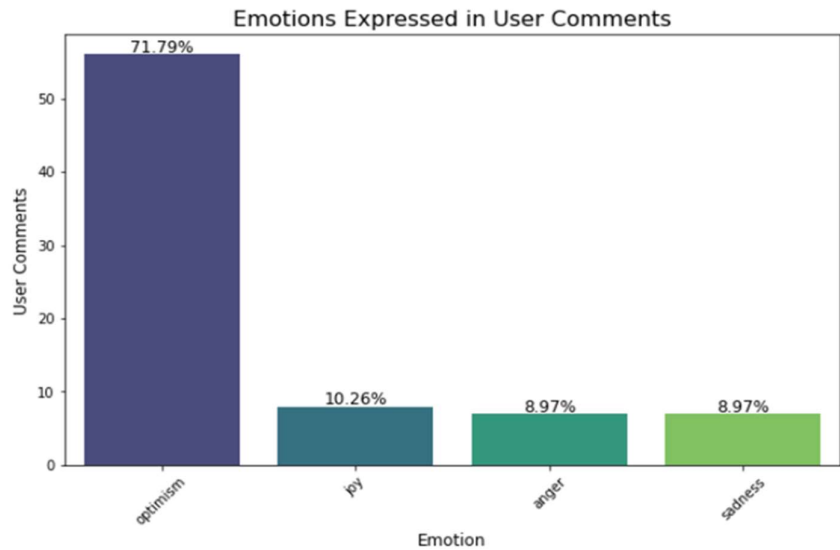

Figure 6. User comments on emotion detection.

2.1. Sentiment in the Open-Ended Questions

There are eight open-ended questions in the survey. The original questions are presented in Table 1, and shortened abbreviations are provided for plotting and writing.

Table 1. Questions from the open-ended section.

Original Question	Shortened Version

How did you find the overall experience of using the Guided Chatbot tool for your Research tasks?	Q1
Did you encounter any difficulties or challenges while using the Guided Chatbot tool for the research? If so, can you provide specific examples?	Q2
What were the most helpful features of the Guided Chatbot in assisting you with your research tasks?	Q3
Were there any limitations or drawbacks to using the Guided Chatbot tool that you would like to see improved?	Q4
Did you feel that the Guided Chatbot tool was able to accurately understand and guide you in your research needs?	Q5
Did you notice any areas where the Guided Chatbot tool could potentially be improved in terms of efficiency or accuracy for research?	Q6
Would you recommend the Guided Chatbot tool to others for similar tasks? Why or why not?	Q7
Is there any additional feedback or suggestions you would like to provide regarding your experience with the Guided Chatbot tool?	Q8

Out of 98 respondents, 85 completed some open-ended questions. To further analyze the sentiment in detail, sentiment values were calculated for each open-ended question. The number of positive, negative, and neutral feedback responses for all eight open-ended questions was computed using the TextBlob algorithms in Python. To visualize the results, all values are displayed in scatter plots based on individual open-ended questions. A total of eight scatter plots were created, one for each of the eight open questions [Figure 7].
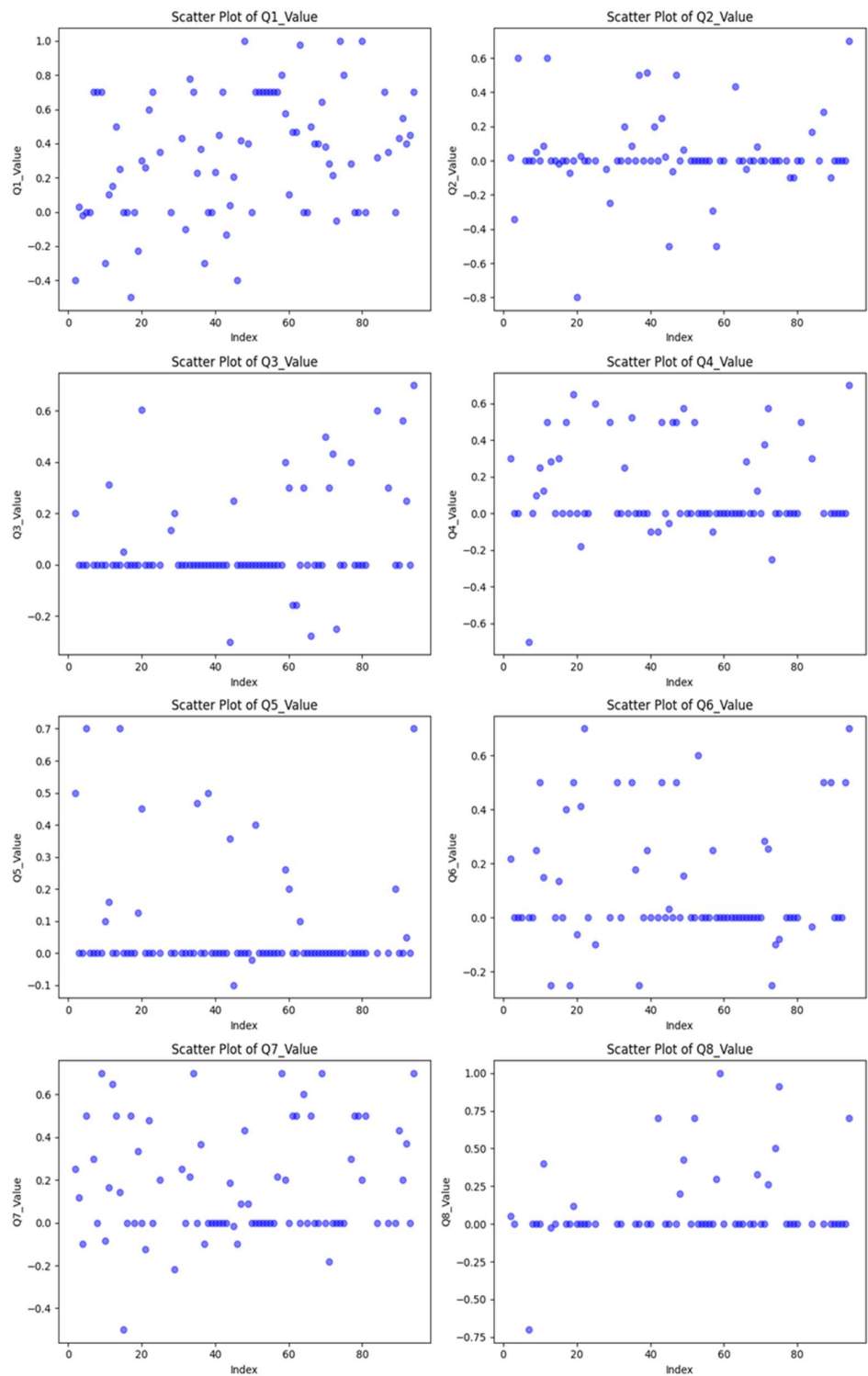
Figure 7. Scatter Plot for Q1 - Q8.

On the Q1 scatter plot, there is no obvious line. On the scatter plots for Q2, Q3, Q4, Q5, Q6, Q7, and Q8, there are obvious lines at the value of 0.00. For these eight scatter plots, there is more positive feedback than negative feedback.

To better compare the numbers, the values of positive, negative, and neutral feedback are listed in Table 2. Additionally, the total count of the feedback and the percentage of positive feedback are calculated and presented in Table 2. According to the comparison table, Q1 has the highest percentage

of positive feedback (70%), indicating that users are generally satisfied with the overall experience of the chatbot. Q7 has a strong 47% positive feedback, suggesting that users are relatively satisfied with the chatbot's overall recommendation potential. Q3, Q5, and Q8 all have relatively low percentages of positive feedback (23% - 25%), showing areas where the chatbot might need improvement based on user responses. Q2 has the lowest percentage of positive feedback (25%), suggesting that users face more difficulties or challenges with the chatbot in the research process. In terms of limitations and efficiency, users show split opinions. 32% positive feedback in Q4 shows some satisfaction with using the chatbot. The rest, with neutral and negative feedback, indicate that limitations in the chatbot should be addressed. 33% of respondents in Q6 have positive feedback, indicating some satisfaction with its efficiency or accuracy. The rest of the respondents suggest improvements should be made in this area.

Table 2. Sentiment Values of Open-Ended Question Feedback.

Question	Positive Feedback	Negative Feedback	Neutral Feedback	Count	Percentage of Positive Feedback
Q1	58	11	14	83	70%
Q2	20	14	45	79	25%
Q3	19	5	59	83	23%
Q4	25	7	47	79	32%
Q5	17	2	64	83	21%
Q6	25	9	41	75	33%
Q7	38	9	34	81	47%
Q8	14	2	46	62	23%

According to the comparison table, Q1 has the highest percentage of positive feedback (70%), indicating that users are generally satisfied with the overall experience of the chatbot. Q7 has a strong 47% positive feedback, suggesting that users are relatively satisfied with the chatbot's overall recommendation potential.

Q3, Q5, and Q8 all have relatively low percentages of positive feedback (23% - 25%), showing areas where the chatbot might need improvement based on user responses. Q2 has the lowest percentage of positive feedback (25%), suggesting that users face more difficulties or challenges with the chatbot in the research process.

In terms of limitations and efficiency, users show split opinions. 32% positive feedback in Q4 shows some satisfaction in using the chatbot. The rest, with neutral and negative feedback, indicate that limitations in the chatbot should be addressed. 33% of respondents in Q6 have positive feedback, indicating some satisfaction with its efficiency or accuracy. The rest of the respondents suggest improvements should be made in this area.

2.2. Questionnaire Responses

The questionnaire contains 14 structured questions. The original questions are presented in Table 3, and shortened versions are provided for plotting and writing.

Table 3. Original Structured questions and the shortened version.

Original Questions	Shortened Version
Specify your gender	Gender
Program	Program
How frequently do you use Generative AI (E.g. ChatGPT) for your work?	Freqency_Use
Which Chatbot provided an effective concept that assisted you most in formulating the research questions and hypotheses?	Which_Chatbot_Effective_Formulate_Questions

Which Chatbot simplified the concept of research methodology?	Which_Chatbot_Simplify_Methodology
Which Chatbot guided you in conceptualizing the research questions and hypotheses?	Which_Chatbot_Conceptualize_Questions
Which Chatbot was your preference in formulating research questions and hypotheses?	Which_Chatbot_Preference_Formulate_Questions
Do you feel Guided Chatbot asked questions that were relevant in formulating the research questions and hypothesis for a novice researcher?	Rate_Formulate_Questions_Relevant
How would you rate the Guided Chatbot to give you more ownership of your idea in formulating the research questions and hypotheses?	Rate_Formulate_Questions_Ownership_Idea
How would you rate the Guided Chatbot to emphasize your creativity in formulating the research questions and hypotheses?	Rate_Formulate_Questions_Creativity
How would you rate the trust in the Guided Chatbot in minimizing misinformation and disinformation towards your research?	Rate_Minimize_Misinformation
How would you rate the Guided Chatbot to assist you in understanding the flow of research methodology?	Rate_Research_Methodology
How would you rate your research method learning outcome using the Guided Chatbot?	Rate_Learning_Outcome
How would you rate your research skills using the Guided Chatbot?	Rate_Research_Skill

Ninety-eight responses were received for this questionnaire. The responses came from almost an equal number of males and females, with 49 males and 46 females participating. This balanced representation ensures that both genders are fairly represented in the data, allowing for a more comprehensive analysis of the responses. The program is an open-ended question. All answers to this question are: 'Undergraduate (Degree)', 'Postgraduate (Masters, PhD)', 'Diploma', 'Post doc' and 'Postgraduate (Masters, PhD)'.

The survey examines the frequency of Generative AI (e.g., ChatGPT) usage in a work setting, categorized into six levels: "Everyday," "Often," "Very Frequently," "Rarely," "Never," and "Never Use." The outcome highlights key usage trends among 98 respondents.

- High-Frequency Usage (56/98): The combined categories of "Everyday," "Often," and "Very Frequently" indicate a significant portion of respondents, with 56 out of 98 individuals (approximately 57.1%) reporting frequent use of ChatGPT for their work. This indicates a notable reliance on Generative AI for daily or regular tasks.
- Low-Frequency Usage (24/98): About 24 respondents (24.5%) reported using ChatGPT "Rarely" for work-related activities. This segment represents users who employ the tool only for occasional or specific purposes.
- Non-Usage (15/98): A smaller segment, comprising 15 respondents (15.3%), indicated that they "Never" or "Never Used" Generative AI tools for their work. This suggests that a minority either does not find the tool relevant or lacks access to it.

Overall, 80 of the 98 respondents (81.6%) acknowledged using Generative AI tools at least occasionally, demonstrating widespread adoption of Generative AI in professional contexts. The majority fall within the high-frequency categories, underscoring the tool's importance and integration into workplace practices. However, the presence of non-users and low-frequency users points to variability in acceptance or need for the technology across different roles or industries. The data is shown in Figure 8.
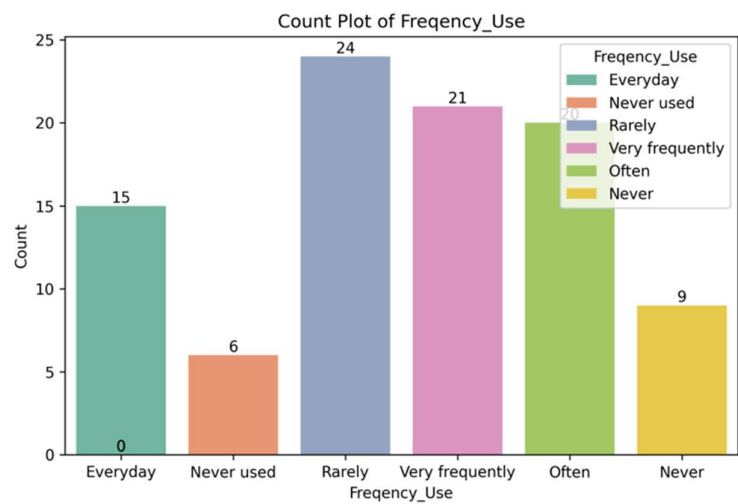
Figure 8. Frequency use of Gen AI.

Which chatbot provided an effective concept that assisted you most in formulating the research questions and hypotheses? There are four options for this structured question. In terms of formulating the research questions and hypotheses, 33 respondents (33.7%) found the Guided Chatbot most effective, while another 33 respondents (33.7%) believed ChatGPT was their primary tool. Both chatbots are equally valued by users for their utility in research-related assistance. Eighteen respondents (18.4%) reported that using both the Guided Chatbot and ChatGPT together was most effective. A high percentage of respondents use the Guided Chatbot and ChatGPT for their research-related assistance. A small segment, 11 respondents (11.2%), chose "Not sure" regarding which chatbot assisted them the most. This highlights a small group that either lacked sufficient experience with the chatbots or found it difficult to distinguish the differences between these two tools. This data is shown in Figure 9.
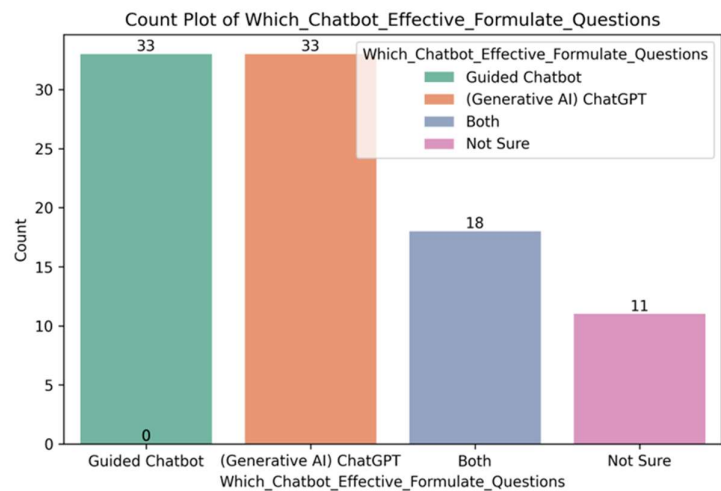
Figure 9. Count of Effectively Formulated Questions.

The question evaluates which chatbot was most effective in simplifying the concept of research methodology based on respondent feedback. From the results, 44 respondents (46.3%) found the Guided Chatbot to be the most effective in simplifying research methodology concepts, 24

respondents (25.3%) preferred ChatGPT, and 12 respondents (12.6%) reported that both tools were the most effective ones. The high percentage for the Guided Chatbot, almost half of the respondents, indicates that the structured and targeted approach of the Guided Chatbot resonates well with users for this purpose. The preference for ChatGPT shows its utility in providing flexible, generative responses for methodology-related questions. The combination of both chatbots suggests that integrating both tools could enhance the functionalities of chatbots in simplifying the concept of research methodology. Fifteen respondents (15.8%) showed "Not sure," reflecting a small segment that either did not use chatbots or were not familiar with this survey question. This data is shown in Figure 10.
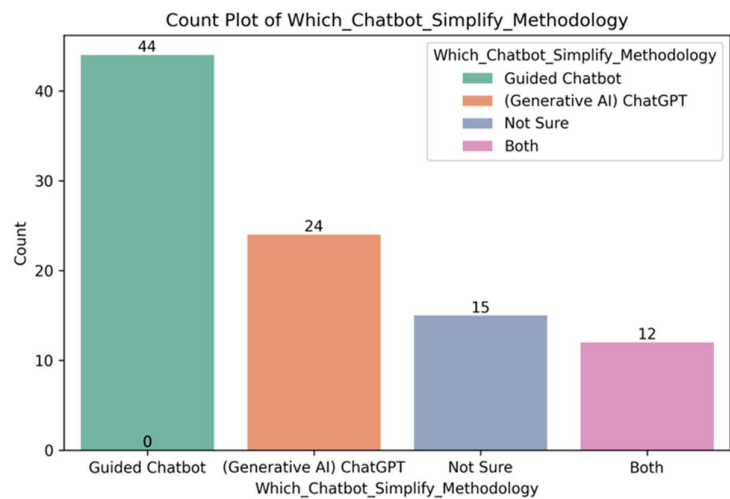
Figure 10. Count of Simplifying Methodology.

Which chatbot guided you in conceptualizing the research questions and hypotheses? The question results show that 46 respondents (48.4%) identified the Guided Chatbot as the most effective tool in conceptualizing research questions and hypotheses. Twenty-one respondents (22.1%) reported that Generative AI was the most effective tool, and 15 respondents (15.8%) reported that both tools were the most effective. The strong preference for the Guided Chatbot underlines the importance of a structured approach in research-related tasks. Future improvements could focus on integrating ChatGPT's flexibility with Guided Chatbot's step-by-step approach. Only 13 respondents (13.7%) indicated that this small group may not use chatbots, may require additional training, or may not understand the difference between the two types of chatbots. This data is shown in Figure 11.
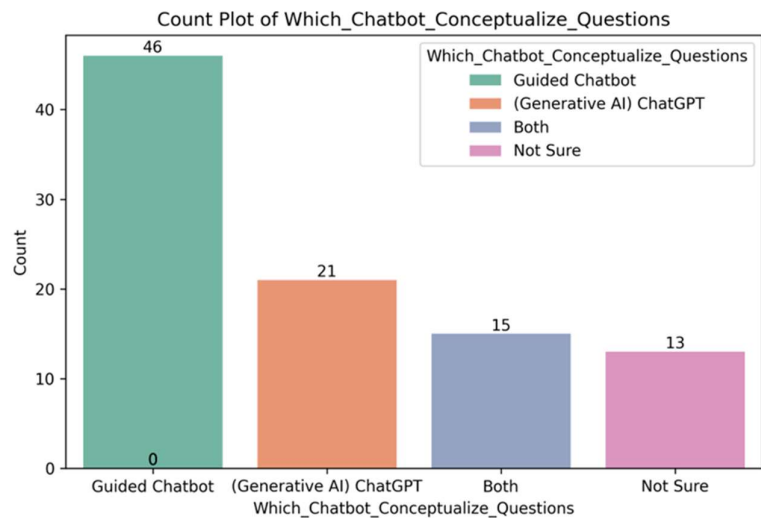
Figure 11. Count of Chatbot Conceptualizing Questions.

The next question investigates user preferences for chatbots in formulating research questions and hypotheses. Based on this question, a majority of 39 respondents (41.1%) preferred the guided chatbot for this task. This suggests the effectiveness of the structured and targeted approach of the guided chatbot, which likely provides clear guidance in formulating research ideas. 31 respondents (32.6%) favored ChatGPT, indicating its popularity for offering flexible, creative, and adaptive suggestions during the research question formulation. 11 respondents (11.6%) reported that a combination of both chatbots was most effective. 14 respondents (14.7%) were "not sure" which chatbot they preferred. From this survey, the guided chatbot is reported as the most preferred tool for formulating research questions and hypotheses, followed closely by ChatGPT. While a smaller group prefers combining both tools, the uncertainty among some respondents suggests an opportunity for further user education and improved design. This data is shown in Figure 12.
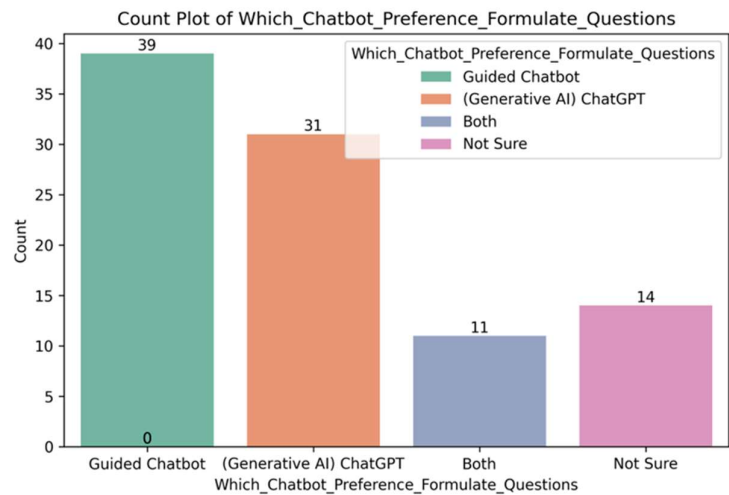

Figure 12. Count of Chatbot Preferences in Formulating Questions.

The next question examines whether the Guided Chatbot posed relevant questions for formulating research questions and hypotheses, particularly for novice researchers. Twenty-eight

respondents (29.5%) strongly agreed that the Guided Chatbot's questions were relevant. This result shows a significant level of confidence among users in the chatbot's ability to provide meaningful and targeted guidance for beginners. Forty-six respondents (48.4%) reported "Agree." This majority response further underscores the chatbot's effectiveness, with nearly half of the participants expressing satisfaction. Twenty respondents (21.1%) were neutral, neither agreeing nor disagreeing with the question. The neutral group of respondents may need additional guidance in understanding this survey question. Only one respondent (1.1%) strongly disagreed, showing that dissatisfaction with the chatbot's relevance is minimal. With seventy-four out of ninety-five respondents (77.9%) either agreeing or strongly agreeing with this specific question, the survey results reveal that guided chatbots pose relevant questions for formulating research questions and hypotheses, particularly for novice researchers. The low level of disagreement (1.1%) and moderate neutrality (21.1%) suggest that while the chatbot is effective for most users, some improvements could enhance its usability for novice researchers. This data is shown in Figure 13.
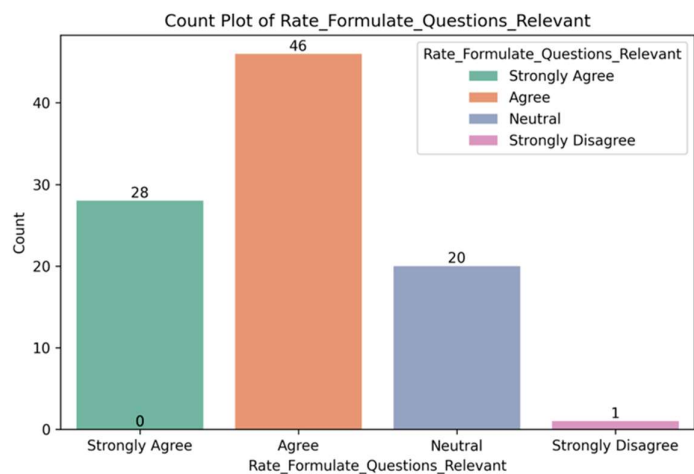

Figure 13. Count of Rate of Formulating Relevant Questions.

The investigation on how well the Guided Chatbot helps users take ownership of their ideas when formulating research questions and hypotheses. Thirty-two respondents (33.7%) rated the Guided Chatbot as excellent in fostering ownership of their ideas. Thirty-nine respondents (41.1%) rated it as good, representing the largest group. Fourteen respondents (14.7%) considered the chatbot to be slightly good in providing ownership. Five respondents (5.3%) remained neutral, suggesting they neither found it effective nor ineffective in enhancing ownership of ideas. Five respondents (5.3%) rated the chatbot as slightly bad, indicating a small group that found it lacking in this aspect. A combined seventy-one out of ninety-five respondents (74.8%) rated the Guided Chatbot as either excellent or good, showcasing strong satisfaction with its ability to help users take ownership of their ideas. Nearly three-quarters of respondents highlighted a strong endorsement of chatbots to support independent thinking and creativity in research formulation.

The presence of 14.7% slightly positive and 10.6% negative or neutral responses indicates the chatbot might not have fully met their expectations in some cases. In this matter, there might be an opportunity for further refinement in how the chatbot supports user autonomy. This data is shown in Figure 14.
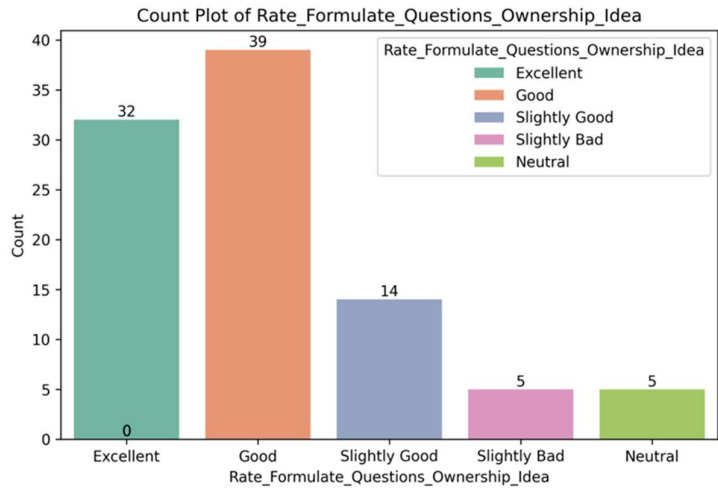
Figure 14. Count of Formulating Questions with Ownership and Ideas.

The question evaluates users' trust in the guided chatbot's ability to minimize misinformation and disinformation regarding research. Twenty-five respondents (26.3%) rated the chatbot as excellent in minimizing misinformation and disinformation. Thirty-eight respondents (40.0%) believed it was good, the largest group. Fourteen respondents (14.7%) reported the chatbot to be slightly good at minimizing misinformation. Fifteen respondents (15.8%) were neutral, implying that these users neither strongly trusted nor distrusted the guided chatbot in this regard. Three respondents (3.2%) believed the chatbot was slightly bad, reflecting a small minority who did not trust it in this context.

A total of 77 out of 95 respondents (81.0%) rated the chatbot positively (excellent, good, or slightly good) for its ability to minimize misinformation and disinformation. This indicates that most respondents have confidence in the chatbot's reliability and accuracy. However, the presence of 15.8% neutral and 3.2% slightly negative feedback (19.0%) suggests that some users remain cautious about its reliability in minimizing misinformation and disinformation regarding research. This indicates room for improvement. This data is shown in Figure 15.

Figure 15. Count of Minimise Misinformation.

The question on how effectively the guided chatbot assists users in understanding the flow of research methodology; the result shows thirty-seven respondents (38.9%) rated the guided chatbot as an excellent tool, thirty-five respondents (36.8%) rated it as good, and twelve respondents (12.6%) considered the chatbot to be slightly good in aiding their understanding of the flow of research methodology. Seven respondents (7.4%) voted for a neutral opinion, and four respondents (4.2%) rated the chatbot as slightly bad, reflecting minimal dissatisfaction.

A total of eighty-four out of ninety-five respondents (88.4%) rated the chatbot positively (excellent, good, or slightly good) in helping them understand the flow of research methodology. These three groups represent a significant proportion of users who were satisfied or very satisfied with the chatbot's assistance. The small percentage of neutral and slightly negative feedback (11.6%) indicates there might be room for potential improvements, such as step-by-step explanations, flowchart demonstrations, interactive guidance, and so on. This data is shown in Figure 16.
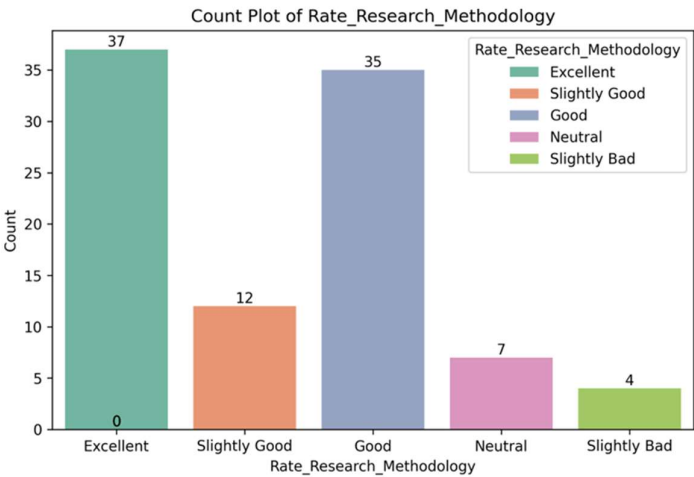

Figure 16. Count of Rate of Research Methodology.

The survey questioned the effectiveness of the Guided Chatbot in supporting users' learning outcomes related to research methods. Thirty-two respondents (33.7%) rated their learning outcomes as excellent. Thirty-nine respondents (41.1%) rated the outcomes as good, which is the largest group. Eleven respondents (11.6%) considered the outcomes slightly good, showing moderate satisfaction. Ten respondents (10.5%) were neutral. Three respondents (3.2%) rated the outcomes as slightly bad, expressing minor dissatisfaction. A combined 82 out of 95 respondents (86.3%) rated their research method learning outcomes positively (excellent, good, or slightly good). Although chatbots are efficient and trustworthy for most users in supporting learning outcomes, 13.7% expressed neutral or slightly negative sentiments. This indicates that there are some gaps to improve. This data is shown in Figure 17.
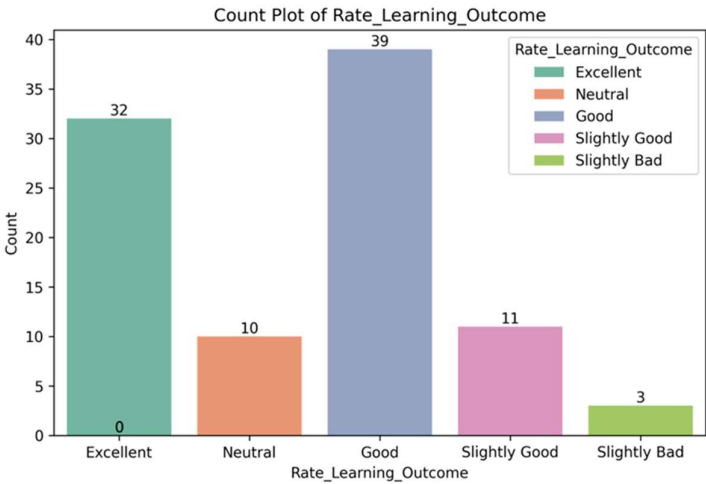
Figure 17. Count on Learning Outcome.

The survey includes a rating question on the user's research skills using the Guided Chatbot. Twenty-three respondents (24.2%) rated their research skills as excellent. Forty-one respondents (43.2%) rated their skills as good, forming the largest group. Thirteen respondents (13.7%) rated their skills as slightly good, demonstrating moderate satisfaction. Thirteen respondents (13.7%) were neutral. Only five respondents (5.3%) rated their skills as slightly bad. Out of 95 respondents, 71 reported that their research skills improved after using the guided chatbot. However, the remaining 14 respondents felt they did not receive significant support from the chatbot, highlighting an area for improvement.

The next question investigated how well the Guided Chatbot helps users in the creativity of their ideas when formulating research questions and hypotheses. Twenty-four respondents (25.3%) rated Guided Chatbot as excellent in fostering the creativity of their ideas. Thirty-nine respondents (41.1%) rated it as good, representing the largest group. Eighteen respondents (18.9%) considered the chatbot to be slightly good in encouraging creativity. Nine respondents (9.4%) remained neutral, suggesting they neither found it effective nor ineffective in enhancing creative ideas. Five respondents (5.3%) rated the chatbot as slightly bad, indicating a small group that found it lacking in this aspect.

A combined seventy-one out of sixty-three respondents (66.3%) rated the Guided Chatbot as either excellent or good, showcasing strong satisfaction with its ability to help users with creative ideas. Nearly three-quarters of respondents highlighted a strong endorsement of chatbots to support independent thinking and creativity in research formulation. The presence of 18.9% slightly positive and 14.8% negative or neutral responses indicates the chatbot might not have fully met their expectations in some cases. In this matter, there might be an opportunity for further refinement in how the chatbot supports user autonomy. This data is shown in Figure 18.
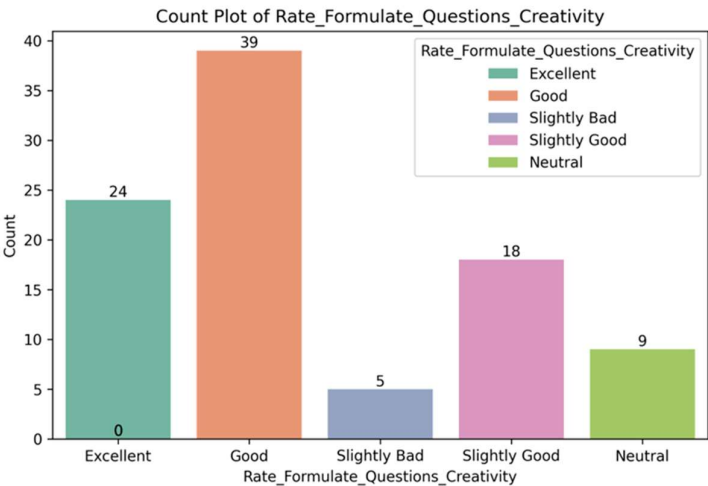
Figure 18. Count of Fostering Creativity.

2.3. Qualitative Evaluation—A Thematic Story

The qualitative analysis of the open-ended questions was conducted as a thematic analysis [14,15]. The eight open-ended questions provide significant insights that capture users' experiences in two themes. Theme 1: Overall experiences and attitudes towards the research-guided chatbot, and Theme 2: Negative experiences and challenges that lead to improvements in the research-guided chatbot.

Theme 1, as shown in *Figure* 21, summarizes the overall experience of using the research-guided chatbot and can be divided into "positive learning outcomes," "positive learning experiences," and "negative experiences." Respondents found the features helpful in stimulating their research ideas and motivating their ability to showcase those ideas. Theme 1 highlights a generally positive reception, particularly from novice researchers. This demographic seems to find the chatbot's features particularly beneficial in demystifying the research process, suggesting that it plays a crucial role in enhancing their learning journey. By breaking down complex phases, the chatbot appears to empower users, allowing them to build confidence in their research abilities. The emphasis on timesaving and user-friendliness suggests that the design and usability of the chatbot align well with users' needs, making it a valuable tool in the research process. The chatbot is informative and helps users understand the research process on a wider scale. It is simple, effective, and easy to use, saving users a lot of time and effort in completing tasks. The examples provided are good. The chatbot's analysis of user errors was explained and helped guide users to the best responses.

The guided chatbot had a formulaic feel that challenged my creativity; however, it was well-designed to understand the research process. ~Postgraduate student.

The most helpful features of the guided chatbot were its ability to quickly organize and summarize information, provide clear explanations, and offer structured guidance for developing research ideas. ~Postgraduate student.

The learning outcome experience using the chatbots was that they accurately understood users' research difficulties and guided them to form research concepts easily, without too much complexity. These helpful features were appreciated by users, leading them to recommend the guided chatbots to other novice researchers. The chatbot understood users' earlier prompts and provided examples that assisted them in formulating their next questions.

The negative experiences were appropriate sentiments from using the first version of the research-guided prototype. Respondents felt there was room for improvement and suggested suitable features to promote learning and enhance user-friendliness.

No, but more examples would be good. While the chatbot provided a solid foundation, there were instances where more in-depth information or specific examples would have been beneficial for certain research tasks. ~Postgraduate student.

Improving the tool's ability to understand and respond to complex or nuanced questions would enhance its efficiency. Additionally, better accuracy in providing relevant and specific answers for research tasks would be beneficial. ~Postgraduate student.

Yes, it needs to better understand human slang, abbreviations, Singlish, and poor punctuation. ~Undergraduate student.
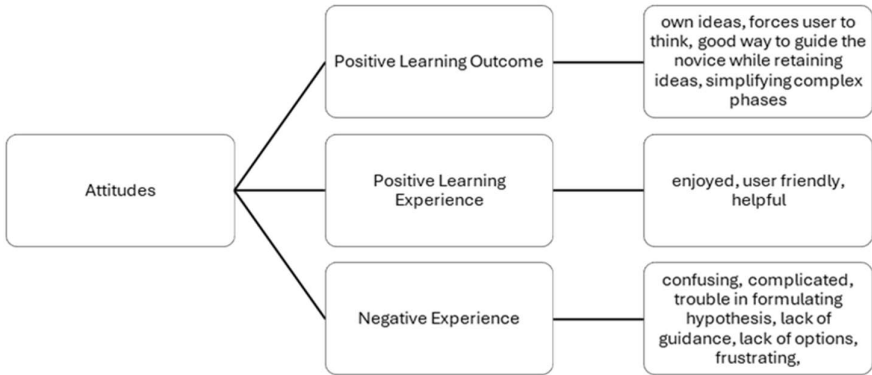
Figure 19. Theme 1 - Overall experiences and Attitudes towards Research-guided chatbot.

Theme 2, as shown in Figure 22, summarizes the conclusions and suggestions for the next version of the research-guided chatbot, emphasizing simplicity as proposed by users. Users have an optimistic view of learning new research skills while retaining their ideas. This view is the main highlight of the chatbot and will apply to the design of future educational chatbots. Negative experiences and challenges are equally important. Understanding these drawbacks can inform future improvements and enhancements to the chatbot. Gathering specific examples of these negative experiences will be crucial for identifying patterns or common issues faced by users. This feedback could guide the development team in refining the chatbot, ensuring it meets the diverse needs of users and minimizes barriers to effective use.

The Guided Chatbot is a good tool for developing skills and understanding a process. It's very clever in providing suggestions and compilations. It allows the development of more personal creativity. It helps to formulate your research idea and correct the thought process as well. ~Postgraduate and undergraduate students.

Additional suggestions from users are to enhance the features for the next version of the research-guided chatbot. The recommendation was relevant for tertiary students to improve their research learning skills.

Yes, I would recommend the Guided Chatbot tool for similar tasks because it provides valuable support in organising and summarising information. However, users should be aware of its limitations with complex queries and relevance and use it alongside other resources for the best results. ~Postgraduate student.

Overall, this thematic analysis not only presents a comprehensive view of users' experiences but also sets the stage for continuous improvement of the chatbot, ultimately benefiting the research community.
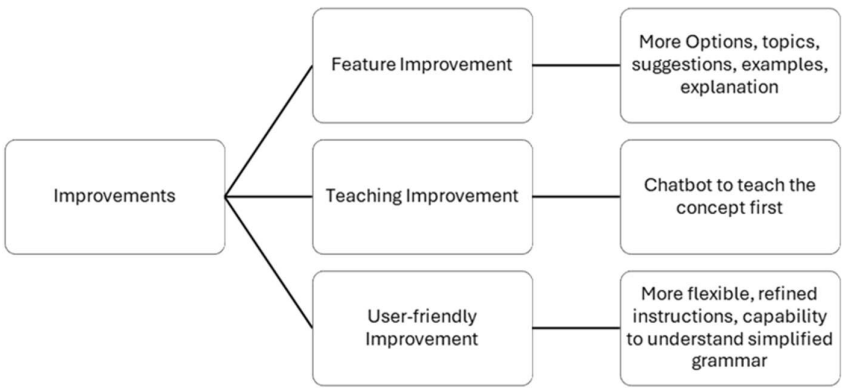
Figure 20. Theme 2 - Negative experiences and challenges lead to improving chatbot.

5. Conclusions and Future Work

Guided chatbots are preferred over generative AI among tertiary students, specifically postgraduate students, in areas of guidance, examples, and idea ownership. Postgraduate students with research experience also found that the guided chatbot is most suitable for novice researchers in practicing the formulation of research questions and hypotheses. Both guided chatbots and generative AI are satisfactory tools for promoting learning for most users. A second version of the guided chatbot is being considered, focusing on areas for improvement for users who rated it as **"neutral"** or **"slightly bad,"** such as providing more support for guiding users with a greater selection of research topics. It is also recommended that future work consider combining guided chatbots with generative AI and investigate how to improve the functionalities of guided chatbots for unsatisfied users. Overall, this user-centered approach can drive improvements and foster greater satisfaction in learning outcomes and experiences. Future research will also explore how chatbot functionalities can be personalized to cater to the distinct needs of postgraduate students through adaptive learning technologies that respond in real time to user inputs and challenges to encourage deeper engagement and satisfaction with the learning process.

References

1. Martinez, F., Collarana, D., Calvaresi, D., Arispe, M., Florida, C., & Calbimonte, J. P. (2023, May). Study-Buddy: A Knowledge Graph-Powered Learning Companion for School Students. In European Semantic

Web Conference (pp. 133-137). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-43458-7_25

2. Kabiljo, M., Vidas-Bubanja, M., Matić, R., & Zivković, M. (2020). Education system in the republic of serbia under COVID-19 conditions: Chatbot-acadimic digital assistant of the belgrade business and arts academy of applied studies. KNOWLEDGE-International Journal, 43(1), 25-30. https://doi.org/10.18690/um.fov.3.2022.44

3. Konecki, M., Konecki, M., & Biskupic, I. O. (2023, April). Using Artificial Intelligence in Higher Education. In CSEDU (2) (pp. 136-141). https://doi.org/10.5220/0012039700003470

4. Duelen, A., Jennes, I., & Van den Broeck, W. (2024, June). Socratic AI Against Disinformation: Improving Critical Thinking to Recognize Disinformation Using Socratic AI. In Proceedings of the 2024 ACM International Conference on Interactive Media Experiences (pp. 375-381). https://doi.org/10.1145/3639701.3663640

5. Yang, A., Xiao, B., Wang, B., Zhang, B., Bian, C., Yin, C., ... & Wu, Z. (2023). Baichuan 2: Open large-scale language models. arXiv preprint arXiv:2309.10305.

6. Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and individual differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

7. Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in education and teaching international, 61(2), 228-239. https://doi.org/10.1080/14703297.2023.2190148

8. Perkins, M. (2023). Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. Journal of University Teaching and Learning Practice, 20(2). https://doi.org/10.53761/1.20.02.07

9. VS Vadivel, S Sohail, Loh, DY., Tan, WJ., Chung LY. (2025). Invert the chatbot and own a research plan – A Chatbot Model for Novice Researchers. Manuscript submitted for publication.

10. Misra, D. P., Gasparyan, A. Y., Zimba, O., Yessirkepov, M., Agarwal, V., & Kitas, G. D. (2021). Formulating hypotheses for different study designs. Journal of Korean Medical Science, 36(50). https://doi.org/10.3346/jkms.2021.36.e338

11. Gu, X., & Krenn, M. (2024). Generation and human-expert evaluation of interesting research ideas using knowledge graphs and large language models. arXiv preprint arXiv:2405.17044.

12. Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014. https://doi.org/10.1609/icwsm.v8i1.14550

13. Yadav, V., & Shukla, P. (2021). cardiffnlp/twitter-roberta-base-emotion. Hugging Face. Retrieved from https://huggingface.co/cardiffnlp/twitter-roberta-base-emotion

14. Vaismoradi, M., Jones, J., Turunen, H., & Snelgrove, S. (2016). Theme development in qualitative content analysis and thematic analysis https://doi.org/10.5430/jnep.v6n5p100

15. Clarke, V., Braun, V., Frith, H., & Moller, N. (2019). Editorial introduction to the special issue: Using story completion methods in qualitative research. Qualitative Research in Psychology, 16(1), 1-20. https://doi.org/10.1080/14780887.2018.1536378