International Journal of Science, Strategic Management and Technology
Volume **02** Issue **04** April-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

Behavior of ChatGPT Response Patterns for Algorithm and Flowchart Questions in Written Assignments

Mrs. Sujata Bachhav
Mrs. Smita Shelar
Department of Computer Science & Application Kaveri College of Arts Science and Commerce

https://doi.org/10.55041/ijsmt.v2i4.118

Cite this Article: Bachhav, S. & Shelar, S. (2026). Behavior of ChatGPT Response Patterns for Algorithm and Flowchart Questions in Written Assignments. International Journal of Science, Strategic Management and Technology, *02*(04). https://doi.org/10.55041/ijsmt.v2i4.118

Abstract

Large Language Models (LLMs) such as ChatGPT demonstrate impressive skills in areas related to Natural Language Processing and code generation [1]. The capacity of ChatGPT to generate coherent, structured, human-like responses is becoming increasingly important in academic settings, particularly for clarifying algorithm design and addressing flowchart-related questions in written assignments.

Studies have demonstrated that ChatGPT can generate well-expressed, hunamoid text passages and dialogues, frequently exhibiting unexpectedly high quality [2]. This is important to be able to understand algorithm descriptions. For algorithm question prompts, ChatGPT usually gives full and exact answers.

This study looks at behavioral response patterns of ChatGPT to the same algorithm and flowchart questions from different people to see if the formatting changes, structural discrepancy, semantics and conceptual consistency of the answers change. This paper analyzes the occurrence of generative variability in structured technical tasks [3], demonstrating that while the logical structure and semantic precision of algorithmic solutions generally remain stable, the syntactic structure may fluctuate across repeated inquiries. This indicates that although the fundamental information remains reliable, its presentation can vary considerably.The results increase our knowledge of the educational implications, repeatability and dependability of technical content created by AI.

Keywords: instructional implications, generative variability, behavioral response patterns, and large language models.

Introduction

ChatGPT and other Large Language Models (LLMs) have demonstrated exceptional performance in natural language processing (NLP) and coding tasks [1], generating responses that are coherent, well-structured, and human-like, making them valuable tools in education and technology. In addition to conversational intelligence, these models also have been observed to be competent in creating algorithmic, explanatory, and structured content such as flowcharts, providing detailed, well-structured, and contextually relevant answers (similar to expert-level responses) to algorithm-based queries, but with variation in responses even with repeated inputs [2]. This study will investigate the behavioral patterns of ChatGPT when queried with repeated algorithm-based and flowchart-based questions and the variability in their responses. The goal of this study is to analyze the variability in responses from ChatGPT to repeated inputs to better understand their behavioral patterns [2].

The findings suggest that syntactical changes in the query structure on the surface level may occur with repeated queries, but the logical structure of the algorithmic responses may stay the same, which may not be the case for flowcharts. The results offer promising insights into the reproducibility, reliability, and pedagogical value of AI-generated content in an educational context.

Literature Review

The rise of Large Language Models such as ChatGPT has impacted several domains in recent years. For example, in the field of education, [5] has discussed the impact of Large Language Models, and a number of studies have investigated how programming students use Large Language Models in various ways, from creating simple question prompts to more advanced question development using multiple rounds of questioning. Interaction with Large Language Models can be employed for a variety of purposes in programming education [6] , from code generation to question creation in basic/Primary programming education.

Researchers have analyzed the chat protocols to look at what has been done with Large Language Models. Large Language Models can be evaluated in their performance at completing a variety of tasks, such as near perfect completion rates on basic programming exams, and questions about academic integrity and students not actually learning the objectives of the course

[4]. As such, there is heightened concern that LLMs are not robust enough to maintain consistency and accuracy of generated outputs for use in learning and teaching contexts [5], particularly for those fields where algorithmic logic and flowchart construction are a major part of the learning and teaching outcomes. Given these capabilities and characteristics of LLMs, a more critical assessment is required of LLMs for tasks that include the nuances of algorithm and flowchart creation, including reliability, repeatability, robustness, and confidence in generating accurate and correct outputs for learning and teaching purposes [1] .

While previous research has focused on whether ChatGPT produces syntactically valid code for competitive programming contests (where it has been found that students who use ChatGPT receive higher scores, but are still prone to errors in the generated code), there is little research about its ability to generate logic or how to construct algorithms and flowcharts, which are foundational to learning before implementation.

The research on student interactions with AI assistants has tried to determine categories of use, from direct requests for code to more collaborative and iterative problem-solving interactions, which more closely resemble interactions between students.The extent to which generative AI can provide accurate and detailed explanations of concepts related to control structures, such as iteration, is debated, and some research indicates human instructors are more effective at explaining complex looping concepts.Furthermore, cognitive demands of algorithmic thinking require students to break problems down into structured and step-by-step solutions, where novice programmers often struggle to formulate appropriate logic for solutions.

Methods used

Research Design

The research analytical method employed is supported by existing research works. The structure of the experiment includes:

1. Submitting the same query for a flowchart and algorithm more than once in different environments by different people.
2. Analyzing results from different trials.
3. Evaluating

- Correctness of logic
- Sequencing steps
- structure of a flowchart
- Formatting variation (Changes)

Evaluation Criteria

Responses were analyzed according to:

- Semantic Stability – Whether the core algorithm logic remains unchanged.
- Structural Consistency – Whether step order and control flow remain similar.
- Syntactic Variation – Differences in wording, formatting, or representation.
- Graphical Representation Variation – Differences in flowchart drawing style An experimental comparative approach was adopted.

Identical assignment-style algorithm and flowchart questions were answered:

- Using ChatGPT
- Without using ChatGPT

A total of 104 responses were collected, 52 in each category.

Error Classification

Responses were categorized into:

Table 1

Detailed Distribution of Response Patterns

Variation in answer	No of Students using ChatGPT	No of Students without using ChatGPT
wrong flowchart Directly written pattern in flowchart	11	2
correct but not written for n rows	5	2
Flowchart without symbols	3	1
No Back arrow in Loop	10	4
no back arrow in loop,not for n rows	5	2
correct	18	41
Total	52	52

Grouped Summary

Table 2

Grouped Categories

Variation in answer	No of Students using ChatGPT	No of Students without using ChatGPT
wrong flowchart Directly written pattern in flowchart	11	2
correct but not written for n rows	5	2
Flowchart without symbols	3	1
Missing back arrow in loop/not for n rows	15	6
correct	18	41
Total	52	52

Results

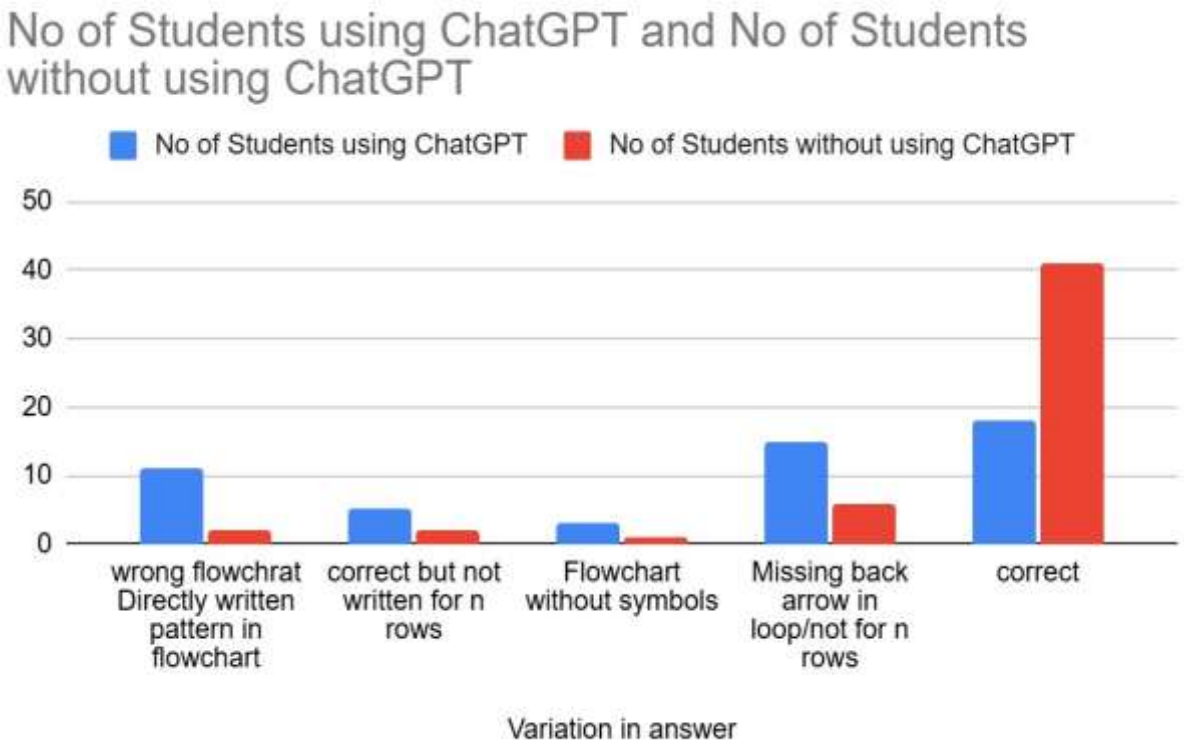

Figure 1

Comparative Distribution of Errors: ChatGPT vs. Without ChatGPT

Pie Chart

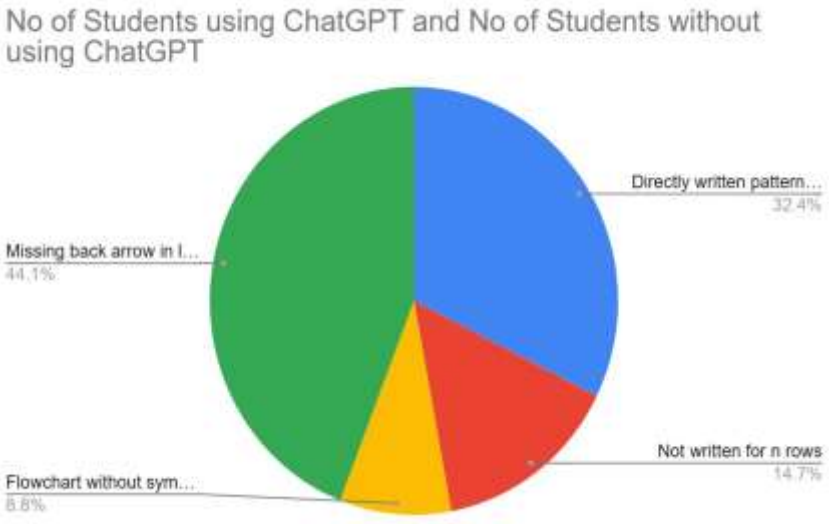

Figure 2

Error Distribution in ChatGPT Responses (Short Labels)

Statistical Hypothesis Testing

In order to further assess the significance of the observed difference in terms of correctness amongst the two groups on comparative measures, a two-proportion Z-test was used.

Purpose of the Test

The purpose of this test is to compare the proportion of correct answers between two groups:

- Group 1: Using ChatGPT
- Group 2: Without using ChatGPT

The test determines whether the difference in correctness rates is statistically significant or simply due to random variation.

Population Proportions

Let:

p_1 = true proportion of correct answers when using ChatGPT

p_2 = true proportion of correct answers without using ChatGPT

These represent the population proportions rather than the observed sample proportions.

Null Hypothesis (H_0)

$H0 : p1 = p2$

This means that there is no difference in the number of correct answers between the two methods. The difference found in the sample (18 correct answers with ChatGPT and 41 correct answers without ChatGPT) is considered to be due to random chance.

Alternative Hypothesis (H_1)

Two possible forms of the alternative hypothesis can be considered.

Two-tailed test

$H_1 : p_1 \neq p_2$

This indicates that there is a statistically significant difference in the proportion of correct answers between the two groups.

One-tailed test

If the expectation is that one group performs better than the other, the hypothesis can be written as:

$H1 : p2 > p1$

which tests whether the group without ChatGPT produces a higher proportion of correct answers.

Alternatively,

$H1 : p1 < p2$

tests whether the correctness rate when using ChatGPT is lower.

Hypotheses Interpretation

In simple terms:

- The null hypothesis is that there is no significant difference between the correctness of the two methods
- The alternative hypothesis is that there is a significant difference between the correctness of the two methods

Hypothesis Test Result

The computed test statistic is as follows:

$Z = -4.56$

with

$p < 0.001$

As the computed probability is considerably smaller than the standard level of significance of 0.05, the null hypothesis can be rejected and the alternate hypothesis is accepted.

This implies that the proportion of correct answers is significantly different between the two groups, in particular, the group that did not use ChatGPT has a much higher correctness rate.

Interpretation

1. **Response Consistency**

The most common problem with the responses of ChatGPT was the lack of loop arrows in its responses (44.1 %).
The wrong logic in the pattern sequencing was the second most common error (32.4%).

2. **Structural Variation**

Significant variations in structure were observed in the responses of ChatGPT.
Although the use of symbolic representation was evident in its responses, the
loop-back mechanism was not sufficiently represented.

3. **Semantic Stability**

In instances where the structural representation was incomplete, the semantic representation was still sufficiently clear to make inferences about the algorithm. However, in instances of wrong logic in pattern representation, the algorithm was significantly altered.

4. **Comparative Findings**

- Without ChatGPT: Stronger structural precision in loop representation, weaker
generalization.
- With ChatGPT: Better formatting and scalability, weaker control-flow completeness.

Discussion

The study findings reveal that there are two different behavioral patterns:
Cognitive Construction (Without ChatGPT)
The answers show deeper contextual understanding and correct loop representation but less consistency in formatting.
Template-Based Generation (With ChatGPT)
ChatGPT shows correct and formatted general structures but lacks explicit control-flow mechanisms such as loop-back arrows. The high occurrence of answers lacking loop-back arrows indicates that the AI model prioritizes textual logic explanation over diagrammatic completeness.

Educational Implications

ChatGPT can be used as a tool to assist in structure but not as a tool to validate or conceptualize. Students must validate and verify the answers produced by AI to ensure correctness in terms of structure. Educators can use this study to create tests that require:

- Explicit loop-back arrows to be verified
- Pattern-specific logic to be validated
- Answers to be checked for completeness

Conclusion

The study proves that ChatGPT produces semantically correct and formatted responses to algorithm questions. However, the study finds that the responses produced by
ChatGPTfor flowchart are incomplete in terms of structure. Comparing the responses: Comparatively:

- Independent responses show better structure.
- AI-assisted responses show better generalization and formatting.

ChatGPT can be used as a tool to assist in structure but must be validated to achieve precision in terms of structure in algorithm and flowchart assignments.

References

[1]https://www.bohrium.com/en/paper-details/algorithmic-learning-assessing-the-potential-of-lar ge-language-models-llms-for-automated-exercise-generation-and-grading-in-educational-settings/11469926940991815 73-2624

[2]https://www.bohrium.com/en/paper-details/evaluation-of-reliability-repeatability-robustness-and-confidence-of-gpt-3-5-and-gpt-4-on-a-radiology-board-style-examination/1000432094112907268-9630

[3]https://www.bohrium.com/en/paper-details/understanding-self-directed-learning-in-ai-assisted-writing-a-mixed-methods-study-of-postsecondary-learners/1006015603112148998-80334

[4]https://link.springer.com/article/10.1186/s13040-023-00339-9

[5]https://www.sciencedirect.com/science/article/pii/S0360131521002438 [6] https://www.nature.com/articles/s41599-025-04471-1