INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 12, Issue 5 - V12I5-1161)
Available online at: https://www.ijariit.com

Generative AI in Statistics Education: Predicting Students' Confidence in Evaluating AI-Generated Answers

Omkar Chandrashekhar Thakur
omkar.thakur@xaviers.edu
Department of Statistics, St. Xavier's College, Empowered Autonomous Institute, Mumbai

ABSTRACT

Students are increasingly using Generative Artificial Intelligence (GenAI) for learning, problem solving and understanding statistical concepts. However, the effective use of such tools requires the ability to critically assess and validate AI-generated solutions. In this study we investigate GenAI use among undergraduate Statistics students, focusing on Perceived Learning Benefits (PLB), Verification Behavior Score (VBS) and confidence in judging AI-generated statistical solutions. Data were acquired from 294 undergraduate students (244 GenAI users) through a structured online questionnaire. Perceived learning advantages and verification behavior were assessed using multi-item measures. A Random Forest classification model was constructed to predict students' confidence in judging AI-generated statistical answers from AI-use patterns, learning advantages, verification behavior, Statistics experience and year of study. We tested the model with a stratified 5-fold cross-validation. The results point to the pedagogical value of GenAI and the need for verification and critical evaluation skills in AI-assisted learning of Statistics.

Keywords: *Generative AI, Statistics Education, Artificial Intelligence Literacy, Student Confidence*

1. INTRODUCTION

The use of Generative Artificial Intelligence (GenAI) is growing in teaching Statistics to allow students to understand concepts, solve problems and interpret statistical findings. Such tools allow for instant, specific, educational support.

However, the solutions created by the AI may contain improper computations, wrong methodology, or misleading interpretations. Students should consequently critically evaluate and verify the answers that AI provides. This is especially important in Statistics since errors might invalidate the results of the analysis.

The study has the following objectives:

1. To investigate the use of Generative AI by undergraduate students studying Statistics.
2. To assess students' perceived learning benefits and verification behaviour when using Generative AI for Statistics.
3. To examine whether perceived learning benefits, verification behaviour and evaluation confidence differ across FY, SY and TY students.
4. To develop a Random Forest classification model for predicting students' confidence in evaluating the statistical correctness of AI-generated answers and identify the most informative predictors.

2. LITERATURE REVIEW

GenAI has grown into a key learning tool that helps with problem-solving, explanations, examples and feedback. But it's important to question AI's answers, as they may be wrong or misleading. This concern is of particular importance in statistics, as incorrect procedures, calculations, or interpretations can lead to wrong results. Previous work examined the advantages, difficulties and adoption of GenAI. Less work has focused on verification behavior and confidence in the assessment of AI-generated statistical responses. This study fills in that gap by looking at these factors among undergraduate statistics students and creating a Random Forest model to guess how confident students in evaluating AI-generated statistical answers.

3. METHODOLOGY

A quantitative cross-sectional survey was conducted among FY, SY and TY Statistics students. Of 294 respondents, 244 reported using GenAI for learning Statistics. The questionnaire measured AI-use patterns, Statistics experience, Perceived Learning Benefits (PLB), Verification Behaviour Score (VBS) and evaluation confidence. PLB and VBS were constructed from five-point Likert items and assessed using Cronbach's alpha. A Random Forest model was developed among AI users using AI-use frequency, PLB, VBS, Statistics experience, Year of study and AI-use purposes as predictors. Performance was evaluated using stratified five-fold cross-validation with accuracy, F1-score and ROC-AUC. Variable importance was assessed using the Random Forest model.

4. RESULTS AND ANALYSIS

4.1 Generative AI Usage Patterns

AI usage patterns by year of study and frequency are presented in Table 1.

Table -1: Profile and Generative AI usage of respondents

Variable	Category	Number	%
AI usage	AI users	244	83.0
	Non-users	50	17.0
Year of study among AI users	FY	73	29.9
	SY	77	31.6
	TY	94	38.5
AI-use frequency	Rarely	22	9.0
	Sometimes	99	40.6
	Often	85	34.8
	Very often	38	15.6

Generative AI was widely used, with 83.0% of respondents reporting its use for learning Statistics; 75.4% of AI users reported using it sometimes or often.

4.2 Perceived Learning Benefits (PLB) and Verification Behaviour Score (VBS)

PLB and VBS were assessed using the constructed measures; descriptive and reliability results are presented in Table 2.

Table -2: Descriptive statistics and reliability of study measures

Measure	Items	Mean	SD	Cronbach's α
Perceived Learning Benefit (PLB)	4	4.05	0.69	0.862
Verification Behaviour Score (VBS)	3	3.41	0.81	0.767
Evaluation Confidence	1	3.07	0.72	—

Students reported high perceived learning benefits (PLB = 4.05) and moderate verification behaviour (VBS = 3.41), with good internal consistency (α = 0.862 and 0.767, respectively).

To further examine the components contributing to these measures, the item-level results are presented in Table 3.

Table -3: Item-wise mean scores for perceived learning benefit and verification behaviour

Construct	Indicator	Mean
PLB	Understand difficult Statistics concepts	4.11
	Understand steps in solving statistical problems	**4.15**
	Interpret statistical results	4.02
	Increase confidence in learning Statistics	3.90
VBS	Verify using another reliable source	3.01
	Check appropriateness of statistical method	**3.68**
	Independently check numerical calculations	3.54

The highest PLB score was observed for understanding the steps involved in solving statistical problems (mean = 4.15). Verification using another reliable source, on the other hand, got the lowest score (mean = 3.01), which suggests that even though students see a lot of benefits from GenAI, external verification of responses generated by AI may need more attention.

4.3 Differences by Year of Study

To examine differences across FY, SY and TY students, the following hypotheses were formulated.

H_{01}: There is no significant difference in perceived learning benefit, verification behaviour and evaluation confidence among FY, SY and TY students.

H_{11}: There is a significant difference in perceived learning benefit, verification behaviour and evaluation confidence among FY, SY and TY students.

Since the study measures are derived primarily from Likert-scale responses and the three groups represent independent samples, the **Kruskal–Wallis test** was used.

Table -4: Comparison of study measures across year of study

Measure	FY Mean	SY Mean	TY Mean	H	p-value
PLB	3.96	4.11	4.06	1.871	0.392
VBS	3.26	3.34	**3.59**	8.623	**0.013**
Evaluation Confidence	2.97	3.00	**3.19**	4.913	0.086

Note: Kruskal–Wallis test; significance level = 0.05.

Verification behaviour differed significantly across academic levels (H=8.623, p=0.013), with TY students showing the highest VBS (3.59), followed by SY (3.34) and FY (3.26). PLB (p=0.392) and evaluation confidence (p = 0.086) did not differ significantly. Thus, academic level was significantly associated with verification behaviour but not PLB or evaluation confidence.

4.4 Random Forest Prediction of Evaluation Confidence

A Random Forest model was developed using AI-use frequency, PLB, VBS, Statistics experience, year of study and AI-use purposes. Students reporting "very confident" or "extremely confident" were classified as having high evaluation confidence.

Table -5: Random Forest classification performance

Measure	Result
AI-user sample	244
Lower/moderate confidence	182 (74.6%)
High confidence	62 (25.4%)
Validation method	Stratified 5-fold cross-validation
Accuracy	**73.4%**
F1-score	**0.488**
ROC-AUC	**0.709**

A Random Forest model was developed to predict high evaluation confidence using AI-use frequency, PLB, VBS, Statistics experience, year of study and AI-use purposes. Stratified five-fold cross-validation yielded an accuracy of 73.4%, F1-score of 0.488 and ROC-AUC of 0.709, indicating moderate predictive performance. PLB, AI-use frequency and VBS were the most informative predictors.

Table -6: Relative importance of predictors in the Random Forest model

Predictor	Relative importance
Perceived Learning Benefit	Highest
AI-use frequency	High
Verification Behaviour	Moderate
Year of study	Lower
Statistics experience	Lower

5. CONCLUSION

The study shows that GenAI is widely used by undergraduate Statistics students and is perceived as beneficial for learning. However, verification behaviour was moderate, particularly for external checking of AI-generated responses. Verification behavior was also very different between academic levels, with TY students showed higher verification behavior. The Random Forest model did a fair job of predicting (ROC-AUC = 0.709), with PLB, AI-use frequency and VBS being the most useful predictions. Combined with critical assessment and thorough validation, GenAI is able to promote Statistics learning.

REFERENCES

[1] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

[2] Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? International Journal of Educational Technology in Higher Education, 16, 39.

[3] Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. International Journal of Educational Technology in Higher Education, 20, 56.

[4] Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2024). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. Computers & Education, 215, 105224.

[5] Wahba, F., Ajlouni, A. O., & Abumosa, M. A. (2024). The impact of ChatGPT-based learning statistics on undergraduates' statistical reasoning and attitudes toward statistics. Eurasia Journal of Mathematics, Science and Technology Education, 20(7), em2468.

[6] Baek, C., Tate, T., & Warschauer, M. (2024). "ChatGPT seems too good to be true": College students' use and perceptions of generative AI. Computers and Education: Artificial Intelligence, 7, 100294.