

ANALYZING THE IMPACT OF CHATGPT USAGE ON LEARNING BEHAVIOUR AND ACADEMIC PERFORMANCE OF MCA AND MMS STUDENTS**Dr. Sulakshana Vispute**

Assistant Professor, MCA Department, DES's NMITD, Mumbai University, India.

Email: sulakshana.vispute@despune.org**Abstract:**

Artificial Intelligence (AI) tools, particularly ChatGPT, have rapidly emerged as widely adopted academic aids among university students. This study examines the effect of ChatGPT usage on the learning behavior and academic performance of postgraduate students pursuing MCA (Master of Computer Applications) and MMS (Master of Management Studies) programmes at an Institute of Technology. A structured questionnaire was administered to 80 students, capturing data on frequency of usage, purpose of use, average daily time spent, learning behaviour via Likert-scale statements, and last semester SGPI. Three analytical techniques—K-Means Clustering, Pearson Correlation Coefficient, and Multiple Linear Regression—were applied to uncover behavioural groupings and performance relationships. The findings reveal that 80% of students use ChatGPT frequently or always, primarily for concept understanding (42.5%) and exam preparation (26.25%). K-Means analysis identified three distinct user clusters. Pearson correlation indicated weak but directionally meaningful associations between usage frequency and SGPI. The regression model identified usage frequency and answer verification as the most positive predictors of academic performance. The study ends with implications for the responsible integration of AI tools in higher education.

Keywords: ChatGPT, Artificial Intelligence, K-Means Clustering, Pearson Correlation, Multiple Linear Regression, Learning Behavior, Academic Performance MMS, SGPI.

1. Introduction

The fast-paced development of Artificial Intelligence (AI) has introduced transformative tools into almost every aspect of human activity, including the education sector. Among the most impactful recent developments is ChatGPT, a large language model developed by OpenAI, which became publicly accessible in 2022-Nov and within two months achieved over 100 million users— one of the fastest-adopted consumer applications in history [1]. In academic settings, students at all levels have adopted ChatGPT for a variety of tasks ranging from essay writing and coding assistance to concept clarification and exam preparation.

This widespread adoption raises critical questions about the nature of student learning. While AI tools offer unprecedented accessibility to information and personalized assistance, they also risk fostering intellectual dependency, diminishing critical thinking, and enabling academic dishonesty. The tension between the favourable and unfavourable aspects of AI-assisted learning necessitates empirical investigation, mainly among masters' students who are expected to demonstrate higher-order thinking and independent research capabilities.

The present study focuses on two postgraduate streams: MCA (Master of Computer Applications), which is inherently technical and coding-intensive, and MMS (Master of Management Studies), which emphasizes analytical and strategic thinking. These two disciplines offer a meaningful comparative lens, as ChatGPT's utility may differ significantly across technical and managerial domains.

The study was conducted at DES's NMITD, where 80 students participated in a structured survey. To explore insights beyond descriptive statistics, three data mining and statistical techniques were applied: K-Means Clustering to identify distinct behavioral groups, Pearson Correlation Coefficient to examine variable relationships with SGPI, and Multiple Linear Regression to model the predictors of academic performance.

A. Objective of the Study

- To determine the frequency and purpose of ChatGPT usage among MCA and MMS students.
- To analyse the impact of ChatGPT on students' learning behaviour using Likert-scale assessments.
- To apply K-Means Clustering to identify distinct ChatGPT user behavioural profiles.
- To examine correlations between ChatGPT usage variables and SGPI using Pearson Correlation.
- To model predictors of academic performance using Multiple Linear Regression.
- To provide recommendations for responsible AI integration in postgraduate education.

B. Scope and Limitations: The study is limited to 80 students from a single institute. The SGPI data is self-reported and may contain recall bias. Causal inferences between ChatGPT usage and academic performance cannot be definitively established through a cross-sectional survey design. The regression model's R^2 of 0.114 suggests that SGPI is influenced by many factors beyond those captured in this survey.

II. Literature Review: The integration of AI technologies in educational settings have gained significant attention from researchers, particularly following the widespread release of ChatGPT in late 2022. Existing literature spans several dimensions, including AI's impact on learning outcomes, student behaviour, academic integrity, and pedagogical transformation.

A. AI and Learning Behaviour: Kasneji et al. [2] provide a comprehensive review of the challenges as well as opportunities that large language models present for education. They argue that while LLMs can serve as personalized tutors and scaffold complex concepts, there is a significant risk of students bypassing the cognitive struggle essential to deep learning. Baidoo-Anu and Ansah [3] found that over-reliance on ChatGPT may impair the development of independent thought, consistent with this study's findings on critical thinking reduction.

B. Academic Performance and AI Tools: Tlili et al. [4] noted that performance improvements associated with AI usage are often task-specific and context-dependent. Students who use AI for concept reinforcement show more sustainable gains compared to those who use it primarily for task completion. Zhai [5] found that AI writing assistants improved assignment quality but reduced students' ability to reproduce work independently—a pattern reflected in this study's regression and clustering findings.

C. Data Mining in Education (Educational Data Mining) Educational Data Mining (EDM) is a growing field that utilizes computational methods to analyze educational data and uncover meaningful patterns. [8]. Clustering method K-Means has been widely used to segment students into behavioural clusters based on learning patterns and engagement metrics [9]. Pearson Correlation used to identify meaningful associations between study habits and academic performance [10]. Multiple Linear Regression is commonly applied to model the predictive relationship between behavioural variables and GPA (performance metrics) [11]. The use of these techniques for analysing AI tool usage in academic is a rapidly growing area of research.

D. Research Gap: While substantial international research exists on ChatGPT in higher education, there is a notable gap in studies applying data mining techniques to examine ChatGPT usage patterns among Indian postgraduate students. The present study addresses this gap by applying K-Means Clustering, Pearson Correlation, and Multiple Linear Regression to survey data from 80 MCA and MMS students.

III. Research Methodology

A. Research Design: This study adopts a descriptive, quantitative, and analytical research design. A structured questionnaire served as the primary data collection instrument. Three analytical methods—K-Means Clustering, Pearson Correlation Coefficient, and Multiple Linear Regression—were applied to the encoded survey data to extract meaningful patterns and relationships.

B. Population and Sample: A convenience sample of 80 postgraduate students from DES's NMITD participated in the survey. Gender distribution: 47 males (58.75%) and 33 females (41.25%). Semester-wise: 51 in Semester II, 27 in Semester IV, and 2 in Semester III.

C. Data Collection Instrument: The questionnaire comprised 13 structured items: (1) demographic information; (2) ChatGPT usage patterns (frequency, purpose, daily time, answer verification); (3) four Likert-scale learning behavior statements; and (4) perceived performance impact and SGPI.

D. Data Encoding: Categorical and ordinal variables were numerically encoded prior to analysis as follows: Usage Frequency (Rarely=1, Sometimes=2, Frequently=3, Always=4); Time Spent (<30 min=1, 30-60 min=2, 1-2 hrs=3, >2 hrs=4); Answer Verification (Never=1, Sometimes=2, Always=3); Likert items (Strongly Disagree=1 to Strongly Agree=5); Perceived Performance (No=1, Not Sure=2, Yes=3). SGPI values above 10 or non-numeric (e.g., 'KT') were excluded, yielding 74 valid records for regression and correlation analyses.

E. Analytical Techniques: Three analytical techniques were applied: (1) K-Means Clustering on all eight encoded behavioural variables to identify student behavioural profiles; (2) Pearson Correlation Coefficient between each usage variable and SGPI to assess linear relationships; (3) Multiple Linear Regression with

SGPI as the dependent variable and usage and learning behaviour variables as independent predictors. All analyses were implemented in Python using the scikit-learn scipy, and pandas libraries.

IV. Data Analysis and Results

A. Demographic Profile of Respondents

Table 1: Demographic Profile of Respondents

Category	Group	Frequency	Percentage (%)
Stream	MCA	46	57.5%
	MMS	34	42.5%
Gender	Male	47	58.75%
	Female	33	41.25%
Semester	Semester II	51	63.75%
	Semester III	2	2.5%
	Semester IV	27	33.75%

B. ChatGPT Usage Patterns

Table 2: ChatGPT Usage Pattern Summary

Variable	Category	Count	%
Usage Frequency	Always	19	23.75%
	Frequently	45	56.25%
	Sometimes	13	16.25%
	Rarely	3	3.75%
Purpose	Concept Understanding	34	42.5%
	Exam Preparation	21	26.25%
	Coding Assistance	15	18.75%
	Assignments	9	11.25%
Time Spent/Day	Report Writing	1	1.25%
	<30 minutes	25	31.25%
	30-60 minutes	36	45.0%
	1-2 hours	13	16.25%
Verify Answers	>2 hours	6	7.5%
	Always	23	28.75%
	Sometimes	51	63.75%
	Never	6	7.5%

C. Learning Behaviour Analysis (Likert Scale)

Table 3: Likert Scale Learning behaviour Responses (SA=Strongly Agree, SD=Strongly Disagree)

Learning Behaviour Dimension	SA (5)	A (4)	N (3)	D (2)	SD (1)
ChatGPT improves understanding of subjects	16 (20%)	48 (60%)	14 (17.5%)	1 (1.25%)	1 (1.25%)
I depend on ChatGPT for assignments	8 (10%)	21 (26.25%)	27 (33.75%)	20 (25%)	4 (5%)
I solve problems myself before ChatGPT	12 (15%)	30 (37.5%)	29 (36.25%)	8 (10%)	1 (1.25%)
ChatGPT reduces critical thinking effort	17 (21.25%)	25 (31.25%)	23 (28.75%)	9 (11.25%)	6 (7.5%)

V. Data Mining Analysis

To extract deeper insights from the survey data, three data mining and statistical techniques were applied: K-Means Clustering, Pearson Correlation Coefficient, and Multiple Linear Regression. These techniques collectively provide a multi-dimensional understanding of how ChatGPT usage patterns relate to learning behaviour and academic performance.

A. K-Means Clustering

i. Algorithm Overview: Machine-learning, an unsupervised algorithm K-Means is partitions a dataset into K distinct, non-overlapping clusters by reducing the sum of squared distances between data points and their respective cluster centroids within each cluster. The algorithm iterates between two steps: (1) each data point is assigned to the nearest centroid, and (2) recalculating centroids as the mean of all assigned points. Convergence is achieved when assignments no longer change between iterations. The algorithm is formally expressed as minimizing the objective function:

$$J = \sum \sum ||x(i) - \mu_k||^2 \text{ for } i \in \text{cluster } k$$

Where $x(i)$ is a data point and μ_k is the centroid of cluster k . In this study, $K=3$ was selected based on the natural grouping expected among low, moderate, and high ChatGPT users. All eight encoded variables (usage frequency, time spent, answer verification, and four learning behaviour Likert items) were standardized using z-score normalization before clustering to prevent scale dominance.

ii. Cluster Profiles: K-Means clustering was applied to all 80 respondents using eight behavioral features. Three distinct student behavioral clusters were identified. Table 4 summarizes the mean values of key variables within each cluster.

Table 4: K-Means Cluster Profiles (K=3)

Variable	Cluster 0 (n=26)	Cluster 1 (n=18)	Cluster 2 (n=36)
Usage Frequency (1-4)	3.00	2.17	3.42
Time Spent/Day (1-4)	2.12	1.44	2.19
LB: Understanding (1-5)	3.42	4.22	4.22
LB: Dependency (1-5)	2.62	2.50	3.78
LB: Self-Solve (1-5)	3.15	4.39	3.42
LB: Critical Thinking (1-5)	3.15	4.00	3.44
Mean SGPI (out of 10)	8.11	8.04	8.22

iii. Cluster Interpretation

Cluster 0 — Moderate Dependent Users (n=26): This group uses ChatGPT frequently (mean=3.00) and spends moderate time on it daily. They show moderate scores across all learning behavior dimensions. Their mean SGPI of 8.11 places them below Cluster 2 but above Cluster 1. Said cluster denotes students who use ChatGPT regularly but have not fully incorporated it into either a dependency or a constructive learning pattern. **Cluster 1 — Selective Independent Learners (n=18):** This is the smallest cluster and exhibits the lowest usage frequency (mean=2.17) and time spent (mean=1.44). Notably, this group records the highest scores for self-directed problem solving (4.39) and critical thinking concern awareness (4.00), indicating self-aware, cautious users. Despite lower ChatGPT engagement, their SGPI (mean=8.04) is marginally lower, suggesting that selective, mindful usage does not always lead to better academic performance. **Cluster 2 — Heavy Dependent Users (n=36):** The largest cluster shows the highest usage frequency (mean=3.42) and the highest dependency score (mean=3.78), indicating significant reliance on ChatGPT for assignments. This group also records the highest mean SGPI (8.22), suggesting that heavy usage shows a slight positive relationship with academic performance in this cohort, possibly because students who use ChatGPT extensively are also more engaged in academic preparation overall. However, their high dependency score raises long-term learning concerns.

B. Pearson Correlation Coefficient

i. Method Overview

The Pearson Correlation Coefficient (r) evaluates how strongly and in what direction two continuous variables are linearly related.. It is defined as:

$$r = \frac{\sum[(x_i - \bar{x})(y_i - \bar{y})]}{\sqrt{[\sum(x_i - \bar{x})^2 \cdot \sum(y_i - \bar{y})^2]}}$$

The coefficient ranges from -1 (perfect negative correlation) to +1 (perfect positive correlation), with 0 indicating no linear relationship. A p-value < 0.05 indicates statistical significance. In this study, each encoded usage and learning behaviour variable was correlated with the students' last semester SGPI (N=74 valid records).

ii. Correlation Results

Table 5: Pearson Correlation Coefficients with SGPI (N=74)

Variable	Pearson r	p-value	Interpretation
Usage Frequency	0.1836	0.1174	Weak positive, not significant
Time Spent per Day	0.0281	0.8119	Negligible, not significant
Answer Verification	0.1751	0.1357	Weak positive, not significant
LB: Understanding	0.0763	0.5180	Negligible, not significant
LB: Dependency	-0.0735	0.5340	Weak negative, not significant
LB: Self-Solve	0.0494	0.6760	Negligible, not significant
LB: Critical Thinking	-0.1708	0.1457	Weak negative, not significant
Perceived Performance	0.1506	0.2003	Weak positive, not significant

iii. Interpretation

None of the variables achieved statistical significance at the $p < 0.05$ threshold. However, directional patterns are informative. Usage frequency ($r=0.1836$) and answer verification ($r=0.1751$) show the strongest positive associations with SGPI, suggesting that regular use of ChatGPT along with output verification is associated with marginally improved academic outcomes. Conversely, LB: Dependency ($r=-0.0735$) and LB: Critical Thinking reduction ($r=-0.1708$) show weak negative associations with SGPI, implying that students who make excessive use of ChatGPT and perceive it as reducing their critical thinking effort tend to have slightly lower academic performance. The absence of statistical significance is likely attributable to the restricted range of SGPI values (mean=8.15, SD=0.66) and the comparatively lesser sample size of 74, which limits statistical power.

C. Multiple Linear Regression

i. Model Specification: Multiple Linear Regression (MLR) was employed to model the relationship between multiple usage and behavioural predictors and SGPI as the outcome variable. The regression equation is expressed as:

$$SGPI = \beta_0 + \beta_1(Freq) + \beta_2(Time) + \beta_3(Verify) + \beta_4(LB1) + \beta_5(LB2) + \beta_6(LB3) + \beta_7(LB4) + \epsilon$$

where β_0 is the intercept, $\beta_1-\beta_7$ are regression coefficients for each predictor, and ϵ is the error term. The model was fitted using Ordinary Least Squares (OLS) on the 74 respondents with valid SGPI scores.

ii. Regression Results

Table 6: Multiple Linear Regression Results — Predictors of SGPI (N=74)

Predictor Variable	Coefficient (β)	Direction	Interpretation
Intercept (β_0)	6.9804	—	Baseline SGPI with all predictors at zero
Usage Frequency (β_1)	+0.2135	Positive	Higher frequency → higher SGPI
Time Spent/Day (β_2)	-0.0203	Negative	More time → slight SGPI decrease
Answer Verification (β_3)	+0.2010	Positive	Verifying answers → higher SGPI
LB: Understanding (β_4)	+0.0936	Positive	Better understanding → higher SGPI
LB: Dependency (β_5)	-0.0570	Negative	Higher dependency → slight SGPI decrease
LB: Self-Solve (β_6)	+0.0516	Positive	Self-solving habit → higher SGPI
LB: Critical Thinking (β_7)	-0.0754	Negative	CT reduction → lower SGPI
R-squared (R^2)	0.1140	—	11.4% variance in SGPI explained

iii. Interpretation

The above model achieved an R^2 of 0.114, indicating that the seven ChatGPT-related predictors describe 11.4% of the variance in SGPI. While modest, this is meaningful given the narrow range of SGPI variation in the sample. The two strongest positive predictors of SGPI are Usage Frequency ($\beta=+0.2135$) and Answer Verification ($\beta=+0.2010$). This proposes that students who use ChatGPT more frequently and critically evaluate its outputs tend to achieve higher academic performance—pointing to a qualitative dimension of AI tool usage that matters beyond mere frequency.

Conversely, LB: Critical Thinking Reduction ($\beta=-0.0754$) and LB: Dependency ($\beta=-0.0570$) are associated with lower SGPI, reinforcing concerns about passive AI reliance. Time Spent per Day ($\beta=-0.0203$) shows a slightly negative coefficient, suggesting that spending excessive time on ChatGPT may marginally detract from academic performance, possibly due to displacement of other study activities. The intercept of 6.98 represents the estimated baseline SGPI when all predictors are at their minimum encoded values, which is a plausible lower bound for the sample. Overall, the regression model confirms that the quality and intentionality of ChatGPT usage—reflected in verification behavior and purposeful frequency—are more important determinants of academic performance than raw usage volume.

VI. Discussion

A. Synthesis of Findings: The three analytical techniques collectively paint a nuanced picture of ChatGPT's role in student academic life. The descriptive findings confirm near-universal adoption (80% frequent/always users) and predominantly constructive use for concept understanding (42.5%). The K-Means clustering reveals that this cohort is not behaviourally homogeneous: three distinct user profiles exist, each with different usage intensities, learning behaviour patterns, and SGPI levels. Importantly, the cluster analysis challenges a simplistic narrative. The cluster with the highest SGPI (Cluster 2, mean=8.22) also exhibits the highest dependency score, while the cluster of selective, independent learners (Cluster 1) has a slightly lower SGPI (8.04). This finding suggests that in a performance-oriented academic culture, engagement with all available tools—including AI—may confer measurable advantages, even if it raises concerns about long-term intellectual development.

B. Implications of Regression and Correlation: The Pearson correlation and regression results consistently highlight answer verification as a critical differentiator. Students who critically engage with ChatGPT outputs—rather than passively accepting them—show stronger positive associations with SGPI. This is an actionable finding: teachers can foster verification habits by creating assignments that require students to provide explanations, justify, and cross-reference AI-provided information. The negative coefficient for critical thinking reduction ($\beta=-0.0754$) in the regression model, while small in magnitude, aligns with theoretical concerns raised in existing literature [2], [3] about the long-term cognitive costs of AI over-reliance. Although these costs are not yet fully reflected in SGPI, they may manifest in higher-level assessments, research capability, and professional performance.

C. Limitations of the Technical Analysis: The regression model's R^2 of 0.114 suggests that most of the variation in SGPI is explained by factors not captured in this survey—including prior educational performance and study behaviours unrelated to ChatGPT usage, teaching quality, and individual aptitude. The absence of statistical significance in Pearson correlations reflects both the restricted range of SGPI in this high-performing sample and the multidimensional nature of academic performance. Future investigations using larger and more varied samples and longitudinal designs would yield more statistically powerful findings.

VII. Conclusion

This research provides analytical and evidence-based findings on the role of ChatGPT in the educational activities of postgraduate MCA and MMS students. Using three complementary analytical techniques—K-Means Clustering, Pearson Correlation, and Multiple Linear Regression—the study moves beyond descriptive statistics to uncover behavioral groupings and performance predictors.

- K-Means Clustering identified three distinct user profiles: Moderate Dependent Users (Cluster 0, n=26), Selective Independent Learners (Cluster 1, n=18), and Heavy Dependent Users (Cluster 2, n=36). The heavy users showed the highest SGPI (8.22) but also the highest dependency, revealing a performance-engagement trade-off.
- Pearson Correlation revealed weak positive associations between usage frequency ($r=0.184$) and answer verification ($r=0.175$) with SGPI, and weak negative associations for dependency and critical thinking reduction, consistent with theoretical expectations.
- Multiple Linear Regression ($R^2=0.114$) identified usage frequency and answer verification as the strongest positive predictors of SGPI, while dependency and critical thinking reduction were associated with lower performance.
- Overall, 80% of students use ChatGPT frequently, 61.25% believe it improves academic performance, and the mean SGPI of 8.15 reflects a high-performing cohort.

A. Recommendations

- Institutions should mandate AI literacy programmes covering model limitations, hallucination risks, and critical evaluation of AI outputs.
- Educators should redesign assessments to include oral defence, in-class problem solving, and tasks requiring independent reasoning that AI cannot substitute.
- Instructors should encourage students to emulate Cluster 1 habits—attempting problems independently before seeking AI assistance—while leveraging the engagement energy of Cluster 2.
- Further studies should utilize larger stratified samples and longitudinal experimental methods to better understand the causal relationship between ChatGPT use and students' academic performance.

References

1. A. Hu, "ChatGPT sets record for fastest-growing user base," Reuters, Feb. 2023.
2. E. Kasneci et al., "ChatGPT for good? On opportunities and challenges of large language models for education," *Learn. Individ. Differ.*, vol. 103, p. 102274, Apr. 2023.
3. D. Baidoo-Anu and L. O. Ansah, "Education in the era of generative AI: Understanding the potential benefits of ChatGPT in promoting teaching and learning," *J. AI*, vol. 7, no. 1, pp. 52–62, 2023.
4. A. Tlili et al., "Is ChatGPT a good teacher coach? Measuring zero-shot performance for scoring and providing actionable feedback," *Human Behav. Emerg. Technol.*, vol. 2023, p. 4237894, 2023.
5. X. Zhai, "ChatGPT user experience: Implications for education," *SSRN Electron. J.*, 2022. doi: 10.2139/ssrn.4312418.
6. I. Dergaa et al., "From human writing to AI generated text: Examining threats of ChatGPT in academic writing," *Biol. Sport*, vol. 40, no. 2, pp. 615–622, 2023.
7. M. Farrokhnia et al., "A SWOT analysis of ChatGPT: Implications for educational practice and research," *Innov. Educ. Teach. Int.*, pp. 1–15, 2023.
8. R. S. J. D. Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions," *J. Educ. Data Min.*, vol. 1, no. 1, pp. 3–17, 2009.
9. C. Romero and S. Ventura, "Educational data mining: A review of the state of the art," *IEEE Trans. Syst. Man Cybern. C*, vol. 40, no. 6, pp. 601–618, Nov. 2010.
10. A. Dutt, M. A. Ismail, and T. Herawan, "A systematic review on educational data mining," *IEEE Access*, vol. 5, pp. 15991–16005, 2017.
11. S. K. Yadav and S. Pal, "Data mining: A prediction for performance improvement of engineering students using classification," *World Comput. Sci. Inf. Technol. J.*, vol. 2, no. 2, pp. 51–56, 2012.

Appendix: Survey Questionnaire

The following questionnaire was administered to 80 MCA and MMS students:

Section A: Demographic Information

- Q1. Select your stream: MCA MMS
Q2. Select Semester: II III IV
Q3. Select Gender: Male Female

Section B: ChatGPT Usage Patterns

- Q4. How often do you use ChatGPT for academic purposes?
 Always Frequently Sometimes Rarely
Q5. For what purpose do you use ChatGPT?
 Concept Understanding Exam Preparation Coding Assignments Report Writing
Q6. Average time spent per day on ChatGPT:
 <30 mins 30–60 mins 1–2 hrs >2 hrs
Q7. Do you verify ChatGPT answers?
 Always Sometimes Never

Section C: Learning Behavior (5-Point Likert Scale: SA / A / N / D / SD)

- Q8. ChatGPT improves my understanding of subjects.
Q9. I depend on ChatGPT to complete assignments.
Q10. I try to solve problems myself before using ChatGPT.
Q11. ChatGPT reduces my critical thinking effort.

Section D: Academic Performance

- Q12. Do you think ChatGPT improves academic performance? Yes No Not Sure
Q13. Last semester SGPI: _