1. Figure Legends

1.1 Tables

Table 1. Shapiro-Wilk Test of Normality: Total Students vs. ChatGPT-3.5

Tests of Normality							
	Students and ChatGPT-3.5	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Total Correct Answers	Students	.059	139	.200*	.991	139	.528
	ChatGPT-3.5	.148	76	<.001	.948	76	.004
*. This is a lower bound of the true significance.							
a. Lilliefors Significance Correction							

Table 2. Independent-Samples Mann-Whitney U Test summary: Total Students vs. ChatGPT-3.5

Independent-Samples Mann-Whitney U Test Summary	
Mann-Whitney U	10042.000
Wilcoxon W	12968.000
Z	10042.000
Standard Error	435.341
Standardized Test Statistic	10.934
Asymptotic Sig.(2-sided test)	<.001

Table 3: Shapiro-Wilk Test of Normality: Year 4 vs. Year 5 vs. ChatGPT-3.5

Tests of Normality						
	Subjects	Kolmogorov-Smirnov[a]			Shapiro-Wilk	
		Statistic	df	Sig.	Statistic	df
Total Correct Answers	Year 4	.078	67	.200*	.987	67
	Year 5	.094	72	.194	.978	72
	ChatGPT-3.5	.148	76	<.001	.948	76

Table 4: Additional Shapiro-Wilk normality tests: Year 4 vs. Year 5 vs. ChatGPT-3.5

Tests of Normality		
	Subject	Shapiro-Wilk[a]
		Sig.
Total Correct Answers	Year 4	.704
	Year 5	.233
	Chat GPT	.004
*. This is a lower bound of the true significance.		
a. Lilliefors Significance Correction		

Table 5: Additional Hypothesis test summary for Independent-Samples Kruskal-Wallis Test: Year 4 vs. Year 5 vs. ChatGPT-3.5

Hypothesis Test Summary	
	Decision
1	Reject the null hypothesis.
a. The significance level is .050.	
b. Asymptotic significance is displayed.	

Table 6. Independent-Samples Kruskal-Wallis Test Summary: Year 4 vs. Year 5 vs. ChatGPT-3.5

Independent-Samples Kruskal-Wallis Test Summary	
Total N	215
Test Statistic	122.445[a]
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	<.001
a. The test statistic is adjusted for ties.	

Table 7: Pairwise comparisons using Dunn's test with Bonferroni correction: Year 4 vs. Year 5 vs. ChatGPT-3.5

Pairwise Comparisons of YearAndChatGPT-3.5					
Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Year 5- Year 4	17.934	10.542	1.701	.089	.267
Year 5- Chat GPT	-105.520	10.214	-10.331	<.001	.000
Year 4- Chat GPT10	-87.587	10.408	-8.416	<.001	.000
Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Asymptotic significance (2-sided tests) are displayed. The significance level is .050.					

Table 8: Kruskal-Wallis Test Ranks: Year 4 vs. Year 5 vs. ChatGPT-3.5

Ranks			
	YearAndChatGPT-3.5	N	Mean Rank
TotalCA	4	67	83.04
	5	72	65.11
	Chat GPT	76	170.63
	Total	215	

Table 9: Shapiro-Wilk Test of Normality for General Surgery: Total Students vs. ChatGPT-3.5

Tests of Normality							
	Students vs ChatGPT-3.5	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
GS	Students	.101	139	.001	.975	139	.011
	ChatGPT-3.5	.210	76	<.001	.917	76	<.001
a. Lilliefors Significance Correction							

Table 10: Independent-Samples Mann-Whitney U Test analysis for General Surgery: Total Students vs. ChatGPT-3.5

Independent-Samples Mann-Whitney U Test Summary	
Mann-Whitney U	9170.000
Wilcoxon W	12096.000
Test Statistic	9170.000
Standard Error	433.043
Standardized Test Statistic	8.978
Asymptotic Sig.(2-sided test)	<.001

Table 11: Shapiro-Wilk Test of Normality for General Medicine: Total Students vs. ChatGPT-3.5

Tests of Normality							
	Students vs ChatGPT-3.5	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
GM	Students	.131	139	<.001	.956	139	<.001
	ChatGPT-3.5	.216	76	<.001	.913	76	<.001
a. Lilliefors Significance Correction							

Table 12: Independent-Samples Mann-Whitney U Test analysis for General Medicine: Total Students vs. ChatGPT-3.5

Independent-Samples Mann-Whitney U Test Summary

Mann-Whitney U	9637.500
Wilcoxon W	12563.500
Test Statistic	9637.500
Standard Error	433.409
Standardized Test Statistic	10.049
Asymptotic Sig.(2-sided test)	<.001

Table 13: Shapiro-Wilk Test of Normality for Traumatology: Total Students vs. ChatGPT-3.5

Tests of Normality							
	Students vs ChatGPT-3.5	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Trauma	Students	.174	139	<.001	.930	139	<.001
	ChatGPT-3.5	.246	76	<.001	.866	76	<.001
a. Lilliefors Significance Correction							

Table 14: Independent-Samples Mann-Whitney U Test analysis for Traumatology: Total Students vs. ChatGPT-3.5

Independent-Samples Mann-Whitney U Test Summary	
Mann-Whitney U	9038.500
Wilcoxon W	11964.500
Test Statistic	9038.500
Standard Error	427.702
Standardized Test Statistic	8.783
Asymptotic Sig.(2-sided test)	<.001