Generative AI Usage Among Different Disciplines at UTM

Khizar Tariq, Rohan Nair, Preet Patel, Nipun Jaiswal, Baraa Shamout
Group Name: Default Name
STA304: Surveys, Sampling and Observational Data

November 11 2024

1 Abstract

Over the past few years, generative AI has emerged as a groundbreaking tool that taken the world of education by storm. In fact, according to a study conducted in Fall of 2023 by Harvard, 51% of young people report having tried generative AI tools for education at some point (Harvard Graduate School of Education. (2024). Teen and young adult perspectives on generative AI). However, the discourse about the usage of generative AI in education is often very polarizing. While some people consider generative AI to be an invaluable resource for learning and completing tasks, others consider it to be a sham. This study aims to analyze a key factor of what makes the discourse polarizing: the discipline generative AI is being used for. To analyze this vital factor, we surveyed students in a third-year statistics surveys and sampling course about their usage of generative AI two disciplines: statistics and the non-statistics discipline they have taken the most courses of. The survey was conducted through a physical questionnaire and Google Forms. Also for simplicity, non-statistic disciplines were grouped into Computer Science, Math, Economics and Other. The results of our study showed that Economics students rated generative AI to be the most useful for grasping course content, Computer Science students reported the greatest proportion of generative AI assistance for assignment completion, and Economics students prefer generative AI the most for help with course content. In response, we can conclude that generative AI is perceived as less useful in explaining theoretical concepts, the tasks in Computer Science are more easily assistable by generative AI, and Economics teaching is the most replaceable by generate AI.

2 Introduction

AI technology integration into educational settings has drawn more attention recently as educators look for creative ways to improve learning outcomes and experiences (Tadiwos et al. 2024). By adapting to the unique demands of different disciplines, generative AI can enable students to gain practical experience, explore ideas more freely, and develop a deeper understanding of their field. Generative AI provides valuable assistance tailored to each discipline's needs: in Computer Science, it supports coding and debugging; in Mathematics, it visualizes abstract ideas and solves equations; in Economics, it enables data analysis and market simulations; and in other fields it can support creative writing, generate ideas for projects, and offer personalized study resources, analysis, creative ideation, and personalized study resources.

Our study explores the perceived usefulness of generative AI as an educational tool across academic disciplines. By examining students' experiences of different academic disciplines on using generative AI in their learning, we aim to understand how AI can enrich the educational experience and help students gain practical skills. To examine these impacts, we conducted a survey in October 2024 with students from the STA304H5 undergraduate course at the University of Toronto Mississauga. The survey was administered through both in-class questionnaires and Google Forms. The survey included students of diverse disciplines ranging from Computer Science, Maths, Economics, etc for comparisons

We aim to study the following research questions:

1. (RQ1) Does the perceived usefulness of generative AI for grasping course content differ across different disciplines?
    - Null Hypothesis: The average of the perceived usefulness of AI by students in the Math and Computer Science discipline is equal to that of students in Economics and other disciplines.
    - Alternative Hypothesis: The average of the perceived usefulness of AI by students in the Math and Computer Science discipline is not equal to that of students in Economics and other disciplines.
2. (RQ2) How does the proportion of assessments completed with the assistance of generative AI differ across academic disciplines?
    - Null Hypothesis: The proportion of assessments completed by AI by students in the Math and Computer Science discipline is equal to that of students in Economics and other disciplines.
    - Alternative Hypothesis: The proportion of assessments completed by AI by students in the Math and Computer Science discipline is not equal to that of students in Economics and other disciplines.
3. (RQ3) Does the preference of using generative AI compared to traditional resources for helping with course content differ across academic disciplines?
    - Null Hypothesis: The preference of using generative AI compared to traditional resources for helping with course content does not differ between [CS and Math] students and [Economics and Other].
    - Alternative Hypothesis: The preference of using generative AI compared to traditional resources for helping with course content does differ between [CS and Math] students and [Economics and Other].

    The structure of the paper is as follows: Section 3 outlines our data collection methods. Section 4 presents our quantitative analysis. Section 5 discusses the qualitative results and addresses our research question. Section 6 covers the limitations of our study. Section 7 concludes our analysis and finally section 8 is the appendix.

3 Methodology

Between October 7th and October 21st, 2024, a survey was conducted to assess the usefulness of generative AI for grasping course content and its role in assignment completion across different disciplines. The survey also aimed to compare the usage of generative AI with traditional academic resources. We used simple random sampling (SRS) to select participants for the survey. To ensure a diverse and representative sample, we attended both lecture sessions to include as many students as possible. Additionally, the survey was made available online to accommodate students who could not attend lectures in person. The final sample size was 55 students. The survey consisted of 9 multiple-choice questions asking participants for their discipline information, view on the usefulness of AI (for statistics courses and other disciplines), confidence in AI accuracy, frequency of AI usage for assignments, and preference for academic help sources.

4 Analysis

For our study, our population size is N = 200 and we plan to collect data using simple random sampling. To determine a sample size to collect, we'll go with the calculation focusing on the mean population parameter. It is assumed that there is an equal proportion among those who think generative AI is highly helpful(greater than or equal to 6 on a scale of 1 to 10) and who think generative AI is hardly

useful(less than or equal to 5 on a scale of 1 to 10). Hence, given a bound of error of 0.116, the sample size calculation ends up being like:

$$\frac{Npq}{(N-1)D+pq} = \frac{(200)(0.5)(0.5)}{(199)\frac{0.116^2}{4}+(0.5)(0.5)} = 54.38 \approx 55$$

Hence, we sampled 55 students for our analysis

We found that 36.4% of our participants took most of their courses from "CS" discipline ($n = 20$), 29% of our participants took most of their courses from "Math" discipline ($n = 16$), 16.4%of our participants took most of their courses from "Economics" discipline ($n = 9$), and 18.2% of our participants took most of their courses from other disciplines ($n = 10$).

To Answer RQ1, we use the Mann-Whitney U Test (for two groups) as our smaller data frames (for example students whose discipline is math and cs) are not normally distributed, so we can not use t test. Additionally, to check if variances are the same, we have to use Levene Test, as we have different data frames that are not normally distributed. The findings are in the table below

	p-value levene test	**p-value wilcox test**
Course Content	0.924	0.3676

Table 1: **Table showing p-values for Levene's test and Mann-Whitney U test for the difference of means of [CS and Math] students perceived usefulness of AI and [Economics and Other] students perceived usefulness of AI for grasping course content**

We also carried out ANOVA(Analysis Of Variances) to see how different the variances of each discipline was. We found the F-values and P-values. Their results are given in the table below

	p-value	**F-value**
CourseContent	0.44	0.908

Table 2: **ANOVA Results for relation between discipline and perceived usefulness of AI for help in Course Content**

To answer RQ2, we use the 2-sample test for equality of proportions of assignments done with the help of AI between [CS and Math] and [Economics and Other] students.

We meet the assumptions np and $n(1-p)$ where p is the proportion is > 5 (for [CS and Math], $np = 36*0.5 = 18 > 5$, $nq = 36*0.5 = 18 > 5$ and for [Economics and Other] $np = 19*0.45 = 8.55 > 5$ and $nq = 19*0.55 = 10.45 > 5$).

We found the Proportions of the two strata, the Chi-Squared statistic, degrees of freedom, 95% Confidence Interval, and p-values for the proportion test. The results are in the table given below

	Prop [CS MATH]	Prop [ECO OTHER]	X-squared	df	LCL	UCL	p-value
GradedAssignments	0.50	0.45	0.7	1	-0.07	0.17	0.4
NonGradedAssignments	0.58	0.57	0.02	1	-0.11	0.13	0.88

Table 3: **Comparison of Proportion of Assignments Completed by AI between [CS and Math] and [Economics and Other] Students Test Results**

To answer RQ3: We carry out the 2-sample test for equality of proportions of students who either only ask generative AI for help in a course or ask generative AI before asking TA's, professors, piazza etc between [CS and Math] students and [Economics and Other] students. We obtain a Chi-square test statistic of 0.48, df = 1, and p = 0.49. We also got a 95% Confidence Interval of (-0.17, 0.35)

5 Results

We found the p-value of the Mann-Whitney U Test (for two groups) for the difference of perceived usefulness of AI for grasping course content between the following two groups: the first being Computer Science and Math students, the second being Economics and Other students to be 0.3676. Both of these values are greater than 0.05(95% CI). So we can not reject the null hypothesis for RQ1.

Additionally, the p-value for ANOVA was found to be 0.44 for perceived usefulness of generative AI for grasping course content with relation to discipline. As we see, these p-values are greater than 0.05, which means we can not reject the fact that different disciplines have different variances around the mean of perceived usefulness of AI.

The p-value for Logistic regression between the perceived usefulness of AI for grasping course content vs each discipline was found to be greater than 0.05 (Tables 3 and 4), so we can not reject the fact that there is no relation between discipline and perceived usefulness of AI

We found the p-value of 2-sample test for equality of proportions of **graded** assignments done with the help of AI between [CS and Math] and [Economics and Other] students to be 0.4 which is greater than 0.05
We found the p-value of 2-sample test for equality of proportions of **non-graded** assignments done with the help of AI between [CS and Math] and [Economics and Other] students to be 0.88 which is greater than 0.05

As both values are greater than 0.05, we can not reject the null hypothesis for RQ2

We found the p-value of 2-sample test for equality of proportions of students who only ask generative AI or ask generative AI before asking TA's, professors, piazza, etc. between [CS and Math] and [Economics and Other] students to be 0.49 which is greater than 0.05
As a result, we can not reject the null hypothesis for RQ3

6 Limitations

In our study, we tried to see how students across different disciplines perceive the usefulness of generative AI, both in their statistics and non statistics courses, and how often students use generative AI for their coursework using Mann-Whitney U Test (for two groups), proportion test, logistic regression and analysis of variances.

Unfortunately, in all our cases, we failed to reject the null hypothesis, so there is no clear relation between a students discipline and their perceived usefulness of generative AI at a 95% Confidence Interval. Most of the students in our sample(STA304 Students) were CS and Math students, and as a result we were limited to only 3 disciplines with more than two people in them. Also, we should have asked questions regarding the proportions of assignments completed with the help of generative AI for students statistics courses, which we unfortunately failed to do. Additionally, a sample size of 56 meant that while comparing two sub data frames (for example CS and Economics), the data was almost never normal and was very small, thus leading to huge variances and no use of t tests.

7 Conclusion

Our study demonstrates that we do not have enough evidence to say that students perceptions and usage of generative AI varies greatly across different academic disciplines. While some disciplines found generative AI useful for understanding course content, others had tasks which are more assistable by AI, and others had a stronger preference for AI help, but none of these preferences were significantly greater than the other. With the conclusion of our study, we can finally answer the research questions

we set out to answer at the start of our study. Through our study we found that we can not reject the fact that the perceived usefulness of generative AI for grasping course content does not differ across different disciplines. We also discovered that we can not reject the fact that the proportion of assessments completed by AI by students in the Math and Computer Science discipline is equal to that of students in Economics and other disciplines. Finally, we determined that we can not reject the fact that the preference of using generative AI compared to traditional resources for helping with course content does not differ between the following groups: the first being Computer Science and Math students, the second being Economics and other students. As generative AI will continue to make a significant impact in education, we know there will be further research regarding this topic, and we have a few suggestions for these future studies. In the future, researchers could increase their sample size and their target population, to get data from more people of different disciplines. Increase in sample size would also help in doing stratified random sampling, given there is enough data in each strata (discipline).

8 Appendix

Harvard Graduate School of Education. (2024). Teen and young adult perspectives on generative AI.

Hiwot, G., Namuduri, S. (2024). The impact of a Large Language Model (LLM): A qualitative study on how students and educators perceive the use of LLMs such as ChatGPT within conventional university education dynamics (Master's thesis). Linnaeus University, Faculty of Technology, Department of Informatics.

```
#changing work directory
setwd("/Users/nipunjaiswal/Downloads")

#Installing libraries
install.packages("readxl")
install.packages("car")

#read the excel file
library(readxl)
df = read_excel("Questionnarie (Responses).xlsx")

#############################################################
# Splitting our data into relevant sub data frames
#############################################################

df_CS <- df[df$Discipline == "CS", ]
df_Math <- df[df$Discipline == "Math", ]
df_Economics <- df[df$Discipline == "Economics", ]
df_Other <- df[df$Discipline == "Other", ]

df_CS_Math <- df[df$Discipline %in% c("CS", "Math"), ]
df_Economics_Other <- df[df$Discipline %in% c("Economics", "Other"), ]


#############################################################################
# Sample Size Computations (SRS)
#############################################################################

N = 200
p = 0.5
q = 1-p
B = 0.116

D = B^2 / 4
(N * p * q) / ((N-1) * D + p * q)

#############################################################################
# Initial Data Analysis
#############################################################################

nrow(df_CS)/nrow(df) #Proportion of students with "CS" Discipline

nrow(df_Math)/nrow(df) #Proportion of students with "Math" Discipline

nrow(df_Economics)/nrow(df) #Proportion of students with "Economics" Discipline

nrow(df_Other)/nrow(df) #Proportion of students with "Other" Discipline


#############################################################################
#(RQ1) Does the perceived usefulness of generative
#AI for grasping course content differ across different disciplines?
#############################################################################
```

```
##############################################################################
#COMPARISON OF MEANS FOR PERCIEVED USEFULNESS FOR GRASPING COURSE CONTENT
#BETWEEN STUDENTS OF [CS AND MATH] AND [ECONOMICS AND OTHER] STUDENTS
##############################################################################

shapiro.test(df_Economics_Other$AIDisciplineCourseContent) #p-value < 0.05 so not normal.
#Proceed to use wilcox test (Mann-Whitney U Test (for two groups))

df_combined <- data.frame( value = c(df_CS_Math$AIDisciplineCourseContent,
df_Economics_Other$AIDisciplineCourseContent),
group = rep(c("Group1", "Group2"),
times = c(length(df_CS_Math$AIDisciplineCourseContent),
length(df_Economics_Other$AIDisciplineCourseContent))) )

# Load the car package
library(car)

# Perform Levene's test for equal variances
leveneTest(value ~ group, data = df_combined) # Checking if variances are same
#(if p-value>0.05, then they are approximately equal).
# We find p-value>0.05 so they are equal variances
wilcox.test(df_CS_Math$AIDisciplineCourseContent, df_Economics_Other$AIDisciplineCourseContent,
alternative = "two.sided", var.equal = TRUE)

#LOGISTIC REGRESSION
#Logistic Regression between predictor discipline and percieved usefulness of AI for course content
df$Discipline <- as.factor(df$Discipline)
model <- glm(AIDisciplineCourseContent/10 ~ Discipline, data = df,)
summary(model)

#ANOVA
anova_results = aov(df$AIDisciplineCourseContent ~ df$Discipline)
summary(anova_results)

boxplot(df$AIDisciplineCourseContent ~ df$Discipline,
data = df, main = "Boxplot of AIDisciplineCourseContent by Discipline",
xlab = "Discipline", ylab = "AIDisciplineCourseContent",
col = c("lightblue", "lightgreen", "lightcoral", "lightyellow"),
border = "black")

stripchart(df$AIDisciplineCourseContent ~ df$Discipline,
data = df, method = "jitter", # Jitters points to prevent overlap
,vertical = TRUE, # Draws the chart vertically
pch = 16, # Symbol type
col = c("blue", "green", "coral", "red"), # Colors for each group
main = "Usefulness of AI by Discipline", xlab = "Discipline", ylab = "Perceived Usefulness of AI")

##############################################################################
#COMPARISON OF MEANS FOR PERCIEVED USEFULNESS FOR HELPING IN ASSIGNMENTS
#BETWEEN STUDENTS OF [CS AND MATH] AND [ECONOMICS AND OTHER] STUDENTS
##############################################################################

shapiro.test(df_Economics_Other$AIDisciplineAssignment) #p-value < 0.05 so not normal.
```

```
#Proceed to use wilcox test (Mann-Whitney U Test (for two groups))

df_combined_2 <- data.frame( value = c(df_CS_Math$AIDisciplineAssignment,
df_Economics_Other$AIDisciplineAssignment),
group = rep(c("Group1", "Group2"),
times = c(length(df_CS_Math$AIDisciplineAssignment),
length(df_Economics_Other$AIDisciplineAssignment))) )

# Load the car package
library(car)

# Perform Levene's test for equal variances
leveneTest(value ~ group, data = df_combined_2) # Checking if variances are same
#(if p-value>0.05, then they are approximately equal).
#We find p-value>0.05 so they are equal variances

wilcox.test(df_CS_Math$AIDisciplineAssignment, df_Economics_Other$AIDisciplineAssignment,
alternative = "two.sided", var.equal = TRUE)

#LOGISTIC REGRESSION
#Logistic Regression between predictor discipline and percieved usefulness of AI for course content
df$Discipline <- as.factor(df$Discipline)
model <- glm(AIDisciplineAssignment/10 ~ Discipline, data = df,)
summary(model)

#ANOVA
anova_results = aov(df$AIDisciplineAssignment ~ df$Discipline)
summary(anova_results)
boxplot(df$AIDisciplineAssignment ~ df$Discipline, data = df,
main = "Boxplot of AIDisciplineCourseContent by Discipline",
xlab = "Discipline", ylab = "AIDisciplineCourseContent",
col = c("lightblue", "lightgreen", "lightcoral", "lightyellow"),
border = "black")

stripchart(df$AIDisciplineAssignment ~ df$Discipline,
data = df, method = "jitter", # Jitters points to prevent overlap
,vertical = TRUE, # Draws the chart vertically
pch = 16, # Symbol type
col = c("blue", "green", "coral", "red"), # Colors for each group
main = "Usefulness of AI by Discipline", xlab = "Discipline", ylab = "Perceived Usefulness of AI")

########################################################################


# OVERALL, THERE IS NO WAY TO REJECT THE FACT THAT THE MEAN PERCIEVED USEFULNESS OF AI
# BY [CS AND MATH] STUDENTS IS THE SAME AS [ECONOMICS AND OTHER] STUDENTS
# FOR BOTH COURSE CONTENT AND ASSIGNMENTS

########################################################################
# (RQ2) How does the proportion of assessments completed with the assistance
#  of generativeAI differ across academic disciplines?
########################################################################

# Proportion of graded assignments done with help of AI
```

```
#for [CS AND MATH] students
mean(df_CS_Math$AIGraded)/5

# Proportion of graded assignments done with help of AI
#for [ECONOMICS AND OTHER] students
mean(df_Economics_Other$AIGraded)/5

#2-sample test for equality of proportions of graded assignments done
#with help of AI for [CS AND MATH] students and Proportion of graded
#assignments done with help of AI for [ECONOMICS AND OTHER] students
prop.test(c(mean(df_CS_Math$AIGraded)*36, mean(df_Economics_Other$AIGraded)*19),
c(5*36,5*19), conf.level = 0.95, correct = FALSE)

# Proportion of non graded assignments done with help of AI
#for [CS AND MATH] students
mean(df_CS_Math$AINonGraded)/5
# Proportion of non graded assignments done with help of AI
#for [ECONOMICS AND OTHER] students
mean(df_Economics_Other$AINonGraded)/5

# 2-sample test for equality of proportions of non graded assignments done
# with help of AI for [CS AND MATH] students and # Proportion of graded
#assignments done with help of AI for [ECONOMICS AND OTHER] students
prop.test(c(mean(df_CS_Math$AINonGraded)*36, mean(df_Economics_Other$AINonGraded)*19),
c(5*36,5*19), conf.level = 0.95, correct = FALSE)

#########################################################################
# (RQ3) Does the preference of using generative AI compared to traditional resources for
#helping with course content differ across academic disciplines?
#########################################################################

#Proportion of CS and Math students who prefer generative AI over
#traditional sources like Ta's, Professors, piazza, etc.
sum(sum(df_CS_Math$AskFirst == "I only ask generative AI for help on course content"),
sum(df_CS_Math$AskFirst == "I ask generative AI before I ask the professors, TA's, piazza, etc"))/3

#Proportion of Economics and Other students who prefer generative AI over
#traditional sources like Ta's, Professors, piazza, etc.
sum(sum(df_Economics_Other$AskFirst == "I only ask generative AI for help on course content"),
sum(df_Economics_Other$AskFirst == "I ask generative AI before I ask the professors, TA's, piazza,

# 2-sample test for equality of proportions between: CS and Math students
#who prefer generative AI over traditional sources like Ta's, Professors, piazza, etc. and
#Proportion of Economics and Other students who prefer generative AI over
#traditional sources like Ta's, Professors, piazza, etc.
prop.test(c(sum(sum(df_CS_Math$AskFirst == "I only ask generative AI for help on course content"),
sum(df_CS_Math$AskFirst == "I ask generative AI before I ask the professors, TA's, piazza, etc")),
sum(sum(df_Economics_Other$AskFirst == "I only ask generative AI for help on course content"),
sum(df_Economics_Other$AskFirst == "I ask generative AI before I ask the professors, TA's, piazza,
c(36,19), conf.level = 0.95, correct = FALSE)
```

9 Generative AI Usage In Our Paper

1. Used ChatGPT for certain R code
    - ChatGPT prompt: "Boxplot for column name AIDisciplineCourseContent on the y axis and 4 different academic disciplines on the x axis"
    - ChatGPT prompt: "Stripchart for column name AIDisciplineCourseContent on the y axis and 4 different academic disciplines on the x axis"
    - ChatGPT prompt: "Logistic Regression Code in R"
    - ChatGPT prompt: "ANOVA code in R"
    - ChatGPT prompt: "How to test comparison of means for non normal data without using t test"
    - ChatGPT prompt: "How to test for equality of variances for different non equal and non normal data frames"
2. Used ChatGPT for paraphrasing Methodology section
    - Methodology before AI

      Between October 7th to October 21st, 2024, a survey was conducted to measure the usefulness of generative AI for grasping course content and the reliance of student in assignment completion across different disciplines, and to compare generative AI's usage versus traditional academic resources. We used simple random sampling (SRS) as our sampling method. To ensure randomness, we attended both lecture times to include as many students as possible in our study. Additionally, we posted the survey online to reach students who were unable to attend lectures in person. Our final sample size is 55 students. The survey consisted of 9 multiple choice questions asking students for the following information: usefulness of AI for statistics courses, discipline outside of statistics, usefulness of AI for other discipline, confidence in AI accuracy, frequency of AI usage for graded/ungraded assignments, and preference for asking AI versus traditional resources.

    - Methodology after AI

      Between October 7th and October 21st, 2024, a survey was conducted to assess the usefulness of generative AI for grasping course content and its role in assignment completion across different disciplines. The survey also aimed to compare the usage of generative AI with traditional academic resources. We used simple random sampling (SRS) to select participants for the survey. To ensure a diverse and representative sample, we attended both lecture sessions to include as many students as possible. Additionally, the survey was made available online to accommodate students who could not attend lectures in person. The final sample size was 55 students. The survey consisted of 9 multiple-choice questions asking participants for their discipline information, view on the usefulness of AI (for statistics courses and other disciplines), confidence in AI accuracy, frequency of AI usage for assignments, and preference for academic help sources.
3. Used ChatGPT for research for Introduction section
4. Used ChatGPT for LaTeX
    - ChatGPT prompt "How to create a 2 by 4 table in LaTeX"