Student Use of Large Language Models (LLMs) and Understanding of Artificial Intelligence (AI)

Annabelle Warner
University of Utah

UUCS-25-004

School of Computing
University of Utah
Salt Lake City, UT 84112 USA

16 April 2025

Abstract

The increasing use of Large Language Models (LLMs) by students in coursework underscores the need for educational systems to adapt, ensuring that students are equipped to use these tools effectively and responsibly. Artificial Intelligence (AI) literacy is a measure of a user's competency surrounding the use, understanding of, and evaluation of AI. Understanding student AI literacy and how students are using AI can offer valuable insights into how educational institutions should adapt in an increasingly AI-driven society. Using existing AI literacy measures, this study explored the relationship between student AI literacy and LLM use. A survey conducted with students at a large, public university (n = 80) revealed significant positive correlations between student AI literacy and self-reported use of large language models (LLMs). Demographic factors, such as student major and gender, were found to influence these correlations and differences in averages among subgroups. The results confirmed that AI literacy levels vary among students from different backgrounds and demographics, yet most students are reporting use of AI tools. This research will aid educational institutions in understanding the current state of student AI literacy and LLM use and provide a starting point for preparing students to use AI tools purposefully.

STUDENT USE OF LARGE LANGUAGE MODELS (LLMs) AND UNDERSTANDING OF ARTIFICIAL INTELLIGENCE (AI)

by

Annabelle Warner

A Senior Honors Thesis Submitted to the Faculty of
The University of Utah
In Partial Fulfillment of the Requirements for the

Honors Degree in Bachelor of Science

In

Data Science

Approved:

Sameer Patil

Sameer Patil
Faculty Thesis Mentor

Mary Hall

Mary Hall
Chair, School of Computing

Thomas C Henderson

Tom Henderson
Departmental Honors Liaison

Monisha Pasupathi, PhD
Dean, Honors College

May 2025

Abstract

The increasing use of Large Language Models (LLMs) by students in coursework underscores the need for educational systems to adapt, ensuring that students are equipped to use these tools effectively and responsibly. Artificial Intelligence (AI) literacy is a measure of a user's competency surrounding the use, understanding of, and evaluation of AI. Understanding student AI literacy and how students are using AI can offer valuable insights into how educational institutions should adapt in an increasingly AI-driven society. Using existing AI literacy measures, this study explored the relationship between student AI literacy and LLM use. A survey conducted with students at a large, public university (n = 80) revealed significant positive correlations between student AI literacy and self-reported use of large language models (LLMs). Demographic factors, such as student major and gender, were found to influence these correlations and differences in averages among subgroups. The results confirmed that AI literacy levels vary among students from different backgrounds and demographics, yet most students are reporting use of AI tools. This research will aid educational institutions in understanding the current state of student AI literacy and LLM use and provide a starting point for preparing students to use AI tools purposefully.

Contents

Abstract ii

1 Introduction 1

2 Related Work 3

2.1 Effects of AI on Student Learning . . . 3

2.2 AI Literacy Tests . . . 4

2.3 Student AI Literacy . . . 4

2.4 Student perceptions and use of LLMs . . . 5

3 Method 7

3.1 Identification of Existing Measurement Tools . . . 7

3.1.1 AI Literacy Tests . . . 7

3.1.2 LLM Use Questionnaire . . . 8

3.2 Pilot Data and Item Refinement . . . 8

3.2.1 Initial Item Refinement . . . 8

3.2.2 Pilot Study . . . 8

3.2.3 Finalized Adapted Questionnaire . . . 9

3.3 Data Collection and Sample Characteristics . . . 9

3.3.1 Data Collection Methods . . . 9

3.3.2 Eligibility and Inclusion Criteria . . . 9

3.3.3 Sample . . . 10

4 Results 11

4.1 AI Knowledge Test . . . 11

4.2 Self-Assessment AI Literacy Test . . . 11

4.3 LLM-use Questions . . . 12

4.4 Relating LLM use to AI Literacy . . . 13

4.5 Factors influencing student perspectives . . . 14

4.5.1 Engineering vs. Non-Engineering Majors . . . 14

4.5.2 Gender . . . 15
4.5.3 Year in school . . . 18
4.5.4 AI or Ethics Course Participants . . . 19
4.6 Student Sentiment Toward LLMs . . . 20

5 Discussion 22
5.1 RQ1: How does student LLM use relate to student AI literacy? . . . 22
5.2 RQ2: What factors influence student perspectives and interactions with LLMs? . . 23
5.2.1 RQ2-A: Engineering vs. Non-Engineering Majors . . . 23
5.2.2 RQ2-B: Females vs. Males . . . 23
5.2.3 RQ2-C: Year in School . . . 24
5.3 Limitations . . . 24

6 Conclusion 26

References 27

Appendix 30

1 Introduction

The recent surge in the availability of Large Language Models (LLMs) has provided students with increased opportunities to utilize these tools for coursework. Trained on vast amounts of existing text data, LLMs take in user prompts and provide catered responses such as textual translations, summaries, or answers to questions, demonstrating a significant advancement in Artificial Intelligence (AI) and Natural Language Processing (NLP) technology. LLMs can be useful tools in education by providing personalized learning experiences, summarizing large texts, or explaining difficult concepts. On the other hand, the use of LLMs can also present challenges. Researchers have discussed a decline in critical thinking and problem-solving skill development with increased reliance on LLMs [8]. Students have expressed concern over these tools returning biased or plagiarized responses [3]. Even the traditional education model has been challenged as genuine student performance becomes increasingly AI-generated [13].

With these concerns in mind, a basic understanding of how LLMs work, how to write effective prompts, and how to critically evaluate their output are becoming increasingly important skills for LLM users to have. Measurement of a user's understanding and skill with using LLMs can take the form of an AI literacy test. Researchers Duri Long and Brian Magerko define AI literacy as "a set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace" [9]. There are many variations of AI literacy tests that vary in who is tested, what they are tested on, and how the test is administered. Some tests focus specifically on students in K-12 and beyond [4, 6], while others focus on working adults in technological fields [11]. The majority of current AI literacy tests focus on self-assessment measurements [6, 11, 14], while a few tests measure literacy through right and wrong answers [4]. To provide a well-rounded approach to measuring AI literacy, this research utilized a combination of both self-assessment and knowledge test components from existing, validated AI literacy tests [4, 14] to gain insight into AI literacy among university students.

In the context of education, measuring AI literacy of students can provide insight into how students understand LLMs and how they use them in coursework or personal projects. These findings are especially relevant as AI becomes increasingly integrated across all sectors of society. To navigate a future workforce shaped by AI, students must be equipped with the skills and knowledge to use it both effectively and responsibly.

In answer to this need for a measurement of AI literacy among students, this research engaged in an exploratory investigation into university student understanding and use of LLMs in coursework. Specifically, we addressed the following research questions:

1. *RQ1: How does student LLM use relate to student AI literacy?*
   (a) *Using self-assessment AI literacy tests?*
   (b) *Using AI knowledge tests?*
2. *RQ2: What demographic factors influence student perspectives and interactions with LLMs?*
   (a) *For engineering vs. non-engineering majors?*
   (b) *Across years in school?*
   (c) *Across gender?*

To answer these questions, we conducted a survey among university students (n=80) to measure AI literacy and record perspectives and use of LLMs. The next section presents related works to this research, providing context for the novelty of this thesis.

2 Related Work

The recent boom in artificial intelligence over the past decade has prompted much research regarding its impacts on all sectors of society. Special interest in the effects of AI in education has emerged as large language models and other generative AI have challenged traditional methods of learning in the classroom. Alongside teachers, teaching assistants, and search engines, students can now turn to LLMs like ChatGPT or Gemini for answers. Additionally, teachers have the opportunity to use LLMs as aids in lesson planning, checking grammar, or grading assignments. Advances in technology such as calculators or the Internet have historically had significant impacts on society, especially impacts on critical thinking skills among students and workers [8]. Thus, many researchers have begun investigating the impacts of AI, especially in the context of education to "make AI a catalyst rather than a replacement for learning" [13]. This section presents relevant research on the impact of AI in education, with discussions on general AI literacy and student-specific AI literacy, as well as student perspectives and understanding of LLMs.

2.1 Effects of AI on Student Learning

To understand the effects of AI on student learning, researchers have looked into how students currently use AI for coursework. A researcher for a 2025 survey on student AI use found that student use of AI is up nearly 30% from last year with 92% of students reporting that they use AI in some form out of the 1,041 students surveyed [3]. This increased use of AI has multifaceted implications for student learning outcomes. Authors of a research paper investigated the effects of ChatGPT and weighed the pros and cons of utilizing the LLM in student learning, concluding with "recommendations for how to address these challenges and ensure that such models are used in a responsible and ethical manner in education" [5]. In another study, researchers explored student use of ChatGPT in an introductory programming course and found that "the use of ChatGPT in programming education statistically significantly increased students' computational thinking skills, programming self-efficacy and motivation" [15]. On the other hand, some researchers have expressed concern over AI overreliance, claiming that AI may "reduce critical engagement" and "deprive the user of the routine opportunities to practice their judgment and strengthen their cognitive musculature" [8]. Overall, there is a general interest and concern over the inevitable growth in student use of AI. Researchers urge for more research on the effects of AI on student learning to leverage the benefits and mitigate challenges [1].

2.2 AI Literacy Tests

One way researchers have explored human interaction and understanding of AI is through AI literacy tests. Similar to its predecessors of digital literacy or data literacy, AI literacy measures a set of competencies regarding the use and understanding of AI [9]. Most research has involved studies with adults, whether experienced in a technological field or not. Extensive research has involved gathering self-assessments of AI literacy from adult subjects. Researchers Pinski and Benlian created and validated an AI literacy test to be used within the corporate workplace as a measurement tool for AI literacy in various information systems roles [11]. Their methods involved expert interviews and the creation of a survey that was refined to 13 items after pilot study results.

Another survey created and deployed by researchers consisted of 34 items grounded in different facets of AI literacy and competency such as application of AI or ethics of AI. The focus of this AI literacy tool was to include psychological competencies as they are “particularly important in the context of pervasive change through AI systems” [2].

Other research on AI literacy focuses on the demographic being measured. For example, Laupichler et al. focused on refining an existing AI literacy test to better measure AI competencies among non-experts [7]. In addition to AI literacy tests geared toward adults, a few tests have emerged that are geared toward student understanding of AI.

2.3 Student AI Literacy

The immense impact of AI on education and the workforce prompts questions about student understanding and preparedness to use these tools at school and at work. Universities and other educational institutions seem like plausible environments for students to increase their AI literacy, however, there is little existing research on student AI literacy. A 2025 survey of undergraduate students found that “while students overwhelmingly believe it is essential to have good AI skills, only 36% have received support from their institution to develop them” [3]. Thus, it is important to begin to include more curriculum on AI and how to use it effectively at educational institutions. Researchers recommend that universities should specifically offer instruction in effective prompt writing to increase the usefulness of AI tools for students [12].

While the amount of research on student AI literacy is small compared to research on AI literacy for adults and the general population, a couple of literacy tests stand out. Researchers in Hong Kong recently published results from the creation and implementation of an AI literacy test among secondary students taking an AI course [6]. The questionnaire consisted of various self-assessment questions related to learning and understanding AI and was created with experts and piloted with a sample of students. Since this study focused on students taking an AI course, there is a need for a tool to gauge general student understanding of AI.

Hornberger et al. addressed this need through an AI literacy test geared toward university students [4]. Additionally, this test was one of the few AI literacy tests that measured AI literacy through a scored questionnaire consisting of right or wrong answers. This research contributed to the field of AI literacy by providing a baseline for general AI literacy among German university students in various majors. The study found that students from a technical background, like computer science, tended to have higher AI literacy than their counterparts. The authors acknowledged the limitation of the research being confined to just German students at a technical university and called for further research to build upon this baseline.

2.4 Student perceptions and use of LLMs

Research on AI literacy is growing with the expansion of AI into all facets of modern society, yet, there remains insight to be gained into how this literacy actually influences the use of AI, especially among students. AI literacy can be a useful tool to discover how prepared students are to use AI tools effectively and ethically. Padiyath et al. explored how social perceptions of students influenced their use of LLMs in an undergraduate programming course [10]. Through the development of a self-reporting questionnaire on LLM usage and self-efficacy, student interviews, and midterm scores, the researchers found significant correlations between student self-efficacy and LLM use. Another study assessed how and what students use AI tools for and what factors drive their willingness to use these tools for university-related tasks [12]. They found that trust in the produced output is a central factor in the use of AI for university-related tasks. Additionally, they found that computer science students reported more positive experiences with AI tools. Along this same line, a 2025 Generative AI student survey found that students in STEM courses reported more enthusiasm for AI [3]. Other findings from this survey included that women and those from lower socio-economic groups tended to have greater concerns about AI and were more likely to say that they have not used AI. Many of these researchers call for further research among different demographics and at different universities to better understand these trends among students.

With all this research and findings in mind, there is an opportunity to explore a novel space relating student LLM use and student AI literacy. Exploration of this relationship may reveal insight into how AI literacy influences LLM use. It will provide meaning to AI literacy results by placing it in the context of actual AI use. This research will answer the call to employ various AI literacy tests and LLM use questionnaires at a different institution with a different demographic of students [4, 10].

3 Method

In order to investigate the research questions, we needed to develop a questionnaire instrument to measure student AI literacy and student LLM use, allowing for exploration into the relationship between the two. In order to accomplish this, we reviewed literature on existing AI literacy tests and designed a questionnaire adapted from multiple of these validated AI literacy tests and LLM use questionnaires. The adaptation process involved adding, removing, and refining items from the existing questionnaires to better align with our research questions and student demographic. Many of the refinements came from feedback and results received from conducting an initial pilot study. After approval from the University of Utah Institutional Review Board (IRB), the questionnaire was made available for University of Utah students to take.

3.1 Identification of Existing Measurement Tools

In order to answer the research questions, we needed to assess existing measurement tools for both AI literacy and LLM usage. AI literacy focuses on general competencies that "enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace" [9]. Thus, an AI literacy test could take several forms, such as a self-reporting questionnaire or a graded knowledge test. On the other hand, an LLM use questionnaire investigates how students interact with and utilize LLMs in their lives through self-assessment questions. Both tools were needed to address the research questions.

3.1.1 AI Literacy Tests

We initiated the questionnaire development by reviewing literature of existing AI literacy measurement tools. We began with a broad search of any research relating to the keyword 'AI literacy' and then narrowed down our search to literature specifically pertaining to AI literacy tests or measurement tools. We found many tests targeted toward a wide demographic [2,7,14]. We further narrowed our search to tests targeted toward students specifically [4,6]. Test types for both demographics are incorporated in the final questionnaire in order to most efficiently measure literacy among students from various backgrounds. The two AI literacy tests we decided to incorporate into our questionnaire were Hornberger et al.'s [4] test that measures AI literacy through a scored knowledge test and Wang et al.'s [14] 12-item self-assessment tool. Both tools provide validated measures of AI literacy.

3.1.2 LLM Use Questionnaire

In order to relate AI literacy to LLM use, we reviewed literature regarding student LLM use in coursework. Padiyath et al. [10] created a questionnaire for students in an undergraduate programming course to explore student LLM use and self-efficacy. We incorporated this questionnaire into our questionnaire in order to provide a way to measure student LLM use.

3.2 Pilot Data and Item Refinement

In the creation of the questionnaire for this study, we performed an initial item refinement to ensure all items aligned with our research questions and objectives of the study. We then conducted a pilot study with members of our research lab and incorporated feedback on the questionnaire.

3.2.1 Initial Item Refinement

Initial item refinement consisted of going through the AI literacy tests and LLM use questionnaire to edit wording to clarify questions or answer choices. We also removed questions that did not align with the purpose of this study, such as questions relating to the environmental impact of AI or student perceptions of their peer's use of AI. We also changed wording in the LLM use questionnaire from "programming course" to "course" because our questionnaire is targeted toward students from all majors, not just computing students. We edited some of the answer choices to include tasks that all students may perform, not just programming tasks, to increase the applicability of the questionnaire questions to students from all majors. A summary of all the edits made to the questionnaire can be found in the appendix.

3.2.2 Pilot Study

The pilot study consisted of undergraduate and graduate students in computing majors (n=7). The purpose of the pilot study was to gather and implement feedback on the questionnaire from a sample group of students that matched the target demographic of the study. Upon analysis of results from the pilot study, we removed and refined items from Hornberger et al.'s test [4] based on how well the pilot test participants performed on it. We evaluated frequently missed questions and removed them from the questionnaire to help the literacy test become more applicable to students from all backgrounds. We also removed questions that had ambiguous wording or answer choices due to translation technicalities, since the test was originally written in German and translated by the authors, or due to the rapid advances in AI impacting correct answers.

3.2.3 Finalized Adapted Questionnaire

The finalized questionnaire included the adapted LLM use questionnaire [10], followed by a self-assessment AI literacy test [14], and lastly included an AI knowledge test [4]. Finally, the questionnaire concluded with demographic questions. The finalized questionnaire is located in the appendix.

3.3 Data Collection and Sample Characteristics

Data collection occurred exclusively on the University of Utah campus and through University of Utah-related technological platforms.

3.3.1 Data Collection Methods

Participants for the study were recruited using various methods across campus. We posted flyers in campus buildings providing information on the study as well as a link and QR-code to participate. We also posted to the University of Utah Reddit page inviting students to participate. In addition, we posted the flyer in three University of Utah related Discord channels. Lastly, we approached professors in multiple departments through Slack and email to invite them to share the flyer with students in class meetings.

3.3.2 Eligibility and Inclusion Criteria

Participants had to be a student at the University of Utah and be 18 years of age or older. Before beginning the questionnaire, participants were prompted to login with their University of Utah single sign-on to proceed. Participants also had to provide their consent, commit to give thoughtful answers, and confirm their age to continue with the questionnaire. The questionnaire included three attention checks spread throughout the questionnaire. Lastly, the questionnaire concluded with a final commitment check and a distraction check.

3.3.3 Sample

We received 131 responses to the questionnaire. Out of the 131 responses, 80 passed all attention checks and met the eligibility criteria. In the sample, the average age of participants was 22, with the youngest participant being 18 years of age and the oldest being 38 years of age. About 57% (n = 45) of participants identified themselves as male, 36% (n = 29) of participants identified as female, and 7% (n = 6) of participants identified as non-binary, self-identified, or preferred not to say. About 27% (n = 22) of participants were freshmen, 14% (n = 11) were sophomores, 23% (n = 19) were juniors, 17% (n = 14) were seniors, and 19% (n = 14) were postgraduate degree students. The majority of students identified as 'White,' which was about 69% (n = 62) of respondents, while the remaining 31% (n = 25) identified as various races and ethnicities, including Asian, Hispanic or Latino, and Black or African American. Respondents came from a variety of educational backgrounds with 38 majors being represented across the sample. The most common major in the sample was Computer Science, with about 27% (n = 21) of respondents belonging to that major.

4 Results

4.1 AI Knowledge Test

The AI literacy scores from the knowledge test were normally distributed according to the Shapiro-Wilk test ($W = 0.97, p = 0.11$). The average score was a 13.43/20, or 67%, with a range of scores from 25% - 95%. The wide range of scores indicates that general knowledge around AI varies drastically between students in the sample, with the average student performing less than satisfactory on the knowledge test according to a standard grading scale. The distribution of scores for the overall sample is shown in Figure 1 below.

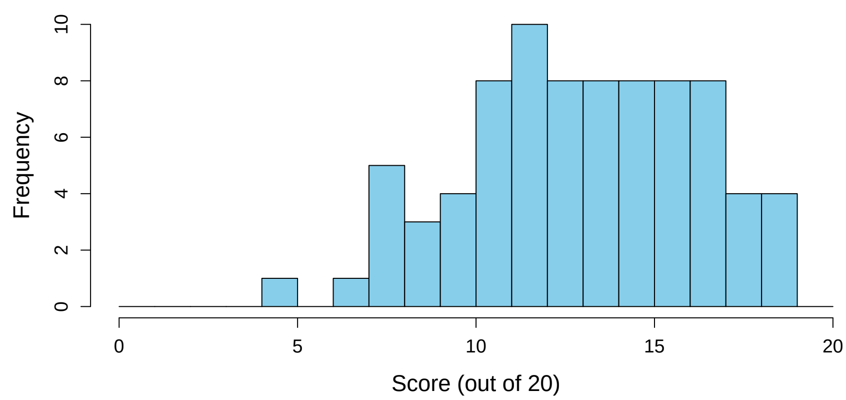

Figure 1: Distribution of AI Knowledge Test Scores

4.2 Self-Assessment AI Literacy Test

On the self-assessment of AI literacy, students scored themselves on a 7-point Likert Scale on various AI competencies. A higher value on the scale corresponded to a higher confidence in the respondent's perception of themselves possessing that competency. The sample average for the test was a 4.9/7, indicating that students tended to agree that they possessed the competencies in the test overall. A distribution of overall scores is shown in Figure 2. The test was also subdivided into four key themes: Awareness, Usage, Evaluation, and Ethics. The corresponding sample averages for each theme were 4.6/7, 5.2/7, 4.9/7, and 5/7 respectively. These results indicate that students were

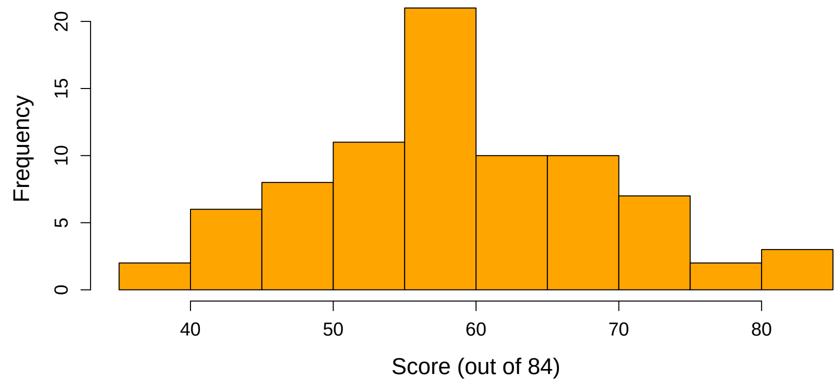

Figure 2: Distribution of Self-Assessment AI Literacy Scores

most confident in their ability to use AI and least confident in their awareness of AI. A distribution of the sample responses is found in Figure 3.

4.3 LLM-use Questions

ChatGPT was the most widely used LLM according to respondents, with 93% (n = 74) of respondents reporting use of ChatGPT for personal projects, coursework, or assignments. Only about 7% (n = 6) of respondents said they have not used any LLMs for these purposes. About 21% (n = 17) of respondents reported using LLMs several times a day, and 25% (n = 20) reported using LLMs several times a week. Only 10% (n = 8) reported no LLM use over the past three months. About 88% (n = 70) of respondents have not purchased enhanced LLM models, while the remaining 12% (n = 10) have made these purchases.

The three most common ways students used ChatGPT were to support their understanding of concepts (75% of students, n = 60), get help with debugging and coding (51% of students, n = 41), and generate ideas (40% of students, n = 32). The top three reported motivations for students to use ChatGPT were for saving time and convenience (68% of students, n = 55), to reduce feelings of stress or frustration when stuck (61% of students, n = 49), and to improve skills or knowledge (56% of students, n = 45). The majority of students reported not feeling over-reliant on tools like ChatGPT (74% of students, n = 59), while 19% (n = 15) of students reported feelings of over-reliance and

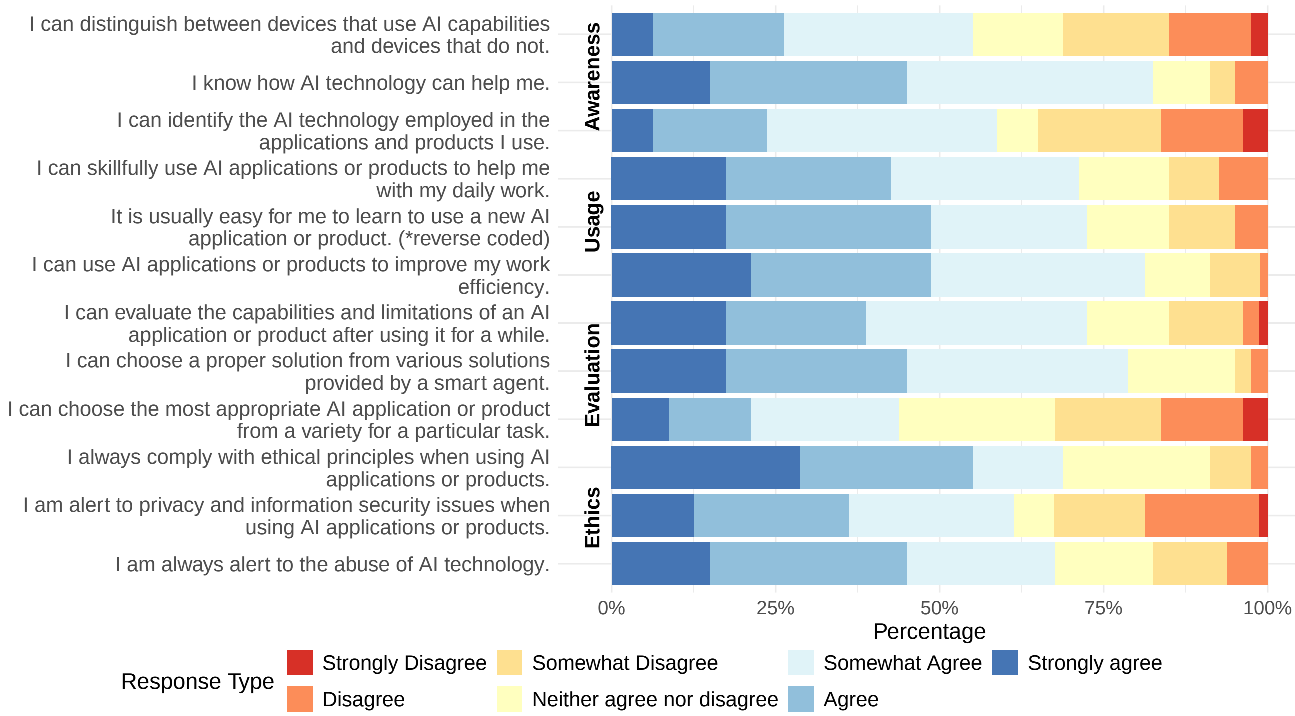

Figure 3: Distribution of Student Self-Assessment AI Literacy Responses (7-Point Likert Scale)

struggles on exams or coursework because of it.

4.4 Relating LLM use to AI Literacy

To answer RQ1 on how student LLM use relates to student AI literacy, we examined correlations between AI literacy test scores and reported LLM use from students. The frequency of use of LLMs was reported by students on a range from multiple times a day to never in the past three months. We assigned ranks to each of these values to correlate with the literacy scores. The distribution of respondents' frequency of use is shown in Figure 4. There was a strong, positive correlation ($r = 0.53, p = 3.28e - 7$) between students self-assessment of AI literacy and their reported LLM use. There also was a positive correlation ($r = 0.246, p = 0.03$) between knowledge test scores and self-assessment scores. Additionally, there was a positive correlation ($r = 0.27, p = 0.01$) between the knowledge test scores and self-assessment scores. The correlation results are shown in Table 1 below.

	Knowledge Test Score	Self-Assessed Score	Use of LLMs Freq.
Knowledge Test Score	-	0.271*	0.246*
Self-Assessed Score	0.271*	-	0.534**
Use of LLMs Freq.	0.246*	0.534**	-

Note: * indicates p<0.05, ** indicates p<0.01

Table 1: Correlation between AI Literacy and LLM Use

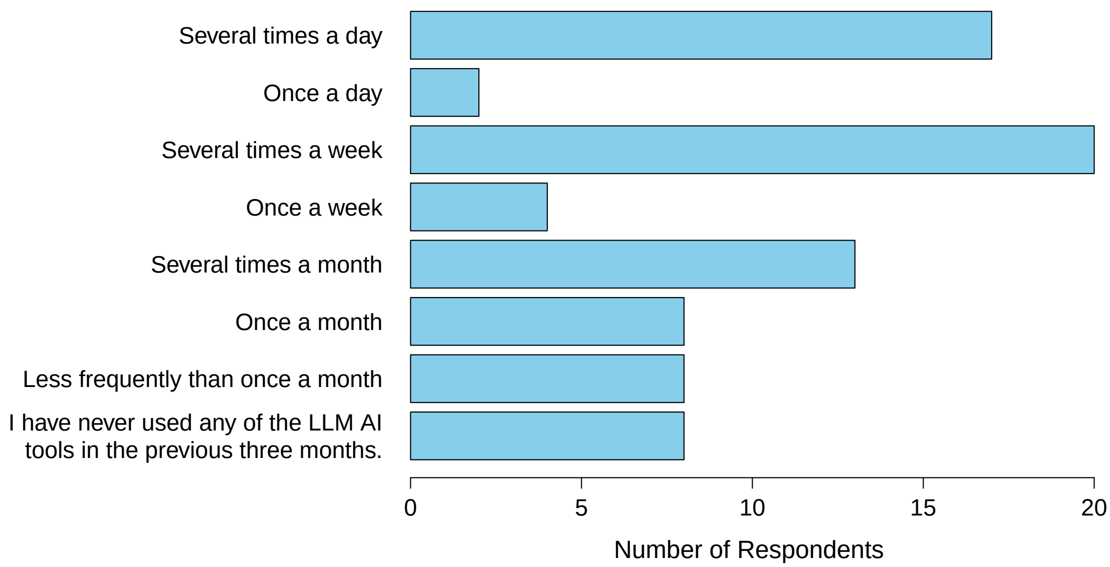

Figure 4: Distribution of Reported LLM Use Frequency

4.5 Factors influencing student perspectives

To answer RQ2 on what factors influence student perspectives and interactions with LLMs, we analyzed specific groups within the sample using various statistical tests, including correlations, t-tests, and ANOVA.

4.5.1 Engineering vs. Non-Engineering Majors

For engineering majors, there was a statistically significant positive correlation (r = 0.33, p = 0.03) between knowledge test scores and LLM use, whereas this correlation was not present among non-engineering majors (r = 0.13, p = 0.46).

For both engineering and non-engineering majors, there was a significant positive correlation

($r = 0.35, p = 0.02$ and $r = 0.7, p = 2.07e-6$) between self-assessment of literacy and LLM use. This correlation was larger and more statistically significant across non-engineering majors.

For engineering majors, there was a statistically significant positive correlation ($r = 0.33, p = 0.03$) between their knowledge test scores and self-assessment scores, whereas this correlation did not exist among non-engineering majors ($r = 0.18, p = 0.3$).

On average, engineering majors scored higher on the knowledge test than non-engineering majors (14.04 vs. 12.66), as well as on the self-assessment test (60.64 vs. 57.71). Engineering majors also reported using LLMs more frequently (5.04 vs. 4.66). T-tests revealed a marginally significant difference in averages for the knowledge test ($t = 1.94, p = 0.05$), with no significant differences observed in the self assessment test ($t = 1.26, p = 0.21$) or frequency of use ($t = 0.74, p = 0.46$). Closer examination of each competency in the self-assessment test revealed one statistically significant difference ($t = 2.64, p = 0.01$): engineering students reported greater confidence than non-engineering students in distinguishing between devices that use AI capabilities and those that do not.

When asked what students primarily used LLM tools for, the top two use cases for both engineering and non-engineering students was for support in learning concepts followed by debugging and coding assistance. For engineering students, the third most common use case was ideation, while non-engineers used it for generating solutions. The distribution of responses is shown in Figure 5. When asked what motivated students to use these tools, engineering students reported for saving time and convenience as their most common motivator, while non-engineering students reported to reduce feelings of stress or frustration when stuck. This distribution is shown in Figure 6.

4.5.2 Gender

For female respondents, there was a significant positive correlation ($r = 0.5, p = 0.005$) between knowledge test scores and LLM use, whereas this correlation was not significant among male respondents ($r = 0.15, p = 0.32$).

For both males and females, there was a significant positive correlation between self-assessment of literacy and LLM use, with the correlation being slightly stronger across female respondents ($r = 0.61, p = 0.0005$) than male respondents ($r = 0.5, p = 0.0007$).

For females, there was a significant positive correlation ($r = 0.52, p = 0.003$) between knowledge

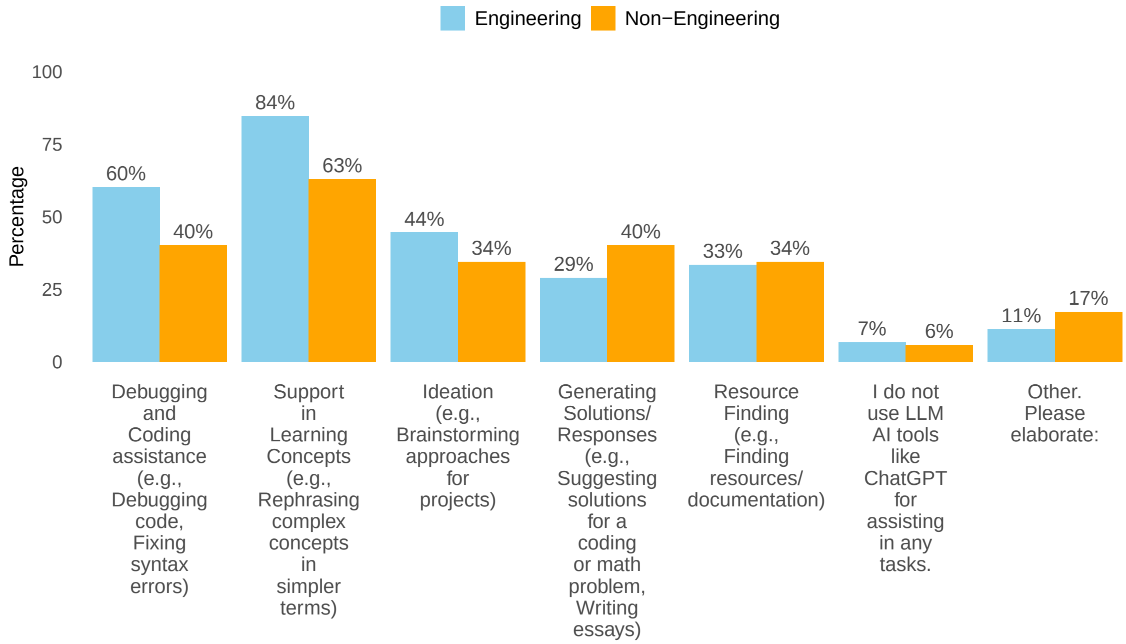

Figure 5: Types of LLM Use by Engineering and Non-Engineering Students

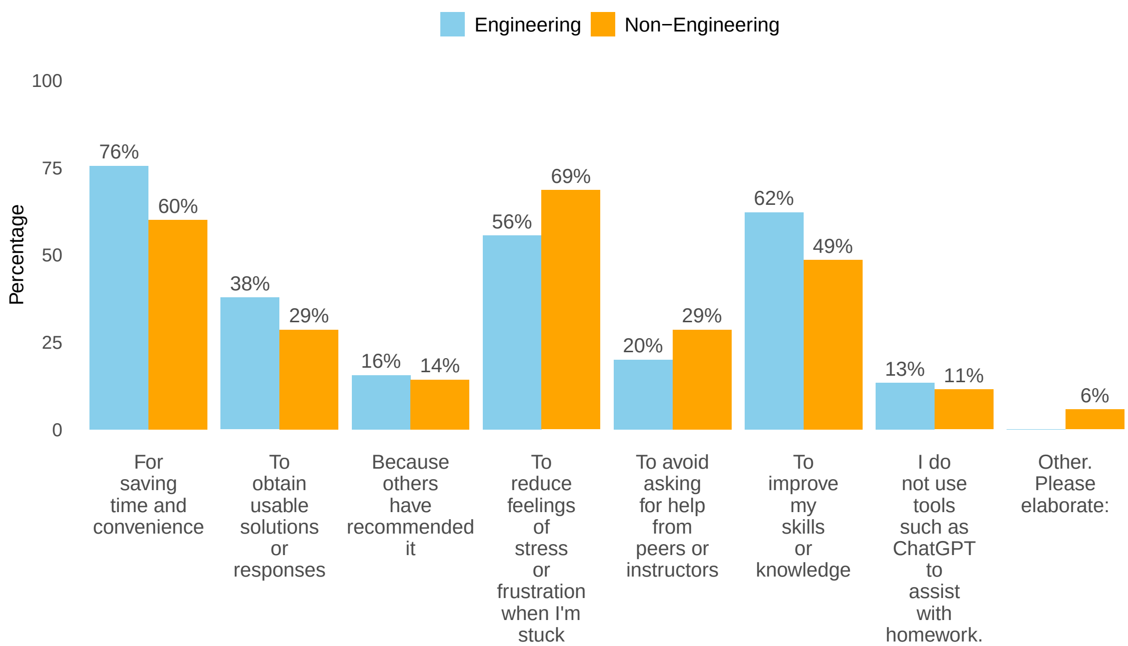

Figure 6: Motivation for LLM Use for Engineering and Non-Engineering Students

test scores and self-assessment scores, whereas this correlation did not exist among males (r = 0.12, p = 0.43). Upon closer examination of individual responses in the self-assessment, we found

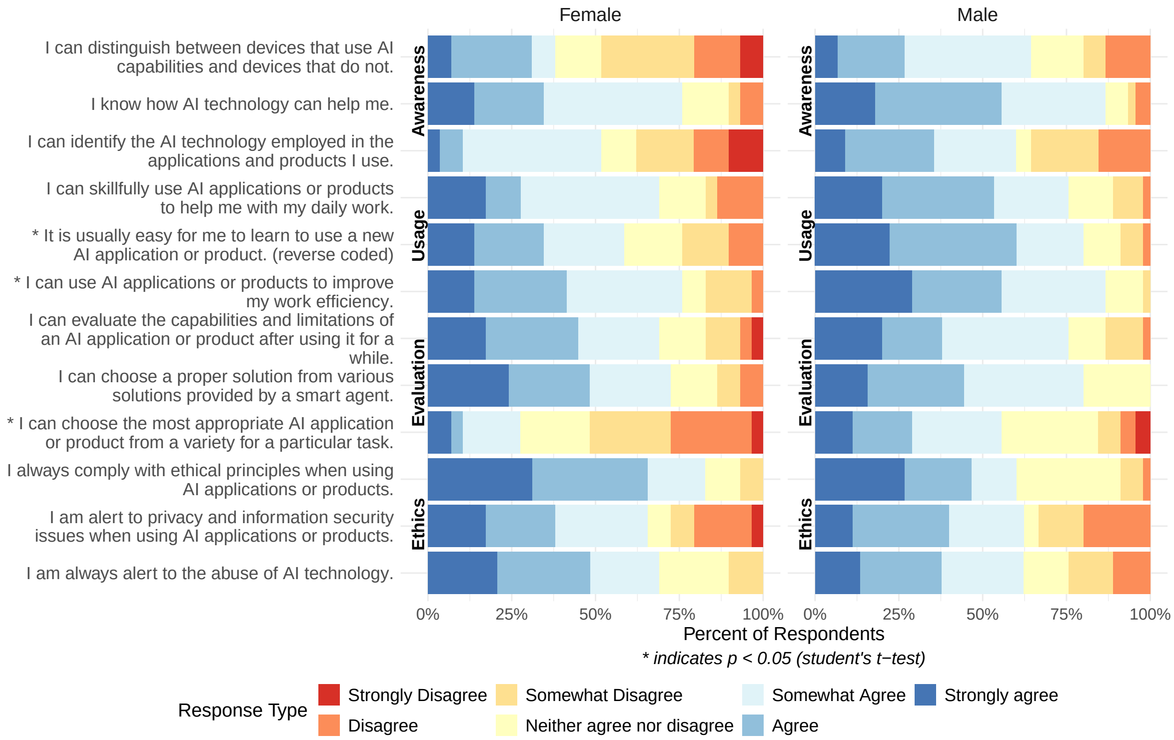

Figure 7: Self-Assessment AI Literacy Responses for Females and Males (7-Point Likert Scale)

that there were significant statistical differences in averages between females and males for the following competencies: "It is usually hard for me to learn to use a new AI application or product" ($t = 2.28, p = 0.02$), "I can use AI applications or products to improve my work efficiency" ($t = -2.06, p = 0.04$), and "I can choose the most appropriate AI application or product from a variety for a particular task" ($t = -2.9, p = 0.005$). The distribution of responses for both groups is shown in Figure 7.

On average, males scored higher on the knowledge test than females (13.64 vs. 12.86, $t = -1, p = 0.32$), as well as on the self-assessment test (61.07 vs. 57.28, $t = -1.52, p = 0.13$). Females and males reported very similar frequency of use of LLMs, with nearly the same average across both groups (4.86 vs. 4.93, $t = -0.13, p = 0.9$). After performing t-tests for all these differences in averages, none of them were found to be statistically significant.

4.5.3 Year in school

We found a significant positive correlation between self-assessment of literacy and LLM use among freshmen ($r = 0.69, p = 0.0004$), juniors ($r = 0.67, p = 0.002$), and seniors ($r = 0.62, p = 0.02$).

Among graduate students, there was a significant positive correlation between knowledge test scores and self-assessment scores ($r = 0.7, p = 0.006$), however, this correlation did not exist among other groups. A closer look at responses for each competency in the literacy test revealed no statistically significant differences between groups. Additionally, we found no correlations across age groups between knowledge test scores and LLM use.

Using ANOVA to test for significant differences in knowledge test scores ($F = 2.18, p = 0.08$), self-assessment scores ($F = 0.34, p = 0.85$), and reported LLM use ($F = 1.17, p = 0.33$) across school years revealed no statistically significant results. This could be due to small group sizes, which do not work well with an ANOVA test, since the largest group was juniors at 19 and the smallest was sophomores at 11. Instead, we used the Kruskal-Wallis test, as it is more appropriate for data with small sample sizes. However, we also found no statistically significant differences between years for the knowledge test ($\chi^2 = 8.55, p = 0.7$), self-assessment ($\chi^2 = 1.41, p = 0.84$), or reported LLM use ($\chi^2 = 4.82, p = 0.31$).

Since the groups were small when divided into the 5 types of students, we decided to compare undergraduate students to graduate students. Among undergraduate students and graduate students, there were no significant correlations between the knowledge test and LLM use frequency. For undergraduates, there was a significant positive correlation ($r = 0.6, p = 1.09e-7$) between self-assessment scores and frequency of LLM use, whereas this was not significant among graduate students ($r = 0.22, p = 0.45$). We used the student's t-test to discern any significant differences in the averages between undergraduate and graduate responses, however, none were found to be statistically significant.

We conducted a closer analysis on how students in different years use LLMs and what motivates their use. The distributions of responses are visualized below, with Figure 8 showing different types of use for LLMs, and Figure 9 showing various motivations for using LLMs. Despite overall similarities in response percentages across academic years for types of LLM use, notable differences include graduate students being the most likely to use LLMs for concept learning and resource

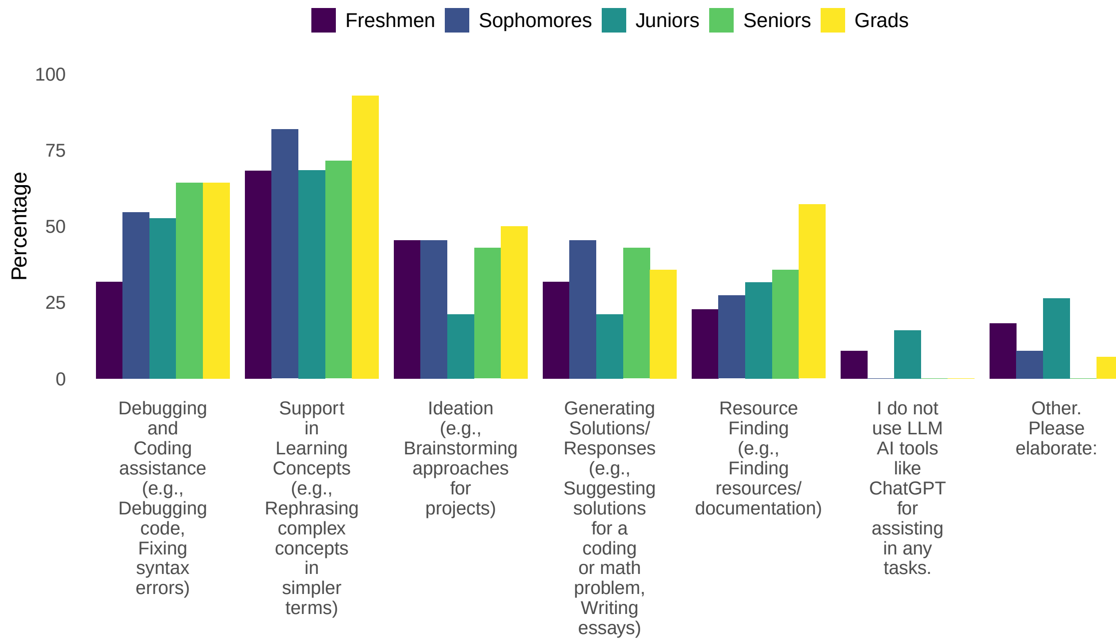

Figure 8: Types of LLM use by Year in School

finding. In fact, the use of LLMs for resource finding appears to increase with each year in school. Analysis of LLM use motivations revealed that freshmen were markedly more likely than other groups to be influenced by recommendations from others.

4.5.4 AI or Ethics Course Participants

There was a statistically significant positive correlation ($r = 0.59, p = 0.003$) between knowledge test scores and self-assessment scores among students who have taken an ethics in technology course, while this correlation did not exist among students who have not taken such a course ($r = 0.17, p = 0.2$). On the other hand, there was a statistically significant positive correlation ($r = 0.33, p = 0.008$) between knowledge test scores and self-assessment scores among students who have not taken an artificial intelligence course, whereas this correlation did not exist among students who have taken an artificial intelligence course ($r = 0.17, p = 0.48$). Further analysis of responses to individual competencies on the self-assessment test revealed no statistically significant differences between groups.

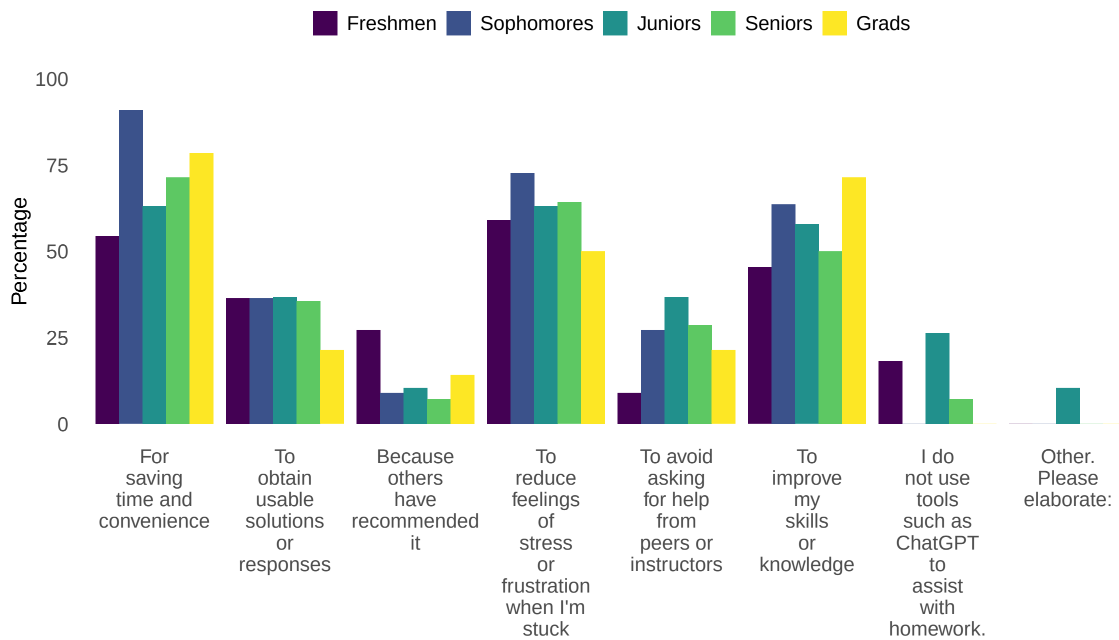

Figure 9: Motivation for LLM use by Year in School

4.6 Student Sentiment Toward LLMs

In the questionnaire, we asked students to rate their perspective on ChatGPT on a 5-point Likert scale. Overall, students tended to agree that they had concerns regarding ChatGPT. There was no statistically significant correlation between this concern and knowledge test scores ($r = -0.006, p = 0.96$). However, there was a small negative correlation ($r = -0.22, p = 0.05$) between this concern and self-assessment scores. The distribution of responses is shown in Figure 10. Overall, the biggest concern for students was that ChatGPT could potentially contain incorrect solutions they don't notice, with 98% ($n = 78$) of students expressing this concern. Student responses were split over whether using ChatGPT for homework provides an unfair advantage with 39% ($n = 31$) disagreeing with this statement and 43% ($n = 34$) agreeing with it.

Additionally, some respondents left comments in the questionnaire that provided further insight into student perspectives on using AI in coursework. One student explained, "When AI came out I was very opposed to it because of how much I saw it being used to cheat in school, but since then I've been using it extensively to help me understand various concepts I'm learning about in differential equations, and I've come to realize that AI has the potential to benefit society more than it harms

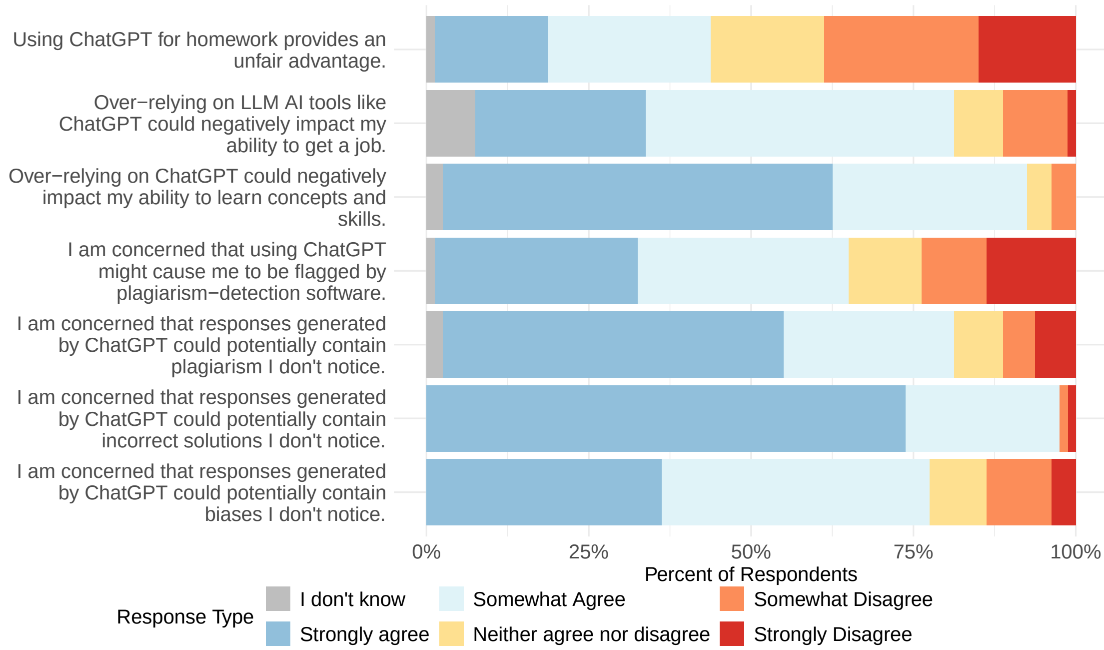

Figure 10: Student Sentiment Toward LLMs (5-Point Likert Scale)

it." Another student described a different perspective saying, "ChatGPT works well enough that I haven't really tried other software. One area I've found it's quite bad at, is math and sometimes coding. It makes mistakes or just invents stuff a lot. I still find it useful though." The full list of additional comments is located in the appendix.

5 Discussion

Our findings suggest that there are relationships between student AI literacy and use of LLMs, both for the overall student sample and for groupings within the sample. Our findings also suggest that multiple statistically significant factors influence student understanding and interactions with LLMs. The methods we used to in our data analysis included correlations, t-tests, and ANOVA.

5.1 RQ1: How does student LLM use relate to student AI literacy?

Upon examining both the knowledge test and self-assessment AI literacy scores, we found that knowledge test scores were positively correlated with student-reported LLM use frequency. This suggests that there is a meaningful relationship between student knowledge of AI and their frequency of LLM use. Furthermore, we found an even stronger positive correlation between the self-assessment scores and LLM use frequency. This finding suggests that students who used LLMs more frequently also tended to report higher confidence in their AI competencies. Additionally, we found a positive correlation between knowledge test scores and self-assessment scores, suggesting that a student's confidence in their AI competencies somewhat aligned with their actual performance on the knowledge test.

Thus, in response to RQ1, student LLM use was positively associated with AI literacy, as measured by both a knowledge test and a self-assessment. This positive correlation suggests that students who more frequently engage with LLMs may develop a strong understanding of AI concepts, or that students with a greater understanding of AI more frequently use these tools. For the correlation between self-assessment scores and frequency, it is possible that students who use LLMs more frequently tended to give themselves higher ratings due to their familiarity with these tools. It is also possible that the relationship goes the other way and that students with a higher AI literacy may feel more comfortable using these tools. However, the correlational nature of the data does not establish causation, and other variables may influence both LLM use and AI literacy. These findings highlight the importance of AI literacy in helping students utilize these tools responsibly in educational settings. Further research is needed to explore the causality of this relationship.

5.2 RQ2: What factors influence student perspectives and interactions with LLMs?

After identifying significant correlations in the overall sample, we examined subgroup patterns to explore potential factors that might influence these relationships.

5.2.1 RQ2-A: Engineering vs. Non-Engineering Majors

One interesting finding for these groups was the statistically significant difference in average knowledge test score between engineering majors and non-engineering majors. This finding aligns with the findings of the initial test performed in Germany where students from technical majors were found to perform better on the AI literacy test [4]. This finding suggests a lack of universal AI literacy across various majors at the university, with engineering majors performing significantly better on average on the AI knowledge test. Another interesting finding from this study was that the difference in self-assessment scores between engineering and non-engineering majors was not statistically significant. This may indicate that non-engineering students assess their AI literacy more confidently than is reflected in their knowledge test performance. Supporting this interpretation, a significant positive correlation was found between self-assessed and knowledge-based AI literacy scores for engineering majors, but not for non-engineering majors—suggesting that engineers tend to evaluate their AI knowledge more in line with their measured performance.

5.2.2 RQ2-B: Females vs. Males

One of the most notable subgroup findings was a positive correlation between AI knowledge test scores and LLM use among female students, a relationship not observed among males. This suggests that, for female students, higher AI literacy was significantly associated with more frequent LLM use, whereas for male students, frequent LLM users may have a lower AI literacy score despite frequent use. Additionally, significant positive correlations existed between self-assessment scores and LLM use for both males and females, with the correlation being slightly stronger across female respondents.

Upon closer examination of individual responses in the self-assessment test, we found significant statistical differences in averages for the following competencies between males and females: "It is usually hard for me to learn to use a new AI application or product," "I can use AI applications or products to improve my work efficiency," and "I can choose the most appropriate AI application or product from a variety for a particular task." Females rated themselves significantly lower than males on these competencies. Two of these competencies make up the 'Usage' construct, suggesting a potential gap in AI literacy education for female students. This finding is similar to a finding in the 2025 Generative AI Student Survey, which found that women are more likely to have concerns about AI and say they have not used AI [3]. However, the causation of this finding remains unknown, calling for future research to confirm the differences between male and female students and investigate what causes them.

5.2.3 RQ2-C: Year in School

Dividing the sample into 5 groups for each year in school did not produce many statistically significant results to discuss, perhaps due to the small subgroup sizes. We did find some positive correlations between self-assessment scores and LLM use among freshmen, juniors, and seniors, but we were unable to understand the cause of this relationship through correlation alone. We also found that this positive correlation existed among undergraduate students, but not among graduate students. However, we could not conclude that the year in school is an influencing factor in AI literacy or reported LLM use among students with these correlations alone. A larger sample would be required to explore the relationship between AI literacy and LLM use among these subgroups.

5.3 Limitations

Perhaps, there are a few key limitations of this research. First, the sample size for this study was small (n=80) and may not be representative of an entire student population. Although statistically significant findings emerged despite the small sample size, future research should aim to validate and extend these results using a larger and more representative sample.

Second, the demographic of participants was limited to enrolled students at a single university. Thus, the findings from surveying this demographic of students may not be representative of the overall student population everywhere. This study should be replicated across other universities and countries to validate these findings.

Third, it is important to note that the statistical methods used in this analysis were correlational, not causal. While the associations found are significant, there is no way to establish the direction of these relationships with the analysis we performed. Reasoning and discussion for these relationships and the causality of them were discussed but not confirmed through any measures.

Lastly, to minimize questionnaire length and reduce the time burden on participants, some questions were systematically removed from the original AI knowledge test, as detailed in the Method section. While this adaptation may have slightly reduced the precision of the AI literacy measure compared to the full version [4], the modification was necessary to promote response quality and mitigate questionnaire fatigue. Although this change may have had some impact on the results, care was taken to preserve the test's core constructs.

6 Conclusion

This study contributes to the growing understanding of the influence of AI in higher education by examining relationships between student AI literacy and LLM use. Through a questionnaire distributed at a large university, we explored associations between students' AI literacy scores and their self-reported frequency of LLM use.

Our findings revealed significant positive correlations between student AI literacy and LLM use as measured by both an AI knowledge test and a self-assessment scale. Furthermore, demographic factors such as student major and gender appeared to influence this relationship. Notably, engineering majors scored significantly higher on the AI knowledge test and showed a positive correlation between their knowledge scores and LLM use—an association not observed among non-engineering majors. Additionally, female students reported statistically significant lower confidence on several AI competencies on the self-assessment scale, despite reporting similar frequencies of LLM use to male students.

These results suggest that while LLM tools are being used by students across all demographics, levels of AI literacy vary. As AI tools become increasingly integrated into education and the workforce, it is crucial that educational institutions integrate AI education more broadly across all disciplines and demographics to ensure that all students are equipped to use these tools effectively and responsibly.

References

[1] Brett A. Becker, Paul Denny, James Finnie-Ansley, Andrew Luxton-Reilly, James Prather, and Eddie Antonio Santos. Programming Is Hard - Or at Least It Used to Be: Educational Opportunities and Challenges of AI Code Generation. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*, SIGCSE 2023, page 500–506, New York, NY, USA, 2023. Association for Computing Machinery. doi:10.1145/3545945.3569759.

[2] Astrid Carolus, Martin J. Koch, Samantha Straka, Marc Erich Latoschik, and Carolin Wienrich. MAILS - Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change- and meta-competencies. *Computers in Human Behavior: Artificial Humans*, 1(2):100014, 2023. URL: https://www.sciencedirect.com/science/article/pii/S2949882123000142, doi:10.1016/j.chbah.2023.100014.

[3] Josh Freeman. Student Generative AI Survey 2025. Technical Report Policy Note 61, Higher Education Policy Institute (HEPI), February 2025. URL: https://www.hepi.ac.uk/wp-content/uploads/2025/02/HEPI-Kortext-Student-Generative-AI-Survey-2025.pdf.

[4] Marie Hornberger, Arne Bewersdorff, and Claudia Nerdel. What do university students know about Artificial Intelligence? Development and validation of an AI literacy test. *Computers and Education: Artificial Intelligence*, 5:100165, 2023. URL: https://www.sciencedirect.com/science/article/pii/S2666920X23000449, doi:10.1016/j.caeai.2023.100165.

[5] Enkelejda Kasneci, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, Stephan Krusche, Gitta Kutyniok, Tilman Michaeli, Claudia Nerdel, Jürgen Pfeffer, Oleksandra Poquet, Michael Sailer, Albrecht Schmidt, Tina Seidel, Matthias Stadler, Jochen Weller, Jochen Kuhn, and Gjergji Kasneci. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103:102274, 2023. URL: https:

//www.sciencedirect.com/science/article/pii/S1041608023000195, doi:10.1016/j.lindif.2023.102274.

[6] Davy Tsz Kit Ng, Wenjie Wu, Jac Ka Lok Leung, and Samuel Kai Wah Chu. Artificial Intelligence (AI) Literacy Questionnaire with Confirmatory Factor Analysis . In *2023 IEEE International Conference on Advanced Learning Technologies (ICALT)*, pages 233–235, Los Alamitos, CA, USA, Jul 2023. IEEE Computer Society. URL: https://doi.ieeecomputersociety.org/10.1109/ICALT58122.2023.00074, doi:10.1109/ICALT58122.2023.00074.

[7] Matthias Carl Laupichler, Alexandra Aster, Nicolas Haverkamp, and Tobias Raupach. Development of the "Scale for the assessment of non-experts' AI literacy" – An exploratory factor analysis. *Computers in Human Behavior Reports*, 12:100338, 2023. URL: https://www.sciencedirect.com/science/article/pii/S2451958823000714, doi:10.1016/j.chbr.2023.100338.

[8] Hao-Ping (Hank) Lee, Advait Sarkar, Lev Tankelevitch, Ian Drosos, Sean Rintel, Richard Banks, and Nicholas Wilson. The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers. In *CHI 2025*, Yokohama, Japan, April 2025. URL: https://www.microsoft.com/en-us/research/publication/the-impact-of-generative-ai-on-critical-thinking-self-reported-reductions-in-cognitive-effort-and-confidence-effects-from-a-survey-of-knowledge-workers/.

[9] Duri Long and Brian Magerko. What is AI Literacy? Competencies and Design Considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–16, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3313831.3376727.

[10] Aadarsh Padiyath, Xinying Hou, Amy Pang, Diego Viramontes Vargas, Xingjian Gu, Tamara Nelson-Fromm, Zihan Wu, Mark Guzdial, and Barbara Ericson. Insights from Social Shaping Theory: The Appropriation of Large Language Models in an Undergraduate Programming Course. In *Proceedings of the 2024 ACM Conference on International Computing Education Research - Volume 1*, ICER '24, page 114–130, New York, NY, USA, 2024. Association for Computing Machinery. doi:10.1145/3632620.3671098.

[11] Marc Pinski and Alexander Benlian. AI Literacy - Towards Measuring Human Competency in Artificial Intelligence. 56th Hawaii International Conference on System Sciences (HICSS-56), Lahaina, HI, 01 2023. doi:10.24251/HICSS.2023.021.

[12] Leonhard Reiter, Moritz Jörling, Christoph Fuchs, , and Robert Böhm and. Student (Mis)Use of Generative AI Tools for University-Related Tasks. *International Journal of Human–Computer Interaction*, 0(0):1–14, 2025. doi:10.1080/10447318.2025.2462083.

[13] Dan Sarofian-Butin. At the Crossroads of Innovation: Embracing AI to Foster Deep Learning in the College Classroom. *EDUCAUSE Review*, July 2024. URL: https://er.educause.edu/articles/2024/7/at-the-crossroads-of-innovation-embracing-ai-to-foster-deep-learning-in-the-college-classroom.

[14] Bingcheng Wang, Pei-Luen Patrick Rau, and Tianyi Yuan and. Measuring user competence in using Artificial Intelligence: Validity and reliability of Artificial Intelligence literacy scale. *Behaviour & Information Technology*, 42(9):1324–1337, 2023. doi:10.1080/0144929X.2022.2072768.

[15] Ramazan Yilmaz and Fatma Gizem Karaoglan Yilmaz. The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Computers and Education: Artificial Intelligence*, 4:100147, 2023. URL: https://www.sciencedirect.com/science/article/pii/S2666920X23000267, doi:10.1016/j.caeai.2023.100147.

Appendix

Open-ended Responses

The following open-ended responses were received from participants in response to the question ”Do you have anything else you would like to add about your LLM use or understanding of AI?”

- I remain doubtful that AI will take significantly more jobs per year than previous innovative technologies like phones or computers. People like the person from NVIDIA (I forget what their position was) who claim AI will take the jobs of many programmers are just stirring the pot. Some programmers may lose jobs, but AI will bring far more programming jobs than it takes. At least in the foreseeable future.

- AI, like many other advancements in technology is a powerful tool that can be used but it shouldn't be supplemented with genuine learning as that would be our failure to continue to strive.

- One question asked something about selecting different AI tools for different purposes. I've only used chatgpt and one other image generating one i forgot the name of. Chatgpt works well enough that I haven't really tried other softwares. One area i've found its quite bad at, is math and sometimes coding. It makes mistakes or just invents stuff a lot. I still find it useful though.

- I don't know a lot about it. I'm very skeptical to use it. Especially when it comes to schoolwork and/or client confidentiality.

- While I do not use AI for coursework, I have used it to attempt to generate ideas and backstories for my own characters. I have found that it is bad at doing so, it is uncreative and often comes up with the exact same ideas for characters.

- When AI came out I was very opposed to it because of how much I saw it being used to cheat in school, but since then I've been using it extensively to help me understand various concepts I'm learning about in differential equations, and I've come to realize that AI has the potential to benefit society more than it harms it.

- I personally believe that AI will overall be a detriment to society. AI is being used by the bourgeois to replace art, music, and creative processes that are natural to human behavior. AI should be used to perform jobs that are too dangerous or too expensive for humans to perform, not replace the things we enjoy. The environmental impact of AI is also going to result in an increase in global climate change. We need to use AI in order to achieve certain goals, but we also need to improve its efficiency and reduce its environmental impact.

- Although AI has evolved extensively through the years, there are some aspects where AI is still dumb. If we convince AI that 2+2 = 3, at some point it agrees to it.

- I primarily use LLMs and AI as tools to learn about particular topics or gain some clarity about the topic. Within some fields of study (e.g., biology, chemistry, psychology) some concepts aren't explained well. LLMs like ChatGPT can be useful in clarifying those topics on demand. Within the coding, the on-demand natural can speed up the debugging process.

- I believe that it can be very helpful or very harmful depending on the use. I use it to understand topics and help me with my personal work but I have become more aware that using it too much is detrimental to my learning. I have come to a fair understanding of how to use it, though, since it helps me through complex situations. In the end the decision of how to implement the answers it gives me is my own and I must take them with a grain of salt.

- It's like Google on steroids. When Google came out, people thought using Google Search gave (unfair) advantage over those who didn't when working on homeworks. It would be impractical to ban students from using LLM AI tools to assist their homeworks. Professors need to revise homeworks so that LLM AI tools can't generate standard solutions.

- I really don't understand AI or how it works

- I get most of my understanding/familiarity with AI from my friend who is much more technologically adept than I am.

- The University better learn to leverage this tool or graduates will be unprepared for today's workplace.

Questionnaire

Screening Question

What is your age? `[Dropdown menu]`

First Commitment Check

We care about the quality of our data. For us to get the most accurate measures in our research, it is important that you thoughtfully answer each question in this questionnaire. As a reminder, your responses are anonymous, and any disclosure of your personal LLM use in coursework will NOT affect your standing with the University of Utah, so you can feel free to answer honestly.

Do you commit to providing thoughtful answers to the questions in this questionnaire?

- ○ Yes, I will provide thoughtful answers.
- ○ No, I will not provide thoughtful answers.
- ○ I cannot promise either way.

LLM Use Questions

The following questions relate to your knowledge and use of LLM AI tools. A reminder that your responses are anonymous and will not affect your standing with the University of Utah, so feel free to answer honestly.

Have you heard of the LLM AI tool ChatGPT?

- ○ Yes
- ○ No
- ○ Maybe

For the purpose of this study, ”LLM AI tools” refers to ChatGPT-like programs.

Have you used any of the following LLM AI tools for personal projects, coursework, or assignments? *(Select all that apply.)*

- □ ChatGPT (OpenAI)
- □ Copilot (Github)
- □ Copilot (Microsoft)
- □ Claude (Anthropic)
- □ Gemini (Google)
- □ Llama (Meta)
- □ I have never used any of these LLM AI tools.
- □ Some other LLM AI tool(s). Please specify: `[Text box]`

How often do you use the tools above for personal projects, coursework, and assignments? Please choose the option that best represents your usage frequency within the previous three months:

- ○ I have never used any of the LLM AI tools in the previous three months.
- ○ Less frequently than once a month
- ○ Once a month
- ○ Several times a month
- ○ Once a week
- ○ Several times a week
- ○ Once a day
- ○ Several times a day

Have you ever purchased credits (paying to get advanced models, paying for accessing some queries) for use of an enhanced LLM AI model (ChatGPT Plus, Github Copilot, etc.)?

- ◦ I was not aware that you could purchase access to enhanced LLM AI models.
- ◦ No, I have not purchased credits for enhanced LLM AI models.
- ◦ Yes, I have purchased credits for enhanced LLM AI models.

If you have purchased credits for LLM AI models, approximately how much money have you spent on enhanced LLM AIs? (For reference, a ChatGPT Plus subscription is currently $20 per month.) If you have not purchased any credits, please enter 0. `[Text box]`

Please indicate your level of agreement with each of the following statements on a scale of 1-7, where 1 = Strongly disagree; 2 = Disagree; 3 = Somewhat disagree; 4 = Neither agree nor disagree; 5 = Somewhat agree; 6 = Agree; 7 = Strongly agree.

- I understand how LLM AI tools like ChatGPT generate responses.
- I know how to use LLM AI tools such as ChatGPT.

Have you ever used LLM AI tools like ChatGPT or Copilot to assist with personal projects (i.e., projects outside of academics)?

- ◦ I was **not aware** I could use LLM AI tools for personal projects.
- ◦ I have **never** used LLM AI tools for personal projects.
- ◦ I have **rarely** used LLM AI tools for personal projects.
- ◦ I use LLM AI tools for personal projects **occasionally**.
- ◦ I use LLM AI tools for personal projects **frequently**.

Have you ever used LLM AI tools like ChatGPT or Copilot to assist with coursework or assignments?

- ◦ I was **not aware** I could use LLM AI tools to assist with assignments.
- ◦ I have **never** used LLM AI tools for assignments.
- ◦ I have **rarely** used LLM AI tools for assignments.

- ○ I use LLM AI tools for assignments **occasionally**.
- ○ I use LLM AI tools for assignments **frequently**.

LLM AI Tools like ChatGPT can cater to different needs. What type of assistance with coursework or assignments do you mainly receive from LLM AI tools like ChatGPT? *(Select all that apply.)*

- ☐ Debugging and Coding assistance (e.g., Debugging code, Fixing syntax errors)
- ☐ Support in Learning Concepts (e.g., Rephrasing complex concepts in simpler terms)
- ☐ Ideation (e.g., Brainstorming approaches for projects)
- ☐ Generating Solutions/Responses (e.g., Suggesting solutions for a coding or math problem, Writing essays)
- ☐ Resource Finding (e.g., Finding resources/documentation)
- ☐ I do not use LLM AI tools like ChatGPT for assisting in any tasks.
- ☐ Other. Please elaborate: `[Text box]`

Students use LLM AI tools such as ChatGPT for a variety of reasons. What motivates you to use LLM AI tools like ChatGPT for help with your coursework and assignments? *(Select all that apply.)*

- ☐ For saving time and convenience
- ☐ To obtain usable solutions or responses
- ☐ Because others have recommended it
- ☐ To reduce feelings of stress or frustration when I'm stuck
- ☐ To avoid asking for help from peers or instructors
- ☐ To improve my skills or knowledge
- ☐ I do not use tools such as ChatGPT to assist with homework.

- ☐ Other. Please elaborate: `[Text box]`

Please indicate your level of agreement with the following statement on a scale of 1-5, where 1 = Strongly disagree; 2 = Somewhat disagree; 3 = Neither agree nor disagree; 4 = Somewhat agree; 5 = Strongly agree.

- I would rather learn how to use an LLM AI tool like ChatGPT to generate solutions and responses than learn the skills to do so myself.

Have you ever felt over-reliant on LLM AI tools like ChatGPT?

- ○ Yes, and I have struggled on an exam because of this over-reliance.
- ○ Yes, I find it hard to complete my homework or other tasks without it.
- ○ No, I have never felt over-reliant on Chat-GPT-like tools.
- ○ Not applicable, as I do not use LLM AI tools in the first place.

If you've ever felt over-reliant on LLM AI tools, have you significantly reduced your usage of these tools as a result?

- ○ Yes, I have reduced my usage after feeling over-reliant.
- ○ No, I have not reduced my usage after feeling over-reliant.
- ○ No, as I have never felt over-reliant.
- ○ Not applicable, as I do not use LLM AI tools much in the first place.

Please indicate your level of agreement with each of the following statements on a scale of 1-5, where 1 = Strongly disagree; 2 = Somewhat disagree; 3 = Neither agree nor disagree; 4 = Somewhat agree; 5 = Strongly agree; I don't know. If you are unsure about your opinion, or do not know enough about the topic to have an opinion, please choose the "I don't know" option.

- I am concerned that responses generated by ChatGPT could potentially contain biases I don't notice.
- I am concerned that responses generated by ChatGPT could potentially contain incorrect solutions I don't notice.

- I am concerned that responses generated by ChatGPT could potentially contain plagiarism I don't notice.
- Using ChatGPT for homework provides an unfair advantage.
- I am concerned that using ChatGPT might cause me to be flagged by plagiarism-detection software.
- For this question, select the fourth option to indicate that you are reading and completing the questionnaire attentively.
- Over-relying on ChatGPT could negatively impact my ability to learn concepts and skills.
- Over-relying on LLM AI tools like ChatGPT could negatively impact my ability to get a job.

AI Literacy Self-Rating Questions

The following questions relate to your perceptions of AI.

Please indicate your level of agreement with each of the following statements on a scale of 1-7, where 1 = Strongly disagree; 2 = Disagree; 3 = Somewhat disagree; 4 = Neither agree nor disagree; 5 = Somewhat agree; 6 = Agree; 7 = Strongly agree.

- I can distinguish between devices that use AI capabilities and devices that do not.
- I know how AI technology can help me.
- I can identify the AI technology employed in the applications and products I use.
- I can skillfully use AI applications or products to help me with my daily work.
- It is usually hard for me to learn to use a new AI application or product.
- I can use AI applications or products to improve my work efficiency.
- Select option two Disagree on the scale for this question.
- I can evaluate the capabilities and limitations of an AI application or product after using it for a while.

- I can choose a proper solution from various solutions provided by a smart agent.
- I can choose the most appropriate AI application or product from a variety for a particular task.
- I always comply with ethical principles when using AI applications or products.
- I am alert to privacy and information security issues when using AI applications or products.
- I am always alert to the abuse of AI technology.

AI Literacy Quantitative Questions

The following questions are about general concepts related to artificial intelligence and machine learning. For each question, please provide the answer you think is most correct.

[NOTE: The questions below were presented in random order.]

Imagine you are chatting with an assistant on the Internet. How could you proceed to find out whether you are interacting with a human or with an AI?

- ○ You can save yourself the effort, since you can no longer distinguish between a human and an AI in written communication.
- ○ Ask a difficult factual question, since only a human can answer it.
- ○ Make a few typing errors in your text, then the AI can no longer understand you, but a human can.
- ○ Make an ironic remark, because this is better understood by humans.

Which of the following interdisciplinary research fields is also a subfield of AI?

- ○ Blockchain
- ○ Natural Language Processing
- ○ Psychology of Learning

- Bioinformatics

What makes AI intelligent?

- AI can walk and talk.
- AI has an artificial brain.
- AI is at least as intelligent as humans.
- AI acts rationally to achieve a particular goal as well as possible.

Why do AI systems behave intelligently?

- They have no feelings that could distract them from their task.
- They think autonomously and pursue their own goals.
- They have been programmed to try to achieve a given goal as well as possible.
- They are built similar to the human brain and therefore have a similar intelligence.

How are humans and AI similar?

- Machine learning procedures are similar to those of human learning.
- Forgetting curves are similar in machine learning and in human learning.
- AI has similar difficulties as humans in learning by heart.
- AI based on neural networks has similar strengths and weaknesses as the human brain.

What are knowledge representations in the field of AI?

- Particular representation of knowledge in order to communicate it to humans through AI.
- Sensors that capture information from the environment.
- Information about the world that can be processed by a computer.
- An algorithm that generates knowledge from data.

How do AI systems make decisions?

- ○ Based on mathematical-logical principles
- ○ Based on links defined by programmers
- ○ Based on quantum entanglement
- ○ Based on artificial intuition

What is a key criterion for the quality of a model in machine learning?

- ○ It can predict the output values of the test data as well as possible
- ○ It contains as few variables as possible
- ○ It is as well adapted as possible to the training data
- ○ The predictions are as unambiguous as possible

How does supervised learning differ from unsupervised learning?

- ○ In supervised learning, the output values of the training data are known.
- ○ In supervised learning, humans must supervise the AI during learning and intervene if necessary.
- ○ In supervised learning all computational steps are documented.
- ○ In supervised learning, stricter legal regulations apply.

Which statement about the steps in the machine learning process is correct?

- ○ The steps of the process are based on behaviorist learning theories.
- ○ The development of machine learning is in part an iterative process.
- ○ The steps of the process of supervised and unsupervised learning are basically the same.
- ○ The steps of the process can be run backwards to generate an (artificial) data set.

Sort the process steps in supervised learning into the correct order by listing the numbers in the order the steps should be taken:

1. Train model with training data
2. Predict test data with the model
3. Collect and prepare data
4. Divide data into training and test data
5. Calculate accuracy of prediction

What should be considered in machine learning when dividing the data into training and test data?

- The data should be divided into parts of as equal size as possible.
- The data should be randomly divided into training and test data sets.
- The test data should be of higher quality than the training data.
- The training and test data should be as different from each other as possible.

To what extent can humans influence the outcome of machine learning?

- Humans can hardly influence the result of machine learning, since it runs automatically.
- The results can only be influenced when selecting the data the model learns with.
- Humans can influence the result during development at several process steps.
- Humans can only influence the interpretation of the results.

What determines the behavior of AI systems?

- AI systems strive for autonomy.
- AI systems pursue a goal that has been given to them by humans.
- AI systems perform behaviors randomly.
- AI systems seek out goals independently and pursue them.

Why can systems based on machine learning obtain good results?

- Their work is often observed by humans and corrected if necessary (”supervised learning”).
- They think similarly to humans, but faster.
- They can draw conclusions from large amounts of data and thus improve their model.
- They are derived from expert systems in which expert knowledge is stored.

What data do AI-based recommender systems of streaming services use?

- The data of every user when using the service
- All data that a user of the service leaves on the Internet
- The data of other users, but not one’s own
- Only one’s own data when using the service

Which of the following is a technology company that is also the name of a fruit?

- Orange
- Peach
- Mango
- Watermelon
- Apple
- Pineapple
- Strawberry

You are testing a machine learning model that is supposed to classify photos of animals. You notice that the model is better at recognizing cats than dogs. What could be the reason for this?

- Dogs are more difficult to recognize than cats, since there are fewer images of dogs on the Internet.

- Small objects (cats) are better recognized than large ones (dogs).
- Most models are generally better at recognizing cats than dogs.
- The training data of the dogs were not representative for all dog breeds.

What is the black box problem?

- AI entails a residual risk that is hard to calculate.
- It is often difficult to determine how AI makes decisions.
- Users are often not informed about the application of AI.
- Many users have little knowledge about AI.

Which societal challenge is frequently mentioned in the context of AI?

- Lack of investment incentives in the educational system
- Chip shortage in industry due to the high computational cost of AI
- High error rate in AI-enabled manufacturing
- Replacement of human workforce by AI

What are central risks in using AI for predictive policing?

- Vulnerability to hacking
- Discrimination against suspects based on origin and status
- Lack of legal certainty in the event of AI breakdown
- Undermining the authority of police officers

Demographic Questions

Finally, please tell us a little bit about yourself.

What is your year of birth?

`[Dropdown menu]`

What is your gender?

- ○ Female
- ○ Male
- ○ Non-binary
- ○ Prefer not to say
- ○ Prefer to self-describe: `[Text box]`

What is your year in school?

- ○ Freshman
- ○ Sophomore
- ○ Junior
- ○ Senior
- ○ Master's or Professional degree student
- ○ Ph.D. student
- ○ Other. Please specify: `[Text box]`

What is your major?

If you have more than one major, provide both. If you are a graduate student, provide your undergraduate major along with your graduate major. `[Text box]`

In your post-secondary education (i.e., education after high school or similar), have you taken or are currently taking a course on AI? *(Select all that apply.)*

- □ I have previously taken a course on AI.

- ☐ I am currently taking a course on AI.
- ☐ I have never taken a course on AI.

In your post-secondary education (i.e., education after high school or similar), have you taken or are currently taking a course on ethics in technology (e.g., ethics in computing, ethics in data science, etc.)? *(Select all that apply.)*

If yes, please provide the name of the course.

- ☐ I have previously taken a course on ethics in technology: `[Text box]`
- ☐ I am currently taking a course on ethics in technology: `[Text box]`
- ☐ I have never taken a course on ethics in technology.

What is your ethnic background? *(Select all that apply.)*

- ☐ American Indian or Native American
- ☐ Asian
- ☐ Black or African American
- ☐ Hispanic or Latino
- ☐ Native Hawaiian or Pacific Islander
- ☐ White
- ☐ Something else. Please specify: `[Text box]`
- ☐ Prefer not to say

Are you an international student?

- ○ No
- ○ Yes

What is your country of origin?

`[Dropdown menu]`

How many years have you lived in the United States?

`[Dropdown menu]`

Additional comments and feedback section

Do you have anything else you would like to add about your LLM use or understanding of AI?

`[Essay-type text box]`

Do you have any feedback on the questionnaire or the research study?

`[Essay-type text box]`

Final Commitment Check

Did you answer all questions according to the provided instructions? Please answer honestly. Your answer has NO consequences for you or the compensation you will receive.

- ○ I answered all questions according to the provided instructions.
- ○ I sometimes chose random answer options because I was not motivated to answer the question or did not know how to answer it.
- ○ I often chose random answer options to finish as quickly as possible.

Could you answer the questionnaire without distractions? Please answer honestly. Your answer has NO consequences for you or the compensation you will receive.

- ○ I completed the study with full attention.
- ○ I sometimes was distracted (by people, noises, etc.).
- ○ I was often distracted (by people, noises, etc.).

Item refinement for knowledge test

Original Question	Answer Choices	Edits Made
In which of these areas is AI typically applied?	a) **Detecting credit card fraud** b) Cryptocurrency mining c) Web tracking d) Encryption for instant messaging services	Removed because there are multiple right answers since AI can also be used in web tracking.
Which of the following systems often use AI?	a) Flight surveillance systems b) Geopositioning systems c) 3D printing systems d) **Inventory management systems**	Removed because multiple answer choices could be correct.
In AI, a distinction can be made between ”weak” and ”strong” AI. ”Weak AI” refers to AI systems that have capabilities in a limited area. ”Strong AI”, on the other hand, is said to be capable of a very broad range of tasks, similar to humans. Which of these examples could be considered strong AI?	a) Intelligent virtual assistant (e.g. Alexa) b) Fully self-driving car c) Powerful search engine (e.g. Google) d) **Strong AI does not exist at the moment**	Removed because weak and strong AI are not key topics within a general understanding of AI. While definitions are provided in the question, this question is not necessary to gauge an average understanding of AI.
What can weak AI NOT do?	a) Make decisions under uncertainty b) **Solve a wide range of tasks** c) Solve a task better than a human d) Learn from unstructured data	Removed because the definition of weak AI given does not clearly explain what weak AI cannot do.

Original Question	Answer Choices	Edits Made
In which task is AI already superior to humans?	a) **Detecting tumors** b) Programming software c) Translating novels d) Designing cancer therapies	Removed because "superior" is a subjective term.
In which areas are humans still superior to AI?	a) Predict extreme weather events from weather data b) **Find a proof for a mathematical theorem** c) Answer quiz questions d) Play poker	Removed because "superior" is a subjective term.
An AI is supposed to divide images into meaningful parts. Input is always an image. Which input-output pair is NOT meaningful:	a) Image → number b) Image → vector c) Image → image d) **Image → video**	Removed because this question is not relevant to an average understanding of AI. The question would also have to define what it means by "meaningful."
What is a key criterion for the quality of a model in machine learning?	a) **It can predict the output values of the test data as well as possible** b) It contains as few variables as possible c) It is well adapted d) The predictions are as unambiguous as possible	Removed because a better answer should refer to generalizing on unseen data.
How can humans influence the outcome of machine learning?	a) Calculation of the accuracy of the prediction b) Randomized division into test and training data c) **Selection of the model** d) Abstraction of the model	Removed because this question could have multiple answers. For option a, accuracy may not be appropriate for unbalanced datasets.

Original Question	Answer Choices	Edits Made
What is the benefit of data visualizations?	a) Maintaining transparency b) Preparation of training for image recognition c) **Communication of results** d) Conducting statistical tests	Removed because data visualization is not under the umbrella of AI knowledge.
Which ethical principles should be considered when developing AI?	a) Holism, fairness, transparency b) **Prevention of harm, transparency, fairness** c) Fairness, holism, prevention of harm d) Transparency, prevention of harm, holism	Removed because this is more of a logic puzzle than a test of AI knowledge.
Which legal challenges do AI applications entail?	a) Users of AI have no option for legal protection b) Protecting the rights of AI itself c) Lawyers do not understand the importance of AI d) **Limited control of AI because of its autonomy**	Removed because the answer choices are subjective and multiple could be correct.

Item refinement for LLM-use questionnaire

Original Question	**Answer Choices**	**Edits Made**
How often do you think your classmates use tools like ChatGPT to help with programming assignments?	a) I don't think my classmates use LLM AI tools for programming assignments. b) I think my classmates rarely use LLM AI tools for programming assignments. c) I think my classmates use LLM AI tools for programming assignments occasionally. d) I think my classmates use LLM AI tools for programming assignments frequently.	Removed because this has already been explored by the source of these questions and will not be explored in my research.
Have you ever used LLM AI tools like ChatGPT or Copilot to assist with personal programming projects (projects outside of academics)?	a) I was not aware I could use LLM AI tools for programming. b) I have never used LLM AI tools for personal programming. c) I have rarely used LLM AI tools for personal programming. d) I use LLM AI tools for personal programming occasionally. e) I use LLM AI tools for personal programming frequently.	Justification for edits: changed programming projects to personal projects since the target survey demographic is programming and non-programming students.

Original Question	Answer Choices	Edits Made
Have you ever used LLM AI tools like ChatGPT or Copilot to assist with programming assignments?	a) I was not aware I could use LLM AI tools to assist with programming assignments. b) I have never used LLM AI tools for programming assignments. c) I have rarely used LLM AI tools for programming assignments. d) I use LLM AI tools for programming assignments occasionally. e) I use LLM AI tools for programming assignments frequently.	Justification for edits: changed programming assignments to assignments since the target survey demographic is programming and non-programming students.
LLM AI Tools like ChatGPT can cater to different programming needs. What type of programming assistance do you mainly receive from LLM AI tools like ChatGPT? (If "Other" , please elaborate.)	a) Debugging and Coding assistance (For example: Debugging code, Fixing syntax errors) b) Support in Learning Concepts (For example: Explaining programming concepts, Rephrasing complex concepts in simpler terms) c) Ideation (For example: Brainstorming approaches for projects) d Generating Solutions (For example: Suggesting solutions for a coding problem) e) Resource Finding (For example: Finding resources/documentation) f) I do not use LLM AI tools like ChatGPT for assisting in programming tasks g) Other. . .	Justification for edits: changed programming assistance to assistance since the target survey demographic is programming and non-programming students.

Original Question	Answer Choices	Edits Made
Students use LLM AI tools such as ChatGPT for a variety of reasons. What motivates you to use LLM AI tools like ChatGPT for help with your programming assignments?	a) For saving time and convenience b) To obtain usable code c) Because others have recommended it d) To reduce feelings of stress or frustration when I'm stuck e) To avoid asking for help from peers or instructors f) To improve my programming skills g) I do not use tools such as ChatGPT to assist with programming homework. h) Other . . .	Justification for edits: changed programming assignments to coursework and assignments since the target survey demographic is programming and non-programming students.
My instructor's policy on the use of LLM AI tools such as ChatGPT in class is clear regarding what is allowed and what is not allowed.	a) Strongly Disagree b) Disagree c) Neutral d) Agree e) Strongly Agree f) I don't remember my instructor's policy on ChatGPT use.	Removed because this is not relevant to the research questions to be explored in this study.
I am concerned about the environmental impacts of using ChatGPT.	5-point Likert scale from strongly disagree to strongly agree	Removed because this is not relevant to the research questions.
LLM AI tools like ChatGPT are being used by my classmates in ways that are against the course policy.	5-point Likert scale from strongly disagree to strongly agree	Removed because this is not relevant to the research questions. Peer usage was explored by this source already.

Item refinement for self-assessment scale

Original Statement	**Response Format**	**Edits Made**
I can distinguish between smart devices and non-smart devices	7-point Likert scale from strongly disagree to strongly agree	Justification for edits: For an AI literacy test, smart devices should be defined by whether or not they have AI capabilities. *I can distinguish between devices that use AI capabilities and devices that do not.*
I do not know how AI technology can help me.	7-point Likert scale from strongly disagree to strongly agree	Justification for edits: Negation removal. *I know how AI technology can help me.*
I am never alert to privacy and information security issues when using AI applications or products.	7-point Likert scale from strongly disagree to strongly agree	Justification for edits: Negation removal. *I am alert to privacy and information security issues when using AI applications or products.*

Name of Candidate: Annabelle Warner

Date of Submission: April 22, 2025