OPEN ACCESS
Received: 19 -03 -2026
Accepted: 04- 06-2026

الآداب
للدراسات اللغوية والأدبية

# Beyond Familiarity: LLM Use, Perceived Accuracy, and Student-Generated Improvements in a University Setting

**Dr. Mohammad Mahyoob Albuhairy** *

mqassem@taibahu.edu.sa

**Abstract:**

Large language models (LLMs) are transforming higher education, yet the factors shaping students' adoption and trust remain insufficiently understood, particularly in underrepresented educational contexts. This study investigates university students' awareness, usage patterns, perceptions, and recommendations regarding LLMs through a cross-sectional survey of engineering, computer science, and data science students. Descriptive and inferential statistical analyses examined adoption patterns and student perceptions. Although 90% of students were familiar with LLMs and 62% had used them for coding tasks, only 51% reported regular daily or weekly use, indicating that perceived usefulness and ease of use alone do not fully explain adoption. They also recognized their value for saving time and supporting learning. Students consistently identified accuracy, source transparency, and verifiability as essential for responsible use, highlighting trust as a key determinant of continued engagement. Building on these findings, the study proposes the Trust-Calibrated Technology Acceptance Model (TC-TAM), extending the Technology Acceptance Model by incorporating perceived trustworthiness as a critical factor influencing LLM adoption. The findings provide practical recommendations for educators, developers, and policymakers to support transparent, equitable, and responsible AI integration while contributing empirical evidence from an underexplored educational setting to advance research on generative AI in higher education.

**Keywords:** large language models, Higher education, Student perceptions, Human-AI interaction, Coding assistance.

* Associate Professor of Computational Linguistics, Department of Languages and Translation, Faculty of Arts and Humanities, Taibah University, Madinah, Saudi Arabia.

**Cite this article as:** Albuhairy, M. M. (2026). Beyond Familiarity: LLM Use, Perceived Accuracy, and Student-Generated Improvements in a University Setting, *Arts for Linguistic & Literary Studies*, *8*(3): 323 -344. https://doi.org/10.53286/fdxrq953

الآداب
للدراسات اللغوية والأدبية

OPEN ACCESS
تاريخ الاستلام: 2026/03/19م
تاريخ القبول: 2026/06/04م

# ما وراء الألفة: استخدام نماذج اللغة الكبيرة، وتقييم دقتها، ومقترحات الطلاب لتحسينها في بيئة جامعية

د. محمد مهيوب البحيري*
Eflu2010@gmail.com

**ملخص:**

تُحدث النماذج اللغوية الكبيرة (Large Language Models- LLMs) تحولًا جوهريًا في التعليم العالي، إلا أن العوامل المؤثرة في تبني الطلاب لها ومستوى ثقتهم بها لا تزال غير مفهومة فهمًا كافيًا، ولا سيما في السياقات التعليمية التي لم تحظَ بتمثيل كافٍ في الأدبيات البحثية. تستقصي هذه الدراسة مستوى وعي طلاب الجامعات بالنماذج اللغوية الكبيرة، وأنماط استخدامها، وتصوراتهم، وتوصياتهم بشأنها، من خلال دراسة مسحية مقطعية شملت طلاب الهندسة وعلوم الحاسوب وعلوم البيانات. واستُخدمت التحليلات الإحصائية الوصفية والاستدلالية لفحص أنماط التبني وتصورات الطلاب. وأظهرت النتائج أن 90% من الطلاب كانوا على دراية بهذه النماذج، وأن 62% منهم استخدموها في مهام البرمجة، إلا أن 51% فقط أفادوا باستخدامها بانتظام يوميًا أو أسبوعيًا، مما يدل على أن المنفعة وسهولة الاستخدام المدركة وحدهما لا تكفيان لتفسير تبني هذه النماذج. كما أدرك الطلاب قيمتها في توفير الوقت ودعم عملية التعلم. وأكدوا باستمرار أن الدقة، وشفافية المصادر، وإمكانية التحقق تمثل متطلبات أساسية للاستخدام المسؤول، مما يبرز الثقة بوصفها عاملًا حاسمًا في استمرار استخدامها. وبناءً على هذه النتائج، تقترح الدراسة نموذج قبول التكنولوجيا المعزز بالثقة (Trust-Calibrated Technology Acceptance Model, TC-TAM)، الذي يوسّع نموذج قبول التكنولوجيا (Technology Acceptance Model, TAM) من خلال دمج الجدارة بالثقة المتصورة بوصفها عاملًا جوهريًا يؤثر في تبني النماذج اللغوية الكبيرة. وتقدم الدراسة توصيات عملية للمعلمين، ومطوري تطبيقات الذكاء الاصطناعي، وصناع السياسات لدعم دمج الذكاء الاصطناعي بصورة شفافة وعادلة ومسؤولة، كما تسهم في إثراء الأدبيات العلمية من خلال تقديم أدلة تجريبية مستمدة من سياق تعليمي لم يحظَ بالدراسة الكافية، بما يعزز البحث في مجال الذكاء الاصطناعي التوليدي في التعليم العالي.

**الكلمات المفتاحية:** النماذج اللغوية الكبيرة، التعليم العالي، تصورات الطلاب، التفاعل بين الإنسان والذكاء الاصطناعي، المساعدة في البرمجة.

* أستاذ اللغويات الحاسوبية المشارك، قسم اللغات والترجمة، كلية الآداب والعلوم الإنسانية، جامعة طيبة، المدينة المنورة، المملكة العربية السعودية.

**للاقتباس:** البحيري، م. م. (2026). ما وراء الألفة: استخدام نماذج اللغة الكبيرة، وتقييم دقتها، ومقترحات الطلاب لتحسينها في بيئة جامعية، *الآداب للدراسات اللغوية والأدبية*، *8*(3): 323- 344 https://doi.org/10.53286/fdxrq953

## 1. Introduction

The rapid adoption of LLMs (Large Language Models) has sparked significant debate about their implementation in higher education. As tools like ChatGPT, DeepSeek, Gemini, and Claude become widely available, students have quickly begun using them for a range of academic purposes, such as retrieving information, coding assistance, summarizing content, and generating ideas (Rasnayaka et al., 2024; Singal and Goyal, 2025; Li & Qiao, 2025). This trend aligns with traditional models of technology acceptance, with perceived usefulness and ease of use as major factors in users' adoption of those technologies (Davis, 1989). However, LLMs pose challenges different from those typically encountered with traditional educational technologies, particularly regarding trust, reliability, and the nature of human-AI collaboration (Bernabei et al., 2023; Glikson & Woolley, 2020). Recently, empirical research has documented differences in AI literacy and engagement across various student populations (Xu et al., 2025; Hossain et al., 2025). This research has also highlighted the potential of LLMs to provide transformative opportunities in academic settings, as well as the pitfalls associated with their deployment (Kasneci et al., 2023; Albuhairy and Algaraady, 2023). Additionally, students' trust in LLM-generated content is developed through both the model's performance and external contextual cues, such as author information and perceived quality (Lyu et al., 2024; Raihan et al., 2025; Akpinar et al., 2026). The dynamic of this trust in LLMs is even more complex in settings where the LLM serves as a partner to the student (Al-Garaady & Albuhairy, 2023; Mirabile et al., 2026). Understanding all these factors will be crucial to providing evidence-based recommendations for the appropriate use of LLMs in higher education.

Even though there is increasing evidence of Language Models' capabilities, most current research has focused either on aggregate usage statistics or on their use for cheating, rather than on productivity gains. Studies show that LLMs can help build coding skills amongst beginner programmers (Kazemitabaar et al, 2023) and improve overall experiences for students and staff due to providing personalised assistance (O'Neill et al, 2025). However, at the same time, there have also been kind of discussions about how LLMs might facilitate academic misconduct by enabling students to commit acts of plagiarism or otherwise compromise authentic assessment (Perkins, 2023) by aiding students who want to be dishonest about their work. This leaves an area that we do not have much information about, and that is how the LLM Users themselves view and evaluate (1) tradeoffs between perceived benefits of using LLMs (for example, access to large amounts of information, better comprehension of subject matter and saved time in doing assignments) compared to the perceived drawbacks of using LLMs (for example, ethical concerns such as being too reliant on using them, potential for inaccuracies in the content they have created), as well as (2) whether the amount of use would appear to differ depending on discipline (for instance, students in AI-related fields appeared to

have much higher levels of engagement than those in computer science or engineering). From this, it is assumed that familiarity with an LLM alone does not explain why some students choose to adopt one, whereas others do not.

This study, hence, looks past the mere measurement of awareness/usage frequency. Using survey data from students in higher education, we consider which LLMs students utilize, along with how their use correlates between the various tools; in addition, we examine the reasoning behind the advantages/disadvantages of the tool(s) being used and consider critical questions regarding their limitations.

**Research Problem**

LLMs quickly found applications in higher education settings, but how students perceive, use, and evaluate these tools remains an open question. Research offers conflicting evidence on the frequency of LLM use and does not explore whether familiarity translates into protracted or sustained use. Nor is it known whether LLMs complement or replace lectures/textbooks, a matter of conflicting views ranging from strong optimism to much skepticism. Student experiences and expectations are underrepresented in the current debate, which is dominated by institutional and instructor perspectives.

Coding assistance is one important application of LLMs in education, yet research is lacking on how students view AI code reliability and which programming tasks they might offload to LLMs. Research indicates increased productivity when students offload difficult tasks (Peng et al., 2025), but less attention is paid to concept-based learning due to the novelty of programming (Lee et al., 2026; Kazemitabaar et al., 2023). Likewise, students have varied experiences with LLM-generated output, ranging from misinformation to context-free, poorly verifiable, or language-barrier-related advice, leading to overconfidence in LLMs (Albuhairy et al., 2023; Hossain et al., 2025). These are topics that require study as part of a student's LLM experience in the HE setting. The aims of the study are to investigate students' perceptions of LLMs as academic tools, student uptake/usage, and evaluation by university students, as they are the direct target audience for these apps. Research provides the basis for an informed, human-centric, and ethical generative AI-based approach that targets higher education.

## 2. Research Objectives

The present study aims to

1. Quantify university students' familiarity with and frequency of use of LLMs across disciplines, and test whether greater awareness is associated with more frequent academic usage.
2. Characterize university students' attitudes toward LLMs as substitutes for or supplements to traditional learning methods, including their perceptions of how LLMs affect comprehension of complex systems or concepts.

3. Assess the nature and extent of university students' use of LLMs for programming tasks, identifying the types of tasks assigned to LLMs and students' ratings of the accuracy of LLM-generated code suggestions.
4. Identify university students' self-reported challenges with LLM use (e.g., inaccurate information, overreliance, lack of context) and their suggestions for improving LLM design and integration, thereby generating user-driven recommendations for future educational applications.

**4. Research Questions**

The study addresses the following research questions, moving from descriptive usage patterns to student perceptions and finally to user-driven recommendations:

RQ1. How familiar are university students with LLMs, how frequently do they use them for academic purposes, and is higher awareness associated with greater frequency of use?

RQ2. How do university students perceive LLMs relative to traditional learning methods, and what are their views on LLM influence on their understanding of complex subject matter?

RQ3. For which programming tasks do university students employ LLMs, and how do they rate the accuracy of the code-related suggestions provided?

RQ4. What challenges (e.g., inaccuracy, overreliance, lack of context) do university students encounter when using LLMs, and what design or pedagogical improvements do they recommend based on their experiences?

**5. Literature Review**

**5.1 Scope and Purpose**

The rapid emergence of large language models as end-user tools has resulted in one of the fastest and most significant technology adoption curves in the history of higher education. Within months of the public release of ChatGPT in November 2022, millions of students began using generative AI in their academic work, often unaware of their institution's support for, or even their instructors' awareness of, its use. The swift pace of adoption has also left educational researchers trying to keep pace. This literature review summarizes empirical and theoretical work from 2022 through 2025 on students' awareness, usage patterns, attitudes, and outcomes related to LLMs in an academic context. Rather than providing a listing of all published works in this field, we structure our review into four major themes: (a) student engagement with LLMs as a means of learning assistance; (b) critical evaluation and AI literacy; (c) ethical issues/over-reliance tradeoffs and the need for institutional infrastructure to guide students' use; and (d) the competing priorities of integrating LLMs versus executing traditional forms of learning. Finally, the current study has been designed to address

existing gaps in the literature, thus providing a basis for an empirical discussion that has thus far been shaped primarily by speculation and institutional concerns.

**5.2 Student Engagement with LLMs: Support for Learning, Critical Evaluation, and Enhanced Performance**

The way learners access educational resources, analyze data, and critique AI-produced material has drastically changed since the introduction of LLMs within higher education. Recent studies show that learners do not merely view LLMs as answer providers but also use them actively to facilitate critical thinking, independent research, and metacognitive development. This diverse engagement further underscores the need for AI literacy to navigate the complexities of AI-generated materials.

**5.2.1 Support for Learning**

Many empirical studies show that LLMs (large language models) support students by providing timely, accessible resources that help them fill knowledge gaps and strengthen their problem-solving skills. O'Neill et al. (2025) reported that students who used LLMs as part of their study process reported enhanced ability to independently resolve complex problems in disciplines requiring iterative reasoning, specifically computer science and engineering. The conversational ability of LLMs- allowing students to further inquire about topics of interest, request additional explanations, and explore topics from different perspectives- appears to foster a supported, inquiry-based learning environment similar to what exists between effective human tutors and their students, although without the social presence of a teacher.

As previously demonstrated by Kasneci et al. (2023), the learning models (LLMs) might be described as "scaffolding" devices to give learners temporary support until they become proficient enough to learn independently. In contrast to other forms of content, such as static textbooks and pre-recorded lectures, the LLMs can adaptively change their response based on what a learner knows at the time they are being asked a question. Consequently, LLMs will provide simpler answers when a learner's knowledge level is lower, and more sophisticated answers when it is known that they are knowledgeable about the material being addressed. It is likely that this ability of LLMs to adaptively change their responses has led participants in this study to report that using LLMs was a significant help in better understanding complex learning concepts (see Section 5.3).

**5.2.2 Critical Evaluation and AI Literacy**

The growing literature on the topic indicates that students should learn to critically evaluate AI-generated content. According to O'Neill et al. (2025), students need to actively engage with LLMs rather than passively receive information from them. Learners need to develop the ability to assess the trustworthiness, truthfulness, and potential biases of LLM-generated content, which requires skills beyond typical information

literacy development. LLM-generated materials create an expectation of authority, regardless of fact-based truth, in contrast to traditional sources of information (e.g., journal articles, authoritative authors). This phenomenon, called "fluency bias" (Akpinar et al., 2026), necessitates explicit pedagogical interventions to develop AI literacy.

The development of AI literacy for learners will require teaching students how to do the following things: (a) identify possible hallucinations or fabrications; (b) seek confirmation of LLM claims by checking with reliable sources; (c) recognize when LLM responses are vague or inconclusive; and (d) create prompts that will yield more accurate and valid responses. Our research confirms this need: students have requested assistance with crafting appropriate citations for their sources and developing their prompting skills, demonstrating their active pursuit of metacognitive scaffolding in accordance with knowledge acquisition theory.

**5.2.3 Enhanced Academic Performance**

Using quantitative research, LLMs have improved performance over earlier work. Hossain et al., (2025) tested a cross-national study of AI literacy with LLM engagement and provided positive correlation statistics with higher cognitive outcomes for users of LLM and demonstrated that when a student uses LLM, he/she often has enhanced cognitive abilities such as writing, syntax & organization/use of credible argumentative evidence; as such; students that engaged regularly with LLM also reported higher confidence in their academic writing ability though the direction of the causal relationship remains unclear. The results also provide caution concerning LLM performance towards deeper levels of learning- e.g., increased conceptual understanding, cognitive analysis, and developing an original argument. These patterns are also evident in the coding education literature, where students using LLMs had lower completion times but, in many instances, lower posttest scores, indicating lower conceptual understanding of coding (Kazemitabaar et al., 2023). As such, the implication may be that LLMs are stronger performance support than effective learning tools and require explicit integration of LLM-based pedagogical strategies that emphasize students' engagement with LLMs to maximize their value.

**5.3.1 Overreliance Risks**

According to Hossain et al. (2025), habitual cognitive offloading to LLMs will likely lead to a failure to develop the very skills that the LLMs are designed to replace, therefore making overreliance one of the greatest risks of using LLMs. Historically, educators have worried that dependence on calculators and spell checkers has caused students to lose important skills. However, the risk of losing skills is exacerbated by the vast array of tasks LLMs can perform (e.g., generating text, summarizing information, translating languages, writing code, and providing explanations), thereby increasing the likelihood that students will lose critical

skills across multiple domains at once. This overreliance issue is even more pronounced for novice learners. Students at the novice level who lack mental models of their domains are unable to verify whether LLM-provided information is accurate. Therefore, LLMs will be most useful to novice learners. However, since novices have not developed mental models of their domains, they would be least able to use LLMs safely. Based on our results, which showed that 32% of respondents identified over-reliance as a major drawback, it appears this tension is recognized by students as well.

In terms of the examination of literature integrity in a post-COVID-19 pandemic environment, Perkins (2023) concluded that traditional definitions of plagiarism and unauthorized aid are not appropriate for LLMs. When a student utilizes an LLM for help with an essay outline, does that constitute cheating or collaboration? When a student employs an LLM to paraphrase a source, does that count as legitimate writing assistance or academic dishonesty? The answers vary from place to place, instructor to instructor, and assignment to assignment, creating confusion for students and inconsistent interpretations of cheating by different faculty members. Hossain et al. (2025) and Perkins (2023) agree that institutions should establish clear ethical guidelines and pedagogical frameworks for LLMs to inform students about their use across different contexts. These pedagogical frameworks should be specific to the disciplinary area of study and the specific assignment and communicated explicitly to students. Many institutions have created "AI use statements" for course syllabi that require students to disclose whether they used any LLM assistance and how they verified the accuracy of the generated text. However, the data indicate that implementing such policies across institutions is inconsistent; one respondent specifically noted that if their institution had an official acknowledgment that using an LLM such as ChatGPT is considered cheating, their use would have been avoided.

**5.4 The Tension Between LLM Integration and Traditional Learning Methods**

A considerable segment of the literature and a considerable concern among academic staff concern the threat posed by LLMs to traditional methods of learning. Educators cite the deep engagement that a student must have with their course material, the struggle, the iteration, and the cognitive friction of working through difficult problems on their own as always necessary. By providing immediate answers, LLMs offer shortcuts around this productive struggle, and such shortcuts may induce surface learning and surface-level task completion. O'Neill et al., however, acknowledged this tension and suggested reframing it. Rather than seeing LLMs as a threat to personalized learning, educators might see them as an opportunity to redesign what those experiences would look like. If the LLM can reliably handle lower-order work (summary, grammar check, standard coding), one could shift class time and testing towards higher-order skills (critical analysis, synthesis, creation, and ethical reasoning). This perspective, sometimes called "AI-mediated pedagogy," sees

the aim not so much as retaining traditional practices unchanged, but rather to adopt change when new capabilities become available.

Indirectly, our findings support this reframing: 11% of the participants in our sample felt that LLMs could replace traditional instruction, but they did believe that LLMs had value as an alternative means to execute some activities. Thus, it seems that the best way to move forward is to integrate the best of both forms of instruction; in other words, to maintain the benefits of traditional methods while augmenting them with AI capabilities or hybridizing instructional approaches rather than replacing them.

**5.5 Gaps in Literature and Rationale for the Present Study**

While LLM use in education is a fast-developing area of research, gaps remain. Research has been heavily concentrated in Western and English-speaking contexts; thus, the results of said studies cannot necessarily be applied elsewhere. This lack of appropriate research is particularly pertinent in Arabic-speaking populations, where language support is poor. Research papers often highlight LLMs' negative aspects, e.g., plagiarism or overreliance, and place less emphasis on the benefits LLMs bring to education. In addition, research reflects educators' voices and the impact on the teaching profession more than students' experiences with LLMs, desirable features, and factors they consider when using such programs. Rarely have studies examined whether familiarity affects the adoption of LLMs, where LLMs are seen as a separate skill, and coding remains under-researched as an area to which LLMs may be applied in education. This article discusses students' opinions of LLMs, their usage patterns, whether LLMs are used for coding, and how students would improve them within an educational context, an area lacking representation in previous studies. The relationship between familiarity and adoption will also be explored alongside how students offload certain processes in learning onto the LLM programs and the perceived levels of accuracy provided by the LLMs for writing code, thus adding to the small existing body of cross-cultural evidence surrounding LLM use within an educational setting, helping to provide an interpretation of their use outside of that within Western societies.

**5.6 Theoretical Framework: Technology Acceptance and the Role of Trust**

The Technology Acceptance Model explains technology adoption using two key determinants: perceived ease of use and perceived usefulness (Davis, 1989), which both explain users' intention and actual use of technology. TAM has also been widely used in educational technology and Generative Artificial Intelligence (GAI): It predicts higher adoption of learning and researching tools that learners believe are easy to use and useful. According to our results, however, TAM is not exhaustive when explaining students' adoption of GAI. Students in our data did seem to find GAI useful in tasks (72% claimed it eased or expedited their assignments alone), it was universally deemed easy to use. Still, only around half (54%) reported regular use (on either a daily/weekly basis), and only around half of those reported having a generally positive

attitude towards GAI. GAI's perceived high ease of use and usefulness still does not result in widespread everyday adoption among undergraduates. Other factors need to be accounted for. Recent work by Jang 2024. suggests that, in the presence of strong information asymmetry, such as in AI, human trust in technologies is a critical indicator of their adoption rate Glikson & Woolley, 2020; Mirabile et al., 2026). Similarly, in a recent study, Hossain et al. (2025) demonstrated that improved literacy about generative AI tools led to more effective LLM use and less human overreliance. In line with such research, this study found that students appreciate LLMs for their utility and use them for benefits, but the LLMs' 'black box' properties or lack of transparency mean they use the tools with caution, owing to concerns over their accuracy, source, or verifiability. Students reported a desire for features like 'citations' or 'confidence intervals and scores for generated content' that helped establish trust towards the content generated with AI, further hinting at the role of 'trust' as an explanatory factor for LLM usage. Based on findings, we put forward the Trust-Calibrated TAM (TC-TAM), an expansion of TAM with the concept of perceived trustworthiness, in terms of 'accuracy', 'source', and 'verifiability' as a moderator between two traditional factors-perceived usefulness and students' actual adoption rate of LLMs. Though the existing literature demonstrates that LLMs have the potential to deliver improved efficiency and learning outcomes, it also concerns academic integrity and potential areas for abuse and lacks studies from developing countries outside the Western world. Our paper seeks to address such shortcomings and gives evidence from the underrepresented setting in the developing world on what LLMs can offer in education, with human trust remaining the backbone of students' adoption behavior.

## 6. Methods

### 6.1 Research Design and Rationale

A cross-sectional survey design was used in this study to explore university students' awareness, attitudes, and usage patterns regarding large language models (LLMs) in academic contexts. Given the exploratory nature of the research questions and the need to capture a snapshot of students' behaviors and perceptions regarding the use of LLMs at a particular point in time, a cross-sectional survey design was the most appropriate method. The survey was conducted approximately two years after the public release of ChatGPT, after the peak novelty associated with the first release had stabilized, but before most institutions had developed clear policies governing the use of LLMs. Given the absence of baseline data for any participants, a longitudinal or experimental design approach would have been inappropriate. The survey design enabled the researchers to quantify the presence of specific phenomena (e.g., familiarity, frequency of use, perceived value of using large language models) and to incorporate students' perspectives through open-ended response options. The design placed more emphasis on breadth than depth, but the inclusion of qualitative items helped offset the pettiness that typically characterizes large-scale survey research.

### 6.2 Study Setting and Participants

The research took place at Taibah University in Saudi Arabia over 4 weeks to examine the use of LLMs in a University environment. The survey was distributed at a specialized engineering and computer science college that offers undergraduate programs taught in English. Students would have the opportunity to engage with LLMs by accessing them in English and using the same institution's language for their survey. The survey was made available through three different methods of communication: emails, course mailing lists for core undergraduate courses, and announcements to student WhatsApp groups maintained for the purpose of communicating with students by program representatives.

**Table 1.**

*Demographic Characteristics of Survey Respondents (N=72)*

| *Characteristic* | *Category* | *Frequency (n)* | *Percentage (%)* |
|---|---|---|---|
| *Gender* | Male | 64 | 88.9 |
| | Female | 8 | 11.1 |
| *Field of Study* | AI & Data Science | 28 | 38.9 |
| | Industrial Engineering | 15 | 20.8 |
| | Computer Science | 13 | 18.1 |
| | Computer Engineering | 6 | 8.3 |
| | Electrical Engineering | 6 | 8.3 |
| | Other | 4 | 5.6 |

Table 1 presents the demographics of the 72 students surveyed in this study. Males accounted for 86.1%, while females made up 13.9%, as some departments admitted only male students. The breakout of survey participants by field of study indicates that AI and Data Science had the highest representation at 38.9%, followed by Industrial Engineering at 20.8%, Computer Science at 18.1%, Computer Engineering at 8.3%, and Electrical Engineering at 8.3%. Altogether, the four engineering fields account for 61.1% of respondents, while the remaining 5.6% were reported as "Other". Overall, this shows a heavy concentration of engineering and computer science respondents in this sample.

We set the inclusion criteria very broadly to capture all levels of LLM familiarity, from no experience to daily use. All participants in the study must be matriculated university students, age 18 or older. There were no restrictions on the area of study, prior experience with LLMs, or proficiency in English. Including students with no experience with LLMs would have artificially increased the estimates of prior experience, making it impossible to analyze the relationship between familiarity and use. 72 participants responded completely to the survey questions considered in this report. No formal power analysis was conducted before data collection because the study was exploratory rather than confirmatory. A post hoc sensitivity analysis showed

that the sample size of 72 was sufficient to detect medium to large effect sizes (Cohen's $h \geq 0.50$ for binomial; Cramér's $V \geq 0.30$ for chi-square tests), with 80% power and $\alpha = 0.05$. This matches the size of effects that we report.

**6.3 Instrumentation and Survey Development**

The survey instrument was created specifically for this study and was based on elements of Davis's (1989) Technology Acceptance Model and other previous educational technology surveys (e.g., Scherer et al., 2019). However, in designing survey instruments, we considered that, due to rapid advancements in LLMs technologies, externally generated categories would likely become obsolete in a short time, so we included both closed-ended and open-ended questions.

**Development process:** The development of the questionnaire involved creating the first draft in English and going through two different processes to clarify each item on the original draft prior to its final approval by all reviewers. In step one, all three faculty members working in educational technology evaluated the first draft whilst assessing overall clarity, relevance, and thoroughness of each item, pointing out some unclear definitions and suggesting additional responses for ambiguous items. In step two, an initial pilot test with 10 undergraduate students (who would not be included in the final population) revealed additional issues with item clarity and confirmed that the average time required for completion was 6–10 minutes. Since there were no major issues with the items identified during the pilot, only minor changes to the wording of two items were made to accurately reflect the diversity of the models being discussed.

**6.4 Data Collection Procedures**

We used Google Forms to set up and host the survey because it is a familiar platform for participants, automatically exports data when conducted online, and meets the university's Data Security Guide. The survey link was sent to potential participants via the noted channels. There were 72 responses to the survey. Participants could access the survey at any time from their own devices (desktops, tablets, or mobile devices) and from locations of their choice, which reflects more naturalistic usage patterns rather than technology use in a laboratory-controlled experimental setting. Participants did not provide any identifying information (name, student ID, email address, or IP address) as part of their response, so their responses would remain anonymous, increasing the likelihood of providing honest responses, especially for items on potentially sensitive issues such as overly relying on outside assistance or concerns regarding academic integrity.

**6.6 Statistical Analysis**

Frequencies and percentages were computed for all categorical variables; means and medians were computed for ordinal variables (e.g., perceived usefulness scores). Using the Wilson score method was used to calculate confidence intervals (95%) for a given proportion.

**Table 2.**

*Descriptive Statistics for Number of LLMs Used per Respondent*

| Statistic | Value |
|---|---|
| **Mean** | 2.46 |
| **Median** | 2 |
| **Mode** | 1 |
| **Standard Deviation** | 1.58 |
| **Variance** | 2.50 |
| **Range** | 1 - 9 |
| **Q1 (25th Percentile)** | 1 |
| **Q3 (75th Percentile)** | 3 |
| **Interquartile Range (IQR)** | 2 |

The descriptive statistics for the LLMs used by each respondent are shown in Table 2. The average number of LLMs used was 2.46 (SD = 1.58), with a median of 2 and a mode of 1. These figures show that most students use 1-3 LLMs. The standard deviation of 1.58 indicates considerable individual variability in the number of LLMs students use, ranging from 1 to 9. The interquartile range (IQR) of 2 (Q1=1; Q3=3) indicates that the middle 50% of students who responded used between 1 and 3 LLMs, suggesting moderate levels of diversification but little extensive diversification.

### 6.6.1 Inferential statistics

Inferential statistical analyses were selected based on the research questions and the measurement levels of the variables. A binomial test was used to determine whether observed proportions differed significantly from a hypothesized proportion of 50% (e.g., awareness of LLMs). Chi-square goodness-of-fit tests assessed whether observed frequency distributions differed from equal expected distributions across multiple categories, with Cramér's V reported as the measure of effect size (small = 0.10, medium = 0.30, large = 0.50).

Associations between categorical variables (e.g., LLM familiarity and usage frequency) were examined using chi-square tests of independence, with Cramér's V used to quantify effect size. Ordinal outcomes between two independent groups (e.g., diploma versus bachelor's degree students) were compared using the Mann–Whitney U test, with rank-biserial correlation (r) reported as the effect size (small = 0.10, medium = 0.30, large = 0.50). This non-parametric test was selected because the ordinal variables could not be assumed to have equal intervals. Spearman's rank-order correlation (ρ) was used to examine monotonic relationships between ordinal variables (e.g., LLM usage frequency and perceived coding accuracy), as it is more appropriate than Pearson's correlation for ordinal data.

## 7. Results

The findings of respondents to the survey are grouped by the four research questions and described in this section. We provide descriptive statistics for all variables, followed by inferential statistics. Of the 72 participants who responded, 62 (86.1%) were male, and 10 (13.9%) were female. Respondents' ages ranged from 19 to 26 years, with most in the 20–21-year-old age category (70.8%). A majority of participants were completing a bachelor's degree (88.9%), while some participants possessed diploma qualifications (11.1%). Regarding the field of study, the largest representation of respondents was from Artificial Intelligence or Data Analysis (29.2%), Industrial Engineering (18.1%), Computer Science (13.9%), and Computer Engineering (9.7%). The remainder of respondents studying other fields (Electrical Engineering, Mechanical Engineering, and Civil Engineering) were poorly represented. The timeframe of respondents' university enrollment (2018 to 2023) shows a growing number of students from 2022 (36.1%).

Figure 1: LLM usage by field of study

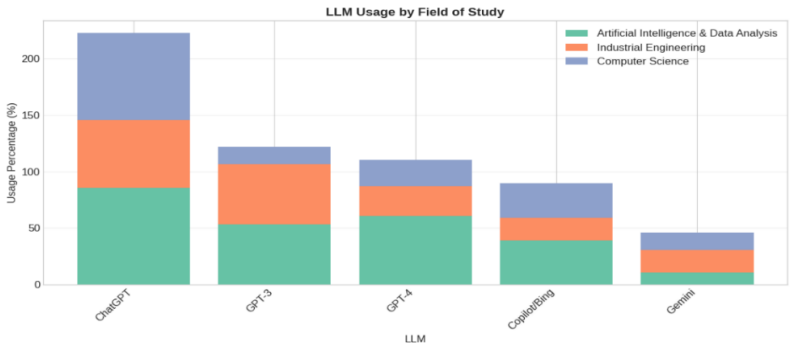


Figure 1 indicates that LLMs use varies by discipline; AI & Data Science students report the widest use (90%), followed by Industrial Engineering (50%) and Computer Science (25%). Industrial Engineering has the most balanced use of the three models (GPT-3, GPT-4, and ChatGPT) at ~50-55%. Computer Science users are lower across the same three models; therefore, they may rely more on other alternatives in their field or adopt a more critical attitude toward using LLMs for coding. Summed use of all models exceeds 100%, indicating that respondents used multiple models. The differences between disciplines could be due to task differences and differences in what is acceptable as knowledge in each field.

### 7.2 RQ1: Awareness, Familiarity, and General Usage Patterns

Almost everyone has heard of large language models, as nearly all (88%) of those responding to the survey acknowledged having at least some familiarity with language models like ChatGPT and Gemini. Thus, that proportion is significantly higher than what chance would predict with a relatively large effect size (Cohen's g = 0.403). Approximately half of respondents reported using ChatGPT (81.9%), while GPT-3 (37.5%) and GPT-4 (31.9%) were reported less frequently. No other LLM was reported to be used by more than 25% of respondents: Copilot (25%), Bard

(11.1%), Gemini (9.7%), Bing Chat (9.7%), and Claude (8.3%). The frequency with which LLMs were used for academic purposes had a broader range of responses than other items on the survey. Weekly (45.8%) was the most common response, followed by monthly (25.0%) and rare (22.2%). Only 5.6% of respondents reported using LLMs daily, while 1.4% reported never using an LLM. Respondents did not provide an equal response in terms of distribution among the frequency categories ($\chi^2$ (4) = 43.17, p < .0001; Cramér's V = .387), but when combined, there was a slight majority (51.4%; 95% CI [39.5%, 63.1%]) of users who used LLMs for daily or weekly purposes.

**Table 3.**

*LLM Usage Frequency and Market Share*

| LLM Platform | Users (n) | Market Share (%) | 95% Confidence Interval |
|---|---|---|---|
| **ChatGPT** | 54 | 75.0 | [63.9%, 83.6%] |
| **GPT-3** | 33 | 45.8 | [34.8%, 57.3%] |
| **GPT-4** | 29 | 40.3 | [29.7%, 51.8%] |
| **Copilot/Bing** | 20 | 27.8 | [18.8%, 39.0%] |
| **Gemini** | 9 | 12.5 | [6.7%, 22.1%] |
| **Bard** | 9 | 12.5 | [6.7%, 22.1%] |
| **Claude** | 6 | 8.3 | [3.9%, 17.0%] |
| **Llama** | 3 | 4.2 | [1.4%, 11.5%] |
| **Poe** | 2 | 2.8 | [0.8%, 9.6%] |

Table 3 shows the usage rates and market share of the 9 Total LLMs platforms used by 72 survey respondents. ChatGPT had the largest number of users at the highest usage rate (75%), followed by GPT-3 (45.8%) and GPT-4 (40.3%). At lower usage rates, Copilot/Bing (27.8%), Gemini (12.5%), and Bard (12.5%) were the most used. At much lower usage rates than the above, Claude (8.3%), Llama (4.2%), and Poe (2.8%) were used. According to the 95% confidence intervals, ChatGPT has a statistically superior market share to all competitors, with the lower bound (63.9%) exceeding the upper bounds of all competitors.

Figure 2: **LLMs Usage Frequency**

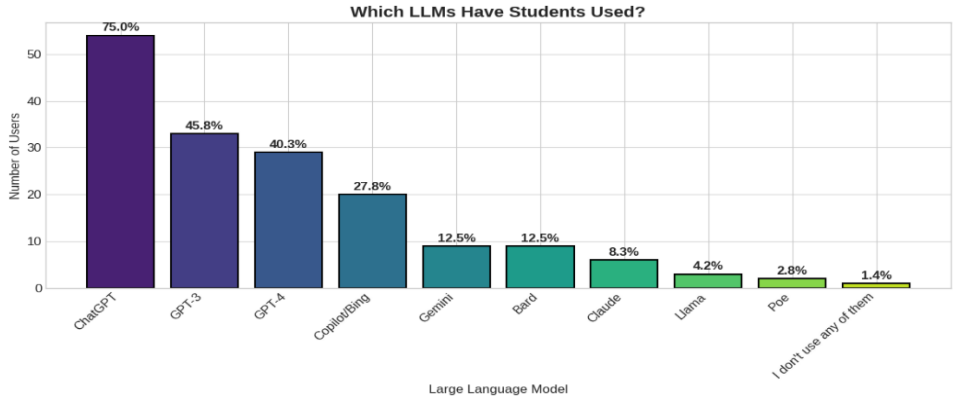

Figure 2 shows that only 1.4% of respondents reported not using any of these models, suggesting near-universal acceptance of LLMs among this student sample. Given the high visibility and familiarity of ChatGPT (GPT3/4), it is likely that these models represent the most significant LLM user group within this sample. By providing students with access to proprietary (Claude, Gemini) and open-source models (Llama), students are likely to use a mix of both for a variety of purposes, such as writing assistance, programming assistance, and information retrieval. The binary "use/non-use" classification used in this study may also mask the degree of variation in the frequency of utilization or for which purpose(s) or by which discipline respondents have utilized LLMs. The findings generally emphasize the need for higher education to no longer engage in debates over whether students use LLMs, but rather to seek insights into how, why, and with what implications for learning and digital literacy, as well as matters of assessment integrity.

An observable and robust relationship has emerged between the parties' level of acquaintance and the frequency of use. Some 70.0% of respondents who self-identified as "very familiar" with LLMs admitted to daily or weekly use. The proportion was markedly lower among those who said they were somewhat or not familiar, at only 28.1%. The association was statistically significant ($\chi^2$ (1) = 12.30, p = .00045) and moderately strong (Cramér's V = 0.413). Very familiar students were almost 6 times as likely to use the tool at least weekly as their less familiar counterparts (odds ratio = 5.96; 95% CI [2.15, 16.53]).

**Table 4.**

*User Typology Classification*

| User Type | Definition | Frequency (n) | Percentage (%) |
|---|---|---|---|
| **Single-LLM Users** | Use exactly 1 LLM | 25 | 34.7 |
| **Dual-LLM Users** | Use exactly 2 LLMs | 18 | 25.0 |
| **Multi-LLM Users (3-4)** | Use 3-4 LLMs | 23 | 31.9 |
| **Power Users (5+)** | Use 5 or more LLMs | 6 | 8.3 |
| **Total** | | 72 | 100.0 |

Table 4 shows the user classifications of the study respondents based on the number of LLMs they report using. The largest proportion of respondents (34.7%, n = 25) were Single LLM Users. The next-largest group was multi-user LLMs, who reported using three to four LLMs (31.9%, n = 23), followed by Dual LLM Users (25.0%, n = 18) and Power LLM Users (8.3%, n = 6). The chi-square goodness-of-fit test confirmed that the user types were different ($\chi^2$ = 12.11; df = 3; p = 0.007), and the large percentage of Single LLM Users provides evidence that many students are not using a diverse set of LLMs in their toolkits.

**Table 5.**

*Chi-Square Goodness of Fit Test for Equal LLM Preferences*

| Models | Observed Selections | Expected Selections (if equal) |
|---|---|---|
| **ChatGPT** | 54 | 18.33 |
| **GPT-3** | 33 | 18.33 |
| **GPT-4** | 29 | 18.33 |
| **Copilot/Bing** | 20 | 18.33 |
| **Gemini** | 9 | 18.33 |
| **Bard** | 9 | 18.33 |
| **Claude** | 6 | 18.33 |
| **Llama** | 3 | 18.33 |
| **Poe** | 2 | 18.33 |
| **Total** | **165** | **165.00** |
| **Test Statistic** | Value | |
| **Chi-square ($\chi^2$)** | 132.65 | |
| **Degrees of freedom (df)** | 8 | |
| **p-value** | < 0.001 | |
| **Cohen's w (effect size)** | 0.897 | |

Table 5 provides results from the chi-square goodness-of-fit test that compares the selections of each of the observed LLMs to their expected selection number based on equal preference for each of the nine LLMs (based on a null hypothesis of equal preference, that is, if there were equal preferences for all LLMs, each would be expected to receive 18.33 selections). The actual selections ranged from 54 for ChatGPT as the most selected LLM to 2 for Poe as the least selected LLM, for a total of 132.65 df=8 and p<0.001, leading to rejection of the null hypothesis. Additionally, the effect size was large (Cohen's w = 0.897), suggesting that preferences among the LLMs differ considerably from equal distributions.

Figure 3: LLM usage Correlation Matrix

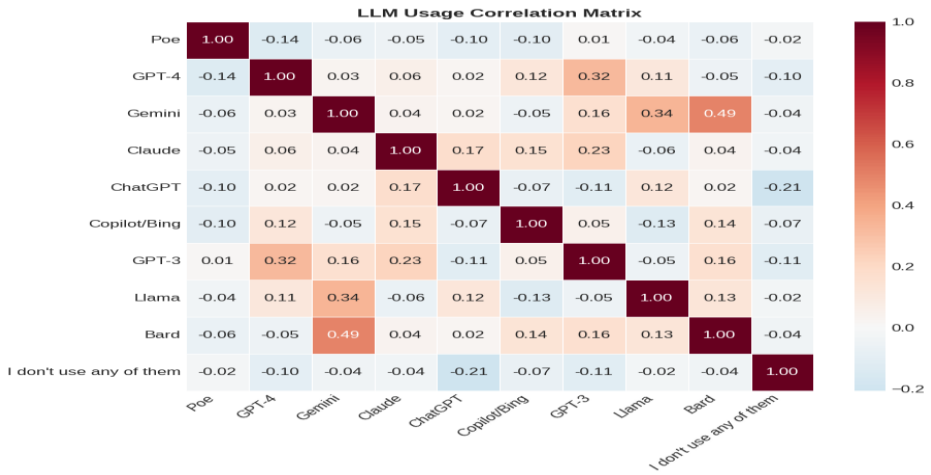

Figure 3 shows a correlation matrix based on students' self-reports of usage for nine different LLMs (see Table 2). The correlations between all possible pairs range from -0.14 to +0.49; for example, the highest correlation (+0.49) occurs between Gemini and Bard, suggesting that students who frequently use one of the Google-based models are also very likely to use the other. It seems that those students might demonstrate some degree of loyalty to the Google brand, or they may do so because they can experiment with similar applications at or about the same time. The second most positively correlated LLMs (+0.32) are GPT-3 and GPT-4; therefore, one can infer that students who have a positive experience with an older OpenAI LLM are very likely to continue using future-generation OpenAI LLMs. On the other hand, there are multiple pairs of LLMs that are negatively correlated, e.g., ChatGPT and GPT-3 (-0.11), ChatGPT and GPT-4 (0.02), and so on for the remaining pairs. The correlations are low for each negatively correlated pair; therefore, the near-zero or slightly negative correlations between ChatGPT and the remaining LLMs (e.g., Claude = +0.17) suggest that students who frequently use ChatGPT do so infrequently with other LLMs. Given that they do not have an application to use as a companion (unless they have sufficient functionality), this may indicate a habitual usage preference. Overall, the data suggests that students are not using LLMs as companions with each other, but when they do use LLMs, they tend to use one brand or type of LLM more than once. Consequently, the findings support the conclusion that student usage of LLMs is primarily (or at least mostly) not related to the LLM's developer, but that students may be forming small, damped groupings for LLMs based on the LLM's developer, which in turn raises some strategic and management implications for future utilization of LLMs in education.

**7.3 RQ2: Beliefs About LLMs and Traditional Learning**

The LLMs were met with uncertainty when respondents were asked whether they could serve as substitutes for textbooks and lectures. 11.1% answered yes, 40.3% no, and the largest group, 48.6%, answered maybe/comparisons of proportions of No responses to the null: 0.50. No statistically significant evidence rejects the alternative hypothesis of no substitution. Thus, implementation seems highly uncertain given evidence of true ambivalence. Looking at the helpfulness of LLMs for creating a better understanding of difficult topics, 16.7% said they were greatly helped, 44.4% reported being moderately helped, 13.9% said they had a little success in getting assistance, and 2.8% thought they didn't fit their situation. A limited number of participants (16.7) did not answer the question, and the combined moderate and great helpers (73.3%) were returned from the adjusted total respondents (44 of 60) but were split.

**7.4 RQ3: Coding-Related Use of LLMs**

In the study, most respondents (62.5%; 95% CI [50.3% – 73.8%]) reported using LLMs to assist with coding tasks. This was statistically significantly greater than 50% (p = 0.005 by binomial test). For these LLM

users, the most reported frequency was weekly (37.8%), followed by rarely (28.9%), monthly (22.2%), and finally daily (8.9%). As for what coding tasks were generally delegated to LLMs, debugging existing code (80.9% of coders) and writing new code (74.5% of coders) were the most regularly delegated to LLMs, followed by explaining code logic to others (63.8% of coders), generating documentation for code (42.6% of coders), and optimizing the performance of code (38.3% of coders). Delegating the task of learning additional programming languages (31.9% of coders) was reported less frequently. Respondents' ratings of the quality of code suggestions provided by the LLMs varied. Overall, all 72 respondents (including those who were not coders) rated LLMs' code suggestions as follows: 11.1% rated code suggestions as excellent, 30.6% rated them as good, 29.2% rated them as fair, and 6.9% rated them as either poor/very poor. Additionally, 22.2% of respondents did not rate any code suggestions provided by LLMs. Among those sampled who rated code suggestions, 41.7% (30 of 72) rated them as excellent or good. A Spearman's rho of -0.09 between the frequency of LLM use for coding and the perceived quality of LLM code suggested that LLM users who used LLMs more frequently were not more or less critical of the quality of the code they produced.

**7.5 RQ4: Perceived Advantages and Disadvantages**

Respondents indicated that "time saving" (72.2%), "greater understanding of content" (48.6%), "vast amounts of information (44.4%), "increased productivity" (43.1%), and "better coding skills" (36.1%) are the most common benefits of these technologies, according to the survey. Conversely, participants indicated that the "lack of contextual understanding" (37.5%), "inaccurate information" (36.1%), and "over-dependency on technology" (31.9%) were the most common reasons for not taking advantage of these innovations. Ethical concerns were noted by 16.7% of respondents, and "lack of source verification" was mentioned by 12.5%. In summary, while LLMs offer many significant functional advantages, they also pose serious risks and therefore deserve a balanced approach to their use in education.

**8. Discussion**

Our results indicate a trend of selective, critical adoption of LLMs among our students rather than unquestioned reliance. Students were familiar with LLMs but did not use them routinely; 90% of students knew of LLMs, but only 51% used them often enough to be considered a regular activity. Prior experience strongly predicted continued use (OR = 5.96, p = 0.00045). However, a high proportion of learners familiar with LLMs chose not to use them, demonstrating a degree of agency rather than adopting the technology solely on perceived benefits. Students are weighing the suitability of each task against their risk perception. This extends the work of O'Neill et al. (2025) to show that familiarity and habitual use occupy separate spaces. Students were also unanimous in rejecting the replacement of traditional education with LLMs. While only 11% supported that notion, a large majority of students (73%) agreed that LLMs enhanced their

comprehension of difficult subjects. This entrenches LLMs as complementary tools for learning, NOT as replacements for the instructor. This view supports O'Neill's perspective of LLMs working as scaffolds and concerns overreliance, expressed through Hossain et al.'s work (2025). The coding results corroborated a perception rooted in calibrated faith. LLMs were widely employed for debugging and code creation, yet students were tentative in their use for more discerning programming tasks. Moreover, because there was no relationship between LLM use frequency and confidence levels, students used LLMs less for complex work ( $\rho$ = -0.09, p = 0.553). These findings accord with the results of Kazemitabaar et al. (2023), rather than those of Peng et al. (2025), the former indicating that gains in productivity need not result in gains in learning in-depth conceptual material. This result highlighted concerns about the quality of education in the Arabic-speaking region, as well as issues related to language. Students saw shortcomings in truthfulness, source clarity, authentication of the sources from which it draws information, and provision of language support for LLMs, citing that provision regarding the Arabic language was paltry, whereas citations helped develop authenticity alongside accountability to the education system. Results illustrate the unsuitability of the traditional TAM model for explaining LLM adoption, as perceived usefulness and ease of use fail to fully account for students' guarded approach to them. Recent research has emphasized the importance of trust in the adoption of AI, and the proposal of the Trust-Calibrated Technology Acceptance Model (TC-TAM), in which trustworthiness (accuracy, source clarity, and authentication) is the moderating variable, rather than the TAM model (Davis, 1989) (Glikson & Woolley, 2020; Mirabile et al., 2026). Greater emphasis on source clarity (through retrieval-augmented generation and appropriate source citations) and appropriate training will result in more aware and informed use and adoption among students.

**9. Conclusion**

The study examines the overlooked absence of students' voices in the LLMs debate and highlights how they approach technology with a combination of caution, pragmatism, and understanding, rather than the unconditional optimism or cynicism associated with most techno-centric higher education LLM narratives. Three important insights emerging from the results are that while students are familiar with LLMs at almost 90%, this does not necessarily translate into habitual use. Most students were not seeking any radical overhaul of their education when presented with a hypothetical proposition; only (11%) admitted to preferring the typical education method, whereas close to half remained ambivalent. Most students (90%) demand that LLMs cite sources because LLMs must uphold, not compromise, existing academic standards. Theoretically, the proposed trust-calibrated TAM (TC-TAM) model, which adds the concept of trustworthiness as a moderator between usefulness and intention to use in existing TAMs. In practice, higher education facilitators are expected to deliver critical literacy training (hallucination detection and cross-check), in which

developers must pay attention to students' identified features (citations, supporting languages, confidence indicators, etc.). The most important limitation of the study is its small, cross-sectional, single-university design, but it opens the door to larger, cross-cultural, longitudinal, and control-group-based studies that test and validate its findings. Students use LLMs as an aid to productivity, not as a replacement for learning, and they are more interested in the trustworthiness side, as suggested by the over-reliance problem. Eventually, the question is: "Given LLMs are going to be there, whether higher education providers and developers possess the sagacity to work in tandem with the students who, in point of fact, are readier than anyone to work with LLMs."

**Disclosure**: The author used Grammarly and ChatGPT for grammar checking and proofreading.

**References**

Akpinar, N. J., Avula, S., Lee, C. J., Dang, B., Razat, K., & Murdock, V. (2026, April). LLM or Human? Perceptions of Trust and Quality in Research Summaries. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems* (pp. 1–25).

Al-Garaady, J., & Albuhairy, M. M. (2023). ChatGPT's capabilities in spotting and analyzing writing errors experienced by EFL learners. *Arab World English Journals, Special Issue on CALL*, (9).

Albuhairy, M. M., Al-Garaady, J., & Alblwi, A. (2023). A proposed framework for human-like language processing of ChatGPT in academic writing. *International Journal of Emerging Technologies in Learning (iJET)*, *18*(14).

Bernabei, M., Colabianchi, S., Falegnami, A., & Costantino, F. (2023). Students' use of large language models in engineering education: A case study on technology acceptance, perceptions, efficacy, and detection chances. *Computers and Education: Artificial Intelligence*, *5*, 100172.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, *13*(3), 319–34.

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, *14*(2), 627–660.

Hossain, S., Khanam, S., Haniya, S., & Nasr, N. R. (2025). AI Literacy and LLM Engagement in Higher Education: A Cross-National Quantitative Study. *arXiv preprint arXiv:2507.03020*.

Jang, C. (2024). Coping with vulnerability: The effect of trust in AI and privacy-protective behaviour on the use of AI-based services. *Behaviour & Information Technology*, *43*(11), 2388-2400

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, *103*, 102274.

Kazemitabaar, M., Chow, J., Ma, C. K. T., Ericson, B. J., Weintrop, D., & Grossman, T. (2023, April). Studying the effect of AI code generators on supporting novice learners in introductory programming. In *Proceedings of the 2023 CHI conference on human factors in computing systems* (pp. 1-23).

Lee, H., Atif, A., & Kang, K. (2026). Analyzing AI utilization in education through learner question types: A constructivist approach. *Australasian Journal of Educational Technology*, *42*(1), 79–96. https://doi.org/10.14742/ajet.10657

Li, R., Li, M., & Qiao, W. (2025). Engineering students' use of large language model tools: An empirical study based on a survey of students from 12 universities. *Education Sciences*, *15*(3), 280.

Mirabile, M., Corazza, G. E., & Alonso-Moral, J. M. (2026). Trust in human–AI collaboration in finance: a bibliometric–systematic literature review. *AI & SOCIETY*, 1–30.

Lyu, W., Wang, Y., Chung, T., Sun, Y., & Zhang, Y. (2024, July). Evaluating the effectiveness of llms in introductory computer science education: A semester-long field study. In *Proceedings of the eleventh ACM conference on learning@ scale* (pp. 63-74).

O'Neill, S., Mulgrew, D., & Bagdasar, O. (2025). On the Use of Large Language Models for Improving Student and Staff Experience in Higher Education. *Open Education Studies*, *7*(1), 20250086.

Peng, Q., Zhang, C., Mangal, R., Pasareanu, C., & Jia, L. (2025, April). Random Perturbation Attack on LLMs for Code Generation. In *2025 IEEE/ACM 4th International Conference on AI Engineering–Software Engineering for AI (CAIN)* (pp. 285–287). IEEE.

Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*, *20*(2), 1-24.

Raihan, N., Siddiq, M. L., Santos, J. C., & Zampieri, M. (2025, February). Large language models in computer science education: A systematic literature review. In *Proceedings of the 56th ACM technical symposium on computer science education V. 1* (pp. 938–944).

Rasnayaka, S., Wang, G., Shariffdeen, R., & Iyer, G. N. (2024, April). An empirical study on usage and perceptions of llms in a software engineering project. In *Proceedings of the 1st international workshop on large language models for code* (pp. 111-118).

Singal, A., & Goyal, S. (2025). Comparative evaluation of AI platforms "Google Gemini 2.5 Flash, Google Gemini 2.0 Flash, DeepSeek V3 and ChatGPT 4o" in solving multiple-choice questions from different subtopics of anatomy. *Surgical and Radiologic Anatomy*, *47*(1), 193.

Scherer, R., Siddiq, F., & Tondeur, J. (2019). The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & education*, *128*, 13-35.

Xu, J., Shen, R., Li, J., & Hu, D. (2025). Artificial intelligence literacy among Chinese Master of Education students: Current situation, influencing factors and generative pathways. *Australasian Journal of Educational Technology*, *41*(6), 1–17. https://doi.org/10.14742/ajet.10656.