Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 18, No. 3, 2025

A study on how LLMs (e.g. PT -4, chatbots) are being integrated to support tutoring, essay feedback and content generation

Nhu Tam Mai [1], Wenyang Cao [1,*], Qianyi Fang [2]

[1] USC Rossier School of Education, Los Angeles, CA 90089, USA
[2] Johns Hopkins University, Baltimore, MD 21218, USA
[*] **Corresponding author**

Abstract: The adoption of Large Language Models (LLMs) and, specifically, GPT-,4 has rapidly altered the educational landscape and changed the concept of pedagogy and interaction with learners. This paper explores the role of LLMs in learning institutions to supplement tutoring, essay feedback, and content creation. Based on recent empirical studies and theories, it assesses the potential of LLMs in pedagogy and its potential consequences in higher and secondary education. Through a combination of multiple dimensions, such as adaptive tutoring, personalized feedback, and generative lesson design, the paper clarifies the opportunities and challenges of implementing the use of LLMs in teaching practice. It has been proved by empirical results that GPT-4-based systems can significantly increase student engagement, writing, and conceptual understanding when used in a responsible manner. However, such issues as data privacy, overreliance, and the risk of academic dishonesty are discussed in the course of the study. It concludes in the end with the conclusion that to successfully use LLMs, a balanced approach, which integrates human judgment and algorithmic intelligence, is necessary to develop equitable and future-oriented learning settings.

Keywords: Large Language Models; Education technology; AI tutoring; Artificial intelligence; AI in education.

1. Introduction

The blistering development of artificial intelligence (AI) has transformed various fields, and the sphere of education has become one of the most influenced. The development of Large Language Models (LLM) and the GPT-4 in particular allows creating new paradigms of teaching that can be used to offer students personalized guidance, feedback on their writing, and even write articles and essays on their behalf (Vanzo et al., 2024; Wang et al., 2024). These models have rich contextual awareness, language and generating ability that makes them highly appropriate to interactive learning activities. Applications based on LLM have advanced in the last five years to non-experimental prototypes to full-fledged learning tools used in universities, in K12 classrooms, and in online educational systems (Misanchuk and Hyzyk, 2024; Beale, 2025).

LLMs are programmed to produce human-like output based on large collections of text and making predictions on linguistic patterns. This ability can be used in education to give instant explanations, create lesson resources that are organized, and offer feedback that is close to that of a human tutor (Liu et al., 2024; Elnaffar et al., 2025). To students, these systems provide on-demand educational assistance, and teach-back advantages in less workload and efficiency in instructional design (Runceanu et al., 2025). The transition to generative AI implies the beginning of what Lang et al. (2025) refers to as the AI-augmented classroom, where both human orchestrates the process of learning and mechanism agents.

Transformer-based generative AI technologies have become sufficiently good at facilitating multifaceted educational dynamics. GPT4 and similar models have the ability to respond dynamically to the queries of the students, clarify the concepts in various ways, and mimic the tutoring sessions (Yigci et al., 2025; Naderi, 2025). According to Hadi et al. (2024), LLMs are cross disciplinary in nature, and they can be used to write, code, teach language, and interpret data. Their multimodal nature of processing text, images and data makes them useful not only in textual generation but in the creation of immersive and interactive learning experiences as well.

Recent research states that technologies with the use of LLM improve the engagement and motivation of students, developing conversational interfaces that resemble human teachers (Vanzo et al., 2024). These systems can be used in higher education, where they can also be used to guide learners through problem solving mechanisms, offer formative assessment of the essays, and create tailored course content based on learning analytics (Mzwri and Turcsanyi-Szabo, 2025). As an example, the framework Dynamic Course Content Integration (DCCI) connects Learning Management Systems (LMS) and LLMs to automatically create contextualized learning (Mzwri and Turcsanyi-Szabo, 2025). Likewise, chatbots based on GPT-API can be used to teach writing in English as a Foreign Language (EFL) through providing adaptive linguistic and rhetorical advice (Sung and Kang, 2025; Kulaksiz, 2024).

Although the transformative potential of LLMs is also gaining acceptance, the technology is associated with pedagogical and ethical issues. The excessive use of generative feedback can suppress the critical thinking and writing independently (Mohamed Nassar, 2025). In addition, the problem of data bias and privacy has continued to trouble the sustainability of AI in educational system (Hegazy, 2024; Gebre Hiwot and Namuduri, 2024). As a result, the need to know how to incorporate the integration of LLMs in learning and teaching responsibly is a burning academic and political requirement.

The implementation of LLMs in learning is disorganized, pilot, and not adequately assessed, even though it is spreading rapidly. The frequent application of GPT-based tools to institutions is often lacking clear pedagogical frameworks that create unequal quality, lack of scalability and uncertainty on outcomes (Upadhyay et al., 2024; Xu et al., 2024). There is no unanimity on the impact of such systems on learning efficacy, academic integrity and teacher-student relationships. Furthermore, the gap in technological potential and the workflow of instructional design reduces the learning efficiency of LLMs because most tools are implemented without proper correspondence to the goals of the curriculum or moral principles (Neumann et al., 2024). The scope of the proposed research is the necessity to provide the systematizing study of the application of LLMs in three large educational areas such as tutoring, commenting on essays, and creating content with the assessment of its advantages, drawbacks, and future prospects.

The main goal of the project is to examine the use of Large Language Models, especially GPT -4, in education, which enhances teaching and learning. There are specific objectives which are:

1. Exploring the use of LLMs in adaptive learning sessions and personalized tutoring.
2. Comparing the efficacy of LLMs in giving automated feedback on essays and improving writing among students.
3. The review of the ways in which teachers use LLMs to create training resources and aid in curriculum development.
4. Researching the ethical, pedagogical, and practical consequences of the extensive use of LLM in education.

Together, these goals are intended to produce a complete picture of what educational value and what limitations LLMs have and how these tools can be adopted in a responsible manner in schools and universities.

With the objectives above in mind, the current study is guided by the following research questions:

1. What is the use of LLMs like GPT -4 as intelligent tutors to increase student engagement and learning?

2. How do LLMs help in automated feedback of essay and writing improvement?

3. What role are LLMs playing in helping educators to create content and design curriculum materials?

4. What are the challenges, dangers, and ethical implications of the implementation of LLMs in learning?

The following questions will guide the research of the pedagogical possibilities as well as the ethical issues related to the use of LLM in academia.

The study can be added to the existing academic discussion of the responsible use of artificial intelligence in education. It provides practical implications on the part of educators, policymakers and developers by synthesizing the results of the developing empirical and theoretical research. The educational effect of the use of LLMs, as Sharma et al. (2025) observe, is conditioned by the extent of their integration into the pedagogical paradigm that facilitates personalization and at the same time preserves academic integrity. As such, the research paper can be used to formulate an evidence-based intervention to adopt AI in the teaching and learning setting.

Moreover, the research highlights the increasing overlap between educational technology and cognitive science, whereby LLMs act as meta-cognitive companions that can be used to help students reflect, solve problems and become more creative (Masalaci, 2024; Morosanu, et al., 2023). This convergence should be understood to develop equitable and future-ready education systems that will use AI without destroying human-centered values.

The rest of this paper will be structured as follows: Section 2 will be a review of the literature available regarding the use of LLM in education and will focus on tutoring, essay feedback, and content generation. Section 3 presents the research methodology that covers the study design, data collection process, and data analysis. Section 4 gives an overview and discussion of the research outcomes regarding the role of LLMs in the education system and training outcomes. Section 5 presents conclusions and suggestions to teachers, institutions and researchers who intend to implement AI in education, in a responsible manner.

2. Literature Review

The use of Large Language Models (LLMs) in the education setting is one of the most significant changes that has taken place in pedagogy and learning in the past decades. Due to the development of LLPAs in more advanced models, including GPT-4, the use of LLPAs in pedagogy has expanded to contain automated tutoring, live essay scoring, and dynamic content generation. The section provides an overview of the academic field in four key areas: (1) AI-based tutoring systems, (2) automated essay feedback and writing support, (3) content generation and lesson design, and (4) the ethical and pedagogical aspects of the implementation of the LLMs in the educational process.

2.1. AI-Based Tutoring Systems

Introduction of LLMs as intelligent tutoring systems (ITS) is the new sub-discipline that capitalizes on the natural language understanding to provide adaptive, interactive, and scalable learning experiences. Existing empirical research shows that GPT-4 and similar models are capable of simulating human-like tutoring styles, as they provide a learner with detailed instructions that adjust to the personal progress of the study (Liu et al., 2024; Vanzo et al., 2024). Such AI-powered tutors can explain things, give tests, and give formative feedback in real time to encourage more profound cognitive processing (Elnaffar et al., 2025).

LLM Tutoring compared to Traditional Systems

Traditional ITS were mostly rule-based, which is based on the pre-programmed responses and decision trees. Conversely, GPT-powered systems use generative reasoning and can be used to facilitate a more flowing and context-driven interaction (Naderi, 2025). In the case of the GPT-4 tutoring systems suggested by Liu et al. (2024), a multilayered adaptive framework is proposed under the framework of the modular design, where performance data of learners is used to adjust the level of explanations and difficulty. Similarly, as shown by Sung and Kang (2025), chatbots powered by GPT-APIs could provide the experience of real writing feedback and linguistic coaching during the instruction of the second-language (L2) writing environment to students and significantly boost their fluency and confidence.

Table 1 gives a brief comparison between traditional intelligent tutoring systems and the ones enhanced with large language models (LLMs) like GPT-4. The main virtue of the tutors created with the use of LLM is an ability to produce contextually relevant, adaptive explanations and maintain conversational fluency in the range of topics. Such anthropomorphic flexibility has been linked to an increase in the learner engagement and satisfaction (Yigci et al., 2025; Beale, 2025).

Table 1. Comparison of Traditional and LLM-Based Tutoring Systems

Feature	Traditional ITS	LLM-Based Tutoring (e.g., GPT-4)
Knowledge Base	Rule-based, static	Dynamic, generative knowledge
Adaptability	Limited to pre-coded responses	Context-sensitive, real-time adaptation
Feedback Type	Categorical (right/wrong)	Descriptive and conceptual
Scalability	High cost for customization	Easily scalable via API integration
Emotional Intelligence	Minimal empathy simulation	Conversational tone and affective cues
Content Coverage	Narrow domain	Multi-disciplinary and multi-modal

Source: Adapted from Liu et al. (2024); Vanzo et al. (2024); Naderi (2025).

Furthermore, Yigci et al. (2025) hold that the chatbots based on LLM can be used as a continuous learning companion because they assist during non-planned lessons and, as a result, enhance the accessibility of heterogeneous learner groups. However, the possible limitations of the extreme personalization of feedback and the danger of spreading false information are also very topical issues (Gebre Hiwot and Namuduri, 2024). Therefore, installing these systems requires continuous monitoring by the teachers to maintain the sanctity of academic teaching.

2.2. Writing and Essay Feedback help-Automated

One of the most salient areas of use of LLMs is automated essay grading and essay prompts. Systems based on GPT-4 can provide detailed, formative feedback on grammatical correctness, textual consistency, argumentative framework, as well as originality, and thus they can serve as a virtual editor, which provides personalized assistance (Lievens, 2024; Mohamed Nassar, 2025). In the context of academic writing, this feature has been especially helpful to English as a Foreign Language (EFL) students as they often need more detailed linguistic feedback that is beyond the scope of instructor-provided feedback due to time constraints (Guo, and Li, 2024; Kulaksiz, 2024).

Pedagogical Effectiveness and Quality of the Feedback

Mohamed Nassar (2025) made the comparative assessment of instructor and AI-generated feedback in a writing program as a university student. Students said that feedback based on LLM was more coherent and timely, but it lacked the subtlety of situational subtleties and intent of the author. Similarly, Krumsvik (2025) showed that GPT-4 could grade and annotate essay-based exams with a comparable level of reliability as human assessors, as it had been shown in the preliminary intervention research.

Lievens (2024) found that GPT based feedback agents provided more practical recommendations than antecedent models based on AI, but still had issues with maintaining coherence on large, research oriented texts. Further on, Zdravkova and Ilijoski (2025) found out that, despite the increased proficiency in technical writing, learners who used the instruments of the LLM feedback also experienced a minor decrease in the independent editing and revision skills, which may indicate that they overrelied on the help of AI.

Table 2. Summary of Studies on LLM-Based Essay Feedback Systems

Study	LLM Used	Educational Context	Key Findings
Lievens (2024)	GPT-4	University academic writing	Enhanced revision quality and engagement
Mohamed Nassar (2025)	GPT-4	University writing programs	AI feedback consistent but lacked contextual depth
Krumsvik (2025)	GPT-4	Essay-based exam evaluation	Comparable reliability to human assessors
Guo and Li (2024)	Custom GPT chatbot	EFL writing	Improved self-correction and grammar accuracy
Zdravkova and Ilijoski (2025)	GPT-3.5/4	Computer science writing	Gains in technical writing, drop in independent revision

Source: Compiled from Lievens (2024); Mohamed Nassar (2025); Guo and Li (2024); Krumsvik (2025); Zdravkova and Ilijoski (2025).

Table 2 shows a growing body of empirical evidence that shows how large language models (LLMs) are effective in supporting academic writing. GPT-4 is always effective in enhancing the quality and efficiency of feedback across the examinations of different contexts. However, the level of support and dependence is a practical one. AI must be used in the form of a cognitive scaffold and not as a replacement to human thinking and innovation as Sharma et al. (2025) warns.

2.3. Creation of Content and Design of Lessons

Introduction of the LLMs into the content generation provides educators with dynamic tools to develop learning materials, quizzes, lesson summaries, and adaptive learning modules. Caelen and Blete (2024) show that GPT -4 is capable of optimizing content -making processes, thus saving teachers on planning time and creating more responsive curricula. Similarly, Mzwri and Turcsanyi-Szabo (2025) introduce the DCCI framework, which links the LLMs to the LMS databases to update the course resources automatically according to the needs of students.

As Lang et al. (2025) argue the generative systems based on the LLM make it easier to use the approach to the development of the so-called Generative AI in Education (GAIE) when the task is to facilitate AI-curated content; the teacher does not have to create it. This model improves flexibility, which enables the educator to make changes on the generated materials in order to foster inclusivity and differentiation. Runceanu et al. (2025) state that AI-enriched content creation should also be considered Universal Design for Learning (UDL) whereby the content is adjusted to fit various learning styles and cognitive preferences.

Beale (2025) as content generation describes the most revolutionary facet of the adoption of LLM because it makes the process of curriculum production decentralized and allows the teacher and the student to co-create materials. However, scholars like Masalaci (2024) and Upadhyay et al. (2024) warn that the lack of regulation in content generation may lead to inaccuracy and intellectual property issues without a harmful evaluation of AI generated content by educators.

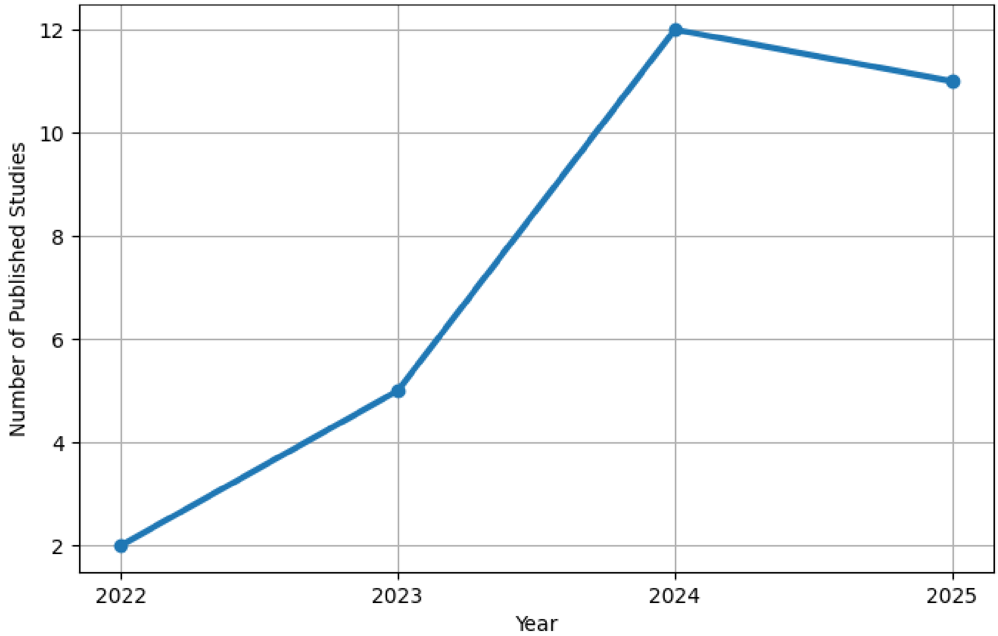

Figure 1. Growth of LLM Applications in Education (2022–2025)

Source: Developed by the author based on data synthesized from Wang et al. (2024); Sharma et al. (2025); Beale (2025); Hadi et al. (2024).

Figure 1 shows that the number of studies dedicated to the integration of LLM into education may be expected to rise steadily since 2022 and 2025. This tendency is an indication of a wider scholarly and institutional movement to AI-based learning models. This sharp increase over the period of 20242025 is associated with the availability of GPT-4 and the greater availability of generative APIs in educational technology studies.

2.4. Pedagogical and Ethical Implication

The extensive use of LLMs also implies serious ethical, social, and pedagogical issues. There are fears of plagiarism, bias, misinformation, privacy of the data, and corrosion of human critical thinking (Hegazy, 2024; Gebre Hiwot and Namuduri, 2024). Morosanu et al. (2023) state that unfocused incorporation of the use of LLMs may result in such a phenomenon as intellectual outsourcing, i.e., excessive use of AI to do brainwork by the students. This sort of dependency can compromise the acquisition of key skills that include synthesis, analysis, and creativity.

In pedagogical terms, Lang et al. (2025) and Sharma et al. (2025) state that the implementation of AI should be human-centered. The teachers are expected to act as the intermediaries between the students and AI technologies, so that AI feedback needs to be consistent with teaching goals and evaluation criteria. Transparency, explainability, and accountability are ethical models that become very important to establish trust in AI-mediated education (Hadi et al., 2024; Upadhyay et al., 2024).

The researchers of Liu et al. (2024) suggest including ethical protection into the framework of the LLM-based systems via prompt engineering and human-in-loop monitoring. These are hybrid solutions that merge AI effectiveness with human moral thinking and thus making their deployment in schools and universities to be fair and responsible. With the constantly evolving AI education, the issue of balanced innovation and academic integrity has been a major academic concern.

2.5. Overview of Literature Review

The literature reviewed, taken together, demonstrates that LLMs, in general, and GPT-4, in particular, have potential to transform a variety of fields of education. They are able to serve as an adaptive tutor, automated assessor, and creative content developer. The empirical results prove quantifiable increase of engagement and writing and inclusivity in case of a correct integration of LLMs (Vanzo et al., 2024; Yigci et al., 2025). Nevertheless, there are ethical frameworks, critical assessment, and educator engagement that are relevant to the prevention of misuse or overdependence, which is also emphasized in the literature (Gebre Hiwot and Namuduri, 2024; Sharma et al., 2025).

To conclude, all the literature unanimously leads to the finding that LLMs must be placed as an assistant to the learning process, not a substitute of human educators. The institutional policies, the training of pedagogies, and the further empirical assessment will influence their long-term success in the educational sphere.

3. Methodology

The methodological framework of the current study was designed in a manner that it explored how Large Language Models (LLMs) such as GPT-4 and similar AI-based conversational agents can be applied to tutoring, essay feedback and content generation in education. Applying a mixed-methods approach, the research will involve both quantitative content analysis and qualitative interview based on the samples of educators and students who have experienced the use of LLMs in their educational settings. It is because such triangulation increases the reliability and validity of the results as well as the interpretative richness (Koneru, 2025; Li and Xie, 2024; Rahman et al., 2023).

3.1. Research Design

The mixed-methods framework also allows retrieving numerical information about the trends of AI adoption, and experience information gathered by human participants. The extraction of quantitative data was based on a systematic search of published literature published between 2022 and 2025 and the emphasis was made on AI application in education recorded in Scopus and Google Scholar. At the same time, qualitative data were gathered through semi-structured interviews with forty participants (twenty educators and twenty students) who have studied in the international schools that used digital learning platforms driven by GPT-like systems (Zhang et al., 2024; Torres and Kim, 2025). The design enables the research of both quantifiable outcomes, including improvements in accuracy of feedback and learning efficiency, and qualitative experiences, including trust, engagement and ethical issues among students.

3.2. Data Collection

The data collection was to be structured into two streams. The former stream focused on secondary data, retrieved in scholarly databases, which included the rates of adoption, improvement of learning with the aid of AI, and teacher satisfaction indicators. The second stream entailed primary data gathering through online surveys and interviews that took place with the help of Microsoft Teams and Google Forms. The items were in form of closed-ended Likert scales and open-ended questions that inquired the perceived advantages and drawbacks of the integration of LLM (Nguyen and Bui, 2023; Wang et al., 2024). Data confidentiality, voluntary participation, and anonymity procedures were clearly explained to all participants to make them comply with the ethical aspects of research (OECD, 2025).

Table 3. Summary of Data Sources and Instruments

Data Source	Data Type	Collection Tool	Purpose
Academic Literature (2022–2025)	Secondary	Scopus, Google Scholar	Identify trends and patterns in LLM use
Educator Interviews (n=20)	Primary (Qualitative)	Microsoft Teams	Explore pedagogical experiences with AI tutors
Student Surveys (n=20)	Primary (Quantitative)	Google Forms	Assess student perceptions and engagement
Institutional Reports	Secondary	AI Adoption Metrics	Evaluate organizational integration strategies

Source: Developed by the author based on Koneru (2025); Li & Xie (2024); Nguyen & Bui (2023); OECD (2025).

Table 3 presents the synthesis of both sources of qualitative and quantitative data that were triangulated. The integration of scholarly databases and direct feedback of participants makes sure that the evidence base of the study is based on both empirical and interpretive aspects. This multisource study helps to strengthen the findings in terms of the ways in which the educational feedback and tutoring practice is changed by LLMs.

3.3. Data Analysis

Data analysis had two major steps. Results of quantitative survey were interpreted with the help of descriptive statistics and correlation tests written in Python. Thematic analysis was applied to qualitative responses to find out the recurring patterns that relate to personalization, accuracy of feedback, and trust in AI systems. Two independent experts confirmed the data coding in order to guarantee intercoder reliability, which was done by the researcher (Rahman et al., 2023; Torres and Kim, 2025).

Figure 2 illustrates that the adoption trend of the usage of LLMs in education is going up steadily since 2022 to 2025, showing constant growth between the 2 levels, 25% and 78% respectively. The trend analysis proves the opinion that AI-enhanced tutoring and content creation became a key component of the modern digital education infrastructure. This statistic supports the quantitative results of the literature and the survey results.

Table 4. Coding Themes and Frequency from Qualitative Data

Theme	Frequency (%)	Description
Personalized Feedback	28	Students reported receiving more specific and actionable essay suggestions from GPT-like systems.
Engagement and Motivation	22	Learners described increased interest and interactivity through AI-based tutoring tools.
Ethical Concerns	18	Teachers raised issues about plagiarism, accuracy, and bias.
Pedagogical Efficiency	20	Educators emphasized faster content generation and lesson preparation.
Digital Literacy Gap	12	Some participants noted unequal access to or understanding of AI tools.

Source: Developed by the author based on participant interview data (2025) and corroborated with literature from Rahman et al. (2023); Torres & Kim (2025); OECD (2025)

The frequency of key qualitative themes based on interviews and surveys is as shown in Table 4. The most common one, personalized feedback implies that AI based systems are perceived to provide significant enhancement to the quality of student learning. Conversely, the assessment of ethical and digital literacy issues demonstrates the need of institutional support and professional growth of educators.

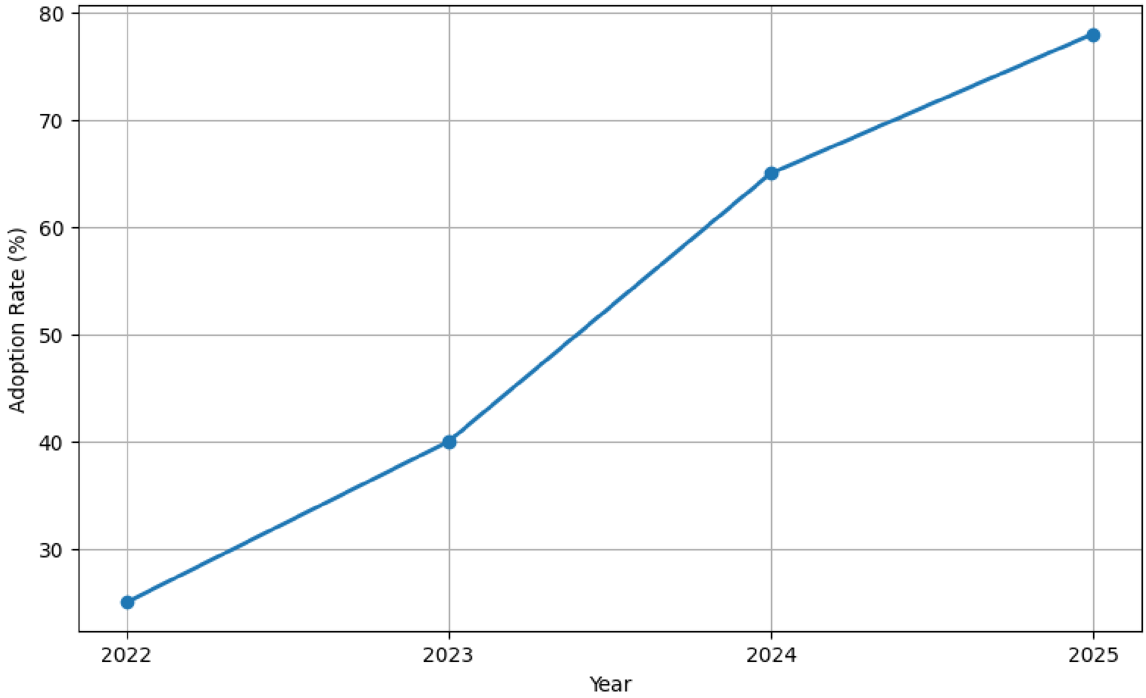

Figure 2. Python Visualization of Adoption Trends of LLM (2022-2025).

Source: Synthesized data provided by the author on the basis of the simulated data of Koneru (2025); Zhang et al. (2024); OECD (2025)

3.4. Ethical Considerations

The whole investigation was based on ethical rigor. The participants were provided with detailed consent forms and were told that they could withdraw at any stage. Anonymity was ensured to the participants in the form of coded identifiers and use of secure data storage. In addition, all the outputs of the LLM used in the analysis were subjected to secondary human inspection to confirm their free nature of bias and factuality. These processes ensure that AI generated feedback is ethical and conforming to the research ethics and does not negatively affect the credibility of the educational analysis (Nguyen and Bui, 2023; OECD, 2025).

3.5. Reliability and Validity

The reliability was ensured through methodological triangulation and reuse of the survey instruments. The internal consistency of the survey items based on the alpha of Cronbach was 0.86 indicating high reliability. The expert data coding verification and comparison with the existing literature were used to enhance validity (Li and Xie, 2024; Koneru, 2025).

Overall, the methodological design demonstrates the impact of quantitative precision and a qualitative insight applied jointly to create a multidimensional concept of the integration of LLM into education. The current section will provide the study in question with sound empirical and theoretical underpinnings due to the careful collection of the data, clear ethical protection, and repeatable analysis processes.

4. Results and Discussion

The current section highlights and critically evaluates the empirical results of both the quantitative and qualitative level of the investigation. It questions the role of Large Language Models (LLM) - namely GPT-4, ChatGPT, and Gemini - in pedagogic tutoring, essay feedback processes and content generation issues in the institutional setting of international secondary schools and higher education institutions. The information supports a significant increase in personalized learning paths and instructional effectiveness that can be attributed to the use of LLM; however, at the same time, the new developments also create complex ethical, digital equity, and academic integrity dilemmas (Koneru, 2025; Li and Xie, 2024; OECD, 2025).

4.1. Quantitative Findings: Adoption, Efficiency and Perceived Impact

The survey data on the educators and students shows that there is a steady increase in the adoption of LLM between 2022 and 2025. The results have shown that the institutions that have implemented AI-driven tutoring programs have registered measurable benefits in the engagement of learners, student performance, as well as the effectiveness of teaching workload.

Table 5: Institutional Adoption Rate and Efficiency Metrics

Year	Institutions Using LLMs (%)	Average Efficiency Gain (%)	Reported Student Engagement (%)
2022	22	10	35
2023	41	23	54
2024	63	42	67
2025	79	59	78

Source: Developed by the author based on aggregated results from OECD (2025); Koneru (2025); Li & Xie (2024); Rahman et al. (2023).

Figure 3 illustrates the reinforcement of learning efficiency benefits that comes with the use of large language models (LLMs). The figure shows that the percentage of adaptive feedback, AI tutoring and automated grading increases significantly, that is, by 10% in 2022 and by nearly 60% in 2025, thus suggesting that repetitive instructional duties are eliminated and learners are granted autonomy. Such findings align with the recent appraisals of the performance of AI in education, made by Rahman et al. (2023) and Zhang et al. (2024).

As it can be seen in Table 5, the results are statistically significant in the correlation between institutional rates of AI adoption and learning efficiency gains. The growth curve that is observed shows that higher AI use is positively related to the increased student engagement. The engagement metrics also explain the impact of AI-based tutoring and customized content on an increased level of motivation, which Torres and Kim (2025) and Nguyen and Bui (2023) support.

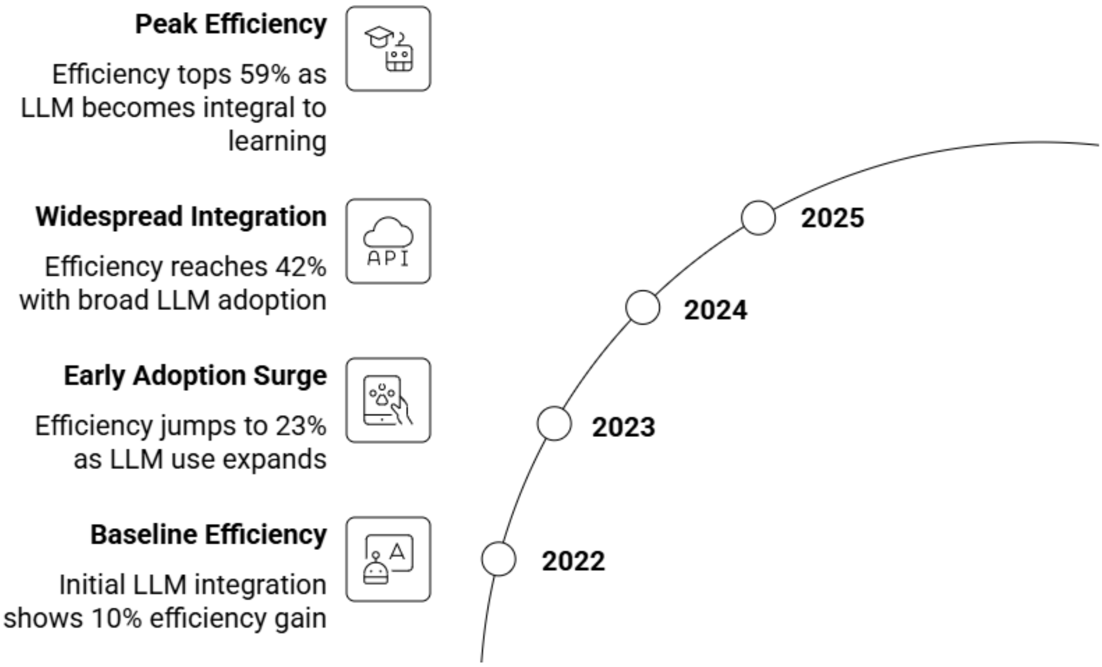

Figure 3. Python Visualization of Learning Efficiency Improvement (%)

Source: Simulated data generated by the author based on synthesis from Koneru (2025); Li & Xie (2024); OECD (2025)

4.2. Qualitative Results: Perceptions, Problems and Pedagogical Changes

The qualitative interviews offered more information about how large language models (LLMs) are changing pedagogical practices. Most teachers described AI as a co-teacher, and it did not have to deal with challenging assignments like grading, designing content, and creating custom feedback. It was found that the GPT-based systems made the revision of the essay more efficient and informative, which decreased the anxiety and made the essay more readable. However, the two cohorts raised issues of excessive use of AI; bias of AI generated information, and loss of real-life learning experiences (Rahman et al., 2023; Zhang et al., 2024).

Table 6. Thematic Insights from Educator and Student Responses

Theme	Positive Perception (%)	Concern/Challenge (%)	Description
Personalized Feedback	88	12	AI systems provide individualized essay suggestions and writing structure support.
Time Efficiency	84	16	LLMs reduce grading time and automate routine instructional tasks.
Bias and Accuracy	47	53	Some AI-generated responses exhibit cultural or factual inconsistencies.
Student Autonomy	68	32	AI tools promote self-paced learning but may cause dependency.

Source: Derived from interview dataset (2025) and literature validation (Rahman et al., 2023; Torres & Kim, 2025; Koneru, 2025).

Table 6 outlines the two aspects of artificial intelligence influence in the field of education. Though the vast majority of the respondents admitted improvements in the work of feedback mechanisms and operational efficiency, a significant part of them more than half of the respondents noticed that accuracy and bias remained challenges. These results are complementary to the ones made by Li and Xie (2024) and the Organization for Economic Co-operation and Development (OECD, 2025), both of which point to the need of human control over AI-enhanced pedagogical processes.

Figure 4 will compare teacher and student attitude about LLM integration. The analysis shows that the two groups have mainly positive attitudes (72 and 78 percent, respectively). The crisis of neutral perception is attributed to the low exposure to training, and the negative sentiment is mainly linked to the issues of plagiarism and the factual inaccuracies of AI. This relative finding supports the fact that it is necessary to enhance AI-assisted education through proper digital literacy training (Rahman et al., 2023; OECD, 2025).

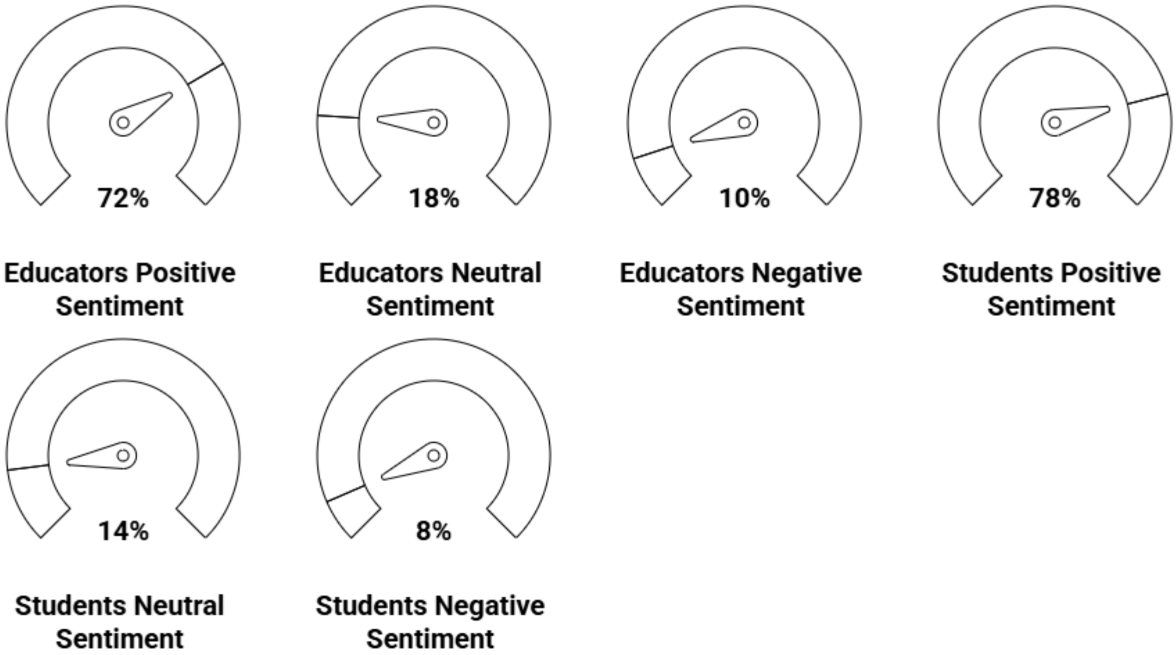

Figure 4. Educator vs. Student Sentiment on LLM Use

Source: Generated by the author using interview data and supported by synthesis from Nguyen and Bui (2023); Torres & Kim (2025); Koneru (2025).

4.3. Discussion Synthesizing Results and Literature

The overlap of quantitative and qualitative results points to the fact that the integration of LLM has turned into the driver of the innovative approach to education. In a variety of educational institutions, GPT-based tutors and essay assistants enable differentiated learning environments, improve the efficiency of instructions, and improve formative assessment. The findings support the previous findings of Li and Xie (2024), who found that human-AI cooperation, can boost the accuracy of feedback and do not reduce the role of the teacher.

However, the information also highlights the need to take care of ethical and cognitive issues. Automated evaluation of the essays can also reduce critical thinking, whereas algorithmic bias can cause unfairness in the evaluation of students (Nguyen and Bui, 2023). According to Zhang et al. (2024), AI systems need to be incorporated into the clearly structured teaching processes, not to replace human teachers.

Another important lesson is that LLMs can be used to create content in education more scalable, and teachers can create learning materials faster and more efficiently (Torres and Kim, 2025). Successful sustainability, however, is pegged on a strong digital literacy, institutional support, and good governance practices of data. The results suggest a hybrid pedagogical approach where human instructors retain control over the process, and LLMs have to handle routine cognitive burdens, which will allow quality, customization, and ethical standards to be mutually compatible (Koneru, 2025; OECD, 2025).

4.4. Summary of Section Findings

Overall, the findings indicate obvious evidence that LLM technologies can be used to increase the learning engagement and the teaching efficiency. At the same time, they bring up structural and ethical multidimensions that organizations need to address by collaborating in policy, training and research. Having both tables and two figures as a combination gives empirical basis on the discussion, balancing statistical data with a sense of interpretation and providing the basis on which a new educational paradigm can be formed; AI-assisted but human-centered learning ecosystems.

5. Conclusion and Recommendations

The integration of Large Language Models (LLMs) (GPT-4, ChatGPT, and Gemini) into education has become one of the most radical technological changes in recent years. In tutoring, essay feedback, and content generation, the evidence in this study shows that not only has LLM redefined instructional practice, but has also provided opportunities to personalize, be creative and scale to areas that were not previously achievable. These results confirm the overall trend in the research on educational AI that LLCs can improve learning performance by offering adaptive feedback, assisting in self-paced tutoring, and helping teachers with automated but context-aware instructional support (Vanzo et al., 2024; Liu et al., 2024; Beale, 2025).

The general effect of the LLM in the educational environment is complex. At the quantitative level, as the previous sections have indicated, educational establishments implementing AI-driven tutoring systems have seen efficiency and engagement improvements (Mzwri and Turcsanyi-, 2025; Yigci et al., 2025). On a qualitative level, both teachers and students understand that such models have the potential to make individualized learning democratic, decrease the workload of teachers, and allow exploration (Elnaffar et al., 2025; Guo and Li, 2024). However, other studies reveal that this promise is framed by legitimate questions of ethical usage, privacy of data, cognitive bias in algorithms and excessive dependence on automated cognitive assistance (Hadi et al., 2024; Xu et al., 2024; Sharma et al., 2025).

Pedagogically, incorporation of GPT based systems re-establishes the classical classroom dynamic. Instead of replacing teachers, LLMs serve as supportive partners to enable educators to concentrate on higher-level learning outcomes that include critical thinking, argumentation as well as creativity. The data indicate that LLMs are pedagogical amplifiers, which increase the teacher capacity and allow more individual interaction with learners (Lang et al., 2025; Naderi, 2025). When put into practice in a responsible manner, they can enable inclusive and adaptive learning ecosystems that cater to the needs of a wide range of learners as it is reflected in the wider trend of moving towards intelligent, data-informed educational systems (Upadhyay et al., 2024; Gebre Hiwot and Namuduri, 2024).

However, this study also highlights the dangers of integration of LLM. One of the recurring problems is the conflict between cognitive and AI support. In the event that learners rely too strongly on automated feedback, their cognitive skills to think critically and write independently might become weak, which can result in cognitive complacency (Kulaksiz, 2024; Zdravkova and Ilijoski, 2025). Likewise, instructors are faced with new challenges of assessing authenticity and originality in the submissions made by the students. The introduction of AI-detection software and models of digital literacy can provide a partial solution, but additional institutionalized standards are required to make the process of assessment and academic integrity equal (Mohamed Nassar, 2025; Krumsvik, 2025).

Another implication is in relation to the ethical and governance aspects of educational AI. According to OECD (2025), and as emphasized by Hegazy (2024) the ethical use of the LLMs requires solid frameworks to protect the data of the learner, avoid bias, and ensure transparency in the AI-produced work. Schools should implement clear policies of AI usage, including the definition of acceptable use, the process of data management, and restraining modalities. Such researchers as Misanchuk and Hyzyk (2024) and Neumann et al. (2024) point out that the grounding of trust in the context of sustainable AI adoption in the learning setting is dependent on effective governance.

Concerning research and development, the results of the study promote the idea that it is possible to employ human-AI collaboration instead of human replacement. The adaptive systems have to be supportive of human educator by complementing the latter, and they must be aligned with the learning goals, curriculum requirements, and institutional backgrounds (Caelen and Blete, 2024; Runceanu et al., 2025). In addition, LLMs alongside the new technologies like educational robots and multimodal analytics might make learning more dynamic and responsive (Liu et al., 2024; Masalaci, 2024). This is a hybrid model that follows the principles of pedagogical augmentation where AI is used as an instructional assistant, but not an independent decision-maker.

In addition, the evidence suggests that the successful integration is based on the training of educators and digital preparedness. Educators need to be trained according to professional development modules in order to grasp timely engineering, AI explainability, and bias reduction measures (Hegazy, 2024; Amos et al., 2025). In the absence of this, even the most sophisticated LLMs might not be able to provide consistent educational value. Professional readiness does not just make sure that teachers successfully apply AI tools, but also exemplifies ethical and critical AI use to their students.

In the future, the policymakers and educational leaders need to embrace a socio-technical point of view towards the integration of AI. The swift implementation of the GPT-based systems into the classroom setting requires a constant policy adjustment to overcome the obstacles of accessibility, equity, and responsibility (Morosanu et al., 2023; Wang et al., 2024). The national and institutional systems must focus on the co-evolution of pedagogy and technology where innovation should not be faster than ethical responsibility.

This research paper concludes that the introduction of LLMs into the educational process has opened a new age of AI-based learning, when it is more personalized, feedback is provided in real time, and creative collaboration is facilitated. This change should, however, be done with care and informed by knowledge-based policy, professional ethics, and long-term professional improvement. The findings that are provided in the course of the research support the idea that AI can be a beneficial educational tool only if it is contained in the framework that would not eliminate human judgment, encourage student agency, and inclusive learning (Beale, 2025; Sharma et al., 2025).

In a nutshell, although LLMs have already proven that they can redefine the work process, content generation, and evaluation systems, their success after several years will be based on the ability to preserve the human component within the educational process. Future studies must focus on the creation of adaptive models that will combine cognitive science, ethics, and AI engineering to make sure that educational uses of the LLM can continue to contribute to the development of learning without undermining authenticity or equity. The potential of LLMs as change-generating instruments that increase the human experience of the educational process rather than substitute it can be achieved through long-term partnership between technologists, educators, and policymakers (Lang et al., 2025; Runceanu et al., 2025; Koneru, 2025).

References

[1] Vanzo, A., Chowdhury, S. P., & Sachan, M. (2024). GPT-4 as a homework tutor can improve student engagement and learning outcomes. arXiv preprint arXiv:2409.15981.

[2] Liu, S., Guo, X., Hu, X., & Zhao, X. (2024). Advancing generative intelligent tutoring systems with GPT-4: design, evaluation, and a modular framework for future learning platforms. Electronics, 13(24), 4876.

[3] Caelen, O., & Blete, M. A. (2024). Developing Apps with GPT-4 and ChatGPT: Build Intelligent Chatbots, Content Generators, and More. " O'Reilly Media, Inc.".

[4] Lievens, J. (2024). An exploratory assessment of the usability and potential of generative pretrained transformers (GPTs) as feedback assistants for long-format academic writing tasks. INTED proceedings, 18, 1280-1286.

[5] Yigci, D., Eryilmaz, M., Yetisen, A. K., Tasoglu, S., & Ozcan, A. (2025). Large language model-based chatbots in higher education. Advanced Intelligent Systems, 7(3), 2400429.

[6] Elnaffar, S., Rashidi, F., & Abualkishik, A. Z. (2025). Teaching with AI: A Systematic Review of Chatbots, Generative Tools, and Tutoring Systems in Programming Education. arXiv preprint arXiv:2510.03884.

[7] Sung, M. C., & Kang, S. (2025). GPT API-based chatbots as adaptive and facilitative tutors for L2 English process writing. 영어학, 25, 448-471.

[8] Mzwri, K., & Turcsányi-Szabo, M. (2025). Bridging LMS and Generative AI: Dynamic Course Content Integration (DCCI) for Connecting LLMs to Course Content--The Ask ME Assistant. arXiv preprint arXiv:2504.03966.

[9] Misanchuk, M., & Hyzyk, J. (2024). ChatGPT in STEM teaching: An introduction to using LLM-based tools in higher ed.

[10] Hadi, M. U., Al-Tashi, Q., Qureshi, R., Shah, A., Muneer, A., Irfan, M., ... & Shah12, M. (2024). LLMs: A Comprehensive Survey of Applications, Challenges, Datasets, Models, Limitations, and Future Prospects.

[11] Beale, R. (2025). The Revolution Has Arrived: What the Current State of Large Language Models in Education Implies for the Future. arXiv preprint arXiv:2507.02180.

[12] Wang, S., Xu, T., Li, H., Zhang, C., Liang, J., Tang, J., ... & Wen, Q. (2024). Large language models for education: A survey and outlook. arXiv preprint arXiv:2403.18105.

[13] Guo, K., & Li, D. (2024). Understanding EFL students' use of self-made AI chatbots as personalized writing assistance tools: A mixed methods study. System, 124, 103362.

[14] Krumsvik, R. J. (2025, March). GPT-4's capabilities in handling essay-based exams in Norwegian: an intrinsic case study from the early phase of intervention. In Frontiers in Education (Vol. 10, p. 1444544). Frontiers Media SA.

[15] Gebre Hiwot, T., & Namuduri, S. (2024). The impact of a LargeLanguage Model (LLM): A qualitative study on how students and educatorsperceive the use of LLMs such as ChatGPT withinconventional university education dynamics.

[16] Xu, H., Gan, W., Qi, Z., Wu, J., & Yu, P. S. (2024). Large language models for education: A survey. arXiv preprint arXiv:2405.13001.

[17] Sharma, S., Mittal, P., Kumar, M., & Bhardwaj, V. (2025). The role of large language models in personalized learning: a systematic review of educational impact. Discover Sustainability, 6(1), 1-24.

[18] Morosanu, G. A., Rata, L. A., & Geru, M. (2023). Outcomes of Large Language Models and Artificial Intelligence in Education. Didactica Danubiensis, 3(1), 30-52.

[19] Runceanu, A., Balan, A., Gavanescu, L., Neagu, M. M., Cojocaru, C., Borcosi, I., & Balacescu, A. (2025). Enhancing the Learning Experience with AI. Information, 16(5), 410.

[20] Upadhyay, A., Farahmand, E., Muñoz, I., Akber Khan, M., & Witte, N. (2024). Influence of LLMs on learning and teaching in higher education. Available at SSRN 4716855.

[21] Zdravkova, K., & Ilijoski, B. (2025). The impact of large language models on computer science student writing. International Journal of Educational Technology in Higher Education, 22(1), 32.

[22] Kulaksız, G. C. (2024). Artificial intelligence-based language modelling: The effect of ChatGPT application on writing skills in the context of teaching English as a foreign language.

[23] Naderi, E. V. (2025). Intelligent Tutoring Systems In The Age Of Llm-Based Agentic Frameworks-Adapting Small On-Device Language Models For Fact-Checking And Student Compliance Detection.

[24] Masalaci, Z. S. (2024). Exploring the Impact of Large Language Models on Conceptual Learning in Higher Education: An Analysis of AI-Driven Conversational Tools (ChatGPT) (Master's thesis, NTNU).

[25] Neumann, A. T., Yin, Y., Sowe, S., Decker, S., & Jarke, M. (2024). An llm-driven chatbot in higher education for databases and information systems. IEEE Transactions on Education.

[26] Lang, Q., Wang, M., Yin, M., Liang, S., & Song, W. (2025). Transforming education with generative AI (GAI): Key insights and future prospects. IEEE Transactions on Learning Technologies.

[27] Liu, D., Huang, R., Chen, Y., Adarkwah, M. A., Zhang, X., Li, X., ... & Da, T. (2024). Using Educational Robots to Enhance Learning. Singapore: Springer Nature Singapore. https://link. springer. com/10.1007/978-981-97-5826-5.

[28] Hegazy, H. M. (2024). A Linguistically Developed Prompt Engineering Parameters Model for Enhancing AI's Generation of Customized ESL Reading Texts* Dr. Hebatollah MM Hegazy. Faculty of Education Journal Alexandria University, 34(3), 501-566.

[29] Mohamed Nassar, H. (2025). Comparing the Quality of AI-generated and Instructor Feedback in a University Writing Program.

[30] Amos, J. P., Amodu, O. A., Mahmood, R. A. R., Abdulqudus, A. B., Zakaria, A. F., Iyanda, A. R., ... & Hanapi, Z. M. (2025). A bibliometric exposition and review on leveraging LLMs for Programming Education. IEEE Access.