






PAPER

Integration of Intelligent Assistants and Adaptive Learning in Engineering Pedagogy: A Case Study in Nanotechnology and Life Sciences

Jorge Cruz-Angeles¹ ,
Mariana E. Elizondo-
García¹ , Genaro
Zavala² , Jorge Membrillo-
Hernández^{1,2}  

¹School of Engineering and Sciences, Tecnológico de Monterrey, Monterrey, Mexico

²Institute for the Future of Education, Tecnológico de Monterrey, Monterrey, Mexico

jmembrillo@tec.mx

ABSTRACT

The proliferation of large language models (LLMs) and generative artificial intelligence has catalyzed an unprecedented pedagogical paradigm shift within higher education. This study investigates the adoption, efficacy, and cognitive impact of these technological tools among undergraduate cohorts in engineering and life sciences. Recognizing that baseline familiarity with AI does not inherently translate into advanced operational competency or prompt engineering literacy, this study evaluates the deployment of both standard LLMs and retrieval-augmented generation (RAG)-based customized assistants. The investigation employs a rigorous dual-phase methodology: an exploratory assessment of technology acceptance using standard ChatGPT, followed by a tightly controlled quasi-experiment evaluating the impact of domain-specific “GPT Custom” mentors on academic performance on complex engineering tasks. The empirical results demonstrate that customized AI assistants significantly improve final academic outcomes, yielding average grade increases of more than 15% on data-intensive analytical assignments. Furthermore, the deployment of customized assistants notably reduced grade variability among students, indicating a homogenization of academic performance that effectively levels the learning environment without compromising rigor. This improvement was statistically amplified when students utilized premium, high-capacity versions of the models for extensive synthesis tasks. Ultimately, the data indicate that while AI offers robust adaptive scaffolding, its efficacy depends on users’ critical-thinking capacities, underscoring the urgent need for educational frameworks that cultivate prompt-engineering literacy and responsible human-AI collaboration.

KEYWORDS

educational innovation, large language models (LLM), generative AI, STEM education, adaptive learning, retrieval-augmented generation (RAG), intelligent tutoring systems (ITS)

Cruz-Angeles, J., Elizondo-García, M. E., Zavala, G., Membrillo-Hernández, J. (2026). Integration of Intelligent Assistants and Adaptive Learning in Engineering Pedagogy: A Case Study in Nanotechnology and Life Sciences. *International Journal of Engineering Pedagogy (iJEP)*, 16(4), pp. 90–103. <https://doi.org/10.3991/ijep.v16i4.61461>

Article submitted 2026-03-11. Revision uploaded 2026-04-17. Final acceptance 2026-04-17.

© 2026 by the authors of this article. Published under CC-BY.

1 INTRODUCTION

The integration of artificial intelligence in education (AIED) represents a critical and highly contested juncture in the evolution of pedagogical methodologies [1]. The field has progressed remarkably since its theoretical and empirical foundations were established in the 1970s, which initially focused on the development of rule-based adaptive learning systems and intelligent tutoring approaches [1]. Early computational models laid the groundwork for tailoring instruction to the specific cognitive requirements of individual learners, establishing a trajectory that has culminated in the contemporary deployment of generative artificial intelligence [2]. Over the subsequent decades, extensive research spanning multiple subject domains demonstrated the feasibility of adaptive instruction, illustrating the immense potential of technology to diagnose student misconceptions, trace knowledge acquisition, and dynamically adjust pedagogical delivery [3], [4], [5].

Recent paradigm-shifting advances in natural language processing (NLP) have facilitated the emergence of large language models (LLMs), such as OpenAI's ChatGPT, which have rapidly and ubiquitously permeated higher education environments [6], [7], [8], [9]. These advanced neural architectures possess the unprecedented capacity to generate coherent text, synthesize complex multi-disciplinary information, and provide adaptive, conversational feedback in real time [7]. However, the rapid adoption of these tools by global student populations has vastly outpaced the development of established pedagogical frameworks governing their use. Consequently, the actual empirical impact of LLMs on academic performance, deep cognitive engagement, and long-term knowledge retention remains a subject of intense academic scrutiny and debate. Integrating ChatGPT and similar generative technologies into education possesses the potential to fundamentally revolutionize traditional teaching approaches by providing students with highly personalized, interactive, and infinitely patient learning experiences [10], [11].

Several prominent authors have recently examined the myriad implications of this technology for teaching, academic research, epistemological validation, the digital transformation of educational institutions, and the complex ethics of machine-assisted learning. Multiple studies indicate that incorporating GPT-based tools into STEM education could offer novel personalized feedback strategies that help students deconstruct and understand complex theoretical concepts by offering real-time explanations and highly illustrative examples [12], [13], [14]. Despite the saturation of theoretical literature, there remains a notable dearth of empirical research quantifying the direct, measurable effects of customized LLMs on specific academic performance metrics [15], [16].

This study addresses this critical gap by investigating the localized deployment of customized, domain-specific AI assistants within undergraduate STEM courses. By transitioning from generic, unconstrained LLM use to implementing customized generative pre-trained transformers that utilize retrieval-augmented generation (RAG), this research evaluates the efficacy of AI as a structured, adaptive learning mentor. The primary objectives of this comprehensive study are manifold: to assess the baseline technological competency of students regarding unconstrained AI tools; to quantify the impact of customized, RAG-enabled AI assistants on academic performance across varying task complexities; to evaluate the specific cognitive and pedagogical challenges associated with human-AI educational interaction; and to determine the differential impact of standard versus premium computational access on student outcomes.

1.1 Theoretical framework and literature review

Historical foundations of adaptive learning in STEM. The historical trajectory of AEID is deeply rooted in the interdisciplinary intersection of computer science, cognitive psychology, and educational theory [17]. The genesis of the field was driven by researchers' ambition to use the educational domain as a controlled "microworld" to model human learning, retention, memory, and cognitive transformation. Foundational adaptive learning systems, such as the SCHOLAR system developed by Carbonell in 1970, and subsequent intelligent tutoring systems (ITS) explored by Auguste in 1985, were designed with the explicit goal of emulating the localized, highly individualized feedback mechanisms inherent in expert human tutoring [1], [2].

Central to these early systems was the foundational concept of student modeling, which aimed to mathematically map a learner's problem-solving strategies and identify specific cognitive deficits or procedural "bugs." Pioneering systems such as DEBUGGY used complex model-tracing techniques to detect fine-grained errors in students' mathematical problem-solving processes [18]. Over time, systems operating on established cognitive theories, such as John R. Anderson's adaptive control of thought-rational (ACT-R) theory, have conceptualized learning as the vital, measurable transition from declarative knowledge to procedural fluency [14]. These systems primarily focused on STEM fields, addressing the persistent dual challenges of accurately diagnosing students' errors in real time and devising adaptive pedagogical interventions that effectively guide students toward mastery. The theoretical underpinnings established by these early rule-based models provide the indispensable architectural context for understanding contemporary personalized learning paradigms.

1.2 The evolution from intelligent tutoring systems to large language models

As the AIED field matured throughout the late 20th and early 21st centuries, the inherent limitations of early ITS became increasingly apparent. Traditional tutoring systems required exhaustive, painstakingly detailed domain-specific programming and struggled to adapt seamlessly to the nuanced, unpredictable, and highly variable nature of human linguistic interaction. The paradigm shift toward modern adaptive systems has been entirely driven by the integration of sophisticated machine learning algorithms and advanced natural language processing [19], [20].

Unlike first-generation systems constrained by rigid, pre-programmed decision trees and limited vocabularies, modern AI tools leverage NLP to comprehend and generate natural language with human-like fluency, thereby facilitating highly intuitive and frictionless user interfaces [21], [22], [23]. This evolution allows educational systems to move far beyond simple binary error detection. Contemporary AI can evaluate complex written responses, identify sophisticated patterns in student reasoning, and provide explanatory feedback that dynamically adapts to affective and motivational fluctuations. Furthermore, the push toward Explainable AI (XAI) enables educators to understand the mechanisms underlying AI decision-making, thereby ensuring pedagogical transparency [24], [25]. LLM represents the current apotheosis of this evolution, offering the flexibility to generate interactive, contextually relevant educational support across an infinite range of domains without the need for manual, rule-by-rule programming [26].

1.3 Retrieval-augmented generation vs. model fine-tuning

While generic LLMs offer vast, generalizable capabilities, their use in rigorous academic environments is frequently compromised by “hallucinations,” the generation of highly coherent, syntactically perfect yet factually inaccurate information. In STEM education, where mathematical precision, empirical accuracy, and factual integrity are paramount, mitigating the risk of hallucination is not merely an optimization but an absolute pedagogical necessity.

To address this critical vulnerability, advanced methodologies such as model fine-tuning and RAG have been developed. Fine-tuning involves permanently adjusting the fundamental neural weights of a base model using massive, domain-specific datasets. A prime example is the development of BioGPT by Luo et al. at Microsoft Research, which was pre-trained on 15 million PubMed abstracts to achieve superior performance in biomedical natural language processing tasks [27]. However, fine-tuning is an immensely complex process requiring substantial computational infrastructure, specialized hardware (such as high-performance GPUs), and deep expertise in neural network architecture, making it highly impractical for standard educational institutions facing budgetary limitations [10], [27].

Conversely, RAG provides a highly efficient, lightweight, and accessible alternative [28], [29]. By dynamically grounding the LLM’s responses in a curated, verified external database at query time, RAG avoids altering the model’s underlying weights. This methodology couples the generative fluency of an LLM with a highly specific retrieval mechanism, ensuring that the AI assistant prioritizes verified course materials, lecture transcripts, and peer-reviewed literature over the generalized, often unreliable internet data on which it was initially trained. The introduction of customizable GPT models (GPT customs) by OpenAI has fundamentally democratized access to RAG technology, allowing standard educators to build highly specialized, domain-accurate virtual assistants capable of delivering precise, adaptive learning experiences without requiring advanced programming skills [30], [31].

2 MATERIALS AND METHODS

The present research employs a comprehensive, dual-phase methodological design to evaluate the integration of generative AI within undergraduate curricula. The investigation combines an exploratory observational study of baseline AI usage with a highly controlled, quasi-experimental evaluation of customized, RAG-enabled AI assistants.

2.1 Phase I: Exploratory integration of standard LLMs

The initial phase of the study was conducted during the Spring 2023 academic semester across a multi-campus higher education institution in Mexico (Tecnologico de Monterrey). The primary objective was to establish a rigorous baseline understanding of students’ familiarity with and acceptance of standard, unconstrained generative AI tools.

Participants and context: The cohort comprised 94 undergraduate students enrolled in an introductory life sciences course titled “Fundamentals of Biological Systems.” The course, delivered online across six distinct institutional campuses, serves as a general education requirement designed to instill foundational methodologies in biological sciences, foster broad cognitive perspectives, and improve critical thinking skills.

Intervention: Students used the free, standard version of ChatGPT (version 3.5) to complete a guided, multi-stage research activity focused on the complex impact of dietary choices on the human gut microbiota. Students were provided with an initial exemplar prompt regarding the nutritional breakdown of common foods. Crucially, the pedagogical design required students to generate preliminary insights using the AI, and subsequently validate those AI-generated claims by cross-referencing them against established, peer-reviewed academic literature and verifiable university databases.

Data collection: Upon completion of the academic module, cognitive engagement, critical reflection, and technological acceptance were systematically evaluated using a 13-item survey, administered via the Qualtrics online platform. The instrument utilized a 10-point Likert scale, ranging from 1 (Strongly Disagree) to 10 (Strongly Agree). Qualitative data were concurrently gathered via open-ended questions about the specific user experience and subsequently analyzed using Python-based natural language processing libraries (Pandas and NLTK). The text was carefully tokenized, and common stop words were filtered to generate descriptive thematic word clouds, isolating the nouns and verbs most indicative of the student experience.

2.2 Phase II: Controlled evaluation of GPT custom assistants

Following the baseline assessment, a more complex, highly controlled quasi-experiment was designed to evaluate the specific, measurable impact of domain-trained AI mentors on academic performance. This phase intentionally transitioned from generic LLM usage to the targeted deployment of adaptive, RAG-enabled customized assistants.

Participants: The sample consisted of 70 fourth-semester undergraduate students enrolled in an in-person Nanotechnology Engineering program. The cohort was divided non-probabilistically into four distinct groups based on participant availability, voluntariness, and the specific tier of the AI tools utilized.

Table 1 delineates the experimental distribution, distinguishing between control groups (utilizing traditional pedagogical methods without AI assistance) and experimental groups (utilizing customized GPT assistants, with a subset granted premium computational access).

Table 1. Distribution of students and access to GPT custom tools

Group	Gender (%)	# Students	Access to GPT Custom	Licensed GPT+
1	59 Male/41 Female	17	No	No
2	58 Male/42 Female	13	No	No
3	60 Male/40 Female	20	Yes	No
4	65 Male/35 Female	20	Yes	Yes

2.3 Design and implementation of retrieval-augmented generation models

The core empirical intervention relied on the construction of highly specialized, RAG-enabled customized GPT assistants. The rigorous development process required extensive prompt engineering and careful knowledge integration to ensure academic precision, maintain a highly academic tone, and eliminate hallucinations.

System prompt engineering: Detailed natural-language instructions were encoded in the GPT Builder interface. These instructions defined the assistant's specific pedagogical persona, the required analytical depth, and the rigid behavioral boundaries. The models were explicitly instructed to function as Socratic mentors, guiding students toward solutions through leading questions rather than simply providing immediate, finalized answers.

Retrieval source configuration: To guarantee absolute factual accuracy, the models were tethered to secure, proprietary databases containing verified course materials. This included exact transcripts of lecture videos, faculty-designed slide decks, rigorous laboratory protocols, and peer-reviewed reference texts.

Model validation: Prior to student deployment, the customized models underwent rigorous usability and accuracy testing by domain faculty to ensure that the generated responses aligned precisely with the curated curriculum and met the required pedagogical standards.

Over 15 distinct customized assistants were developed to support various facets of the academic program. Table 2 provides a representative overview of the highly specialized models deployed during the experimental phase.

2.4 Data collection and statistical processing

Academic performance data, including grades for individual laboratory reports, complex figure generation, and comprehensive final examinations, were extracted directly from the institutional learning management system (LMS, CANVAS).

Due to the inherent non-normal distribution frequently observed in highly segmented academic grading datasets, statistical comparisons between the control and experimental cohorts were executed using the non-parametric Mann-Whitney U test. This robust statistical approach was used to determine whether there were significant differences in academic performance between groups that used GPT Custom and those that did not, without assuming a standard Gaussian distribution of the data. The entire statistical analysis was conducted in R (version 4.3.2) using the 'stats' package, with a standard alpha level of $p < 0.05$ for statistical significance.

3 RESULTS

3.1 Phase I Outcomes: Technology acceptance and cognitive engagement

The exploratory integration of standard ChatGPT yielded profound insights into baseline student competencies. The pedagogical requirement to cross-reference AI-generated outputs with traditional academic literature revealed pervasive inaccuracies in the foundational LLM's dataset, particularly in highly nuanced areas of nutritional biochemistry. This deliberate exposure to AI fallibility served as a critical catalyst for the development of student skepticism and analytical rigor, proving that AI outputs cannot be accepted unquestioningly in scientific domains.

Table 2. Description and application spectrum of customized retrieval-augmented generation assistants

GPT Model Designation	Primary Description and Function	Verified RAG Data Sources	Extrapolated Pedagogical Applications
Class GPT (Transcript Based)	A general assistant is designed to resolve conceptual doubts regarding explicit class content.	Video recordings, lecture notes, and full session transcripts.	Universal applicability to any course utilizing recorded lectures.
Exam Problem Solver	An analytical assistant guiding students through the logic of complex mathematical/engineering problems.	Solved problem sets, historical exam banks, step-by-step rubrics.	Highly applicable to all quantitative STEM disciplines.
Study Guide & Exam Prep	An interactive study partner capable of quizzing students to assess knowledge retention.	Comprehensive course materials, syllabus, and core competency lists.	Universal applicability for summative assessment preparation.
Final Interview Simulator	Socratic simulator for practicing high-stakes oral examinations and technical interviews.	Historical database of oral exam questions and exemplary answer rubrics.	Applicable to capstone defenses and professional certification prep.
Laboratory Report Architect	Assistant guiding the synthesis of empirical data into formal academic reporting structures.	Technical equipment manuals, procedural documents, and evaluation rubrics.	Essential for all practical laboratory and empirical research courses.
Figure & Data Visualizer	Guidance tool for the correct formatting, styling, and descriptive titling of scientific graphs.	Institutional style guides, peer-reviewed examples of data visualization.	Applicable to all disciplines requiring complex data representation.

Table 3. Questions present in the final surveys

Evaluated Feature	Items	Item ID
Technology acceptance and usage	During the development of the activity involving ChatGPT:	
	1. I understood the professor’s instructions clearly.	Q1
	2. ChatGPT allowed me to complete the activity more quickly.	Q2
	3. ChatGPT was useful.	Q3
	4. The interaction with ChatGPT was clear and understandable.	Q4
	5. Learning to use ChatGPT was easy for me.	Q5
Support in the process of analysis and reflection	The activity involving ChatGPT:	
	6. ChatGPT was user-friendly for carrying out the activity.	Q6
	7. Facilitated the analysis of information.	Q7
	8. Helped me reflect on the acquired learning.	Q8
	9. Helped create a self-interpretation of the concepts.	Q9
	10. Kept me focused.	Q10
	11. Challenged me to use my judgment.	Q11
	12. Challenged me to look for information from other sources.	Q12
13. Made me question the credibility of the information given by ChatGPT.	Q13	

The post-intervention survey demonstrated overwhelmingly positive responses regarding immediate acceptance of technology. For questions regarding basic usability (Q1–Q6), the vast majority of participants selected “Strongly Agree,”

indicating that the interface was highly intuitive and significantly accelerated initial data gathering.

However, the data on higher-order cognitive engagement presented a notable, concerning dichotomy. Table 3 outlines the specific survey items targeting analysis and reflection. While students universally agreed that the tool was useful, items Q9 (“Helped create a self-interpretation of the concepts”) and Q10 (“Kept me focused”) showed a much broader range of opinions, including the emergence of “Strongly Disagree” responses. This suggests that a statistically significant subset of students struggled to use AI to foster deep, independent conceptual interpretation, feeling that the tool did not adequately challenge their independent thought processes.

Crucially, qualitative analysis of the activity submissions highlighted a profound deficit in prompt engineering literacy. Despite possessing the pedagogical freedom to query the AI expansively and dynamically, the overwhelming majority of students restricted their interactions solely to the rudimentary example prompt provided by the faculty. The thematic word cloud generated from open-ended feedback corroborated this finding; terms such as “Simple,” “Easy,” “Rapid,” and “Tool” dominated the discourse, while words indicating complex analysis or debate were entirely absent. This indicates that students primarily viewed the LLM as an accelerated search engine rather than a dynamic, conversational analytical partner.

3.2 Phase II outcomes: Academic performance across experimental cohorts

The highly controlled deployment of RAG-enabled GPT custom assistants yielded massive, statistically significant gains in academic performance. A direct non-parametric comparison between the cohorts utilizing the customized assistants (Groups 3 and 4) and the control cohorts relying on traditional methods (Groups 1 and 2) revealed distinct statistical advantages for the AI-assisted students across a variety of highly complex engineering tasks.

Table 4 presents the exact results of the Mann-Whitney U tests, highlighting the specific academic activities in which the use of GPT Custom yielded statistically significant improvements.

Table 4. Statistical comparison of academic performance (control vs. GPT custom cohorts)

Academic Activity	Mann-Whitney U	p-Value	Statistical Significance
X-Ray Diffraction (XRD) Report	102.5	1.26E-09	Highly significant
XRD Figure Generation	209	3.53E-06	Highly significant
Scientific Introduction Writing	210	1.19E-05	Highly significant
FTIR Figure Generation	236	1.47E-05	Highly significant
Comprehensive Final Report	265	1.28E-04	Highly significant
Calibration Curve Figure	277	6.60E-04	Highly significant
Combined DSC and FTIR Report	367	9.10E-03	Significant
DSC Figure Generation	420.5	1.92E-02	Significant
UV-Vis Spectroscopy Report	472.5	2.69E-01	Not significant
Final Oral Interview	577	9.74E-01	Not significant

Notes: Significance labels are descriptive: “highly significant” denotes $p < 0.001$ and “significant” denotes $0.001 \leq p < 0.05$. All inferential conclusions use $\alpha = 0.05$.

The empirical data indicate that the deployment of customized AI mentors improved outcomes in data-intensive reporting and visual characterization, as well as in theoretical synthesis. For example, the very low p -value for the XRD report (1.26E-9) reflects a clear improvement in students' ability to synthesize complex crystallographic data when supported by an AI mentor. Conversely, activities that relied heavily on spontaneous verbal articulation (final oral interview) showed no statistically significant variance, suggesting that the cognitive benefits of the AI mentor were concentrated primarily in the preparatory and written analytical phases of learning, rather than in real-time verbal recall.

Furthermore, deep variance analysis via box-plot mapping revealed a dramatic reduction in grade disparity across the experimental cohorts. The interquartile range of grades for students utilizing GPT Custom was significantly narrower than that of the control group. The customized AI assistants effectively raised the foundational performance level of the entire cohort, yielding far more consistent academic outcomes regardless of a student's prior individual capability.

3.3 The impact of premium AI capabilities on complex tasks

To ascertain whether the underlying computational capacity of the LLM impacted academic outcomes, an isolated analysis was conducted comparing Group 3 (standard GPT Custom) and Group 4 (GPT Custom powered by premium ChatGPT+ licenses). The premium licenses provided enhanced contextual memory windows, sophisticated data analysis plugins, and significantly higher query rate thresholds. Table 5 details the statistical comparison between the free and premium GPT tiers.

Table 5. Statistical comparison of performance by AI computational tier (Free vs. Premium)

Academic Activity	W-Statistic	p-Value	Statistical Significance
X-Ray Diffraction (XRD) Report	100	0.00038	Highly Significant
Combined DSC and FTIR Report	77.5	0.00074	Highly Significant
Comprehensive Final Report	90	0.00287	Highly Significant
Scientific Introduction Writing	112.5	0.01640	Significant
Calibration Curve Figure	144.5	0.06635	Marginal/Not Significant
DSC Figure Generation	153	0.08434	Not Significant
UV-Vis Spectroscopy Report	225	0.49382	Not Significant
XRD Figure Generation	185	0.86944	Not Significant
Final Oral Interview	195	0.89724	Not Significant
FTIR Figure Generation	181	0.98543	Not Significant

Notes: Significance labels are descriptive: "highly significant" denotes $p < 0.001$ and "significant" denotes $0.001 \leq p < 0.05$. All inferential conclusions use $\alpha = 0.05$.

The Mann-Whitney U test results indicate that the premium capabilities provided a statistically significant advantage exclusively within extensive, long-form synthesis tasks. Specifically, the comprehensive experimental reports ("XRD Report," "DSC and FTIR Report," and "Final Report") saw notable grade elevations. However, for discrete visual tasks (e.g., generating specific figures) or brief conceptual interactions, the premium license offered zero statistical advantage over the standard

customized model. This suggests that the premium model's expanded context window is highly beneficial for maintaining narrative coherence across long documents but mathematically unnecessary for discrete, short-form tasks.

4 DISCUSSION

4.1 Interpretation of the findings and homogenization of performance

The empirical data confirm that customized, RAG-enabled LLMs are highly effective mechanisms for adaptive learning in rigorous STEM environments. The most profound and pedagogically transformative finding of this study is the observed homogenization of academic performance. The dramatic reduction in grade variability (standard deviation) among cohorts utilizing GPT Custom indicates that these systems function as highly personalized, infinitely patient cognitive scaffolds. By providing real-time, context-accurate interventions tailored to the individual's exact pacing needs, the AI mentors effectively elevate the performance baseline, acting as a great equalizer for students with varying levels of foundational preparation.

This phenomenon directly addresses one of the most persistent historical challenges within educational theory: Benjamin Bloom's "2 Sigma Problem." Bloom posited that students receiving personalized, one-to-one tutoring perform two standard deviations better than students in traditional classroom environments. Historically, providing personalized tutoring at scale has been economically and logistically impossible. The current data strongly suggests that LLMs, when appropriately constrained and focused via RAG, successfully fulfill this historical mandate, offering scalable, individualized mastery learning that genuinely raises the academic floor without lowering the ceiling for advanced students.

4.2 Cognitive engagement and the necessity of prompt engineering literacy

While the academic gains in the experimental phase are irrefutable, the findings from Phase I expose a critical, systemic vulnerability in current engineering pedagogy: a profound lack of prompt engineering literacy. The observation that undergraduate students predominantly treated a highly dynamic LLM as a static search query engine, refusing to deviate from a single provided example prompt, reveals a significant metacognitive deficit.

The successful integration of AI into complex engineering curricula must go far beyond mere access to software. Effective interaction with an LLM requires foundational domain knowledge; if a student does not understand the underlying principles of a discipline, they lack the cognitive architecture required to formulate complex, multi-layered queries. The AI cannot scaffold reasoning if the student cannot initiate the intellectual dialogue. Therefore, a critical implication of this study is the urgent necessity for academic institutions to formally integrate "AI interaction mechanics" and advanced prompt engineering as foundational core competencies, taught alongside traditional research methodologies and statistical analysis.

4.3 Scalability, cost-benefit allocation, and democratization

The specific statistical findings regarding premium versus standard AI tiers present a highly actionable paradigm for institutional resource allocation. The data

demonstrate that premium LLM capabilities benefit only extensive, long-form synthesis tasks. Consequently, educational entities must strategically deploy resources; providing expensive premium AI licenses may be highly cost-effective and pedagogically justified for capstone design courses, graduate thesis writing, or advanced research seminars requiring massive data synthesis. However, for introductory courses focused on discrete conceptual acquisition or specific task execution, standard, free-tier customized models are entirely sufficient.

Furthermore, the utilization of RAG models democratizes access to advanced AI in education. Because RAG relies on external database retrieval rather than computationally expensive model fine-tuning (which requires massive GPU clusters), institutions with limited budgets can easily build highly sophisticated, error-free AI tutors. This ensures that the benefits of personalized AI tutoring are not restricted exclusively to elite, highly funded research institutions.

5 CONCLUSION

The integration of RAG-customized LLM represents a transformative evolutionary leap in engineering pedagogy. This study conclusively demonstrates that domain-specific AI assistants significantly enhance academic performance, foster deeply adaptive learning environments, and substantially reduce outcome disparities across heterogeneous student cohorts. The technology effectively democratizes personalized tutoring, providing scalable solutions to historical educational challenges. However, realizing the full potential of these technological tools requires an immediate paradigm shift in educational strategy. Institutions must move beyond merely providing AI software and actively cultivate students' prompt-engineering literacy and critical analytical skills to effectively orchestrate these advanced systems. Ultimately, the future of highly effective engineering education lies not in replacing human educators but in seamless, ethical, and highly skilled collaboration between human cognition and artificial intelligence.

6 ACKNOWLEDGEMENTS

The authors acknowledge The Writing Lab, the Institute for the Future of Education, and Tecnológico de Monterrey for their support in producing this manuscript.

7 REFERENCES

- [1] J. Carbonell, "AI in CAI: An artificial-intelligence approach to computer-assisted instruction," *IEEE Trans. Man Mach. Syst.*, vol. 11, no. 4, pp. 190–202, 1970. <https://doi.org/10.1109/TMMS.1970.299942>
- [2] D. Auguste, "Intelligent tutoring systems," *Artificial Intelligence*, vol. 26, no. 2, pp. 233–238, 1985. [https://doi.org/10.1016/0004-3702\(85\)90033-5](https://doi.org/10.1016/0004-3702(85)90033-5)
- [3] V. J. Shute and J. Psocka, "Intelligent tutoring systems: Past, present, and future," in *Handbook of Research on Educational Communications and Technology*, D. Jonassen, Ed., New York: Macmillan, 1996, pp. 570–600.
- [4] B. P. Woolf, J. E. Beck, C. Eliot, and M. K. Stern, "Growth and maturity of intelligent tutoring systems: A status report," *Semantic Scholar*, 2001. <https://api.semanticscholar.org/CorpusID:53397321>

- [5] K. VanLEHN, “The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems,” *Educational Psychologist*, vol. 46, no. 4, pp. 197–221, 2011. <https://doi.org/10.1080/00461520.2011.611369>
- [6] J. Crawford, M. Cowling, and K.-A. Allen, “Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI),” *JUTLP*, vol. 20, no. 3, 2023. <https://doi.org/10.53761/1.20.3.02>
- [7] I. García-Martínez, J. M. Fernández-Batanero, J. Fernández-Cerero, and S. P. León, “Analysing the impact of artificial intelligence and computational sciences on student performance: Systematic review and meta-analysis,” *J. New Approaches Educ. Res.*, vol. 12, no. 1, pp. 171–197, 2023. <https://doi.org/10.7821/naer.2023.1.1240>
- [8] F. J. García-Peñalvo, “La percepción de la Inteligencia Artificial en contextos educativos tras el lanzamiento de ChatGPT: disrupción o pánico,” *Educ. Knowl. Soc.*, vol. 24, p. e31279, 2023. <https://doi.org/10.14201/eks.31279>
- [9] C. Stokel-Walker, “AI bot ChatGPT writes smart essays—should professors worry?” *Nature*, 2022. <https://doi.org/10.1038/d41586-022-04397-7>
- [10] G. Yenduri *et al.*, “GPT (Generative Pre-Trained Transformer)—A comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions,” *IEEE Access*, vol. 12, pp. 54608–54649, 2024. <https://doi.org/10.1109/ACCESS.2024.3389497>
- [11] N. Anderson *et al.*, “AI did not write this manuscript, or did it? Can we trick the AI text detector into generated texts? The potential future of ChatGPT and AI in sports & exercise medicine manuscript generation,” *BMJ Open Sport Exerc Med.*, vol. 9, no. 1, p. e001568, 2023. <https://doi.org/10.1136/bmjsem-2023-001568>
- [12] A. Tlili *et al.*, “What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education,” *Smart Learn. Environ.*, vol. 10, no. 1, 2023. <https://doi.org/10.1186/s40561-023-00237-x>
- [13] R. Winkels, “A new framework for describing and designing intelligent tutoring systems,” in *Artificial Intelligence in Higher Education*, in Lecture Notes in Computer Science, V. Mařík, O. Štěpánková, and Z. Zdráhal, Eds., vol. 451, Springer Berlin Heidelberg, 1990, pp. 230–243. https://doi.org/10.1007/3-540-52952-7_29
- [14] J. R. Anderson, A. Corbett, J. Fincham, D. Hoffman, and R. Pelletier, “General principles for an intelligent tutoring architecture,” in *Cognitive Approaches to Automated Instruction*, V. J. Shute and J. W. Regian, Eds., Hillsdale, NJ, USA: Lawrence Erlbaum Associates, 1992, pp. 81–106.
- [15] M. Ali, I. B. A. Wahab, H. Z. Huri, and M. S. Yusoff, “Personalised learning in higher education for health sciences: A scoping review protocol,” *Syst. Rev.*, vol. 13, no. 1, p. 99, 2024. <https://doi.org/10.1186/s13643-024-02478-4>
- [16] K. R. Koedinger, J. R. Anderson, W. H. Hadley, and M. A. Mark, “Intelligent tutoring goes to school in the big city,” *International Journal of Artificial Intelligence in Education*, vol. 8, pp. 30–43, 1997. <https://doi.org/10.1184/R1/6470153.V1>
- [17] K. Hilali, M. Chergui, and A. Ammoumou, “Adaptive learning systems: A comprehensive overview and identification of challenges,” in *2023 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD)*, Rabat, Morocco: IEEE, 2023, pp. 1–7. <https://doi.org/10.1109/ICTMOD59086.2023.10438128>
- [18] J. S. Brown and R. R. Burton, “Diagnostic models for procedural bugs in basic mathematical skills,” *Cognitive Science*, vol. 2, no. 2, pp. 155–192, 1978. https://doi.org/10.1207/s15516709cog0202_4
- [19] R. Azevedo *et al.*, “Lessons learned and future directions of MetaTutor: Leveraging multichannel data to scaffold self-regulated learning with an intelligent tutoring system,” *Front. Psychol.*, vol. 13, p. 813632, 2022. <https://doi.org/10.3389/fpsyg.2022.813632>

- [20] V. Alevén *et al.*, “Example-tracing tutors: Intelligent tutor development for non-programmers,” *Int. J. Artif. Intell. Educ.*, vol. 26, no. 1, pp. 224–269, 2016. <https://doi.org/10.1007/s40593-015-0088-2>
- [21] Y. Walter, “Embracing the future of Artificial Intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education,” *Int. J. Educ. Technol. High. Educ.*, vol. 21, no. 1, p. 15, 2024. <https://doi.org/10.1186/s41239-024-00448-3>
- [22] A. Extance, “ChatGPT has entered the classroom: How LLMs could transform education,” *Nature*, vol. 623, no. 7987, pp. 474–477, 2023. <https://doi.org/10.1038/d41586-023-03507-3>
- [23] A. Y. Q. Huang, O. H. T. Lu, and S. J. H. Yang, “Effects of Artificial Intelligence–Enabled personalized recommendations on learners’ learning engagement, motivation, and outcomes in a flipped classroom,” *Computers & Education*, vol. 194, p. 104684, 2023. <https://doi.org/10.1016/j.compedu.2022.104684>
- [24] T. Susnjak, “ChatGPT: The end of online exam integrity?” *arXiv preprint arXiv no:2212.09292*, 2022. <https://doi.org/10.48550/ARXIV.2212.09292>
- [25] M. E. Elizondo-García, H. Hernández-De la Cerda, I. G. Benavides-García, P. Caratozzolo, and J. Membrillo-Hernández, “Who is solving the challenge? The use of ChatGPT in mathematics and biology courses using challenge-based learning,” *Frontiers in Education*, vol. 10, 2025. <https://doi.org/10.3389/educ.2025.1417642>
- [26] N. V. Covington and O. Vruwink, “ChatGPT in undergraduate education: Performance of GPT-3.5 and identification of AI-generated text in introductory neuroscience,” *Int. J. Artif. Intell. Educ.*, vol. 35, no. 2, pp. 627–650, 2024. <https://doi.org/10.1007/s40593-024-00427-9>
- [27] R. Luo *et al.*, “BioGPT: Generative pre-trained transformer for biomedical text generation and mining,” *arXiv preprint arXiv no:2210.10341*, 2022. <https://doi.org/10.48550/ARXIV.2210.10341>
- [28] M. Sevgi, F. Antaki, and P. A. Keane, “Medical education with large language models in ophthalmology: Custom instructions and enhanced retrieval capabilities,” *Br J Ophthalmol.*, p. bjo-2023-325046, 2024. <https://doi.org/10.1136/bjo-2023-325046>
- [29] Y. S. Kiyak and A. A. Kononowicz, “Case-based MCQ generator: A custom ChatGPT based on published prompts in the literature for automatic item generation,” *Medical Teacher*, vol. 46, no. 8, pp. 1018–1020, 2024. <https://doi.org/10.1080/0142159X.2024.2314723>
- [30] C. P. Chu, “ChatGPT in veterinary medicine: A practical guidance of generative artificial intelligence in clinics, education, and research,” *Front. Vet. Sci.*, vol. 11, p. 1395934, 2024. <https://doi.org/10.3389/fvets.2024.1395934>
- [31] M. Almasre, “Development and evaluation of a custom GPT for the assessment of students’ designs in a typography course,” *Education Sciences*, vol. 14, no. 2, p. 148, 2024. <https://doi.org/10.3390/educsci14020148>

8 AUTHORS

Jorge Cruz-Angeles is a Full-time Professor at the School of Engineering and Sciences, Tecnológico de Monterrey, Monterrey Campus. Tecnológico de Monterrey, Monterrey Campus, Monterrey, Mexico. His research focuses on the application of Artificial Intelligence in higher education, with particular emphasis on generative AI to support teaching and learning processes in science and engineering (E-mail: jecruzangeles@tec.mx).

Mariana E. Elizondo-García is a Lecturer of the School of Engineering and Sciences at the Vice Presidency of Educational Innovation and Academic Regulations at the Monterrey Campus of the Tecnológico de Monterrey, Monterrey, Mexico.

Her research focuses on artificial intelligence, new educational models, digital education, and impact measurement (E-mail: mariana.elizondo@tec.mx).

Genaro Zavala is the former Associate Director of the Institute for the Future of Education Research Lab, Tecnológico de Monterrey, Monterrey, Nuevo León, Mexico. His research focuses on several areas, including conceptual understanding in physics and engineering, active learning methodologies, interdisciplinary STEM education, advanced technological tools in educational practices, and artificial intelligence applications in education (E-mail: genaro.zavala@gmail.com).

Jorge Membrillo-Hernández, Ph.D., is a Research Professor at the Institute for the Future of Education and the School of Engineering and Sciences at the Tecnológico de Monterrey, Monterrey, Mexico. His research interests include Challenge-based Learning, Sustainable Education, Artificial Intelligence, and STEM education (E-mail: jmembrillo@tec.mx).