

Article

Perceived Learning vs. Engagement in AI-Assisted Homework: A Comparative Study of ChatGPT Use Across High School, University, and Teachers in Sonora, Mexico (2024–2025)

Raquel Torres-Peralta ^{1,*}, Federico Cirett-Galán ¹, María del Carmen Heras-Sanchez ², Karla Lerma-Molina ¹
and Ilse Espinoza-Flores ¹

¹ Facultad Interdisciplinaria de Ingeniería, Universidad de Sonora, Hermosillo 83000, Mexico; federico.cirett@unison.mx (F.C.-G.); karlaarlethlerma@gmail.com (K.L.-M.); ilseesflo27@gmail.com (I.E.-F.)

² Facultad Interdisciplinaria de Ciencias Exactas y Naturales, Universidad de Sonora, Hermosillo 83000, Mexico; carmen.heras@unison.mx

* Correspondence: raquel.torres@unison.mx

Abstract

This study examines how generative AI is adopted and experienced across educational levels in Sonora, Mexico, and whether students' perceived learning aligns with engagement behaviors during AI-assisted homework. We analyze survey data from 2024–2025 covering 1477 participants (high school and university students and teachers) from public and private institutions, including adoption, perceived learning and time savings, help-seeking preferences (teachers vs. ChatGPT vs. Google), and ethical concerns. To move beyond self-reports alone, we introduce a Learning Engagement Index (LEI; 0–1) based on three student behaviors when using ChatGPT to complete academic tasks: reading AI responses, modifying outputs, and integrating personal ideas. Adoption was widespread but consistently higher in university than in high school for both students and teachers. University students reported slightly higher perceived learning and greater time savings. LEI scores were generally high and higher among university students, indicating more frequent engagement behaviors such as reading and adapting AI outputs rather than copying them. However, perceived learning showed only weak alignment with LEI, suggesting that students' self-assessments do not consistently track the engagement actions measured by the index. A complementary GitHub Copilot Free (version GPT-4) experiment ($n = 16$) indicated faster task completion and improved task completeness, while also highlighting the risk of reduced algorithmic reasoning when AI suggestions are used uncritically. Overall, the findings point to the need for pedagogical approaches that emphasize guided use, verification practices, and assessment designs that more directly evidence learning in AI-mediated settings.



Academic Editors: Diego Vergara and Pablo Fernández-Arias

Received: 15 December 2025

Revised: 29 January 2026

Accepted: 3 February 2026

Published: 28 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: generative AI; ChatGPT; GitHub Copilot; educational technology; comparative study

1. Introduction

Generative artificial intelligence (GenAI) systems such as ChatGPT and GitHub Copilot are rapidly becoming embedded in academic routines, reshaping how students search for information, draft responses, and complete assignments. Unlike earlier educational technologies centered on retrieval or practice, conversational models can generate fluent,

context-aware outputs that closely resemble human-produced explanations, raising fundamental questions about learning processes, assessment practices, and the evolving role of human instruction [1–5].

A central concern in this emerging landscape is the potential gap between students' perceived learning and their actual mastery when GenAI is used to complete assignments. Because these tools can reduce the effort required to produce polished outputs, students may report high learning or productivity even when deeper processing (understanding, integrating, and applying concepts) is limited. This issue is especially salient because homework has traditionally been treated as a routine mechanism associated with learning achievement, although evidence indicates that the relationship between homework and achievement depends on factors such as time and frequency and may be non-linear [6]. When GenAI changes how homework is produced, completion or apparent quality may no longer function as a straightforward proxy for topic mastery. Consequently, beyond measuring adoption and self-reported benefits, there is a growing need to examine how students engage with AI outputs—whether they read, transform, and integrate generated content into their own reasoning—so that perceived learning can be interpreted within a clearer behavioral context [4,5].

Recent research has documented the growing adoption of ChatGPT in education, highlighting perceived benefits such as time efficiency, accessibility, and academic support, alongside concerns related to academic integrity, overreliance, and evaluation fairness [1–3,5]. Evidence focused on teachers indicates both interest in the pedagogical potential of GenAI and persistent uncertainty regarding how to regulate its use and assess learning fairly [1,2]. Prior work conducted with university students and teachers in Mexico suggests that GenAI is already embedded in everyday academic practices, while institutional guidance and shared evaluation frameworks remain limited [1,2,7–10]. These findings reinforce that educational impact cannot be understood solely through usage metrics and perceived outcomes, but also requires attention to how these tools are interpreted, negotiated, and regulated within specific educational contexts [4,5].

Educational context matters because GenAI use is shaped by institutional norms, assessment practices, and students' stage of academic development. Upper-secondary (high school) and university students encounter different task demands and expectations regarding autonomy, academic integrity, and critical evaluation of information. Teachers at each level likewise face distinct constraints and incentives when deciding whether—and how—to incorporate or restrict GenAI in classroom practice. Yet much of the available empirical evidence on GenAI in education originates from North America, Europe, or East Asia, and educational contexts in Latin America differ in meaningful ways, including variability in technological infrastructure, institutional resources, teacher training, and students' access to digital tools [4,11]. Structural inequalities, uneven connectivity, and diverse educational trajectories can shape how technologies are adopted and experienced, underscoring the need for context-sensitive research that reflects local conditions and challenges [4]. Moreover, existing studies have focused predominantly on higher education, leaving less evidence on how upper-secondary students engage with GenAI and how these practices compare with university settings in Latin America.

In addition to general-purpose tools such as ChatGPT, domain-specific systems such as GitHub Copilot introduce further complexity in technical disciplines. AI-assisted programming tools may influence problem-solving approaches, task completion time, and perceptions of competence, raising distinct questions about learning outcomes and skill development. Comparative analyses that consider both types of tools across educational levels and roles remain limited, particularly in Latin American contexts [4,5,8,9,11].

This study addresses these gaps through a comparative analysis of GenAI use among high school students, university students, and teachers in Sonora, Mexico. We combine descriptive evidence with targeted inferential analyses to (i) characterize adoption and perceived benefits and (ii) test whether perceived learning aligns with engagement behaviors during AI-assisted work. To operationalize engagement, we propose a Learning Engagement Index (LEI) that summarizes students' reported actions when using ChatGPT to complete academic tasks—reading AI outputs, modifying them, and integrating personal ideas—allowing an empirical comparison between perceived learning and engagement practices within the same sample.

This study pursues five objectives:

1. Characterize and compare the adoption of ChatGPT (and similar GenAI tools) among students and teachers across high school and university levels.
2. Perceived outcomes: Compare perceived learning (Q1) and perceived time savings among students across educational levels.
3. Help-seeking preferences: Describe students' preferred sources for resolving academic doubts (teachers vs. ChatGPT vs. search engines/other sources) and interpret reported reasons for these choices.
4. Engagement and alignment: Construct the Learning Engagement Index (LEI) from engagement behaviors during AI-assisted task completion (Q2–Q4) and test whether perceived learning (Q1) aligns with these engagement actions.
5. Complementary task-based evidence: Contextualize survey findings with a small GitHub Copilot protocol to capture task-based indicators of productivity and completeness in programming-related work [4,5,8,9,11].

Results are organized to mirror these objectives. We first present adoption patterns for students and teachers and then report perceived learning and time savings among students. We next describe help-seeking preferences and related reasoning. We then introduce and analyze the LEI and assess its association with perceived learning. Finally, we summarize the complementary evidence from the GitHub Copilot protocol and discuss implications for pedagogy, assessment, and responsible integration of GenAI in educational settings.

2. Background and Related Work

The emergence of generative artificial intelligence (AI) tools—most notably ChatGPT—has rapidly reshaped educational discourse worldwide. Since its public release in late 2022, ChatGPT has attracted sustained attention from educators, students, policymakers, and researchers, prompting a growing body of literature that seeks to understand its pedagogical potential, limitations, and risks. While early discussions were largely speculative or conceptual, recent studies increasingly adopt empirical and mixed-method approaches to examine how generative AI is being incorporated into teaching and learning practices across disciplines and educational levels.

2.1. Generative AI in Education: Opportunities and Risks

Systematic reviews and large-scale analytical studies consistently emphasize the dual nature of generative AI in education. On the one hand, large language models are recognized for their capacity to support personalized learning, provide immediate feedback, assist with content generation, and reduce instructors' workload [5,12,13]. On the other hand, these same studies warn of significant challenges related to academic integrity, over-reliance, erosion of foundational skills, and the difficulty of evaluating authentic learning when AI-generated content is involved [5,14,15]. Kasneci et al. [5] and Montenegro-Rueda et al. [13] argue that the educational value of ChatGPT depends less on the technology itself than on the pedagogical frameworks governing its use. Similarly, Memarian and

Doleck [12] note that much of the early literature relies on conceptual commentary rather than rigorous empirical evidence, calling for studies that move beyond perception toward measurable learning outcomes.

2.2. Teacher Perceptions and Pedagogical Adaptation

A substantial body of research has focused on educators' perceptions of generative AI and their readiness to integrate it into teaching practice. Early studies report a cautious but increasingly pragmatic stance among instructors. Alarcón Llontop et al. [1] document initial perceptions of university teachers, showing that while many recognize ChatGPT's usefulness for instructional preparation, they express concern about plagiarism, student dependence, and the absence of institutional guidance. These findings are reinforced by Arguello et al. [2], who report high adoption rates among higher education faculty, particularly for planning and content development, alongside persistent ethical reservations. Qualitative and mixed-method studies further highlight that teachers' responses are shaped by disciplinary culture and pedagogical identity. Mateus et al. [16], in a study spanning Mexico, Peru, and Spain, show that educators recognize the potential of generative AI to transform teaching practices but differ in how they envision its alignment with higher-order learning objectives. More recently, García Sánchez and Carrillo Guerrero [17] report largely positive evaluations of ChatGPT among health sciences faculty in Mexico, particularly regarding efficiency and practicality, while noting lower confidence in accuracy and critical thinking outcomes. Across these studies, a recurring theme is the need for professional training and clear institutional frameworks to support responsible AI integration.

2.3. Student Perceptions, Usability, and Satisfaction

Parallel research has examined students' perceptions, acceptance, and patterns of use of generative AI. Quantitative studies consistently show positive attitudes toward ChatGPT, driven by its immediacy, accessibility, and perceived usefulness [18–20]. Gil-Vera [21], using a structural equation modeling approach, demonstrates a direct and positive relationship between perceived usability and satisfaction among university students, suggesting that ease of use and functional value strongly shape students' overall appraisal of the tool. Importantly, the study emphasizes that satisfaction does not necessarily imply deep learning, underscoring the importance of pedagogical mediation and explicit guidance on responsible use. Other studies reveal a more ambivalent student stance. Chávez et al. [18] and [20] report that while students value ChatGPT as a learning aid, they also express concerns about reduced effort, superficial understanding, and dependence on AI-generated answers. The evidence indicates that student engagement with generative AI is neither uniform nor uncritical, but evolves alongside academic experience and instructional context.

Adoption rates alone provide an incomplete picture of generative AI's educational impact; what matters is how students engage with AI outputs during academic work. Prior work on GenAI in education suggests that learning benefits are more plausible when students actively evaluate, elaborate on, and integrate AI-generated content rather than reproducing it mechanically—for example, through verification, critical use, and reflective integration [5,12,13,19]. In line with this perspective, we operationalize engagement using self-reported behaviors—reading responses, editing/modifying outputs, and adding one's own ideas—so that perceived learning can be contrasted against action-oriented indicators of engagement.

2.4. Empirical and Experimental Evidence on Learning Outcomes

Beyond perceptions, a growing but still limited body of work has examined the impact of generative AI on learning outcomes. Malinka et al. [14] show that ChatGPT performs adequately on routine academic tasks but struggles with activities requiring

deeper critical reasoning, reinforcing the argument that AI should complement rather than replace cognitive effort. Díaz Vera et al. [22], in a comparative experimental study within an undergraduate computing course, report statistically significant performance gains for students who used ChatGPT during instructional activities, alongside high levels of satisfaction and perceived improvement in comprehension. Comparative evaluations of multiple generative AI models further indicate that no single system is optimal across all educational objectives. Kerimbayev et al. [23] demonstrate substantial variation among models such as GPT-4, Copilot, and others in terms of content quality, adaptability, and performance, highlighting the importance of aligning AI selection with specific pedagogical goals. These results resonate with domain-specific studies on GitHub Copilot in engineering education, which report efficiency gains and improved task completion when AI assistance is used under structured and supervised conditions.

2.5. Identified Gaps and Contribution of the Present Study

Despite the rapid expansion of this literature, several gaps remain. Most existing studies focus on a single educational level, discipline, or stakeholder group, and rely predominantly on cross-sectional designs. Comparative analyses that simultaneously examine secondary and higher education contexts, integrate both student and teacher perspectives, and account for changes over time remain scarce—particularly in Latin American settings. The present study addresses these gaps by offering a comparative, multi-level analysis of generative AI use among high school and university students and teachers in northern Mexico, complemented by experimental evidence from GitHub Copilot in engineering education. By situating local empirical findings within the broader international literature, this work contributes to a contextualized and developmentally informed understanding of how generative AI is reshaping learning practices, assessment challenges, and ethical considerations across educational stages.

3. Methodology

This study adopts a comparative and mixed-methods approach to analyze how generative AI tools, specifically ChatGPT and GitHub Copilot, are used and perceived by students and teachers across high school and university settings in Sonora, Mexico. The methodological design integrates quantitative data from large-scale surveys with qualitative insights from open-ended responses and a controlled experimental component. This structure enables both statistical comparison and interpretive understanding of how AI technologies influence learning, teaching, and cognitive engagement across different educational contexts. The section that follows describes the study's scope, participants, instruments, and analytical procedures in detail.

3.1. Design and Scope

This study employs a comparative mixed-methods design that integrates quantitative and qualitative analyses to explore how students and teachers use, perceive, and integrate generative artificial intelligence tools, particularly ChatGPT and GitHub Copilot, in educational settings across Sonora, Mexico. The research focuses on two educational levels, high school and university, encompassing institutions in Hermosillo, Ciudad Obregón and Guaymas. The high school component includes several public and private preparatory schools in Hermosillo, while the university component covers three public higher-education institutions and a private university. Although differences between public and private universities may affect access to technology and institutional policy, this study focuses on comparing educational levels and user roles—students and teachers—regarding their interaction with generative AI tools in learning and teaching. The study was conducted between

October 2024 and May 2025, covering two academic terms, and constitutes an extension of research initiated in 2024 [10,24]. It incorporates and builds upon previously reported findings on the use of generative AI tools, including GitHub Copilot, while expanding the analysis through additional survey data from students and teachers across educational levels and a broader comparative framework. The study involved a total of 1477 participants, drawn from both upper secondary (high school) and higher education contexts in the state of Sonora, Mexico. Participants included students and teachers from public and private institutions, allowing for a comparative perspective across educational levels.

3.2. Participants

At the upper secondary (high school) level, 602 students participated in the survey. These students were enrolled in ten preparatory schools located in Hermosillo and Guaymas, including both public and private institutions. The student sample comprised 52.2% women ($n = 314$), 45.7% men ($n = 275$), and 2.2% other/prefer not to say ($n = 13$). In addition, 42 high school teachers from the same institutions completed the questionnaire. Among them, 64.3% were women ($n = 27$) and 35.7% were men ($n = 15$). Most participating teachers were affiliated with subjects in the areas of science, mathematics, or technology, reflecting the disciplinary composition of the participating schools.

University Participants

At the higher education level, 765 students responded to the survey. These participants were enrolled at three public universities and one private university, located in the cities of Hermosillo, Ciudad Obregón and Guaymas. The gender distribution within this group was 59.5% women ($n = 455$), 37.1% men ($n = 284$), and 3.4% other/prefer not to say ($n = 26$). The university-level teacher sample included 68 instructors, who completed the survey electronically. Of these, 57.4% were men ($n = 39$), 41.2% were women ($n = 28$), and 1.5% other/prefer not to say ($n = 1$). The majority of university instructors were affiliated with engineering, computer science, and social sciences, aligning with the academic programs represented in the student sample. Table 1 summarizes the demographic distribution of participants across educational levels and roles.

Table 1. Demographic distribution of participants across educational levels and roles.

Group	<i>n</i>	Women, <i>n</i> (%)	Men, <i>n</i> (%)	Other/Prefer Not to Say, <i>n</i> (%)
High school students	602	314 (52.2)	275 (45.7)	13 (2.2)
University students	765	455 (59.5)	284 (37.1)	26 (3.4)
High school teachers	42	27 (64.3)	15 (35.7)	0 (0.0)
University teachers	68	28 (41.2)	39 (57.4)	1 (1.5)

Note: Percentages are calculated within each row and may not sum to 100 due to rounding. “Other/Prefer not to say” includes any gender responses different from “Women” and “Men” (e.g., non-binary, prefer not to say).

Ethical Considerations:All participation was anonymous and voluntary. Respondents were informed of the academic nature of the research and their right to withdraw at any time. No personally identifiable information was collected.

3.3. Instruments and Variables

Data were collected through four complementary instruments designed to capture quantitative patterns and qualitative perspectives on the educational use of generative artificial intelligence.

Each instrument was adapted to the characteristics of its target population—high school students, university students, teachers, and engineering students—in an experimen-

tal context and was reviewed by an interdisciplinary academic team to ensure conceptual clarity and consistency.

3.3.1. High School Student Questionnaire

A 19-item digital questionnaire was administered personally in classrooms by collaborators of the research team, using a QR code to ensure that all respondents were verified students. The survey was not accessed or distributed through external channels. The instrument combined Likert-scale items (0–10) evaluating perceived learning, trust in AI accuracy, time savings, and attitudes toward teacher replacement; multiple-choice questions on usage motives and preferred sources for resolving academic doubts; and open-ended prompts on advantages, risks, and misuse. Core variables included adoption level, learning perception, productivity effects, trust and dependence, and preference for human versus AI assistance. This mixed-response design enabled a multidimensional understanding of students' experiences with AI-assisted learning.

3.3.2. University Student Questionnaire

The university survey comprised a 28-item online instrument completed independently between October 2024 and May 2025. It extended the high school version by examining students' integration of ChatGPT into academic and personal contexts, as well as their evolving ethical awareness. The questionnaire covered frequency and types of use, perceived learning support (Likert 0–10), time efficiency, task improvement, and ethical or professional considerations, alongside open-ended reflections on motivations, limitations, and responsible use. Core variables included use frequency, perceived learning gains, efficiency metrics, motivational profiles, and ethical/professional awareness. Administering the survey in two waves enabled observation of attitudinal changes across the academic year.

3.3.3. Teacher Questionnaire (High School and University)

The teacher survey consisted of a 20-item electronic instrument assessing educators' experiences with generative AI, their classroom policies, and their own use of ChatGPT for instructional design and evaluation. It included closed-ended questions on usage frequency and perceived benefits and drawbacks, Likert-scale items (0–10) on attitudes toward AI integration and ethical understanding, and open-ended prompts regarding detection of AI-generated student work and pedagogical adjustments. Key variables included AI use in teaching tasks, attitudes toward student use, detection strategies, ethical awareness, and reported instructional adaptations. Teachers' responses provided critical insight into institutional readiness and the pedagogical challenges associated with integrating generative AI into educational practice.

3.3.4. Learning Engagement Index (LEI)

To complement self-reported learning with indicators of students' active engagement during AI-assisted work, we constructed a Learning Engagement Index (LEI) using three survey items that capture how students interact with ChatGPT-generated outputs: (Q2) whether students read AI responses rather than copying them verbatim when using the tool to complete academic tasks, (Q3) the extent to which students modify AI-generated answers, and (Q4) whether students complement AI-generated answers with their own ideas. These items were selected because they reflect engagement behaviors commonly associated with deeper processing (e.g., comprehension monitoring, elaboration, and integration) and can be operationalized consistently across educational levels within the survey design.

3.3.5. Analytic Sample and Exclusions

LEI was computed only for respondents providing valid responses to all three items (Q2–Q4). Records were excluded if any of Q2–Q4 were missing/NA, and if any of the three items indicated non-use of ChatGPT (Q2: “I do not use ChatGPT”; Q3: “I do not use it”; Q4: “I do not use it”). After applying these criteria, the LEI analytic sample retained $n = 1108$ respondents. For analyses comparing LEI with perceived learning (Q1; 0–10), only respondents with a valid numeric Q1 response were retained ($n = 1105$).

3.3.6. Coding and Normalization

Responses were coded to reflect increasing levels of engagement. For Q2 (reading AI responses), responses were scored on a 0–3 scale as follows: 3 = “I always read the full responses”, 2 = “I read part of the text before copying and pasting”, 1 = “It depends on the task; sometimes I read them and sometimes I do not”, and 0 = “I do not read the response; I only copy and paste”. For Q3 (modification of AI responses), responses were scored on a 0–4 scale as follows: 4 = “I modify almost the entire response”, 3 = “I modify a considerable portion of the response”, 2 = “I modify the response a little”, 1 = “I barely modify the response”, and 0 = “I do not modify the response”. For Q4 (integration of own ideas), responses were coded as binary: 1 = “Yes” and 0 = “No”. Each component was normalized to a 0–1 range by dividing by its maximum possible score (Q2/3, Q3/4, Q4/1).

3.3.7. Learning Engagement Index (LEI) Computation

The LEI was defined as the mean of the three normalized components, yielding a continuous score ranging from 0 to 1, where higher values indicate greater engagement with AI outputs:

$$LEI = \frac{Q2_{norm} + Q3_{norm} + Q4_{norm}}{3}, \quad LEI \in [0, 1],$$

LEI was summarized using descriptive statistics and compared across educational level and institutional sector using non-parametric tests, given the bounded nature of the index. To examine whether students’ perceived learning aligned with engagement behaviors, we estimated both bivariate associations (Spearman correlations between LEI and Q1) and adjusted linear models with Q1 as the outcome and LEI as a predictor, controlling for educational level, institutional sector, age, and survey wave where available. All analyses were interpreted as associations and not as evidence of causality.

3.3.8. Qualitative Analysis

Open-ended responses were analyzed from Q26.1 (“Justify your previous response”), an open-text follow-up to Q26 (“Would you agree to block the use of ChatGPT?”), and Q29 (“Anything else you would like to add about ChatGPT in your education”). A thematic coding approach was used to identify recurring ethical and cognitive themes (e.g., academic honesty, privacy, authorship, verification, bias, and perceived dependence). To quantify the prevalence of “perceived dependence,” we constructed a composite indicator at the respondent level: a participant was coded as mentioning dependence if the theme appeared in at least one of their open-ended responses (Q26.1 and/or Q29). The denominator for prevalence estimates included only respondents who provided at least one valid open-ended response (non-empty text) in Q26.1 and/or Q29; non-responses were excluded.

3.3.9. GitHub Copilot Experimental Protocol

A complementary experimental protocol was implemented with information systems engineering students to examine GitHub Copilot as a domain-specific AI assistant for programming. Using a controlled between-groups design, 16 participants (8 per

group) completed a two-hour Python data-processing task, with one group receiving Copilot/ChatGPT support and the other working without AI tools. Data were collected through pre-task surveys, performance metrics (completion time, feature implementation, code modularity/readability), and post-task reflections. Key variables included efficiency, task completeness, code quality, and perceived difficulty.

4. Results

This section presents and interprets the results obtained from the quantitative and qualitative analyses of students' and teachers' experiences with generative AI tools. The discussion integrates findings from high school and university participants, considering both educational levels and user roles. The results are organized around key analytical dimensions: adoption and frequency of use, perceived learning and productivity, ethical awareness, and the evolving relationship between human instruction and AI-assisted learning. The experimental data on GitHub Copilot use among information systems engineering students are incorporated to illustrate how domain-specific applications of AI contribute to technical skill development and educational performance. The results, considered jointly, provide a comprehensive view of how generative AI is reshaping educational practices, perceptions, and interactions across contexts.

4.1. Overview of Adoption and Usage Patterns

Across both educational levels, the use of generative AI tools was widespread, with adoption patterns differing between high school and university respondents. In both groups, ChatGPT emerged as the predominant tool, whereas GitHub Copilot was reported mainly by university students enrolled in computer science and engineering programs. At the high school level, 76.7% of students (462/602) reported having used an AI assistant such as ChatGPT. Their open-ended descriptions typically characterized use as occasional and task-specific, most commonly for clarifying concepts, checking writing, or obtaining support with homework. Among university students, 93.1% (712/765) reported having used an AI assistant, and many described ChatGPT as a regular or integrated part of their study routines. This difference in adoption by educational level was statistically significant ($\chi^2(1) = 72.73$, $p = 1.49 \times 10^{-17}$), with a small-to-moderate association ($\phi = 0.231$) and an estimated adoption gap of 16.3 percentage points (95% CI: 12.5–20.2) favoring university respondents. For more details, see Figure 1.

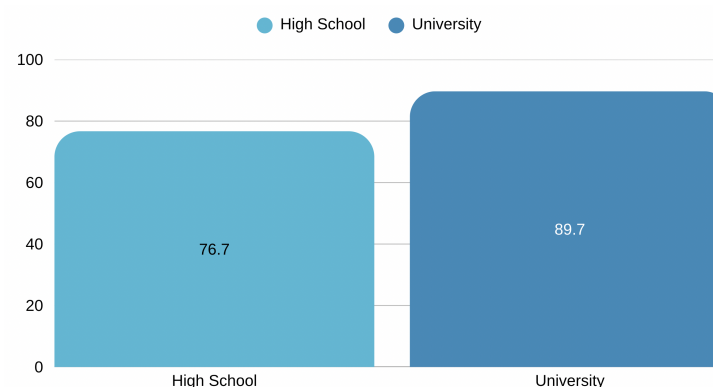


Figure 1. ChatGPT adoption by education level.

While this difference may suggest a shift from more exploratory or instrumental use in high school to more systematic and goal-oriented use in higher education, these results must be interpreted with caution. The surveys were administered at different points in time, and therefore, the contrast between educational levels cannot be attributed

solely to academic maturity or grade level. It remains possible that broader temporal trends—such as increased exposure to generative AI tools, greater normalization of their use, or evolving institutional attitudes—also contributed to the higher reported integration among university students. Consequently, the observed pattern highlights the need for longitudinal monitoring to determine whether differences in use reflect developmental trajectories, temporal effects, institutional context, or a combination of these factors.

Teacher adoption exhibited clear differences by educational level. Among high school instructors, 33.3% (14/42) reported using ChatGPT or similar tools to generate teaching materials, often citing limited time, lack of institutional guidance, or low confidence in output accuracy. In contrast, among university instructors, 69.1% (47/68) reported using generative AI for tasks such as preparing instructional materials, editing texts, translating documents, and designing exercises or assessments. This difference in adoption between educational levels was statistically significant ($\chi^2(1) = 13.46$, $p = 2.44 \times 10^{-4}$), with a moderate association ($\phi = 0.350$). The estimated adoption gap was 35.8 percentage points (95% CI: 17.8–53.8), indicating substantially higher uptake among university instructors. Nonetheless, teachers at both levels emphasized the need for clear ethical guidelines and professional training to support responsible integration. See Figure 2.

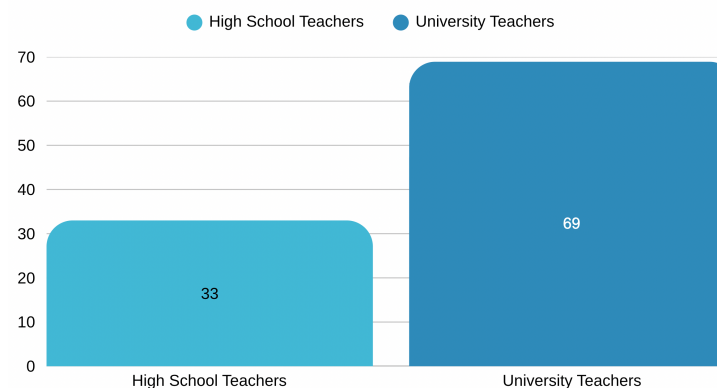


Figure 2. Percentage of teachers reporting adoption of ChatGPT or other generative AI tools (by educational level).

The results suggest that exposure and educational maturity significantly influence adoption. University students and teachers tend to integrate AI into more complex cognitive and creative tasks, while high school users—both students and educators—remain in earlier stages of exploration and adaptation. This pattern supports the notion of a developmental continuum of AI literacy, where usage evolves from pragmatic experimentation to intentional and reflective application as academic experience deepens.

4.2. Perceived Learning and Productivity

Perceptions of learning and productivity reveal both the educational potential and the cognitive risks associated with integrating generative AI into academic settings. Quantitative results indicate higher self-perceived learning gains among university students compared with high school students, alongside a progressive normalization of AI-assisted study practices. Among high school students, the average self-reported learning score on a 0–10 scale was 5.75 ($n = 584$; excluding non-responses), suggesting a moderate perceived benefit.

Qualitative responses further indicated that many students viewed the tool as a convenient aid to “understand explanations in simpler words” or to “get ideas when the topic is confusing,” rather than as a means of deep learning. Consequently, reported use tended to be instrumental, with students relying on ChatGPT primarily to complete assignments or verify answers rather than to explore concepts in depth.

University students reported a higher average perceived learning score of 6.30 ($n = 745$; excluding non-responses). This difference in educational level was statistically significant (Welch's t -test: $t = 3.38$, $p = 7.4 \times 10^{-4}$), indicating modestly greater perceived cognitive benefit among university respondents. Refer to Figure 3.

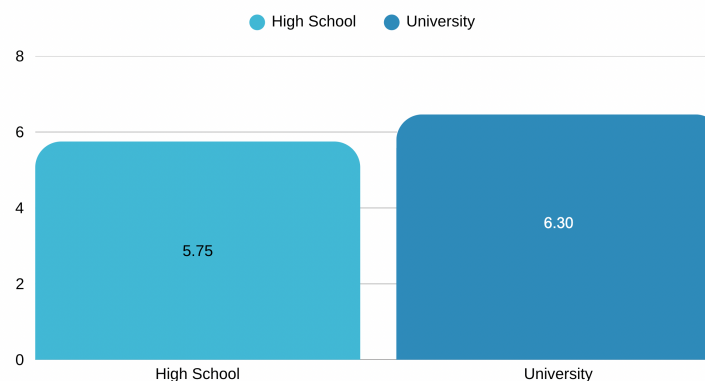


Figure 3. Perceived learning from ChatGPT by educational level (self-reported scores on a 0–10 scale).

Many described ChatGPT as a complementary learning partner that supports summarization, reformulation, and conceptual expansion of course materials. Importantly, their narratives reflect an emerging awareness of their own learning processes: students used the tool to monitor understanding, identify gaps, rehearse explanations, and reinforce knowledge. Within the context of this study, these patterns appear consistent with elements of self-regulated learning, whereby AI is incorporated into cycles of checking, adjusting, and consolidating comprehension. While these findings cannot establish a generalized shift, they indicate that at least among this group, generative AI is beginning to be used in ways that align with reflective and intentional learning strategies.

Perceived productivity followed a similar trajectory. Among high school students who reported using ChatGPT ($n = 456$; excluding non-responses and “No lo utilizo”), reported time savings per task were distributed across the full set of categories: ≤ 15 min (36.0%, 164/456), 15 min–1 h (43.2%, 197/456), 1–3 h (15.4%, 70/456), and >3 h (5.5%, 25/456). In contrast, university users ($n = 668$) more frequently reported larger time savings (1–3 h: 32.6%, 218/668; >3 h: 13.5%, 90/668), while 14.5% (97/668) reported savings under 15 min and 39.4% (263/668) reported savings of 15 min–1 h. The distribution of reported time savings categories differed significantly by educational level ($\chi^2(3) = 103.15$, $p = 3.27 \times 10^{-22}$). See Figure 4.

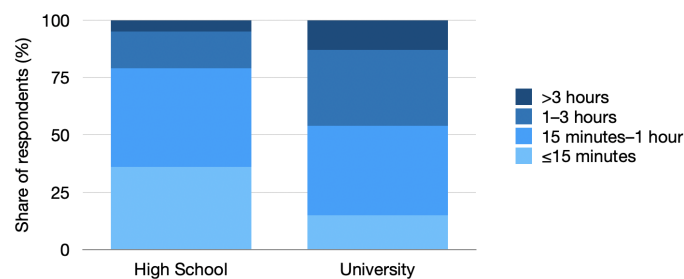


Figure 4. Perceived time savings from ChatGPT use by high school and university students.

Teacher perspectives further contextualize these patterns. Among high school instructors ($n = 42$), 33.3% (14/42) reported using ChatGPT or similar tools to generate teaching materials, mainly for idea generation or drafting written resources, although many expressed uncertainty regarding reliability and accuracy. In contrast, among university instructors ($n = 68$), 69.1% (47/68) reported employing generative AI for tasks such as

preparing teaching resources, editing or translating texts, and designing assessments. This difference in adoption by educational level was statistically significant, indicating substantially higher uptake among university instructors. While several acknowledged substantial time savings, others questioned whether increased speed necessarily translates into higher pedagogical quality or deeper student learning.

These findings suggest that perceived learning and productivity vary across educational levels and may be associated with students' increasing academic experience. In this sample, university students appear to move beyond purely procedural uses of AI and engage with it in ways that resemble more reflective and strategic learning practices, potentially supporting metacognitive monitoring and time management. However, because the surveys were administered at different moments in time and in distinct educational contexts, these patterns should not be interpreted as evidence of a developmental progression. Rather, they reflect observed tendencies within the groups studied. Similarly, teachers' adoption patterns show signs of evolving professional engagement with AI—particularly in instructional design and administrative tasks—while also revealing ongoing concerns about quality, ethics, and the authenticity of student work. These dynamics highlight emerging shifts in practice, though further longitudinal research is needed to determine whether such patterns represent broader or stable trends in educational environments.

4.3. Ethical Awareness and Perceived Dependence

Ethical awareness and perceptions of dependence are central to understanding how students and teachers navigate the role of generative AI in educational settings. Although adoption was high, participants varied widely in their capacity to evaluate the ethical, cognitive, and social implications of using tools such as ChatGPT and GitHub Copilot. The qualitative evidence suggests that ethical literacy does not arise solely from exposure to technology but is shaped by the pedagogical and institutional contexts through which learners are guided to question and interpret AI outputs. This section is based on open-ended responses from Q26.1 (a written justification prompted after the closed-ended Yes/No item Q26, "Would you agree to block the use of ChatGPT?") and Q29 ("Anything else you would like to add about ChatGPT in your education"), with themes coded at the respondent level and excluding non-responses.

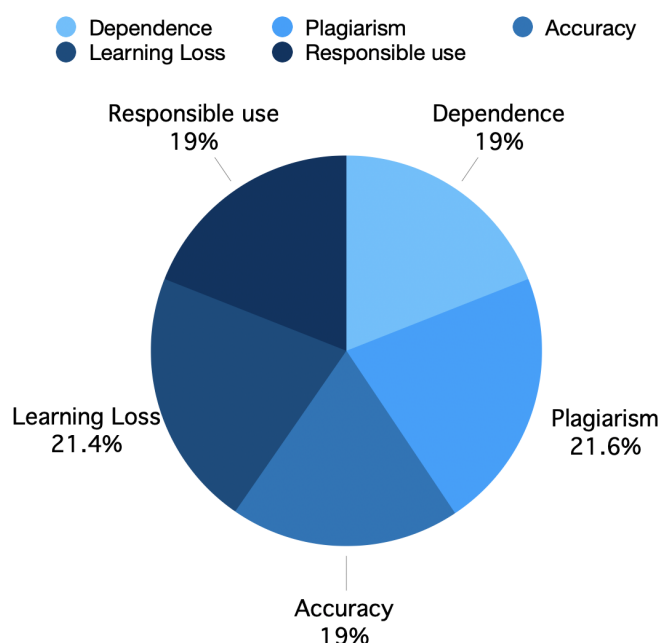
Among high school students, ethical reasoning tended to be pragmatic rather than principled. Many framed generative AI as a legitimate aid for completing assignments as long as the results appeared correct or visually acceptable, and relatively few explicitly referenced academic honesty, data privacy, or authorship. Their judgments often reflected a utilitarian logic that prioritized efficiency and correctness over process or integrity. Even so, a notable proportion of high school respondents (40.7% of those who provided at least one valid open-ended response in Q26.1 and/or Q29) raised concerns about cognitive dependence—frequently expressed as the idea that heavy reliance on ChatGPT could "make students lazy" or reduce independent learning—signaling an emerging awareness of dependence-related risks, albeit often articulated in behavioral rather than explicitly ethical terms.

Table 2 summarizes the comparative patterns in students' ethical awareness and perceived dependence across cohorts (high school 2024, university 2024, university 2025), highlighting a shift from task-oriented use toward more reflective framing and broader ethical reasoning.

Table 2. Comparison of Ethical Awareness and Perceived Dependence among Students.

Dimension	High School Students (2024)	University Students (2024)	University Students (2025)
View of AI use	Task-oriented; valued for convenience and speed	Functional; helpful for clarity and structure but requiring verification	Reflective; understood as a cognitive partner rather than a replacement
Ethical awareness	Minimal; ethicality equated with “not copying”	Emerging; focus on plagiarism and authorship	Broader ethical reasoning including fairness, data bias, and responsibility
Perceived dependence	Situational; used when teachers are unavailable	Cognitive; some report difficulty starting tasks without AI	Managed; conscious effort to balance AI use with personal reasoning
Perception of accuracy	High trust; AI perceived as infallible	Conditional trust; students verify outputs with other sources	Critical trust; recognition of factual and contextual errors
Attitude toward teachers	Teachers seen as authority and emotional support	Teachers valued for expertise but sometimes replaced for efficiency	Teachers revalued for human judgment, empathy, and contextual understanding
Main ethical tension	Shortcut vs. effort	Authorship vs. assistance	Responsibility vs. overreliance

University students’ responses generally reflected a more developed and conceptually nuanced ethical perspective. In addition to concerns commonly centered on plagiarism and authorship, many respondents introduced broader considerations such as the need for human verification, transparency regarding sources, and the possibility of biased or misleading outputs. They also emphasized that fluent or persuasive language does not guarantee accuracy, pointing to a growing recognition that responsible use requires critical evaluation of both the content produced and the conditions under which it is generated. See Figure 5.

**Figure 5.** Main ethical tensions identified by students across educational levels.

Teachers across both levels reflected these tensions but approached them differently. High school instructors emphasized concerns about assessment validity and the difficulty of identifying AI-generated work without institutional guidance or technological support. University instructors, on the other hand, adopted a more systemic perspective. They viewed AI as an inevitable component of higher education and stressed the need for curric-

ular strategies that teach students to critically evaluate AI outputs, cite them appropriately, and articulate their own reasoning alongside machine-generated content. Across both groups, the consensus was that the challenge is not to prohibit AI, but to adapt pedagogy so that ethical and reflective use becomes an integral part of learning.

Ethical Knowledge Models

The comparative evidence points to a gradual maturation of ethical reasoning as students advance along their educational pathways and gain more experience with AI-mediated learning. Quantitatively, self-reported ethical knowledge (Q27; ordered: No < A little < Yes; excluding non-responses) differed by educational level ($\chi^2(2) = 21.12$, $p = 2.59 \times 10^{-5}$), Cramér's $V = 0.127$). University students were more likely to report full ethical knowledge (Yes: 47.1%, 348/739) than high school students (Yes: 38.5%, 219/569), whereas high school respondents more frequently selected "A little" (33.4%, 190/569 vs. 22.2%, 164/739). An ordered logit model (proportional odds) estimated a positive association between university level and higher ethical-knowledge categories (OR = 1.17, 95% CI: 0.96–1.43), although this effect was modest and did not reach conventional statistical significance ($p = 0.129$).

Consistent with these quantitative patterns, the qualitative coding of open-ended responses (using the same thematic analysis procedure applied across the study, with three human reviewers and consensus resolution) suggests a shift from an instrumental orientation—centered on convenience, efficiency, and task completion—to a more reflective stance in which students increasingly question the limits of AI outputs and the role of personal judgment in evaluating them. High school narratives tended to emphasize compliance with visible rules (e.g., avoiding direct copying) rather than deeper concerns about learning processes or knowledge construction, whereas university responses more often foregrounded responsibility, transparency, fairness, and the need for verification. These findings support the interpretation of ethical and cognitive development around generative AI as an educational continuum with distinct implications for early-stage pedagogy.

At the same time, this emerging sophistication also brings its own challenges. While early users tend to rely on AI to fill specific informational gaps, more advanced users may risk deferring the start of their own thinking—turning to ChatGPT before attempting to engage with a task independently. This pattern, which could be described as prompt-dependent learning, highlights an important pedagogical concern: how to cultivate autonomy, curiosity, and persistence in a context where instant answers are always available. The implications are particularly urgent for the earliest educational stages. If the capacity for ethical discernment and critical evaluation develops gradually, it becomes essential to introduce discussions of AI use from the primary and lower-secondary levels, not merely at the university level. Doing so would require that curricula and textbooks include content on responsible AI engagement, authorship, data transparency, and the social implications of automation. Classroom activities should incorporate AI tools with guidance and awareness, enabling students to experiment safely while understanding both potential and limitation.

Equally critical is addressing the structural inequality that shapes access to these technologies. In many vulnerable populations, limited connectivity, resource scarcity, and digital illiteracy exacerbate educational disparities. Without targeted policies for equitable access and teacher training, AI integration risks reinforcing existing hierarchies rather than democratizing knowledge.

Ethical and cognitive development around generative AI must be treated as an educational continuum, beginning in early schooling and extending through higher education. The challenge for institutions and policymakers lies not only in regulating use but in culti-

vating critical, inclusive, and ethically conscious digital citizens capable of engaging with emerging technologies responsibly and reflectively.

4.4. Trust and Teacher–AI Preference

Patterns of help-seeking among students reveal a mixed support ecology in which teachers, ChatGPT, and Google coexist as complementary sources of academic support. Rather than a simple substitution dynamic, students appear to evaluate these options based on emotional comfort, epistemic expectations, and practical constraints. In high school, teachers remained the most frequently preferred source of clarification (45.2%, 270/597; excluding non-responses), followed by Google (25.6%, 153/597) and ChatGPT (21.8%, 130/597); the remaining responses were distributed across other options (7.4%, 44/597). At the university level, teachers were likewise the most frequently preferred source (46.4%, 353/761), while a modest shift toward ChatGPT was observed (25.1%, 191/761), and Google was selected by 21.2% (161/761); other options accounted for 7.4% (56/761). When comparing the distribution of the three main categories (Teacher vs. ChatGPT vs. Google; excluding “Other”), differences by educational level were small and not statistically significant ($\chi^2(2) = 4.55, p = 0.103$; Cramér’s $V = 0.060$), indicating broadly similar preference structures across levels (See Figure 6).

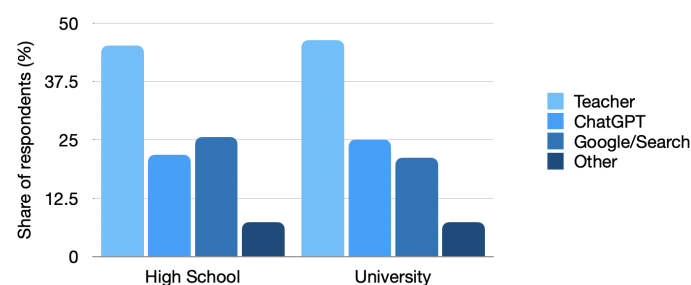


Figure 6. Preferred sources for resolving academic doubts by High School and university students.

Within the university subsample, preferences were also stable across survey years. Comparing 2024 ($n = 303$ valid responses) and 2025 ($n = 458$), teachers remained the most frequently preferred source in both cohorts (46.5% vs. 46.3%). Preference for ChatGPT increased slightly in 2025 (23.4% vs. 26.2%), while preference for Google decreased marginally (21.8% vs. 20.7%); “other” responses remained infrequent (8.3% vs. 6.8%). However, the overall distribution did not differ significantly by year ($\chi^2(3) = 1.20, p = 0.754$; Cramér’s $V = 0.040$), suggesting that year-to-year changes were modest. Qualitative analysis of students’ open-ended justifications (coded using the same thematic methodology applied throughout the study, with three independent reviewers and consensus resolution) clarifies the motives underlying these preferences. High school students who preferred teachers emphasized trust in expertise, clarity, and the ability to provide personalized, context-specific guidance; the affective dimension was prominent, with comments indicating that teachers “understand how I learn,” “give concrete examples,” or “explain in a way that feels real,” underscoring the continued importance of relational trust.

Students who preferred ChatGPT highlighted immediacy, simplicity, and the absence of judgment—particularly among those reporting shyness or discomfort asking questions in class—valuing step-by-step explanations, unlimited repetition, and a sense of psychological safety, while simultaneously expressing concerns about accuracy, logical errors, and overreliance. Google was frequently framed as a complementary option valued for breadth of sources, multimedia explanations, and the ability to contrast perspectives.

University students often described a more differentiated, situational use of resources (e.g., ChatGPT for initial structuring, Google for exploration, and teachers for validation and contextual refinement), although fully consolidated hybrid strategies were not consistently dominant. Teachers' perspectives aligned with these tensions: high school instructors emphasized concerns about accuracy, assessment integrity, and classroom dynamics, whereas university instructors framed AI tools as part of an evolving academic environment requiring new forms of pedagogical mediation. Across both levels, educators noted that their role increasingly involves helping students evaluate, verify, and contextualize AI-generated information rather than merely transmitting content.

4.5. Learning Engagement Index and Its Relationship with Perceived Learning

If perceived learning reflects substantive learning gains, it should be positively associated with students' engagement behaviors during AI-assisted work. We examine this alignment by relating perceived learning (Q1; 0–10) to the Learning Engagement Index (LEI; 0–1), which captures three engagement behaviors when students use ChatGPT to complete academic tasks: reading AI responses rather than copying them verbatim, modifying generated content, and integrating one's own ideas (Q2–Q4). The coding scheme, normalization procedure, and index definition are reported in Table 3, and the analytic sample flow is summarized in Table 4. After excluding records with missing responses on Q2–Q4 and responses indicating non-use of ChatGPT, the LEI analytic sample comprised $n = 1108$ respondents; analyses linking LEI to perceived learning further required a valid numeric Q1 response ($n = 1105$).

Table 3. Learning Engagement Index (LEI): coding and normalization scheme. LEI captures active engagement behaviors when using ChatGPT: reading AI responses (Q2), modifying AI responses (Q3), and complementing AI responses with one's own ideas (Q4). Responses indicating non-use of ChatGPT in any LEI item were excluded.

Item	Response Option (Translated)	Raw Score	Normalization
Q1. Perceived learning	Self-reported perceived learning from ChatGPT (0–10 scale)	0–10	Not part of LEI. Used to examine (Q1 vs. LEI).
Q2. Reading AI responses	I always read the full responses	3	$Q2_{norm} = Q2/3$
	I read part of the text before copying and pasting	2	
	It depends on the task, sometimes I read the responses and sometimes I don't	1	
	I don't read the response, I just copy and paste	0	
	I don't use ChatGPT	Excluded	
Q3. Modifying AI responses	I modify almost the entire response	4	$Q3_{norm} = Q3/4$
	I modify a considerable portion of the response	3	
	I modify the response a little	2	
	I barely modify the response	1	
	I don't modify the response	0	
	I don't use it	Excluded	
Q4. Complementing with own ideas	Yes	1	$Q4_{norm} = Q4$
	No	0	
	I don't use it	Excluded	

Table 4. Analytic sample selection. The LEI analytic sample excludes records with missing values in any LEI item (Q2–Q4) and excludes any record indicating non-use of ChatGPT in any LEI item.

Step	Criterion	<i>n</i> Excluded	<i>n</i> Retained
0	Total records in dataset	–	1367
1	Exclude missing/NA in any of Q2–Q4	25	1342
2	Exclude “I don’t use ChatGPT” (Q2) OR “I don’t use it” (Q3) OR “I don’t use it” (Q4)	234	1108
3	For Q1–LEI association: require numeric Q1 (0–10)	3	1105

Step 2 exclusions are not mutually exclusive across items; a respondent may indicate non-use in more than one LEI item.

Across the analytic sample, LEI values were generally high, with an overall mean of 0.829 (SD = 0.155) and a median of 0.833 (Table 5). Group comparisons indicate that LEI differs by educational level: university students displayed higher engagement than high school students (Table 6). By contrast, LEI differences by institutional sector were not statistically significant in this dataset. Wave comparisons were restricted to university respondents; within university, LEI did not differ significantly by survey wave (Table 6).

Table 5. Descriptive statistics for the Learning Engagement Index (LEI).

<i>n</i>	Mean	SD	Median	Q1	Q3	Min	Max
1108	0.829	0.155	0.833	0.771	0.917	0.111	1.000

Table 6. Learning Engagement Index (LEI) group comparisons. Group differences were tested using Mann–Whitney U tests; effect size is Cliff’s delta (δ). LEI ranges from 0 to 1 (higher values indicate greater engagement).

Comparison	Group	<i>n</i>	Mean $\pm$ SD	Median [Q1, Q3]	<i>p</i> -Value	Cliff’s δ
Educational level	High school	440	0.775 $\pm$ 0.180	0.833 [0.722, 0.917]	1.86×10^{-21}	−0.327
	University	668	0.864 $\pm$ 0.123	0.917 [0.833, 0.917]		
Institutional sector	Public	870	0.831 $\pm$ 0.152	0.917 [0.806, 0.917]	0.349	0.038
	Private	238	0.820 $\pm$ 0.163	0.833 [0.750, 0.917]		
Survey wave (university only) *	2024	257	0.866 $\pm$ 0.108	0.917 [0.833, 0.917]	0.464	−0.032
	2025	411	0.863 $\pm$ 0.132	0.917 [0.833, 0.917]		

* Wave comparison is restricted to university respondents because high school records do not contain valid timestamps in this dataset.

We then compared perceived learning (Q1; 0–10) with the LEI. Spearman correlations indicated that the association between Q1 and LEI was weak overall and not statistically significant in the pooled sample (Table 7). In high school, the Q1–LEI correlation reached statistical significance but remained very small in magnitude, while in university the correlation was non-significant (Table 7). These results suggest that students’ self-reported learning is not consistently aligned with the engagement behaviors captured by the LEI (reading, modifying, and complementing AI responses), at least as measured by these survey items.

Table 7. Correlation between perceived learning (Q1; 0–10) and learning engagement (LEI; 0–1). Spearman’s rank correlation (ρ).

Sample	<i>n</i>	Spearman ρ	<i>p</i> -Value
Overall	1105	0.052	0.0836
High school only	440	0.104	0.0295
University only	665	0.031	0.4277

To examine whether this conclusion was robust to basic adjustment, we estimated parsimonious regression models with perceived learning (0–10) as the outcome. In the pooled sample, LEI was positively but not statistically significantly associated with perceived learning after adjusting for educational level (Table 8 Model A). The same conclusion held when restricting the analysis to university respondents and additionally adjusting for survey wave (2025 vs. 2024) (Table 8, Model B). Across models, explained variance remained low, consistent with the interpretation that perceived learning reflects a subjective appraisal that may depend on factors not captured by the specific engagement behaviors operationalized in the LEI. These findings should be interpreted cautiously and do not support causal claims. The absence of a clear statistical relationship does not imply that engagement behaviors are unimportant for learning, nor does it demonstrate that students’ perceived learning is inaccurate. Rather, within the scope of this study, the results indicate no strong empirical linkage between perceived learning and the engagement actions summarized by the LEI. Establishing whether—and under what conditions—these actions translate into measurable learning gains would require more direct outcome assessments (e.g., controlled pre/post testing, performance-based evaluations, or rubric-based expert scoring), which are outside the present study’s scope. Future work will address this limitation through evaluation designs that can test learning outcomes more directly in AI-mediated contexts.

From an educational standpoint, the observed Q1-LEI mismatch has two practical implications. First, as student use of generative AI continues to expand, it becomes increasingly important to monitor not only whether students use these tools, but how they engage with them while completing academic tasks—specifically, the extent to which they read, edit, and integrate AI outputs rather than relying on superficial reproduction. Students should therefore be encouraged to recognize that perceived learning may not reliably reflect deeper cognitive processing, particularly when generative systems reduce the effort required to produce polished answers. Second, these findings reinforce the need to rethink assessment practices: as AI becomes embedded in everyday study routines, take-home assignments may no longer function as a straightforward indicator of topic mastery. Instead, the role of homework may shift toward guided practice, iterative feedback, and formative learning, while robust demonstrations of mastery may require redesigned evaluations that prioritize demonstrable understanding and make students’ reasoning processes more visible.

Table 8. Association between perceived learning (0–10) and the Learning Engagement Index (LEI; 0–1). OLS models with HC3 robust standard errors. Coefficients (β) and 95% confidence intervals are reported. Model A (pooled): Learning~LEI + educational level. Model B (university only): Learning~LEI + survey wave (2025 vs. 2024).

Predictor	Model A (Pooled)		Model B (University Only)	
	β [95% CI]	<i>p</i>	β [95% CI]	<i>p</i>
LEI (0–1)	0.690 [−0.187, 1.567]	0.123	0.659 [−0.685, 2.004]	0.337
University (vs. High school)	−0.128 [−0.393, 0.137]	0.343	—	—
Survey wave 2025 (vs. 2024)	—	—	−0.198 [−0.522, 0.125]	0.229
<i>n</i>	1105		665	
<i>R</i> ²	0.003		0.004	

Building on these implications, the next step is to examine how students navigate academic support choices in practice—particularly when they face uncertainty or conceptual difficulty. If perceived learning does not systematically track engagement actions, then

students' help-seeking preferences and the reasons underlying them become especially relevant for instructional design and policy.

4.6. GitHub Copilot Experiment

To complement the survey-based analysis of ChatGPT, a controlled experiment was conducted to examine the educational implications of GitHub Copilot, a domain-specific generative AI assistant designed to support programming tasks. This experiment provided empirical evidence of how AI-mediated assistance influences performance, learning perception, and cognitive effort among information systems engineering students. The study involved 16 participants divided equally into two groups: one using GitHub Copilot and ChatGPT as support tools, and a control group completing the same Python-based programming task without AI assistance. Both groups had comparable levels of programming proficiency, ensuring that differences in outcomes could be attributed primarily to the presence or absence of AI support. The experimental task consisted of reading, processing, and updating data from a CSV file within a two-hour timeframe, allowing measurement of time efficiency, feature implementation, and code quality.

Quantitative results demonstrated clear advantages for AI-assisted participants. The AI-supported group achieved significantly faster completion times ($p = 0.033$) and higher rates of feature implementation ($p = 0.012$) compared with the non-assisted group (see Figure 7). Furthermore, qualitative evaluation of code readability and modularity indicated higher coherence and structure among AI-assisted submissions. These results align with prior findings in software engineering research, suggesting that AI coding tools improve productivity without compromising output quality, particularly for intermediate-level programmers.

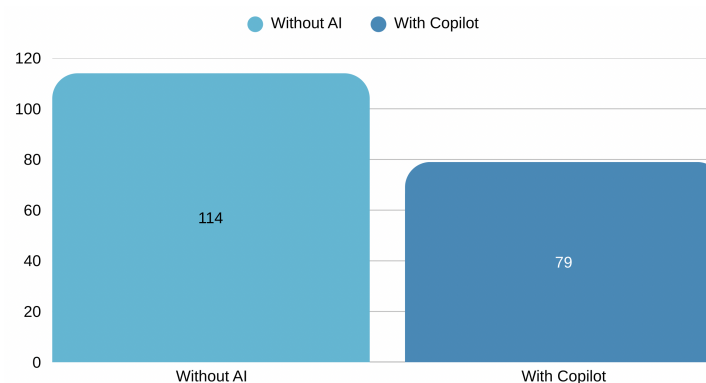


Figure 7. GitHub Copilot experiment results: Task completion time.

Post-experiment surveys revealed that AI-assisted participants reported lower frustration levels, attributing their efficiency to real-time code suggestions and problem-solving guidance. By contrast, students in the control group cited time pressure, difficulty recalling syntax, and “getting stuck without guidance” as key challenges. Importantly, however, both groups rated their learning experience similarly, indicating that while AI assistance enhanced performance, it did not necessarily replace the cognitive processes of understanding and debugging essential to programming education. See Figure 8 for more detail.

The scatter plot (Figure 9) shows the relationship between students' self-reported perceived learning (Q1; 0–10) and the Learning Engagement Index (LEI; 0–1), which summarizes three engagement actions during ChatGPT-assisted task completion (reading responses, modifying outputs, and integrating one's own ideas). The fitted trend line is nearly flat and the points are widely dispersed, indicating that higher engagement behaviors do not translate into consistently higher perceived learning scores. This visual pattern is consistent with the weak correlation and the adjusted models reported in the LEI

analyses, suggesting limited alignment between subjective learning ratings and the specific engagement actions captured by the index.

These findings illustrate that domain-specific generative AI tools can serve as effective cognitive scaffolds when used with clear pedagogical structure. Students benefit from the immediate feedback loop provided by Copilot, which supports experimentation and iterative learning.

Yet, the experiment also underscores the risk of reduced cognitive engagement: some participants reported relying on Copilot's code completions without fully analyzing the underlying logic, suggesting potential erosion of algorithmic reasoning skills if the use of AI tools is not critically mediated by instructors.

The integration of Copilot findings with the broader survey results reinforces a central insight of this study: the impact of AI on learning is context-dependent. While Chat-GPT broadens accessibility to general knowledge and conceptual explanations, Copilot demonstrates how specialized AI can enhance professional skill development in technical disciplines. In both cases, the pedagogical value of AI emerges not from automation itself but from the guided interaction between human reasoning and machine assistance.

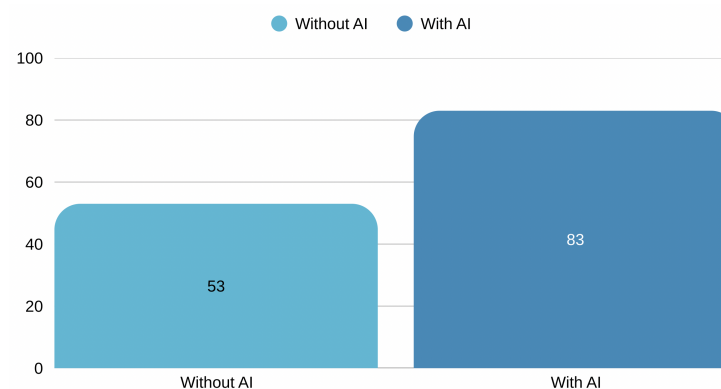


Figure 8. GitHub Copilot experiment results: Programming task completeness.

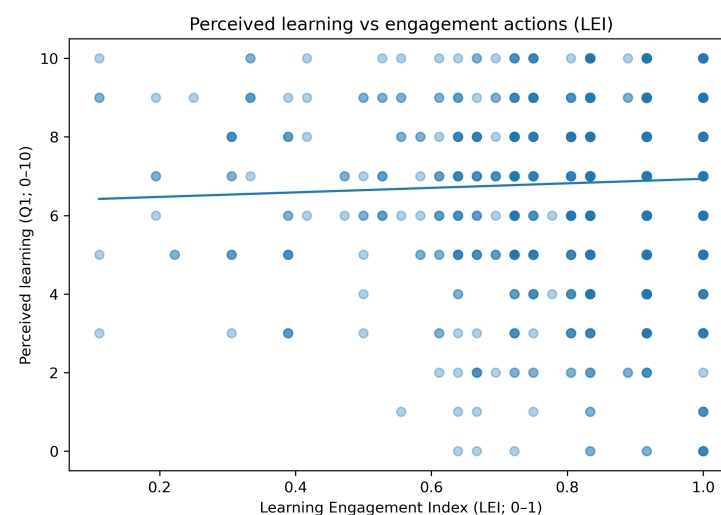


Figure 9. Scatter plot of perceived learning (Q1; 0–10) versus Learning Engagement Index (LEI; 0–1). Lines represent fitted trends with 95% confidence intervals. Because many respondents share similar Q1 and LEI values, points overlap; shading/darker regions represent higher observation density rather than a separate category.

4.7. Cross-Level Findings

The multi-layered comparative design of this study—encompassing high school students, university cohorts from two academic periods (2024 and 2025), teachers at both

levels, and an experimental evaluation of GitHub Copilot—reveals a coherent developmental trajectory in the adoption, perception, and educational integration of generative AI tools. Across all student groups, ChatGPT emerged as the predominant AI resource, yet the ways in which learners used and understood the tool varied substantially by educational stage. High school students tended to approach generative AI as a task-oriented utility, employing it primarily for completing assignments, resolving immediate doubts, or simplifying explanations. Their ethical and cognitive frameworks were correspondingly narrow, grounded in notions of convenience, correctness, and behavioral compliance. Teachers at this level expressed concern about the challenge of maintaining academic integrity in environments where students lacked the maturity to critically evaluate AI-generated content.

University students exhibited more structurally integrated patterns of use, reflecting both higher digital fluency and greater academic autonomy. In the 2024 cohort, students described ChatGPT as an efficient problem-solving companion, though their ethical reflections remained closely tied to plagiarism and authorship. By 2025, however, students articulated a more sophisticated understanding of generative AI, recognizing not only its potential but also its limitations in contextual accuracy, the importance of verification, and the value of balancing AI-generated synthesis with their own reasoning. See Figure 10.

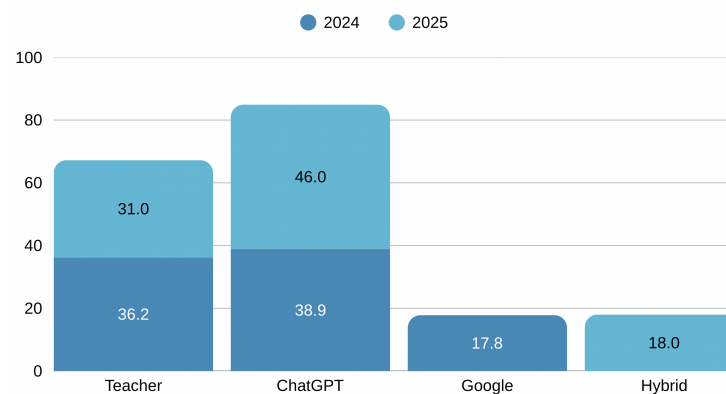


Figure 10. Preferred sources for resolving academic doubts by University students.

Teacher responses mirrored these differences. High school instructors tended to view AI as a threat to assessment validity and classroom authority, while university instructors increasingly framed AI as a pedagogical partner requiring structured integration, ethical guidance, and curricular adaptation. Across both groups, the lack of formal institutional training or policy was evident, reinforcing the need for systemic approaches to responsible AI instruction.

The GitHub Copilot experiment provided an additional layer of insight by illustrating how specialized generative tools shape learning in technically demanding domains. AI-assisted students demonstrated significantly higher productivity and code quality, yet the results also revealed the risk of cognitive dependence when AI suggestions replace rather than support algorithmic thinking. This reinforces a central theme emerging across educational levels: the pedagogical value of AI does not lie in automation alone but in the mediated interplay between human judgment and machine assistance.

The evidence presented here indicates that the integration of generative AI in education follows a developmental and contextual gradient. Younger learners benefit from the accessibility and clarity of AI tools but require stronger ethical scaffolding and teacher mediation. University students demonstrate increasing sophistication in balancing AI assistance with human expertise but face nuanced challenges related to overreliance, authorship, and critical evaluation. The Copilot experiment highlights the transformative potential

of AI in skill-based learning environments while underscoring the need for intentional, reflective, and ethically grounded instructional design.

This cross-level synthesis underscores that the educational impact of generative AI is neither uniform nor linear. Instead, it is shaped by students' cognitive maturity, disciplinary context, teacher preparation, and the structural conditions of access and inequality—factors that must be addressed collectively in policy, curriculum design, and teacher training to ensure equitable and responsible integration moving forward.

4.7.1. Equity and Variability in Access to Generative AI

Qualitative responses revealed notable variability in students' access to devices, reliable internet connectivity, and opportunities to engage with generative AI tools. These differences were not attributable to specific institutions, but rather to personal, household, and contextual conditions. Some students reported seamless access through personal devices, while others faced unstable connectivity or limited availability outside the classroom.

Such disparities influence not only how frequently students are able to use generative AI, but also their opportunities to gain experience in using these tools to streamline academic processes and improve the quality of their work—skills that are increasingly essential today and likely to remain so in the future. At the same time, the findings underscore the importance of balancing AI-assisted work with the development of problem-solving abilities, critical thinking, and the acquisition and mastery of disciplinary knowledge. Without coordinated institutional support, teacher preparation, and inclusive policies, uneven access risks reinforcing existing educational inequalities and limiting students' future competitiveness in academic and professional contexts.

4.7.2. Assessment, Fairness, and Responsible Use

A central concern expressed by teachers relates to assessment and fairness in AI-mediated learning environments. Educators questioned how to evaluate student performance when some learners use generative AI to produce high-quality work while others do not have access to, or choose not to use, such tools. This raises unresolved questions about what should be assessed in contemporary education: the final product, the effort invested, the knowledge acquired, or the responsible use of generative AI to generate solutions. Knowledge acquisition, however, is not optional; without it, students lack the capacity to judge the validity, feasibility, or correctness of AI-generated outputs, placing core educational objectives at risk.

Importantly, the present study captures perceptions and practices, but does not directly measure learning outcomes. Determining whether generative AI ultimately enhances or hinders learning requires systematic evaluation of knowledge acquisition, skill development, and long-term cognitive effects—an issue repeatedly emphasized by participating teachers. These challenges highlight an ethical dilemma that goes beyond grading equity and directly shapes how academic standards, competencies, and preparedness for future work are defined.

5. Conclusions

This study provides a comparative view of how generative AI tools—particularly ChatGPT and GitHub Copilot—are being adopted, perceived, and negotiated by students and teachers across high school and university contexts in Sonora, Mexico. The findings show widespread use, alongside differences in purpose, intensity, and interpretation that vary by educational level, role, and context; these contrasts should not be interpreted deterministically as “developmental progression,” as adoption and practices are also shaped by timing, exposure, and institutional conditions. Students across both levels value generative

AI for its immediacy and academic support, while teachers emphasize both its practical utility and the challenges it introduces for assessment integrity, ethical boundaries, and the absence of clear institutional guidance. Rather than rejecting AI, many teachers describe a shift in their role toward supporting interpretation, judgment, and critical engagement with AI-generated content.

A key contribution of this revised analysis is the introduction of a Learning Engagement Index (LEI) that operationalizes engagement behaviors during AI-assisted work (reading responses, modifying outputs, and integrating one's own ideas). Comparing LEI with perceived learning (Q1) reveals a critical tension: students who report learning with ChatGPT do not necessarily report the engagement actions typically associated with deeper processing of AI-generated content. The Q1-LEI relationship is weak and not consistently significant across groups, indicating that perceived learning is not reliably aligned with the engagement practices captured by the LEI. This does not establish that students are not learning; rather, it underscores that perceived learning cannot be assumed to reflect mastery without direct outcome assessment. This perception-engagement gap has urgent implications for evaluation. If students can produce high-quality homework outputs with reduced cognitive effort, then take-home assignments may no longer serve as a straightforward proxy for topic mastery, and assessment designs must place greater emphasis on reasoning, verification, and demonstrable understanding. At the same time, disparities in access to connectivity, devices, and opportunities to develop AI-related competencies highlight the risk that generative AI may amplify existing inequalities without deliberate institutional action.

The central challenge is not whether generative AI will be used in education, but how educational systems can adapt teaching and assessment quickly enough to preserve meaningful mastery in an AI-enabled learning environment. Future work within this research line will therefore evaluate whether students' perceived learning aligns with demonstrable content mastery through more concrete assessments of knowledge and understanding.

Author Contributions: Conceptualization, R.T.-P. and F.C.-G.; Methodology, R.T.-P.; Data curation, M.d.C.H.-S., I.E.-F. and K.L.-M.; Formal analysis, R.T.-P. and F.C.-G.; Writing—original draft preparation, R.T.-P.; Writing—review and editing, R.T.-P. and F.C.-G.; Supervision, R.T.-P.; Project administration, R.T.-P.; visualization, I.E.-F. and K.L.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical review and approval were waived for this study as it was based on anonymous, non-invasive questionnaires administered for educational research purposes. No personal data were collected, participation was voluntary, and no foreseeable risks were associated with the study.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data analyzed in this study are publicly available in an open-access repository. The surveys can be found at Zenodo via the following Digital Object Identifier (DOI): <https://doi.org/10.5281/zenodo.18626799>.

Acknowledgments: During the preparation of this manuscript/study, the authors used ChatGPT 5.2 for the purposes of editing the manuscript. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
LLM	Large Language Model

References

- Alarcón Llontop, L.R.; Pasapera Ramírez, S.; Torres-Mirez, K. The ChatGPT application: Initial perceptions of university teachers. In *Proceedings of 21st LACCEI International Multi-Conference for Engineering, Education, and Technology*; LACCEI: Boca Raton, FL, USA, 2023. [\[CrossRef\]](#)
- Arguello, F.C.; Banda, J.E.M.; Chamorro, M.I.; Jiménez, A.B.; Álvarez, L. Analysis of teachers' perception on the use and impact of ChatGPT in contemporary education. *Rev. Gest. Soc. Ambient.* **2024**, *18*, e08162. [\[CrossRef\]](#)
- Bittle, S.; El-Gayar, O. Generative AI and Academic Integrity in Higher Education: A Systematic Review and Research Agenda. *Information* **2025**, *16*, 296. [\[CrossRef\]](#)
- Holmes, W.; Miao, F. *Guidance for Generative AI in Education and Research*; UNESCO Publishing: Paris, France, 2023.
- Kasneci, E.; Sessler, K.; Küchemann, S.; Bannert, M.; Dementieva, D.; Fischer, F.; Gasser, U.; Groh, G.; Günnemann, S.; Hüllermeier, E.; et al. ChatGPT for good? On opportunities and challenges of large language models for education. *Learn. Individ. Differ.* **2023**, *103*, 102274. [\[CrossRef\]](#)
- Chin, J.M.-C.; Lin, H.-C.; Chen, C.-W. Homework and learning achievements: How much homework is enough? *Educ. Stud.* **2022**, *48*, 408–423. [\[CrossRef\]](#)
- Cárdenas Gándara, T.d.J.; Cornejo Álvarez, J.F.; Frausto Loera, M.Á. Inteligencia artificial generativa en el contexto universitario: Un estudio de seguimiento sobre la percepción docente. *Transdigital* **2025**, *6*, e484. [\[CrossRef\]](#)
- Gerónimo Bautista, E.; Rodríguez Ruiz, J.G.; Ochoa Rosas, E. Artificial Intelligence (AI) in the Mexican University Context: An Exploratory Analysis from the Researcher's Context. *Cienc. Lat. Rev. Cient. Multidiscip.* **2025**, *9*, 7757–7772. [\[CrossRef\]](#)
- Jácome Onofre, P.; Orona Hinojos, J.M.; Reyes Morales, G.; Macias Ávila, E.; Flores Leal, A.; Alvarez Jimenez, B. Educational Innovation in Mexico: Enhancing Learning with Artificial Intelligence, Applying ChatGPT in Higher Education Institutions, Professor-Student Collaboration. *Cienc. Lat. Rev. Cient. Multidiscip.* **2024**, *8*, 6210–6228. [\[CrossRef\]](#)
- Torres-Peralta, R.; Cirett-Galán, F.; Lerma-Molina, K.; Espinoza-Flores, I. Análisis comparativo del uso de herramientas de IA generativas como ChatGPT entre docentes y estudiantes universitarios. In *Proyecto V Libro Digital Investigación: Inteligencia Artificial en la Docencia y la Investigación*, 1st ed.; Aguirre-Boza, Á., Giménez-Castillo, B., Eds.; Zenodo: Geneva, Switzerland, 2025; Volume 5, pp. 154–175. [\[CrossRef\]](#)
- Cantú-Ortiz, F.J.; Sánchez, N.G.; Garrido, L.; Terashima-Marin, H.; Brena, R.F. An Artificial Intelligence Educational Strategy for the Digital Transformation. *Int. J. Interact. Des. Manuf. (IJIDeM)* **2020**, *14*, 1195–1209. [\[CrossRef\]](#)
- Memarian, B.; Doleck, T. ChatGPT in education: Methods, potentials and limitations. *Comput. Hum. Behav. Artif. Hum.* **2023**, *1*, 100022. [\[CrossRef\]](#)
- Montenegro-Rueda, M.; Fernández-Cerero, J.; Fernández-Batanero, J.M.; López-Meneses, E. Impact of the implementation of ChatGPT in education: A systematic review. *Computers* **2023**, *12*, 153. [\[CrossRef\]](#)
- Malinka, K.; Peresini, M.; Firc, A.; Hujnák, O.; Janus, F. On the educational impact of ChatGPT. In *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1*; ACM Digital Library: New York, NY, USA, 2023; pp. 47–53. [\[CrossRef\]](#)
- Rahman, M.M.; Watanobe, Y. ChatGPT for education and research: Opportunities, threats, and strategies. *Appl. Sci.* **2023**, *13*, 5783. [\[CrossRef\]](#)
- Mateus, J.C.; Cappello, G.; Lugo, N.; Guerrero-Pico, M. Communication educators facing the arrival of generative artificial intelligence. *Digit. Educ. Rev.* **2024**, *45*, 106–114. [\[CrossRef\]](#)
- García Sánchez, J.; Carrillo Guerrero, L. Evaluación docente del uso de ChatGPT en la enseñanza universitaria del área de la salud. *Rev. Investig. Tecnol. Inf.* **2025**, *13*, 84–95. [\[CrossRef\]](#)
- Chávez, M.Y.; Ugalde, F.; Aguilar, M.; Jiménez, S.; Mejia-Medina, D.A.; Juárez-Ramírez, R. Exploratory study in students' perception in the use of ChatGPT. In *Proceedings of the 2023 11th International Conference in Software Engineering Research and Innovation (CONISOFT)*; IEEE: Piscataway, NJ, USA, 2023; pp. 124–131. [\[CrossRef\]](#)
- Romero-Rodríguez, J.-M.; Ramírez-Montoya, M.-S.; Buenestado-Fernández, M.; Lara-Lara, F. Use of ChatGPT at university as a tool for complex thinking. *J. New Approaches Educ. Res.* **2023**, *12*, 323–339. [\[CrossRef\]](#)
- Shoufan, A. Exploring students' perceptions of ChatGPT. *IEEE Access* **2023**, *11*, 38805–38818. [\[CrossRef\]](#)
- Gil-Vera, V. Uso de ChatGPT por estudiantes universitarios: Un análisis relacional. *Form. Univ.* **2024**, *17*, 129–140. [\[CrossRef\]](#)

22. Vera, J.P.D.; Hojas, D.S.P.; Sarmiento, Z.J.F.; Ramírez, A.K.R.; Mora, D.V.M. Estudio comparativo experimental del uso de ChatGPT y su influencia en el aprendizaje. *Rev. Univ. Guayaquil* **2023**, *137*, 51–63. [[CrossRef](#)]
23. Kerimbayev, N.; Menlibay, Z.; Garvanova, M.; Djaparova, S.; Jotsov, V. A comparative analysis of generative AI models for improving learning process in higher education. In *Proceedings of the IEEE International Conference on Automatics and Informatics*; IEEE: Piscataway, NJ, USA, 2024; pp. 271–276. [[CrossRef](#)]
24. Cirett-Galán, F.; Torres-Peralta, R.; Navarro Hernández, R.F.; Ochoa Hernández, J.L.; Contreras-Rivera, S.; Estrada-Ríos, L.A.; Machado-Encinas, G. Assessing the use of GitHub Copilot on students of engineering of information systems. *Int. J. Softw. Eng. Knowl. Eng.* **2024**, *34*, 1717–1734. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.